

11-10-1972

# Postcensal Population Estimates for Oregon Counties: An Evaluation of Selected Methods

Guy Jeffrey Barnes  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Demography, Population, and Ecology Commons](#), and the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

**Let us know how access to this document benefits you.**

---

## Recommended Citation

Barnes, Guy Jeffrey, "Postcensal Population Estimates for Oregon Counties: An Evaluation of Selected Methods" (1972). *Dissertations and Theses*. Paper 1588.  
<https://doi.org/10.15760/etd.1587>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

AN ABSTRACT OF THE THESIS OF Guy Jeffrey Barnes for the  
Master of Science presented November 10, 1972.

Title: Postcensal Population Estimates for Oregon Counties:  
An Evaluation of Selected Methods

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

[REDACTED]

James E. Weiss, Chairman

[REDACTED]

Earle H. MacCannell

[REDACTED]

Barry D. Lebowitz

This study evaluates the results of three widely used methods for preparing postcensal estimates of counties. The methods are Census Bureau's Component Method II, the Ratio Correlation Method and the Bogue-Duncan Composite Method. Hypotheses based upon empirical generalizations from previous comparative studies are tested. Statistical tools used are Average Percent Deviation (without regard to sign) and Standard Deviation of Percent Errors. Directional bias and frequency of extreme error are also examined. Evaluations are conducted of the accuracy of the estimates for groups of counties stratified in terms of density and growth rate dimensions. With few exceptions, Ratio Correlation produces consistently better results. The ecological fallacy is illustrated in the application

of national migration assumptions to groups of constituent counties. Averaging the results of different methods does not produce appreciably greater accuracy. Other techniques may be useful in Oregon as benchmarks upon which to evaluate the reasonableness of Ratio Correlation estimates. Efforts in Oregon should be directed toward developing additional and/or more refined data series to be used in Ratio Correlation.

POSTCENSAL POPULATION ESTIMATES FOR OREGON COUNTIES:  
AN EVALUATION OF SELECTED METHODS

By

GUY JEFFREY BARNES

A thesis submitted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE  
in  
SOCIOLOGY

Portland State University  
1972

TO THE OFFICE OF GRADUATE STUDIES:

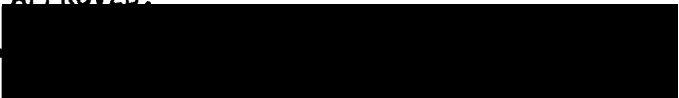
The members of the Committee approve the thesis of  
Guy Jeffrey Barnes presented November 10, 1972.

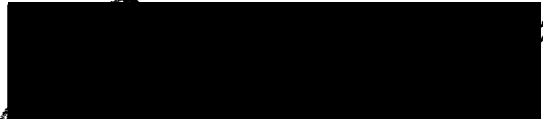
  
James E. Weiss, Chairman

  
Earle H. MacCannell

  
Barry D. Lebowitz

APPROVED:

  
Don C. Gibbons, Head, Department of Sociology

  
David T. Clark, Dean of Graduate Studies

## ACKNOWLEDGEMENTS

Without the patience and encouragement of my wife, Nancy, this work would not have been accomplished.

I am especially grateful to Jim Weiss, whose guidance and understanding in multiple roles was an essential and, I'm sure, difficult task.

Barry Lebowitz and Earle MacCannell made valuable contributions to the clarification and continuity of the statement of the problem.

Wes Brenner and other members of the staff of the Center for Population Research and Census provided valuable assistance.

I am indebted also to the Office of Graduate Studies for the one quarter extension to my seven year statute of limitation.

I am indebted to my parents, relatives and friends for the continual encouragement they afforded me.

## TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENT. . . . .	iii
LIST OF TABLES . . . . .	vi
CHAPTER	
I INTRODUCTION AND BACKGROUND. . . . .	1
Need for Population Estimates . . . . .	1
Different Classes of Estimates. . . . .	2
General Considerations in Selecting Data and Method. . . . .	3
Some Generalizations. . . . .	5
Limitations of Comparative Studies. . . . .	8
Hypotheses. . . . .	10
II SELECTED METHODS AND DATA. . . . .	11
Component Method II . . . . .	11
Ratio Correlation Method. . . . .	15
The Composite Method. . . . .	17
Averaging the Results of Various Methods. . . . .	19
Evaluating the Accuracy of Methods. . . . .	20
III MODIFIED COMPONENT METHOD II . . . . .	22
Procedural Steps. . . . .	22
Evaluation of the Results . . . . .	26

CHAPTER	PAGE
IV THE RATIO CORRELATION METHOD . . . . .	34
Procedural Steps. . . . .	34
Evaluation of the Results . . . . .	35
V THE COMPOSITE METHOD . . . . .	42
Procedural Steps. . . . .	42
Evaluation of the Results . . . . .	44
VI AVERAGING RESULTS. . . . .	59
VII SUMMARY AND CONCLUSIONS. . . . .	64
SOURCES CONSULTED. . . . .	70



## LIST OF TABLES

TABLE		PAGE
I	Summary of Deviations of 1970 Modified Component Method II Estimates for Oregon Counties. . . . .	27
II	Typology for Oregon Counties, 1970 . . . . .	28
III	Summary of Percent Deviations of 1970 Component Method II Estimates for Oregon Counties, by Type . . . . .	29
IV	Evaluation of 1970 Modified Component Method II Estimates for Oregon Counties, by Type . . . . .	31
V	Summary of Deviations of 1970 Ratio Correlation Estimates for Oregon Counties. . . . .	36
VI	Summary of Percent Deviations of 1970 Ratio Correlation Estimates for Oregon Counties, by Type . . . . .	37
VII	Evaluation of 1970 Ratio Correlation Estimates for Oregon Counties, by Type . . . . .	39
VIII	Summary of Deviations of 1970 Composite Estimates for Oregon Counties. . . . .	45
IX	Summary of Percent Deviations of 1970 Composite Estimates for Oregon Counties, by Type . . . . .	46
X	Evaluation of 1970 Composite Estimates for Oregon Counties, by Type. . . . .	48
XI	Summary of Percent Deviations of 1970 Composite Age Group Estimates for Oregon Counties. . . . .	49
XII	Evaluation of 1970 Composite Age Group Estimates for Oregon Counties, by Type . . . . .	52
XIII	Comparison of Actual and Estimated Age Specific Birth Rates, State of Oregon, 1970 . . . . .	53
XIV	Comparison of Actual and Estimated Age/Sex Specific Death Rates, State of Oregon, 1970 . . . . .	55
XV	Summary of Percent Deviations of 1970 Composite Estimates for Oregon Counties Adjusted to Method II Estimated State Total. . . . .	56

TABLE		PAGE
XVI	Evaluation of 1970 Adjusted Composite Estimates for Oregon Counties, by Type . . . . .	57
XVII	Summary Comparison of Percent Deviations of Selected 1970 Average Estimates for Oregon Counties . . . . .	60
XVIII	Comparative Evaluation of 1970 Selected Average Estimates for Oregon Counties, by Type . . . . .	61
XIX	Comparative Evaluation of All 1970 Selected Estimates for Oregon Counties, by Type . . . . .	65

## CHAPTER I

### INTRODUCTION AND BACKGROUND

#### I. NEED FOR POPULATION ESTIMATES

The purpose of this work is to evaluate the accuracy of selected methods of developing postcensal population estimates. The need to develop postcensal estimates arises from the fact that, at the present time, the most complete and reliable source of information on population is the most recent census. Population continues to change, sometimes quite rapidly, however, and often five or ten year benchmarks are not adequate for current purposes. Public health officials, for instance, need a current population base(s) in order to compute meaningful vital rates. Public and private planners, as well as market researchers, need current data to provide the bases for rational deployment of services and locating of investments. Often in the absence of current data, decision makers find themselves developing solutions to problems of the past. Although state and local governments sometimes conduct special censuses of their areas, these sparse data do not by any means meet public needs. Moreover, the method of complete enumeration is expensive, laborious and time-consuming. Also, this method is not applicable to past and future dates.

In order to meet the need for basic population data more fully, a wide array of estimating techniques has been developed. The sample survey, while it is an estimating method, is subject to the same limitations as a census, with exceptions occurring sometimes only as a matter of degree.

Analytic techniques which employ vital statistics, migration data and other data symptomatic of population change and mathematical methods, however, are relatively inexpensive to apply and can be used to prepare estimates for past, present and future dates.

## II. DIFFERENT CLASSES OF ESTIMATES

There are three basic types of estimates, viewed in the most general sense. Each presents its own set of methodological problems and, as a rule, achieves varying degrees of reliability. They are:

1) Intercensal Estimates, which are for dates in between complete enumerations and which use these counts in their determination; 2) Postcensal Estimates, which are for a date after a prior census, do not take into account any census after the estimate date. However, they often utilize other data covering the period between the previous census and the estimate date; 3) Projections, which are for dates after a prior census, usually the last one. The projection has an estimate date which is usually for some point in the future and is distinguished from the previous types in that no current data are available to be considered. In general, intercensal estimates are considered interpolations, while postcensal estimates and projections are extrapolations (1, p. 211).

Estimates vary in several other respects. They may be developed for various types of areas, ranging from entire countries to major geographical subdivisions, all the way down to areas within cities. While estimates for the total population are most common, estimates for various classes of a population differentiated by such criteria as age, sex, race, ethnicity, employment status, etc., may be developed. Definition of residency is also another important aspect of population

estimates. A person may be either a de jure (legal or voting, etc.) resident of an area, or he may be a de facto (usually, physically present) resident. Estimates tend to follow the same definition of residency as was used in the last census of the area, for the sake of comparability (2, p. 1).

### III. GENERAL CONSIDERATIONS IN SELECTING DATA AND METHOD

The most important factor determining the choice of the method(s) to be used in preparing a given population estimate is the type and quality of data available for this purpose. Despite the standardized logic, most methods require considerable modification before they can be applied to the available data. A method tailored to one area's configuration of input data will not necessarily be suitable in another data environment, whether a different area or the same one at a later time. The "state of the art" hinges less on methodological sophistication than on the "state of the data," their accuracy and reliability, and how readily they can be adapted to use (3, p. 2).

There are basically two types of data on which population estimates may be based: 1) Direct data, and 2) Indirect data. In both cases, the concern is with data for the base date and for the period between the base date and the estimate date. Whether a data series is classified as direct or indirect depends on the content of the information and how it is used in a particular method. For example, direct data may be (a) those from censuses or population registers, and (b) recorded data on the components of population change (i.e., statistics on births, deaths and migration) when these data are used to measure these phenomena themselves. Indirect data includes series on school enrollment,

employment, voter registration, passenger car registration and birth and death statistics when used to measure total population change rather than to measure natural increase directly. Estimates of population in households often rely heavily on such data as building permits, conversions, demolitions, electric and water hookups. It should be apparent that data of a given type may be direct for one kind of estimate and indirect for another and that there is no rigid dividing line between the two classes of data. Data on school enrollment represent indirect data if they are being used to estimate the total population and direct data if they are used to estimate the population of school age. Both direct and indirect data may be used in combination in preparing a given population estimate.

The usefulness of indirect data for population estimation depends largely on the extent to which they are influenced by factors other than population size and distribution. Changes in school enrollment may be due to changes in school district boundaries, laws regarding compulsory school attendance, availability of facilities and, of course, changes in the number of children of school age. Employment and housing construction vary with economic conditions as well as with population and households.

The data to be used must be evaluated carefully. The coverage of the latest census is especially important. An understanding of definitions and collection procedures is important. For instance, a spurious population change between two dates would be indicated if a population estimate based on school enrollment data used fall term head count one year and school year average daily membership on another estimate date. The method used in collecting the data may give important indications as

to the consistency of a series and the likelihood of over- or under-counting. For example, private schools may not be required to file reports with the state department of education. Special precautions must be taken to insure that all private school enrollments are included or population estimates which rely heavily on school enrollment could be seriously affected (2, pp. 3-4).

