Lawrence Berkeley National Laboratory

Title
LABOR MARKET PROJECTIONS MODEL (LMPM) REPORT ON SECOND YEAR ACTIVITIES

Permalink
https://escholarship.org/uc/item/7xs2q1wx

Author
Schroeder, E.

Publication Date
1977-08-01

LABOR MARKET PROJECTIONS MODEL (LMPM) REPORT ON SECOND YEAR ACTIVITIES

E. Schroeder, W. Benson, H. Holmes,
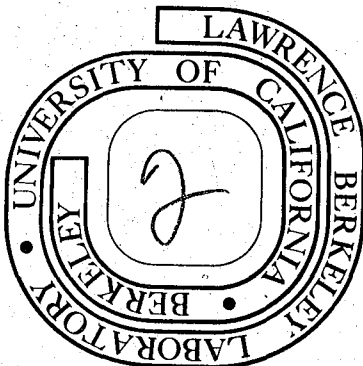F. Marchat, E. Williams, and P. Wood

August 1977

LABOR MARKET PROJECTIONS MODEL (LMPM) REPORT ON SECOND YEAR ACTIVITIES

Schroeder, E., Benson, W., Holmes, H.,
Marchat, F., Williams, E., Wood, P.

Lawrence Berkeley Laboratory
University of California
Berkeley, California

LABOR MARKET PROJECTIONS MODEL (LMPM)
REPORT ON SECOND YEAR ACTIVITIES

I.   Introduction

The affiliated State Employment Security Agencies (SESAs) of the Department of Labor - Employment and Training Administration (DOL-ETA) supply local planners in CETA projects with estimates of persons in need of manpower services. To be of maximum use, these estimates should reflect "current" local needs. This information is not used to determine funding levels, but instead enables local planners to make intelligent estimates of local need. These estimates are then used to distribute funds which have been previously allocated. The 1970 Census was an excellent source of such information when it first came out. However, with each passing year, it becomes more out-of-date and less useful.

In collaboration with DOL-ETA, Lawrence Berkeley Laboratory (LBL) has developed an estimation procedure through its Labor Market Projections Model (LMPM). The goal of the LMPM project is to provide a comprehensive modeling system to be used by local SESA analysts. LMPM is a computerized model that integrates data from a number of sources (1970 Census, CPS, migration, etc.) and for a variety of geographic areas (states, SMSAs and prime sponsors). A variety of extrapolations and predictions can be made from this data by local SESA analysts, according to their needs.

The LMPM is an ongoing project that was started at LBL in the summer of 1975. This paper will report on the current status of LMPM after two years of work and what should be done in the third year.

II. Historical Background

For the past four or five years, the Socio-Economic Environmental Demographic Information System (SEEDIS) being developed at LBL has been a valuable resource to DOL-ETA and the U.S. Energy Research and Development Administration (ERDA). (Footnote 1) By the summer of 1975, SEEDIS projects had resulted in the development of data base management techniques, manipulation of huge quantities of data, interactive retrieval of data, utilization of data gathered for one purpose for ancillary uses, digital encoding of geographical boundaries, and computer simulation and modeling.

In collaboration with DOL-ETA, LBL started to investigate and develop facilities for updating existing data (especially the 1970 Census) and for creating a computer based interactive display and mapping capability. This work was related to, and a logical extension of SEEDIS.

In the first year of the LMPM project, most of the effort at LBL was spent in updating certain tabulations of the 1970 Fourth Count Census by using the latest March Current Population Survey (CPS) tapes (1975). Updated Census estimates were computed by multiplying the 1970 Census data by the ratio of target year CPS data to 1970 CPS data and normalizing the results.

to maintain consistency. (These steps are more fully described
in Appendix I.) Essential to this computation was a careful
comparison of Census definitions with CPS definitions. Before
the CPS data could be compared with the 1970 Census, the CPS
data had to be tabulated to correspond to the Census tables of
interest. (Footnote 2.)
Census updates for these tables were done for all states, SMSAs,
and prime sponsors, as that level of geographic detail is of
most interest to local analysts. The resultant reports were
mailed to the fifty states in January 1976.

The rest of this paper will deal with the improvements and
refinements made in the second year, and with suggestions for
proposed activities during the third year. However, the accom-
plishments of the second year may be summarized as follows.
First, the Census updates via the CPS tapes were repeated with
a few new tabulations and with the geoareas involved updated
to the 1976 definitions. Second, the Guidelines for Projecting
Planning Year Supply Information and Persons in Need of Manpower
Services were computerized. This necessitated the collection
and integration of data from several diverse sources to form
a comprehensive data base. Elements from this data base were
then combined and manipulated to yield population, labor force
and unemployment projections for the desired geoareas. In
January 1977, the reports, consisting of the CPS updates and

the projections, were mailed to the fifty states.

III. Census Update Using the Current Population Survey Tapes

Relatively current socio-economic indicators at the local level were obtained by updating the 1970 census using the most recent available (March 1975) CPS tapes. The method relies upon a careful comparison of Census data categories with items in the CPS data. Once the proper correspondences are made, tabulations equivalent to the Census tables that are of interest are generated from the CPS data. By using the CPS tabulations for 1970 and for 1975 and the Census tables for 1970, it is possible to estimate a 1975 Census table by using proportions. This methodology was developed by DOL-ETA and is described in more detail in Appendix I.

