OSU COLLEGE OF BUSINESS

economic research series number 3

NASA-63 (04)

GPO PRICE $ ______

CFSTI PRICE(S) $ ______

Hard copy (HC) 4.00

Microfiche (MF) 1.00

ff 653 July 65

AN EVALUATION OF ALTERNATIVE TECHNIQUES FOR ESTIMATING COUNTY POPULATION IN A SIX-STATE AREA

FACILITY FORM 602

N66 20031 (ACCESSION NUMBER)

(THRU)

140 (PAGES)

1 (CODE)

CR 70833 (NASA CR OR TMX OR AD NUMBER)

34 (CATEGORY)

OKLAHOMA STATE UNIVERSITY
STILLWATER, OKLAHOMA

AN EVALUATION OF ALTERNATIVE TECHNIQUES FOR ESTIMATING COUNTY POPULATION IN A SIX-STATE AREA

AN EVALUATION OF ALTERNATIVE TECHNIQUES FOR ESTIMATING COUNTY POPULATION IN A SIX-STATE AREA

by

Richard W. Poole
James D. Tarver
David White
William R. Gurley

College of Business
Oklahoma State University
Stillwater, Oklahoma

printed by

The Publishing and Printing Department
Oklahoma State University
Stillwater, Oklahoma

1966

price $3.50

Copies can be ordered from Business Extension, College of Business, Oklahoma State University, Stillwater, Oklahoma.

ACKNOWLEDGEMENTS

The authors are indebted to many persons at the local, state and national levels for their cooperation in various phases of this three-year study. Some supplied input data, others technical advice, others contributed toward the execution of the actual computations; others efficiently handled the numerous administrative details which made the actual work possible.

First of all, the professional staffs of the National and State and Local Population Estimates and Projections branches, Population Division, Bureau of the Census, contributed invaluable advice, unpublished data, and substantial amounts of time on various technical aspects in making these county population estimates possible. Grateful acknowledgement is due to the following persons: Henry S. Shryock, Jr., Assistant Chief, Population Division; Jacob S. Siegel, Chief, National Population Estimates and Projections Branch, and to Donald S. Akers, National Population Estimates and Projections Branch; to Meyer Zitter, Chief, State and Local Population Estimates and Projections Branch, and to Donald E. Starsinic, State and Local Population Estimates and Projections Branch. The Population Division, Bureau of the Census, also provided unpublished 1950 and 1960 population estimates of metropolitan counties in the United States which were used in the preliminary tests of accuracy.

Dr. John F. Putnam, Educational Records and Reports, School Administration Branch, Office of Education, Department of Health, Education, and Welfare, assisted in securing comparable annual county school enrollments in the six-state area. Moreover, officials in the state departments of public instruction in the six states supplied the annual school enrollments collected by their agencies. Credit is due the following persons who cooperated in supplying data: Frank W. Cannaday, Supervisor of Statistics, Arkansas State Department of Education; Arthur C. Anderson, Director, Research and Publications, Iowa State Department of Public Instruction; Lyell Ocobock, Accountant, Statistical Services Section, Kansas State Department of Public Intsruction; Bernard H. Voges, Director, School Finance and Statistics, Missouri State Department of Education; W. A. Schindler, Director of Statistical Services, State of Nebraska, Department of Education; and Clarence L. DeWees, Director, Finance Division, Oklahoma State Board of Education.

Acknowledgement is also made for the cooperation of the parochial school superintendents who supplied annual county school enrollments for Catholics, Lutherans, and Seventh-Day Adventists. Numerous school superintendents at the local, state and district offices graciously supplied tabulations of unpublished data required for this study.

In addition, area and national offices of the Bureau of Indian Affairs furnished annual elementary enrollments for Federal Indian schools.

Officials in each of the six state departments of vital statistics provided annual county resident births and deaths and made special tabulations of county deaths, by age. Those rendering generous assistance in this phase of data collection were the following persons: Miss Frances E. Kimzey, Principal Clerk, Statistical Division, Arkansas State Board of Health; Loren E. Chancellor, Director, Division of Vital Statistics, Iowa State Department of Health; Irvin G. Franzen, Director, Division of Vital Statistics, The Kansas State Board of Health; W. W. Marshall, Jr., Director, Statistical Services, Division of Health of Missouri; Miss Freda Theis, Director, Department of Health, Nebraska State Bureau of Vital Statistics; Professor Margaret F. Shackelford, Biostatistical Unit, Department of Preventive Medicine and Public Health, University of Oklahoma Medical Center (formerly, Director, Division of Statistics, Oklahoma State Department of Health).

Acknowledgement is made to the personnel offices of the five branches of the Armed Forces (Departments of the Air Force, Army, Navy, Marine Corps and Coast Guard) for providing the number of servicemen stationed at each base in the six-state area on each estimate date. Some data were supplied by local military commanders and some by regional and national offices.

Mr. Paul Ijams, Assistant Secretary, Kansas State Board of Agriculture, furnished annual Kansas county population enumerations taken by the assessors in that State.

Credit is due Professor Margaret F. Shackelford for her contribution in programming the two composite methods of estimating county population — the Bogue-Duncan Composite Method and the Census Variation of the Composite Method. Professor Shackelford also com-

puted the April 1, 1960 county population estimates for Oklahoma by the Bogue-Duncan Composite Method.

Three programmers who contributed significantly in the actual computational phases of the county population estimates were Connie Boggs, Jeanie Hill, and Susan Thomas.

Finally, the administrative staffs of the Midwest Research Institute, Kansas City, Missouri and the Research Foundation, Oklahoma State University, deserve special mention for their efficient and courteous handling of the numerous tasks involving contractual arrangements. Among the many staff members of both agencies who expedited the day-to-day matters were James A. Alcott, Director of Economic Development, Midwest Research Institute; Dr. Marvin T. Edmison, Director, Research Foundation, Oklahoma State University; Miss Gretchen Pulver, Administrative Assistant, and Miss Eloise Dreessen, Office Manager, Research Foundation. Oklahoma State University.

Richard W. Poole
James D. Tarver
David White
William R. Gurley

Stillwater, Oklahoma
November, 1965

FOREWORD

This report is Number 3 of the Economic Research Series published by the College of Business, Oklahoma State University. The series is intended to make the results of significant faculty research, which may be of interest to many people, available to the general public as well as to professional colleagues. It follows, then, that studies in this series are not published on a regular time schedule.

The publications to date in the series relate to local, state and regional problems. In Report Number 1, *The Oklahoma Economy,* the authors provided an over-all integrated view and analysis of Oklahoma's economic development since statehood, with emphasis on the period since 1929. The study also identified promising areas for research and provided some background for such research. In Report Number 2, *Public Welfare in Oklahoma,* the author compared costs and trends in public welfare programs in Oklahoma with those in seven neighboring states.

The present study provides a detailed evaluation of alternative techniques for estimating county population. The evaluation tests were conducted for 564 counties in a six-state area. The methodological and empirical findings will be of particular interest to people in business, government and education who use county population estimates.

Richard W. Poole, Dean
College of Business
Oklahoma State University

November, 1965

CONTENTS

TABLES

PLATES

Chapter I

THE SIX-STATE STUDY

This publication is a by-product of a three-year regional research project encompassing the states of Arkansas, Iowa, Kansas, Missouri, Nebraska, and Oklahoma. It embodies only that part of the research relating to accuracy tests which were used to evaluate alternative methods of estimating county population. The major concern of this chapter is to provide a brief description of the overall six-state pilot project.

The six-state project was carried on with support through the Midwest Research Institute in Kansas City, Missouri (Subcontract No. 2571-1 under Prime Contract No. NASr-63(04)). It has been a cooperative project involving faculty members at the University of Arkansas, the University of Iowa, the University of Kansas, the University of Missouri, the University of Nebraska, Oklahoma State University, and the University of Oklahoma. The principal investigators have been W. Nelson Peach, Richard W. Poole, and James D. Tarver. The project was initiated formally in February, 1962.

A Basic Problem for Regional Analysis

Attention to regional analysis has grown in the post-World War II period. As a result, it has become well recognized that a most important shortcoming in regional analysis is the lack of adequate, comparable, reliable, comprehensive data on units smaller than the state level. During recent decades considerable progress has been made in improving economic and social data at the national and state levels. This improvement has made possible a corresponding improvement in the decision-making process by a wide variety of public and private agencies. Unfortunately, this program has not been paralleled by a comparable improvement in data for the areas smaller than the state.

Even where data for units smaller than the state government have been improved in a particular state, the regional analyst faces almost insuperable problems when he moves across state lines. Each of the 50 states has its own body of tax and spending laws. Some states have income taxes on persons and corporations; others do not. Some states have sales taxes; others do not. Even in the case of two states having taxes on income and sales, the taxes will vary with respect to such factors as rates and coverage. There are wide differences in the administrative machinery for handling statistical data among the various states. There are big differences in the interest shown in statistical data among agencies within a particular state and among the states.

The regional analyst finds himself in the unenviable position of having developed a body of skills and techniques but lacking the raw materials in the form of good data inputs on which to test his models. He is somewhat like a builder with highly skilled craftsmen and tools but without the steel, brick, lumber and nails required for putting up a building.

In summary, a basic problem confronting the regional investigator and/or decision maker is inadequate, incomparable, or nonexistent data. We believe that one of the essential next steps in regional analysis is the generation and collection of reasonably uniform, comprehensive data in a systematic framework for units smaller than the state level. The six-state project, of which this volume is one part of the resulting research, represents a modest effort along these lines. Our reference to the project as a pilot program reflects our conviction that such a framework and data collection system for regional analysis must eventually be nationwide. From the inception of the project, it has been our hope that the cooperative six-state pilot project will provide support for the emergence of a national program.[1]

---

[1]Recently regional economics was given an important boost by the Federal Government. A Regional Economics Division in the U.S. Department of Commerce was activated in early 1964. The primary function of the new division is to develop and maintain measures that reflect the current economic situation in the various regions of the nation and to provide a means of tracing regional economic development. To carry out this function, the division will measure and analyze factors responsible for geographic differences in levels of economic activity and in rates of economic growth and development. The new division prepares annual estimates of personal income by states. It is also in process of preparing estimates of personal income received by residents of Standard Metropolitan Statistical Areas where about 3/4 of the income of the nation is received. The division will also prepare quarterly estimates of personal income by states on a seasonally adjusted basis. Also in preparation are estimates of personal income in multi-county areas covering the entire nation. These estimates, prepared for selected years only, will be in about the same industrial and type-of-income detail as those now published by states. Another important activity of the division has been the completion of one phase of an analysis of changes in employment by counties between 1940, 1950 and 1960. In this analysis, which is carried out separately for 32 industries in each county, factors underlying changes in employment are identified and measured statistically. This analysis is now being prepared for publication in a nine-volume series covering all counties in the nation.

Two Basic Concepts

Before proceeding further, it will be advantageous to briefly consider the following two concepts: (1) the concept of region, and (2) the concept of regional building blocks. The overall research framework for the six-state project is based on our interpretation of these two concepts.

Region. Analytically we do not regard "region" as a static concept. We reject any regional classification scheme that is defined in terms of inflexible geographic boundaries. What constitutes a region for a mineral resource problem will, in all probability, require a different geographic composition than problems associated with water resources, or marketing, or agriculture. There is no unique regional classification scheme that satisfies the variety of problems facing the regional investigator and/or decision maker. Thus, in contrast with static delimitation schemes we argue for a dynamic (functional) interpretation of region —one which recognizes the existence of an infinite number of overlapping, as well as independent, regions. The geographic extent of a region is shaped by the nature of the problem under study. In short, we operate on the proposition that to be meaningful, **any region must be functionally defined in terms of the problems at hand.**[2]

[2]For a sample of the literature dealing with the regional concept, see: Richard B. Andrews, "Mechanics of the Urban Economic Base: The Problem of Base Area Delimitation," *Land Economics,* XXX, (Madison, Wisconsin: University of Wisconsin, 1954), pp. 309-319; Donald J. Bogue, *State Economic Areas,* U. S. Department of Commerce, Bureau of the Census, (Washington, D. C., 1951), pp. 1-6; Donald J. Bogue, "The Need for an International System of Regions and Subregions," *Papers and Proceedings of the Regional Science Association,* I, (1955), pp. P1-P9; George H. Borts, "An Approach to Measuring Regional Growth Differentials," *Papers and Proceedings of the Regional Science Association,* IV, (1958), pp. 207-220; Joseph L. Fisher, "Concepts in Regional Economic Development," *Papers and Proceedings of the Regional Science Association,* I (1955), pp. W-1 thru W-20; Morris E. Garnsey, "The Dimensions of Regional Science," *Papers and Proceedings of the Regional Science Association,* II, (1956), pp. 27-39; Walter Isard, "Regional Science, The Concept of Region, and Regional Structure, " *Papers and Proceedings of the Regional Science Association,* II, (1956), pp. 13-26; Walter Isard and Guy Freutel, "Regional and National Product Projections and Their Interrelations," *Long-Range Economic Projection, Studies in Income and Wealth,* XVI, National Bureau of Economic Research, (Princeton: Princeton University Press, 1954), pp. 427-471; Walter Isard, et. al., *Methods of Regional Analysis: An Introduction to Regional Science,* (New York: John Wiley & Sons, Inc., 1960), pp. 322-324; A. Losch, "The Nature of Economic Regions," *Southern Economic Journal,* V, (July 1938-April 1939), pp. 71-78; Harvey S. Perloff, "Problems of Assessing Regional Economic Progress," *Regional Income, Studies in Income and Wealth,* XXI, National Bureau of Economic Research, (Princeton: Princeton University Press, 1957), pp. 37-62; Harvey S. Perloff, Edgar S. Dunn, Jr., Eric E. Lampard, and Richard F. Muth, *Regions, Resources, and Economic Growth,* Resources for the Future, Inc., (Baltimore: Johns Hopkins Press, 1960), pp. 4-8; Josiah C. Russell, "The Metropolitan City Region of the Middle Ages," *Journal of Regional Science,* II, No. 2, (1960), pp. 55-70; Charles M. Tiebout, "A Method of Determining Incomes and Their Variations in Small Regions," *Papers and Proceedings of the Regional Science Association,* I, (1955), pp. F1-F12; Morris B. Ullman and Robert C. Klove, "The Geographic Area in Regional Economic Research," *Regional Income, Studies in Income and Wealth,* XXI, National Bureau of Economic Research, (Princeton: Princeton University Press, 1957), pp. 87-109; U. S. Bureau of the Census, *U. S. Censuses of Population and Housing*: 1960, *Geographic Identification Code Scheme,* United States PHC (2)-1, (Washington, D. C., 1961), p. V; U.S.D.A., Economic Research Service, *Farm Income* 1949-1962, *A Supplement to the Farm Income Situation for July* 1963, FIS-191 Supplement, (August, 1963), p. 6; U. S. Department of Commerce, Office of Business Economics, *Personal Income by States Since* 1929, (Washington, D. C., 1956), (Frontispiece); Rutledge Vining, "The Region as an Economic Entity and Certain Variations to be Observed in the Study of Systems of Regions," *Papers and Proceedings of the American Economic Association,* XXXIX, (May, 1949), pp. 90-92; Rutledge Vining, "Delimitation of Economic Areas: Statistical Conceptions in the Study of the Spatial Structure of an Economic System," *Journal of the American Statistical Association,* XLVIII, (March, 1953), pp. 44-64; Andrzej Wrobel, "Regional Analysis and the Geographic Concept of Region," *Regional Science Association Papers,* VIII, (European Congress, The Hague, 1961), pp. 37-41.

Building Blocks. As previously indicated unavailable, inadequate, and incomparable subnational data constitute a major obstacle for the regional decision maker or investigator.[3] For example, aggregate data are not readily available for the Arkansas River Basin, or the Hugoton natural gas field, or the Tri-State lead and zinc mining district, or for the marketing area of a major oil company. Rather than collect economic progress data on these and other overlapping regions separately, we advocate its collection in terms of a micro unit that is small enough to serve as a building block, yet large enough to qualify as a workable statistical unit. The identification and use of such a micro building block enables the aggregating of data for an infinite number of regions.

We have selected the county as the basic micro unit for the collection of economic progress data. The basis for this selection is threefold. First, the county's relatively small size qualifies it as a regional building block. Given comparable data on a county basis, any user can put together as many counties as may be required for the problem at hand. Second, we have more pertinent time series data available for the county than for any other local unit. Third, changes in boundaries are not expected to disrupt the continuity or historical validity of our county building block data. It seems likely that any future changes in the county unit will be toward a consolidation of existing counties rather than a further subdivision.

The Building Block Data

The types of building block data needed to facilitate regional analysis and decision making were determined in consultation with other regional investigators within and without the six-state area; private state-local civic, planning, and development groups; business firms; and appropriate federal-state-local government agencies. Given our budget and manpower limitations, it was impossible to incorporate all suggested data items. Thus, through a series of conferences, data priorities were established. The resulting framework and system of data collection for the six-state pilot program logically subclassified into two broad categories: principal measures of economic progress, and supporting measures of economic progress. The former category includes measures previously unavailable on a reasonably uniform basis for all 564 counties. These principal measures are personal income and population. The generation of these data required the greatest inputs of manpower. They also involved the major methodological problems. Given these two

[3]Werner Hochwald, Editor, *Design of Regional Accounts*, (Baltimore: Johns Hopkins Press, 1961).

measures, the user can compute per capita personal income for any desired grouping of counties. Per capita rather than total personal income is the best single measure of economic progress since higher standards of living do not necessarily result from increased total personal income. If population grows more rapidly than does total personal income, economic well-being will decline. This phenomenon occurs in many underdeveloped regions of the world. Herein lies the explanation for our designation of population, as well as total personal income, as a principal measure of economic progress.

"Income measures are the best starting point for an economic appraisal because (1) income shows how economic activities pay off, (2) income payments are closely related to the economic welfare of the people, and (3) it is possible to break down total income into payments from various sources, which can be related to the major types of economic activity in an area."[4] Dr. W. Nelson Peach of the University of Oklahoma had the principal responsibility for coordinating the preparation of the county personal income estimates. The state project directors in the other five states were Dr. Darwin W. Daicoff, University of Kansas; Dr. Robert N. McMichael, University of Arkansas; Dr. Robert W. Paterson, University of Missouri; Dr. Wallace C. Peterson, University of Nebraska; and Dr. Lewis E. Wagner, State University of Iowa.[5]

The personal income estimates were prepared on an annual basis for the 1950-1962 period for each of the 564 counties in the six-state area. Three tables have been prepared for each county.[6] Table 1 presents county personal income annually by major component: (1) wages, salaries and other labor income, (2) proprietor income, (3) property income, and (4) transfer payments. Table 2 presents annual county personal income by broad industrial source: (1) farm income, (2) government subclassified into (a) federal and (b) state and local, and (3) private nonfarm income. Table 3 presents annual county wages and salaries by major industrial source: (1) farm, (2) mining, (3) contract construction, (4) manufacturing, (5) wholesale and retail trade, (6) finance, insurance and real estate, (7) public utilities including transportation and communication, (8) services, and (9) government.

---

[4]*Comparative Economic Progress in the Southeast,* as quoted in Harvey S. Perloff, "Problems of Assessing Regional Economic Progress," *Regional Income,* Vol. 21, National Bureau of Economic Research (Princeton: Princeton University Press, 1957), p. 42.

[5]Glenn H. Miller Jr. initiated the work in Kansas prior to moving to Boston to complete requirements for the Ph.D. Vincent E. Cangelosi directed the work in Arkansas prior to leaving for a year's post-doctoral study under a National Science Foundation grant. Conrad Stucky directed the work in Iowa before accepting a Ford Foundation assignment in Lebanon.

[6]For an illustration of the format of these tables and other data see: W. Nelson Peach, Richard W. Poole, and James D. Tarver, *County Building Block Data For Regional Analysis: Oklahoma,* (Research Foundation, Oklahoma State University, 1965).

Given our concepts of (1) region and (2) the county as a building block, we decided not to incorporate situs adjustments into the county data, even though such adjustments were computed for counties in several states. The situs problem is primarily associated with counties in and near Standard Metropolitan Statistical Areas where large numbers of workers commute to the central county.[7] We are concerned with the county as a building block. Thus, when we add data on a group of counties to construct a region, the problem of situs washes out.

The supporting economic progress data are designed to assist the regional investigator or decision maker in analyzing the trends revealed by our principal measures of economic progress. They include information on such aspects of each county's economy as agriculture, mining, wholesale trade, retail trade, manufacturing and banking. Also, data are presented for selected years on social characteristics such as education, housing, race and age distribution of the population.

A situation often overlooked by some academicians and statisticians is that many businessmen, civic leaders and governmental-legislative officials are unaware of the nature and scope of existing data. One objective of our supporting economic progress data series is to acquaint such decision makers with data availability. To facilitate this process, detailed source notes and explanations for each data component of this series were prepared in a readable form for use by the layman.[8] Further, groups knowledgeable with respect to data sources often find it necessary to go back over the same source materials and spend much time hashing and rehashing the same set of data. Even if a person in one state put this kind of data together, almost surely his method would differ from the way it would be put together by an individual in another state. It is our conviction that once such information is available on a comparable basis, a large number of people will be free to allocate more time to analysis as well as to concentrate on other areas of study.

A Case Study

To illustrate the use of our building block data, let us briefly examine a ten-county rural region in the six-state area. As the following comments indicate, the ten counties comprise a depressed area.

---

[7]The problem becomes acute in such areas as Kansas City and St. Louis. *Net* commuting figures for selected counties in Illinois and Missouri to St. Louis County and St. Louis City are as follows: St. Clair County, Illinois, 15,285; Madison County, Illinois, 4,591; Monroe County, Illinois, 1,050; Clinton County, Illinois, 358; Randolph County, Illinois, 322; Jefferson County, Missouri, 2,195; Lincoln County, Missouri, 693; St. Francois County, Missouri, 605; and Washington County, Missouri, 302. A similar magnitude of commuting exists in Kansas City. For example, net commuting from Johnson County, Kansas to Kansas City, Missouri, is 12,239.

[8]W. Nelson Peach, *et al., Source Notes and Explanations To County Building Block Data For Regional Analysis: Arkansas, Iowa, Kansas, Missouri, Nebraska and Oklahoma,* (Research Foundation, Oklahoma State University, 1965).

During the period 1950-1962 personal income in the United States almost doubled, but in our ten-county region personal income increased some 70 percent. Nationwide, about 20 percent of personal income comes from government (federal, state and local). In our ten-county area 44 percent comes from government. For the United States transfer payments account for about 8 percent of personal income. In our ten-county area the figure is around 30 percent. Nationwide, manufacturing wages and salaries account for some 23 percent of personal income. But in the ten county area only 5 percent comes from manufacturing wages and salaries.

Low income translates into substandard housing and inadequate sanitary facilities not only in large cities but also in rural America, as reflected by the following aggregated data on our ten low-income rural counties. Of 49,590 housing units only 50 percent have flush toilets. Fifty-one percent of the housing units have no bathing facilities. Only 49 percent of the units have hot and cold piped water inside the structure.

Our supporting economic progress data provide insight into problems which should be dealt with by programs designed to assist such regions of poverty. To illustrate, although median age for our ten-county low-income rural region does not vary significantly from the national median, the age distribution of the population is bimodal. This reflects the high out-migration of the population in the productive age groups. While the ten counties have experienced a 30 percent decline in population since 1930 (from 204,256 to 143,552), the population 55 years and over has **increased** 96 percent, whereas the population in the 20-34 age group has **declined** 59 percent. The 20-34 age group declined from 43,981 to 18,116, whereas the number of persons 55 years and over increased from 18,399 to 36,012. This redistribution of population by age groups explains, in part, the relative growth of the transfer payment component of personal income in the region. When the foregoing age characteristics are combined with the area's median educational level of slightly more than eight years, one becomes painfully aware of the problems confronting this depressed region. While traditional development programs may raise such a region's level of personal income, there is no assurance that the people most needing improved incomes will be able to participate to any measurable extent in the newly created economic opportunities.

Organization of This Publication

The remainder of this document embodies only that part of the research connected with the accuracy tests which were used to evaluate alternative methods of estimating county population. A brief sketch of the chapter organization follows.

Chapter II is devoted to a description of alternative methods of estimating county population. Chapter III reviews previous accuracy tests which have been made of the alternative methods of estimating county population. Chapter IV deals with the statistical model developed as a part of the six-state study to evaluate the accuracy of alternative county population estimation techniques. Chapter V describes five preliminary tests which were undertaken with the statistical model. Chapter VI deals with the derivation of the county population estimates that were required for the six-state evaluation tests. Chapter VII deals with the six-state accuracy tests of alternative techniques of estimating county population. Chapter VIII provides a critique of procedures conventionally used in determining accuracy and those employed in the six-state study, as well as a discussion of other appropriate parametric and nonparametric tests. Chapter IX, Summary and Conclusions, completes the study.

Chapter II

COUNTY POPULATION ESTIMATION TECHNIQUES

The Federal government takes complete censuses of county population every ten years, in years ending in 0. Between decennial censuses, the Bureau of the Census conducts special population censuses of counties only at their request and expense. Only two of the 50 states take population censuses. Kansas takes an annual census of county population as of January 1 each year and Massachusetts conducts a census in years ending in 5.

The Bureau of the Census prepares annual July 1 population estimates of states, but it does not make annual population estimates of counties, except in special cases. Since annual county population estimates are necessary for the computation of yearly birth and death rates, public assistance and welfare loads, and for the allocation of state tax revenues in certain states, numerous state and local agencies have had to develop their own annual intercensal population estimates to meet local needs. The county population estimation work in the various states proceeded largely on a local basis for many years without any direct guidance from the Bureau of the Census. Consequently, a variety of diverse estimation methods employing numerous calculation procedures have been used at one time or another by various local or state agencies.

Only the most conventional and standardized methods will be discussed in this chapter.[1] The first four major sections of this chapter describe the following four widely used estimation techniques for

[1]Henry S. Shryock, Jr., "Development of Postcensal Population Estimates for Local Areas," with discussion by John N. Webb and Ormond C. Corry, National Bureau of Economic Research, "Studies in Income and Wealth," Vol. XXI, *Regional Income* (Princeton: Princeton University Press, 1957), pp. 377-399; Jacob S. Siegel, "Status of Research on Methods of Estimating State and Local Population," *Proceedings of the Social Statistics Section, American Statistical Association* (Washington, D.C., 1960), pp. 172-179; U. S. Bureau of the Census, *Current Population Reports, Population Estimates*, "Local Population Estimates Prepared by State and City Agencies: Mail Survey of 1960," Series P-25, No. 244, March, 1962.

April 1, 1960 estimates: **First,** Component Method II; **second,** the Vital Rates Method; **third,** the Bogue-Duncan Method; and **fourth,** the Census Variation of the Composite Method. The fifth major section of this chapter provides a somewhat briefer discussion of the other methods which either have had restricted use or else are now in the process of being perfected.

Component Method II

This method, which was developed by the Population Division of the Bureau of the Census over a relatively long period, is one of the most extensively used methods of making annual county population estimates. It first appeared in published form in 1956.[2] Component Method II has been very popular and has been applied by the Bureau of the Census and by state and local agencies in preparing annual state, city, and county population estimates.[3]

Component Method II involves a laborious computational procedure and is one of several component methods of making county population estimates. The primary feature of Component Method II, which differentiates it from all other component methods, is the procedure employed in estimating net civilian migration. In estimating this component, it compares the reported number of elementary school children on each estimate date with the expected number of elementary school children surviving from the appropriate age group in the last decennial census.

Component Method II is designed specifically for making postcensal estimates of the total population of counties. It does not provide estimates by age, sex, and race categories, although separate total white and non-white populations can be estimated whenever the basic county input data are available. The generalized formula for computing the April 1, 1960 county population estimates by Method II is as follows:

$$P_{ijlmok} = P_{ok} + B_{ik} - D_{ik} + M_{jik} + M_{lik} + P_{mik} + \epsilon_{ijlmok}, \qquad (1)$$

where P_{ijlmok} is the April 1 county population estimate of the i^{th} estimate year (1960), for the k^{th} county; P_{ok} is the civilian population at the last decennial census date (April 1, 1950) for the k^{th} county; B_{ik} and D_{ik} respectively, are the number of births and the number of deaths occurring between the last decennial census and the i^{th} estimate date for the

---

[2]U. S. Bureau of the Census, *Current Population Reports, Population Estimates,* "Illustrative Example of a Method of Estimating the Current Population of Subdivisions of the United States," Series P-25, No. 133, March, 1956; "Estimates of the Population of States: July 1, 1950 to 1956," Series P-25, No. 165, November, 1957; and "Illustrative Procedure of the Census Bureau's Component Method II For Estimating Current Population," March, 1965.

[3]James D. Tarver and Jeanie Hill, *IBM 650 Program Instructions For Estimating the Current Population of Subdivisions of the United States By Bureau of the Census' Method II and the Vital Rates Method,* Oklahoma Agricultural Experiment Station Processed Series P-344, Stillwater, March, 1960; and James D. Tarver, "Computer Programs for Estimating and Projecting County, City, and Other Local Subdivisional Populations," *Behavioral Science,* 8 (April, 1963), pp. 165-168.

k^{th} county; M_{jik} is an estimate of the net civilian migration between the last decennial census and the i^{th} estimate date in the k^{th} county; M_{lik} is an estimate of the net movement of civilians into the Armed Forces between the last decennial census and the i^{th} estimate date for the k^{th} county; P_{mik} is the number of persons in the Armed Forces stationed in the k^{th} county on the i^{th} estimate date; and ϵ_{ijlmok} is the error in the measurement of the six specified components, plus all other measurement errors.

The six components which constitute the basic input data for the April 1, 1960 county population estimates in the six-state region for the evaluation tests, which are described in Chapter VI of this publication, were obtained from the following sources: **First,** the civilian population (P_0) of each (k^{th}) county at the last decennial census date (P_{ok}) was obtained from the published April 1, 1950 Federal decennial censuses of population; **second,** the annual county resident deaths and resident live births, which are the second and third components, were obtained from the departments of vital statistics in each state; **third,** the number of persons in the Armed Forces stationed in each county (P_{mik}) on each estimate date was obtained from the five branches of the Armed Forces —Departments of the Air Force, Army, Navy, Marine Corps and Coast Guard. Some figures came from national, regional and local military establishments; **fourth,** the net movement of civilians into the Armed Forces (M_{lik}) from each k^{th} county for the April 1, 1960 estimate date was calculated, using the male population 18-24 years of age on the last decennial census date and the state net losses to the Armed Forces, which were computed by the Bureau of the Census; and **fifth,** the net civilian migration (M_{jik}) for each k^{th} county between the last decennial census and the April 1, 1960 estimate date was obtained by comparing the actual number of elementary school children in grades 2-8 on each estimate date with the expected number surviving to ages 7.5 to 15.49. In estimating the net civilian migration between the last decennial census and each i^{th} estimate date, Component Method II uses national migration factors which represent the ratio of the migration rate of the total population to the migration rate of the elementary school population for each estimation period.

