SIMULATING POPULATION AND EMPLOYMENT CHANGE FOR U.S. METROPOLITAN AND RURAL AREAS

Peter M. Allaman

ABSTRACT

This paper reports on a computer simulation model of migration and employment change in 315 areas which together constitute the contiguous United States. In the process of constructing this model, an extensive database of 1960 and 1970 social and economic data was assembled at the county level from the Census of Population and Housing and procedures were developed for aggregating these data to more meaningful functional groupings of counties. This provided measures of levels of activities in each area. In addition to such aggregate databases, a facility was developed for working with large files of individual records from the 1/100 Public Use Sample, the one percent sample of Social Security records in the Continuous Work History Sample, and a file of 3.5 million businesses which was compiled by merging two years of data from the Dun and Bradstreet Corporation. These data were generally used to estimate the gross flows of the various actors in the system. A model of population and employment change was constructed which accepted initial stocks of people and employment in each region in 1960 or 1970. Based on characteristics of the areas, estimates of vital processes, migration, and employment change were calculated and equilibrated. The simulation moved forward on an annual basis for up to 15 years. Interactive methods were developed so that outside users -- perhaps unsophisticated in the use of computers -- could access both the model results and the databases using a simple set of commands over a time-sharing network.

INTRODUCTION

In this paper we discuss a computer simulation model of migration and employment change which was constructed at MIT by David Birch, Elizabeth Martin, and the author. In the course of the project we assembled an extensive database which permitted the analysis of both individual and aggregate behavior, developed a simulation model which made yearly, disaggregated population and employment estimates for each of 315 areas, and developed a system to permit others to access both our simulation results and historical data. What follows is an overview of the methodology, selected substantive results, the problems we faced, and the lessons we learned. Detailed descriptions of the results and procedures may be found in Birch, Allaman, Martin [5, 6, 8].

There were several reasons why we undertook the project:

- While the vital processes of births and deaths are relatively stable and predictable with simple models -- such as cohort survival techniques -- migration is perhaps the most important yet most poorly understood process of change for local areas. In order to understand a local area, it is necessary to understand the external forces which impinge upon it.
- The country is at a turning point in its development -- old trends are changing, several individual metropolitan areas are beginning to decline in size, and there is movement to less congested nonmetropolitan areas of the country. As Table 1 shows, the growth differentials which favored metropolitan counties in the 1960s have begun to favor not only adjacent nonmetropolitan counties, but also more remote and sparsely populated areas. This conclusion is also supported by examining growth rates of places arranged by distance from an SMSA or degree of rurality (see [6], pp. 138-147). In such a situation, simple models, relying on stable trends of the 1960s, may produce misleading results.
- While a number of studies and models had been produced at the time (late 1974), none seemed to use a simultaneous framework to provide detailed, disaggregated estimates of migration, population and employment change on an annual basis for all substate areas in the U.S. for short- to medium-range time horizons (3 to 15 years).
- The conjunction of the availability of large-scale files of individual records, powerful computer hardware, and software techniques for manipulating large-scale databases made it possible to look at the components of gross -- not simply net -- change on a routine and relatively inexpensive basis, and facilitated the implementation of a disaggregate model.

CATEGORIZATION

The first set of problems we faced involved choosing categorization schemes for geographic units, person and employment types. The problems were intensified because we were working with such a diverse collection of data sets. Our categories needed to be meaningful in terms of the problem, but also to span adequately the categories provided in all of the relevant data sets.

Table 1

POPULATION CHANGE BY METROPOLITAN STATUS AND DENSITY OF NONADJACENT COUNTIES

County Type	Annual Compound Growth Rate 1960-1970	1970-1974
Metropolitan Counties	1.6	.8
Adjacent Nonmetro-Counties	.7	1.5
Nonadjacent Nonmetro-Counties with pop. density per sq. mi.		
150 and over	1.0	1.1
100 to 149	.8	1.2
75 to 99	.4	1.0
50 to 74	.3	1.2
25 to 49	.2	1.2
10 to 24	-.3	1.2
Less than 10	-.2	1.5

SOURCE: Table 7-5 of [6].