During the past year a generalized tabulation routine was written. This enabled easy specification of CPS tabulations once the comparison of data items with the 1970 Census was made. (This routine is described in detail in (Reference 2). In order to limit the size of the resultant reports and to eliminate data that was not relevant to the planning of the local analysts, several Census tabulations were combined into two new tables. Thus on just one or two pages, the analyst sees the data in which s/he is interested - s/he no longer has to flip through several pages getting a few items from

each page.

IV.   Automation of the Guidelines

This past year, work was carried out to computerize parts I,
II, and III of the Guidelines for Projecting Planning Year
Supply Information and Persons in Need of Manpower Services.
(Footnote 3).   -   The purpose of the guidelines is to pro-
vide a methodology to project population, labor force, and
unemployment for each planning area for January 1978. To be
of use, the projections for each category are broken down by
sex, race, and age. The resultant figures are used by DOL-ETA
to determine the persons in need of manpower services. The
geoareas of interest to DOL-ETA are mainly states, SMSAs, and
prime sponsors. However, the procedure used is flexible
enough so that these same projections could be developed for
any county or sum of counties.

The Guidelines as developed by DOL-ETA use a refined cohort-
component method to determine population projections.   This
involves the separate projection of mortality, fertility and
net migration for five year age cohorts to give population by
age, by sex and by race for a particular area. (Footnote 4.)
Using population data from the Bureau of the Census, these projec-
tions are then normalized to add up to a control total for that
area. The component method makes explicit the assumptions

regarding the different components of population growth and can thus give considerable insight into the way the population changes. The labor force and unemployment projections are based on the population projections and national changes in the structure of the labor force and in unemployment. The estimated 1978 labor force and unemployment projections are then normalized to agree with state provided control totals.

In order to implement this model, data from several different data sources had to be collected and integrated into a comprehensive data base, GISFILE. The population on which the projections are based comes from Tabulation 17 of the 1970 Fourth Count Census. This tabulation contains 1970 population by age, by race and by sex. The mortality data used are national survival rates for 1970-1975 by age, by sex. (Footnote 5). The fertility data used are national fertility rates by age, by race. (Footnote 6 .) The migration data used are 1960-1970 net migration by age, by sex, by race at the county level. (Footnote 7 .) Thus, for each area of interest and for which Tabulation 17 of the Fourth Count Census exists and for which the 1960-1970 net migration data exists, it is possible to project a population distribution by age cohort, by sex and by race. Since the Bureau of the Census has developed projections of resident population in January 1978 at the local level (Footnote 8), these numbers can be applied

to the distributions derived above to give population by age, by race, by sex in January 1978.

The labor force composition projections and the unemployment projections utilized in the model are based on national changes in labor force and unemployment from 1970 to 1978, the labor force and unemployment situation in the local area in 1970, the local population in 1978 as previously derived, and the state provided figures for labor force participation and unemployment in the local area in 1978. In order to get data reflecting the labor force and unemployment at the local level in 1970, Tabulations 54 and 55 of the 1970 Fourth Count Census were used. An age breakdown for unemployment was possible only if the area was in a SMSA in 1970, as only Tabulation 1150 of the Sixth Count Census contains data on unemployment by age, by sex, by race.

In summary, the data sources currently being used to project population, labor force, and unemployment to January 1978 are as follows:

  1) 1970 Census, fourth count tabulations 17, 54, and 55

  2) 1970 Census, sixth count tabulation 1150

  3) 1960-1970 County Migration by race, by sex, by age

  4) Total population control estimates for January 1978, by county

5) Labor Force participation rates for January 1978

6) National Fertility Rates by age, by race

7) National Survival Rates by age, by sex

V. Future Directions

At the present time, LMPM implements two methods for obtaining socio-economic indicators at the local level - 1) updating the 1970 Census using the most recent March CPS tapes, and 2) projecting the 1970 Census based on several diverse data sources. There are several improvements and refinements that could be made to the existing models that would greatly increase their usefulness and accuracy.

Additional data sources should be included in the data base. This is probably the model's most obvious weakness at present. The existing tabulation facility should be utilized to determine whether available sample data can be used, especially the Social Security samples. The use of administrative data bases, particularly those such as ESARS which are already installed at LBL, should be investigated.

If additional data sources are included, it would be fruitful to apply more sophisticated methodologies to the data. Time

series analysis and other approaches should be studied for the CPS updates. A literature search should be initiated to determine the current status and approaches to population modeling. Based on the results of these studies, LBL would investigate alternative methodologies for population projection applied to labor force and unemployment estimates.

At the present, both models exist as batch programs. Interactive access should be provided so that the model may be accessed at any time and the values used in the model can be changed. National Office analysts could then perform sensitivity analyses by varying the data, or they could substitute more accurate data values based on local knowledge. It should also be possible to re-define interactively the various geographic areas involved.

Further, an interface to the graphic analysis system, CHART, and to the mapping display system, CARTE, should be provided. CHART combines two basic capabilities - exploratory data analysis, and data presentation in the form of matrix displays, and other graphics forms such as bar charts, pie charts, and line graphs. Most people have trouble assimilating even small amounts of data in tabular format. It takes less effort and is more effective to use familiar graphic representations. Interaction is emphasized as both exploring the data for meaningful patterns and presenting clear and convincing charts are interactive pro-

cesses. CARTE provides several types of displays based on geographic base files and geographically-coded data. CARTE contains a variety of data integration and manipulation facilities to allow the interactive analysis of geographic data by real-time displays of thematic maps. CHART facilitates quantitative comparison of variables in a variety of ways and CARTE is most useful in making comparisons across geographic areas.