The Vital Rates Method

The Vital Rates Method is perhaps the most popular of all standardized estimation techniques because of its simplicity. Professor Bogue[4]

[4]Donald J. Bogue, "A Technique for Making Extensive Population Estimates," *Journal of the American Statistical Association,* 45 (June, 1950), pp. 149-163.

originally developed this method, and the Bureau of the Census further refined it by introducing a number of modifications in computational procedures.[5]

The Vital Rates Method assumes that changes in the number of births and deaths of counties reflect similar changes in the size of the total county population. Specifically, it assumes that the changes in the birth and death rates in each county are proportionate to the changes in the state birth and death rates, respectively, between the last Federal decennial census date and the postcensal estimate date. The generalized formula used in estimating the April 1, 1960 population of counties by the Vital Rates Method is as follows:

$$P_{imok} = \frac{\dfrac{\frac{1}{2}(B_{ik}+B_{ik-1})}{B_{(R)ok}+(B_{(R)in}-B_{(R)on})} + \dfrac{\frac{1}{2}(D_{ik}+D_{ik-1})}{D_{(R)ok}+(D_{(R)in}-D_{(R)on})}}{2} + P_{mik} + \epsilon_{imok}, \quad (2)$$

where B_{ik} and D_{ik} are, respectively, the number of births and deaths for the i^{th} (1960) estimate year for the k^{th} county; B_{ik-1} and D_{ik-1} are, respectively, the number of births and deaths for the year (1959) preceding the April 1, 1960 estimate date in the k^{th} county; $B_{(R)ok}$ and $D_{(R)ok}$ are, respectively, the crude birth and death rates for the o^{th} base period (the average 1949-1950 rates) for the k^{th} county; $B_{(R)in}$ and $D_{(R)in}$ are, respectively, the crude birth and death rates for the i^{th} (1960) estimate year for the n^{th} state; $B_{(R)on}$ and $D_{(R)on}$ are, respectively, the crude birth and death rates for the o^{th} base period for the n^{th} state; P_{mik} is the number of persons in the Armed Forces in the k^{th} county on the i^{th} estimate date; and ϵ_{imok} is the error in the measurement of the specified components, plus all other measurement errors.

The Vital Rates Method makes two separate estimates of the civilian population of each county by applying estimated birth and death rates; it then averages the population figures of the two different estimates; finally, it adds the number of military personnel stationed in each county on the estimate date to obtain the estimated total county population.

The following eight major calculations were involved in computing the April 1, 1960 county population estimates by the Vital Rates Method: **First,** the crude birth and death rates (rates per 1,000 population) for each county and the state were estimated for the base period, using 1949

[5]Bureau of the Census, *Current Population Reports, Population Estimates,* "Estimates of the Population of Continental United States, By Regions, Divisions, and States, and of Alaska, Hawaii, Puerto Rico, The Canal Zone and the Virgin Islands: July 1, 1950 to 1953," Series P-25, No. 97, August, 1954.

and 1950 resident live births and deaths and the estimated April 1, 1950 civilian populations; **second,** the 1960 crude birth and death rates of the civilian population for the state were then estimated; **third,** each county's estimate-year (1960) resident live births were averaged with the corresponding births for the previous year (1959) and deaths were averaged in a similar manner; **fourth,** then, state crude birth and death rates during the base period 1949-1950 were subtracted from the corresponding rates in the 1960 estimate year; **fifth,** the 1949-1950 to 1960 state change in crude birth and death rates were added to each county's 1949-1950 crude birth and death rates, respectively, to obtain "expected" 1960 county birth and death rates of the civilian population for each estimate year; **sixth,** the April 1, 1960 civilian population estimates were made by dividing the "expected" 1960 county birth and death rates into the "expected" births and deaths for the estimate year (described in third above); **seventh,** the two separate April 1, 1960 county civilian population estimates computed by birth rates and by death rates, respectively, were then averaged; and **eighth,** the number of military personnel stationed in each county on April 1, 1960 was added to obtain April 1, 1960 estimates of the total population by the Vital Rates Method.

The computational procedure for making the annual July 1, 1950 to July 1, 1959 county population estimates varies slightly from that outlined for calculating the April 1, 1960 population estimates.[6]

The Bogue-Duncan Composite Method

In 1959 Professor Bogue developed an ingenious procedure to obtain annual county population estimates by age, sex, and race characteristics.[7] Since this method incorporated various computational features of other existing estimation techniques, he used the term "Composite" to identify this newly developed estimation method.

The Bogue-Duncan Composite Method estimates the April 1, 1960 county population by age, race, and sex characteristics from school enrollment data, births, deaths, and Armed Forces strength figures, using specified indices for the base year (1949-1950) and for the estimate year (1960). The detailed county population estimates for April 1, 1960 are then summed to obtain the total estimated county population for all ages.

---

[6]In estimating the mid-year intercensal county population for say July 1, 1959, the procedure in step three above calls for a proportional adjustment of an average of the total 1958 and 1959 county births and deaths, respectively, to the corresponding state births and deaths in the estimate year (1959). However, in the county estimates for April 1, 1960, the average 1959 and 1960 county births and deaths are used without adjustment to the state total in 1960.

[7]Donald J. Bogue and Beverly Duncan, "A Composite Method for Estimating Postcensal Population of Small Areas by Age, Sex, and Color," National Office of Vital Statistics, *Vital Statistics—Special Reports,* Volume 47, Number 6, August 24, 1959.

Even though the Bogue-Duncan Composite Method originally provided detailed population estimates by 5- and 10-year age groups, various age groupings may be used. For example, a recent abbreviated version of this method provides estimates for the following five broad age groups: 0-4, 5-17, 18-44, 45-64, and 65 years of age and over.[8] The population estimates for each of the five age groups are derived from the following basic data: the number of children under 5 years of age on April 1, 1960 are estimated from ratios of children under 5 to females aged 18-44 (Bogue originally used females aged 20-34) ; the number of children 5 to 17 years of age are estimated by a component method, using school enrollment or school census data; the number of persons 18-44 are estimated by the fertility ratio (births to women aged 18-44) and sex ratios; and the number of persons 45-64 and 65 years of age and over are estimated by the number of deaths and death rates. The military personnel residing in each county on April 1, 1960, by age, sex, and race, are then added to the civilian population estimates to obtain the total April 1, 1960 population estimates, by age, race, and sex.

Writing a complete generalized formula for the computation of the April 1, 1960 county population estimates by the Bogue-Duncan Composite Method is a rather tedious undertaking, for various rates and ratios enter into the actual calculations for all of the broad age groups. Since one can vary the number of age groups, the specific formula can be written in various ways.

The detailed computational procedure is carefully outlined in the original publication[9] and a modified version for five broad age groups was drafted by the Staff of the Study Group on Postcensal Population Estimates, Public Health Conference on Records and Statistics, to serve as a guide for the computation of county population estimates for the states of Montana, Ohio, Oregon, and Pennsylvania.[10]

The Census Variation of the Composite Method

The Bureau of the Census has simplified the original Bogue-Duncan Composite Method and has modified the estimation procedures to increase the accuracy of the composite estimation technique. Its version

[8]Jacob S. Siegel, "Status of Research on Methods of Estimating State and Local Population," *Proceedings of the Social Statistics Section,* American Statistical Association (Washington, D. C., 1960), pp. 172-179.

[9]See reference cited in footnote 7.

[10]Jacob S. Siegel, Chief, National Population Estimates and Projections Branch, Population Division, Bureau of the Census, served as chairman of the Study Group. He prepared a series of memos outlining the recommended computational procedures for each of the major estimation methods and provided various national and state "controls" needed in the actual calculations. These memos are rather exhaustive and, altogether, constitute the most detailed account which has ever been drafted of computations for the specified methods.

of the Bogue-Duncan Composite Method, which provides county population estimates for five broad age groups by color, has been descriptively labeled as the Census Variation of the Composite Method.[11]

The Census Variation of the Composite Method uses the Component Method II procedure and school enrollment data to estimate the county population under 5, 5-17, and 18-44 years of age. It employs deaths and death rates to estimate the county population 45 to 64 and 65 years of age and over. The military personnel stationed in the county on April 1, 1960, by age, are then added to the estimates for each of the five separate age groups. Then, these five different age categories are summed to obtain the total April 1, 1960 county population estimates.

Since the generalized formula for calculating the April 1, 1960 population of counties by the Census Variation of the Composite Method requires extensive sets of rates, ratios, and national conversion factors for all five broad age groups, the precise formula will not be written, for it is too cumbersome to specify concisely in its entirety. The detailed computational procedure has been carefully outlined elsewhere.[12]

Other Methods

In addition to the four methods discussed above, there are numerous other techniques which have received rather extensive use.

Bureau of the Census Component Method I,[13] which determines the county net migration rate on each estimate date from changes in local school-age population as compared with changes in the national school-age population, has been used rather extensively, for its computational procedure is much simpler than that for Component Method II.

Component Methods I and II are the most popular of all so-called Migration-and-Natural Increase estimation methods. In other versions of the Migration-and-Natural Increase Method, net migration is estimated by various procedures, using either school data or migration in a previous census decade.[14]

The Natural Increase Method, which estimates county population by adding the natural increase (births minus deaths) to the population

[11]See reference cited in footnote number 8.

[12]See footnote number 10 and the reference cited in footnote number 8.

[13]U. S. Bureau of the Census, *Current Population Reports, Population Estimates*, "Illustrative Examples of Two Methods of Estimating the Current Population of Small Areas," Series P-25, No. 20, May, 1949; and "Suggested Procedures for Estimating the Current Population of Counties," Series P-47, No. 4, April, 1947.

[14]U. S. Bureau of the Census, *Current Population Reports, Population Estimates*, "Current Status of Population Estimates Prepared by State Agencies," Series P-25, No. 116, June, 1955; and "Local Population Estimates Prepared by State and City Agencies: Mail Survey of 1960," Series P-25, No. 244, March, 1962.

enumerated in the last Federal decennial census, assumes that the net migration during the estimation period is zero.[15] This method has not been widely used.

The Censal Ratio Method has been used occasionally in estimating county population.[16] The procedure involves the multiplication of current symptomatic data by the ratio of the population to the basic indicator (school data, births, deaths, and numerous other variables) at the last population census. In some instances the ratio is based upon a time series trend. The Censal Ratio Method may be a simple one, using only a basic indicator, such as the change in the number of occupied dwelling units,[17] or it may be a complex one involving several basic indicators, weighted in various ways.

The Proration Method assumes either that the current population of counties is distributed proportionately throughout the state as at the time of the last Federal decennial census or that the ratio of population to symptomatic data is the same for all counties.[18] Annual county population estimates are made either on the basis of the population at the last census or upon current symptomatic data, such as school data, births, deaths, or other variables. Some have characterized this estimation procedure as an apportionment or ratio technique.

The following two extrapolative methods have been used by many agencies in preparing annual county population estimates: Arithmetic Extrapolation and Geometric Extrapolation.[19] The Arithmetic Extrapolation Method assumes that the yearly population change of a county during the estimation period is identical to the average yearly change during a recent period, usually that of the last decennial census period. On the other hand, the Geometric Extrapolation Method assumes that the annual rate of population change during the estimation period is identical to that of a recent period. Both of these extrapolation methods exclude all available annual data for each year since the last Federal decennial census.

The Bureau of the Census is perfecting the Regression Method which may prove to be simpler to apply than the tedious Component Method II or one of the Composite Methods and may also prove to be a more

---

[15] See the two references cited in footnote number 14.

[16] See the two references cited in footnote number 14.

[17] U. S. Bureau of the Census, *Current Population Reports, Population Estimates*, "Estimates of the Population of the Standard Metropolitan Areas of Houston, Milwaukee, St. Louis, and Washington, D. C.: January 1, 1956," Series P-25, No. 137, May, 1956; and "Estimates of the Population of the New Orleans Standard Metropolitan Area: July 1, 1956," Series P-25, No. 156, April, 1957.

[18] See the two references cited in footnote number 14.

[19] *Ibid.*

accurate estimation technique. State population estimates have been prepared by the Regression Method in combination with Component Method II[20]. Although the Bureau of the Census has not prepared county population estimates by the Regression Method, this method is practically the same procedure as the Ratio-Correlation Method used by others in the preparation of county population estimates in the States of Michigan and Washington.[21] However, it is much to early to speculate whether the Regression Method will eventually replace other methods of estimating county population.

As formulated by the Bureau of the Census, the Regression Method relates changes in selected series of data to changes in population. The six series of data employed in the 1960-1964 state population estimates were as follows: births, deaths, elementary school enrollments, number of Federal individual income tax returns filed, passenger automobile registrations, and employees on nonagricultural payrolls.[22]

Summary

This chapter has described the computational features of the major standardized methods which have been employed in estimating the population of counties. The four most extensively used methods discussed were Component Method II, the Vital Rates Method, the Bogue-Duncan Composite Method, and the Census Variation of the Composite Method.

Other estimation methods which received briefer mention in this chapter were Component Method I, other Migration-and-Natural Increase methods, the Natural Increase Method, the Censal Ratio Method, the Proration Method, the Arithmetic and Geometric Extrapolation methods, and the Regression Method. The last named method appears to be a promising technique and is currently being perfected by the Bureau of the Census.

---

[20]Meyer Zitter and Henry S. Shryock, Jr., "Accuracy of Methods of Preparing Postcensal Population Estimates for States and Local Areas," Volume 1, *Demography* (1964), pp. 227-241; and U. S. Bureau of the Census, *Current Population Reports, Population Estimates*, "Estimates of the Population of States; July 1, 1963, with Preliminary Estimates for July 1, 1964," Series P-25, No. 289, August, 1964.

[21]David Goldberg, Allen Feldt, and J. William Smit, "Estimates of Population Change in Michigan: 1950-1960," *Michigan Population Studies No. 1*, Program For Research in Population and Human Ecology (Ann Arbor: The University of Michigan, 1960); and Robert C. Schmitt and Albert H. Crosetti, "Accuracy of the Ratio-Correlation Method For Estimating Postcensal Population," *Land Economics*, Vol. XXX, No. 3 (August, 1954), pp. 279-280.

[22]See the reference cited in footnote number 20.

Chapter III

PREVIOUS TESTS OF ACCURACY

A number of reports comparing the accuracy of various methods of making postcensal county population estimates have been published. This chapter is designed to provide a summary of these studies. This will be accomplished in the following manner. First, the four most comprehensive studies will be identified and their findings briefly reviewed. Second, inconsistency of findings among the four studies will be discussed.

Findings of Four Comprehensive Studies

The four studies identified as the most comprehensive are designated as follows: Study I—Bureau of the Census evaluation tests for selected metropolitan counties; Study II—evaluations by the Study Group on Postcensal Population Estimates, Public Health Conference on Records and Statistics, for counties in the four states of Montana, Ohio, Oregon and Pennsylvania; Study III—the Schmitt-Crosetti evaluations for counties in the State of Washington; and Study IV—the Goldberg-Balakrishnan evaluations for counties in the State of Michigan. Table 1 provides information relevant to each of these four studies.

Tests of 1960 population estimates for counties in Tennessee, Virginia, Wisconsin, and other states, which are not reviewed in this chapter, showed results that are similar to those of the four studies discussed below.[1]

---

[1] For further reference to the accuracy of various estimation techniques for counties in these and other states, see: Jacob S. Siegel, "Status of Research on Methods of Estimating State and Local Population," *Proceedings of the Social Statistics Section, American Statistical Association* (Washington, D.C., 1960), pp. 172-179; U. S. Bureau of the Census, *Current Population Reports, Population Estimates*, "Estimates of the Population of the Standard Metropolitan Areas of Houston, Milwaukee, St. Louis, and Washington, D.C.: January 1, 1956," Series P-25, No. 137, 1956; and "Estimates of the Population of the New Orleans Standard Metropolitan Area: July 1, 1956," Series P-25, No. 156, 1957; U. S. Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, National Vital Statistics Division, The Public Health Conference on Records and Statistics, *Preliminary Report of the Study Group on Postcensal Population Estimates*, Document No. 520.6–6/11/62; and Meyer Zitter and Henry S. Shryock, Jr., "Accuracy of Methods of Preparing Postcensal Population Estimates for States and Local Areas," Volume 1, *Demography* (1964), pp. 227-241.

TABLE 1
AVERAGE ABSOLUTE PERCENTAGE DEVIATIONS OF SELECTED 1950 AND 1960 COUNTY POPULATION ESTIMATES, PREPARED BY ELEVEN DIFFERENT METHODS IN FOUR EVALUATION STUDIES

Method	STUDY I		STUDY II				STUDY III	STUDY IV
	Selected Metropolitan Counties in the United States		Counties in Four Selected States, 1960				Washington Counties, 1950[5]	Michigan Counties, 1960[6]
	1950[1] (1)	1960[2] (2)	Montana[3] (3)	Ohio[3] (4)	Oregon[3] (5)	Pennsylvania[4] (6)	(7)	(8)
	Average Percentage Deviations							
Component Method I	9.2	NA	NA	NA	NA	NA	NA	NA
Component Method II	6.6	5.9	14.6	6.9	4.5	4.5	NA	4.7
Vital Rates	6.3	5.1	13.5	4.7	5.6	2.7	11.5	6.3
Arithmetic Extrapolation	18.3	NA	10.4	5.5	12.8	NA	16.3	NA
Geometric Extrapolation	16.4	NA	NA	NA	NA	NA	NA	NA
Natural Increase	15.7	NA	NA	NA	NA	NA	NA	NA
Bogue-Duncan Composite	NA	NA	NA	3.4	NA	NA	NA	5.1
Ratio-Correlation	NA	NA	NA	NA	NA	NA	7.4	3.5
Apportionment	NA	NA	NA	NA	NA	NA	28.9	NA
Proration	NA	NA	NA	NA	NA	NA	8.7	NA
Censal Ratio	NA	NA	NA	NA	NA	NA	7.0	NA

The notation NA means not applicable.
[1] See the reference cited in Footnote number 2.
[2] See the reference cited in Footnote number 3.
[3] See the reference cited in Footnote number 4.
[4] See the reference cited in Footnote number 5.
[5] See the references cited in Footnote number 6.
[6] See the references cited in Footnote numbers 7, 8, and 9.

Study I. The April 1, 1950[2] and the April 1, 1960[3] estimates for selected metropolitan counties, which were prepared by the Bureau of the Census, show that the Vital Rates Method gave smaller average percentage errors than did Component Method II (Table 1, columns 1 and 2). However, as shown in Table 1, column 1, both the Vital Rates Method and Component Method II gave consistently smaller average percentage errors in estimating the April 1, 1950 population of selected metropolitan counties than did Component Method I, the Natural Increase Method, the Arithmetic Extrapolation Method, and the Geometric Extrapolation Method.

[2] Henry S. Shryock, Jr., "Development of Postcensal Population Estimates for Local Areas," with discussion by John N. Webb and Ormond C. Corry, National Bureau of Economic Research, Vol. XXI, *Regional Income*, (Princeton: Princeton University Press, 1957), pp. 377-399.

[3] Meyer Zitter and Henry S. Shryock, Jr., "Accuracy of Methods of Preparing Postcensal Population Estimates for States and Local Areas," Volume 1, *Demography* (1964), pp. 227-241.

Study II. In a comparison of April 1, 1960 postcensal county population estimates in four states conducted by the Study Group on Postcensal Population Estimates, Public Health Conference on Records and Statistics[4], the following contradictory results, based upon average percentage errors by selected estimation methods, were obtained: **First**, the most accurate methods for Ohio counties, in the order of their precision, were the Bogue-Duncan Composite Method, the Vital Rates Method, the Arithmetic Extrapolation Method, and Component Method II; **second**, the most accurate methods for Oregon counties, in the order of their precision, were Component Method II, the Vital Rates Method, and the Arithmetic Extrapolation Method; **third**, the most accurate methods for Montana counties, in the order of their precision, were the Arithmetic Extrapolation Method, the Vital Rates Method, and Component Method II; and **fourth**, an examination of two methods of estimating Pennsylvania county population revealed that the Vital Rates Method was superior to the Component Method II.[5]

Study III. In an analysis of April 1, 1950 population estimates for the 39 counties in the State of Washington, the Censal Ratio Method was the most accurate, followed, in order, by the Ratio-Correlation Method, the Proration Method, the Vital Rates Method, the Arithmetic Extrapolation Method, and the Apportionment Method.[6]

Study IV. In an evaluation of April 1, 1960 population estimates for the 83 counties in the State of Michigan, the Ratio-Correlation Method was the most accurate, followed in order, by Component Method II, the Bogue-Duncan Composite Method, and the Vital Rates Method.[7]

Inconsistency of Findings Among Studies

The foregoing publications which analyzed county population estimates failed to demonstrate conclusively which method was most precise. Moreover, the findings were inconsistent from one study to another. Nevertheless, the average percentage error in postcensal popula-

---

[4] U. S. Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, National Vital Statistics Division, the Public Health Conference on Records and Statistics, *Preliminary Report of the Study Group on Postcensal Population Estimates*, Document No. 520.6–6/11/62.

[5] The April 1, 1960 Pennsylvania county population estimates were prepared under the auspices of the Study Group on Postcensal Population Estimates, but were published separately; see Commonwealth of Pennsylvania, Department of Internal Affairs, Bureau of Statistics, "Report of Tests Made of Vital Rates and Census II Methods of Estimating County Population" (mimeographed).

[6] Robert C. Schmitt, "Short-Cut Methods of Estimating County Population," *Journal of the American Statistical Association*, 47 (June, 1952), pp. 232-238; and Robert C. Schmitt and Albert H. Crosetti, "Accuracy of the Ratio-Correlation Method for Estmiating Postcensal Population," *Land Economics*, 30 (August, 1954), pp. 279-281.

[7] David Goldberg and T. R. Balakrishnan, "A Partial Evaluation of Four Estimation Techniques," *Michigan Population Studies No. 2*, Program for Research in Population and Human Ecology (Ann Arbor: The University of Michigan, 1961). The April 1, 1960 county population estimates were based upon extrapolations of July 1, 1958 county population estimates.

tion estimates generally declined as population size increased and as the rate of population change during the previous census decade increased. Component Method II was generally more accurate (in terms of average percentage errors) than the Vital Rates Method for (1) metropolitan counties having central cities, (2) counties with large populations, and (3) counties with small rates of population change in the previous census decade. On the other hand, the Vital Rates Method was generally superior to Component Method II for (1) counties with small populations, (2) counties with large rates of population increase in the previous decade, and (3) for suburban counties.

In summary, the findings of the four studies of postcensal county population estimates reviewed above were contradictory. In eight separate comparisons, six different estimation methods were shown to be most precise. Three of the eight comparisons indicated that the Vital Rates Method was superior (see the average percentage deviations for the 1950 and 1960 metropolitan county estimates and the 1960 Pennsylvania county estimates in Table 1); one comparison indicated that the Bogue-Duncan Composite Method was superior (see the average percentage deviations for the 1960 Ohio county estimates in Table 1); another comparison indicated that Component Method II was superior (see the average percentage deviations for the 1960 Oregon county estimates in Table 1); another indicated that the Arithmetic Extrapolation Method was superior (see the average percentage deviations for the 1960 Montana county estimates in Table 1); another indicated that the Censal Ratio Method was superior (see the average percentage deviations for the April 1, 1950 Washington county estimates in Table 1); and the eighth comparison indicated that the Ratio-Correlation Method was superior (see the average percentage deviations for the April 1, 1960 Michigan county estimates in Table 1).[8] It is impossible to determine from these conflicting conclusions whether any one of the six methods purported to be the most precise was actually more accurate than any of the other methods discussed.

Apparently, there are three major reasons for the inconclusive findings on the accuracy of the estimation methods. First, the quality of the basic data, especially county school enrollments, varies from state to state. This greatly affects the reliability of Component Method II estimates. Second, in some of the computations the estimation methods were modified. For example, some county estimates were adjusted to

[8] David Goldberg, Allen Feldt, and J. William Smit, "Estimates of Population Change in Michigan, 1950-1960," *Michigan Population Studies No. 1*, Program for Research in Population and Human Ecology (Ann Arbor: The University of Michigan, 1960).

independent state population estimates. On the other hand, some county estimates were not adjusted to independent state population estimates. **Third,** another explanation for the inconclusive findings of the published evaluation tests lies in the statistical procedure employed in gauging the accuracy of different estimation methods. Briefly, the conventional method of determining the accuracy of the various methods of estimating county population is to obtain the numerical difference between the postcensal county population estimate and the enumerated population from the Federal decennial census of population for each county. Then, the absolute percentage deviation (obtained by dividing the numerical difference by the enumerated county population) is computed for each county. The procedure is to compute the average (mean) absolute percentage deviations of each estimation method for all counties in a particular state or area, and in some cases the variance of the absolute percentage deviations, the number of positive deviations, and the number of percentage deviations exceeding some level. The weaknesses and deficiencies inherent in the conventional procedure for determining the accuracy of alternative estimation methods will be given in Chapter VIII, "Evaluation of Accuracy Tests."

Some of the published studies indicated that the averaging of two or more independent methods of approximately equal precision tended to provide more accurate estimates than either one of the methods used individually.[9]

Summary

The four major comprehensive studies discussed in this chapter did not conclusively show that any one of the most widely used methods of estimating county population is more accurate than any of the other methods tested. The divergent findings were due to at least three major factors: **First,** variability in the quality of the basic data, especially county school enrollments; **second,** lack of uniformity in estimation procedures; and **third,** statistical procedures used in determining the accuracy of various estimation methods.

[9] Jacob S. Siegel, "Status of Research on Methods of Estimating State and Local Population," *Proceedings of the Social Statistics Section, American Statistical Association* (Washngton, D. C., 1960), pp. 172-179; and Meyer Zitter and Henry S. Shryock, Jr., "Accuracy of Methods of Preparing Postcensal Population Estimates for States and Local Areas," Volume 1, *Demography* (1964), pp. 227-241.

Chapter IV

THE STATISTICAL MODEL FOR TESTS OF ACCURACY

As indicated in the last chapter, published tests of accuracy do not conclusively show any one technique of estimating county population to be consistently more precise than other techniques. Presumably, variations in accuracy among the various estimation techniques which employ uniform basic data inputs and estimation procedures may be explained in terms of differences in population density, in metropolitan classification, and in changes in the number of births and deaths. Accordingly, the most precise estimation technique may vary with the pertinent characteristics of each county. The basic problem is, therefore, to determine which estimation method is most accurate for each type of county, given each county's specified characteristics.

This chapter is devoted to the development of a statistical model for analyzing the accuracy of four alternative methods of estimating county population. In chapter V, five different variations of this model will be employed in some preliminary tests of accuracy. Specifically, the model will be used to evaluate the accuracy of selected methods of estimating (1) the 1950 and 1960 population of selected metropolitan counties in the United States, and (2) the 1960 population of counties in Montana, Ohio, Oregon, and Pennsylvania. Then, in Chapter VII three different variations of the model will be employed to test the accuracy of selected methods of estimating the 1960 population of counties in Arkansas, Iowa, Kansas, Missouri, Nebraska, and Oklahoma. Thus, this chapter is limited to the formulation of the basic statistical model, whereas Chapters V and VII are devoted to an application of the model.

The first major section of this chapter is devoted to the definition of terms; the second section to specific questions which will be answered in the tests of accuracy; the third section to the statistical model; and the fourth major section to the tests of hypotheses.

Definition of Terms

The following five terms are basic to the statistical model developed here:

1. A *census value* is the total number of people living in a specified county, as enumerated by a Federal decennial census.

2. An *estimator* is a specific formula (method) designed to estimate a census value. The problem described in this chapter involves a comparison of four different estimators (estimation methods).

3. An *estimate* is the numerical value obtained by substituting actual input data into each of the four estimator formulas.

4. The *mean square error* is the mean of the squares of the errors. A specific estimate of a census value is compared with the census value itself. The difference between the two is defined as the *error* of that estimate. Since each county estimate is compared with a census value for the same county, there will be an equal number of errors and counties. A mean square error can be found for each of the four estimators to be compared, and the estimator with the smallest mean square error will tend to fall closest to the "true" or census value.

5. *Bias* is the difference between the average of all the estimates made by a particular estimator and the average of all census values. If the average of all the estimates is exactly the same as the average of all the census values, the estimator is said to be unbiased. If the averages are unequal, the difference between them is defined as the *bias*. If the average for an estimator is larger than the census average (hence, the bias is positive), the estimate will more often be too high than too low, and conversely.

Specific Questions To Be Answered in the Tests of Accuracy

In general, the estimators will be evaluated by comparing their mean square errors. That estimator having the smallest mean square error will be considered the most "accurate" estimator. If no genuine difference in accuracy can be detected on this basis, then the estimators will be compared with respect to bias. However, it is unnecessary to consider bias further in this chapter in the development of the statistical model.

The requisite data for testing the accuracy of the four estimation methods in the six-state area were prepared, using the errors in the April

1, 1960 county population estimates as the basic data for the statistical model developed here. For each county in the six-state area, the following five items of information had been tabulated prior to making the April 1, 1960 population estimate:

1. Metropolitan status of counties at the time of the previous Federal census, with counties being classified into the following three categories: (1) metropolitan counties with central cities; (2) metropolitan counties without central cities; and (3) nonmetropolitan counties.

2. Population density of counties at the previous Federal census. Population density will be employed as a covariable (concomitant variable) in some analyses and as a variable in other analyses, with counties being classified into one of the following five density groups: (1) under 250 people per square mile; (2) 250-499; (3) 500-749; (4) 750-999; and (5) 1,000 and over.

3. Change in births from the previous Federal census year to the estimate year (ratio of births in the estimate year, 1960, to those in the base year, 1950).

4. Change in deaths from the previous Federal census year to the estimate year (ratio of deaths in the estimate year, 1960, to those in the base year, 1950).

5. The actual census value (April 1, 1960 enumerated population) was recorded.

The above three concomitant variables (2, 3, 4) and two variables (1, 2) had been obtained for the specific purpose of explaining differences in the accuracy of the four estimation methods.