GEOGRAPHY

All of the data were collected at the county level and could be aggregated to any meaningful grouping of counties. We initially used State Economic Areas (SEAs), 512 subdivisions of states for the U.S., but abandoned them, since they were too numerous for our purposes, out of date in terms of the current definitions of metropolitan areas, and forced an arbitrary adherence to state lines, which was troublesome in a large number of metropolitan areas which crossed state lines. We thus turned to the 173 functional economic areas delineated by the Bureau of Economic Analysis (BEA) as providing a geographically more meaningful scheme. BEA areas were designed to delineate "commuter sheds" around economic centers (Standard Metropolitan Areas -- SMSAs -- or large cities) and constitute a practical attempt to define functionally integrated economic areas [12] which collectively span the United States. Since we wished to capture urban-rural migration flows, and since SMSAs were popular units of analysis, we subdivided BEA areas which contained SMSAs into metropolitan and nonmetropolitan parts.

This resulted in 315 areas for the contiguous United States which we termed migration areas (MAs). It is interesting to note that one-third of the MAs cross state lines. A more extensive discussion of the rules followed for delineating the areas and a definition of each area by county is contained in [3]. A number of files of geographic mapping information were merged to develop the capability to aggregate all of the disparate sources of information, each with its own state and county numbering scheme, into the Migration Areas.

Although predictions were made at the MA level, groupings of MAs were also used in various places in the model, and in summary tabulations. The groupings used were: 1) the four census regions, as approximated by MAs, and 2) five area types, based on size and metropolitan character. They are displayed in Table 2.

Table 2

GROUPINGS OF MIGRATION AREAS

Census Regions

1. Northeast
2. Northcentral
3. South
4. West

Area Types

1. Supercity (New York, Chicago, San Francisco, Los Angeles)
2. Regional Centers (22 largest SMSA's not in 1)
3. Large Metro (SMSAs over 350,000 not in 1 and 2)
4. Small Metro (SMSAs under 350,000)
5. Rural (all nonmetropolitan MAs)

SOURCE: Pages 31-34 of [8].

POPULATION TYPES

Having settled on a geography, it was necessary to choose population and employment types. The use of disparate data sources, each with different definitions of the variables, posed problems to be overcome. In order to keep the dimensionality of the model within bounds, a parsimonious set of categories was needed which best discriminated migrants from nonmigrants. We used the 1/1000 Public Use Sample (PUS) for 1970 to determine the best set of categories for this use.

Age, sex and race are frequently used variables in migration analysis. As is well known, interstate migration rates rise rapidly after high school graduation to a peak in the mid-20s, then decrease until a slight bump occurs around retirement age. The task was to choose a set of age breakpoints which best captured the variability in these rates in a small number of categories and was consistent with the differing breakpoints used in the 1960 and 1970 census data. Blacks tend to have lower rates than nonblacks for all ages, but the shape of the curve is generally the same.

Based on the analysis of PUS data, however, it appeared that education better discriminated migrants from nonmigrants than did sex (see [8], pp. 6-11). While census data, used for the calculation of levels of activity, contained education, the social security data, used for the calculation of flows, did not. We thus created an age/income proxy for education which served fairly well. This proxy was developed by analyzing the 1970 PUS and setting income breakpoints for each age group which best captured the underlying educational distribution. This mapping is shown in Table 3; see [7] for greater detail. There were thus 30 population types, the Cartesian product of the five age, two race, and three education categories, as given in Table 4.

Table 3

STATUS VARIABLE ON A ONE-TO-THREE SCALE DEFINED BY AGE AND INCOME

Earnings ($1000)	Age 16-24	25-34	35-64	65 +
0-4	1	1	1	1
4-8	2	2	2	2
8-10	3	2	2	2
10-14	3	3	3	2
14+	3	3	3	3

SOURCE: Table 6 of [7].

Table 4

POPULATION CATEGORIES EMPLOYED

Age

Level	Years
1.	0-14
2.	15-24
3.	25-34
4.	35-64
5.	65+

Racial Status

Level	Race
1.	Nonblack
2.	Black

Educational Status

Level	Years Completed Equivalent
1.	<12
2.	12-15
3.	16+

SOURCE: Page 20 of [6].

EMPLOYMENT

On the employment side, we avoided elaborate industrial classification schemes. The work of Bergsman [10], which grouped the Standard Industrial Classification (SIC), a product-oriented classification, by employment location patterns revealed that two-thirds of all employees fell into one large undifferentiated category and that the remainder were scattered among relatively specialized industries. Based on this, we sought a simple classification based on occupational homogeneity. We again analyzed the 1970 PUS, tabulating twelve industry groups by four occupation groups. It was clear that farming, with a high proportion of laborers, was distinct from manufacturing and other industry and these distinct from trade and service. Table 5 defines the categories in terms of SIC codes. For data analysis the five categories farming, manufacturing, other industry, trade and services were used. In the model itself, the five categories were aggregated still further into three: 1) agriculture, 2) manufacturing and other industry, and 3) trade and services.