VI. Conclusion

After two years of effort, Lawrence Berkeley Laboratory, in cooperation with DOL-ETA, has developed two computerized models to help supply analysts in CETA projects with estimates of persons in need of manpower services. The first model updates certain tabulations of the 1970 Fourth Count Census by using the latest CPS tapes (March 1975) and the 1970 CPS tapes. It is thus possible to update these 1970 Census tabulations to March 1975. The second model computerizes the Guidelines for Projecting Planning Year Supply Information and Persons in Need of Manpower Services, saving several man-years of tedious hand calculations. For each area of interest, this model gives population, labor force and unemployment in January 1978, by age cohort, by race, by sex. In late 1976, both models were run for all states, SMSA's and prime sponsors. The results were mailed to the State Employment Security Agencies in each of the fifty states.

FOOTNOTES

Footnote 1 -   Austin, Kranz, and Quong, <u>An Overview of the LBL</u>
<u>Socio-economic-environmental demographic informa-</u>
<u>tion System.</u> March 1975, LBL 3699

Footnote 2 -   Schroeder, Tabulation of the Current Population
Survey Tapes, Dec. 1976, UCID #3973

Footnote 3-   <u>Guidelines for Projecting Planning Years Supply</u>
<u>Information and Persons in Need of Manpower Services</u>,
Employment and Training Administration, U.S. Depart-
ment of Labor, Draft: 3/12/76

Footnote 4-   Shyrock, Henry S., Siegel, Jacob S., and Associates,
<u>The Methods and Materials of Demography</u>, U.S. Depart-
ment of Commerce, 1973

Footnote 5-   Current Population Reports, Population Estimates and
Projections Series P-25, No. 493, December 1972,
table 8-5

Footnote 6-   Vital Statistics of the United States 1971, Volume
1 - Natality; Table 1-6; U.S. Department of H.E.W.

Footnote 7-   Gladys K. Bowles, "Net Migration of the Population,
1960-70 by Age, Sex, and Color," U.S. Department
of Agriculture, Economic Research Service; Univer-
sity of Georgia, Institute for Behavioral Research;
and National Science Foundation, Research Applied
to National Needs, cooperating

Footnote 8-   Projections of Resident Population for January 1978

for the Manpower Administration, Cam Gibson, Bureau

of the Census. Projections were done for the follow-

ing geographic areas.

1) U.S. Total

2) States - all

3) Counties - all

4) For New England - all MCD's

5) For New Jersey and New York - all cities

and MCD's with 50,000+ population in 1973

and balance of county

6) For other states (42 in total) - all cities

with 50,000+ population in 1973 and balance of

county

APPENDIX I

U.S. DEPARTMENT OF LABOR
MANPOWER ADMINISTRATION*
WASHINGTON, D.C. 20210

TO: Research and Analysis Chief, State Employment Security Agency

This package provides updated information on selected socio-
economic indicators as well as base 1970 decennial census
data for these same indicators for use in developing Annual
Planning Reports (APRs), formerly Annual Manpower Planning
Reports (AMPRs). This package is part of the Guidelines
for Projecting Planning Year Supply Information being
transmitted to the State ES agencies through the ETA
regional offices.

Insert these materials in Section VI-D, the Area 1973 Updated
Indicators Package, of these guidelines. Included in this
package is a glossary of terms and methods used in the
updating, sixteen updated tabulations, and an additional
tabulation presenting 1970 unemployment information.

See Section VI of the aforementioned guidelines for a
complete description of the contents and guidelines on how
this information should be used.

*New Name: Employment and Training Administration

METHODOLOGY

This methodology was developed by the Employment and Training Administration of the United States Department of Labor. The tables are an attempt to update the 1970 Census using the 1970 and 1975 Current Population Surveys (CPS). The CPS data was tabulated to obtain tables similar to the Census tables. Some entries in the Census tables had to be combined in some cases. The method used is to treat all the tables as percents. Then the ratio of the 1975 CPS to the 1970 CPS is used as an updating factor for the 1970 Census table of percents

1975 Updated Table (approximate percents before normalization)

$$= 1970 \text{ Census (in percents)} \times \frac{1975 \text{ CPS (in percents)}}{1970 \text{ CPS (in percents)}}$$

These approximate percents are then adjusted (normalized) so that they total 100%. This computation was performed separately for each ethnic group (Total, White, Black, and Other).

EXAMPLE

This data is from Table 37 for the State of Illinois using the national CPS tabulations. Only the Total ethnic group is shown. After the CPS tabulation the following data is available:

	1970 Census (Illinois)	1970 CPS (U.S. Total)	1975 CPS (U.S. Total)
Vietnam Veteran	226,862	3,991,987	7,128,192
WW II and/or Korea	1,016,865	17,070,901	15,810,633
WW I	84,007	1,343,149	848,581
Other Service	216,365	4,117,184	4,074,484
Non-Veteran	2,064,802	37,667,074	43,237,153

The tables are then converted to percents:

	1970 Census (Illinois)	1970 CPS (U.S. Total)	1975 CPS (U.S. Total)
Vietnam Veteran	6.29	6.22	10.03
WW II and/or Korea	28.18	26.59	22.24
WW I	2.33	2.09	1.19
Other Service	6.00	6.41	5.73
Non-Veteran	57.21	58.68	60.81

The approximate percentages were then computed:

1975 Updated Table Before Normalization

Vietnam Veteran	10.13
WW II and/or Korea	23.56
WW I	1.33
Other Service	5.36
Non-Veteran	59.29
Total	99.67

Note that the converted table does not sum to 100%. Each value is therefore normalized (divide by .9967) so that the entries sum to 100%.