The two major questions which the evaluation test proposes to answer are the following:

1. Which of the above four variables (1, 2, 3, 4) is most closely related to estimator accuracy?

2. Will the utilization of all the variables and concomitant variables provide a better comparison of estimator accuracy than use of the best one or two alone?

All except one of the following five questions deal with the relationship between estimator accuracy and the concomitant variables listed above:

First, is there any consistent difference in the accuracy of the four estimation methods used, without reference to the concomitant variables?

Second, is one estimator most accurate for metropolitan counties while another is best for nonmetropolitan counties?

Third, is the best estimator for a given county affected by the population density or metropolitan classification of that county?

Fourth, is the best estimator for a given county affected by the change in births (ratio of 1960 to 1950 births) in that county?

Fifth, is the best estimator for a given county affected by the change in deaths (ratio of 1960 to 1950 deaths) in that county?

The Statistical Model

In the statistical analysis formulated here, the common logarithms of the positive difference ($\log_{10}$ of the absolute difference) between the April 1, 1960 postcensal county population estimates and the April 1, 1960 Federal decennial census enumerations were taken. The formula used to obtain the logarithms is as follows:

$$U_{ijk} = \log|c_{ijk} - v_{ijk}|, \quad (1)$$

where c_{ijk} is the April 1, 1960 county population estimate for estimation method i ($i = 1, 2, 3, 4$), metropolitan classification j ($j = 1, 2, 3$), and the k^{th} county within the j^{th} metropolitan classification ($k = 1, 2, \ldots, n_j$); v_{ijk} is the corresponding census value (April 1, 1960 census enumeration) for that county; and U_{ijk} is the common logarithm of the positive difference between the estimate for a county and the corresponding census value for that county. The positive difference was used because the precision of the estimate (that is, the distance from the "true" or census value) was desired, and a logarithmic transformation makes this variable approximately normally distributed.

A multiple covariance regression model will be used in computing different analyses of variance of the logarithms, and F- and t-tests will be made in testing for significant differences. The mathematical model employed is as follows:

$$U_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \alpha_i x_{jk} + \beta_i y_{jk} + \gamma_i z_{jk} + \epsilon_{ijk}, \quad (2)$$

where a_i is the fixed effect on the accuracy due to estimator i ($i = 1, 2, 3, 4$); b_j is the fixed effect on accuracy due to metropolitan classification j ($j = 1, 2, 3$); $(ab)_{ij}$ is the fixed interaction effect, showing the relationship between a_i and b_j. If the accuracy of the estimator depends upon which of the three metropolitan classifications a county is in, this interaction effect will differ from zero; x_{jk} is the April 1, 1950 population density of county k, within metropolitan classification j; y_{jk} is the ratio of the number of 1960 to 1950 births in county k, within metropolitan classification j; z_{jk} is the ratio of the 1960 to 1950 deaths in county k, within metropolitan classification j; α_i, β_i, and γ_i are partial regression coefficients associated with x_{jk}, y_{jk}, and z_{jk}, respectively, for each of the four estimation methods; μ is the overall mean; and ε_{ijk} is the errror term. This multiple covariance model assumes that the epsilons (ε's) are normally distributed, with a mean of zero and a variance of sigma squared [$\varepsilon_{ijk} \sim N(0, \sigma^2)$].

Tests of Hypotheses, With Formulas

The model given above (Formula 2) may be more complex than the county population estimation data will justify. Moreover, some of the proposed tests will vary according to the complexity of the model. For this reason, the only formulas which will be given here are for the tests to ascertain this complexity. The formulas and derivations designed to answer the questions outlined in the above section were given in an unpublished memorandum.[1]

The relevant questions which must be answered in order to determine the complexity of the model are the following:

1. Does interaction exist between estimator accuracy and
   (a) metropolitan status?
   (b) population density?
   (c) ratio of 1960 to 1950 births?
   (d) ratio of 1960 to 1950 deaths?

2. Which of the four variables add nothing more to what is known through other variables?

A statistical test corresponding to each of these five questions will be given in the remaining part of this chapter.

First, we shall give some definitions by explaining the notations introduced in Formulas 1 and 2.

We shall let

$$X_{j\cdot} = \sum_{k=1}^{n_j} x_{jk}, \quad Y_{j\cdot} = \sum_{k=1}^{n_j} y_{jk}, \quad Z_{j\cdot} = \sum_{k=1}^{n_j} z_{jk} \qquad (3)$$

[1]David White, "Proofs for the Report on Statistical Analyses of County Population Estimates of Six Midwestern States," Research Foundation, Oklahoma State University, February, 1963 (mimeographed).

PLATE I - REGRESSION ESTIMATES FOR THE ADJUSTED ERROR SUM-SQUARES

$$\Big(\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j}\Big)\hat\alpha_1 + 0\cdot\hat\alpha_2 + 0\cdot\hat\alpha_3 + 0\cdot\hat\alpha_4 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\beta_1 + 0\cdot\hat\beta_2 + 0\cdot\hat\beta_3 + 0\cdot\hat\beta_4 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_1 + 0\cdot\hat\gamma_2 + 0\cdot\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{1jk}X_{jk} - \sum_j \frac{U_{1j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + \Big(\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j}\Big)\hat\alpha_2 + 0\cdot\hat\alpha_3 + 0\cdot\hat\alpha_4 + 0\cdot\hat\beta_1 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\beta_2 + 0\cdot\hat\beta_3 + 0\cdot\hat\beta_4 + 0\cdot\hat\gamma_1 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_2 + 0\cdot\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{2jk}X_{jk} - \sum_j \frac{U_{2j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + 0\cdot\hat\alpha_2 + \Big(\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j}\Big)\hat\alpha_3 + 0\cdot\hat\alpha_4 + 0\cdot\hat\beta_1 + 0\cdot\hat\beta_2 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\beta_3 + 0\cdot\hat\beta_4 + 0\cdot\hat\gamma_1 + 0\cdot\hat\gamma_2 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{3jk}X_{jk} - \sum_j \frac{U_{3j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + 0\cdot\hat\alpha_2 + 0\cdot\hat\alpha_3 + \Big(\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j}\Big)\hat\alpha_4 + 0\cdot\hat\beta_1 + 0\cdot\hat\beta_2 + 0\cdot\hat\beta_3 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\beta_4 + 0\cdot\hat\gamma_1 + 0\cdot\hat\gamma_2 + 0\cdot\hat\gamma_3 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_4 = \sum_{j,k} u_{4jk}X_{jk} - \sum_j \frac{U_{4j\cdot}X_{j\cdot}}{n_j}$$

$$\Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\alpha_1 + 0\cdot\hat\alpha_2 + 0\cdot\hat\alpha_3 + 0\cdot\hat\alpha_4 + \Big(\sum_{j,k} Y_{jk}^2 - \sum_j \frac{Y_{j\cdot}^2}{n_j}\Big)\hat\beta_1 + 0\cdot\hat\beta_2 + 0\cdot\hat\beta_3 + 0\cdot\hat\beta_4 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_1 + 0\cdot\hat\gamma_2 + 0\cdot\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{1jk}Y_{jk} - \sum_j \frac{U_{1j\cdot}Y_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\alpha_2 + 0\cdot\hat\alpha_3 + 0\cdot\hat\alpha_4 + 0\cdot\hat\beta_1 + \Big(\sum_{j,k} Y_{jk}^2 - \sum_j \frac{Y_{j\cdot}^2}{n_j}\Big)\hat\beta_2 + 0\cdot\hat\beta_3 + 0\cdot\hat\beta_4 + 0\cdot\hat\gamma_1 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_2 + 0\cdot\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{2jk}Y_{jk} - \sum_j \frac{U_{2j\cdot}Y_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + 0\cdot\hat\alpha_2 \quad \Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\alpha_3 + 0\cdot\hat\alpha_4 + 0\cdot\hat\beta_1 + 0\cdot\hat\beta_2 + \Big(\sum_{j,k} Y_{jk}^2 - \sum_j \frac{Y_{j\cdot}^2}{n_j}\Big)\hat\beta_3 + 0\cdot\hat\beta_4 + 0\cdot\hat\gamma_1 + 0\cdot\hat\gamma_2 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{3jk}Y_{jk} - \sum_j \frac{U_{3j\cdot}Y_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + 0\cdot\hat\alpha_2 + 0\cdot\hat\alpha_3 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\hat\alpha_4 + 0\cdot\hat\beta_1 + 0\cdot\hat\beta_2 + 0\cdot\hat\beta_3 + \Big(\sum_{j,k} Y_{jk}^2 - \sum_j \frac{Y_{j\cdot}^2}{n_j}\Big)\hat\beta_4 + 0\cdot\hat\gamma_1 + 0\cdot\hat\gamma_2 + 0\cdot\hat\gamma_3 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\gamma_4 = \sum_{j,k} u_{4jk}Y_{jk} - \sum_j \frac{U_{4j\cdot}Y_{j\cdot}}{n_j}$$

$$\Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\alpha_1 + 0\cdot\hat\alpha_2 + 0\cdot\hat\alpha_3 + 0\cdot\hat\alpha_4 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\beta_1 + 0\cdot\hat\beta_2 + 0\cdot\hat\beta_3 + 0\cdot\hat\beta_4 + \Big(\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j}\Big)\hat\gamma_1 + 0\cdot\hat\gamma_2 + 0\cdot\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{1jk}Z_{jk} - \sum_j \frac{U_{1j\cdot}Z_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\alpha_2 + 0\cdot\hat\alpha_3 + 0\cdot\hat\alpha_4 + 0\cdot\hat\beta_1 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\beta_2 + 0\cdot\hat\beta_3 + 0\cdot\hat\beta_4 + 0\cdot\hat\gamma_1 + \Big(\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j}\Big)\hat\gamma_2 + 0\cdot\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{2jk}Z_{jk} - \sum_j \frac{U_{2j\cdot}Z_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + 0\cdot\hat\alpha_2 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\alpha_3 + 0\cdot\hat\alpha_4 + 0\cdot\hat\beta_1 + 0\cdot\hat\beta_2 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\beta_3 + 0\cdot\hat\beta_4 + 0\cdot\hat\gamma_1 + 0\cdot\hat\gamma_2 + \Big(\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j}\Big)\hat\gamma_3 + 0\cdot\hat\gamma_4 = \sum_{j,k} u_{3jk}Z_{jk} - \sum_j \frac{U_{3j\cdot}Z_{j\cdot}}{n_j}$$

$$0\cdot\hat\alpha_1 + 0\cdot\hat\alpha_2 + 0\cdot\hat\alpha_3 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\alpha_4 + 0\cdot\hat\beta_1 + 0\cdot\hat\beta_2 + 0\cdot\hat\beta_3 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\hat\beta_4 + 0\cdot\hat\gamma_1 + 0\cdot\hat\gamma_2 + 0\cdot\hat\gamma_3 + \Big(\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j}\Big)\hat\gamma_4 = \sum_{j,k} u_{4jk}Z_{jk} - \sum_j \frac{U_{4j\cdot}Z_{j\cdot}}{n_j}$$

$$[(\sum_{j,k} X_{jk}^2 - \frac{X_{..}^2}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})]\hat{\alpha}_1 \quad - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})\hat{\alpha}_2 \quad - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})\hat{\alpha}_1 + [(\sum_{j,k} X_{jk}^2 - \frac{X_{..}^2}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})]\hat{\alpha}_2 \quad - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})\hat{\alpha}_1 \quad - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})\hat{\alpha}_2 + [(\sum_j X_{jk}^2 - \frac{X_{..}^2}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})] \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})\hat{\alpha}_1 \quad - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})\hat{\alpha}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}^2}{n_j} - \frac{X_{..}^2}{n.})\hat{\alpha}_3 + [(\sum_j X_{jk}^2 - \frac{X_{..}^2}{n.}) - \ldots$$

$$[(\sum_{j,k} X_{jk} Y_{jk} - \frac{X_{..}Y_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})]\hat{\alpha}_1 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\alpha}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\alpha}_1 + [(\sum_{j,k} X_{jk} Y_{jk} - \frac{X_{..}Y_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})]\hat{\alpha}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\alpha}_1 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\alpha}_2 + [(\sum_{j,k} X_{jk} Y_{jk} - \frac{X_{..}Y_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\alpha}_1 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\alpha}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\alpha}_3 + [(\sum_{j} X_{jk} Y_{jk} - \ldots$$

$$[(\sum_{j,k} X_{jk} Z_{jk} - \frac{X_{..}Z_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})]\hat{\alpha}_1 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\alpha}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\alpha}_1 + [(\sum_{j,k} X_{jk} Z_{jk} - \frac{X_{..}Z_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})]\hat{\alpha}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\alpha}_1 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\alpha}_2 + [(\sum_{j,k} X_{jk} Z_{jk} - \frac{X_{..}Z_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n} \ldots$$

$$-\frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\alpha}_1 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\alpha}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\alpha}_3 + [(\sum_{j,k} X_{jk} Z_{j} \ldots$$

Note: $X_{j.}Y_{j.}/n_j$, as typed in the memo,

/

PLATE II - REGRESSION ESTIMATES

$$\frac{\ldots^2}{n_\cdot})\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}^2}{n_j} - \frac{X_{\cdot\cdot}^2}{n_\cdot})\hat{\hat\alpha}_4 + [(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot}) - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})]\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j$$

$$\frac{\ldots^2}{n_\cdot})\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}^2}{n_j} - \frac{X_{\cdot\cdot}^2}{n_\cdot})\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_1 + [(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot}) - \frac{1}{4}(\sum_j \frac{X\ldots}{n\ldots}$$

$$\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}^2}{n_j} - \frac{X_{\cdot\cdot}^2}{n_\cdot})\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{Y_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_2 + [(\sum_{j,k} X_{j\ldots}$$

$$-(\sum_j \frac{X_{j\cdot}^2}{n_j} - \frac{X_{\cdot\cdot}^2}{n_\cdot})]\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_2 - \frac{1}{4}(\sum_j \frac{X\ldots}{}$$

$$-\frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_4 + [(\sum_{j,k} Y_{jk}^2 - \frac{Y_{\cdot\cdot}^2}{n_\cdot}) - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} - \frac{Y_{\cdot\cdot}^2}{n_\cdot})]\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j}$$

$$\frac{\ldots Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} - \frac{Y_{\cdot\cdot}^2}{n_\cdot})\hat{\hat\beta}_1 + [(\sum_{j,k} Y_{jk}^2 - \frac{Y_{\cdot\cdot}^2}{n_\cdot}) - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} -$$

$$\frac{Y_{\cdot\cdot}}{n_\cdot})]\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} - \frac{Y_{\cdot\cdot}^2}{n_\cdot})\hat{\hat\beta}_1 + [(\sum_{j,k} Y_{jk}^2 - \frac{Y_{j\cdot}^2}{n_j}) - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} - \frac{Y_{\cdot\cdot}^2}{n\ldots}$$

$$\frac{\ldots Y_{\cdot\cdot}}{n_\cdot}) - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Y_{\cdot\cdot}}{n_\cdot})]\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} - \frac{Y_{\cdot\cdot}^2}{n_\cdot})\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} - \frac{Y_{\cdot\cdot}^2}{n_\cdot})\hat{\hat\beta}_2 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}^2}{n_j} -$$

$$\frac{X_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_4 + [(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot}) - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j} - \frac{Y_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})]\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z\ldots}{n_j}$$

$$\frac{X_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j} - \frac{Y_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_1 + [(\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}) - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j}$$

$$\ldots)]\hat{\hat\alpha}_3 - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j} - \frac{Y_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j} - \frac{Y_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_2 + [(\sum_{j,k} Y_{jk}Z\ldots$$

$$-\frac{X_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot}) - \frac{1}{4}(\sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j} - \frac{X_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})]\hat{\hat\alpha}_4 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j} - \frac{Y_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_1 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j} - \frac{Y_{\cdot\cdot}Z_{\cdot\cdot}}{n_\cdot})\hat{\hat\beta}_2 - \frac{1}{4}(\sum_j \frac{Y_{j\cdot}Z\ldots}{n_j}$$

is the same as $\frac{X_{j\cdot}Y_{j\cdot}}{n_j}$ on this chart, etc.

FOR THE ADJUSTED INTERACTION SUM-S

$$\frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\hat{\beta}}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\hat{\beta}}_4 + [[\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{n.}) -$$

$$- \frac{X_{..}Y_{..}}{n.})]]\hat{\hat{\beta}}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\hat{\gamma}}_1 + [[$$

$$Y_{jk} - \frac{X_{..}Y_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})]]\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\hat{\gamma}}_1 - \frac{1}{4}($$

$$\frac{Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})\hat{\hat{\beta}}_3 + [[\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{n}) - \frac{1}{4}(\sum_j \frac{X_{j.}Y_{j.}}{n_j} - \frac{X_{..}Y_{..}}{n.})]]\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n.})\hat{\hat{\gamma}}_1 - \frac{1}{4}($$

$$- \frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}^2}{n_j} - \frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}^2}{n_j} - \frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_4 + [[\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{n.}) - \frac{1}{4}(\sum_j$$

$$\frac{Y_{..}^2}{n.})]]\hat{\hat{\beta}}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}^2}{n_j} - \frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}^2}{n_j} - \frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\gamma}}_1 + [[\sum_{j,}$$

$$)]]\hat{\hat{\beta}}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}^2}{n_j} - \frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}^2}{n_j} - \frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\gamma}}_1 - \frac{1}{4}(\sum_j$$

$$\frac{Y_{..}^2}{n.})\hat{\hat{\beta}}_3 + [[\sum_{j,k} Y_{jk}^2 - \frac{Y_{..}^2}{n.}) - \frac{1}{4}(\sum_j \frac{Y_{j.}^2}{n_j} - \frac{Y_{..}^2}{n.})]]\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\gamma}}_1 - \frac{1}{4}(\sum_j$$

$$- \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\beta}}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\beta}}_4 + [[\sum_{j,k} Z_{jk}^2 - \frac{Z_{..}^2}{n.}) - \frac{1}{4}(\sum_j$$

$$- \frac{Y_{..}Z_{..}}{n.})]]\hat{\hat{\beta}}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{Z_{j.}^2}{n_j} - \frac{Z_{..}^2}{n.})\hat{\hat{\gamma}}_1 + [[$$

$$- \frac{Y_{..}Z_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})]]\hat{\hat{\beta}}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{Z_{j.}^2}{n_j} - \frac{Z_{..}^2}{n.})\hat{\hat{\gamma}}_1 - \frac{1}{4}($$

$$- \frac{Y_{..}Z_{..}}{n.})\hat{\hat{\beta}}_3 + [[\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{n.}) - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n.})]]\hat{\hat{\beta}}_4 - \frac{1}{4}(\sum_j \frac{Z_{j.}^2}{n_j} - \frac{Z_{..}^2}{n.})\hat{\hat{\gamma}}_1 - \frac{1}{4}($$

QUARES

$$\frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})]\hat{\gamma}_1 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j}$$

$$\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j.}Z_{j.}}{n_j}) - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})]\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} -$$

$$\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})\hat{\gamma}_2 + [(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{n_.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})]\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})$$

$$\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.})\hat{\gamma}_3 + [(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{n_.}) - \frac{1}{4}(\sum_j \frac{X_{j.}Z_{j.}}{n_j} - \frac{X_{..}Z_{..}}{n_.}$$

$$\frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})]\hat{\gamma}_1 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.}$$

$$Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{n_.}) - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})]\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.}$$

$$\frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})\hat{\gamma}_2 + [(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{n_.}) - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})]\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.}$$

$$\frac{Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n_.})\hat{\gamma}_3 + [(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{n_.}) - \frac{1}{4}(\sum_j \frac{Y_{j.}Z_{j.}}{n_j} - \frac{Y_{..}Z_{..}}{n}$$

$$\frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})]\hat{\gamma}_1 - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.}).$$

$$\sum_{j,k} Z^2_{jk} - \frac{Z^2_{..}}{n_.}) - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})]\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}$$

$$\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}_2 + [(\sum_{j,k} Z^2_{jk} - \frac{Z^2_{..}}{n_.}) - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})]\hat{\gamma}_3 - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}_4$$

$$\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}_2 - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})\hat{\gamma}_3 + [(\sum_{j,k} Z^2_{jk} - \frac{Z^2_{..}}{n}) - \frac{1}{4}(\sum_j \frac{Z^2_{j.}}{n_j} - \frac{Z^2_{..}}{n_.})]\hat{\gamma}$$

$$\ldots - \frac{X_{..}Z_{..}}{n.}\Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} X_{jk} u_{1jk} - \frac{X_{..}U_{1..}}{n.} - \sum_j \frac{X_{j.}U_{1j.}}{n_j} + \frac{X_{..}U_{...}}{4n.}$$

$$\ldots \frac{X_{..}Z_{..}}{n.}\Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} X_{jk} u_{2jk} - \frac{X_{..}U_{2..}}{n.} - \sum_j \frac{X_{j.}U_{.j.}}{n_j} + \frac{X_{..}U_{...}}{4n.}$$

$$\ldots \hat{\hat{\gamma}}_4 = \sum_{j,k} X_{jk} u_{3jk} - \frac{X_{..}U_{3..}}{n.} - \sum_j \frac{X_{j.}U_{.j.}}{n_j} + \frac{X_{..}U_{...}}{4n.}$$

$$\ldots \Big]\hat{\hat{\gamma}}_4 = \sum_{j,k} X_{jk} u_{4jk} - \frac{X_{..}U_{4..}}{n.} - \sum_j \frac{X_{j.}U_{.j.}}{n_j} + \frac{X_{..}U_{...}}{4n.}$$

$$\ldots \Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} Y_{jk} u_{1jk} - \frac{Y_{..}U_{1..}}{n.} - \sum_j \frac{Y_{j.}U_{.j.}}{n_j} + \frac{Y_{..}U_{...}}{4n.}$$

$$\ldots \Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} Y_{jk} u_{2jk} - \frac{Y_{..}U_{2..}}{n.} - \sum_j \frac{Y_{j.}U_{.j.}}{n_j} + \frac{Y_{..}U_{...}}{4n.}$$

$$\ldots \Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} Y_{jk} u_{3jk} - \frac{Y_{..}U_{3..}}{n.} - \sum_j \frac{Y_{j.}U_{.j.}}{n_j} + \frac{Y_{..}U_{...}}{4n.}$$

$$\ldots \Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} Y_{jk} u_{4jk} - \frac{Y_{..}U_{4..}}{n.} - \sum_j \frac{Y_{j.}U_{.j.}}{n_j} + \frac{Y_{..}U_{...}}{4n.}$$

$$\ldots \hat{\hat{\gamma}}_4 = \sum_{j,k} Z_{jk} u_{1jk} - \frac{Z_{..}U_{1..}}{n.} - \sum_j \frac{Z_{j.}U_{.j.}}{n_j} + \frac{Z_{..}U_{...}}{4n.}$$

$$\ldots = \sum_{j,k} Z_{jk} u_{2jk} - \frac{Z_{..}U_{2..}}{n.} - \sum_j \frac{Z_{j.}U_{.j.}}{n_j} + \frac{Z_{..}U_{...}}{4n.}$$

$$\ldots = \sum_{j,k} Z_{jk} u_{3jk} - \frac{Z_{..}U_{3..}}{n.} - \sum_j \frac{Z_{j.}U_{.j.}}{n_j} + \frac{Z_{..}U_{...}}{4n.}$$

$$\ldots \hat{\hat{\gamma}}_4 = \sum_{j,k} Z_{jk} u_{4jk} - \frac{Z_{..}U_{4..}}{n.} - \sum_j \frac{Z_{j.}U_{.j.}}{n_j} + \frac{Z_{..}U_{...}}{4n.}$$

5

also,

$$U_{ij} = \sum_{k=1}^{n_j} u_{ijk}, \quad U_{i..} = \sum_{j=1}^{3} \sum_{k=1}^{n_j} u_{ijk}$$

$$U_{.j.} = \sum_{i=1}^{4} \sum_{k=1}^{n_j} u_{ijk} \quad U_{...} = \sum_{i=1}^{4} \sum_{j=1}^{3} \sum_{k=1}^{n_j} u_{ijk} \tag{4}$$

and finally,

$$n. = \sum_{j=1}^{3} n_j\,. \tag{5}$$

Since the error term which will be used is the same for all tests, the error term will be given first. To obtain it, one computes the following:

$$\sum_{i=1}^{4} \sum_{j=1}^{3} \sum_{k=1}^{n_j} u^2_{ijk} \tag{6}$$

and subtracts from it a term denoted as follows:

$$R[\mu, a, b, (ab), \alpha, \beta, \gamma]. \tag{7}$$

To obtain this term, one employs the system of equations given in Plate I. First, one calculates the following expression:

$$\sum_i \hat{\alpha}_i [\sum_{j,k} (x_{jk} - \bar{x}_{j.}) (u_{ijk} - \bar{u}_{ij.})] + \sum_i \hat{\beta}_i [\sum_{j,k} (y_{jk} - \bar{y}_{j.}) (u_{ijk} - \bar{u}_{ij})]$$

$$+ \sum_i \hat{\gamma}_i [\sum_{j,k} (z_{jk} - \bar{z}_{j.}) (u_{ijk} - \bar{u}_{ij})] = R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}), \tag{8}$$

which is the "sum of squares due to the concomitant variables." This can be obtained directly, without computing the individual values for the $\hat{\alpha}_i$'s, $\hat{\beta}_i$'s, and $\hat{\gamma}_i$'s from the Abbreviated Doolittle method. Then,

$$R[\mu, a, b, (ab), \alpha, \beta, \gamma] = \sum_{i,j} U^2_{ij.}/n_j + R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \tag{9}$$

and the error sum of squares is as follows:

$$\sum_{ijk} u^2_{ijk} - R[\mu, a, b, (ab), \alpha, \beta, \gamma], \tag{10}$$

which has $4n. - 24$ degrees of freedom.

Turning next to the tests corresponding to the two major questions (1 and 2) given at the beginning of this section (Tests of Hypotheses, With Formulas) we shall determine the relative complexity of the model. First, we shall begin with the first question stated above.

Question 1 (a). Test for the existence of interaction between estimator accuracy and metropolitan status. The hypothesis is expressed as follows:

$$H_o\text{: } (ab)_{ij} = 0, \text{ for all } i, j. \qquad (11)$$

To make the test, one computes a mean square for this hypothesis and divides it by the error mean square, to obtain an F statistic with six and 4n.—24 degrees of freedom.

The sum of squares for the hypothesis is computed as described below, and denoted by $R(ab|\mu, a, b, \alpha, \beta, \gamma)$. This term is equal to the following:

$$R[\mu, a, b, (ab), \alpha, \beta, \gamma] - R(\mu, a, b, \alpha, \beta, \gamma), \qquad (12)$$

where the first term was defined in Formula 9 and the computations for the second term are given below.

From the system of equations on Plate II, one computes the following:

$$\sum_{i} \hat{\alpha}_i \Big(\sum_{j,k} x_{jk} u_{ijk} - X..U_{i}../n. - \sum_{j} X_j.U._j./n_j + X..U.../4n.\Big)$$
$$+ \sum_{i} \hat{\beta}_i \Big(\sum_{j,k} y_{jk} u_{ijk} - Y..U_{i}../n. - \sum_{j} Y_j.U._j./n_j + Y..U.../4n.\Big)$$
$$+ \sum_{i} \hat{\gamma}_i \Big(\sum_{j,k} z_{jk} u_{ijk} - Z..U_{i}../n. - \sum_{j} Z_j.U._j./n_j + Z..U.../4n.\Big) = R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}), \qquad (13)$$

as before. This can be obtained by the Doolittle method, without calculating the terms individually. Then,

$$R(\mu, a, b, \alpha, \beta, \gamma) = -U^2.../4n. + \sum_{i} U^2_{i}../n. + \sum_{j} U^2._j./4n_j + R(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}), \qquad (14)$$

which is the sum of squares to test the hypothesis and denoted by $R[(ab)|\mu, a, b, \alpha, \beta, \gamma]$, which is Formula 7. This, in turn, becomes the following term:

$$\sum_{i,j} U^2_{ij}./n_j - \sum_{i} U^2_{i}../n. - \sum U^2._j./4n_j + U^2.../4n. + R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}). \qquad (15)$$

The last two quantities can be readily calculated on a high-speed computer. The others can be easily calculated with a desk calculator. These results can be precisely summarized in an analysis of variance table, as shown in Table 2, where the statistical test is also given.

The tests for the last three questions under I above are identical, differing only in the specific covariable used for the test. These three questions were as follows:

Does interaction exist between estimator accuracy, and
1 (b) population density?
1 (c) ratio of 1960 to 1950 births?
1 (d) ratio of 1960 to 1950 deaths?

We shall first consider the test for interaction between estimator accuracy and population density. The corresponding formulas for the remaining two tests for questions 1 (c) and 1 (d) will then be given briefly, without repeating the instructions.

Question 1 (b). The statistical form of the null hypothesis that there is no interaction between estimator accuracy and population density is as follows:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha. \tag{16}$$

To make this test, we employ the system of equations on Plate III, and from it compute the following:

$$\Big[\sum_{j,k} U_{\cdot jk}\, x_{jk} - \sum_{j} U_{\cdot j \cdot} X_{j \cdot}/n_j\Big]\tilde{\alpha}^* + \sum_{i}\Big[\sum_{j,k} y_{jk}\, u_{ijk} - \sum_{j} Y_{j \cdot} U_{ij \cdot}/n_j\Big]\tilde{\beta}^*_i$$
$$+ \sum_{i}\Big[\sum_{i,k} z_{jk}\, u_{ijk} - \sum_{j} Z_{j \cdot} U_{ij \cdot}/n_j\Big]\tilde{\gamma}^*_i = R(\tilde{\alpha}^*, \tilde{\beta}^*_i, \tilde{\gamma}^*_i). \tag{17}$$

The subscript i is used for β^* and γ^* to indicate that the hypothesis eliminates all subscripts on the α^*'s. This is calculated using the Abbreviated Doolittle method, without obtaining the individual values for the α^*'s, $\tilde{\beta}^*_i$'s and $\tilde{\gamma}^*_i$'s, as in the preceding tests. Then, the test is as follows:

$$\frac{[R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\tilde{\alpha}^*, \tilde{\beta}^*_i, \tilde{\gamma}^*_i)]/3}{\text{Error M.S.}} \sim F_{(3, 4n.-24)}. \tag{18}$$

Then one compares the above calculated number with the tabulated value in an F table, using 3 and 4n.-24 degrees of freedom. The error mean square was given in Formula 10 and in Table 2.