Table 5

INDUSTRIAL CLASSIFICATION EMPLOYED

Category	2-Digit SIC Codes	Line of Business
Agriculture	1-7	Farming
Manufacturing	20-39	Manufacturing
Other Industry	8-9	Forestry and Fishing
	10-14	Mining
	15-17	Construction
	40-49	Transportation and Utilities
Trade	50-59	Wholesale and Retail Trade
Service	60-67	Finance, Insurance, Real Estate
	70-89	Services
	91-97	Government
	99	Unclassified

SOURCE: Table 4-2 of [6].

DATA PROCESSING AND ANALYSIS

A considerable amount of data processing was required in order to implement the model. As mentioned, our strategy was to start from an initial state describing the stock of persons or jobs at a particular place and time, to predict gross flows based on the properties of the areas, and to use these gross flows in turn to modify the initial state of the area. Data for both gross flows and level of activity were needed.

POPULATION

On the population side this required a database which tabulated numbers of people by age, race, and education status. Unfortunately only two-way classifications of the census data were available and it was necessary to estimate the requisite three-way distribution. The approach was: 1) to tabulate the required three-way distribution for each MA from the two million records of the 1970 1/100 County Group Public Use Sample File; 2) to obtain the requisite marginal distributions (by age, race or education) from census county data for 1960 and 1970; and 3) to synthesize the results of steps 1 and 2 using the elementary but successful technique of matrix balancing [11]. Following this procedure, military, military family, and group quarters counts were removed from the tables, since in each case these persons' movements were institutionally determined and we did not wish to attempt to predict them.

Gross flow data between each pair of MAs for 1967-68, 1969-70, and 1970-71 for the 30 population types were obtained from tabulations provided us of the Continuous Work History Sample, 1 percent of the records of the workforce covered by Social Security. Table 6 gives a taste of summary results we obtained from tabulations for census regions. As can be seen, the data were revealing that, by 1967-68, the net flow of blacks into the Northeast had virtually stopped, although the same thing was not happening in the North Central or West. Migration into the South appears to substitute whites for blacks and higher educated people for lesser educated people. The West is growing across the board. While these results are not new, they confirm the work of others and hold out the promise that an accurate and up-to-date measurement system can be developed from the data source.

Table 6

RATE AND COMPOSITION BY TYPE OF MIGRATION FLOWS BY REGION (In Percent)

	Race		Education Status			
Flow	Non-Black	Black	1	2	3	Rate
			Northeast			
In	91.5	8.5	48.0	39.6	12.3	5.8
Out	91.9	8.1	47.5	39.5	13.0	6.0
			North Central			
In	92.5	7.5	52.3	37.6	10.1	7.7
Out	94.5	5.5	50.6	39.2	10.2	8.0
			South			
In	85.8	14.2	58.3	34.0	7.7	11.0
Out	83.7	16.3	59.3	33.4	7.3	10.9
			West			
In	96.0	4.0	46.2	41.7	12.1	11.7
Out	96.9	3.1	47.1	40.8	12.1	10.9
			U.S.			
In	90.8	9.2	52.3	37.6	10.1	8.7
Out	90.8	9.2	52.3	37.6	10.1	8.7

SOURCE: Adapted from Table 1 in [9].

A further finding which strongly influenced the model design is the strong degree of association between gross in- and out-migration which there is for most areas. Table 6 shows for regions what is also borne out for MAs; the distributions of in- and out-migration rates are close, and net migration is typically small, centering on zero, of course. If one plots for MAs the ratio of net migration to gross out-migration, in 40 percent of the MAs the difference is an order of magnitude and in all but about 5 percent the gross rate is at least double the net rate. This implies strong equilibrating forces -- such as effective information flows -- to keep the rates in balance. Changes in the mix of migration can cause regions to change, but the aggregate flows move together.