#### IV. SOME GENERALIZATIONS

1. More accurate estimates can generally be developed for large areas than for smaller areas. As a rule, more direct data, data of better quality and more information on how to adjust these data for deficiencies is available for the larger areas, such as the states, than for smaller areas. The size of small populations may fluctuate widely, making accurate estimation for these areas at least extremely difficult, if not impossible. The shutting down of a single factory may sharply reduce the population of a small county without having any discernible effect on the size of the population of the state. It is advisable, therefore, to adjust estimates for a group of contiguous small areas to an independently derived estimate of their area total.

2. The total population of an area can be estimated with more reliability than estimates for subpopulations of the area (i.e., demographic characteristics). Fewer and poorer data are available usually which would allow estimates to be made of the population classified by age, sex or other characteristics. Estimates of these characteristics should be adjusted to the area total (i.e., estimates of males and females should be adjusted to the estimate of total population and estimates of the female age distribution would be adjusted to the

adjusted estimate of total females).

3. Average error varies directly with rates of population growth. It is lowest among slowly growing counties, followed by rapidly growing counties, followed by counties losing population (4, 5, 6, 7, 8, 9, 10, 11).

4. In general, assuming that the available data are of high quality, direct data usually produce more accurate estimates than indirect data produce. The closer the base data are to the actual population being estimated, or change in the population size since some base date, the less manipulation of the base data is required and the estimate would be expected to contain less error. In reality, since any data may be defective, criteria of selection of base data include accuracy, completeness, internal consistency and recency of the data. Use of direct data and methods which parallel actual demographic processes will, on the average, produce more accurate estimates than the use of indirect data or methods.

5. It is reasonable to assume that the poorer the quality of the data and the longer the estimating period, the less reliable are the resulting population estimates. There is no conclusive evidence to support this assumption, however (4, 5, 6, 7, 8, 9, 10, 11).

6. Mathematical methods which incorporate no data related to the estimated date are to be avoided in the interest of accuracy. Interpolation, whether to fill holes in a data series or to produce intercensal estimates, is more reliable than extrapolation (2, pp. 4-5).

7. Other desirable features can be built into the design of an estimating procedure. Not all errors that can arise are manageable but some should be anticipated. For instance, one should be aware that a compounding logic is involved in composite methods. The risk of magnify-



ing errors is present because an estimate of one segment of the population is used to estimate another segment.

8. Although no single method has, by comparative study, shown consistently greater accuracy (4, 5, 6, 7, 8, 9, 10, 11), an estimate may be checked by comparing it with other estimates based on reasonable but different (ideally, independent) data and assumptions. If two estimates differ significantly from one another, there is reason to doubt both. If they are quite similar, one may have more confidence in each. The key is that the estimate to be compared employed primarily different data series and methods. Often, the best estimate is a weighted average of such alternative estimates. A self-correcting character has been imputed to this procedure and comparative evidence has shown that consistently lower average error can be attained by averaging together estimates made by different methods (4, 5, 6, 7, 8, 9, 10, 11).

9. The data environment is continually changing and expanding methods of estimating population should be sufficiently flexible in design as to readily incorporate new sources of information having possible estimation applications.

10. The value of estimates depends upon not only how accurate they are, but also upon how much information they provide. The effective delivery of family planning, health care, education and other community services is influenced by the size and distribution of their respective target populations. For example, the effective locating of free day care centers necessitates having some idea where children who are members of one parent families live, in addition to other kinds of parameter estimates for areas. Tradeoffs between precision and detail often are

possible and desirable. Batteries of estimating techniques should include components which are flexible and responsive to the needs of users in addition to those which are fixed by principles of methodology (3, pp. 4-5). In general, an estimating program should be based on the demand for various kinds of statistics, the availability and quality of the basic data, the amount of work necessary to produce the estimates and the resources (i.e., funds, personnel and time) available.

#### V. LIMITATIONS OF COMPARATIVE STUDIES

Many of the above generalizations are based upon numerous studies which have attempted to gauge the comparative accuracy of various methods. The results of these tests can be appraised jointly in several ways. In a very general and categorical sense, they have served to weed out the least accurate kinds of estimating approaches. However, most available evidence is still clouded by inconsistencies and problems of comparability. The conclusions are highly tentative in terms of formulating a basis for establishing a battery of estimates for an area not included in the study(s).

Briefly, the reasons (3, pp. 24-25) for the lack of comparability include the following: 1) The data used vary in type and in quality from one population to another; 2) Modifications are often necessary, due to the above, which induce variations and restrict comparability to an unknown extent; 3) Different universes contain different proportions of sub-areas whose characteristics (e.g., large or fast growing) produce differential bias in the overall results; 4) Adequate statistical appraisals of the results are rarely conducted. For instance, mean absolute percent deviation is used exclusively and/or no attempt is made

to control for various differences in characteristics of sub-areas (e.g., urban-rural, rate of growth).

It is suggested, therefore, that a battery of estimating techniques be developed for a specified universe and its constituent parts which utilizes the best available data sets for the specified estimating period. The methods selected for evaluation below for Oregon counties utilize resident livebirths, resident deaths, passenger car registrations, voter registrations and elementary school enrollment. These data are the "cleanest" available for the period. They also meet the important criterion of being tied directly or indirectly to actual demographic processes. Total resident livebirth and death registrations provide a measure of natural increase in population. An increase in resident livebirths usually indicates an increase in the population 15-44 years of age, who bear the prime responsibility for these events. An increase in deaths of older persons is often indicative of an influx of older persons into a given area.

An increase in passenger car registrations is suggestive of an increase in persons who usually drive cars. These are primarily persons who are ages 16-74 and who also represent a large segment of the total population. Voter registrations vary directly with changes in the population ages 21 and over (until 1970). Finally, elementary school enrollment reflects directly (by law) children ages 6-14 and indirectly population in other age segments, primarily those persons of an age range most likely to be that of parents of elementary school age children. Relationship to other age segments may be determined empirically for a specified time period.

In addition, it is necessary to control for relevant differences between sets of counties in tailoring methods for each type. The results of tests of the following hypotheses should meet this criterion.

## VI. HYPOTHESES

1. Average error should be lower in large or metropolitan or primarily urban counties.
2. Rapidly growing counties and counties losing population should experience errors with the most directional bias with the former being underestimated most consistently and the latter being overestimated most consistently.
3. Average error varies directly with rates of population growth. Slowly growing counties should have the lowest average error.
4. Stable or slowly growing counties can be estimated with a higher degree of reliability (i.e., lower standard deviation of errors) than rapid growth or population loss counties.
5. Total population estimates should be more accurate than age group estimates, especially where direct data is used only to obtain part of the age group and a compound logic is used to obtain the remainder.
6. Estimates which rely heavily on data representing a small number of events (e.g., age-specific deaths to estimate population ages 45 and over) will be less accurate than those using data series with magnitudes approaching the population size.
7. Consistently lower errors should be attained by averaging estimates made by different methods.
8. No single method should provide consistently greater accuracy.

## CHAPTER II

### SELECTED METHODS AND DATA

Three methods have been selected for evaluation. They are: 1) a modified version of U. S. Census Bureau's Component Method II, 2) the Ratio Correlation Method, and 3) the Bogue-Duncan Composite Method. Averages of their results will also be evaluated. Comparative evidence suggests these approaches should produce the most accurate results (4, 5, 6, 7, 8, 9, 10, 11).

#### I. COMPONENT METHOD II

Component Method II has been developed as a product of the Census Bureau's vast experience in preparing population estimates for the nation's geographic subdivisions. It has been used by many researchers who required estimates for regions, states and counties (7, 8, 9, 10, 11, 13). Population change is broken down into components of natural increase and net migration. Natural increase is measured through vital statistic records, whereas net migration is estimated by means of a symptomatic logic (13, p. 11). The formula is:

$$P_i = P_0 + B_i - D_i + M_i - M_a + A_i + e \quad (2.1)$$

where:  $P_i$  = estimated population on postcensal date  $i$ ,

$P_0$  = civilian resident population at time  $0$ ,

$B_i$  = resident livebirths during interval  $0$  to  $i$ ,

$D_i$  = resident civilian deaths during interval  $0$  to  $i$ ,

$M_i$  = estimated net civilian migration during the interval  $0$  to  $i$ ,

$M_a$  = estimated net movement of civilians into the Armed Forces during interval 0 to i,

$A_i$  = number of persons in the Armed Forces stationed in the study area at time i,

e = all measurement errors.

Data requirements for Component Method II are:

1. Resident births and deaths for the estimate period,
2. Elementary school enrollment by grade for the base year and the estimate year,
3. Net movement between civilian and military populations during the estimate period,
4. Institutional population as of last census date,
5. Population by age according to last census date.

Resident births and deaths can be obtained from appropriate official sources of vital statistics whose jurisdiction includes the study area.

Population change due to the stationing of Armed Forces ( $A_i$ ), as well as net movement of civilians into the military ( $M_a$ ), can be accounted for directly from Department of Defense statistics.

Institutional population and population by age are found in the appropriate printed report from the last census.

The crucial data series, however, is school data because of its mainstay role in the procedure for estimating net civilian migration ( $M_i$ ). The procedure relies heavily on the close and consistent relationship between net migration of elementary school age children and the corresponding movement by the total population.

This involves comparing school data on the estimate date with expected enrollment based on the age-specific survivors from the last census. The difference between actual and expected enrollments is assumed to represent net migration of school age children. A school age

migration rate is determined. A migration ratio, based on the age distribution of migrants recorded in the Current Population Survey, is used to adjust the school age rate to a rate which applies to the total civilian noninstitutional population. For April 1, 1970, the inter-county migration ratio is .9193 (14).

School data in any of various forms, whether they be school census, school enrollment or school attendance data, might be best as a Component Method II input, depending upon the study area. School censuses are actual enumerations of children of school age (including sometimes preschoolers, as well) conducted by school authorities in certain states. School enrollment data represent counts of children enrolled in school. In the United States they are generally available for all important political units (states, cities and counties) in one or more of several forms (i.e., beginning fall enrollment, year-end enrollment, average annual enrollment). Average daily attendance and average daily membership represent other types of school enrollment data.

Theoretically, school census data with carefully defined age limits would be most useful for estimating purposes. However, in the few states where school censuses are taken, the quality of enumeration is usually quite poor and varies appreciably from year to year. Grade data are generally more dependable than school census data for measuring year-to-year population changes. While enrollment data by age are, in principle, preferable to data by grade, they are usually not available. Since the school data series serves to measure changes in the population of school age, its coverage must be restricted to those ages where attendance is virtually complete, i.e., elementary grades (excluding kindergarten but including any special and ungraded classes on the ele-

mentary level). High school enrollment data are unsatisfactory since many children drop out of high school and the drop-out rates vary from year to year. The grade coverage must be the same from year to year and the figures must relate to the same date in the school year (2, p. 26).

As indicated above, the migration rate for the total population of the study area is the product of the computed school age rate for the study area times the appropriate national average interstate or inter-county total population to school age population migration ratio. There are important assumptions in this procedure: 1) that there has been no change since the last census in the ratio of the population of elementary school age to the number enrolled in the elementary grades, and 2) that the ratio of the net migration rate of the total population to the migration rate of the school age population for a given postcensal period, for a given local area, corresponds to that for gross interstate or intercounty migrants in the United States for the same period. In view of the very high proportion of children attending school, the first assumption would give rise to relatively little error. However, geographic variations in the ratio of the migration rate of the total population to the migration rate of the elementary school age population reflecting area differences in the age composition of migrants, are far more important. The movement of certain groups in the population are poorly represented by the movements of school age children (e.g., single men and women, new migrants to suburbs, institutional population, and college students). On the other hand, the civilian population of an area with many families headed by military personnel may be overestimated since their school age children will be included in the figure on school enrollment. This assumption, therefore, can give rise to rather large



errors in the population estimates. However, this type of error is, in general, reduced by a pro rata adjustment of the initial estimates of net migration of population for a set of local areas (e.g., counties) to an independent estimate for the parent area (i.e., state) (2, p. 28).