1975 Updated Table After Normalization

Vietnam Veteran	10.17
WW II and/or Korea	23.64
WW I	1.33
Other Service	5.37
Non-Veteran	59.49

If these percentages are used to obtain estimated counts of individuals for the particular areas, care must be exercised to use an estimate of the correct universe, e.g., Civilian Males 16 Years Old and Over, and the correct ethnic group, Total, White, Black or Other.

The 1975 Census update has been computed using national CPS data and also using a regional breakdown of CPS data. The Census has not been updated using local data, since the CPS samples are not identified by local area. The CPS regional breakdowns are given below:

ID	States
1	Connecticut
2	Maine, NH., Vt. R.I., Massachusetts
3	New York
4	New Jersey
5	Pennsylvania
6	Ohio
7	Indiana
8	Illinois
9	Michigan, Wisconsin
10	Minn., Iowa, N.D., S.D., Nebraska, Kansas, Mo.
11	District of Columbia
12	Va., W. Va., Maryland, Delaware
13	North Carolina, South Carolina, Georgia
14	Florida
15	Kentucky, Tennessee
16	Alabama, Mississippi
17	Arkansas, Louisiana, Oklahoma
18	Texas
19	Montana, Idaho, Wyoming, Colorado, New Mexico Utah, Arizona, Nevada
20	Washington, Oregon, Alaska, Hawaii
21	California

DISCUSSION

The following considerations should be kept in mind when interpreting these results. The best results would be obtained when

1) All three tables cover the same region.
2) All three tables were constructed exactly the same way.
3) No zeroes are present in the 1970 CPS table
4) No disclosure suppression exists in the 1970 Census table.
5) Close agreement occurs between the 1970 Census and the 1970 CPS percents.

(1) When the three tables do not cover the same region, the user must be aware of local trends and effects. Since the CPS data can be aggregated only to statewide levels, this will usually be the case.

(2) Differences between Census and CPS definitions are noted in the Glossary.

(3) Zeroes in the 1970 CPS table cause arithmetic problems in the ratio computation. Such cases are indicated by an asterisk in the corresponding cell of the updated table of percents. These cells were treated as zeroes during normalization.

(4) Disclosure suppressions are ignored during computation. Cells with disclosure suppression are indicated by asterisks beside the data. Because disclosure suppression affects relative percentages, other cells are also affected (but not marked).

(5) The user should ensure that the agreement between the 1970 Census and the 1970 CPS data is satisfactory.

TYPES OF INCOME

Any Other Income is composed of the CPS terms "Dividends", "Unemployment Compensation", and "Alimony" and is changed from the Census term "All Other Income".

EXPLANATION OF METHODOLOGY USED IN SPECIFIC TABLES

Tables 64, 65, 70, 81, and 90 are converted to percents using the universe in the title (by race) as a base, i.e. the sum of all lines or rows.

Tables 87, 88, 98, and 99 are converted to percents using total persons in the corresponding geographic region (by race) as a base. These figures are derived from Table 17 for the census tables.

Tables 64 and 65. Males 16 Years Old and Over by Weeks Worked in 1969 and Age.

In this table, the Census Bureau counts both civilian and military personnel in the term "Males", while the CPS counts only civilians. Therefore' this table shows only Civilian Males 16 Years and Over.

Table 81. Count of Persons 14 Years Old and Over with Income

This table is not a cross tabulation. It is assumed to mean Count of Persons 14 Years Old and Over with Some Income by Sex. For example, persons can be shown to receive one or more of the six types of income shown.

Table 98. Related Children Under 18 Below Poverty Level as Percent of All Persons.

This table has been derived from Census Table 85. Percentages have been computed using total persons in the geographic region (by race) rather than the universe in the title (by race) as a base. The base has been derived from Table 17. The updated tables of percents are normalized in two steps. First the major breakdown by age is normalized so that the percentages for the two age groups sum to the total such persons percentages. Then the breakdown by type of family is normalized so that, for example, the percents for male head/under 6 years old and female head/under 6 years old sum to percents computed in the first step.

Table 99. Persons Below Poverty Level as a Percent of all Persons

This table has been derived from Census Tables 83 and 89. The racial breakdowns for persons in families pertain to the race of

the Head of Household.   Percentages have been computed using
total persons in the geographic region (by race) rather than
the universe in the title (by race) as a base.   The base has
been derived from Table 17.   The updated tables of percents
are normalized so that the percents for persons in families
and unrelated persons below poverty level sum to the percents
for total persons below poverty level.

GLOSSARY

Generally, the terms used in this report will be those used by the Bureau of the Census in the Census Users Guide Tabulations. In some cases, however, a direct correspondence of terms with those used by the Current Population Survey (CPS) was not possible. Those exceptions are listed below.

Also listed are those tables where the concepts used in devising the Census tables are not the same as those used by the CPS, or where some explanation of the table may be needed.

OCCUPATION

Professional, technical and managerial workers is defined in the CPS to be the two categories, "Professional, Technical, etc. Workers", and Managers, Officials, Proprietors".

Craftsmen, foremen, and kindred workers is defined in the CPS to be "Craftsmen".

Farm Workers are defined by the CPS to be the two categories "Farmers and Farm Managers" and "Farm Laborers and Foremen".

Service Workers, Except Private Household is equivalent to the CPS term "All Other Service Workers".