EMPLOYMENT

Employment levels were also taken from the 1960 and 1970 censuses of population. The employment change rate data were taken from a combination of Dun and Bradstreet data, corrected with the use of County Business Patterns data. We initially acquired the complete Dun and Bradstreet (D&B) files for the United States as they stood on December 31, 1969 and 1972. By merging the two files it was possible to define components of change by "births," "deaths," "new listings," and firms which existed in the file at both points in time (see Table 7). By retaining only numeric information such as numbers of employees, geographic codes, sales, etc., we were able to reduce the file to one-quarter of its original size; by then representing each numeric field with a bit-string large enough to accommodate only the largest possible value for that variable, we were able to reduce the file by another four-to-one ratio. Thus, 48 tapes of information were reduced to a little more than three tapes for which the 3.5 million records could be analyzed repeatedly at relatively small cost.

Errors in coding employment in the original files caused large percentage changes in some of our estimates of change; we thus had to develop editing rules to discard that .04 percent of the cases with improbable employment shifts in order to stabilize the estimates. This was accomplished by deleting cases with: 1) births, deaths, or new listings over certain absolute employment sizes; 2) certain minimum sizes but with growth over 300 percent or decline over 70 to 85 percent in the 1970-1972 period (these rules varied by SIC); 3) absolute change over 15,000; 4) employment in 1969, a power of 10 of employment in 1972 (to eliminate shifts of characters within the number fields). These rules were developed after analyzing over 10,000 cases of extreme employment change and are presented in detail in [1].

Table 8 is an example of initial tabulations produced from the data. It contains the change rates by component for all industries in metropolitan and rural MAs by regions. Several points are apparent:

- The amount of gross change (sum of the magnitude of components) is high, ranging from 35 percent in the metropolitan North Central to 57 percent in the rural West. Roughly 40 percent of the jobs in the economy were eliminated and replaced in the three-year period.
- Gross change is far larger than net change, just as it was for population migration. Clearly, there is much flux beneath the relatively smooth exterior of net change.
- When measured by the range of the rates in each column, the processes of decline are more stable than those of growth. Areas seem to lose employment at a relatively steady rate; whether they grow or decline is determined by births and expansions.
- Growth in rural areas is higher than it is in metropolitan areas, as we also saw for population growth.
- Migration rates are low. In part this is due to bias in the D&B recording system, causing our estimates of births and deaths to be somewhat higher. Partly it is due to multiplant firms "moving" by differential expansion in branches. But the most plausible explanation is that individual entrepreneurs move and start new, independent businesses [4].

Table 7

DEFINITION OF COMPONENTS OF EMPLOYMENT CHANGE

Firms Which Existed in 1969 and 1972

Same Area in Both Years	
No change	Same employment in both years.
Expansion	An increase in the number of employees.
Contraction	A decrease in the number of employees.
Different Area in 1969 and 1972	
Inmigration	The presence in 1972 of a firm in another area in 1969.
Outmigration	The presence in another area in 1972 of a firm in the area in 1969.

Firms Which Did Not Exist in Both Years

Death	The disappearance from the file of a firm with a particular DUNS number.
Birth	The appearance in the 1972 file of a firm with a new DUNS number, for which the year started was 1970-1972.
New Listing	The appearance in the 1972 file of a firm with a new DUNS number, for which the year started was not 1970-1972.

SOURCE: Page 64 of [8].

The net change which is shown in Table 8 is understated. This arises because of lags in the D&B reporting system. Deaths of firms seem to be picked up sooner than births, since the system is primarily oriented for credit reporting, and new firms are small. Death rates are understated by about one sixth or less, while birth rates can be

Table 8

AVERAGE COMPONENTS OF EMPLOYMENT CHANGE FOR METROPOLITAN AND RURAL AREAS BY REGION, 1970-72
(In Percent)

Region	Growth Processes				Decline Processes				Gross Change[1]	Net Change[2]
	Births	Expand.	Inmig.	Sub-Total	Deaths	Contrct.	Outmig.	Sub-Total		
Metropolitan										
Northeast	4.5	10.9	0.2	15.6	-12.1	-9.9	-0.3	-22.3	37.9	-6.7
North Central	4.5	10.6	0.2	15.3	-10.2	-9.3	-0.2	-19.7	35.0	-4.4
South	7.8	13.6	0.3	21.7	-11.6	-9.9	-0.2	-21.7	43.4	-0.0
West	6.7	14.2	0.3	21.2	-13.9	-10.7	-0.2	-24.8	46.0	-3.6
Rural										
Northeast	5.4	13.6	0.3	19.3	-13.2	-8.2	-0.2	-21.6	40.9	-2.3
North Central	6.6	15.9	0.3	22.8	-12.7	-8.5	-0.2	-21.4	44.2	1.4
South	7.4	15.6	0.4	23.4	-13.9	-8.0	-0.2	-22.1	45.5	1.3
West	10.4	21.6	0.3	32.3	-14.8	-9.6	-0.2	-24.6	56.9	7.7
Range[3]	5.9	11.0	0.2	17.0	4.6	2.7	0.1	5.1		

[1]Gross Change = Growth Subtotal - Decline Subtotal.