Question 1 (c). The corresponding test for interaction between estimator accuracy and the ratio of 1960 to 1950 births involves the system of equations of Plate IV, from which we obtain the following term:

PLATE III – REGRESSION ESTIMATES FOR TESTING INTERACTION BETW. POPULATION DENSITY & METROP. STATUS.

$$
\begin{aligned}
&\Big(\sum_{j,k}X_{jk}^2-\sum_j\tfrac{X_{j\cdot}^2}{n_j}\Big)\tilde\alpha^*+\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\beta_1^*+\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\beta_2^*+\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\beta_3^*+\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\beta_4^*\\
&\quad+\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_1^*+\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_2^*+\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_3^*+\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_4^*=\sum_{j,k}U_{\cdot jk}X_{jk}-\sum_j\tfrac{U_{\cdot j\cdot}X_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\alpha^*+\Big(\sum_{j,k}Y_{jk}^2-\sum_j\tfrac{Y_{j\cdot}^2}{n_j}\Big)\tilde\beta_1^*+0\cdot\tilde\beta_2^*+0\cdot\tilde\beta_3^*+0\cdot\tilde\beta_4^*+\Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\tfrac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_1^*+0\cdot\tilde\gamma_2^*+0\cdot\tilde\gamma_3^*+0\cdot\tilde\gamma_4^*=\sum_{j,k}u_{1jk}Y_{jk}-\sum_j\tfrac{U_{1j\cdot}Y_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\alpha^*+0\cdot\tilde\beta_1^*+\Big(\sum_{j,k}Y_{jk}^2-\sum_j\tfrac{Y_{j\cdot}^2}{n_j}\Big)\tilde\beta_2^*+0\cdot\tilde\beta_3^*+0\cdot\tilde\beta_4^*+0\cdot\tilde\gamma_1^*+\Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\tfrac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_2^*+0\cdot\tilde\gamma_3^*+0\cdot\tilde\gamma_4^*=\sum_{j,k}u_{2jk}Y_{jk}-\sum_j\tfrac{U_{2j\cdot}Y_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\alpha^*+0\cdot\tilde\beta_1^*+0\cdot\tilde\beta_2^*+\Big(\sum_{j,k}Y_{jk}^2-\sum_j\tfrac{Y_{j\cdot}^2}{n_j}\Big)\tilde\beta_3^*+0\cdot\tilde\beta_4^*+0\cdot\tilde\gamma_1^*+0\cdot\tilde\gamma_2^*+\Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\tfrac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_3^*+0\cdot\tilde\gamma_4^*=\sum_{j,k}u_{3jk}Y_{jk}-\sum_j\tfrac{U_{3j\cdot}Y_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\tfrac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde\alpha^*+0\cdot\tilde\beta_1^*+0\cdot\tilde\beta_2^*+0\cdot\tilde\beta_3^*+\Big(\sum_{j,k}Y_{jk}^2-\sum_j\tfrac{Y_{j\cdot}^2}{n_j}\Big)\tilde\beta_4^*+0\cdot\tilde\gamma_1^*+0\cdot\tilde\gamma_2^*+0\cdot\tilde\gamma_3^*+\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\gamma_4^*=\sum_{j,k}u_{4jk}Y_{jk}-\sum_j\tfrac{U_{4j\cdot}Y_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\alpha^*+\Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\tfrac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\beta_1^*+0\cdot\tilde\beta_2^*+0\cdot\tilde\beta_3^*+0\cdot\tilde\beta_4^*+\Big(\sum_{j,k}Z_{jk}^2-\sum_j\tfrac{Z_{j\cdot}^2}{n_j}\Big)\tilde\gamma_1^*+0\cdot\tilde\gamma_2^*+0\cdot\tilde\gamma_3^*+0\cdot\tilde\gamma_4^*=\sum_{j,k}u_{1jk}Z_{jk}-\sum_j\tfrac{U_{1j\cdot}Z_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\alpha^*+0\cdot\tilde\beta_1^*+\Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\tfrac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\beta_2^*+0\cdot\tilde\beta_3^*+0\cdot\tilde\beta_4^*+0\cdot\tilde\gamma_1^*+\Big(\sum_{j,k}Z_{jk}^2-\sum_j\tfrac{Z_{j\cdot}^2}{n_j}\Big)\tilde\gamma_2^*+0\cdot\tilde\gamma_3^*+0\cdot\tilde\gamma_4^*=\sum_{j,k}u_{2jk}Z_{jk}-\sum_j\tfrac{U_{2j\cdot}Z_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\alpha^*+0\cdot\tilde\beta_1^*+0\cdot\tilde\beta_2^*+\Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\tfrac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\beta_3^*+0\cdot\tilde\beta_4^*+0\cdot\tilde\gamma_1^*+0\cdot\tilde\gamma_2^*+\Big(\sum_{j,k}Z_{jk}^2-\sum_j\tfrac{Z_{j\cdot}^2}{n_j}\Big)\tilde\gamma_3^*+0\cdot\tilde\gamma_4^*=\sum_{j,k}u_{3jk}Z_{jk}-\sum_j\tfrac{U_{3j\cdot}Z_{j\cdot}}{n_j}\\[6pt]
&\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\tfrac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\alpha^*+0\cdot\tilde\beta_1^*+0\cdot\tilde\beta_2^*+0\cdot\tilde\beta_3^*+\Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\tfrac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde\beta_4^*+0\cdot\tilde\gamma_1^*+0\cdot\tilde\gamma_2^*+0\cdot\tilde\gamma_3^*+\Big(\sum_{j,k}Z_{jk}^2-\sum_j\tfrac{Z_{j\cdot}^2}{n_j}\Big)\tilde\gamma_4^*=\sum_{j,k}u_{4jk}Z_{jk}-\sum_j\tfrac{U_{4j\cdot}Z_{j\cdot}}{n_j}
\end{aligned}
$$

Note: $X_{j\cdot}Y_{j\cdot}/n_j$, as typed in the memo, is the same as $\frac{X_{j\cdot}Y_{j\cdot}}{n_j}$ on this chart, etc.

PLATE IV - REGRESSION ESTIMATES FOR TESTING INTERACTION BETW. BIRTH CHANGES & METROP. STATUS

$$(\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j})\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + (\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\beta}^* + (\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_1^* + 0\cdot\tilde{\gamma}_2^* + 0\cdot\tilde{\gamma}_3^* + 0\cdot\tilde{\gamma}_4^* = \sum_{j,k} u_{1jk}X_{jk} - \sum_j \frac{U_{1j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + (\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j})\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + (\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\beta}^* + 0\cdot\tilde{\gamma}_1^* + (\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_2^* + 0\cdot\tilde{\gamma}_3^* + 0\cdot\tilde{\gamma}_4^* = \sum_{j,k} u_{2jk}X_{jk} - \sum_j \frac{U_{2j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + (\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j})\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + (\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\beta}^* + 0\cdot\tilde{\gamma}_1^* + 0\cdot\tilde{\gamma}_2^* + (\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_3^* + 0\cdot\tilde{\gamma}_4^* = \sum_{j,k} u_{3jk}X_{jk} - \sum_j \frac{U_{3j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + (\sum_{j,k} X_{jk}^2 - \sum_j \frac{X_{j\cdot}^2}{n_j})\tilde{\alpha}_4^* + (\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\beta}^* + 0\cdot\tilde{\gamma}_1^* + 0\cdot\tilde{\gamma}_2^* + 0\cdot\tilde{\gamma}_3^* + (\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_4^* = \sum_{j,k} u_{4jk}X_{jk} - \sum_j \frac{U_{4j\cdot}X_{j\cdot}}{n_j}$$

$$(\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\alpha}_1^* + (\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\alpha}_2^* + (\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\alpha}_3^* + (\sum_{j,k} X_{jk}Y_{jk} - \sum_j \frac{X_{j\cdot}Y_{j\cdot}}{n_j})\tilde{\alpha}_4^* + 4(\sum_{j,k} Y_{jk}^2 - \sum_j \frac{Y_{j\cdot}^2}{n_j})\tilde{\beta}^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_1^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_2^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_3^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\gamma}_4^* = \sum_{j,k} U_{jk}Y_{jk} - \sum_j \frac{U_{j\cdot}Y_{j\cdot}}{n_j}$$

$$(\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\beta}^* + (\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j})\tilde{\gamma}_1^* + 0\cdot\tilde{\gamma}_2^* + 0\cdot\tilde{\gamma}_3^* + 0\cdot\tilde{\gamma}_4^* = \sum_{j,k} u_{1jk}Z_{jk} - \sum_j \frac{U_{1j\cdot}Z_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + (\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\beta}^* + 0\cdot\tilde{\gamma}_1^* + (\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j})\tilde{\gamma}_2^* + 0\cdot\tilde{\gamma}_3^* + 0\cdot\tilde{\gamma}_4^* = \sum_{j,k} u_{2jk}Z_{jk} - \sum_j \frac{U_{2j\cdot}Z_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + (\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\beta}^* + 0\cdot\tilde{\gamma}_1^* + 0\cdot\tilde{\gamma}_2^* + (\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j})\tilde{\gamma}_3^* + 0\cdot\tilde{\gamma}_4^* = \sum_{j,k} u_{3jk}Z_{jk} - \sum_j \frac{U_{3j\cdot}Z_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + (\sum_{j,k} X_{jk}Z_{jk} - \sum_j \frac{X_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\alpha}_4^* + (\sum_{j,k} Y_{jk}Z_{jk} - \sum_j \frac{Y_{j\cdot}Z_{j\cdot}}{n_j})\tilde{\beta}^* + 0\cdot\tilde{\gamma}_1^* + 0\cdot\tilde{\gamma}_2^* + 0\cdot\tilde{\gamma}_3^* + (\sum_{j,k} Z_{jk}^2 - \sum_j \frac{Z_{j\cdot}^2}{n_j})\tilde{\gamma}_4^* = \sum_{j,k} u_{4jk}Z_{jk} - \sum_j \frac{U_{4j\cdot}Z_{j\cdot}}{n_j}$$

Note: $X_j.Y_j./n_j$, as typed in the memo, is the same as $\frac{X_{j\cdot}Y_{j\cdot}}{n_j}$ on this chart, etc.

$$\sum_i \Big[\sum_{j,k} x_{jk} u_{ijk} - \sum_j X_{j\cdot} U_{ij\cdot}/n_j \Big] \tilde{\alpha}^*_i + \Big[\sum_{j,k} U_{\cdot jk}\, y_{jk} - \sum_j U_{\cdot j\cdot} Y_{j\cdot}/n_j \Big] \tilde{\beta}^*$$
$$+ \sum_i \Big[\sum_{j,k} z_{jk} u_{ijk} - \sum_j Z_{j\cdot} U_{ij\cdot}/n_j \Big] \tilde{\gamma}^*_i = R(\tilde{\alpha}^*_i, \tilde{\beta}^*, \tilde{\gamma}^*_i). \qquad (19)$$

The test for the null hypothesis that there is no interaction is then:

$$\frac{[R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\tilde{\alpha}^*_i, \tilde{\beta}^*, \tilde{\gamma}^*_i)]/3}{\text{Error M.S.}} \sim F_{(3,4n.-24)}. \qquad (20)$$

Question 1 (d). The test for interaction in the case of the ratio of the 1960 to 1950 deaths involves the system of equations on Plate V, from which we calculate the following:

$$\sum_i \Big[\sum_{j,k} x_{jk} u_{ijk} - \sum_j X_{j\cdot} U_{ij\cdot}/n_j \Big] \tilde{\alpha}^*_i + \sum_i \Big[\sum_{j,k} y_{jk} u_{ijk} - \sum_j Y_{j\cdot} U_{ij\cdot} \Big] \tilde{\beta}^*_i$$
$$+ \Big[\sum_{j,k} U_{\cdot jk}\, z_{jk} - \sum_j Z_{j\cdot} U_{\cdot j\cdot}/n_j \Big] \tilde{\gamma}^* = R(\tilde{\alpha}^*_i, \tilde{\beta}^*_i, \tilde{\gamma}^*). \qquad (21)$$

The test of the hypothesis is as follows:

$$\frac{[R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\tilde{\alpha}^*_i, \tilde{\beta}^*_i, \tilde{\gamma}^*)]/3}{\text{Error M.S.}} \sim F_{(3,4n.-24)}, \qquad (22)$$

which completes the tests for the four questions under 1 (a, b, c, d) stated at the beginning of this section.

Question 2. *Which variables add nothing more to what is known through other variables?*

To answer this question, one tests each variable separately by obtaining a sum of squares in which each specified variable has been ignored in the statistical model. The sum of squares obtained in this manner is then compared with the sum of squares obtained by using all the variables, and the difference is the numerator sum of squares in the F-test. This computational procedure will be described in detail for metropolitan classification, and for one of the three covariables (population density); the corresponding formulas for the other two covariables will be given briefly, without repeating the instructions.

Question 2(a). Does the metropolitan classification tell one anything more about estimator accuracy than do the covariables? We obtain the following term:

PLATE V – REGRESSION ESTIMATES FOR TESTING INTERACTION BETW. DEATH CHANGES & METROP. STATUS

$$\Big(\sum_{j,k}X_{jk}^2-\sum_j\frac{X_{j\cdot}^2}{n_j}\Big)\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + \Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\beta}_1^* + 0\cdot\tilde{\beta}_2^* + 0\cdot\tilde{\beta}_3^* + 0\cdot\tilde{\beta}_4^* + \Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{1jk}X_{jk}-\sum_j\frac{U_{1j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + \Big(\sum_{j,k}X_{jk}^2-\sum_j\frac{X_{j\cdot}^2}{n_j}\Big)\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + 0\cdot\tilde{\beta}_1^* + \Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\beta}_2^* + 0\cdot\tilde{\beta}_3^* + 0\cdot\tilde{\beta}_4^* + \Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{2jk}X_{jk}-\sum_j\frac{U_{2j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + \Big(\sum_{j,k}X_{jk}^2-\sum_j\frac{X_{j\cdot}^2}{n_j}\Big)\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + 0\cdot\tilde{\beta}_1^* + 0\cdot\tilde{\beta}_2^* + \Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\beta}_3^* + 0\cdot\tilde{\beta}_4^* + \Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{3jk}X_{jk}-\sum_j\frac{U_{3j\cdot}X_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + \Big(\sum_{j,k}X_{jk}^2-\sum_j\frac{X_{j\cdot}^2}{n_j}\Big)\tilde{\alpha}_4^* + 0\cdot\tilde{\beta}_1^* + 0\cdot\tilde{\beta}_2^* + 0\cdot\tilde{\beta}_3^* + \Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\beta}_4^* + \Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{4jk}X_{jk}-\sum_j\frac{U_{4j\cdot}X_{j\cdot}}{n_j}$$

$$\Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + \Big(\sum_{j,k}Y_{jk}^2-\sum_j\frac{Y_{j\cdot}^2}{n_j}\Big)\tilde{\beta}_1^* + 0\cdot\tilde{\beta}_2^* + 0\cdot\tilde{\beta}_3^* + 0\cdot\tilde{\beta}_4^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{1jk}Y_{jk}-\sum_j\frac{U_{1j\cdot}Y_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + \Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + 0\cdot\tilde{\beta}_1^* + \Big(\sum_{j,k}Y_{jk}^2-\sum_j\frac{Y_{j\cdot}^2}{n_j}\Big)\tilde{\beta}_2^* + 0\cdot\tilde{\beta}_3^* + 0\cdot\tilde{\beta}_4^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{2jk}Y_{jk}-\sum_j\frac{U_{2j\cdot}Y_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + \Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\alpha}_3^* + 0\cdot\tilde{\alpha}_4^* + 0\cdot\tilde{\beta}_1^* + 0\cdot\tilde{\beta}_2^* + \Big(\sum_{j,k}Y_{jk}^2-\sum_j\frac{Y_{j\cdot}^2}{n_j}\Big)\tilde{\beta}_3^* + 0\cdot\tilde{\beta}_4^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{3jk}Y_{jk}-\sum_j\frac{U_{3j\cdot}Y_{j\cdot}}{n_j}$$

$$0\cdot\tilde{\alpha}_1^* + 0\cdot\tilde{\alpha}_2^* + 0\cdot\tilde{\alpha}_3^* + \Big(\sum_{j,k}X_{jk}Y_{jk}-\sum_j\frac{X_{j\cdot}Y_{j\cdot}}{n_j}\Big)\tilde{\alpha}_4^* + 0\cdot\tilde{\beta}_1^* + 0\cdot\tilde{\beta}_2^* + 0\cdot\tilde{\beta}_3^* + \Big(\sum_{j,k}Y_{jk}^2-\sum_j\frac{Y_{j\cdot}^2}{n_j}\Big)\tilde{\beta}_4^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}u_{4jk}Y_{jk}-\sum_j\frac{U_{4j\cdot}Y_{j\cdot}}{n_j}$$

$$\Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\alpha}_1^* + \Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\alpha}_2^* + \Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\alpha}_3^* + \Big(\sum_{j,k}X_{jk}Z_{jk}-\sum_j\frac{X_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\alpha}_4^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\beta}_1^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\beta}_2^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\beta}_3^* + \Big(\sum_{j,k}Y_{jk}Z_{jk}-\sum_j\frac{Y_{j\cdot}Z_{j\cdot}}{n_j}\Big)\tilde{\beta}_4^* + \Big(\sum_{j,k}Z_{jk}^2-\sum_j\frac{Z_{j\cdot}^2}{n_j}\Big)\tilde{\gamma}^* = \sum_{j,k}U_{\cdot jk}Y_{jk}-\sum_j\frac{U_{\cdot j\cdot}Y_{j\cdot}}{n_j}$$

Note: $X_{j\cdot}Y_{j\cdot}/n_j$, as typed in the memo, is the same as $\frac{X_{j\cdot}Y_{j\cdot}}{n_j}$ on this chart, etc.

TABLE 2
ANALYSIS OF COVARIANCE TO TEST FOR THE INTERACTION OF METROPOLITAN CLASSIFICATION AND ESTIMATION METHODS

Source of Variation	Degrees of Freedom (d.f.)	Sum of Squares [SS(U)]	Reduction in Sum of Squares [R(SS)]	Adjusted Sum of Squares [SS (Adj.)]
Total	tn.	$\sum_{ijk} u^2_{ijk} = SS_{total}$		
Mean (μ)	1	$U^2_{...}/tn. = SS_M$		
Metropolitan Classification (Blocks) (b_j)	b	$\sum_{j} U^2_{.j.}/tn_j = SS_B$		
Estimators (Treatments) (a_i)	t	$\sum_{i} U^2_{i..}/n. = SS_T$		
Interaction $(ab)_{ij}$	tb—t—b+1	$\sum_{ij} U^2_{ij.}/n_j - SS_T - SS_B + SS_M = SS_I$		$SS_I(Adj) = SS_{I+E}(Adj) - SS_E(Adj)$
Error	tn. — tb	$\sum_{ijk} u^2_{ijk} - \sum_{ij} U^2_{ij.}/n_j = SS_E$	$R(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, (See Formula 8)	$SS_E - R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = SS_E(Adj)$
I + Error		$SS_{I+E} = SS_I + SS_E$	$R(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$, (See Formulas 14 & 15)	$SS_{I+E} - R(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) = SS_{I+E}(Adj)$

The test, then, is the quotient of the following two quantities:
Numerator: $A = [SS_{I+E}(Adj) - SS_E(Adj)] \;/\; [(tb-t-b+1)-3^*]$. This term is obtained from the Analysis of Covariance, using the difference between the I + E and the Error rows, in the SS (Adj) column $= SS_I(Adj) \;/\; [(tb-t-b+1)-3^*]$.
Denominator: B $= [SS_E - R(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})] \;/\; (tn.-2tb)$. Then, $A/B \sim F_{[(b-1)\ (t-1)-3^*,\ tn.-2tb]}$.

*Note: The number "3" refers to the number of covariables in the model; it will change, as one uses more (or fewer) covariables.

$$R(u, a, \alpha, \beta, \gamma) = \sum_i U^2_{i..}/n. + R(\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\hat{\gamma}}), \qquad (23)$$

where $R(\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\hat{\gamma}})$ is calculated from the set of equations on Plate VI, as follows:

$$R(\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\hat{\gamma}}) = \sum_i \hat{\hat{\alpha}}_i \left(\sum_{jk} x_{jk}u_{ijk} - X..U_{i..}/n. \right) + \sum_i \hat{\hat{\beta}}_i \left(\sum_{jk} y_{jk}u_{ijk} - Y..U_{i..}/n. \right)$$

$$+ \sum_{j,k} \hat{\hat{\gamma}}_i \left(\sum z_{jk}u_{ijk} - Z..U_{i..}/n. \right). \qquad (24)$$

Then, the sum of squares necessary to test the hypothesis is as follows:

$$R[\mu, a, b, (ab), \alpha, \beta, \gamma] - R(\mu, a, \alpha, \beta, \gamma)$$

$$= \sum_{i,j} U^2_{ij\cdot}/n_j - \sum_i U^2_{i..}/n. + R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\hat{\gamma}}), \qquad (25)$$

which has $\sum_{i,j} (n_j - 1) - (t-1) = tn.-bt-t+1$ degrees of freedom. The error sum of squares is as usual (see Analysis of Covariance Table 3).

Question 2 (b). Does the population density in the previous census year tell one anything more about estimator accuracy than do all of the other covariables and variables? Moreover, does metropolitan classification tell anyone more about estimator accuracy than do all of the covariables and variables? We proceed to calculate $R[\mu, a, b, (ab), \beta, \gamma]$

$$= \sum_{ijk} u^2_{ijk} - \sum_{i,j} U^2_{ij\cdot}/n_j + R(\tilde{\tilde{\beta}}, \tilde{\tilde{\gamma}}), \qquad (26)$$

where $R(\tilde{\tilde{\beta}}, \tilde{\tilde{\gamma}})$ is computed from the set of equations on Plate I, by deleting the top four equations and the terms involving the $\hat{\alpha}_i$'s in the remaining equations. Then, $R(\tilde{\tilde{\beta}}, \tilde{\tilde{\gamma}})$ is calculated from the remaining set in the usual fashion, as follows:

$$R(\tilde{\tilde{\beta}}, \tilde{\tilde{\gamma}}) = \sum_i \tilde{\tilde{\beta}}_i \left(\sum_{jk} y_{jk}u_{ijk} - \sum_j Y_{j\cdot}U_{ij\cdot}/n_j \right)$$

$$+ \sum_i \tilde{\tilde{\gamma}}_i \left(\sum_{jk} z_{ik}u_{ijk} - \sum_j Z_{j\cdot}U_{ij\cdot}/n_j \right). \qquad (27)$$

The sum of squares for the test of the hypothesis is as follows:

$$R[\mu, a, b, (ab), \alpha, \beta, \gamma] - R[\mu, a, b, (ab), \beta, \gamma]$$

$$= R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\tilde{\tilde{\beta}}, \tilde{\tilde{\gamma}}), \qquad (28)$$

with 4 degrees of freedom. The error sum of squares is the usual term.

PLATE VI - REGRESSION ESTIMATES FOR TESTING FOR REDUNDANCY OF METROP. STATUS

$$-\Big(\sum_{j,k} X_{jk}^2 - \frac{X_{..}^2}{4n.}\Big)\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{1jk}X_{jk} - \frac{U_{1..}X_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + \Big(\sum_{j,k} X_{jk}^2 - \frac{X_{..}^2}{4n.}\Big)\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{2jk}X_{jk} - \frac{U_{2..}X_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + \Big(\sum_{j,k} X_{jk}^2 - \frac{X_{..}^2}{4n.}\Big)\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{3jk}X_{jk} - \frac{U_{3..}X_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + \Big(\sum_{j,k} X_{jk}^2 - \frac{X_{..}^2}{4n.}\Big)\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{4jk}X_{jk} - \frac{U_{4..}X_{..}}{4n.}$$

$$-\Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + \Big(\sum_{j,k} Y_{jk}^2 - \frac{Y_{..}^2}{4n.}\Big)\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{1jk}Y_{jk} - \frac{U_{1..}Y_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + \Big(\sum_{j,k} Y_{jk}^2 - \frac{Y_{..}^2}{4n.}\Big)\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{2jk}Y_{jk} - \frac{U_{2..}Y_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + \Big(\sum_{j,k} Y_{jk}^2 - \frac{Y_{..}^2}{4n.}\Big)\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{3jk}Y_{jk} - \frac{U_{3..}Y_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + \hat{\hat{\alpha}}_3 + \Big(\sum_{j,k} X_{jk}Y_{jk} - \frac{X_{..}Y_{..}}{4n.}\Big)\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + \Big(\sum_{j,k} Y_{jk}^2 - \frac{Y_{..}^2}{4n.}\Big)\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{4jk}Y_{jk} - \frac{U_{4..}Y_{..}}{4n.}$$

$$-\Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + \Big(\sum_{j,k} Z_{jk}^2 - \frac{Z_{..}^2}{4n.}\Big)\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{1jk}Z_{jk} - \frac{U_{1..}Z_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + \Big(\sum_{j,k} Z_{jk}^2 - \frac{Z_{..}^2}{4n.}\Big)\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{2jk}Z_{jk} - \frac{U_{2..}Z_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\alpha}}_3 + 0\cdot\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\beta}}_3 + 0\cdot\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + \Big(\sum_{j,k} Z_{jk}^2 - \frac{Z_{..}^2}{4n.}\Big)\hat{\hat{\gamma}}_3 + 0\cdot\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{3jk}Z_{jk} - \frac{U_{3..}Z_{..}}{4n.}$$

$$-\ 0\cdot\hat{\hat{\alpha}}_1 + 0\cdot\hat{\hat{\alpha}}_2 + 0\cdot\hat{\hat{\alpha}}_3 + \Big(\sum_{j,k} X_{jk}Z_{jk} - \frac{X_{..}Z_{..}}{4n.}\Big)\hat{\hat{\alpha}}_4 + 0\cdot\hat{\hat{\beta}}_1 + 0\cdot\hat{\hat{\beta}}_2 + 0\cdot\hat{\hat{\beta}}_3 + \Big(\sum_{j,k} Y_{jk}Z_{jk} - \frac{Y_{..}Z_{..}}{4n.}\Big)\hat{\hat{\beta}}_4 + 0\cdot\hat{\hat{\gamma}}_1 + 0\cdot\hat{\hat{\gamma}}_2 + 0\cdot\hat{\hat{\gamma}}_3 + \Big(\sum_{j,k} Z_{jk}^2 - \frac{Z_{..}^2}{4n.}\Big)\hat{\hat{\gamma}}_4 = \sum_{j,k} u_{4jk}Z_{jk} - \frac{U_{4..}Z_{..}}{4n.}$$

TABLE 3

ANALYSIS OF COVARIANCE TO TEST FOR THE USEFULNESS OF METROPOLITAN CLASSIFICATION

Source of Variation	Degrees of Freedom (d.f.)	Sum of Squares [SS (U)]	Reduction in Sum of Squares [R (SS)]	Adjusted Sum of Squares [SS (Adj.)]
Total	tn.	$\sum_{ijk} u^2_{ijk} = SS_{total}$		
Mean (μ)	1	$U^2.../tn. = SS_M$		
Metropolitan Classification (Blocks) (b_j)	b	$\sum_{j} U._j./tn_j = SS_B$		$SS_{B+I+E}(Adj) — SS_E(Adj) = SS_{B+I}(Adj)$
Estimators (Treatments) (a_i)	t	$\sum_{i} U^2_{i}../n. = SS_T$		
Interaction $(ab)_{ij}$	tb—t—b+1	$\sum_{ij} U^2_{ij}/n_j — SS_T — SS_B + SS_M = SS_I$		
Error	tn. — bt	$\sum_{ijk} u^2_{ijk} — \sum_{ij} U^2_{ij}./n_j = SS_E$	$R(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ (see Formula 8)	$SS_E — R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = SS_E(Adj)$
B + I + E	tn. —t+1	$SS_B + SS_J + SS_E = SS_{B+I+E}$	$R(\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\hat{\gamma}})$ (see Formula 24)	$SS_{B+I+E} — R(\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\hat{\gamma}}) = SS_{B+I+E}(Adj)$

The test is the quotient of the following two quantities:

Numerator: $A = [SS_{B+I+E}(Adj) — SS_E(Adj)] / (tb—t+1)$. This term is obtained from the Analysis of Covariance, using the difference between the B+I+E, and the Error rows in the SS (Adj) column $= SS_{B+I}(Adj) / (tb—t+1)$.

Denominator: $B = [SS_E — R(\hat{\alpha}, \hat{\beta}, \hat{\gamma})] / (tn.—2tb)$. Then, $A/B \sim F_{(tb-t+1,\ tn.-2tb)}$.

The test for the ratio of the number of 1960 to 1950 births is based upon computations from the set of equations on Plate I, by deleting the middle four equations, and the terms involving the $\hat{\beta}_i$'s in the remaining equations. Then, we obtain the following expression:

$$R(\tilde{\tilde{\alpha}}, \tilde{\tilde{\gamma}}) = \sum_{i} \tilde{\tilde{\alpha}}_i \Big(\sum_{jk} x_{jk}u_{ijk} - \sum_{j} X_{j\cdot}U_{ij\cdot}/n_j\Big) + \sum_{i} \tilde{\tilde{\gamma}}_i \Big(\sum_{jk} z_{jk}u_{ijk} - \sum_{j} Z_{j\cdot}U_{ij\cdot}/n_j\Big). \qquad (29)$$

The sum of squares for the test of the hypothesis is as follows:

$$R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\tilde{\tilde{\alpha}}, \tilde{\tilde{\gamma}}), \qquad (30)$$

with 4 degrees of freedom.

To test for the usefulness of the ratio of the number of 1960 to 1950 deaths over and above the other variables, the equations are obtained by deleting the bottom four equations on Plate I, and deleting the terms involving $\hat{\gamma}_i$'s in the remaining equations. The numerator sum of squares for the test of the hypothesis is as follows:

$$R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - R(\tilde{\tilde{\alpha}}, \tilde{\tilde{\beta}}), \text{ where} \qquad (31)$$

$$R(\tilde{\tilde{\alpha}}, \tilde{\tilde{\beta}}) = \sum_{i} \tilde{\tilde{\alpha}}_i \Big(\sum_{j,k} x_{jk}u_{ijk} - \sum_{j} X_{j\cdot}U_{ij\cdot}/n_j\Big) + \sum_{i} \tilde{\tilde{\beta}}_i \Big(\sum_{j,k} y_{jk}u_{ijk} - \sum_{j} Y_{j\cdot}U_{ij\cdot}/n_j\Big), \qquad (32)$$

which has 4 degrees of freedom.