## II. RATIO CORRELATION METHOD

A more versatile approach to the use of symptomatic data is to estimate the total population of an area directly by mathematically relating changes in several indicator data series to changes in population by means of a multiple regression equation. Kinds of symptomatic data which have been used in the application of this method in the United States are vital events, elementary school data, motor vehicle registrations, voter registrations, employment, bank deposits, income tax returns, and others. Since population is usually limited to the decennial census, variables used in these equations typically are expressed as ten year ratios of change in a set of study areas' shares of a parent "population" (5, 6, 7, 8, 9, 10, 11).

Ratio correlation estimation proceeds in two stages. First, a multiple regression equation is constructed to express the relationship between (a) the change over the previous intercensal period in an area's share of the total for the parent area for several symptomatic data series and (b) the change in an area's share of the population of the parent area.

In the first stage of the process, an equation is estimated for  $n$  independent variables and takes the general form

$$Y = A + B_1 X_1 + B_2 X_2 + \dots + B_n X_n + e \quad (2.2)$$

Observation of ten year relationships among  $Y$  and  $X_i$  refers to some uni-

verse of contiguous geographic units, such as counties within a state. The equation in Chapter IV illustrates how change in population of Oregon counties between 1950 and 1960 relates to changes in births, school enrollment, registered voters and passenger car registrations in those counties during the same period.

Once the estimating relationship has been established, it is used to convert postcensal changes in the symptomatic data into corresponding estimates of population change.

A necessary assumption, of course, is that the estimating relationship remains invariant over time. The adequacy of this assumption is dependent on the size of the multiple correlation coefficient ( $R^2$ ) and the number of variables used. High multiple correlation coefficients for two consecutive decades would suggest that the degree of association is not changing very rapidly. In such a case, the regression based on the last intercensal period should be applicable to the current intercensal period. Also, if the symptomatic indicators maintain interrelationships of the same order of magnitude during the current postcensal period, confidence in the equation is enhanced. With the use of high speed computers, many combinations of variables can be included.

A further assumption is that deficiencies in coverage in the basic data series will remain constant, or change very little in the postcensal period. Often extensive investigation is required to justify this confidence.

Ratio correlation has several advantages. First, it provides the statistical capabilities associated with this type of model: quantitative measures of goodness of fit, variance explained and statistical differences. Second, the model is flexible in that it is suited to a

changing data environment. Finally, variables are weighted on empirical rather than intuitive grounds.

One apparent drawback centers around when new change measures can be used. Because of the need to construct the initial estimating relationship, new variables that become available during the postcensal period itself cannot be incorporated into the model. For example, the release of county medicare statistics beginning in 1966 cannot be used to develop estimates until change in these statistics can be related to change in population (3, pp. 19-22).

### III. THE COMPOSITE METHOD

A characteristic of the two methods described thus far is that they are single methods, not combinations of methods which measure change in the total population of an area directly rather than in terms of its principal components. However, composite techniques build on the assumption that most symptomatic measures and methods attain peak sensitivity within specific age brackets. The logic of composite estimation is to assemble a complement of indicators that jointly measure all age groups in the population.

Any composite model has to reconcile two conflicting objectives: flexibility and generality. On the one hand, sources of data available for a specific population should be used to maximum advantage. On the other hand, a general composite design has the advantage of being widely applicable. Moreover, its precision can be assessed more thoroughly in a variety of situations (3, p. 17).

The Bogue-Duncan Composite Method (4) strikes a balance between these two objectives. The method utilizes three widely available sources

of data--school enrollment, births and deaths--to measure the respective components of the population which are most likely to generate these events. Many alternatives are possible. Typically, the number of deaths of persons 45 years old and over by age group is used to estimate the population in this age range; the number of females in the childbearing ages (15 to 44 years), which, in turn, is used to estimate the number of males in this age range; school enrollment data are used to estimate the population of school ages (5 through 14 years), and the number of births in the previous 5-year period, in conjunction with migration information generated for the above component, are used to estimate the population under five years of age.

Composite estimation has some advantages. It can provide considerable compositional detail, limited only by the level of detail in the input data. Moreover, its modular design facilitates additional refinement as other, perhaps more reliable, symptomatic indicators become available.

On the other hand, this approach is more laborious to carry out and requires more refined data (e.g., deaths by age) when compared to the other procedures described. A practical difficulty with its application for small areas (e.g., counties in the U. S.) is the absence of required data on a regular, current basis. The U. S. National Center for Health Statistics does not, for example, tabulate resident deaths by age for areas within states. Moreover, the lag in reporting of their data is not always predictable. Some state health departments have tabulated such figures for the counties and larger cities in their state, however.

The details of the composite method may be varied according to the kinds and quality of data available. One such variation is presented

in Chapter V.

#### IV. AVERAGING THE RESULTS OF VARIOUS METHODS

The averaging of methods may be employed as a basis for improving the accuracy of population estimates (2, p. 33). Ideally, the methods to be averaged should employ different indicators or essentially different procedures and assumptions. This ideal is seldom met. Averaging the results of two methods may affect the accuracy of population estimates in two ways: 1) It may reduce the risk of an extreme error, and 2) It may partly offset opposite biases characteristic of the types of estimates to be averaged. Averaging these results of three or more methods entails the risk of introducing systematic bias (compounding error) due to a majority of the results being highly intercorrelated. The methods to be averaged may be selected subjectively or on the basis of various quantitative indications given by studies of the accuracy of methods. Because of the problems of comparative studies cited above, this decision should be made empirically on this basis of results of different methods for the same study area. For example, two methods which have relatively low average errors, but which have opposite biases, may be considered good candidates for averaging. The existence of opposite biases is indicated by a low or negative correlation between percent errors for the geographic units in a distribution (e.g., counties) according to the two methods. The methods to be averaged may be given the same weights or different weights. These too may be determined subjectively or quantitatively on the basis of evaluation studies of comparable areas or, for the same areas at an earlier date.

## V. EVALUATING THE ACCURACY OF METHODS

A variety of methods of estimating populations of geographic subdivisions in the U. S. has been discussed with relatively little explicit concern for evaluating their relative accuracy. Any consideration of the accuracy of methods depends on the availability of an adequate standard against which to judge the methods and the establishment of criteria of accuracy. We may consider as general criteria of the accuracy of postcensal estimates one or more of the following: 1) consistency of the estimate with a previous census count, 2) agreement with a subsequent census count or 3) agreement with another type of independent estimate. Securing an adequate standard to represent the "true" population is often difficult or impossible. A special registration or a population register may provide more correct figures than the census. Or, a subsequent census may not be quite comparable with the previous census which provides the population base for the estimate. Because of the lack of a standard to judge the absolute validity of most estimates, and because of the implicit character of most estimates both as comparable to the last census and as anticipating the results of the next census, the common practice is to employ the later census as a standard for evaluating postcensal estimates.

The principal measures of accuracy used have been: 1) the mean percent deviation of the estimates from the census counts over all geographic subdivisions estimated for the study area (the errors being taken without regard to sign); 2) the number of deviations of 5 to 10 percent, or more; 3) the number of positive and negative deviations; 4) and the standard deviation of the percent errors taken without regard

to sign.

The first measure describes average error, the second, the frequency of extreme error, the third the presence of upward or downward bias (or long-run inaccuracy) in the method. If appreciably more than half of the estimates for counties have positive or negative deviations from the census counts, we would attribute an upward or downward bias to the set of estimates. The fourth measure (standard deviation of percent errors) is also useful as a measure of extreme errors and the degree of efficiency or reliability of the method. It refers to, in other words, the degree of confidence that, on a given trial, the procedure will attain a stated level of precision (2, pp. 41-42).

## CHAPTER III

### MODIFIED COMPONENT METHOD II

This chapter contains a procedural outline of Component Method II, application of the method to the 36 counties of Oregon, an evaluation of the results and the extent to which the original (working) hypotheses are supported.

#### I. PROCEDURAL STEPS

For the sake of clarity, the following is presented in outline form.

- A. County Population = Total resident population of county, April 1, 1960 + Natural increase April 1, 1960 to March 31, 1970 + Net migrants April 1, 1960 to April 1, 1970.
- B. Total Resident population of each county at the time of the last federal census (1960).
- C. Natural Increase of each county, April 1, 1960 to March 31, 1970.  
From Vital Statistics Summaries (15) obtain by county:
  1. Resident livebirths
  2. Resident deaths
  3. Natural Increase (births minus deaths).
- D. Estimate of total net migrants April 1, 1960 to April 1, 1970, based on relationships between total (public and non-public) school enrollment grades 1-8 and cohort population ages 6.00 - 14.99:
  1. Survivors of births April 1, 1960 to March 31, 1964,



ages 0 - 1 through ages 4 - 5 on April 1, 1960, all to ages 6.00 - 14.99 on April 1, 1970.

a. Stationary populations (16) by single years

0 - 14.

b. Survival rates for state and counties (17):

i. Rates for births ('60-'64) to ages 6 - 9.99 ('70) are simply the single year stationary populations for ages 6 - 7, 7 - 8, 8 - 9 and 9 - 10, respectively, as they appear in column Lx of the Life Table.

ii. Rates for under 1 to 4.99 ('60) to ages 10 - 14.99 ('70) are derived by dividing stationary population by 10 - 11 by stationary population age 0 - 1, age 11 - 12 by age 1 - 2, and so on, through dividing age 14 - 15 by age 4 - 5.

iii. All survival rates should be carried to five decimal places as reflects a base of 100,000.

c. Survivors (for the State and each county).

i. Apply survival rate of births to age 6 - 7 to births occurring April 1, 1963 to March 31, 1964 and so on in annual increments through the rate of births to age 9 - 10 to births occurring April 1, 1960 to March 31, 1961.

ii. Apply rates for single years of age starting with rate of under 1 to age 10 - 11 to age under 1 on April 1, 1960, and so on, through

the rate of age 4 - 5 to 14 - 15 to age 4 - 5 on April 1, 1960 to get survivors on April 1, 1970 who are ages 10 - 11, 11 - 12, 13 - 14, and 14 - 15.

- iii. Sum single-year survivors (derived in items i - ii) to obtain an estimate of ages 6.00 - 14.99 on April 1, 1970. This estimate, of course, assumes that there has been zero net migration in this cohort since April 1, 1960.

2. Net Migration Rates (for State and each county).

- a. Compute a ratio (six places) of population sum of ages 6.00 - 14.99 on April 1, 1960 to the sum of public and non-public regular unhandicapped enrollment (Average Daily Membership) in grades 1-8, Spring, 1960 (18).
- b. Apply the ratio (item a) to the sum of public and non-public regular unhandicapped enrollment in grades 1-8, Spring, 1970 (18) to obtain an estimate of school age (6.00 - 14.99) population. This assumes the relationship between school age population and enrollment in grades 1-8 is unchanged.
- c. The difference between this estimate (item b) and the estimate obtained from surviving the base population (item 1:c. iii) provides an estimate of the net migration that has occurred in the cohort during the period.
- d. The raw net migration rate (i.e., for the school age population) is derived by dividing the estimated net

migrants (item c) by the base population, which is the sum of births (April 1, 1960 to March 31, 1964) and ages 0 through 6.99 (April 1, 1960).

e. The total population net migration rate is the product of the raw migration rate times the intercounty migration ratio. This factor was developed by the U. S. Bureau of the Census from relationships obtained on the basis of the Current Population Survey (13, 14).

3. Estimates of Net Migration (independent for State and for counties).

a. Base population for computing estimated net migrants for total population is the algebraic sum of the total enumerated population on April 1, 1960 plus 1/2 the natural increase (births minus deaths) April 1, 1960 to March 31, 1970 minus the total special population on April 1, 1960. Special population is defined as a unique population, which exceeds 3 per cent of the total population, changes in which can be said to reflect almost entirely net migration, and finally, data for which are available on both the census and estimate dates. In Oregon, military station strength does not meet this criterion.

b. The estimate of total net migrants for the period is the product of the total net migration rate (item 2:e) and the base population (item 2).