[2]Net Change = Growth Subtotal + Decline Subtotal.

[3]Range = Column Maximum - Column Minimum.

NOTE: These are raw figures, unadjusted for sampling biases, for illustration only. Birth rates tend to be two to three times higher than shown, and death rates 20 percent higher in magnitude. Net change is consequently understated.

SOURCE: Adapted from Table 3 in [2].

understated by a factor of as much as two or three in a three-year period, depending on the industry and area. Growing areas and industries suffer more from this underreporting. For this reason, we relied on the more stable estimates of net change derived from Country Business Patterns for 1970 and 1972, and used these data to adjust our birth figures prior to modeling.

Currently under development is an algorithm to correct for the sampling biases in the Dun and Bradstreet file without resorting to the use of County Business Patterns information. This algorithm essentially simulates a) births and deaths of firms in the "real world"; b) Dun and Bradstreet's sampling of this world; c) our own discrete snapshots at two- or three-year intervals. Correction factors which may be applied to observed birth and death rates are the result. However, this algorithm was not used in the results reported here.

THE SIMULATION MODEL

The model which we developed was fairly complex and implemented in FORTRAN on an IBM 370/C168. It is thus not easy to summarize the model with a few expressions in a high level simulation language or to present easily interpretable flowgraphs. We will confine ourselves to sketching the outlines of the algorithm and refer the interested reader to the fuller description as contained in [6]. The model first estimates employment change by MA, equilibrates this change, then estimates natural increase and migration, and finally equilibrates between the migration and employment estimates. It does this on a yearly basis.

EMPLOYMENT CHANGE

Before estimating employment at the MA level, we impose regional employment totals by industry by a variant of shift share methodology. Total U.S. employment, which has been growing in a linear fashion for over a decade, is extrapolated; any other method of estimating total U.S. employment could be inserted. Since the shares of employment by industry have been changing at very slow, stable rates, this change is extrapolated and applied to U.S. employment totals to arrive at industry totals. Likewise, since regional shares of each industry have been changing in a smooth manner over time, these smoothly varying shares are applied to the industry estimates to derive region by industry employment totals. The difference from year to year is net change by region and industry.

In estimating employment location, we had initially wished to implement a simultaneous equations model with income change an endogenous determinant of employment change. Due to the unavailability of satisfactory initial conditions for income, a different strategy was adopted. Average yearly employment loss rates for the 1970-72 period were predicted using equations for each industry and region which were linear in the parameters and fit using ordinary least squares. Explanatory variables were properties of each MA. Typical explanatory variables were population growth, median family income, unemployment, population density, percent clerical skills, percent with a high school education, outmigration, and level of employment by industry. Significant coefficients of determination for the rate equations ranged from .2 to .7; those for levels of losses ranged from .7 to .99.

The sum of these losses within a region, when subtracted from the exogenously determined net change, leaves the net employment gains by region. Since the employment-gain rates estimated from the Dun and Bradstreet data were much less reliable than the employment-loss data, shares of the regional employment gain were distributed to MAs instead of predicting employment gains by MA, as was done for employment losses by MA. The share of employment gains going to any MA was predicted as a function of its characteristics at the previous point in time. Typical variables used to predict shares were total population, percent high school graduates, median family income, population potential, and employment in farming, industry, and service. The estimation of these functions was accomplished using an algorithm which minimized the sum of absolute errors.

WORK FORCE EQUILIBRATION

Since the estimation was done cross-sectionally, the parameter estimates suffered from cross-section bias -- they tended to represent long-run equilibrium states more than short-run adjustments and when used in a time-series simulation tended to overpredict the change which might occur in each year. As a result, it was necessary to constrain the change which did occur; this was done by using unemployment rates to take up the slack and to pass information about unbalanced growth on to the population routines.