Summary

A statistical model was developed in this chapter to test the accuracy of four different methods of estimating county population. The procedure formulated for the tests of accuracy employed the common logarithms of the absolute differences between April 1 county population estimates and April 1 federal decennial census enumerations. Then F- and t-tests were used to test for significant differences.

The statistical model is a multiple covariance model containing two independent variables, one two-factor interaction term, and three covariables. The independent variables are estimation methods and the metropolitan classification in the previous census decade; the one two-factor interaction is for estimation method and metropolitan classification; and the three concomitant variables are population density in the previous census decade, ratio of 1960 to 1950 births, and ratio of 1960 to 1950 deaths. The tests of the hypotheses for each of these independent variables and covariables are presented fully in this chapter.

Chapter V

PRELIMINARY TESTS OF ACCURACY

This chapter applies specified variations of the basic mathematical model formulated in Chapter IV in evaluating the accuracy of selected methods of estimating county population. In addition, this chapter reports the results of the preliminary tests of accuracy which were made of selected county population estimates prepared by other agencies.

The first major part of this chapter analyzes the accuracy of various methods of estimating the April 1, 1950 and April 1, 1960 population of metropolitan counties in the United States. The estimates used in this preliminary test were prepared by the Bureau of the Census and were summarized in Table 1, columns 1 and 2.

The second major part of this chapter analyzes the accuracy of various methods of estimating the April 1, 1960 population of counties in Montana, Ohio, Oregon, and Pennsylvania. The estimates used in this preliminary test were prepared by the Study Group on Postcensal Population Estimates, Public Health Conference on Records and Statistics, and were summarized in Table 1, columns 3, 4, 5, and 6.

The basic purpose of the preliminary tests of accuracy was two-fold: *First,* to determine whether one of the methods used in preparing the April 1, 1950 and April 1, 1960 county population estimates summarized in Table 1, columns 1 to 7, was superior to the other methods tested. Conflicting findings had been reported in the published evaluation studies which presented the tests of accuracy (see Chapter III); and *Second,* to establish the complexity and appropriateness of the basic statistical model formulated in Chapter IV and to experiment with variations of that statistical model.

1950 and 1960 Population Estimates of Metropolitan Counties

Three separate analyses of the metropolitan county population estimates were undertaken in the preliminary tests of accuracy: the first analyzed errors in the April 1, 1950 county population estimates; the second analyzed errors in the April 1, 1960 county population estimates; and the third analyzed errors in the combined 1950 and 1960 county population estimates prepared by identical methods. Each of the three analyses employed a specified variation of the basic statistical model given in Chapter IV.

1950 Estimates. Errors in the April 1, 1950 population estimates prepared by the Bureau of the Census for 102 metropolitan counties by the following five methods were analyzed: Component Method II, Component Method I, the Vital Rates Method, the Arithmetic Extrapolation Method, and Geometric Extrapolation Method. The mathematical model employed in this evaluation test was the following:

$$U_{ijk} = \mu + \alpha_i + \tau_j + \beta_i Y_{jk} + \gamma_i Z_{jk} + \varepsilon_{ijk}, \qquad (1)$$

where α_i is the fixed effect on the accuracy due to estimation method i, with $i = 1, \ldots, 5$; τ_j is the fixed effect on the accuracy due to the population density classification in the previous census year (April 1, 1940), with $j = 1, \ldots, 5$ (1 = counties with less than 250 people per square mile; 2 = 250-499; 3 = 500-749; 4 = 750-999; and 5 = 1,000 people and over per square mile); Y_{jk}, a covariable, is the ratio of the number of 1960 to the 1950 births of the k^{th} county, within population density classification j; Z_{jk}, a covariable, is the ratio of the number of 1960 to the 1950 deaths of the k^{th} county, within population density classification j; β_i and γ_i are partial regression coefficients associated with Y_{jk} and Z_{jk}, respectively, for each of the five estimation methods, α_i; μ is the overall mean; U_{ijk} is the logarithm of the absolute difference of the April 1, 1950 census enumeration and the April 1, 1950 population estimate by the i^{th} method, in the j^{th} population density classification, for the k^{th} county. This multiple covariance model assumes that the epsilons (ε_{ijk}'s) are normally distributed, with a mean of zero and a variance of sigma squared [$\varepsilon \sim N(O, \sigma^2)$].

The least squares method of solving simultaneous equations was employed to test the following six hypotheses:

1. H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_5$.
2. H_0: $\tau_1 = \tau_2 = \ldots = \tau_5$.
3. H_0: $\beta_1 = \beta_2 = \ldots = \beta_5 = \beta$.
4. H_0: $\gamma_1 = \gamma_2 = \ldots = \gamma_5 = \gamma$.
5. H_0: $\beta = 0$.
6. H_0: $\gamma = 0$.

The procedure employed in testing each of the six hypotheses was to compute the reduction in the sum of squares which each independent variable and covariable explained, after first adjusting for the effect of every other independent variable and covariable in the model. For example, in testing for the significance of the reduction in the sum of squares in model one (formula 1) due to α (hypothesis number 1), the following computation was made: $R(\alpha|\mu, \tau, \beta, \gamma) = R(\mu, \alpha, \tau, \beta, \gamma) - R(\mu, \tau, \beta, \gamma)$. All of the adjusted sums of squares shown in Table 4 were computed in this manner, using the entire computational model. This procedure of testing hypotheses gives a precise measurement of the influence of each independent variable and covariable in the model and provides an exact test of the hypotheses.[1]

The analysis of variance of the logarithms of the absolute differences between the April 1, 1950 population estimates and the enumerated populations of 102 metropolitan counties showed no significant differences in estimation accuracy among the five estimation methods (Table 4). The mean log differences for the five methods, in the order of their accuracy, were as follows: Component Method II, 4.006; the Vital Rates Method, 4.025; Component Method I, 4.187; the Geometric Extrapolation Method, 4.526; and the Arithmetic Extrapolation Method, 4.540.

Population density of the counties in 1940 was the only significant independent variable related to estimator accuracy.[2] The accuracy of the 1950 county population estimates was consistently greater for the sparsely than for the densely populated counties, regardless of which of the estimation methods was used. As population density in 1940 increased, the error in the population estimates increased. The mean log differences of counties in the five classes of 1940 population densities were as follows: Under 250 people per square mile, 3.910; 250-499, 4.163; 500-749, 4.415; 750-999, 4.569; and 1,000 people per square mile and over, 5.109.

Therefore, only one of the six hypotheses tested in Table 4 for model one, hypothesis number 2, was rejected.

1960 Estimates. Errors in April 1, 1960 population estimates prepared by the Bureau of the Census for 132 counties in 46 of the large SMSA's by the Component Method II and the Vital Rates Method were

---

[1] Franklin A. Graybill, *An Introduction to Linear Statistical Models,* Volume I (New York: McGraw-Hill Book Company, Inc., 1961), Chapters 6 and 13.

[2] The two calculated variance ratios of 4.78 and 4.32 in Table 4, were not considered significant because the small error mean square of .2079 makes the tests extremely sensitive to very small differences in population estimation errors. Moreover, these two calculated variance ratios were relatively small compared to the calculated F of 79.77 for the covariable population density in 1940.

TABLE 4
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1950 POPULATION ESTIMATES OF 102 COUNTIES IN SELECTED 1950 SMA'S, PREPARED BY COMPONENT METHOD II, COMPONENT METHOD I, THE VITAL RATES METHOD, AND THE ARITHMETIC EXTRAPOLATION AND GEOMETRIC EXTRAPOLATION METHODS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	510	9,444.9488		
R (μ)	1	9,242.1343		
R (due to Model $\mid \mu$)	26	102.3772		
Estimation Methods R[α (adjusted)]	4	1.2886	0.3222	1.55
Population Density R[τ_j (adj.)]	4	66.3379	16.5845	79.77**
Ratio of 1960 to 1950 births R[β (adj.)]	5	4.9696	0.9939	4.78[1]
Ratio of 1960 to 1950 deaths R[γ (adj.)]	5	1.1694	0.2339	1.13
R[(β_i) — R(β) adj.)][2]	4	0.6130	0.1532	<1
R[(γ_i) — R(γ) adj.)]	4	3.5912	0.8978	4.32[1]
Error	483	100.4373	0.2079	

* One and two asterisks indicate significance at the five and one percent levels, respectively.

[1] Judged nonsignificant.

[2] Sum of squares were computed in the following manner: $R(\mu, \alpha_i, \tau_j, \beta_i, \gamma_i) - R(\mu, \alpha_i, \tau_j, \beta, \gamma_i)$, which may be interpreted as the sum of squares attributable to the model $Y = \mu + \alpha_i + \tau_j + \beta_i Y_{jk} + \gamma_i Z_{jk} + \epsilon_{ijk}$ minus the sum of squares attributable to the model $Y = \mu + \alpha_i + \tau_j + \beta Y_{jk} + \gamma_i Z_{jk} + \varepsilon_{ijk}$, which is used to test hypothesis number 3.

analyzed by a modified version of the basic statistical model formulated in Chapter IV. The mathematical model employed was the following:

$$U_{ijk} = \mu + \alpha_i + \tau_j + \beta Y_{jk} + \gamma Z_{jk} + \varepsilon_{ijk}, \qquad (2)$$

where α_i, τ_j, Y_{jk}, and Z_{jk} were previously defined in model 1 (formula 1); in this model $\alpha_i = 1$ and 2; β and γ, previously defined in model 1, are partial regression coefficients associated with the covariables Y_{jk} and Z_{jk}.

The following four hypotheses were tested in this analysis:

1. H_0: $\alpha_1 = \alpha_2$.
2. H_0: $\tau_1 = \tau_2 = \ldots = \tau_5$.
3. H_o: $\beta = 0$.
4. H_o: $\gamma = 0$.

The analysis of variance of the April 1, 1960 population estimates of the selected metropolitan counties revealed no significant differences between the Vital Rates and Component Method II estimates, although the Vital Rates Method was slightly more accurate (the log means were 4.005 for the Vital Rates Method and 4.094 for Component Method II). Population density of the counties in 1950 was closely related to estimator accuracy, with the estimation method being more accurate in the sparsely inhabited counties than in the densely inhabited counties (Table 5). The log means ranged from 3.687 for counties with fewer than 250 people per square mile in 1950 to 4.472 for counties with 1,000 and over people per square mile in 1950. This finding corroborates that noted for metropolitan counties for the year 1950 (Table 4).

The ratio of the 1960 to 1950 deaths had a significant effect upon the accuracy of county population estimation in 1960 (Table 5). As the number of deaths increased, the estimation errors increased.

Two of the four hypotheses tested in Table 5 for model two, hypotheses numbers 2 and 4, were rejected.

1950 and 1960 Estimates Combined. The 1950 population estimates of the 102 metropolitan counties and the 1960 estimates of the 132 metropolitan counties computed both by the Vital Rates Method and Component Method II were combined into one analysis to determine whether the two estimation methods were more accurate in one year than in another and to determine whether one estimation method might prove more precise in one year, whereas the other method might prove more precise in another year.

The following mathematical model was used in this analysis:

$$U_{ijnk} = \mu + \alpha_i + \lambda_m + (\alpha\lambda)_{im} + \tau_j + \beta Y_{jmk} + \gamma Z_{jmk} + \varepsilon_{ijmk}, \quad (3)$$

where α_i, τ_j, Y_{jmk}, Z_{jmk}, β, and γ have the same definition as in Model 2; λ_m is the fixed effect on the accuracy due to the years, with m = 1 (1950) and 2 (1960); and $(\alpha\lambda)_{im}$ is a fixed effect, showing the interaction between α_i and λ_m (estimation methods and years interaction).

The following six hypotheses were tested:

1. H_0: $\alpha_1 = \alpha_2$.
2. H_0: $\lambda_1 = \lambda_2$.
3. H_0: $(\alpha\lambda) = 0$.
4. H_0: $\tau_1 = \tau_2 = \ldots = \tau_5$.
5. H_0: $\beta = 0$.
6. H_0: $\gamma = 0$.

Table 6 shows that there was a significant difference in the accuracy of the estimation methods between the two census years, 1950 and 1960. The 1960 population estimates were more accurate for the metropolitan counties than were the 1950 population estimates, with the calculated regression coefficients for 1950 and 1960, respectively, being $\hat{\lambda}_1 = 1.194$ and $\hat{\lambda}_2 = 0$. Nevertheless, there was no significant difference between the accuracy of the two methods in the year 1950, the year 1960, or both years 1950 and 1960 combined. Although the Vital Rates Method gave smaller errors in county population estimates for 1950 and 1960 combined than did Component Method II, the differences were not significant. The log means were 4.014 for the Vital Rates Method and 4.056 for Component Method II. Moreover, the results indicated that the two-factor interaction of estimation methods and years was not significant.

The accuracy of the 1950 and 1960 county population estimates decreased consistently as population density of counties in the previous decade increased. In addition, the precision of the 1950 and 1960 county population estimates decreased as the ratio of the 1960 to 1950 deaths increased. These two findings were consistent with those for the 1960 metropolitan county results shown in Table 5.

Thus, three of the six hypotheses tested in Table 6 for Model 3, hypotheses numbers 2, 4, and 6, were rejected.

Since all of the counties included in the above three tests were in the metropolitan areas of the United States, the main effect of metropolitan classification, which was included in the basic statistical model given in Chapter IV, was deleted. The variable population density, classified into five groups, was used instead, for both variables measure practically the same thing.

TABLE 5
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF 132 COUNTIES IN 46 SELECTED 1960 SMSA'S, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	264	4,448.3767		
R (μ)	1	4,328.9965		
R (due to Model $\mid \mu$)	7	41.2257		
Estimation Methods R[α (adjusted)]	1	.5254	.5254	1.72
Population Density R[τ (adj.)]	4	34.3185	8.5796	28.10**
Ratio of 1960 to 1950 Births				
R[β (adj.)]	1	.0570	.0570	<1
Ratio of 1960 to 1950 Deaths				
R[γ (adj.)]	1	6.6015	6.6015	21.62**
Error	256	78.1545	.3053	

* One and two asterisks indicate significance at the five and one percent levels, respectively.

TABLE 6
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1950 AND THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN 1950 SMA'S AND 1960 SMSA'S, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	468	7,816.4168		
R (μ)	1	7,618.5884		
R (due to model $\mid \mu$)	9	64.2056		
Estimation Methods R [α (adjusted)]	1	.5185	.5185	1.78
Years R [λ (adj.)]	1	14.7345	14.7345	50.50**
Estimation Methods $\times$ Years Interaction				
R [$\alpha\lambda$ (adj.)]	1	.3309	.3309	1.13
Population Density R [τ (adj.)]	4	56.7122	14.1781	48.59**
Ratio of 1960 to 1950 Births R [β (adj.)]	1	.0195	.0195	<1
Ratio of 1960 to 1950 Deaths R [γ (adj.)]	1	6.0379	6.0379	20.69**
Error	458	133.6228	.2918	

*One and two asterisks indicate significance at the five and one percent levels, respectively.

1960 Population Estimates of Montana, Ohio, Oregon, and Pennsylvania Counties

Population estimates of the 247 counties in these four states were prepared in conjunction with the specially organized Study Group on Postcensal Population Estimates. April 1, 1960 estimates for Ohio counties were computed by four different methods: Component Method II, the Vital Rates Method, Bogue-Duncan Composite Method, and the Arithmetic Extrapolation Method; estimates for Oregon and Pennsylvania counties were computed by three different methods: Component Method II, the Vital Rates Method, and the Arithmetic Extrapolation Method; and estimates for Pennsylvania counties were computed by two methods: Component Method II and the Vital Rates Method.

The following mathematical model was used in analyzing errors in the 1960 postcensal county population estimates in these four states:

$$U_{ink} = \mu + \alpha_i + \omega_n + \delta_i X_{nk} + \beta_i Y_{nk} + \gamma_i Z_{nk} + \varepsilon_{ink}, \qquad (4)$$

where α_i stands for estimation methods, with i = 1, 2, 3, and 4; ω_n is the fixed effect on the accuracy due to states, with n = 1, 2, 3, and 4; X_{nk}, a covariable, is the population density in the previous census year (April 1, 1950) of the kth county within the nth state; Y_{nk}, Z_{nk}, β_i, and γ_i were defined previously in model one (formula 1).

The following eight hypotheses were tested:

1. H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$.
2. H_0: $\omega_1 = \omega_2 = \omega_3 = \omega_4$.
3. H_0: $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta$.
4. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$.
5. H_0: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma$.
6. H_0: $\delta = 0$.
7. H_0: $\beta = 0$.
8. H_0: $\gamma = 0$.

The analysis of variance indicated no significant differences in the accuracy of the four estimation methods, even though the Composite Method had the smallest error of the four methods (Table 7). For all four states combined, the mean log differences for the four methods were as follows: the Composite Method, 3.013; the Vital Rates Method, 3.056; the Arithmetic Extrapolation Method, 3.075; and Component Method II, 3.186.

The 1960 county population estimates of these four states showed that Method II had the largest errors of the four different methods, although the difference was not significant. Undoubtedly, the major reason for this result was that reliable county school enrollments did not exist in Ohio and Pennsylvania, the two states having the largest overall estimation errors. In both states, Component Method II estimates were prepared using school census data rather than enrollment data.[3]

Table 7 shows there was a definite difference in the accuracy of the April 1, 1960 county population estimates among the four states. Montana county population estimates were most accurate (a log mean of 2.774), followed, in order, by Oregon (3.037), Pennsylvania (3.205), and Ohio (3.230), with the overall log mean of counties in the four states being 3.098. T-tests indicated that the errors in the county estimates for Montana were significantly smaller than for counties in Oregon, Ohio, and Pennsylvania; also that the errors in the Oregon county estimates were smaller than for counties in Ohio.

One of the primary factors which accounted for the differences in estimation accuracy of counties in the four states was the marked variability in the average number of inhabitants in each county, for errors in county population estimates increase directly with the population size of the county, as was shown in Tables 4, 5, and 6. Montana is predominantly a rural state, with an average of only 12,049 residents per county in 1960. Oregon had an average of 49,130 persons per county in 1960. On the other hand, the counties in Ohio and Pennsylvania were densely populated, averaging 110,300 and 168,946 residents, respectively, per county in 1960. Therefore, the larger errors in Pennsylvania and Ohio county population estimates as compared to Montana and Oregon were consistent with expectations. However, the larger errors for Ohio than for Pennsylvania counties may have been attributable, at least in part, to faulty school census data used in preparing Component Method II county population estimates.

The errors in the population estimates of Montana, Oregon, Pennsylvania, and Ohio counties increased directly with increasing population density in 1950 and directly as the ratios of 1960 to 1950 births and deaths increased (Table 7). Furthermore, the analysis showed these three covariables (population density, ratio of 1960 to 1950 births, and ratio of 1960 to 1950 deaths) did not differ significantly for each of the four estimation methods, as the partial regression coefficients were approximately the same for each estimation method. Accordingly, four of the eight hypotheses tested in Table 7 for Model 4, hypotheses numbers 2, 6, 7, and 8, were rejected.

---

[3] U. S. Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, National Vital Statistics Division, The Public Health Conference on Records and Statistics, *Preliminary Report of the Study Group on Postcensal Population Estimates*, Document No. 520.6–6/11/62, pp. 4-6.

In a second analysis of the 1960 population estimates, the 67 Pennsylvania counties were deleted to provide a precise comparison of the accuracy of Component Method II, the Vital Rates Method, and the Arithmetic Extrapolation Method (Table 8). The exclusion of Pennsylvania counties altered somewhat the findings summarized in Table 7.

The mathematical model used in evaluating the estimation errors in the 1960 population of counties in Montana, Ohio, and Oregon was the following:

$$U_{ink} = \mu + \alpha_i + \omega_n + (\alpha\omega)_{in} + \delta_i X_{nk} + \beta_i Y_{nk} + \gamma_i Z_{nk} + \varepsilon_{ink}, \quad (5)$$

where the symbol α_i stands for estimation methods, with $i = 1, 2,$ and 3; the symbol ω_n stands for the classification of states, with $n = 1, 2,$ and 3. All of the other covariables and the one two-factor interaction have been defined previously.

Nine hypotheses were tested in this model. The first eight corresponded with those tested for model 4, except for hypotheses numbers 1 and 2, where the number of methods changed from 4 to 3 and the number of states changed from 4 to 3. The ninth hypothesis tested for Model 5 was the following:

9. H_0: $(\alpha\omega)_{in} = 0$.

Table 8 indicates that there were significant differences in the accuracy of the estimation methods in the three states. The mean log differences for counties in Montana were 2.774, in Oregon 3.037, and in Ohio 3.303. T-tests revealed that the Montana county population estimates had smaller errors than the Oregon and Ohio county population estimates.

The three estimation methods are of approximately equal precision, with the following log means: 3.038 for the Vital Rates Method; 3.075 for the Arithmetic Extrapolation Method; and 3.141 for Component Method II.

Two of the three covariables (the ratio of 1960 to 1950 births, and the ratio of 1960 to 1950 deaths) exerted a significant influence upon estimator accuracy. The relationship was positive, as previously established in other tests. Moreover, Table 8 indicates that the two-factor interaction between estimation methods and states was not significant.

Three of the nine hypotheses tested in Table 8 for Model 5, hypotheses numbers 2, 7, and 8, were rejected.

TABLE 7
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN MONTANA, OHIO, OREGON, AND PENNSYLVANIA, PREPARED BY COMPONENT METHOD II, THE VITAL RATES METHOD, THE BOGUE-DUNCAN COMPOSITE METHOD, AND BY THE ARITHMETIC EXTRAPOLATION METHOD

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	762	7,651.5216		
R (μ)	1	7,396.1405		
R (due to Model $\mid \mu$)	18	30.0000		
States R [ω (adjusted)]	3	15.8283	5.2761	15.35**
Estimation Methods R [α (adjusted)]	3	1.4591	0.4864	1.42
Population Density in 1950 [δ (adj.)]	1	14.0147	14.0147	37.50**
Ratio of 1960 to 1950 Births [β (adj.)]	1	4.8447	4.8447	12.96**
Ratio of 1960 to 1950 Deaths [γ (adj.)]	1	11.9388	11.9388	31.95**
R [(δ_1) — R(δ) adj.]	3	1.1431	0.3810	1.11
R [(β_1) — R(β) adj.]	3	1.3598	0.4533	1.32
R [(γ_1) — R(γ) adj.]	3	1.2781	0.4260	1.14
Error	743	225.3811	0.3737	

* One and two asterisks indicate significance at the five and one percent levels, respectively.

TABLE 8
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN MONTANA, OHIO, AND OREGON, PREPARED BY COMPONENT METHOD II, THE VITAL RATES METHOD, AND THE ARITHMETIC EXTRAPOLATION METHOD

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	540	5,356.8721		
R (μ)	1	5,138.6290		
R (due to Model $\mid \mu$)	17	61.0375		
States R [ω (adjusted)]	2	12.0175	6.0088	19.95**
Estimation Methods R [α (adj.)]	2	5.0403	2.5202	8.37[1]
Estimation Methods × States Interaction R [$\alpha\omega$ (adj.)]	4	7.9933	1.9983	6.63[1]
Population Density in 1950 R [δ (adj.)]	1	2.1992	2.1992	7.30[1]
Ratio of 1960 to 1950 Births R [β (adj.)]	1	3.9486	3.9486	13.11**
Ratio of 1960 to 1950 Deaths R [γ (adj.)]	1	7.0970	7.0970	23.56**
R [(δ_1) — R(δ) adj.]	2	3.6492	1.8246	6.06[1]
R [(β_1) — R(β) adj.]	2	2.4408	1.2204	4.05[1]
R [(γ_1) — R(γ) adj.]	2	0.0855	0.0428	<1
Error	522	157.2056	0.3012	

* One and two asterisks indicate significance at the five and one percent levels, respectively.
[1] Judged nonsignificant.

Summary

This chapter analyzed errors in the 1950 and 1960 population estimates of counties which were computed by selected estimation methods. In this preliminary analysis, five multiple covariance regression models were used in computing analyses of covariance of the common logarithms of the absolute differences between the April 1 county population estimates and census enumerations.[4] In the five tests, a total of six different estimation methods were employed. The major results of the five separate analyses are summarized as follows: *First,* no significant differences were found in the accuracy of the various methods of estimating county population; *second,* errors in county population estimates increased directly as the population density of the counties in the previous census decade increased; *third,* the 1960 population estimates of metropolitan counties by the Vital Rates Method and Component Method II were more accurate than the 1950 county population estimates by the same two methods; *fourth,* generally, errors in county population estimates became progressively larger as the ratios of the 1960 to 1950 births and deaths increased; and *fifth,* there were significant differences in the accuracy of the county population estimates among the four states studied.

---

[4] William R. Gurley, David White, and James D. Tarver, "The Accuracy of Selected Methods of Preparing Postcensal County Population Estimates," Journal No. 86, *Estadística* (in press).

Chapter VI

COUNTY POPULATION ESTIMATES DEVELOPED FOR THE SIX-STATE TESTS OF ACCURACY

This chapter describes the development of the April 1, 1960 county population estimates required for the six-state tests of accuracy. The first part of the chapter reviews the existing censuses of population and existing population estimates of counties in this area. The availability of these censuses and estimates had to be established before one could determine precisely the specific county estimates which had to be prepared. The second part of this chapter describes the procedures used to generate the necessary April 1, 1960 county population estimates needed for the six-state tests of accuracy.

Enumerations and Existing Estimates, 1950-1962

This section provides a brief summary of the nature and scope of county population enumerations and estimates for the six-state area as of the time this project was initiated. This brief review, which covers the 1950-1962 period, is divided into two parts. The first part deals with Federal and state census enumerations of county population between April 1, 1950 and July 1, 1962. The second part summarizes various sets of county population estimates existing at the beginning of the project.

Census Enumerations. The only censuses of population available for counties in the five States of Arkansas, Iowa, Missouri, Nebraska, and Oklahoma between 1950 and 1962 were the two Federal decennial censuses for April 1, 1950 and 1960. In the sixth state, Kansas, these two Federal censuses were supplemented by annual county population enumerations conducted in accordance with the General Statutes of the State of Kansas.

Two Federal decennial censuses of population, which provide complete counts of the residents of counties, were taken during the 1950-1962 period, one for April 1, 1950 and one April 1, 1960. Between decennial censuses, the Bureau of the Census conducts special censuses of local areas only at the request and expense of local governments. Between April 1, 1950 and April 1, 1960, the Bureau of the Census conducted special censuses for 48 Arkansas towns and cities and for two "places" in Iowa.[1] Between April 1, 1960 and July 1, 1962, special censuses were conducted for 14 Arkansas towns and cities.[2] However, the Bureau of the Census conducted no special censuses for any of the 564 counties in the six-state area between April 1, 1950 and July 1, 1962.

The only other population censuses for counties in the six-state area were the annual enumerations taken in the State of Kansas. The General Statutes of that State require that each deputy assessor make an annual enumeration of the inhabitants in his assessing district. The annual enumerations were taken as of March 1 during the period 1950 to 1959 and as of January 1, beginning with the year 1960. These annual county population enumerations were published in mimeographed form.[3]

The Kansas population counts are relatively complete, but the county enumerations are not strictly comparable with those taken in the Federal decennial censuses. The State Statutes require that the county of residence of college students, inmates of state institutions and hospitals, and servicemen quartered on federal military reservations be determined in a different manner.[4] In contrast with the Federal decennial population censuses of 1950 and 1960, the Kansas Census allocates college students to the county of their residence, not to the county in which they attend college; inmates of institutions are classified by the county of their residence, not by the county in which they are institutionalized; and servicemen living on federal military reservations in the state are not counted as residents of the counties in which they are based.

---

[1] *United States Census of Population: 1960 United States Summary, Number of Inhabitants* (Washington, D. C.: U. S. Department of Commerce, Bureau of the Census, 1961), Table 40.

[2] U. S. Bureau of the Census, *Current Population Reports, Special Censuses*, Series P-28, Numbers 1276, 1285, and 1316, Published in March, 1961, March, 1962, and February, 1963, respectively.

[3] The mimeographed releases are published by the Kansas State Board of Agriculture, Topeka, Kansas.

[4] *Kansas Population Schedule for 1962 for Deputy Assessors: Instructions to Deputy Assessors*, Kansas State Board of Agriculture, Topeka, Kansas.

Existing Estimates. The Bureau of the Census makes annual July 1 population estimates for each state.[5] However, the Census Bureau does not make annual county population estimates (except at the expense of requesting agencies), due to insufficient manpower and resources.[6] Various sets of population estimates have been prepared by state agencies for counties in all six states except Kansas, which conducts an annual census. These estimates apply to various dates during the year and have not been prepared on a uniform basis. Thus, at the time this project was initiated, annual July 1 population estimates for counties in this six-state area, based upon uniform estimation procedures, did not exist for the 1950-1962 period.[7]

The Nebraska county population estimates for December 31 were based upon a ratio-type method, using a trend series involving five ratios for each county, weighted as follows: drivers' licenses 3, head tax 3, school census 2, total vote 1, and vital statistics 1.[8]

Annual April 1 population estimates for Oklahoma counties have been published regularly by the University of Oklahoma Bureau of Business Research since 1951.[9] These estimates were based upon a modified Bureau of the Census Component Method I procedure. Members of the Oklahoma Committee for Population Estimates, an informal group drawn from business firms and public agencies, have prepared

---

[5] U. S. Bureau of the Census, *Current Population Reports, Population Estimates,* "Preliminary Intercensal Estimates of the Population of States, July 1, 1950 to 1959," Series P-25, No. 229, May, 1961; and "Estimates of the Population of States: July 1, 1963, with Preliminary Estimates for July 1, 1964," Series P-25, No. 289, August, 1964.

[6] January 1, 1956 population estimates were made by the Bureau of the Census for counties in the St. Louis Standard Metropolitan Area, using the "dwelling unit" censal ratio method: see U. S. Bureau of the Census, *Current Population Reports, Population Estimates,* "Estimates of the Population of the Standard Metropolitan Areas of Houston, Milwaukee, St. Louis, and Washington, D. C.: January 1, 1956," Series P-25, No. 137, May, 1956.