Other is derived from the Census term "Last worked 1959 or Earlier", and the CPS term "No Previous Full Time Work Experience".

VETERAN STATUS

Korean War And/Or World War II is derived from the Census terms "Korean War", "Korean War and World War II", and "World War II", and the CPS terms "Korean War" and "World War II".

LABOR FORCE STATUS

Employed is equivalent to the CPS terms "Working", and "With a Job".

Unemployed is equivalent to the CPS term "Looking".

Not in Labor Force is equivalent to the CPS terms "Housework", "School", "Unable to Work", and "Other".

1- TERMINOLOGY

The objective of the updating procedure is to get estimates of the current values for certain tabulations of interest, according to a geographical breakdown consisting of relatively small geographical areas (preferably counties, to match the 1970 census geographical breakdown).

The values in the tabulations may be the number of individuals, or other enumerable items in a given category. A category is defined by parameters (like race, income category, occupation, age, sex, geographical area, and so on.) A k-dimensional partition is a breakdown into disjoint categories, where each category is defined by the values of k parameters. (For example, a given population could be classified according to sex only, which yields a one-dimensional partition; or according to sex and age, which would be a two-dimensional partition.)

This way of formalizing the classification treats geographical area as just another parameter, which simplifies the structure. On the other hand, parameters can always be assumed to take on a finite number of possible values (if necessary by using intervals instead of actual values).

2- THE LMPM MODEL

2-1 Description of the model (see ref. 1)

The basic assumption is that for a given category, the proportion of individuals in this category (relative to the total population) changes at the same rate from time t to time t+s for the CPS and for the census.

let $q_x(t)$ denote the proportion of individuals in category x at time t for the CPS, and let $P_x(t)$ denote the

analog for the census. Then the basic assumption can be described by the following equation

$$\frac{P_x(s+t)}{P_x(t)} = \frac{q_x(s+t)}{q_x(t)}$$

which yields a value for $P_x(s+t)$. These values are estimates for proportions, and their sum over all categories for a given population should be one. This is usually not the case, because there may be deviations from the model. An improved estimate for the census proportion for category x at time s+t is computed by dividing the $P_x(s+t)$ value above by the sum over all categories (x) of the $P_x(s+t)$. Denote this estimate by $P_x(s+t)$, then

$$P_x(s+t) = \frac{P_x(s+t)}{\sum_y P_y(s+t)}$$

Of course, the categories here are supposed to be mutually exclusive. (In some cases the data are available in such a way that this assumption is not satisfied. Then the scaling procedure is not appropriate.) On the other hand, the categories are also supposed to be exhaustive, that is, nobody is left out. (This is not satisfied when there is disclosure suppression. In this case the scaling is not appropriate either.)

2.2 Problems with the LMPM model

2.2.1 Admissibility

(An estimator is called admissible when its range is the same as the range of the random variable being estimated.)

The estimator yielding the value

$$P_x(s+t) = P_x(t) \; \frac{q_x(s+t)}{q_x(t)}$$

is already not admissible: nothing guarantees that $P_x(s+t)$ is between zero and one.

The scaled estimate $P_x(s+t)$ will be admissible. However, it is hard to tell how good an estimate it is (even when the assumption of mutually exclusive categories is satisfied), because the joint behavior of the base year census value and of the CPS ratio is not well understood and might introduce biases.

On the other hand, there is the problem of disclosure suppression, which introduces an additional perturbation.

2.2.2 Correlation between parameters

First, there is the dependence which results from the fact that each individual out of the (finite) population is put into one of a finite number of categories. The number of items in the categories are correlated. The distribution of the vector corresponding to the numbers in each category is multinomial (if we do not take any other dependence into account, see below). The correlation matrix is simple and well-known.

Second, category-defining parameters are usually not independent. For example, age and income are related. (This is one of the more complex aspects of the problem.)

2.3 Testing the model

The complex structure described above may or may not be relevant; however, it is not being brought in explicity by the

LMPM model. In any case, one should have some way of checking
how well the model fits reality. One way might be to use a test
set, for which everything is known. Then one would compare the
actual values with those predicted by the model. For example:

    a.   Using 1970 as base year, compute model predictions
        for computation year, and check those with a survey.
        (Example: the Trends Alert Survey.)

    b.   Using 1960 as base year, and 1970 as computation year,
        compare the output of the model with the 1970 census
        values.

Another way might be to use known figures or published estimates
for various errors to compute an estimate for the error on the
updated census proportion. One would take into account.

    a.   The base year CPS sampling errors

    b.   The computation year CPS sampling errors

    c.   The base year CPS/census bias

This has limitations, some of which are:

    1.   The geographical areas must coincide for Census and CPS.
        More generally, the definition of the categories must
        be the same. This implies that the CPS geographical
        areas, which are bigger, will have to be the ones used.

    2.   The bias between CPS and Census in the base year may
        include non-sampling errors in the collection of
        census data. (The CPS is designed so that it is very
        much less likely to incur this kind of error.)

    3.   This is again a method of estimating the value of
        the error, and it should also be tested for accuracy.

Some research is currently in progress at the Census Bureau to evaluate related models[b], and it might be of use for this one.   Further reference to this will be made below.

3.   ALTERNATIVE APPROACHES

Post-censal estimation has been a problem of interest for many years, and there is a large body of literature about it. A bibliography is given at the end of this report. Ways of attacking the problem usually fall into one of the following categories.