E. Estimate of total population.

1. The estimate of total population on April 1, 1970, then, is the sum of total population on April 1, 1960 minus the Special

population on April 1, 1960 plus total net migrants during the period (item D:3:b) plus total natural increase during the period plus the Special population on April 1, 1970.

2. The ratio of the independent State estimate to the sum of the independent county estimates is used to develop final county estimates which, of course, sum to the State estimate.

## II. EVALUATION OF THE RESULTS

Table I presents a summary of deviations of method estimates from 1970 census counts. While, in general, errors of larger magnitudes occur in larger counties, large percent deviations occur more frequently in smaller counties. Table I also provides the data for the county classification scheme in Table II. While the Metropolitan, Other Urban and Rural classification bases are apparent, the growth rate types are scaled relative to the State of Oregon, which is growing at almost twice the rate of the entire United States. The lower cutoff of +19.5 percent for the Moderate to Fast Growth category is convenient in that the county with the next highest rate of change is Harney at +7.0 percent, the upper limit of the Stable to Slow Growth county classification. Only two counties in this class actually "lost" population, but -1.6 percent and -0.4 percent is not much in ten years. Six out of ten counties in the Population Loss category experienced declines in excess of ten percent.

Errors for Metropolitan counties range from -7.54 percent to +3.32 percent, as shown by Table III and there is an equal number of plus and minus deviations. Only one Other Urban county is underestimated as the

TABLE I  
 Summary of Deviations  
 of 1970 Modified Component Method II Estimates  
 for Oregon Counties

State	1970 Census	1970 Method II Estimate	Absolute Deviation	Percent Deviation	Percent Change 1960 to 1970	Percent Urban
State	2,091,533	2,094,997	3,464	0.16	18.2	67.1
Baker	14,919	17,341	2,442	16.23	-13.7	62.7
Benton	53,776	51,334	-2,442	- 4.55	37.3	65.4
Clackamas	166,088	170,579	4,491	2.70	46.9	62.7
Clatsop	28,473	29,105	632	2.21	4.0	51.4
Columbia	28,790	30,179	1,389	4.82	28.6	21.6
Coos	56,515	58,759	2,244	3.97	2.8	51.3
Crook	9,985	10,313	328	3.28	5.9	41.1
Curry	13,006	14,021	1,015	7.80	- 7.0	20.9
Deschutes	30,442	30,603	161	0.52	31.8	57.3
Douglas	71,743	77,382	5,639	7.86	4.8	33.9
Gilliam	2,342	2,826	484	20.66	-23.7	--
Grant	6,996	7,499	503	7.18	- 9.4	--
Harney	7,215	7,234	19	0.26	7.0	45.6
Hood River	13,187	14,699	1,512	11.46	- 1.6	30.3
Jackson	94,533	96,317	1,784	1.88	27.8	55.2
Jefferson	8,548	9,074	526	6.15	19.9	--
Josephine	35,746	36,652	906	2.53	19.5	51.9
Klamath	50,021	51,034	1,013	2.02	5.4	63.0
Lake	6,343	6,981	638	10.05	-11.4	42.6
Lane	215,401	212,381	-3,020	- 0.46	31.0	69.7
Lincoln	25,755	24,796	- 959	- 3.73	4.5	47.4
Linn	71,914	71,764	- 150	- 0.21	22.2	39.8
Malheur	23,169	25,656	2,487	10.73	1.8	39.5
Marion	151,309	156,344	5,035	3.32	25.2	67.0
Morrow	4,465	4,872	407	9.11	- 8.3	--
Multnomah	554,668	533,778	-20,890	- 3.77	6.5	97.4
Polk	35,349	35,457	108	0.30	33.3	58.8
Sherman	2,139	2,308	169	7.90	-12.6	--
Tillamook	18,034	18,368	334	1.85	- 5.4	22.1
Umatilla	44,923	46,258	1,335	2.97	1.3	49.4
Union	19,377	20,670	1,293	6.67	6.6	49.8
Wallowa	6,247	6,623	376	6.01	-12.0	--
Wasco	20,133	21,941	1,808	8.98	- 0.4	51.8
Washington	157,920	146,015	-11,905	- 7.54	71.2	74.5
Wheeler	1,849	2,141	292	15.79	-32.1	--
Yamhill	40,273	43,693	3,480	8.63	23.8	41.4

TABLE II  
 Typology for Oregon Counties  
 1970

County Type	Number of Counties	Explanation
All	36	Total
Metropolitan	6	Federally Defined SMSA Counties
Other Urban	9	Urban Population $\geq$ 50%, 1970
Rural	21	Urban Population < 50%, 1970
Moderate to Fast Growth	13	Ten Year Change $\geq$ +19.5% (State = +18.2%)
Stable to Slow Growth	13	Ten Year Change $\geq$ -1.6% $\leq$ +7.0%
Population Loss	10	Ten Year Change $\geq$ -5.4%

TABLE III

Summary of Percent Deviations  
of 1970 Component Method II Estimates  
for Oregon Counties, by Type

County	Metro- politan Counties	Other Urban Counties	Rural Counties	Moderate to Fast Growth Counties	Stable and Slow Growth Counties	Popula- tion Loss Counties
Baker		16.23				16.23
Benton		-4.55		-4.55		
Clackamas	2.70			2.70		
Clatsop		2.21			2.21	
Columbia			4.82	4.82		
Coos		3.97			3.97	
Crook			3.28		3.28	
Curry			7.80			7.80
Deschutes		0.52		0.52		
Douglas			7.86		7.86	
Gilliam			20.66			20.66
Grant			7.18			7.18
Harney			0.26		0.26	
Hood River			11.46		11.46	
Jackson		1.88		1.88		
Jefferson			6.15	6.15		
Josephine		2.53		2.53		
Klamath		2.02			2.02	
Lake			10.05			10.05
Lane	-0.46			-0.46		
Lincoln			-3.73		-3.73	
Linn			-0.21	-0.21		
Malheur			10.73		10.73	
Marion	3.32			3.32		
Morrow			9.11			9.11
Multnomah	-3.77				-3.77	
Polk	0.30			0.30		
Sherman			7.90			7.90
Tillamook			1.85			1.85
Umatilla			2.97		2.97	
Union			6.67		6.67	
Wallowa			6.01			6.01
Wasco		8.98			8.98	
Washington	-7.54			-7.54		
Wheeler			15.79			15.79
Yamhill			8.63	8.63		

deviations for these counties range from -4.55 percent to 16.23 percent. All but two of these nine counties are within five percent of the census count. Of the twenty-one Rural counties, only two are negative deviations and the range is -3.73 percent to +20.66 percent.

The growth rate typology reveals that Moderate to Fast Growing counties are underestimated in four of thirteen cases, with deviations from -7.54 percent to +8.63 percent. Stable and Slow Growth counties also tend to be overestimated (11 of 13 cases) and have deviations ranging from -3.77 percent to +11.46 percent. All ten Population Loss counties are overestimated by Method II, most quite badly, as the errors range from +1.85 percent to 20.66 percent. Only one county came within five percent.

While the Method II State total estimate is quite close to the census (+0.16 percent), it has an upward bias in thirty of thirty-six counties. The six underestimates are of sufficient magnitudes to compensate statewide.

An old census folklore adage holds that if a census count falls within three percent of "reality," as inferred by the post-enumeration survey, the operation is considered a success. A similar upper limit of five percent has become the traditional benchmark for dichotomizing good and bad population estimates. According to this criterion, the overall performance of Method II is not very encouraging since almost half (17 of 36) of the counties deviate by more than five percent, as can be seen in Table IV. Moreover, the average error for all counties is 5.95 percent. As a reliable estimator of the population of all Oregon counties, Method II leaves something to be desired in that it is necessary to include all counties that deviate in absolute percents



TABLE IV  
 Evaluation of 1970 Modified Component  
 Method II Estimates for Oregon Counties,  
 by Type

County Type	Number of Counties	Number of Positive Deviations	Number of Deviations $\geq 5\%$	Number of Deviations $\geq 10\%$	Mean % Deviation	Standard Deviation
All	36	30	17	6	5.95	4.82
Metropolitan	6	3	1	0	3.02	2.65
Other Urban	9	8	2	1	4.77	4.94
Rural	21	19	14	5	7.29	4.90
Moderate to Fast Growth	13	9	3	0	3.35	2.82
Stable to Slow Growth	13	11	5	2	5.22	3.55
Population Loss	10	10	9	4	10.26	5.63

ranging from 1.13 to 10.77 in order to capture the "normal" 68 percent of the cases.

Method II fares better on Metropolitan counties. Half the deviations are positive and hence no directional bias. Only one county has an "excessive" deviation and the deviations are clustered much closer than overall, reflecting highest reliability of any subset. The performance of the method is progressively worse on Other Urban and Rural counties, respectively, although only two of nine Other Urban deviations exceed five percent. While two-thirds of the Rural counties exceed five percent and almost one-fourth are over ten percent, the estimates are reliable to about the same degree as the Other Urban county estimates.

As for the typology which groups counties according to growth rate, the method performs best on the Moderate to Fast Growing counties. While nine of thirteen are overestimated, this is the least biased of the three classifications. In addition, only three estimates in this class deviate in excess of five percent and there are none over ten percent. Reliability in this class is highest of the three as shown by a standard deviation of 2.82 compared with 3.55 (Stable to Slow Growth) and 5.63 (Population Loss). Method II overestimated all ten Population Loss counties, all but one excessively, with the average error at 10.26 percent.

The results of the above support the following conclusions regarding the hypotheses tested:

1. Average error is lower for Metropolitan and primarily urban counties.
2. Counties with the highest growth rates are underestimated more than other classes and Population Loss counties are

all overestimated but the overall upward bias of this method render any conclusion regarding this hypothesis highly tentative.

3. Average error does not vary directly with growth rate due to an inversion in the average error for Moderate to Fast Growth counties (low of 3.55 percent). However, average error for Stable to Slow Growth counties is much lower than that of Population Loss counties. These mixed findings are somewhat inconclusive.
4. No support is found for the hypotheses that Stable or Slow Growth counties can be estimated with a higher degree of reliability than rapid gain or loss counties. The standard deviation for this class falls between the other two.

## CHAPTER IV

### THE RATIO CORRELATION METHOD

This chapter contains the procedural steps used in the Ratio Correlation Method, its application to Oregon's 36 counties, an evaluation of the results and the extent to which the hypotheses are supported.

#### I. PROCEDURAL STEPS

The dependent variable (Y) in the regression equation (2.2) represents the ratio of a county's share of the State total population in 1960 to its share in 1950; that is,

$$\frac{\text{Proportion of State population in County } i, 1960}{\text{Proportion of State population in County } i, 1950}$$

The independent variables ( $X_1, X_2$ , etc.), for example, passenger car registrations, are expressed in a corresponding manner,

$$\frac{\text{Proportion of State passenger cars in County } i, 1960}{\text{Proportion of State passenger cars in County } i, 1950}$$

The correlation coefficients between the independent variables and the dependent variable are as follows:

$$\begin{aligned} X_1, \text{ Cars (21)} &= .9388 \\ X_2, \text{ Voters (22)} &= .9356 \\ X_3, \text{ Births (15)} &= .9745 \\ X_4, \text{ School Enrollment (18)} &= .9770 \end{aligned}$$

The multiple correlation coefficient for these variables,  $R^2_{Y.X_1X_2X_3X_4}$  is .9868. The corresponding regression equation is

$$Y = .1192 + .2066X_1 + .2456X_2 + .2703X_3 + .1651X_4 + e \quad (4.1)$$

where error (e) is assumed to be distributed randomly.