The job of this part of the submodel was thus to constrain the employment change within reasonable limits and to transmit information about these constraints forward to other submodels. It did this by: a) computing regional unemployment rates; b) comparing expected base unemployment at the regional level and raising or lowering the base as a result; c) comparing potential excess workforce with base unemployment at the regional level and raising or lowering the basic unemployment rate at the MA level as a result; d) normalizing total gains at the MA level to sum to regional gains; e) damping growth by industry at the MA level to avoid wide divergence of the individual industries within each MA.

The change for most (70 percent) MAs is adjusted by less than 15 percent. While there is a tendency to cut back, the net growth for a region must be completely allocated within that region, so that some regions must take up the slack and be increased. The adjustments are similar from year to year in each MA and there tends not to be much association between the size of a region and the adjustment performed.

POPULATION CHANGE AND MIGRATION

Births and deaths were estimated by applying type-specific rates derived from national averages to the population categories. Birth rates were trended downward and death rates remained constant. Nothing in the methodology precludes using MA-specific rates, although this was not done in the current implementation. The shifting among age and education categories was likewise done in a straightforward manner.

Outmigration from each area was estimated by applying an average U.S. rate of outmigration by population type as derived from the social security data and adjusted by a factor specific to each MA; this was performed for the 24 adult population categories. The migration of children was estimated as a constant proportion of the outmigration of adults.

Behavioral equations were developed to predict the inmigration of persons to MAs. This was done in two steps. First, migrants of each type chose a region and an area type within that region. Outmigrants by type from each origin MA were first allocated to destination region by a factor again derived from social security data and modified so that those with higher education migrated greater distances, older migrants sought the South, etc. This was done separately for each race. The area type sought was determined as a function of the destination region and the area type in the origin region from which the migrant came. This was also done by race and attempted to include the effect of previous experience and counterflows from one region to another.

Second, migrants chose an MA within each region and area type combination. This was done for the six education by race types, since little variation was found by age. These migrants are distributed to the MAs in each region based on the characteristics of the MA. The methodology used was the same as used for allocating shares of employment gains to MAs. Since there were six education/race types and nineteen region/area type combinations, there were 114 equations to be fit in order to accomplish this distribution. Typical properties of MAs which were used as exogenous variables were total population, college educated population, minority population, job opportunities, percent professional, percent of housing units built before 1939 and after

1959, percent of persons born in a different state, and population growth from 1960 to 1970. It is impossible to summarize the results of the fitting in a few sentences. However, the more highly educated tended to emphasize size while the less well educated emphasized job opportunities; blacks were quite sensitive to black concentrations; growing areas attracted migrants disproportionately; and people generally tended to avoid older areas of the South.

POPULATION CONSTRAINTS

As it stands at the end of the employment submodel, the unemployment rate is artificially low (in growing areas) because the workforce upon which it is based is that of the previous year. It is up to the population routines to update the workforce based upon natural increase and migration. At that point, the unemployment estimates are then back up to more realistic levels.

As in the case of employment growth, population growth is modified by a workforce constraint. In the case of population, however, the logic is reversed -- migrants are sensitive to job opportunities and unemployment rates (whereas employers were sensitive to the availability of labor). The base unemployment rates are then adjusted to reflect the difference between the expected and the actual unemployment rate were no correction made. Before the correction, however, the "pressure" of any excess inflow is remembered, and the relative magnitudes of the base unemployment rates adjusted accordingly.

The population adjustments are, like the employment adjustments, clustered around zero, consistent from year to year, bear little relation to size, and bear little relation to employment adjustments.

ACCOUNTING

After the basic business and migrant location decisions have been made, a set of accounting procedures relates events in one submodel to the others and checks on crude balances within the regions. U.S. labor force participation rates are updated as are unemployment rates.

VALIDATION AND PROJECTION

In developing the simulation, we had several measures of error. First were goodness of fit statistics for the fitted equations. For the fitting of rates of employment loss, the significant coefficients of determination ranged from about .2 to .7; when rates were integrated to levels the fits ranged from .7 to .99. The equivalent of R^2 for the shares of employment gain were from .5 to .98. Fits for the population immigration equations were high by cross-sectional fitting standards.

A second test was to simulate the system from 1960-1970. The average absolute percent errors of "prediction" for 1970 were as presented in Table 9. For population, the error rises as one descends to finer geographic or substantive detail. For employment, only the MA errors are meaningful, and are slightly higher than those for population.