3.1 Multiple-regression type models.

For all these models, one starts from a base year and wants to obtain results for the computation year.   In updating problems, some data are available both for the base year and the computation year.   (The LMPM model is of this type.)   In prediction models, hypotheses about time evolution have to be made.   (Sometimes such hypotheses also have to be made for updating models.)   Estimates are computed by means of a regression equation which uses as predictors the available data.

Examples of such methods are:

a.   The LMPM model

b.   The ratio-correlation estimates (as described in ref. 4,5, 34)

c.   Eugene P. Ericksen's multiple regression estimates (see ref. 9, 10, 11, 23, 32)

In (a), the predictor variable is the CPS ratio, and it is used as a straight estimate of the census ratio to be predicted.

In (b) and (c), there are more than one predictor ("sympto-
matic indicators"). The regression model in (b) uses census
data (from previous censuses) to calculate the coefficients
to use on the latest census. In (c) a current sample is used
(the CPS) to compute the regression equation supposed to pre-
dict a census ratio value. This use of a sample introduces
an additional error term for sampling error in the regression
equation, and a bias for localization for areas smaller than
the CPS areas.

Both (b) and (c) have the problem that the correlation be-
tween predictors is high while the number of predictor variables
is small. Finally, it turns out that the correlation between
CPS and census observed on 1970 counts is of the same order as
the smallest of the correlations between the predictors and
the census (which would seem to say that the CPS is not a very
good census predictor itself). However, (c) turns out to be
preferable to (b). This may be related to the fact that (as
R. Royall said in ref. 32) (b) estimates an out-of-date-equation,
even though the estimates in (c) are less precise; and also
that the estimate obtained in (b) is used as one of the pre-
dictors in the (c) model. For details on evaluation and com-
parisons, see ref. 11.

3.2 Proportional allocation models.

This refers mainly to the work of M.E. Gonzalez and others
at the US Bureau of Census. This was described by Dr. Harold

Nisselson, Bureau of the Census in a letter [b] as follows
(slightly condensed).

"A common practice is to compute the (CPS) ratio for
some large level of geography, and then to apply this
ratio to update census estimates for lower levels of
geography. The problem of evaluating these estimates
is rather difficult. One method to do so is in the
form of applying the procedure to census data, or up-
dating an earlier census (say 1960) to a later one
(say 1970) and comparing to the census of that year.
The first of these is illustrated by a paper by Waksberg
and Gonzalez (1973) (see ref. 42), and also for esti-
mates of dilapidated housing with all plumbing facili-
ties. A paper by Gonzalez and Hoza (16) applies this
method to county estimates of unemployment.

A different method is that of calculating a global
measure from the (CPS) survey itself. This procedure
is the only one yielding current estimates of accuracy,
but it requires precise estimates of the variances and
the sample estimates themselves at the level of geogra-
phy of interest (see Ericksen's papers). The Gonzales-
Hoza paper also considers Ericksen-type estimators".

Essentially, in the Gonzalez paper (15) global estimates
are allocated to smaller geographical areas in proportion
to their size. This introduces two stages of estimation:

a. The size of the cell

b.   The bias introduced by proportional estimation (given

the cell size).

Assuming the answer to (a) is known, one could check whether the biases are randomly distributed around the proportional-to-cell-size estimate; this can be done for census years.   Next, if one assumes that the distribution of the bias is stationary, (b) can be obtained.

3.3   Surveys

3.3.1   The CPS

An estimate that comes to mind for the actual census proportion for a category is to use the CPS proportion for the category. This is assuming that the categories match for CPS and census. This estimate is preferable to a ratio estimate when the correlation between the CPS values for base year and computation year are fairly low (less than .5).   (Suggested by Dr. Nisselson).

3.3.2   The Trends-Alert Survey (1975)

(See ref. 45 for a detailed description.)   This was designed for the express purpose of updating the 1970 census in the Washington metropolitan area.   It was done by means of a carefully planned survey designed to match the Public Use Sample of the 1970 Census.   The results appear to be satisfactory.   For example, they showed unexpected trends in population changes, family types, etc.   They also matched closely the results obtained from other available data (for example, school enrollment). The problem with this method is its cost, which makes it feasible for limited projects only.

3.4   Other methods

3.4.1   The Census Bureau's Composite Method and Component Method II   (2, 40)

These are computational procedures using several estimates from different sources and pooling them together. They have been evaluated in comparison with Erickson's method (see ref. 11) and in general the results are not as good (except in a few cases).

3.4.2   Economic prediction models

These usually are complex prediction models designed to be used in a particular context. A relevant one is the National Planning Association model for projection of income distribution, which is related to the objective of the LMPM model development. See ref. 44 for a description of this model (and ref. 3, 12, 26 for other interesting models.)

3.4.3   Stein-James estimators

This is a new and more general approach mentioned by Dr. Nisselson. It is described in a paper by Madow and Hansen (1975) (see ref. 28). Some research in this area is being done by Madow at the Census Bureau, one objective being to develop estimates for which approximate confidence intervals can be constructed.[b]

4.   POSSIBLE OTHER DIRECTIONS

4.1   Data bases

Ericken mentions in one of his papers that one might want

to start with data that seem to be related to the problem being
treated. Here are a few, aside from automobile registrations
since this has already been mentioned.

4.1.1 The Employment Security Automated Reporting Systems

      (ESARS)

This is an "automatic" data gathering tool from the Employ-
ment Security offices located throughout the nation, otherwise
known as State Employment Offices. It describes the population
using these agencies.