Estimates for 1970 (April) are prepared by substituting in the equation appropriate data for the 1960-70 period. For example, the value  $X_1$  for a given county in 1970 is computed as:

$$\frac{\text{Proportion of State passenger cars in County } i, 1970}{\text{Proportion of State passenger cars in County } i, 1960}$$

Values for other independent variables are derived in similar fashion. When the equation is solved for each county, the results represent estimates of the following form:

$$\frac{\text{Proportion of State population in County } i, 1970}{\text{Proportion of State population in County } i, 1960}$$

The ratio so computed for each county is applied to each county's proportion of the State population in 1960, as shown by the 1960 census, to arrive at its estimated proportion of the State population in 1970. The 1970 proportions for all counties are summed and adjusted to add to 100 percent. These proportions are applied to the estimated resident population for the State for April 1, 1970 (in this case, Method II estimate), yielding an estimate of the resident population of each county on April 1, 1970.

## II. EVALUATION OF THE RESULTS

Table V summarizes the deviations of the Ratio Correlation estimates of total population from the 1970 census counts. The errors range from -11.46 percent to +8.11 percent. Errors of larger absolute magnitude occur, in general, in the larger counties. The pattern of percent deviations presents no clear distinctions. Plus deviations occur in nearly half the counties (16 of 36), revealing no directional bias.

From Table VI we can see that the Metropolitan counties experience deviations from -11.46 percent to +5.95 percent. Four of six are under-

TABLE V  
 Summary of Deviations  
 of 1970 Ratio Correlation Estimates  
 for Oregon Counties

State	1970 Census	1970 Ratio Correlation Estimate	Absolute Deviation (Est.-Cens.)	Percent Deviation
Baker	14,919	15,755	836	5.60
Benton	53,776	49,503	-4,273	-7.95
Clackamas	166,088	161,764	-4,324	-2.61
Clatsop	28,473	28,857	384	1.34
Columbia	28,790	30,642	1,852	6.43
Coos	56,515	57,126	611	1.08
Crook	9,985	9,419	- 566	-5.67
Curry	13,006	12,562	- 444	-3.42
Deschutes	30,442	29,064	-1,378	-4.53
Douglas	71,743	72,730	987	1.37
Gilliam	2,342	2,510	168	7.17
Grant	6,996	6,916	- 80	-1.15
Harney	7,215	7,339	124	1.71
Hood River	13,187	13,516	329	2.49
Jackson	94,533	91,640	-2,893	-3.07
Jefferson	8,548	8,528	- 20	-0.24
Josephine	35,746	34,922	- 824	-2.31
Klamath	50,021	50,408	387	0.77
Lake	6,343	6,277	- 66	-1.05
Lane	215,401	206,609	-8,792	-4.08
Lincoln	25,755	25,576	- 179	-0.70
Linn	71,914	69,056	-2,858	-3.98
Malheur	23,169	23,816	647	2.79
Marion	151,309	158,258	6,949	4.59
Morrow	4,465	4,203	- 262	-5.87
Multnomah	554,668	587,687	33,019	5.95
Polk	35,349	31,301	-4,048	-11.46
Sherman	2,139	2,118	- 21	-0.99
Tillamook	18,034	17,796	- 238	-1.32
Umatilla	44,923	45,396	473	1.05
Union	19,377	19,743	366	1.88
Wallowa	6,247	6,152	- 95	-1.53
Wasco	20,133	20,606	473	2.34
Washington	157,920	145,917	-12,003	-7.61
Wheeler	1,849	1,999	150	8.11
Yamhill	40,273	39,289	- 984	-2.30

TABLE VI

Summary of Percent Deviations  
of 1970 Ratio Correlation Estimates  
for Oregon Counties, by Type

County	Metro- politan Counties	Other Urban Counties	Rural Counties	Moderate to Fast Growth Counties	Stable and Slow Growth Counties	Popula- tion Loss Counties
Baker		5.60				5.60
Benton		-7.95		-7.95		
Clackamas	-2.61			-2.61		
Clatsop		1.34			1.34	
Columbia			6.43	6.43		
Coos		1.08			1.08	
Crook			-5.67		-5.67	
Curry			-3.42			-3.42
Deschutes		-4.53		-4.53		
Douglas			1.37		1.37	
Gilliam			7.17			7.17
Grant			-1.15			-1.15
Harney			1.71		1.71	
Hood River			2.49		2.49	
Jackson		-3.07		-3.07		
Jefferson			-0.24	-0.24		
Josephine		-2.31		-2.31		
Klamath		0.77			0.77	
Lake			-1.05			-1.05
Lane	-4.08			-4.08		
Lincoln			-0.70		-0.70	
Linn			-3.98	-3.98		
Malheur			2.79		2.79	
Marion	4.59			4.59		
Morrow			-5.87			-5.87
Multnomah	5.95				5.95	
Polk	-11.46			-11.46		
Sherman			-0.99			-0.99
Tillamook			-1.32			-1.32
Umatilla			1.05		1.05	
Union			1.88		1.88	
Wallowa			-1.53			-1.53
Wasco		2.34			2.34	
Washington	-7.61			-7.61		
Wheeler			8.11			8.11
Yamhill			-2.30	-2.30		

estimates. Four of nine Other Urban counties are underestimated, but the overall pattern is tighter than the subset above with a range from -7.95 to +5.60 percent. Directional bias of the method is not revealed by examination of the deviations for Rural counties (9 plus, 11 minus). The range for this class is similar to those above, going from -5.67 to +8.11.

When controlled for rate of growth, we see that the deviations for Moderate to Fast Growth counties fall below census counts in most cases (11 of 13) and also vary widely (-11.46 percent to +6.43 percent). Deviations for Stable to Slow Growth counties cluster in closer proximity to zero, ranging from -5.67 percent to +5.95 percent, but with a definite upward bias (11 of 13 cases). Strangely, this method underestimates Population Loss counties in seven of ten cases and the deviations range from -5.87 percent to +8.11 percent.

Overall, according to the five percent criterion discussed above, Ratio Correlation produces "bad" estimates in 10 of 36 cases. However, only one of these exceeds ten percent. The average deviation of 3.51 percent reflects the bulk of low errors the tightest of which are confirmed by a standard deviation of 2.69. Table VII illustrates.

For Metropolitan counties the method is not as good as overall with deviations in three of six counties exceeding five percent. The average error of 6.05 percent also is excessive.

The picture improves considerably in Other Urban counties in which only two of nine counties deviate by five percent or more and none exceed ten percent. The 3.22 average deviation is low as is the standard deviation of 2.38.



TABLE VII  
 Evaluation of 1970 Ratio Correlation Estimates  
 for Oregon Counties, by Type

County Type	Number of Counties	Number of Positive Deviations	Number of Deviations $\geq 5\%$	Number of Deviations $\geq 10\%$	Average Percent Deviation	Standard Deviation
All	36	16	10	1	3.51	2.69
Metropolitan	6	2	3	1	6.05	3.15
Other Urban	9	5	2	0	3.22	2.38
Rural	21	9	5	0	2.92	2.36
Moderate to Fast Growth	13	2	4	1	4.70	2.99
Stable to Slow Growth	13	11	2	0	2.24	1.71
Population Loss	10	3	4	0	3.62	2.81

This technique works even better in Rural counties, where only five of twenty-one estimates are off by five percent or more, and none are ten percent or more. This approach is most efficient (reliable) for this typology suggesting an inverse relationship between population density and stability of interrelationships among the selected variables.

The growth rate typology reveals that, while only 4 of 13 Moderate to Fast Growth counties exceed five percent deviations, the method is not too accurate for this group whose average deviation is 4.70 percent. This is helped in part by the one county over ten percent. Deviations are fairly consistent, however, as seen by the standard deviation of 2.99

Ratio Correlation performs best on Stable to Slow Growth counties, by every available standard. Only two of 13 exceed five percent (none over ten percent). The remainder range from  $-.70$  percent to  $+2.49$  percent, accounting for the low average deviation (2.24 percent) and standard deviation (1.71).

Population Loss counties mirror the overall picture.

The results of the above support the following conclusions regarding the hypotheses tested:

1. Suggest that large and metropolitan counties cannot always be estimated with more accuracy.
2. Support the hypothesis that fast growing counties should be consistently underestimated.
3. Do not support the hypothesis that Population Loss counties should be consistently overestimated.
4. Support the hypothesis that average error varies directly

with rates of population growth. Stable and Slow Growth counties have the lowest average error.

5. Support the hypothesis that Stable or Slow Growth counties can be estimated with a higher degree of reliability than Fast Growth or Population Loss counties.

## CHAPTER V

### THE COMPOSITE METHOD

This chapter contains the procedural steps used in the Composite Method to develop separate estimates of the population ages 45 years and older, ages 15 to 44 years, and 0 to 14 years. The method is applied to the 36 counties of Oregon, its results are evaluated and the degree to which the hypotheses are supported is assessed.

#### I. PROCEDURAL STEPS

##### A. Population 45 years old and over.

1. Compute age-sex specific death rates (15) by ten year age groups for the census year (1960), on the basis of the census population (19), starting with the population 45 to 54 years and ending with the population 85 years old and over, for the U. S. and for the State.
2. Compute the corresponding death rates for the country for the estimate year (15).
3. Prepare estimates of the age-sex specific death rates for the State for the estimate year, on the assumption that the change in the death rate for the State from the census year was of the same ratio as for the U. S. as a whole.
4. Compute the estimated population for the State on the estimate date in each age-sex group by dividing the number of deaths for each group during the estimate year by the estimated specific death rate obtained above.

5. For Oregon counties, deaths are available by age only.

Also, the small number of deaths in some counties makes use as bases for estimates somewhat questionable. Therefore, the procedure is modified so that estimates for counties are prepared using total age groups and three year (centered) averages of age-specific deaths. The estimates are adjusted to the State estimate developed above.

B. Population 15 to 44 years of age.

1. Estimates of the number of females 15 to 44 years old are first developed in a manner corresponding to steps (1) through (5) above using data on the number of births, by age of mother (15), and the number of females 15 to 44 years of age (19). As above, this method assumes that the change in age-specific birth rates for the State and its counties was of the same ratio as for the U. S. during the period.
2. The ratio of the number of males to females at the last census in the State and each county in this age range adjusted for the change in this ratio for the U. S. as a whole during the period (20) is used to arrive at an estimate of the number of males in the State and each county. The county estimates are summed and adjusted to agree with the State estimates.
3. Estimates of male and females are summed to yield an estimate of the total population 15 to 44 years of age.

C. Population under 15 years of age.

1. The population 6-14 years is derived earlier as part of the Component Method II procedure which applies the censal ratio to school enrollment on the estimate date.
2. Add births six years prior to the estimate date.
3. Subtract deaths to the cohort for the same period.
4. Add an estimate of net migration.
  - a. A migration exposure factor of .5 is used under the assumption that the births and deaths are exposed to only one-half the migration probability period.
  - b. The factor used to convert the school age migration rate to the rate for the population under six years of age is .3630 (14).

- D. An estimate of total population is derived by summing estimates A, B, and C.

## II. EVALUATION OF THE RESULTS

Table VIII reveals that while the Composite State estimate is 3.61 percent below the census count, the over- and underestimates are evenly split among counties. Of course, the negative magnitudes are larger overall and, specifically, in the larger counties. Errors range from -11.67 percent to +48.67 percent.