[7] Four commercial firms published annual county and/or city population estimates, but their precise estimating techniques are unknown, except that the methods vary from area to area. Sales Management, Inc., publishes January 1 population estimates in its annual *Survey of Buying Power;* Standard Rate and Data Service annually publishes January 1 and July 1 county population estimates; Editor and Publisher Company, Inc., annually publishes January 1 county population estimates in its *Market Guide;* and Rand McNally and Company publishes January 1 population estimates of cities classified as principal business centers in its annual *Commercial Atlas and Marketing Guide.* Apparently, these four firms have published no tests of the accuracy of their estimation methods.

[8] "Nebraska County Population Estimates, 1961," *University of Nebraska News, Business in Nebraska* (Lincoln, Nebraska: Bureau of Business Research, College of Business Administration, March, 1962), Number 210. The year-end 1962 and 1963 county population estimates were published in "Population Estimates For 1963," *University of Nebraska News, Business in Nebraska,* Number 234, March, 1964. The four county population estimates for 1952, 1954, 1956, and 1958 were for July 1 each year (See, for example, "Nebraska County Population, 1954," *University of Nebraska News, Business In Nebraska,* Number 131, August, 1955.

[9] "County Population Estimates," *Oklahoma Business Bulletin,* Volume 26, Number 8 (Norman, Oklahoma: The University of Oklahoma Bureau of Business Research, College of Business Administration), August, 1959; and four dittoed unnumbered releases of the population of Oklahoma counties on April 1, 1961, 1962, 1963, and 1964. Annual July 1, 1940 to 1947 county population estimates prepared by component Method I were published in the following report: Francis R. Cella, *Population Shifts of Oklahoma Counties,* 1940-47, Studies in Business and Economics, Number Two, Bureau of Business Research, College of Business Administration (Norman: The University of Oklahoma), October, 1948. The 1945 and 1955 Oklahoma County population estimates, the 1956 and 1957 county estimates, and the 1962 county estimates, respectively, were also published in the *Statistical Abstract of Oklahoma,* 1956, 1957, and 1962, Bureau of Business Research, College of Business Adminisration, University of Oklahoma, Norman.

various series of county estimates by several methods on an experimental basis.[10]

Annual population estimates for counties in Missouri for July 1 and December 1 have been prepared by the Missouri Division of Health, beginning with December 31, 1960.[11] These estimates were made by a component method, using natural increase and net migration based upon the 1950-1960 decade.[12] One set of Missouri county population estimates were prepared for July 1, 1959, using the Bogue-Duncan Composite Method.[13] Apparently, this was the only year for which population estimates were developed by this method for Missouri counties.

Annual July 1 population estimates for Iowa counties have been prepared by the Division of Vital Statistics, Iowa State Department of Health. The county and city population estimates were made by a modified Method II procedure.[14]

Annual July 1 population estimates for counties in Arkansas have been prepared by the Bureau of Business and Economic Research, University of Arkansas. The county population estimates were made by a modified Method II procedure.[15]

Development of Needed County Population Estimates

Since no uniformly prepared April 1, 1960 population estimates for all 564 counties in the six-state area existed at the time this research was initiated, it was necessary to prepare the actual 1960 estimates by alternative techniques. Moreover, because published tests of accuracy had not conclusively shown that any one of the conventional methods of estimating postcensal county population was superior to the other methods tested, the planned evaluation test called for the development of four separate sets of April 1, 1960 population estimates for counties in the six-state area, employing the following four most highly recommended

---

[10] Two publications by this committee give July 1, 1959 county population estimates for Oklahoma prepared by an average of Component Method II and the Vital Rates Method: Oklahoma State Committee For Population Estimates, "Estimates as of July 1, 1959," June 22, 1960 (mimeographed); and James D. Tarver, *County Population Trends in Oklahoma, 1950-1959*, Oklahoma State University Agricultural Experiment Station Processed Series P-351, May, 1960.

[11] "Population Estimates," *Health Facts*, Vol. II, No. 11-12 (Jefferson City, Missouri: Missouri Division of Health, Statistical Services), November-December, 1962; other estimates are published in the same series, Vol. III, No. 4, April, 1963; Volume III, No. 8, August, 1963; Volume IV, No. 2, February, 1964; and Volume IV, No. 3, March, 1964.

[12] "Estimates of the Components of Population Change, Missouri: 1950-1960," *Health Facts*, Vol. II, No. 7 (Jefferson City, Missouri: Missouri Division of Health, Statistical Services), July, 1962.

[13] Thomas C. Dundon, "A Method For Estimating the Population of Missouri Counties," *Business and Economic Review*, Vol. 2 (Columbia: University of Missouri, Business and Public Administration Research Center), January-February, 1961, pp. 10-15.

[14] "Iowa Civilian Population Estimates," dittoed release by the Division of Vital Statistics, Iowa State Department of Health, Des Moines, for July 1, 1961, 1962, and 1963.

[15] "Arkansas 1961 Population Estimates," "Revised Arkansas 1962 Population Estimates," and "Arkansas 1963 Population Estimates," dittoed releases by the Bureau of Business and Economic Research, University of Arkansas, Fayetteville.

methods: Component Method II, the Vital Rates Method, the Bogue-Duncan Composite Method, and the Census Variation of the Composite Method. Then the objective was to employ the basic statistical model developed in Chapter IV and proceed to determine which of the four estimation methods gave the most accurate set of April 1, 1960 county population estimates. The results of the evaluation tests are presented in the next chapter.

In the spring of 1962 a complete list of data inputs required to make April 1, 1960 county population estimates by each of the four selected estimation methods was prepared. Next, plans for the collection of the basic input data needed in making the county population estimates in the six-state area were carefully drafted, and the actual compilation of these data proceeded at the most rapid pace possible. The requisite data on civilian and total population, by age-sex characteristics, which were published in the April 1, 1950 and April 1, 1960 Federal decennial censuses of population for each county were tabulated, coded, and punched on IBM cards. The Population Estimates and Projections Branch, Population Division, Bureau of the Census, was requested to provide the various state "controls", "components", and other required computational figures for making the April 1, 1960 population estimates by each of four estimation methods.

The other basic county input data were assembled from the following three sources:

First, the armed forces strength data for each county on April 1, 1950 and April 1, 1960 were obtained from the two appropriate Federal decennial censuses of population.

Second, the necessary resident live births and resident deaths, by age, for counties were obtained from the state departments of vital statistics in all states in which the data were available. Special tabulations had to be made to obtain the resident county deaths by age.

Third, the required county school enrollments in public schools in grades 2-8 were obtained from the departments of public instruction in each state. Special tabulations were required to obtain comparable public school enrollments in the six states. Parochial school enrollments were obtained either from the state departments of public instruction or from local, state, and regional offices of the following three religious groups: Catholics, Lutherans, and Seventh-Day Adventists. Also, elementary school enrollments were obtained for all Federal Indian schools located in the six-state area.

Since the compilation of the requisite comparable data for counties in each state became a formidable task, a conference on county population estimates in the six-state area was held at the Midwest Research Institute, Kansas City, Missouri, on July 19 and 20, 1962 to expedite the collection of the required input data needed to make the April 1, 1960 county population estimates by each of the four methods. The objectives of the conference were as follows: to facilitate the collection of county public and private school enrollment data on a uniform basis, to discuss county population estimation and evaluation plans outlined for this project, to review the existing county population estimates being prepared in the six-state area, to discuss the annual Kansas county census enumerations, and to obtain the co-operation of the private and public agencies at the local, state, and national levels, in supplying the data necessary for preparing the required county population estimates. Representatives of the Midwest Research Institute, The Bureau of the Census, Office of Education (HEW), religious bodies, state departments of public instruction and vital statistics, the Kansas State Board of Agriculture, Kansas City Planning Commission, and staff members of state universities in the six states participated in the conference.[16]

Then, computer programs were written to calculate April 1, 1960 county population estimates by each of the four specified estimation methods. The two programs which calculate estimates by the two composite methods were written under the supervision of Professor Margaret F. Shackelford, Department of Preventive Medicine and Public Health, Biostatistical Unit, The University of Oklahoma Medical Center.

The procedure established for making April 1, 1960 county population estimates required that the year 1950 be used as the base year. To make April 1, 1960 county population estimates by Component Method II and the Vital Rates Method, it was necessary to have births and deaths reported by county of residence for 1949, 1950, 1959, and 1960. The needed county vital statistics, by place of residence, were not available in Arkansas for the first two years. Thus, special tabulations of the data had to be developed.

To make April 1, 1960 county population estimates by the two composite methods, it was necessary to have the 1949-1950 and 1959-1960 county death rates of the population 45-64 and 65 years of age and over, or, at least, the figures for each of the two years 1950 and 1960. Deaths, by age, were tabulated for counties in four of the six states. However,

[16] The proceedings of the conference are reported in "Conference on County Population Estimates in the Midwest Region, at the Midwest Research Institute, Kansas City, Missouri, July 19 and 20, 1962," Research Foundation, Oklahoma State University, July, 1962 (mimeographed).

it was impossible to obtain the 1949 and 1950 resident county deaths, by age, in Arkansas and Nebraska. Thus, the April 1, 1960 population estimates by the Bogue-Duncan Composite Method and the Census Variation of the Composite Method for counties in Arkansas and Nebraska could not be computed.

After determining that it was possible to prepare April 1, 1960 postcensal population estimates for all 564 counties in the six states by only two of the four methods (the Vital Rates Method and Component Method II), it was necessary to limit the evaluation of accuracy in the six-state area to only these two methods. However, for the State of Oklahoma, April 1, 1960 county population estimates were computed by the Bogue-Duncan Composite Method, in addition to estimates prepared by the other two methods.

Accordingly, April 1, 1960 population estimates were calculated for all 564 counties in the six-state area by the Vital Rates Method and Component Method II, and a third set of April 1, 1960 population estimates were prepared for the counties in Oklahoma by the Bogue-Duncan Composite Method.

Summary

The only April 1 censuses of population available for counties in the six-state area betwen 1950 and 1962 were the two Federal decennial censuses for April 1, 1950 and 1960. In Kansas, these two Federal censuses were supplemented by annual March 1, 1950 to 1959 and annual January 1, 1960 to 1962 county population enumerations conducted in accordance with the General Statutes of the State of Kansas.

Since no uniformly prepared April 1, 1960 population estimates existed for counties in the six states at the time this project was initiated, it was necessary to develop the various sets of 1960 estimates for the tests of accuracy. April 1, 1960 population estimates for counties in Arkansas and Nebraska could not be prepared by the Bogue-Duncan Composite Method and the Census Variation of the Composite Method due to unavailability of 1949 and 1950 resident county deaths by age. Thus, it was necessary to restrict the tests of accuracy in the entire six-state area to only two of the four methods—Component Method II and the Vital Rates Method. For counties in the State of Oklahoma April 1, 1960 population estimates were also prepared by the Bogue-Duncan Composite Method. The accuracy of these specified methods of estimating the April 1, 1960 Federal decennial census enumerations in the entire six-state area and in the State of Oklahoma are reported in the following chapter.

Chapter VII

THE SIX-STATE TESTS OF ACCURACY

The first two major parts of this chapter analyze two sets of county population estimates: First, the errors in the April 1, 1960 population estimates of counties in Arkansas, Iowa, Kansas, Missouri, Nebraska, and Oklahoma prepared by Component Method II and the Vital Rates Method; and second, errors in April 1, 1960 population estimates of counties in Oklahoma prepared by Component Method II, the Vital Rates Method, and the Bogue-Duncan Composite Method. It was impossible to compute April 1, 1960 county population estimates by the two composite methods for all six states, due to the unavailability of the requisite data for counties in Arkansas and Nebraska. The third major section of this chapter applies the findings of the tests of accuracy in the actual development of annual July 1, 1950-1962 county population estimates.

1960 Population Estimates of Counties in the Six States

Tables 9 through 14 show the April 1, 1960 county population estimates in each of the six states by each of the two methods and the actual numeric and logarithmic differences.

The following mathematical model was employed in analyzing the errors in the April 1, 1960 population estimates of counties in the entire six-state area prepared by Component Method II and the Vital Rates Method:

$$U_{inlk} = \mu + \alpha_i + \omega_n + (\alpha\omega)_{in} + \theta_l + (\alpha\theta)_{il} + \beta Y_k + \gamma Z_k + \varepsilon_{inlk}, \quad (1)$$

where α_i is the fixed effect on the accuracy due to estimation method i, with $i = 1$ and 2; ω_n is the fixed effect on the accuracy due to states, with $n = 1, \ldots, 6$; θ_l is the fixed effect on the accuracy due to the metropolitan classification in the previous census decade (April 1, 1950), with 1 (metropolitan counties with central cities), 2 (metropolitan counties without central cities), and 3 (nonmetropolitan counties); $(\alpha\omega)_{in}$ and

$(\alpha\theta)_{il}$ are fixed effects, showing the interactions between α and ω and α and θ, respectively. Should the accuracy of the estimation method depend upon which state or metropolitan classification a county is in, the interaction effect will differ from zero. Y_k, a covariable, is the ratio of the 1960 to the 1950 births in the k^{th} county; Z_k, a covariable, is the ratio of the 1960 to the 1950 deaths in the k^{th} county; β and γ are, respectively, partial regression coefficients associated with the covariables Y_k and Z_k; μ is the overall mean; and U_{inlk} is the logarithm of the absolute difference for the k^{th} county for estimation method i.

The least squares method of solving simultaneous equations was employed to test the following seven hypotheses:

1. H_o: $\alpha_1 = \alpha_2$.
2. H_o: $\omega_1 = \omega_2 = \ldots = \omega_6$.
3. H_o: $\theta_1 = \theta_2 = \theta_3$.
4. H_o: $(\alpha\omega) = 0$.
5. H_o: $(\alpha\theta) = 0$.
6. H_o: $\beta = 0$.
7. H_o: $\gamma = 0$.

The seven hypotheses were used to determine whether each of the five independent variables and two two-factor interactions exert a significant influence upon the accuracy of the county population estimates.

Tables 9-14 give the errors in population estimates separately for each county and Tables 15 and 16 summarize the errors in the April 1, 1960 population estimates of the 564 counties in the six-state area prepared by Component Method II and the Vital Rates Method. The Vital Rates Method was slightly more accurate than Component Method II, with the respective log means being 2.762 and 2.810. The overall log mean for both methods was 2.786. However, the F-test computed in Table 17 revealed that this small difference between the two estimation methods was not significant since the calculated variance ratio was only 1.07.

The two estimation methods were most accurate for counties in Nebraska, followed, in order, by counties in Kansas, Iowa, Missouri, Arkansas, and Oklahoma. The log means of the six states were, in the order specified, as follows: 2.6022, 2.6534, 2.8468, 2.8551, 2.8803, and 2.9180 (Table 15). However, there was considerable variation in the estimation errors for counties in each of the six states, as reflected in the standard deviations of the errors and in the coefficients of variation (Table 15).

The sum of squares of the log differences attributable to the main effect of "states" (after adjusting for all other independent variables) was only 6.2969 (Table 17). Hence, the calculated variance ratio of 4.37 for "states" was not considered significant because the small error mean square of .2881 made the variance ratio test extremely sensitive to very small differences in estimation errors. Moreover, this variance ratio was relatively small compared to the highly significant calculated F ratios of 60.02 and 29.13 (Table 17).

The absolute errors in population estimates increased directly as the population size of counties increased. The magnitude of the estimation errors did not vary directly with population density in the six states, although the two states which had the smallest errors also had the smallest number of people per county and the lowest population densities. Counties in each of the six states are relatively sparsely inhabited, varying from an average of 18 persons per square mile in Nebraska to 63 in Missouri in 1960.

Next, the two-factor interaction of estimation methods by "states" was tested and found not to be significant since the calculated variance ratio was only 1.75 (Table 17). The log means in Table 15 show that the Vital Rates Method gave consistently more accurate population estimates than Component Method II for counties in every state, except in Kansas, even though the differences were not significant.

The 1960 population estimates were significantly more accurate for the 542 nonmetropolitan counties in 1950 than for the 16 metropolitan counties with central cities and for the 6 metropolitan suburban counties without central cities (Table 17 shows that the variance ratio was 29.13, which was significant at the one percent level). However, the t-tests indicated that the errors in the population estimates of the metropolitan counties with central cities did not differ significantly from the errors in the population estimates for the metropolitan suburban counties. Indeed, both the Vital Rates Method and Component Method II gave smaller estimation errors for the nonmetropolitan counties than for the two classes of metropolitan counties (Table 16). Thus, the 1960 population estimates were much more accurate for the sparsely populated than for the densely populated counties.

Table 16 reveals that the Vital Rates Method was slightly more accurate than Component Method II in estimating the 1960 population of nonmetropolitan and suburban metropolitan counties, but that Component Method II was more accurate than the Vital Rates Method in estimating the populations of metropolitan counties with central cities. However, due to the variability of the errors and their small differences, the two-factor interaction of estimation methods and metropolitan classification was not significant at the five percent level (the calculated F was only 1.03—Table 17).

Errors in the April 1, 1960 population estimates increased as the ratio of 1960 to 1950 births increased, with the calculated variance ratio of 60.02 being significant at the one percent level (Table 17). Although estimation errors increased as the ratio of the 1960 to 1950 deaths increased, the partial regression coefficient does not differ significantly from zero; and the calculated F of 1.49 was not significant at the five percent level.

Thus, only two of the seven hypotheses tested in Table 17 for model 1, hypotheses numbers 3 and 6, were rejected in the analysis of errors in the April 1, 1960 population estimates of the 564 counties in the six-state area.

1960 Population Estimates of Counties in Oklahoma

Table 14 gives the April 1, 1960 population estimates for each of the 77 counties in the State of Oklahoma, as well as the actual numeric and logarithmic errors of the estimates prepared by the following three methods: Component Method II, the Vital Rates Method, and the Bogue-Duncan Composite Method. Table 18 summarizes these errors.

Two multiple covariance models were employed to analzye the errors in the April 1, 1960 county population estimates in Oklahoma. The following two mathematical models were employed:

$$U_{ik} = \mu + \alpha_i + \delta X_k + \beta Y_k + \gamma Z_k + \varepsilon_{ik}, \text{ and} \quad (2)$$

$$U_{lk} = \mu + \theta_l + \varepsilon_{lk}, \quad (3)$$

where α_i is the fixed effect on the accuracy due to estimation method i, with i=1, 2, and 3; θ_l is the fixed effect on the accuracy due to the metropolitan classification in the previous census decade (April 1, 1950), with $l = 1$, 2, and 3; X_k, a covariable, is the population density of the k^{th} county in the previous census (April 1, 1950); Y_k, a covariable, is the ratio of the 1960 to 1950 births in the k^{th} county; Z_k, a covariable, is the ratio of the 1960 to 1950 deaths in the k^{th} county; and δ, β, and γ, are, respectively, partial regression coefficients associated with the covariables X_k, Y_k, and Z_k.

The least squares method of solving simultaneous equations was employed to test the following five hypotheses:

1. H_o: $\alpha_1 = \alpha_2 = \alpha_3$.
2. H_o: $\theta_1 = \theta_2 = \theta_3$.
3. H_o: $\delta = 0$.
4. H_o: $\beta = 0$.
5. H_o: $\gamma = 0$.

Table 18 indicates that the Bogue-Duncan Composite Method was slightly the most accurate method for estimating the April 1, 1960 population of Oklahoma counties, followed, in order, by the Vital Rates Method, and Component Method II. However, the variation in the Bogue-Duncan Composite Method population estimates was highest ($s = .636390$), followed, in order, by the Vital Rates Method and Component Method II; and the coefficients of variation followed the same order as the standard deviations (Table 18). Due to variability of the errors for each of the estimation methods in the 77 counties and their small mean differences, the calculated variance ratio was not significant at the five percent level (Table 19).

The accuracy of the April 1, 1960 Oklahoma population estimates was directly associated with the population density of the counties in 1950 for the errors increased as the population density of counties increased (Table 19). The calculated variance ratio of 21.52 was significant at the one percent level.

The mathematical model used in computing the analysis of variance shown in Table 19 did not include a metropolitan classification of counties in 1950, since this variable measured practically the same thing as did population density in 1950. Nevertheless, errors in the three estimation methods were tabulated by metropolitan classification to provide a test for Model 3 (Table 20).

Table 18 indicates that the Vital Rates Method was most accurate for the two counties in Oklahoma which were in metropolitan areas in 1950, followed, in order, by the Bogue-Duncan Composite Method and Component Method II. For the 75 nonmetropolitan counties in 1950, the Composite Method was most accurate, followed, in order, by the Vital Rates Method and Component Method II. Table 20 shows that the errors in the population estimates for nonmetropolitan counties were significantly smaller than for metropolitan counties, with the calculated variance ratio of 25.10 being significant at the one percent level. However, the errors in the estimation methods did not differ significantly within each of the two metropolitan classifications.

The other covariable which was closely related to estimator accuracy was the ratio of the 1960 to 1950 births, with the calculated variance ratio of 33.44 being significant at the one percent level (Table 19). As the number of births in the county increased between 1950 and 1960, the accuracy of estimation decreased. The ratio of 1960 to 1950 deaths had no appreciable influence on estimator accuracy.

Therefore, only two of the hypotheses tested in Table 19 for Model 2, hypotheses numbers 3 and 4, were rejected. In addition, the F-test based upon the results given in Table 20 rejected hypothesis number 2 [$\theta_1 \neq \theta_3$; since there were no θ_2 (suburban metropolitan) counties in Oklahoma in 1950].

Application of Findings to Annual County Population Estimates, 1950-1962

The analytical procedure formulated for testing the April 1, 1960 population estimates of the 564 counties in the six-state area called for an evaluation of the accuracy of four of the most widely used methods of estimating county population described previously in Chapter II. After determining which of the estimation methods was most accurate for estimating the April 1, 1960 population of the 564 counties, the plan was then to employ the most accurate method in making annual July 1, 1950 to 1962 population estimates for each of the 564 counties.

Due to the unavailability of basic input data, it was necessary to restrict the tests of accuracy for all counties in the six-state area to only two methods, the Vital Rates Method and Component Method II (see the section "Development of Needed County Population Estimates" in Chapter VI).

Since the April 1, 1960 tests of accuracy for the 564 counties in the six-state area indicated no significant differences in the accuracy of the Vital Rates Method and Component Method II, either method could have been appropriately chosen for use in preparing July 1 population estimates for all 564 counties in the six-states each year from 1950 to 1962. Method II was selected for use in making the midyear county population estimates, not on the basis of its superiority but rather because it does provide detailed components of population change lacking in the other method.

After the annual county population estimates were computed by Method II, they were proportionately adjusted to sum to the Bureau of the Census' official state July 1 population estimate each year.[1]

As a final check, the annual county population estimates for Oklahoma were submitted to local civic and business leaders through chambers of commerce, for review.

There were some inherent problems of uniformity and comparability in the basic input data used in estimating the 1950-62 annual July 1 population of counties in the six states.

The number of persons in the Armed Forces stationed in each county on each estimate date was obtained from the five branches of the Armed Forces—Departments of the Air Force, Army, Navy, Marine Corps, and Coast Guard. Some annual strength figures were secured from national, regional, and local military commanders. It was impossible to obtain the Armed Forces strength on each estimate date by county of residence. In most counties, this factor probably does not greatly affect current population estimates. However, in a few instances where large military installations, with many servicemen living off the bases, are situated near two or more counties, rather large errors may occur in the annual population estimates due to errors in the military components. In a few instances, certain military strength data could not be obtained from any source.[2]

State vital statistics departments allocate births and deaths to the county of residence. Although errors occur in the allocation of county of residence, they are small except for a few unusual counties. Some state laws require that institutional deaths be allocated to the county of previous residence, while the Federal decennial census enumerates institutional populations as residents of the county in which they are institutionalized. Since institutional deaths are not reported separately, it is impossible to adjust for these small differences.

The data which were the most difficult to secure on a comparable, uniform basis between 1949 and 1963 were annual county school enrollments (public and private) in grades 2-8. The following factors affect the quality of the data, both over time and among states: *First,* it was impossible to obtain school enrollments by county of residence since school districts do not follow county boundaries, and school district reorganization affects county enrollments; *second,* pupil accounting and reporting procedures differ among the states and change over time, thus creating a

[1]U. S. Bureau of the Census, *Current Population Reports, Population Estimates,* "Revised Estimates of the Population of States and Components of Population Change, 1950 to 1960," Series P-25, No. 304, April, 1965; and "Estimates of the Population of States: July 1, 1963, with Preliminary Estimates for July 1, 1964," Series P-25, No. 289, August, 1964.

[2]In every case where strength figures for a particular branch of the five Armed Forces for a particular year were unavailable, they were estimated and tied-in with the reported figures for other years.

lack of uniformity in enrollment data; *third,* enrollments of public and private schools have variable reporting and accounting dates; and *fourth,* it was impossible to obtain accurate resident county enrollments in grades 2-8 due to the existence of ungraded classes and special and Federal Indian schools which attract children from various counties and states.

Each of the basic components which was employed in preparing the annual county population estimates was carefully checked to obtain the greatest possible degree of comparability and uniformity throughout the 1950-1962 period. In view of the limitations which the basic data contain, the annual population estimates for the counties appear reasonable and consistent with the known population shifts which have occurred during this period.

Summary

In this chapter variations were used of the basic statistical model given in Chapter IV to determine whether one of the methods employed in estimating the April 1, 1960 population of counties in the six-state area was superior to the other methods tested. Tests of accuracy were made first for the Vital Rates Method and Component Method II for all 564 counties in the six-state area. Tests of accuracy were then made for the Vital Rates Method, Component Method II, and the Bogue-Duncan Composite Method for the 77 counties in Oklahoma. Every test conducted revealed no significance in the accuracy of the estimation methods tested. On the basis of these findings, Component Method II was then employed to make annual July 1, 1950 to 1962 population estimates for all counties in the six-state area. The annual county population estimates for each of the six states were then proportionately adjusted to add to the official annual July 1 state population estimates prepared by the Bureau of the Census.

Finally, in the third major section of this chapter some of the data limitations and problems of uniformity and comparability which affect the annual 1950-1962 county population estimates were discussed.