A computerized version of these reports is available at LBL.
The data are published quarterly or monthly by geographical areas.
At the time of this writing the geographical areas were the
states, but a finer breakdown was being considered for implemen-
tation within a few months for at least some of the states.
(For details on ESARS see Ref. 20). These files contain infor-
mation about characteristics of the labor force using the agency
such as race, age, sex, wages, use of the job counseling, and
a lot more. These data would not be as universal as census
data as they apply to a restricted population. However, they
seem to be relevant to specific applications, and give both
overall and fairly localized information. One might be able
to use them for general purposes, if one could find out, inde-
pendently, how representative of the general population the
ESARS population is.

4.1.2 Social Security data

The Social Security Administration publishes annual files designed for public use, like the 1% continuous work-history sample (see references 22, 27, 29). This contains employment, migration, wage and age data (and possibly more) for the population under social security.

The same remark should be made as for the ESARS population (it does not include the whole US population) and one might try and get around it the same way, maybe with better success as this population is much larger.

4.1.3 Tax, finance, and other data

Sources might be

a. The Internal Revenue Service

b. The Office of Revenue Sharing

c. The Census Bureau

For details see ref 8, 21, 13

Other sources might be large data banks (like the one at Oakridge National Laboratory, see ref. 30), state and national data (drivers licenses and automobile registrations are usually computerized and are an example of this).

4.2 Methodology development

Since the problem is one that the Census Bureau has been much involved with, and the data are often Census Bureau data and in many cases confidentiality restrictions apply, it would be best to conduct this type of research in conjunction with

the Census Bureau, or at least to keep close ties with them.
The most recent development there seems to be Madow's research
on Stein-James estimators (see 3.4.3 above).

Keeping the above in mind, there is a problem which has been
mentioned only in passing and might be looked into, that is,
the time evolution of the system. This is referred to by R.
Royall in Ref. 32 for example. He suggests a probabalistic
model (Markov?) which might be an interesting approach to con-
sider.

Another approach might be a time series analysis of the
CPS data, which are gathered monthly for the basic data and
quarterly or yearly for other specific data. Such an analysis
might indicate trends, which might be applied to census data.

BIBLIOGRAPHY

1- Allmaier-Woods. CPS Updating of Poverty-Income Data (Preprint,
        9/18/74, LBL/CSAM, UC Berkeley)

2- Bogue, Donald J; Duncan, Beverely. Composite Method for Estimat-
        ing Post-Censal Population by Sex, Age and Color; Vital
        Statistics, Special Reports, Selected Studies, Vol. 47,
        No. 6, Aug. 24, 1959

3- Cohen, Malcolm S; Hill, C. Russell; Shapiro, Harold T. The LMIS
        Model: An Econometric Model of local urban labor markets;
        American Statistical Association, 1974. Proceedings of
        the Business and Economic Statistics Section, 8-17.

4- Crosetti, Albert H; Schmidt, Robert C. Accuracy of the ratio-
        correlation method for estimating post censal populations;
        Land Economics, Aug. 1954, 279-281

5- Crosetti, Albert H; Schmidt, Robert C. A Method of Estimating
        the Intercensal Population of Counties; Journal of the
        American Statistical Association, 51 (Dec. 1956), 5 87-
        590

6- Current Population Survey, Annual Demographic File 1973-1975
        Appendix A (Reprint, LBL/CSAM, UC Berkeley)

7- Dempsey, Richard E. State and Local Labor Market Manpower
        Projections Program: An evaluation of the past and a
        look ahead; American Statistical Association, 1974
        Proceedings of the Business and Economic Statistics Sec-
        tion, 18-27.

8- Eisen, M; Helfand, H., Robinson E., Sizkind, D. The Quality
        of Census Bureau Economic Time Series Indicators - Past
        Performance and Planned Improvements, American Statistical
        Association, 1975; Proceedings of the Business and Economic
        Statistics Section 190-199

9- Ericksen, Eugene P. A method for combining sample survey data
        and symptomatic indicators to obtain population estimates
        for local areas; Demography, Vol. 10, No. 2, May 1973, 137-
        160.

10- Ericksen, Eugene P. Recent developments in estimation for local
        areas; American Statistical Association, Proceedings of
        the Social Statistics Section, 1973, 37-41.

11- Ericksen, Eugene P. A regression method for estimating popula-
          tion changes of local areas; Journal of the American Sta-
          tistical Association, Volume 69, Number 348, Dec. 1974,
          867-875.

12- Folk, Hugh. The Illinois Employment Model; American Statistical
          Association, 1974; Proceedings of the Business and Economic
          Statistics Section, 1-7.

13- Fromm, Gary. Discussion on the papers of Hendry and Eisen
          American Statistical Association, 1974. Proceedings of
          the Business and Economic Statistics Section, 200-201

14- Goldberg, David; Vao, V.R.; Namboodiri, N.F. A test of the
          accuracy of ratio correlation population estimates. Land
          Economics; Feb. 1964, 100-102

15- Gonzalez, Maria Elena. Use and Evaluation of Synthetic Estimates
          American Statistical Association, Proceedings of the Social
          Statistics Section, 1973, 33-41.

16- Gonzales and Hoza. To appear in Journal of the American Statistical
          Association.