Table 9

AVERAGE ABSOLUTE PERCENT ERROR
SIMULATED VERSUS ACTUALS
(In Percent)

Geographic Level	Totals 1970	Totals 1975[1]	Marginals 1970
	Population		
U.S.	.2	1.0	.8
Regions	.2	1.0	1.8
MAs	5.8	2.9	7.5
	Employment		
U.S.	[2]		.7
Regions	[2]		.7
MAs	6.5		9.9

[1]Only population totals are available for 1975.

[2]Not applicable in 1970.

SOURCE: Table 7-1 and p. 150 of [7].

Another check was to simulate from 1970 to 1975 and compare population totals with census estimates. For 1975, the MA errors are smaller than for 1970, partly because the rate data used to fit the equations were for the late 1960s to early 1970s period. The results indicated higher percent error for small areas, low estimates for retirement communities and those with high migration rates, and high estimates for some poorer rural southern areas and those with large dependence on military or governmental employment.

The simulation results to 1985 show a tendency for the trends of the 1970s to continue, although this is not uniform. Table 10 shows historical and simulated annual compound growth rates by region. Growth rates for the South and West are high, and those for the Northeast and Northcentral low. However, the growth rate trends for the former two areas seem to be leveling off in the simulation while those of the latter two are rising somewhat. At the MA level, southern growth will spread to smaller areas, many large northern metropolitan areas will experience slow growth to decline, but many small or rural areas will grow; there are exceptions to these generalizations, however. A complete set of forecasts for all areas is provided in [7].

Table 10

POPULATION ANNUAL COMPOUND GROWTH RATES HISTORICAL (1950-1975) AND SIMULATED (1975-1985) BY REGION AND U.S. (In Percent)

Time Period	North-east	North-central	South	West	U.S.
1950-1955	1.4	1.8	1.3	4.0	1.8
1955-1960	1.2	1.3	2.1	3.7	1.8
1960-1965	1.2	1.0	1.7	2.9	1.5
1965-1970	0.7	0.9	1.2	1.8	1.1
1970-1975	0.2	0.4	1.8	1.8	1.0
1975-1985	0.7	0.8	2.0	1.9	1.4

SOURCE: Derived from Tables 7-2, 7-3, and 7-8 in [7].

ACCESS

In the course of alternative model specification and scenario development, the need for a system which would help analyze the results of each simulation quickly and easily became apparent. A series of programs for displaying model results were developed. Later, outside users became interested in using both the original data, and the model results. A system called ACCESS was developed to fill this need. By utilizing a question answering format, it allows users with little computer expertise to tabulate the large and detailed unit record databases, as well as to display the historical and forecast values for MAs. The system is self-instructing, does not permit careless or very expensive mistakes on the part of the user, is inexpensive, general purpose, and accessible through a computer network. A number of users experimented with this system and provided valuable feedback. This system is described in [5], and is operational at MIT's Program on Neighborhood and Regional Change.

LESSONS LEARNED

Several lessons learned in the process of model development and use are summarized below:

- In fields such as migration analysis where the level of formal, theoretical development is low and where the existing models (such as gravity models) are essentially nonbehavioral, there is wide scope for systems which attempt adequately to measure the fundamental behavioral units and to forecast their behavior, if only on a simplified basis. An eclectic approach is an asset in such a theoretically ill-defined area.
- The choice of proper categories for use in modeling is an important first step to ensure the adequate representation of behavior with reasonable dimensionality.
- The use of disaggregate databases, which allows the definition of one's own categories -- not those chosen by others for other purposes -- is essential for both categorization and adequate behavioral modeling. Where there is no choice, ways of mapping between available categories and those desired are critical.
- Measurement systems based on the ability to perform large-scale data processing cheaply and efficiently are an important cornerstone to such an approach. Further, the data analysis techniques must be able to operate on the output of the simulation as well as the input. Few systems achieve this, and those which do are not oriented toward large-scale data processing applications.
- On a substantive level, how the equilibration of the system occurs is of paramount importance, yet very little research, or even systematic thought, has gone into this question. In both migration and employment change, the gross change which does occur is much larger (often by an order of magnitude) than the net change which occurs. In order for such a situation to occur, potent equilibrating mechanisms -- such as extensive information flows -- must be present. However, little is known about these mechanisms. A formal investigation might indeed yield important insights.
- Even if the conceptual nature of system equilibration were solved the "cross-sectioning bias" problem will necessitate constraining the simulation in some way. The tendency for simulations to seek long-run states too quickly has been noted in several other major regional models ([13], p. 229) and requires methodological attention and the development of better time-series databases.