Composite underestimates all but one of the Metropolitan counties as seen in Table IX. Other Urban counties are estimated best of any density type except in one case (-10.99 percent). The remaining eight counties deviate by less than five percent. Five of nine are plus deviations. With the exception of Wheeler County (+48.67 percent), the

TABLE VIII  
 Summary of Deviations  
 of 1970 Composite Estimates  
 for Oregon Counties

	1970 Census	1970 Composite Estimate	Absolute Deviation	Percent Deviation
<b>State</b>	<b>2,091,533</b>	<b>2,016,062</b>	<b>-75,471</b>	<b>- 3.61</b>
Baker	14,919	14,645	- 274	- 1.84
Benton	53,776	47,866	- 5,910	-10.99
Clackamas	166,088	146,701	-19,387	-11.67
Clatsop	28,473	27,050	- 1,423	- 4.99
Columbia	28,790	28,718	- 72	- 0.25
Coos	56,515	58,559	2,044	3.62
Crook	9,985	10,100	115	1.15
Curry	13,006	12,098	- 908	- 6.98
Deschutes	30,442	31,162	720	2.37
Douglas	71,743	72,318	575	0.80
Gilliam	2,342	2,260	- 82	- 3.50
Grant	6,996	7,357	361	5.16
Harney	7,215	7,837	622	8.62
Hood River	13,187	12,675	- 512	- 3.88
Jackson	94,533	92,701	- 1,832	- 1.94
Jefferson	8,548	9,088	540	6.32
Josephine	35,746	37,383	1,637	4.58
Klamath	50,021	50,412	391	0.78
Lake	6,343	6,025	- 318	- 5.01
Lane	215,401	194,663	-20,738	- 9.63
Lincoln	25,755	24,859	- 896	- 3.48
Linn	71,914	73,231	1,317	1.83
Malheur	23,169	24,010	841	3.63
Marion	151,309	158,786	7,477	4.94
Morrow	4,465	4,458	- 7	- 0.16
Multnomah	554,668	531,492	-23,176	- 4.18
Polk	35,349	31,585	- 3,764	-10.65
Sherman	2,139	2,007	- 132	- 6.17
Tillamook	18,034	16,786	- 1,248	- 6.92
Umatilla	44,923	46,244	1,321	2.94
Union	19,377	19,747	370	1.91
Wallowa	6,247	5,706	- 541	- 8.66
Wasco	20,133	21,033	900	4.47
Washington	157,920	142,455	-15,465	- 9.79
Wheeler	1,849	2,749	900	48.67
Yamhill	40,273	41,296	1,083	2.69

TABLE IX

Summary of Percent Deviations  
of 1970 Composite Estimates  
for Oregon Counties, by Type

County	Metro- politan Counties	Other Urban Counties	Rural Counties	Moderate to Fast Growth Counties	Stable and Slow Growth Counties	Popula- tion Loss Counties
Baker		-1.84				-1.84
Benton		-10.99		-10.99		
Clackamas	-11.67			-11.67		
Clatsop		-4.99			-4.99	
Columbia			-0.25	-0.25		
Coos		3.62			3.62	
Crook			1.15		1.15	
Curry			-6.98			-6.98
Deschutes		2.37		2.37		
Douglas			0.80		0.80	
Gilliam			-3.50			-3.50
Grant			5.16			5.16
Harney			8.62		8.62	
Hood River			-3.88		-3.88	
Jackson		-1.94		-1.94		
Jefferson			6.32	6.32		
Josephine		4.58		4.58		
Klamath		0.78			0.78	
Lake			-5.01			-5.01
Lane	-9.63			-9.63		
Lincoln			-3.48		-3.48	
Linn			1.83	1.83		
Malheur			3.63		3.63	
Marion	4.94			4.94		
Morrow			-0.16			-0.16
Multnomah	-4.18				-4.18	
Polk	-10.65			-10.65		
Sherman			-6.17			-6.17
Tillamook			-6.92			-6.92
Umatilla			2.94		2.94	
Union			1.91		1.91	
Wallowa			-8.66			-8.66
Wasco		4.47			4.47	
Washington	-9.79			-9.79		
Wheeler			48.67			48.67
Yamhill			2.69	2.69		



smallest county in the State with 1849 population, the range of deviations for Rural counties is -8.66 percent to +8.62 percent. There is no directional bias in this class (11 plus, 10 minus).

There is an inversion in the growth rate typology in terms of directional bias. Composite does not clearly underestimate, as expected, in the Moderate to Fast Growth class, but does in eight of ten cases in the Population Loss class. While the Stable to Slow Growth counties are overestimated in nine of thirteen cases, the deviations are less than five percent, except in one case (+8.62 percent).

The results shown in Table X do not commend the Composite Method very highly. Only in two classes does the average deviation fall below five percent. Part of the problem lies with the large error in Wheeler County. The statistics in parenthesis do not include Wheeler County and illustrate how any county, regardless of size, can bias the results. Even so, the "controlled" results are not too exciting. The Stable to Slow Growth counties is where the method is strongest as it produces deviations of five percent or more in only one of thirteen cases. The average deviation of 3.42 percent and standard deviation of 2.10 percent suggest the method might be applied to this class of counties in the future with a reasonable degree of confidence.

At this point it should be stated that the more extensive analysis in this chapter is not intended to favor the Composite Method over the others selected. However, the nature of the Composite Method does lend itself to a more specific assessment of the contributions to estimate error.

Table XI reveals that the largest contribution to the error in the State estimate is in the estimate of the 15-44 age group (-7.05 percent).

TABLE X  
 Evaluation of 1970 Composite Estimates  
 for Oregon Counties, by Type

County Type	Number of Counties	Number of Positive Deviations	Number of Deviations $\geq 5\%$	Number of Deviations $\geq 10\%$	Average Percent Deviation	Standard Deviation
All	36	17	14	4	5.98 (4.76)	7.98 (3.23)
Metropolitan	6	1	4	2	8.48	3.13
Other Urban	9	5	1	1	3.95	3.01
Rural	21	11	9	1	6.13 (4.00)	10.08 (2.64)
Moderate to Fast Growth	13	6	6	3	5.97	4.09
Stable to Slow Growth	13	9	1	0	3.42	2.10
Population Loss	10	2	7	1	9.31 (4.93)	14.06 (2.69)

TABLE XI

Summary of Percent Deviations  
of 1970 Composite Age Group Estimates  
for Oregon Counties

	0-14 Actual	0-14 Estimated	Percent Deviation	15-44 Actual	15-44 Estimated	Percent Deviation
State	569,727	562,812	- 1.21	846,415	786,742	- 7.05
Baker	4,138	4,292	3.72	5,213	4,801	- 7.90
Benton	12,470	12,415	- 0.44	29,599	23,936	-19.13
Clackamas	49,239	47,230	- 4.08	67,471	57,571	-14.67
Clatsop	6,876	6,796	- 1.16	10,502	9,101	-13.34
Columbia	8,690	8,310	- 4.37	11,087	11,373	2.58
Coos	16,624	16,425	- 1.20	22,232	23,636	6.32
Crook	2,788	2,627	- 5.77	3,782	3,558	- 5.92
Curry	3,649	3,662	0.36	4,650	3,764	-19.05
Deschutes	8,528	8,342	- 2.18	11,610	10,747	- 7.43
Douglas	21,523	21,602	0.37	28,125	28,277	0.54
Gilliam	663	729	9.95	897	836	- 6.80
Grant	1,955	1,949	- 0.30	2,643	2,443	- 7.57
Harney	2,092	2,185	4.45	2,899	2,613	- 9.87
Hood River	3,580	3,585	0.14	4,741	4,443	- 6.29
Jackson	25,147	24,693	- 1.81	37,151	36,990	- 0.43
Jefferson	2,798	3,224	15.23	3,366	3,108	- 7.66
Josephine	9,353	9,548	2.08	12,144	13,760	13.31
Klamath	14,438	14,193	- 1.70	20,542	20,522	- 0.10
Lake	1,797	1,702	- 5.29	2,456	2,423	- 1.34
Lane	60,015	59,909	- 0.18	95,879	79,030	-17.57
Lincoln	6,204	6,212	0.13	8,551	7,690	-10.07
Linn	21,575	20,386	- 5.51	28,544	30,340	6.29
Malheur	7,064	7,088	0.34	8,722	8,771	0.56
Marion	41,980	42,386	0.97	59,492	64,464	8.36
Morrow	1,246	1,148	- 7.87	1,607	1,762	9.65
Multnomah	136,842	138,325	1.08	221,404	201,609	- 8.94
Polk	9,321	8,682	- 6.86	14,776	13,079	-11.48
Sherman	556	560	0.72	773	751	- 2.85
Tillamook	4,917	4,661	- 5.21	6,386	5,450	-14.66
Umatilla	12,080	11,906	- 1.44	17,301	17,417	0.67
Union	5,101	4,906	- 3.82	7,864	8,262	5.06
Wallowa	1,637	1,548	- 5.44	2,232	1,983	-11.16
Wasco	5,496	5,680	3.35	7,447	7,495	0.64
Washington	47,898	44,809	- 6.45	68,068	56,862	-16.46
Wheeler	482	431	-10.58	686	829	20.85
Yamhill	10,965	10,666	- 2.73	15,573	17,046	9.46

TABLE XI (con't.)  
 Summary of Percent Deviations  
 of 1970 Composite Age Group Estimates  
 for Oregon Counties

	45+ Actual	45+ Estimated	Percent Deviation
State	675,391	666,508	- 1.32
Baker	5,568	5,552	- 0.29
Benton	11,707	11,515	- 1.64
Clackamas	49,378	41,900	-15.16
Clatsop	11,095	11,153	0.52
Columbia	9,013	9,035	0.24
Coos	17,659	18,498	4.75
Crook	3,415	3,915	14.64
Curry	4,707	4,672	- 0.74
Deschutes	10,304	12,073	17.17
Douglas	22,095	22,439	1.56
Gilliam	782	695	-11.13
Grant	2,398	2,965	23.64
Harney	2,224	3,039	36.65
Hood River	4,866	4,647	- 4.50
Jackson	32,235	31,018	- 3.78
Jefferson	2,384	2,756	15.60
Josephine	14,249	14,075	- 1.22
Klamath	15,041	15,697	4.36
Lake	2,090	1,900	- 9.09
Lane	59,507	55,724	- 6.36
Lincoln	11,000	10,957	- 0.39
Linn	21,795	22,505	3.26
Malheur	7,383	8,151	10.40
Marion	49,837	51,936	4.21
Morrow	1,612	1,548	- 3.97
Multnomah	196,422	191,558	- 2.48
Polk	11,252	9,824	-12.69
Sherman	810	696	-14.07
Tillamook	6,731	6,675	- 0.83
Umatilla	15,542	16,921	8.87
Union	6,412	6,579	2.60
Wallowa	2,378	2,175	- 8.54
Wasco	7,190	7,858	9.29
Washington	41,954	40,784	- 2.79
Wheeler	681	1,489	118.65
Yamhill	13,675	13,584	- 0.67

While the errors in remaining two age groups are close at the State level (-1.21 percent and -1.32 percent), a glance at the county deviations reveals quite a scatter in the 45 and over age group. The range is from -15.14 percent to +118.65 percent (Wheeler County).

Table XII reveals that the only adequate methodological approach in the Composite Method is the application of Component Method II to the 0-14 age group. Average error for the other two age groups falls below five percent in only one class of counties, and then, just barely. The 0-14 estimates are reasonably accurate overall and for all classes (none  $\geq$  five percent average deviation). The method performs especially well in Other Urban counties and in Stable to Slow Growth counties. In the former, all nine counties deviate by less than five percent with overall accuracy being reflected in the low average deviation of 1.96 percent. This method is highly reliable in this class as shown by the standard deviation of 1.04. Only one of thirteen cases in the latter class exceeds five percent. Comparable accuracy is reflected by the average deviation of 1.92 percent and the standard deviation of 1.83.

Part of the problem with the 15-44 age group estimates is revealed by Table XIII. That the deviations among age groups between actual and estimated birth rates do not compensate is obvious. Note that the biggest deviations occur in the 20-24 and the 25-29 age groups, those with the highest fertility. Births are events of relatively small magnitude compared with the age group producing them. This, in conjunction with rapidly declining birth rates, can produce rather large errors with only small variations in the numbers of births. Apparently, errors in the estimated sex ratios were not of a compensating nature.