TABLE 9
SUMMARY OF ERRORS IN THE POPULATION ESTIMATES OF COUNTIES IN ARKANSAS, APRIL 1, 1960, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

		Component Method II			Vital Rates Method		
County	April 1, 1960 Census Population (1)	April 1, 1960 Population Estimates (2)	Difference (Col. 1 — Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	April 1, 1960 Population Estimates (5)	Difference (Col. 1 — Col. 5) (6)	Common Logarithms of the Absolute Differences (7)
Arkansas	23,355	23,062	293	2.466	21,580	1,775	3.249
Ashley	24,220	21,590	2,630	3.419	22,639	1,581	3.198
Baxter	9,943	8,530	1,413	3.150	9,196	747	2.873
Benton	36,272	35,237	1,035	3.014	38,867	—2,595	3.414
Boone	16,116	14,981	1,135	3.054	17,266	—1,150	3.060
Bradley	14,029	12,958	1,071	3.029	12,213	1,816	3.259
Calhoun	5,991	6,165	— 174	2.240	5,539	452	2.655
Carroll	11,284	10,624	660	2.819	11,552	— 268	2.428
Chicot	18,990	17,850	1,140	3.056	19,792	— 802	2.904
Clark	20,950	19,400	1,550	3.190	21,799	— 849	2.928
Clay	21,258	22,218	— 960	2.982	20,800	458	2.660
Cleburne	9,059	8,923	136	2.133	9,172	— 113	2.053
Cleveland	6,944	7,102	— 158	2.198	6,750	194	2.287
Columbia	26,400	25,266	1,134	3.054	29,251	—2,851	3.454
Conway	15,430	15,974	— 544	2.735	13,207	2,223	3.346
Craighead	47,303	48,638	—1,335	3.125	41,743	5,560	3.745
Crawford	21,318	22,376	—1,058	3.024	20,721	597	2.775
Crittenden	47,564	52,045	—4,481	3.651	44,494	3,070	3.487
Cross	19,551	19,444	107	2.029	18,723	828	2.918
Dallas	10,522	9,729	793	2.899	9,662	860	2.934

TABLE 9 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Desha	20,770	21,483	— 713	2.853	21,262	— 492	2.691
Drew	15,213	14,691	522	2.717	14,526	687	2.836
Faulkner	24,303	23,390	913	2.960	24,049	254	2.404
Franklin	10,213	10,713	— 500	2.698	10,407	— 194	2.287
Fulton	6,657	7,104	— 447	2.650	6,680	— 23	1.361
Garland	46,697	40,838	5,859	3.767	45,707	990	2.995
Grant	8,294	8,945	— 651	2.813	9,508	—1,214	3.084
Greene	25,198	26,803	—1,605	3.205	24,261	937	2.971
Hempstead	19,661	17,280	2,381	3.376	19,636	25	1.397
Hot Spring	21,893	24,714	—2,821	3.450	23,604	—1,711	3.233
Howard	10,878	10,651	227	2.356	10,936	— 58	1.763
Independence	20,048	19,393	655	2.816	19,960	88	1.944
Izard	6,766	6,835	— 69	1.838	7,250	— 484	2.684
Jackson	22,843	23,176	— 333	2.522	20,363	2,480	3.394
Jefferson	81,373	77,733	3,640	3.561	80,268	1,105	3.043
Johnson	12,421	12,586	— 165	2.217	14,074	—1,653	3.218
Lafayette	11,030	10,808	222	2.346	13,070	—2,040	3.309
Lawrence	17,267	23,070	—5,803	3.763	16,245	1,022	3.009
Lee	21,001	22,398	—1,397	3.145	19,609	1,392	3.143
Lincoln	14,447	13,731	716	2.854	16,752	—2,305	3.362
Little River	9,211	9,508	— 297	2.472	9,952	— 741	2.869
Logan	15,957	16,172	— 215	2.332	17,790	—1,833	3.263
Lonoke	24,551	23,841	710	2.851	23,811	740	2.869
Madison	9,068	9,614	— 546	2.737	8,593	475	2.676
Marion	6,041	5,425	616	2.789	6,026	15	1.176

TABLE 9 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Miller	31,686	31,385	301	2.478	30,222	1,464	3.165
Mississippi	70,174	75,841	—5,667	3.753	73,252	—3,078	3.488
Monroe	17,327	18,133	— 806	2.906	18,185	— 858	2.933
Montgomery	5,370	6,030	— 660	2.819	5,174	196	2.292
Nevada	10,700	10,417	283	2.451	10,614	86	1.934
Newton	5,963	6,889	— 926	2.966	7,882	—1,919	3.283
Ouachita	31,641	31,097	544	2.735	29,975	1,666	3.221
Perry	4,927	4,917	10	1.000	6,029	—1,102	3.042
Phillips	43,997	47,238	—3,241	3.510	40,526	3,471	3.540
Pike	7,864	7,715	149	2.173	10,208	—2,344	3.369
Poinsett	30,834	31,844	—1,010	3.004	29,854	980	2.991
Polk	11,981	11,049	932	2.969	11,361	620	2.792
Pope	21,177	19,822	1,355	3.131	23,662	—2,485	3.395
Prairie	10,515	11,417	— 902	2.955	11,713	—1,198	3.078
Pulaski	242,980	232,636	10,344	4.014	249,991	—7,011	3.845
Randolph	12,520	12,775	— 255	2.406	13,048	— 528	2.722
St. Francis	33,303	43,149	—9,846	3.993	33,043	260	2.414
Saline	28,956	16,219	12,737	4.105	30,995	—2,039	3.309
Scott	7,297	9,699	—2,402	3.380	7,267	30	1.477
Searcy	8,124	11,930	—3,806	3.580	9,027	— 903	2.955
Sebastian	66,685	55,283	11,402	4.056	64,995	1,690	3.227
Sevier	10,156	9,540	616	2.789	9,663	493	2.692
Sharp	6,319	10,696	—4,377	3.641	8,843	—2,524	3.402
Stone	6,294	6,369	— 75	1.875	6,231	63	1.799
Union	49,518	48,936	582	2.764	51,121	—1,603	3.204

TABLE 9 (CONTINUED)

County		Component Method II			Vital Rates Method		
	April 1, 1960 Census Population (1)	April 1, 1960 Population Estimates (2)	Difference (Col. 1 – Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	April 1, 1960 Population Estimates (5)	Difference (Col. 1 – Col. 5) (6)	Common Logarithms of the Absolute Differences (7)
Van Buren	7,228	7,086	142	2.152	8,104	— 876	2.942
Washington	55,797	50,094	5,703	3.756	55,615	182	2.260
White	32,745	34,021	—1,276	3.105	33,863	—1,118	3.048
Woodruff	13,954	15,788	—1,834	3.263	14,504	— 550	2.740
Yell	11,940	12,055	— 115	2.060	11,390	550	2.740
STATE	1,786,272	1,773,084	13,188	218.144	1,795,662	—9,390	213.907

TABLE 10
SUMMARY OF ERRORS IN THE POPULATION ESTIMATES OF COUNTIES IN IOWA, APRIL 1, 1960, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

County		Component Method II			Vital Rates Method		
	April 1, 1960 Census Population (1)	April 1, 1960 Population Estimates (2)	Difference (Col. 1 – Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	April 1, 1960 Population Estimates (5)	Difference (Col. 1 – Col. 5) (6)	Common Logarithms of the Absolute Differences (7)
Adair	10,893	8,788	2,105	3.323	11,206	— 313	2.495
Adams	7,468	6,614	854	2.931	8,205	— 737	2.867
Allamakee	15,982	17,636	—1,654	3.218	16,083	— 101	2.004
Appanoose	16,015	16,351	— 336	2.526	18,483	—2,468	3.392
Audubon	10,919	10,889	30	1.477	9,731	1,188	3.074
Benton	23,422	25,165	—1,743	3.241	23,050	372	2.570
Black Hawk	122,482	126,957	—4,475	3.650	118,501	3,981	3.599
Boone	28,037	30,619	—2,582	3.411	26,468	1,569	3.195
Bremer	21,108	25,586	—4,478	3.651	21,600	— 492	2.691
Buchanan	22,293	25,228	—2,935	3.467	23,477	—1,184	3.073
Buena Vista	21,189	21,205	— 16	1.204	20,728	461	2.663
Butler	17,467	18,568	—1,101	3.041	16,535	932	2.969
Calhoun	15,923	17,711	—1,788	3.252	15,907	16	1.204
Carroll	23,431	23,033	398	2.599	24,115	— 684	2.835
Cass	17,919	19,568	—1,649	3.217	19,601	—1,682	3.225
Cedar	17,791	21,103	—3,312	3.520	15,988	1,803	3.255
Cerro Gordo	49,894	49,879	15	1.176	48,387	1,507	3.178
Cherokee	18,598	19,726	—1,128	3.052	18,596	2	.301
Chickasaw	15,034	15,407	— 373	2.571	14,293	741	2.869
Clarke	8,222	8,712	— 490	2.690	8,311	— 89	1.949

TABLE 10 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Clay	18,504	19,685	—1,181	3.072	16,338	2,166	3.335
Clayton	21,962	21,580	382	2.582	21,894	68	1.832
Clinton	55,060	58,925	—3,865	3.587	52,819	2,241	3.350
Crawford	18,569	19,752	—1,183	3.072	19,884	—1,315	3.118
Dallas	24,123	24,733	— 610	2.785	23,876	247	2.392
Davis	9,199	9,869	— 670	2.826	10,083	— 884	2.946
Decatur	10,539	10,004	535	2.728	9,819	720	2.857
Delaware	18,483	17,497	986	2.993	17,351	1,132	3.053
Des Moines	44,605	44,401	204	2.309	42,508	2,097	3.321
Dickinson	12,574	12,546	28	1.447	12,595	— 21	1.322
Dubuque	80,048	76,811	3,237	3.510	80,037	11	1.041
Emmet	14,871	14,358	513	2.710	13,723	1,148	3.059
Fayette	28,581	29,826	—1,245	3.095	29,197	— 616	2.789
Floyd	21,102	20,972	130	2.113	21,329	— 227	2.356
Franklin	15,472	16,020	— 548	2.738	15,507	— 35	1.544
Fremont	10,282	10,767	— 485	2.685	9,973	309	2.489
Greene	14,379	15,822	—1,443	3.159	13,760	619	2.791
Grundy	14,132	14,530	— 398	2.599	14,626	— 494	2.693
Guthrie	13,607	16,935	—3,328	3.522	13,702	— 95	1.977
Hamilton	20,032	21,097	—1,065	3.027	20,531	— 499	2.698
Hancock	14,604	16,153	—1,549	3.190	13,670	934	2.970
Hardin	22,533	24,674	—2,141	3.330	21,502	1,031	3.013
Harrison	17,600	18,676	—1,076	3.031	19,443	—1,843	3.265
Henry	18,187	19,489	—1,302	3.114	17,829	358	2.553
Howard	12,734	13,735	—1,001	3.000	13,292	— 558	2.746

TABLE 10 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Humboldt	13,156	12,681	475	2.676	14,602	—1,446	3.160
Ida	10,269	10,050	219	2.340	9,520	749	2.874
Iowa	16,396	19,728	—3,332	3.522	15,699	697	2.843
Jackson	20,754	22,067	—1,313	3.118	21,553	— 799	2.902
Jasper	35,282	34,420	862	2.935	35,779	— 497	2.696
Jefferson	15,818	14,246	1,572	3.196	15,852	— 34	1.531
Johnson	53,663	45,577	8,086	3.907	53,010	653	2.814
Jones	20,693	22,235	—1,542	3.188	19,466	1,227	3.088
Keokuk	15,492	16,984	—1,492	3.173	14,925	567	2.753
Kossuth	25,314	26,445	—1,131	3.053	24,727	587	2.768
Lee	44,207	45,014	— 807	2.906	41,534	2,673	3.426
Linn	136,899	135,042	1,857	3.273	131,899	5,000	3.698
Louisa	10,290	11,218	— 928	2.967	9,519	771	2.887
Lucas	10,923	10,931	— 8	.903	11,634	— 711	2.851
Lyon	14,468	15,008	— 540	2.732	15,755	—1,287	3.109
Madison	12,295	12,843	— 548	2.738	13,050	— 755	2.877
Mahaska	23,602	21,501	2,101	3.322	23,706	— 104	2.017
Marion	25,886	26,944	—1,058	3.024	26,743	— 857	2.932
Marshall	37,984	38,143	— 159	2.201	38,996	—1,012	3.005
Mills	13,050	14,948	—1,898	3.278	13,586	— 536	2.729
Mitchell	14,043	14,390	— 347	2.540	14,123	— 80	1.903
Monona	13,916	14,224	— 308	2.488	15,066	—1,150	3.060
Monroe	10,463	10,189	274	2.437	11,313	— 850	2.929
Montgomery	14,467	15,261	— 794	2.899	14,813	— 346	2.539
Muscatine	33,840	32,590	1,250	3.096	36,646	—2,806	3.448

TABLE 10 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
O'Brien	18,840	19,608	— 768	2.885	17,321	1,519	3.181
Osceola	10,064	10,083	— 19	1.278	10,958	— 894	2.951
Page	21,023	20,806	217	2.336	21,807	— 784	2.894
Palo Alto	14,736	16,420	—1,684	3.226	14,612	124	2.093
Plymouth	23,906	25,399	—1,493	3.174	23,585	321	2.506
Pocahontas	14,234	14,955	— 721	2.857	14,621	— 387	2.587
Polk	266,315	271,274	—4,959	3.695	259,540	6,775	3.830
Pottawattamie	83,102	87,057	—3,955	3.597	80,817	2,285	3.358
Poweshiek	19,300	20,661	—1,361	3.133	19,651	— 351	2.545
Ringgold	7,910	9,792	—1,882	3.274	8,725	— 815	2.911
Sac	17,007	19,935	—2,928	3.466	17,261	— 254	2.404
Scott	119,067	124,256	—5,189	3.715	115,151	3,916	3.592
Shelby	15,825	14,176	1,649	3.217	18,553	—2,728	3.435
Sioux	26,375	26,387	— 12	1.079	27,314	— 939	2.972
Story	49,327	42,737	6,590	3.818	48,768	559	2.747
Tama	21,413	24,557	—3,144	3.497	19,081	2,332	3.367
Taylor	10,288	11,709	—1,421	3.152	10,164	124	2.093
Union	13,712	15,006	—1,294	3.111	14,147	— 435	2.638
Van Buren	9,778	11,708	—1,930	3.285	10,408	— 630	2.799
Wapello	46,126	48,270	—2,144	3.331	44,734	1,392	3.143
Warren	20,829	24,428	—3,599	3.556	21,619	— 790	2.897
Washington	19,406	20,324	— 918	2.962	19,936	— 530	2.724
Wayne	9,800	10,234	— 434	2.637	10,126	— 326	2.513
Webster	47,810	48,427	— 617	2.790	49,257	—1,447	3.160
Winnebago	13,099	14,129	—1,030	3.012	12,279	820	2.913

TABLE 10 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Winneshiek	21,651	20,541	1,110	3.045	21,427	224	2.350
Woodbury	107,849	108,411	— 562	2.749	105,574	2,275	3.356
Worth	10,259	11,853	—1,594	3.202	9,238	1,021	3.009
Wright	19,447	21,428	—1,981	3.296	17,779	1,668	3.222
STATE	2,757,537	2,834,852	—77,315	289.728	2,732,522	25,015	273.947

TABLE 11
SUMMARY OF ERRORS IN THE POPULATION ESTIMATES OF COUNTIES IN KANSAS, APRIL 1, 1960, PREPARED BY COMPOENT METHOD II AND THE VITAL RATES METHOD

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 — Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 — Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Allen	16,369	15,545	824	2.915	16,549	— 180	2.255
Anderson	9,035	8,649	386	2.586	8,353	682	2.833
Atchison	20,898	22,470	—1,572	3.196	21,682	— 784	2.894
Barber	8,713	8,557	156	2.193	7,538	1,175	3.070
Barton	32,368	31,914	454	2.657	32,432	— 64	1.806
Bourbon	16,090	15,579	511	2.708	15,632	458	2.660
Brown	13,229	12,442	787	2.895	13,383	— 154	2.187
Butler	38,395	38,664	— 269	2.429	39,305	— 910	2.959
Chase	3,921	3,649	272	2.434	3,701	220	2.342
Chautauqua	5,956	5,815	141	2.149	6,050	— 94	1.973
Cherokee	22,279	22,101	178	2.250	24,102	—1,823	3.260
Cheyenne	4,708	4,175	533	2.726	5,062	— 354	2.549
Clark	3,396	3,399	— 3	.477	3,373	23	1.361
Clay	10,675	9,966	709	2.850	9,683	992	2.996
Cloud	14,407	14,627	— 220	2.342	15,919	—1,512	3.179
Coffey	8,403	8,150	253	2.403	7,347	1,056	3.023
Comanche	3,271	3,082	189	2.276	3,778	— 507	2.705
Cowley	37,861	38,446	— 585	2.767	41,194	—3,333	3.522
Crawford	37,032	33,356	3,676	3.565	37,898	— 866	2.937
Decatur	5,778	5,577	201	2.303	6,164	— 386	2.586

TABLE 11 (CONTINUED)

County		Component Method II			Vital Rates Method		
	April 1, 1960 Census Population (1)	April 1, 1960 Population Estimates (2)	Difference (Col. 1 – Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	April 1, 1960 Population Estimates (5)	Difference (Col. 1 – Col. 5) (6)	Common Logarithms of the Absolute Differences (7)
Dickinson	21,572	20,180	1,392	3.143	21,757	— 185	2.267
Doniphan	9,574	9,560	14	1.146	9,815	— 241	2.382
Douglas	43,720	37,063	6,657	3.823	43,034	686	2.836
Edwards	5,118	4,941	177	2.247	6,134	—1,016	3.006
Elk	5,048	4,432	616	2.789	5,629	— 581	2.764
Ellis	21,270	19,411	1,859	3.269	21,179	91	1.959
Ellsworth	7,677	7,196	481	2.682	8,133	— 456	2.658
Finney	16,093	14,912	1,181	3.072	14,579	1,514	3.180
Ford	20,938	19,823	1,115	3.047	21,071	— 133	2.123
Franklin	19,548	18,473	1,075	3.031	20,686	—1,138	3.056
Geary	28,779	27,841	938	2.972	31,570	—2,791	3.445
Gove	4,107	4,213	— 106	2.025	4,832	— 725	2.860
Graham	5,586	5,759	— 173	2.238	6,883	—1,297	3.112
Grant	5,269	5,110	159	2.201	5,042	227	2.356
Gray	4,380	4,204	176	2.245	3,719	661	2.820
Greeley	2,087	1,870	217	2.336	1,649	438	2.641
Greenwood	11,253	10,780	473	2.674	11,813	— 560	2.748
Hamilton	3,144	2,982	162	2.209	3,574	— 430	2.633
Harper	9,541	9,278	263	2.419	8,663	878	2.943
Harvey	25,865	24,993	872	2.940	24,372	1,493	3.174
Haskell	2,990	2,385	605	2.781	2,344	646	2.810
Hodgeman	3,115	3,048	67	1.826	2,855	260	2.414
Jackson	10,309	10,145	164	2.214	10,362	— 53	1.724
Jefferson	11,252	11,746	— 494	2.693	12,350	—1,098	3.040
Jewell	7,217	6,895	322	2.507	8,444	—1,227	3.088

TABLE 11 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Johnson	143,792	133,829	9,963	3.998	113,005	30,787	4.488
Kearny	3,108	3,065	43	1.633	3,339	— 231	2.363
Kingman	9,958	10,061	— 103	2.012	10,403	— 445	2.648
Kiowa	4,626	4,500	126	2.100	4,714	— 88	1.944
Labette	26,805	25,999	806	2.906	27,029	— 224	2.350
Lane	3,060	2,806	254	2.404	2,650	410	2.612
Leavenworth	48,524	54,536	—6,012	3.779	45,819	2,705	3.432
Lincoln	5,556	5,612	— 56	1.748	5,233	323	2.509
Linn	8,274	8,159	115	2.060	8,114	160	2.204
Logan	4,036	3,784	252	2.401	3,980	56	1.748
Lyon	26,928	22,087	4,841	3.684	27,259	— 331	2.519
McPherson	24,285	23,696	589	2.770	22,436	1,849	3.266
Marion	15,143	14,135	1,008	3.003	15,786	— 643	2.808
Marshall	15,598	15,353	245	2.389	18,403	—2,805	3.447
Meade	5,505	5,169	336	2.526	5,696	— 191	2.281
Miami	19,884	21,234	—1,350	3.130	20,694	— 810	2.908
Mitchell	8,866	9,082	— 216	2.334	8,511	355	2.550
Montgomery	45,007	42,548	2,459	3.390	42,958	2,049	3.311
Morris	7,392	7,330	62	1.792	6,551	841	2.924
Morton	3,354	3,033	321	2.506	2,782	572	2.757
Nemaha	12,897	12,938	— 41	1.612	14,792	—1,895	3.277
Neosho	19,455	18,565	890	2.949	20,182	— 727	2.861
Ness	5,470	5,374	96	1.982	4,446	1,024	3.010
Norton	8,035	7,864	171	2.232	9,808	—1,773	3.248
Osage	12,886	13,474	— 588	2.769	13,712	— 826	2.916

TABLE 11 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Osborne	7,506	7,081	425	2.628	7,184	322	2.507
Ottawa	6,779	6,958	— 179	2.252	6,882	— 103	2.012
Pawnee	10,254	9,991	263	2.419	11,324	—1,070	3.029
Phillips	8,709	8,966	— 257	2.409	7,898	811	2.909
Pottawatomie	11,957	12,137	— 180	2.255	12,219	— 262	2.418
Pratt	12,122	11,653	469	2.671	11,550	572	2.757
Rawlins	5,279	5,296	— 17	1.230	5,171	108	2.033
Reno	59,055	57,593	1,462	3.164	58,816	239	2.378
Republic	9,768	9,525	243	2.385	10,150	— 382	2.582
Rice	13,909	13,735	174	2.240	15,351	—1,442	3.158
Riley	41,914	33,542	8,372	3.922	44,023	—2,109	3.324
Rooks	9,734	9,183	551	2.741	10,535	— 801	2.903
Rush	6,160	6,416	— 256	2.408	6,418	— 258	2.411
Russell	11,348	11,123	225	2.352	10,634	714	2.853
Saline	54,715	52,755	1,960	3.292	61,393	—6,678	3.824
Scott	5,228	4,468	760	2.880	4,948	280	2.447
Sedgwick	343,231	322,025	21,206	4.326	335,598	7,633	3.882
Seward	15,930	13,663	2,267	3.355	14,498	1,432	3.155
Shawnee	141,286	137,574	3,712	3.569	150,715	—9,429	3.974
Sheridan	4,267	4,207	60	1.778	3,588	679	2.831
Sherman	6,682	6,523	159	2.201	7,032	— 350	2.544
Smith	7,776	7,171	605	2.781	7,691	85	1.929
Stafford	7,451	7,847	— 396	2.597	7,557	— 106	2.025
Stanton	2,108	2,123	— 15	1.176	2,477	— 369	2.567
Stevens	4,400	3,910	490	2.690	4,106	294	2.468

TABLE 11 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Sumner	25,316	26,889	—1,573	3.196	26,305	— 989	2.995
Thomas	7,358	6,842	516	2.712	7,469	— 111	2.045
Trego	5,473	5,473	---	.000	6,416	— 943	2.974
Wabaunsee	6,648	7,073	— 425	2.628	6,573	75	1.875
Wallace	2,069	2,208	— 139	2.143	2,225	— 156	2.193
Washington	10,739	10,972	— 233	2.367	11,714	— 975	2.989
Wichita	2,765	3,366	— 601	2.778	2,835	— 70	1.845
Wilson	13,077	15,556	—2,479	3.394	12,858	219	2.340
Woodson	5,423	6,241	— 818	2.912	5,670	— 247	2.392
Wyandotte	185,495	231,776	—46,281	4.665	192,653	—7,158	3.854
STATE	2,178,611	2,149,497	29,114	271.275	2,180,412	—1,801	285.939

TABLE 12
SUMMARY OF ERRORS IN THE POPULATION ESTIMATES OF COUNTIES IN MISSOURI, APRIL 1, 1960, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Adair	20,105	16,372	3,733	3.572	20,929	— 824	2.915
Andrew	11,062	11,592	— 530	2.724	11,383	— 321	2.506
Atchison	9,213	8,629	584	2.766	9,590	— 377	2.576
Audrain	26,079	26,386	— 307	2.487	23,791	2,288	3.359
Barry	18,921	19,929	—1,008	3.003	22,124	—3,203	3.505
Barton	11,113	10,518	595	2.774	10,716	397	2.598
Bates	15,905	16,401	— 496	2.695	17,431	—1,526	3.183
Benton	8,737	9,088	— 351	2.545	7,821	916	2.961
Bollinger	9,167	8,782	385	2.585	9,203	— 36	1.556
Boone	55,202	49,583	5,619	3.749	59,563	—4,361	3.639
Buchanan	90,581	82,996	7,585	3.879	96,790	—6,209	3.793
Butler	34,656	34,723	— 67	1.826	33,481	1,175	3.070
Caldwell	8,830	9,298	— 468	2.670	8,899	— 69	1.838
Callaway	23,858	23,827	31	1.491	23,625	233	2.367
Camden	9,116	8,527	589	2.770	7,396	1,720	3.235
Cape Girardeau	42,020	36,377	5,643	3.751	37,431	4,589	3.661
Carroll	13,847	13,440	407	2.609	14,165	— 318	2.502
Carter	3,973	3,817	156	2.193	4,255	— 282	2.450
Cass	29,702	34,509	—4,807	3.681	30,281	— 579	2.762
Cedar	9,185	9,837	— 652	2.814	9,643	— 458	2.660

TABLE 12 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 − Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 − Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Chariton	12,720	12,130	590	2.770	13,311	— 591	2.771
Christian	12,359	11,671	688	2.837	11,555	804	2.905
Clark	8,725	7,765	960	2.982	8,738	— 13	1.113
Clay	87,474	88,376	— 902	2.955	72,238	15,236	4.182
Clinton	11,588	11,486	102	2.008	11,130	458	2.660
Cole	40,761	34,206	6,555	3.816	41,810	—1,049	3.020
Cooper	15,448	15,698	— 250	2.397	16,213	— 765	2.883
Crawford	12,647	11,861	786	2.895	12,840	— 193	2.285
Dade	7,577	7,354	223	2.348	7,678	— 101	2.004
Dallas	9,314	7,735	1,579	3.198	8,610	704	2.847
Daviess	9,502	9,079	423	2.626	9,264	238	2.376
De Kalb	7,226	6,110	1,116	3.047	6,642	584	2.766
Dent	10,445	8,580	1,865	3.270	10,928	— 483	2.683
Douglas	9,653	7,781	1,872	3.272	9,589	64	1.806
Dunklin	39,139	43,945	—4,806	3.681	40,418	—1,279	3.106
Franklin	44,566	43,538	1,028	3.011	45,535	— 969	2.986
Gasconade	12,195	12,469	— 274	2.437	11,455	740	2.869
Gentry	8,793	8,305	488	2.688	8,844	— 51	1.707
Greene	126,276	115,839	10,437	4.018	120,016	6,260	3.796
Grundy	12,220	10,896	1,324	3.121	11,500	720	2.857
Harrison	11,603	11,181	422	2.625	13,610	—2,007	3.302
Henry	19,226	19,348	— 122	2.086	19,504	— 278	2.444
Hickory	4,516	5,857	—1,341	3.127	4,737	— 221	2.344
Holt	7,885	7,222	663	2.821	8,579	— 694	2.841
Howard	10,859	10,351	508	2.705	10,946	— 87	1.939

TABLE 12 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Howell	22,027	21,277	750	2.875	20,791	1,236	3.092
Iron	8,041	8,812	— 771	2.887	8,539	— 498	2.697
Jackson	622,732	606,698	16,034	4.205	641,069	—18,337	4.263
Jasper	78,863	75,888	2,975	3.473	80,916	—2,053	3.312
Jefferson	66,377	67,992	—1,615	3.208	62,044	4,333	3.636
Johnson	28,981	31,883	—2,902	3.462	28,695	286	2.456
Knox	6,558	6,570	— 12	1.079	8,017	—1,459	3.164
Laclede	18,991	19,789	— 798	2.902	19,497	— 506	2.704
Lafayette	25,274	25,859	— 585	2.767	24,225	1,049	3.020
Lawrence	23,260	21,797	1,463	3.165	23,505	— 245	2.389
Lewis	10,984	11,168	— 184	2.264	10,871	113	2.053
Lincoln	14,783	14,584	199	2.298	16,702	—1,919	3.283
Linn	16,815	15,083	1,732	3.238	17,086	— 271	2.432
Livingston	15,771	14,151	1,620	3.209	15,409	362	2.558
McDonald	11,798	11,188	610	2.785	13,265	—1,467	3.166
Macon	16,473	16,691	— 218	2.338	16,560	— 87	1.939
Madison	9,366	9,874	— 508	2.705	9,452	— 86	1.934
Maries	7,282	6,528	754	2.877	7,829	— 547	2.737
Marion	29,522	28,819	703	2.846	29,490	32	1.505
Mercer	5,750	5,205	545	2.736	6,047	— 297	2.472
Miller	13,800	13,628	172	2.235	15,880	—2,080	3.318
Mississippi	20,695	20,931	— 236	2.372	18,530	2,165	3.335
Moniteau	10,500	8,563	1,937	3.287	10,711	— 211	2.324
Monroe	10,688	10,719	— 31	1.491	11,576	— 888	2.948
Montgomery	11,097	10,945	152	2.181	11,548	— 451	2.654

TABLE 12 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Morgan	9,476	10,558	—1,082	3.034	9,352	124	2.093
New Madrid	31,350	31,807	— 457	2.659	31,133	217	2.336
Newton	30,093	29,084	1,009	3.003	27,054	3,039	3.482
Nodaway	22,215	17,877	4,338	3.637	22,769	— 554	2.743
Oregon	9,845	9,323	522	2.717	10,534	— 689	2.838
Osage	10,867	7,771	3,096	3.490	11,282	— 415	2.618
Ozark	6,744	7,385	— 641	2.806	7,978	—1,234	3.091
Pemiscot	38,095	44,233	—6,138	3.788	42,017	—3,922	3.593
Perry	14,642	14,052	590	2.770	15,818	—1,176	3.070
Pettis	35,120	37,706	—2,586	3.412	35,862	— 742	2.870
Phelps	25,396	22,903	2,493	3.396	26,065	— 669	2.825
Pike	16,706	15,857	849	2.928	17,574	— 868	2.938
Platte	23,350	25,600	—2,250	3.352	23,368	— 18	1.255
Polk	13,753	13,501	252	2.401	12,770	983	2.992
Pulaski	46,567	54,289	—7,722	3.887	40,655	5,912	3.771
Putnam	6,999	6,592	407	2.609	7,768	— 769	2.885
Ralls	8,078	6,017	2,061	3.314	7,519	559	2.747
Randolph	22,014	21,425	589	2.770	22,622	— 608	2.783
Ray	16,075	16,955	— 880	2.944	17,330	—1,255	3.098
Reynolds	5,161	6,207	—1,046	3.019	4,793	368	2.565
Ripley	9,096	8,921	175	2.243	9,695	— 599	2.777
St. Charles	52,970	50,527	2,443	3.387	48,956	4,014	3.603
St. Clair	8,421	7,698	723	2.859	8,962	— 541	2.733
St. Francois	36,516	38,536	—2,020	3.305	35,196	1,320	3.120
St. Louis	703,532	740,271	—36,739	4.565	801,890	—98,358	4.992

TABLE 12 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
St. Louis City	750,026	722,413	27,613	4.441	1,210,309	—460,283	5.663
Ste. Genevieve	12,116	11,188	928	2.968	5,659	6,457	3.810
Saline	25,148	25,890	— 742	2.870	24,254	894	2.951
Schuyler	5,052	5,802	— 750	2.875	4,962	90	1.954
Scotland	6,484	6,256	228	2.357	6,550	— 66	1.819
Scott	32,748	28,144	4,604	3.663	31,913	835	2.921
Shannon	7,087	5,605	1,482	3.170	7,988	— 901	2.954
Shelby	9,063	9,567	— 504	2.702	8,904	159	2.201
Stoddard	29,490	30,440	— 950	2.977	27,659	1,831	3.262
Stone	8,176	8,766	— 590	2.770	7,579	597	2.775
Sullivan	8,783	8,135	648	2.811	9,218	— 435	2.638
Taney	10,238	9,323	915	2.961	9,724	514	2.710
Texas	17,758	19,855	—2,097	3.321	18,752	— 994	2.997
Vernon	20,540	18,517	2,023	3.305	20,508	32	1.505
Warren	8,750	8,359	391	2.592	9,112	— 362	2.558
Washington	14,346	15,489	—1,143	3.058	14,453	— 107	2.029
Wayne	8,638	8,353	285	2.454	9,204	— 566	2.752
Webster	13,753	13,167	586	2.767	14,573	— 820	2.913
Worth	3,936	3,885	51	1.707	3,551	385	2.585
Wright	14,183	14,315	— 132	2.120	14,467	— 284	2.453
STATE	4,319,813	4,266,620	53,193	334.394	4,880,762	—560,949	322.287

TABLE 13
SUMMARY OF ERRORS IN THE POPULATION ESTIMATES OF COUNTIES IN NEBRASKA, APRIL 1, 1960, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

County		Component Method II			Vital Rates Method		
	April 1, 1960 Census Population (1)	April 1, 1960 Population Estimates (2)	Difference (Col. 1 – Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	April 1, 1960 Population Estimates (5)	Difference (Col. 1 – Col. 5) (6)	Common Logarithms of the Absolute Differences (7)
Adams	28,944	28,278	666	2.823	27,295	1,649	3.217
Antelope	10,176	10,912	— 736	2.866	11,247	—1,071	3.029
Arthur	680	824	— 144	2.158	898	— 218	2.338
Banner	1,269	1,668	— 399	2.600	1,345	— 76	1.880
Blaine	1,016	1,090	— 74	1.869	1,083	— 67	1.826
Boone	9,134	9,846	— 712	2.852	9,150	— 16	1.204
Box Butte	11,688	11,658	30	1.477	11,223	465	2.667
Boyd	4,513	4,625	— 112	2.049	3,782	731	2.863
Brown	4,436	4,153	283	2.451	4,062	374	2.572
Buffalo	26,236	25,086	1,150	3.060	23,588	2,648	3.422
Burt	10,192	10,846	— 654	2.815	10,289	— 97	1.986
Butler	10,312	10,649	— 337	2.527	11,780	—1,468	3.166
Cass	17,821	18,921	—1,100	3.041	18,494	— 673	2.828
Cedar	13,368	14,864	—1,496	3.174	12,666	702	2.846
Chase	4,317	3,729	588	2.769	3,279	1,038	3.016
Cherry	8,218	8,187	31	1.491	8,256	— 38	1.579
Cheyenne	14,828	16,419	—1,591	3.201	13,658	1,170	3.068
Clay	8,717	10,809	—2,092	3.320	8,237	480	2.681
Colfax	9,595	10,735	—1,140	3.056	11,089	—1,494	3.174
Cuming	12,435	13,250	— 815	2.911	12,064	371	2.569
Custer	16,517	17,525	—1,008	3.003	16,762	— 245	2.389
Dakota	12,168	13,069	— 901	2.954	11,646	522	2.717
Dawes	9,536	8,135	1,401	3.146	8,518	1,018	3.007
Dawson	19,405	19,392	13	1.113	18,521	884	2.946
Deuel	3,125	3,574	— 449	2.652	2,545	580	2.763