17- Gurney, Margaret and Jewett, Robert S. Constructing Orthogonal
          replications for variance estimation; Journal of the
          American Statistical Association Vol. 70, No. 352, Dec.
          1975, 819-821

18- Hansen, Morris H., Hurwitz, William N., Madow, William G.
          Sample Survey Methods and Theory, Vol. 1. John Wiley,
          New York, 1953

19- Hawkes, William J. Discussion on the papers of the Eisen and
          Hendry; American Statistical Association, 1974, Proceedings
          of the Business and Economic Statistics Section, 202-204

20- Heckman, B. and Gee, Wen-Sue. Guide to ESARS retrieval, LBL/CSAM
          UC Berkeley, March 1977

21- Hendry, M.A.; Eisen, M.; Nisselsen, H; Wikoff, J.R. An evalua-
          tion of the 1972 Economic Censuses: Looking Towards 1977.
          American Statistical Association, 1975 Proceedings of
          the Business and Economic Statistics Section, 184-189

22- Hirschberg, David A. The Continuous Work-History Sample; American
          Statistical Association, 1974 Proceedings of the Business
          and Economic Statistics Section, 54-56

23- Kaitz, Hyman B. Comments on the papers of Gonzalez and Ericksen American Statistical Association, Proceedings of the Social Statistics Section, 1973, p. 44-45.

24- Kaitz, Hyman B. Discussion on the papers of Cohen, Dempsey and Folk, American Statistical Association, 1974; Proceedings of the Business and Economic Statistics Section, 28-30

25- Keyfitz, Nathan. Estimates of Sampling variance where two units are selected from each stratum. Journal of the American Statistical Association, Dec. 1957, 503-510

26- Lin, Steven A.Y. A Synthesizied Regression for Interregional Comparisons. American Statistical Association, 1974 Proceedings of the Business and Economic Statistics Section

27- Lucas, Trudi; Brown, Lawrence. Combined comments on the papers of Hirschberg, Zeisset and Nelsen; American Statistical Association, 1974 Proceedings of the Business and Economic Section, 69-71

28- Madow, William G., Hansen, Morris H. On statistical methods of estimation in sample surveys. Paper presented at the International Association of Survey Statisticians meeting at Warsaw, 1975 (Preprint)

29- Nelson, Kathryn P. Public data files for socio-economic research: Problems and opportunities with social security sample data. American Statistical Association, 1974 Proceedings of the Business and Economic Statistics Section, 66-68

30- Loebl, A.S., Malthouse, N.S., Shonka, D.B., Ogle, M.C., Johnson, M.L. Regional Information Group (RIC) Energy, Environmental, and Socio economic Data Bases and Associated Software at Oak Ridge National Laboratory, National Technical Information Service, US Department of Commerce, 5285 Port Royal Road, Springfield, Virginia 22161

31- Robson, D.S. Application of multivariate polykays to the theory of unbiased ratio-type estimation, Journal of the American Statistical Association, Dec. 1957, 511-522

32- Royall, Richard. Discussion on the papers of Gonzalez and Ericksen, American Statistical Association, Proceedings of the Social Statistics Section, 1973, 42-43.

33- Rudolph, Steven M. Census Data for General Revenue-Sharing, American Statistical Association, 1974 Proceedings of the

34- Snow, E.C. The Application of the Method of Multiple Correlation to the Estimation of Post-censal Populations. Journal of the Royal Statistical Society, 74 (may 1911), 575-620

35- Toikka, Richard J. Discussion on the papers of Cohen, Dempsey and Folk. American

36- U.S. Bureau of the Census. An annotated list of papers and reports[a] Census Bureau Methodological Research, 1975

37- US Bureau of the Census. Concepts and methods used in manpower statistics from the current population survey, BLS Report No. 313; Current Population Reports, Series P-23, No. 22, Washington DC; US Government Printing Office, June 1967

38- US Bureau of the Census. The Current Population Survey; A report on methodology. Technical Paper No. 7, 1963

39- US Bureau of the Census. Federal-State Cooperative Programs for Local Population Estimates; Test results, April 1, 1970. Current Population Reports, Series P-26, No. 2, 1973

40- US Bureau of the Census. Methods of Population Estimation: Part I Illustrative Procedure of the Census Bureau's Component Method II, Current Population Reports, Population Estimates, Series P-25, No. 339, June 6, 1966

41- US Department of Labor. Guidelines for Projecting Planning Year Supply Information: Section VI; 1973 Updated Indicators Package, Employment and Training Administration (Manpower Administration), Washington, D.C.

42- Waksberg, Joseph and Gonzalez, Maria Elena. Estimation of the error of synthetic estimates;paper presented at the International Association of Survey Statisticians, Vienna (Austria), Aug. 18-25, 1973. (Preprint)

43- Von Furstenberg, G.M. Stabilization characteristics of unemployment insurance, American Statistical Association, 1975 Proceedings of the Business and Economic Statistics Section, 10-22.

44- Von lee, Jose; Stone, Betty Jane. A computer-based technique for projecting income distribution. American Statistical Association, 1975 Proceedings of the Business and Economic Statistics Section, 378-382

45- Washington Center for Metropolitan Studies. Trends Alert; 1717 Massachusetts Ave. N.W., Washington, D.C. 20036

46- Zeisset, Paul T; Carbaugh, Larry W. Microdata from the Current
        Population Survey. American Statistical Association, 1974
        Proceedings of the Business and Economic Statistics Section,
        57-63.

FOOTNOTES

a. The Census Bureau's bibliography (Ref. 36) was the starting point for the bibliography above.

b. A letter from Dr. Harold Nisselson (Feb. 4, 1977) provided the information about current and recent work at the Census Bureau. Enclosed with this letter were the preprints quoted in 43 and 28 in the bibliography above.