TABLE XII

Evaluation of 1970 Composite Age Group Estimates  
for Oregon Counties, by Type

County Type and Age Group	Number of Counties	Number of Positive Deviations	Number of Deviations ≥ 5%	Number of Deviations ≥ 10%	Average Percent Deviation	Standard Deviation
<u>All</u>	36					
0-14		14	11	2	3.53	3.46
15-44		13	27	12	8.47	5.88
45 +		17	16	11	10.47	18.68
					(7.37)	(7.83)
<u>Metropolitan</u>	6					
0-14		2	2	0	3.27	2.94
15-44		1	6	4	12.91	3.90
45 +		1	3	2	7.28	5.38
<u>Other Urban</u>	9					
0-14		3	0	0	1.96	1.04
15-44		3	6	3	7.62	6.67
45 +		5	2	1	4.78	5.43
<u>Rural</u>	21					
0-14		9	9	2	4.29	4.07
15-44		9	15	5	7.57	5.64
45 +		11	11	8	13.81	24.17
					(8.57)	(9.21)
<u>Moderate to Fast Growth</u>	13					
0-14		3	4	1	4.07	4.00
15-44		5	11	6	10.37	5.71
45 +		5	5	4	6.52	6.27
<u>Stable to Slow Growth</u>	13					
0-14		7	1	0	1.92	1.83
15-44		6	8	2	5.25	4.46
45 +		10	5	3	7.77	9.67
<u>Population Loss</u>	10					
0-14		4	6	1	4.94	3.77
15-44		2	8	4	10.18	6.40
45 +		2	6	4	19.10	34.81
					(8.03)	(7.69)

TABLE XIII  
Comparison of Actual and Estimated  
Age-Specific Birth Rates  
State of Oregon, 1970

Age Group	Births Per 1,000 Women	
	Actual	Estimated
15-19	58.9	61.4
20-24	167.5	172.5
25-29	139.4	130.5
30-34	58.3	57.1
35-39	21.7	23.8
40-44	5.4	6.6

The accuracy of the estimate of the State total of persons ages 45 and over is explained by the close estimates of death rates shown in Table XIV. Even though these rates are not changing rapidly, the numbers are of such a small magnitude that the large errors can result from the assumption that rates for counties change at the same ratio to the State, as well as from small variations in deaths, even if the death rates are well estimated.

Tables XV and XVI reflect the Composite estimates adjusted to the Method II State estimate. This was done for the purpose of comparability in the averaging of results below. About the only comment worth making is that, surprisingly, the accuracy worsened, slightly, across counties even though the adjustment brought them much closer to the State total. This is due to the effect of the positive deviations increasing more overall than the negative deviations decreased.

The results above support the following conclusions regarding the hypotheses tested:

1. Partially support the hypothesis that more accurate estimates can be obtained for larger and more urban counties. The inversion the Metropolitan counties contradicts a linear relationship, however.
2. Contradict the hypothesis regarding directional bias. Slightly more than half the Moderate to Fast Growth counties are overestimated and 80 percent of the Population Loss counties are underestimated.
3. Support the hypothesis that average error varies directly with rates of population growth. Slowly growing and Stable counties have the lowest average error.



TABLE XIV  
 Comparison of Actual and Estimated  
 Age/Sex Specific Death Rates  
 State of Oregon, 1970

Age Group	Deaths Per 1,000 Persons			
	Male		Female	
	Actual	Estimated	Actual	Estimated
45-54	8.32	8.83	4.32	4.33
55-64	20.19	21.53	9.54	9.44
65-74	43.75	42.37	22.54	22.09
75-84	96.74	92.70	59.86	60.18
85 +	172.62	186.28	144.95	152.42

TABLE XV

Summary of Percent Deviations  
of 1970 Composite Estimates for Oregon Counties  
Adjusted to Method II Estimated State Total

	1970 Census	1970 Composite Adjusted	Percent Deviation
State	2,091,533	2,094,997	
Baker	14,919	15,218	2.00
Benton	53,776	49,740	- 7.51
Clackamas	166,088	152,445	- 8.21
Clatsop	28,473	28,109	- 1.28
Columbia	28,790	29,842	3.65
Coos	56,515	60,852	7.67
Crook	9,985	10,495	5.11
Curry	13,006	12,572	- 3.34
Deschutes	30,442	32,382	6.37
Douglas	71,743	75,149	4.75
Gilliam	2,342	2,348	0.26
Grant	6,996	7,645	9.28
Harney	7,215	8,144	12.88
Hood River	13,187	13,171	- 0.12
Jackson	94,533	96,331	1.90
Jefferson	8,548	9,444	10.48
Josephine	35,746	38,847	8.68
Klamath	50,021	52,386	4.73
Lake	6,343	6,261	- 1.29
Lane	215,401	202,285	- 6.09
Lincoln	25,755	25,832	0.30
Linn	71,914	76,098	5.82
Malheur	23,169	24,950	7.69
Marion	151,309	165,003	9.05
Morrow	4,465	4,633	3.76
Multnomah	554,668	552,300	- 0.43
Polk	35,349	32,822	- 7.15
Sherman	2,139	2,086	- 2.48
Tillamook	18,034	17,443	- 3.28
Umatilla	44,923	48,055	6.97
Union	19,377	20,520	5.90
Wallowa	6,247	5,929	- 5.09
Wasco	20,133	21,857	8.56
Washington	157,920	148,033	- 6.26
Wheeler	1,849	2,857	54.52
Yamhill	40,213	42,913	6.71

TABLE XVI  
 Evaluation of 1970 Adjusted Composite Estimates  
 for Oregon Counties, by Type

County Type	Number of Counties	Number of Positive Deviations	Number of Deviations $\geq 5\%$	Number of Deviations $\geq 10\%$	Average Percent Deviation	Standard Deviation
All	36	23	21	3	6.66 (5.29)	7.02 (3.21)
Metropolitan	6	1	5	0	6.20	3.05
Other Urban	9	7	5	0	5.41	3.01
Rural	21	15	11	3	7.32 (4.96)	10.18 (3.43)
Moderate to Fast Growth	13	8	11	1	6.76	2.25
Stable to Slow Growth	13	10	7	1	4.93	5.13
Population Loss	10	5	3	1	8.53 (3.42)	16.54 (5.22)

4. Support the hypothesis that Stable or Slow Growth counties can be estimated with a higher degree of reliability than Fast Growth or Population Loss counties.
5. Partially support the hypothesis that total population estimates should be more accurate than age group estimates, especially where direct data is used only to obtain part of the age group and a compound logic is used to obtain the remainder. The exception of high accuracy in the 0-14 age group is explained by the next conclusion.
6. Support the hypothesis that estimates which rely heavily on data representing a small number of events will be less accurate than those using data series with magnitudes approaching the population size.
7. Following the point above, the assumption that the relationship between school enrollment and school age population holds over time better than the relationship between school age migration and total migration is supported by the relatively greater accuracy of the 0-14 estimates compared with Method II estimates of total population. Part of this is due, no doubt, to the application of the national average intercounty migration ratio to all Oregon counties.

## CHAPTER VI

### AVERAGING RESULTS

Due to the absence of a precedent of a previous study which would indicate accuracy characteristics of the selected methods when applied to Oregon counties, two sets of averages are presented here which combine the results of methods which employ essentially different indicators, procedures and/or assumptions. The first set merges the results of Component Method II and Ratio Correlation. The second set of averages is of the results of Ratio Correlation and Composite (adjusted). Table XVII presents a summary comparison by county.

We can see from Table XVIII that both sets of averages offset partly the directional bias of their constituent estimates. Both, however, retain an upward bias. Average I overestimates consistently in four of six classes of counties. Average II is somewhat better with chronic overstatement in two of six categories.

Both averages have a similar number of extreme errors ( $\geq 5$  percent), although three Average I estimates exceed ten percent compared with one case for Average II. The bulk of the five percent plus errors for both sets occur in Rural and Moderate to Fast Growth counties, although five of twelve of these Average II excesses occur in Metropolitan counties (class total = 6).

That averaging offsets variances in at least one of the methods can be seen in the overall average deviations of both sets. Both fall below the magic five percent. It will be recalled that the comparable statistics for both Component Method II and Composite exceeded this standard.

TABLE XVII

Summary Comparison of Percent Deviations  
of Selected 1970 Average Estimates  
for Oregon Counties

	1970 Census	$\bar{X}_1$ (Meth II and Ratio Correlation)	$\bar{X}_2$ (Ratio Correlation and Composite 2)	% Deviations $\bar{X}_1$	% Deviations $\bar{X}_2$
State	2,091,533				
Baker	14,919	16,548	15,486	10.92	3.80
Benton	53,776	50,419	49,622	- 6.24	- 7.72
Clackamas	166,088	166,172	157,105	0.05	- 5.41
Clatsop	28,473	28,981	28,483	1.78	0.04
Columbia	28,790	30,410	30,242	5.63	5.04
Coos	56,515	57,943	58,989	2.53	4.38
Crook	9,985	9,866	9,957	- 1.19	- 0.28
Curry	13,006	13,291	12,567	2.19	- 3.38
Deschutes	30,442	29,834	30,723	- 2.00	0.92
Douglas	71,743	75,056	73,940	4.62	3.06
Gilliam	2,342	2,668	2,429	13.92	3.71
Grant	6,996	7,207	7,280	3.02	4.06
Harney	7,215	7,286	7,741	0.98	7.29
Hood River	13,187	14,107	13,343	6.98	1.18
Jackson	94,533	93,979	93,986	- 0.59	- 0.58
Jefferson	8,548	8,801	8,986	2.96	5.12
Josephine	35,746	35,787	36,885	0.11	3.19
Klamath	50,021	50,721	51,397	1.40	2.75
Lake	6,343	6,629	6,269	4.51	- 1.17
Lane	215,401	209,495	204,447	- 2.74	- 5.09
Lincoln	25,755	25,186	25,704	- 2.21	- 0.20
Linn	71,914	70,410	72,577	- 2.09	0.92
Malheur	23,169	24,736	24,383	6.76	5.24
Marion	151,309	157,301	161,630	3.96	6.82
Morrow	4,465	4,537	4,418	1.61	- 1.05
Multnomah	554,668	560,733	569,994	1.09	2.76
Polk	35,349	33,379	32,061	- 5.57	- 9.30
Sherman	2,139	2,213	2,102	3.46	- 1.73
Tillamook	18,034	18,082	17,619	0.27	- 2.30
Umatilla	44,923	45,827	46,726	2.01	4.01
Union	19,377	20,206	20,131	4.28	3.89
Wallowa	6,247	6,387	6,040	2.24	- 3.31
Wasco	20,133	21,273	21,231	5.66	5.45
Washington	157,920	145,966	146,975	- 7.57	- 6.93
Wheeler	1,849	2,070	2,428	11.95	31.31
Yamhill	40,213	41,491	41,101	3.18	2.21

TABLE XVIII

Comparative Evaluation of 1970 Selected Average Estimates  
for Oregon Counties, by Type

County Type	Number of Counties	Number of Positive Deviations	Number of Deviations $\geq 5\%$	Number of Deviations $\geq 10\%$	Average Percent Deviation	Standard Deviation
<u>All</u>	36					
Average I		27	10	3	3.84	3.29
Average II		22	12	1	4.32	5.17
<u>Metropolitan</u>	6					
Average I		3	2	0	3.50	2.80
Average II		2	5	0	6.05	2.20
<u>Other Urban</u>	9					
Average I		6	3	1	3.47	3.50
Average II		7	2	0	3.20	2.49
<u>Rural</u>	21					
Average I		18	5	2	4.10	3.45
Average II		13	5	1	4.31	6.46
<u>Moderate to Fast Growth</u>	13					
Average I		6	4	0	3.28	2.41
Average II		7	8	0	4.56	2.81
<u>Stable to Slow Growth</u>	13					
Average I		11	3	0	3.19	2.19
Average II		11	3	0	3.12	2.24
<u>Population Loss</u>	10					
Average I		10	3	3	5.41	4.91
Average II		4	1	1	5.58	9.11

Average I fares better of the two sets overall. It is "one-half" percent more accurate overall and substantially more reliable (S.D. =  $3.29 < 5.17$ ).