In summary, we have attempted to document in this paper the importance of efficient computer and data processing techniques to the development of a large-scale migration and employment change model. This was true in the initial stages of variable definition, continued by providing insights which guided the model design process, and, of course, was essential in actually implementing the model.

ACKNOWLEDGEMENTS

The model described here was a collaborative effort of David Birch, Principal Investigator; Peter Allaman; and Elizabeth Martin. The research was performed initially at the Joint Center for Urban Studies of MIT and Harvard University and has continued at MIT's Program on Neighborhood and Regional Change. It was funded by the U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Contract No. H-2212R, and the Economic Development Administration of the U.S. Department of Commerce, Interagency Agreement No.

1AA-H-56-76. A grant to the author from the Ford Foundation aided the preparation of this paper. David Birch provided helpful comments, as did a referee. The statements made in this paper are the responsibility of the author, however, and do not necessarily reflect the views of any other person or organization.

BIBLIOGRAPHY

1. Allaman, Peter. "Using the Dun and Bradstreet Data for Research on Regional Economic Development." MIT-Harvard Joint Center for Urban Studies, Cambridge, Mass., 1975. Inter-Area Migration Project Working Paper No. 4.

2. Allaman, Peter, and David Birch. "Components of Employment Change for Metropolitan and Rural Areas in the United States by Industry Group, 1970-72." MIT-Harvard Joint Center for Urban Studies, Cambridge, Mass., 1975. Inter-Area Migration Project Working Paper No. 8.

3. Allaman, Peter, and David Birch. "Definition of Migration Areas Used in the Inter-Area Migration Project." MIT-Harvard Joint Center for Urban Studies, Cambridge, Mass., 1975. Inter-Area Migration Project Working Paper, No. 2.

4. Birch, David. "Regional Differences in Factor Costs: Labor, Land, Capital, and Transportation." MIT Program on Neighborhood and Regional Change, Cambridge, Mass., 1978.

5. Birch, David, Peter Allaman, and Elizabeth Martin. *ACCESS: A System for Easily Accessing and Analyzing Large Data Files*. MIT-Harvard Joint Center for Urban Studies, Cambridge, Mass., 1977.

6. Birch, David, Peter Allaman, and Elizabeth Martin. *A Model of Population and Employment Change for Metropolitan and Rural Areas in the United States*. MIT Program on Neighborhood and Regional Change, Cambridge, Mass., 1978.

7. Birch, David, Peter Allaman, and Elizabeth Martin. "Comparability of Census and Social Security Measures of Migration: A Perceptual Variable." MIT-Harvard Joint Center for Urban Studies, Cambridge, Mass., 1975. Inter-Area Migration Project Working Paper No. 1.

8. Birch, David, Peter Allaman, and Elizabeth Martin. *Data Acquisition and Reduction for a Model of Population and Employment Change for Metropolitan and Rural Areas in the United States*. MIT-Harvard Joint Center for Urban Studies, Cambridge, Mass., 1977.

9. Birch, David, Peter Allaman, and Elizabeth Martin. "Level and Composition of Migration Streams into and Out of Metropolitan and Rural Areas." MIT-Harvard Joint Center for Urban Studies, Cambridge, Mass., 1975. Inter-Area Migration Project Working Paper No. 3.

10. Bergsman, Joel, Peter Greenston, and Robert Healy. "A Classification of Economic Activities Based on Location Patterns," in *Journal of Urban Economics*, Vol. 2, No. 1, 1975.

11. McFadden, D., S. Cosslett, G. Duquay, and W. Jung. *Demographic Data for Policy Analysis*, Vol. VIII of Urban Travel Demand Forecasting Project, Phase I, Final Report Series. The Institute of Transportation Studies, University of California, Berkeley, Calif., 1977.

12. Olsen, Richard J., and G. W. Westley. *County Code Equivalents for Regional Economic Analysis*. Oak Ridge National Laboratory, Oak Ridge, Tenn., 1978. ORNL-NSF-EP-35.

13. Olsen, Richard J. "MULTIREGION: A Socioeconomic Computer Model for Labor Market Forecasting," Chapter 10, in Hansen, Niles M., ed. *Human Settlement Systems: International Perspectives on Structure, Change and Public Policy*. Ballinger, Cambridge, Mass., 1978.