Both sets are fairly efficient in the Other Urban and Stable and Slow Growth counties. On this criterion, Average I performs fairly well also, and quite better than Average II, in Metropolitan and in Moderate to Fast Growth counties.

Both averages produce consistent deviations in most classes of counties. Average II, however, has especially low reliability in Rural and Population Loss counties.

The above results support the following conclusions regarding the hypotheses tested:

1. Averaging reduces extreme error in most cases.
2. Opposite bias is not controlled to the extent of even distribution.
3. Support in three cases and contradict in one the hypothesis that average error should be lower in large or metropolitan or primarily urban counties.
4. Fail to support in most cases the directional bias hypothesis that expects more of it in population change counties. Partial support is found, however, in that Average I overestimated all ten Population Loss counties.
5. Support the hypothesis that Slow Growth or Stable counties have the lowest average error. Average I performs almost as well on Moderate to Fast Growth counties.
6. Support the hypothesis that Stable or Slow Growth counties can be estimated with a higher degree of relia-



bility than Fast Growth or Population Loss counties.

## CHAPTER VII

### SUMMARY AND CONCLUSIONS

Table XIX shows clearly that Ratio Correlation is the most accurate of the selected methods over all Oregon counties. No other approach produces less directional bias, less extreme deviations, lower average deviation or a higher degree of reliability.

The six Metropolitan counties are best represented by Component Method II as judged by its performance on all the accuracy criteria.

Comparable accuracy is achieved in Other Urban counties by both Ratio Correlation and Average II (Ratio Correlation and Composite 2), although Ratio Correlation does not produce the directional bias of the average.

Ratio Correlation is also superior in the Rural counties based on its results which produce virtually no consistent directional bias or extreme errors. The method scores extremely well on overall accuracy and reliability, especially in light of the various aberrations produced by the other approaches.

The best results in Moderate to Fast Growth counties is produced by Average I (Method II and Ratio Correlation), although the Method II results are similar and adequate. Average I is slightly more reliable and evidences no particular directional bias.

All the approaches produce consistent overestimates in the Stable to Slow Growth counties. On the other evaluative criteria, however, Ratio Correlation performs very well. The best single performance occurs here with an average deviation of 2.24 percent and tightly

TABLE XIX

Comparative Evaluation of All 1970 Selected Estimates  
for Oregon Counties, by Type

County Type and Method	Number of Counties	Number of Positive Deviations	Number of Deviations ≥ 5%	Number of Deviations ≥ 10%	Average Percent Deviation	Standard Deviation
<u>All</u>	36					
1. Method II		30	17	6	5.95	4.82
2. Ratio Corre.		16	10	1	3.51	2.69
3. Composite 1		17	14	4	5.98	7.98
4. Composite 2		23	21	3	6.66	7.02
5. $\bar{X}_1(X_1 X_2)$		27	10	3	3.84	3.29
6. $\bar{X}_2(X_2 X_4)$		22	12	1	4.32	5.17
<u>Metropolitan</u>	6					
1. Method II		3	1	0	3.02	2.65
2. Ratio Corre.		2	3	1	6.05	3.15
3. Composite 1		1	4	2	8.48	3.13
4. Composite 2		1	5	0	6.20	3.05
5. $\bar{X}_1(X_1 X_2)$		3	2	0	3.50	2.80
6. $\bar{X}_2(X_2 X_4)$		2	5	0	6.05	2.20
<u>Other Urban</u>	9					
1. Method II		8	2	1	4.77	4.94
2. Ratio Corre.		5	2	0	3.22	2.38
3. Composite 1		5	1	1	3.95	3.01
4. Composite 2		7	5	0	5.41	3.01
5. $\bar{X}_1(X_1 X_2)$		6	3	1	3.47	3.50
6. $\bar{X}_2(X_2 X_4)$		7	2	0	3.20	2.49
<u>Rural</u>	21					
1. Method II		19	14	5	7.29	4.90
2. Ratio Corre.		9	5	0	2.92	2.36
3. Composite 1		11	9	1	6.13	10.08
4. Composite 2		15	11	3	7.32	10.18
5. $\bar{X}_1(X_1 X_2)$		18	5	2	4.10	3.45
6. $\bar{X}_2(X_2 X_4)$		13	5	1	4.31	6.46

TABLE XIX (con't.)

Comparative Evaluation of All 1970 Selected Estimates  
for Oregon Counties, by Type

County Type and Method	Number of Counties	Number of Positive Deviations	Number of Devia- tions ≥ 5%	Number of Devia- tions ≥10%	Average Percent Deviation	Standard Deviation
<u>Moderate to Fast Growth</u>	13					
1. Method II		9	3	0	3.35	2.82
2. Ratio Corre.		2	4	1	4.70	2.99
3. Composite 1		6	6	3	5.97	4.09
4. Composite 2		8	11	1	6.76	2.25
5. $\bar{X}_1$ ( $X_1$ $X_2$ )		6	4	0	3.28	2.41
6. $\bar{X}_2$ ( $X_2$ $X_4$ )		7	8	0	4.56	2.81
<u>Stable to Slow Growth</u>	13					
1. Method II		11	5	2	5.22	3.55
2. Ratio Corre.		11	2	0	2.24	1.71
3. Composite 1		9	1	0	3.42	2.10
4. Composite 2		10	7	1	4.93	5.13
5. $\bar{X}_1$ ( $X_1$ $X_2$ )		11	3	0	3.19	2.19
6. $\bar{X}_2$ ( $X_2$ $X_4$ )		11	3	0	3.12	2.24
<u>Population Loss</u>	10					
1. Method II		10	9	4	10.25	5.63
2. Ratio Corre.		3	4	0	3.62	2.81
3. Composite 1		2	7	1	9.31	14.06
4. Composite 2		5	3	1	8.53	16.54
5. $\bar{X}_1$ ( $X_1$ $X_2$ )		10	3	3	5.41	4.91
6. $\bar{X}_2$ ( $X_2$ $X_4$ )		4	1	1	5.58	9.11

clustered results evidenced by a standard deviation of 1.71. Very adequate results are obtained also from the unadjusted Composite and both sets of averages.

The difficulty of obtaining accurate population estimates for counties which are losing population is well illustrated by the widely discrepant results produced by all approaches with the exception of Ratio Correlation. While the method overestimates at a rate of 70 percent and four of ten excessive deviations is not outstanding, the average deviation and the standard deviation are fairly low.

The results of this study warrant the following conclusions:

1. They do not support the hypothesis that average error should be lower in large or Metropolitan or in primarily Urban counties. In fact, the best overall method, Ratio Correlation, produces just the opposite results.
2. Do not support the hypothesis that Moderate to Fast Growth counties should be consistently underestimated and Population Loss counties consistently overestimated.
3. In general, Stable and Slow Growth counties have the lowest average error.
4. In general, Stable or Slow Growth counties are not estimated with an appreciably higher degree of reliability than Moderate to Fast Growth counties. Estimates for both these classes are much more reliable than those for Population Loss counties.
5. Total population estimates are more accurate, in general, than estimates for age groups which employ a compound logic to get an estimate of one component of the group

from an estimate of another component.

6. The accuracy of the 0-14 age group estimates, the inaccuracy of the 45 and over age group and accurate performance of the Ratio Correlation method are, in part, related directly to the closeness in magnitude of the indicators and the target population.
7. Challenge the validity of assuming factors which apply to the whole also apply to smaller, constituent parts (i.e., the applicability of the national average inter-county migration ratio to Oregon counties, or, an across-the-board ratio of change in vital rates).
8. Do not support the hypothesis that consistently lower errors should be attained by averaging estimates made by different methods.
9. Do not support the hypothesis that no single method should provide consistently greater accuracy. Ratio Correlation is better overall and in four of six county classes.
10. Finally, the results of this study establish the importance of, at least, area specific evaluations of various methods prior to selection of the most appropriate battery of estimates for that area and its constituent parts.

In conclusion, it is suggested that, while other comparative studies have not unveiled a method which is consistently more accurate than others, a major research emphasis be placed on obtaining additional and/or better data series for use in the Ratio Correlation Method. Also,

where reporting lag is a problem, efforts to extend data series into the near future using a variety of sophisticated models might prove to be fruitful.

The other methods studied did not perform as well on the bases of these tests, suggesting that strong caution be exercised in the application of their results. However, it is felt that their primary utility is in the form of providing benchmarks or reasonable limits from which to assess the face validity of various Ratio Correlation combinations.

## SOURCES CONSULTED

- 1 Jaffe, A. J., Handbook of Statistical Methods for Demographers, U. S. Bureau of the Census, Washington, D. C., 1960, 211-227.
- 2 Shryock, Henry S. and Jacob S. Siegel, The Methods and Materials of Demography, U. S. Department of Commerce, Bureau of the Census (in press, 1972). Preliminary draft, Chapter 23, 1-42.
- 3 Morrison, Peter A., Demographic Information for Cities: A Manual for Estimating and Projecting Population Characteristics, The Rand Corporation, 1970, 2-22.
- 4 Bogue, Donald J. and Beverly Duncan, "A Composite Method for Estimating Postcensal Population of Small Areas by Age, Sex, and Color," U. S. National Office of Vital Statistics, Vital Statistics - Special Reports, 47 (6), August 24, 1969.
- 5 Schmitt, Robert C. and Albert H. Crossetti, "Accuracy of Ratio-Correlation Method for Estimating Postcensal Population," Land Economics, 30 (3), August, 1954, 279-280.
- 6 Goldberg, David, et al., "Estimates of Population Change in Michigan: 1950-60," Michigan Population Studies, No. 1, Ann Arbor, Michigan: University of Michigan, 1960, 36-39.
- 7 U. S. Department of Health, Education and Welfare, Public Health Conference of Records and Statistics, Preliminary Report of the Study Group on Postcensal Population Estimates, June 11, 1962.
- 8 Zitter, Meyer and Henry S. Shryock, Jr., "Accuracy of Methods of Preparing Postcensal Population Estimates for States and Local Areas," Demography, 1 (1), 1964, 227-241.
- 9 Goldberg, David and T. R. Balakrishnan, "A Partial Evaluation of Four Estimating Techniques," Michigan Population Studies, No. 2, Ann Arbor, Michigan: University of Michigan, 1961.
- 10 Hillery, Jr., George A., "Postcensal Population Estimates for Small Areas: An Evaluation of Selected Techniques," KAFS RS-17, Lexington, Ky.: University of Kentucky, Agricultural Experimental Station, September, 1962.
- 11 Poole, Richard W., et al., "An Evaluation of Alternative Techniques for Estimating County Population in a Six-State Area," Economic Research Series, No. 3, Stillwater, Oklahoma: Oklahoma State University, 1965.
- 12 Spiegelman, Mortimer, Introduction to Demography, Cambridge, Mass.: Howard University Press, 1968, 393-417.



- 13 U. S. Bureau of Census, Current Population Reports, Series P-25, No. 339, June 6, 1966.
- 14 U. S. Bureau of the Census, Population Estimates and Projections Branch, Unpublished U. S. Controls, 1972.
- 15 Vital Statistics Summary, State of Oregon, Department of Health, Portland, Oregon.
- 16 U. S. Department of Health, Education and Welfare, Public Health Service, Oregon, State Life Tables: 1959-1961.
- 17 Survival rates computed for the total population were applied to the counties as well.
- 18 Department of Education, State of Oregon.
- 19 U. S. Department of Commerce, Bureau of the Census, "General Population Characteristics, Oregon," PC(1)-B, for U. S. and Oregon, 1960.
- 20 U. S. Bureau of the Census, "Projections of the Population of the United States, by Age and Sex (Interim Revisions): 1970-2020," Current Population Reports, Series P-25, No. 448, U. S. Government Printing Office, Washington, D. C., 1970.
- 21 Motor Vehicles Division, State of Oregon.
- 22 Secretary of State, State of Oregon.