TABLE 13 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Dixon	8,106	8,236	— 130	2.113	6,907	1,199	3.078
Dodge	32,471	33,035	— 564	2.751	31,828	643	2.808
Douglas	343,490	359,592	—16,102	4.206	350,050	—6,560	3.816
Dundy	3,570	10,196	—6,626	3.821	3,192	378	2.577
Fillmore	9,425	10,346	— 921	2.964	8,836	589	2.770
Franklin	5,449	5,005	444	2.647	5,900	— 451	2.654
Frontier	4,311	4,275	36	1.556	3,944	367	2.564
Furnas	7,711	7,581	130	2.113	7,600	111	2.045
Gage	26,818	30,236	—3,418	3.533	27,162	— 344	2.536
Garden	3,472	3,241	231	2.363	3,160	312	2.494
Garfield	2,699	2,752	— 53	1.724	2,524	175	2.243
Gosper	2,489	1,911	578	2.761	2,442	47	1.672
Grant	1,009	980	29	1.462	1,365	— 356	2.551
Greeley	4,595	5,868	—1,273	3.104	4,978	— 383	2.583
Hall	35,757	38,724	—2,967	3.472	32,103	3,654	3.562
Hamilton	8,714	8,838	— 124	2.093	7,471	1,243	3.094
Harlan	5,081	4,765	316	2.499	5,412	— 331	2.519
Hayes	1,919	1,803	116	2.064	1,515	404	2.606
Hitchcock	4,829	4,492	337	2.527	5,117	— 288	2.459
Holt	13,722	13,982	— 260	2.414	13,878	— 156	2.193
Hooker	1,130	1,077	53	1.724	1,237	— 107	2.029
Howard	6,541	7,059	— 518	2.714	6,575	— 34	1.531
Jefferson	11,620	11,176	444	2.647	10,803	817	2.912
Johnson	6,281	6,789	— 508	2.705	6,716	— 435	2.638
Kearney	6,580	7,766	—1,186	3.074	5,726	854	2.931

TABLE 13 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)
Keith	7,958	8,743	— 785	2.894	7,229	729	2.862
Keya Paha	1,672	1,821	— 149	2.173	1,919	— 247	2.392
Kimball	7,975	9,724	—1,749	3.242	8,204	— 229	2.359
Knox	13,300	14,325	—1,025	3.010	13,643	— 343	2.535
Lancaster	155,272	153,245	2,027	3.306	168,684	—13,412	4.127
Lincoln	28,491	31,132	—2,641	3.421	27,063	1,428	3.154
Logan	1,108	1,583	— 475	2.676	1,677	— 569	2.755
Loup	1,097	1,144	— 47	1.672	1,141	— 44	1.643
McPherson	735	496	239	2.378	421	314	2.496
Madison	25,145	26,050	— 905	2.956	25,352	— 207	2.315
Merrick	8,363	9,715	—1,352	3.130	8,119	244	2.387
Morrill	7,057	6,991	66	1.819	7,426	— 369	2.567
Nance	5,635	5,445	190	2.278	5,917	— 282	2.450
Nemaha	9,099	8,051	1,048	3.020	7,938	1,161	3.064
Nuckolls	8,217	7,497	720	2.857	7,299	918	2.962
Otoe	16,503	14,609	1,894	3.277	16,073	430	2.633
Pawnee	5,356	4,846	510	2.707	5,334	22	1.342
Perkins	4,189	4,528	— 339	2.530	4,147	42	1.623
Phelps	9,800	10,607	— 807	2.906	9,359	441	2.644
Pierce	8,722	10,163	—1,441	3.158	8,849	— 127	2.103
Platte	23,992	24,840	— 848	2.928	24,211	— 219	2.340
Polk	7,210	7,257	— 47	1.672	6,387	823	2.915
Red Willow	12,940	12,265	675	2.829	12,449	491	2.691
Richardson	13,903	13,874	29	1.462	14,723	— 820	2.913
Rock	2,554	2,768	— 214	2.330	2,589	— 35	1.544

TABLE 13 (CONTINUED)

County		Component Method II			Vital Rates Method		
	April 1, 1960 Census Population (1)	April 1, 1960 Population Estimates (2)	Difference (Col. 1 – Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	April 1, 1960 Population Estimates (5)	Difference (Col. 1 – Col. 5) (6)	Common Logarithms of the Absolute Differences (7)
Saline	12,542	13,213	— 671	2.826	13,365	— 823	2.915
Sarpy	31,281	29,967	1,314	3.118	30,263	1,018	3.007
Saunders	17,270	20,189	—2,919	3.465	18,300	—1,030	3.012
Scotts Bluff	33,809	32,840	969	2.986	34,267	— 458	2.660
Seward	13,581	16,444	—2,863	3.456	13,119	462	2.664
Sheridan	9,049	9,879	— 830	2.919	7,992	1,057	3.024
Sherman	5,382	6,151	— 769	2.885	6,101	— 719	2.856
Sioux	2,575	2,102	473	2.674	3,279	— 704	2.847
Stanton	5,783	5,795	— 12	1.079	5,995	— 212	2.326
Thayer	9,118	9,296	— 178	2.250	9,153	— 35	1.544
Thomas	1,078	1,042	36	1.556	1,187	— 109	2.037
Thurston	7,237	6,940	297	2.472	7,208	29	1.462
Valley	6,590	8,139	—1,549	3.190	6,962	— 372	2.570
Washington	12,103	13,280	—1,177	3.070	12,480	— 377	2.576
Wayne	9,959	9,566	393	2.594	10,249	— 290	2.462
Webster	6,224	6,694	— 470	2.672	6,467	— 243	2.385
Wheeler	1,297	1,244	53	1.724	1,000	297	2.472
York	13,724	15,103	—1,379	3.139	13,306	418	2.621
STATE	1,411,330	1,467,572	—56,242	245.135	1,414,764	—3,434	238.874

TABLE 14

SUMMARY OF ERRORS IN THE POPULATION ESTIMATES OF COUNTIES IN OKLAHOMA, APRIL 1, 1960, PREPARED BY COMPONENT METHOD II, THE VITAL RATES METHOD, AND THE BOGUE-DUNCAN COMPOSITE METHOD

		Component Method II			Vital Rates Method			Bogue-Duncan Composite Method		
County	April 1, 1960 Census Population (1)	April 1, 1960 Population Estimates (2)	Difference (Col. 1 – Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	April 1, 1960 Population Estimates (5)	Difference (Col. 1 – Col. 5) (6)	Common Logarithms of the Absolute Differences (7)	April 1, 1960 Population Estimates (8)	Difference (Col. 1 – Col. 8) (9)	Common Logarithms of the Absolute Differences (10)
Adair	13,112	13,605	— 493	2.692	14,061	— 949	2.977	14,761	—1,649	3.217
Alfalfa	8,445	8,348	97	1.986	7,531	914	2.960	7,685	760	2.880
Atoka	10,352	9,898	454	2.657	9,518	834	2.921	9,563	789	2.897
Beaver	6,965	7,203	— 238	2.376	6,418	547	2.737	6,731	234	2.369
Beckham	17,782	16,314	1,468	3.166	21,246	—3,464	3.539	19,121	—1,339	3.126
Blaine	12,077	11,781	296	2.471	12,929	— 852	2.930	11,804	273	2.436
Bryan	24,252	22,254	1,998	3.300	23,850	402	2.604	22,924	1,328	3.123
Caddo	28,621	29,012	— 391	2.592	28,377	244	2.387	27,599	1,022	3.009
Canadian	24,727	23,300	1,427	3.154	25,486	— 759	2.880	23,636	1,091	3.037
Carter	39,044	40,976	—1,932	3.286	37,754	1,290	3.110	36,981	2,063	3.314
Cherokee	17,762	16,670	1,092	3.038	18,825	—1,063	3.026	17,285	477	2.678
Choctaw	15,637	14,477	1,160	3.064	17,568	—1,931	3.285	15,518	119	2.075
Cimarron	4,496	4,563	— 67	1.826	4,597	— 101	2.004	5,185	— 689	2.838
Cleveland	47,600	45,182	2,418	3.383	50,495	—2,895	3.461	49,727	—2,127	3.327
Coal	5,546	4,597	949	2.977	6,003	— 457	2.659	5,544	2	.301
Comanche	90,803	81,815	8,988	3.953	104,720	—13,917	4.143	106,961	—16,158	4.208
Cotton	8,031	7,644	387	2.587	8,340	— 309	2.489	8,137	— 106	2.025
Craig	16,303	16,090	213	2.328	17,429	—1,126	3.051	17,172	— 869	2.939
Creek	40,495	39,884	611	2.786	41,065	— 570	2.755	40,154	341	2.532
Custer	21,040	19,999	1,041	3.017	23,249	—2,209	3.344	22,041	—1,001	3.000

TABLE 14 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 — Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 — Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)	Bogue-Duncan Composite Method: April 1, 1960 Population Estimates (8)	Bogue-Duncan Composite Method: Difference (Col. 1 — Col. 8) (9)	Bogue-Duncan Composite Method: Common Logarithms of the Absolute Differences (10)
Delaware	13,198	13,636	— 438	2.641	13,304	— 106	2.025	11,751	1,447	3.160
Dewey	6,051	6,244	— 193	2.285	6,886	— 835	2.921	6,694	— 643	2.808
Ellis	5,457	4,757	700	2.845	5,890	— 433	2.636	5,373	84	1.924
Garfield	52,975	48,103	4,872	3.687	52,963	12	1.079	52,139	836	2.922
Garvin	28,290	30,001	—1,711	3.233	28,461	— 171	2.232	29,306	—1,016	3.006
Grady	29,590	28,253	1,337	3.126	30,997	—1,407	3.148	30,034	— 444	2.647
Grant	8,140	7,631	509	2.706	8,263	— 123	2.089	7,261	879	2.943
Greer	8,877	7,797	1,080	3.033	10,924	—2,047	3.311	9,533	— 656	2.816
Harmon	5,852	6,791	— 939	2.972	7,479	—1,627	3.211	6,342	— 490	2.690
Harper	5,956	6,010	— 54	1.732	7,115	—1,159	3.064	6,568	— 612	2.786
Haskell	9,121	9,324	— 203	2.307	15,817	—6,696	3.825	10,163	—1,042	3.017
Hughes	15,144	13,328	1,816	3.259	16,598	—1,454	3.162	14,225	919	2.963
Jackson	29,736	36,816	—7,080	3.850	42,632	—12,896	4.110	42,899	—13,163	4.119
Jefferson	8,192	7,859	333	2.522	9,992	—1,800	3.255	9,488	—1,296	3.112
Johnston	8,517	9,322	— 805	2.905	9,009	— 492	2.691	8,898	— 381	2.580
Kay	51,042	53,175	—2,133	3.328	47,731	3,311	3.519	49,125	1,917	3.282
Kingfisher	10,635	11,506	— 871	2.940	12,652	—2,017	3.304	11,702	—1,067	3.028
Kiowa	14,825	13,644	1,181	3.072	17,596	—2,771	3.442	16,227	—1,402	3.146
Latimer	7,738	7,045	693	2.840	8,458	— 720	2.857	9,475	—1,737	3.239
Le Flore	29,106	30,656	—1,550	3.190	27,011	2,095	3.321	27,214	1,892	3.276

TABLE 14 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Difference (Col. 1 – Col. 2) (3)	Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Difference (Col. 1 – Col. 5) (6)	Common Logarithms of the Absolute Differences (7)	Bogue-Duncan Composite Method: April 1, 1960 Population Estimates (8)	Difference (Col. 1 – Col. 8) (9)	Common Logarithms of the Absolute Differences (10)
Lincoln	18,783	18,317	466	2.668	18,909	— 126	2.100	19,352	— 569	2.755
Logan	18,662	17,340	1,322	3.121	19,644	— 982	2.992	18,956	— 294	2.468
Love	5,862	5,801	61	1.785	6,342	— 480	2.681	5,597	265	2.423
McClain	12,740	13,607	— 867	2.938	13,985	—1,245	3.095	13,272	— 532	2.725
McCurtain	25,851	25,039	812	2.909	25,948	— 97	1.986	26,946	—1,095	3.039
McIntosh	12,371	11,794	577	2.761	11,786	585	2.767	12,378	— 7	.845
Major	7,808	6,280	1,528	3.184	7,878	— 70	1.845	6,930	878	2.943
Marshall	7,263	7,286	— 23	1.361	6,559	704	2.847	6,596	667	2.824
Mayes	20,073	21,704	—1,631	3.212	20,638	— 565	2.752	20,391	— 318	2.502
Murray	10,622	11,124	— 502	2.700	10,807	— 185	2.267	10,069	553	2.742
Muskogee	61,866	64,288	—2,422	3.384	61,716	150	2.176	59,566	2,300	3.361
Noble	10,376	11,071	— 695	2.841	12,025	—1,649	3.217	11,031	— 655	2.816
Nowata	10,848	12,410	—1,562	3.193	11,648	— 800	2.903	11,378	— 530	2.724
Okfuskee	11,706	11,543	163	2.212	11,793	— 87	1.939	11,430	276	2.440
Oklahoma	439,506	434,025	5,481	3.738	435,537	3,969	3.598	426,922	12,584	4.099
Okmulgee	36,945	34,314	2,631	3.420	35,885	1,060	3.025	35,203	1,742	3.241
Osage	32,441	33,270	— 829	2.918	30,029	2,412	3.382	30,795	1,646	3.216
Ottawa	28,301	27,448	853	2.930	25,615	2,686	3.429	25,279	3,022	3.480
Pawnee	10,884	10,704	180	2.255	10,989	— 105	2.021	10,714	170	2.230
Payne	44,231	34,279	9,952	3.997	44,640	— 409	2.611	43,727	504	2.702

TABLE 14 (CONTINUED)

County	April 1, 1960 Census Population (1)	Component Method II: April 1, 1960 Population Estimates (2)	Component Method II: Difference (Col. 1 – Col. 2) (3)	Component Method II: Common Logarithms of the Absolute Differences (4)	Vital Rates Method: April 1, 1960 Population Estimates (5)	Vital Rates Method: Difference (Col. 1 – Col. 5) (6)	Vital Rates Method: Common Logarithms of the Absolute Differences (7)	Bogue-Duncan Composite Method: April 1, 1960 Population Estimates (8)	Bogue-Duncan Composite Method: Difference (Col. 1 – Col. 8) (9)	Bogue-Duncan Composite Method: Common Logarithms of the Absolute Differences (10)
Pittsburg	34,360	31,532	2,828	3.451	33,645	715	2.854	31,972	2,388	3.378
Pontotoc	28,089	26,741	1,348	3.129	28,902	— 813	2.910	30,083	—1,994	3.299
Pottawatomie	41,486	40,553	933	2.969	42,630	—1,144	3.058	40,489	997	2.998
Pushmataha	9,088	9,380	— 292	2.465	9,659	— 571	2.756	9,609	— 521	2.716
Roger Mills	5,090	2,875	2,215	3.345	5,510	— 420	2.623	4,838	252	2.401
Rogers	20,614	22,312	—1,698	3.229	20,584	30	1.477	19,715	899	2.953
Seminole	28,066	25,955	2,111	3.324	31,985	—3,919	3.593	30,852	—2,786	3.444
Sequoyah	18,001	18,498	— 497	2.696	17,590	411	2.613	17,992	9	.954
Stephens	37,990	41,047	—3,057	3.485	34,856	3,134	3.496	34,331	3,659	3.563
Texas	14,162	13,938	224	2.350	14,791	— 629	2.798	15,700	—1,538	3.186
Tillman	14,654	14,126	528	2.722	15,586	— 932	2.969	14,730	— 76	1.880
Tulsa	346,038	374,989	—28,951	4.461	328,932	17,106	4.233	335,360	10,678	4.028
Wagoner	15,673	13,864	1,809	3.257	13,209	2,464	3.391	13,262	2,411	3.382
Washington	42,347	44,419	—2,072	3.316	36,474	5,873	3.768	38,880	3,467	3.539
Washita	18,121	19,195	—1,074	3.031	21,493	—3,372	3.527	20,276	—2,155	3.333
Woods	11,932	9,895	2,037	3.308	12,325	— 393	2.594	11,487	445	2.648
Woodward	13,902	12,523	1,379	3.139	14,102	— 200	2.301	14,420	— 518	2.714
STATE	2,328,284	2,318,996	9,288	226.316	2,363,885	—35,601	223.058	2,323,474	4,810	220.786

TABLE 15
SUMMARY OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN THE SIX-STATE AREA, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

Estimation Method	State						Total, 6 States
	Arkansas	Iowa	Kansas	Missouri	Nebraska	Oklahoma	
Total Absolute Log Differences							
Method II	218.144	289.728	271.275	334.394	245.135	226.316	1584.992
Vital Rates Method	213.907	273.947	285.939	322.287	238.874	223.058	1558.012
Total	432.051	563.675	557.214	656.681	484.009	449.374	3143.004
Means of the Absolute Log Differences							
Method II	2.9086	2.9265	2.5836	2.9078	2.6359	2.9392	2.8103
Vital Rates Method	2.8521	2.7671	2.7232	2.8025	2.5685	2.8969	2.7624
Total	2.8803	2.8468	2.6534	2.8551	2.6022	2.9180	2.7864
Standard Deviations of the Errors							
Method II	.58123	.59835	.69188	.57609	.61438	.53032	.62008
Vital Rates Method	.56496	.58154	.53129	.67184	.52664	.58622	.58927
Total	.57193	.59391	.61933	.62665	.57164	.56017	.60508
Coefficients of Variation of the Errors[1]							
Method II	19.98	20.45	26.78	19.81	23.31	18.04	22.06
Vital Rates Method	19.81	21.02	19.51	23.97	20.50	20.24	21.33
Total	19.86	20.86	23.34	21.95	21.97	19.20	21.72

[1] $V = \frac{s}{\bar{x}}(100)$, where the coefficient of variation gives the standard deviation expressed as a percentage of the mean.

TABLE 16
SUMMARY OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN THE SIX-STATE AREA, CLASSIFIED BY ESTIMATION METHOD AND METROPOLITAN STATUS, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

Estimation Method	Metropolitan Classification: Metropolitan Counties with Central Cities	Metropolitan Counties without Central Cities	Nonmetropolitan Counties	Total, 6 States
	Total Absolute Log Differences			
Method II	60.321	23.330	1501.341	1584.992
Vital Rates Method	63.454	23.302	1471.256	1558.012
Total	123.775	46.632	2972.597	3143.004
	Means of the Absolute Log Differences			
Method II	3.779	3.888	2.770	2.810
Vital Rates Method	3.966	3.884	2.714	2.762
Total	3.868	3.886	2.742	2.786
	Standard Deviations of the Errors			
Method II	.42093	.63289	.58935	.62008
Vital Rates Method	.52008	.67306	.53829	.58927
Total	.51851	.65511	.56482	.60508
	Coefficients of Variation of the Errors[1]			
Method II	11.17	16.28	21.28	22.07
Vital Rates Method	13.02	17.33	19.83	21.33
Total	13.41	16.86	20.60	21.72

[1] $V = \frac{s}{\bar{x}}(100)$, where the coefficient of variation gives the standard deviation expressed as a percentage of the mean.

TABLE 17
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN ARKANSAS, IOWA, KANSAS, MISSOURI, NEBRASKA, AND OKLAHOMA, PREPARED BY COMPONENT METHOD II AND THE VITAL RATES METHOD

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	1128	9,170.1269		
R (μ)	1	8,757.5125		
R (due to Model $\mid \mu$)	17	92.8500		
Estimation Methods R[α (adjusted)]	1	.3072	.3072	1.07
States R[ω (adj.)]	5	6.2969	1.2594	4.37[1]
Estimation Methods X States Interaction R[$\alpha\omega$ (adj.)]	5	2.5210	.5042	1.75
Metropolitan Classification R[θ (adj.)]	2	16.7834	8.3917	29.13**
Estimation Methods X Metropolitan Classification Interaction R[$\alpha\theta$ (adj.)]	2	.5923	.2962	1.03
Ratio of 1960 to 1950 Births R[β (adj.)]	1	17.2911	17.2911	60.02**
Ratio of 1960 to 1950 Deaths R[γ (adj.)]	1	.4306	.4306	1.49
Error	1110	319.7644	.2881	

*One and two asterisks indicate significance at the five and one percent levels, respectively.
[1] Judged nonsignifcant.

TABLE 18
SUMMARY OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN OKLAHOMA, PREPARED BY THE VITAL RATES METHOD, COMPONENT METHOD II, AND THE BOGUE-DUNCAN COMPOSITE METHOD (MODEL 3)

	Metropolitan Classification			
Estimation Method	Metropolitan Counties with Central Cities	Metropolitan Counties without Central Cities	Nonmetropolitan Counties	Total
	Total Absolute Log Differences			
Method II	8.199	---	218.117	226.316
Vital Rates Method	7.831	---	215.227	223.058
Bogue-Duncan Composite Method	8.127	---	212.659	220.786
Total	24.157	---	646.003	670.160
	Means of the Absolute Log Differences			
Method II	4.100	---	2.908	2.939
Vital Rates Method	3.916	---	2.870	2.897
Bogue-Duncan Composite Method	4.064	---	2.835	2.867
Total	4.026	---	2.871	2.901
	Standard Deviations of the Errors			
Method II	.511238	---	.497951	.530324
Vital Rates Method	.449012	---	.572272	.586215
Bogue-Duncan Composite Method	.050200	---	.613354	.636390
Total	.317344	---	.561492	.585779
	Coefficients of Variation of the Errors[1]			
Method II	12.47	---	17.12	18.04
Vital Rates Method	11.47	---	19.94	20.24
Bogue-Duncan Composite Method	1.24	---	21.64	22.20
Total	7.88	---	19.56	20.19

[1] $V = \frac{s}{\bar{x}}(100)$, where the coefficient of variation gives the standard deviation expressed as a percentage of the mean.

TABLE 19
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN OKLAHOMA, PREPARED BY THE VITAL RATES METHOD, COMPONENT METHOD II, AND THE BOGUE-DUNCAN COMPOSITE METHOD (MODEL 2)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	231	2,023.1398		
R (μ)	1	1,944.2182		
R (due to Model | μ)	5	21.1006		
Estimation Methods R[α(adjusted)]	2	.2007	.1003	<1
Population Density in 1950				
R[δ (adj.)]	1	5.5307	5.5307	21.52**
Ratio of 1960 to 1950 Births				
R[β (adj.)]	1	8.5938	8.5938	33.44**
Ratio of 1960 to 1950 Deaths				
R[γ (adj.)]	1	.0950	.0950	<1
Error	225	57.8209	.2570	

*One and two asterisks indicate significance at the five and one percent levels, respectively.

TABLE 20
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN OKLAHOMA, PREPARED BY THE VITAL RATES METHOD, COMPONENT METHOD II, AND THE BOGUE-DUNCAN COMPOSITE METHOD (MODEL 3)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Total	231	2,023.1398		
R (μ)	1	1,944.2182		
Corrected Total	230	78.9216		
Metropolitan Classification	1	7.7969	7.7969	25.10**
Error	229	71.1247	.3106	

*One and two asterisks indicate significance at the five and one percent levels, respectively.

Chapter VIII

EVALUATION OF ACCURACY TESTS

This chapter examines the conventional procedures employed in tests of accuracy and discusses alternative procedures appropriate for determining the accuracy of two or more estimation methods. The first major section of this chapter is a critique of conventional evaluation procedures. The second major section explores alternative evaluation tests appropriate for determining estimation accuracy.

Critique of Conventional Evaluation Procedures

Chapter III demonstrated that the findings of published evaluations of the accuracy of alternative methods of estimating county populations were inconclusive. Apparently, the contradictory nature of the conflicting published results lay, at least in part, in the statistical procedures and assumptions used to determine the efficacy of different estimation methods.

Published evaluation tests established the accuracy of various estimation methods in the following manner: **First,** the numerical differences between the April 1 postcensal county population estimates and the enumerated population from the Federal decennial census for each county were computed (see Columns 3 and 6, Table 21); **second,** the absolute percentage deviations by each estimation method were calculated for each county by dividing the numerical differences by the enumerated county population (Columns 4 and 7, Table 21); **third,** the average (mean) absolute percentage deviations of each estimation method for counties were computed for all specified counties in an area or state; and **fourth,** the variance of the absolute percentage deviations, the number of positive deviations, and the number of percentage deviations exceeding some level (5 or 10 percent) were calculated.

The conventional method of determining the accuracy of various estimation methods is illustrated in Table 21, where a simple case of 5 hypothetical counties is employed. In this illustrative example, the April 1 postcensal county population estimates were prepared by two alternative Methods A and B, using the previous decennial census year as the base year. Then, the numerical differences between the postcensal county population estimates and the enumerated population from the Federal decennial census were obtained, as shown in Table 21.

Table 21 shows that the means of the absolute percentage deviations of the five hypothetical counties were 52.6 percent for Method A and 34.0 percent for Method B. According to the conventional procedure of using the mean absolute deviations in determining the accuracy of two or more estimation methods, Method B would be judged superior to Method A. This would be an erroneous conclusion, since the "true" mean errors were—1.1 for Method A and—14.3 for Method B (Table 21).

In Table 21, Method A was actually the more accurate estimation method for the five hypothetical counties, for it overestimated their total populations by only 40,000 persons, which was an overestimation of only 1.1 percent. On the other hand, Method B overestimated the total population of the five counties by 501,500 persons, which was an overestimation of 14.3 percent. Thus, Method A was a more accurate estimation technique than Method B, and was, therefore, the superior method.

The procedures inherent in the statistical model formulated in Chapter IV and applied in Chapters V and VII for testing the accuracy of selected methods of estimating postcensal county population departed from the conventional procedure followed in published evaluation tests. In the procedure developed in Chapter IV and employed in Chapters V and VII, the common logarithms of the positive difference ($\log_{10}$ of the absolute difference) between the postcensal county population estimates and Federal decennial census enumerations were taken. The positive (absolute) differences were used, for a logarithmic transformation made this variable approximately normally distributed. Finally, the multiple covariance model formulated in Chapter IV, as well as different variations of it, was applied in Chapters V and VII in computing analyses of variance of the logarithms. F- and t-tests were then made in testing each of the stated null hypotheses, particularly the one that there were no differences in the accuracy of the various estimation methods.

Table 21 reveals that the means of the common logarithms of the absolute differences also indicated that Method A was superior to Method

B, since the two respective log means were 4.241 and 4.320. Moreover, the variance of the common logarithms for Method A was smaller than for Method B, being .762 compared to 1.468, respectively. Therefore, the findings based upon common logarithmic differences corroborated the conclusion previously reached which was based upon the actual numerical differences for the five counties (Table 21).

One of the most serious inadequacies of the conventional method of determining accuracy is that identical percentage deviations do not reflect identical numerical errors in county population estimates, since the denominators of counties in the United States vary from a few hundred population in sparsely inhabited rural counties to several million inhabitants in large metropolitan counties. Thus, percentage deviations depend to a greater extent upon the population size of a county rather than upon the actual estimation error. Consequently, the absolute percentage deviations tend to decline proportionately as the population size of counties increases, irrespective of the estimation technique employed.

Perhaps the greatest weakness of the conventional procedure is that it is impossible to obtain the accurate mean percentage deviations for all counties in a state which have unequal populations without returning to the original county data, adding the actual numerators and denominators of all counties, then dividing.[1] In Table 21, the "true" means (both numerical and percentage) give entirely different conclusions than the means of the absolute percentage differences. Whenever this occurs, F- and t-tests based upon absolute percentage deviations of counties give erroneous results.

The above analysis has rather clearly shown (Table 21) that the mean absolute percentage deviations provided inaccurate conclusions about estimation accuracy for an example in which the population of the five hypothetical counties varied greatly. On the other hand, this example indicated that the common logarithms of the absolute differences were valid measures of estimation errors. Accordingly, variance ratio (F-tests) and t-tests based upon logarithmic transformations are efficient parametric tests for determining significant differences.

Alternative Procedures Appropriate for Determining Estimation Accuracy

Three of the possible appropriate measures of estimation errors (common logarithms of the absolute differences, the absolute differences, and

[1] George W. Snedecor, *Statistical Methods*, Fifth Edition (Ames: The Iowa State College Press, 1956), pp. 32-34.

the numerical differences with signs) were employed in Table 22 to determine whether there were significant differences in the accuracy of three methods in estimating the April 1, 1960 population of counties in Oklahoma (see Table 14 for the actual 1960 county population estimates, numeric, and logarithmic differences). The calculated variance ratios in Table 22 gave identical conclusions for each of the three measures; consequently, there were no significant differences in the accuracy of the three estimation methods of estimating the April 1, 1960 population of counties in Oklahoma.

One can employ the median test, the bivariate extension of the Mann-Whitney U test, and various other nonparametric tests in determining the accuracy of the three different methods of estimating the April 1, 1960 population of Oklahoma counties. However, since all nonparametric tests are less powerful than parametric tests, one can be confident that nonparametric tests will never detect significant differences when parametric tests fail to show significance. For this reason, there is no point in proceeding to employ the less efficient nonparametric tests in evaluating the 1960 population estimates of Oklahoma counties, for all of them will give identical conclusions as did the three parametric tests in Table 22. Even so, one of the most powerful nonparametric tests was computed on the same set of Oklahoma county population estimates data for illustrative purposes only. The Kruskal-Wallis one-way analysis of variance technique based upon the ranks of the absolute differences was applied to the three sets of April 1, 1960 Oklahoma county population estimates (Table 23). This nonparametric test, of course, confirms the conclusions previously established by the parametric tests.

Summary

This chapter gave a critique of the conventional procedure followed in published tests of the accuracy of various estimation methods and has shown that this method gave wrong conclusions in the illustrative example. Also, this chapter described various parametric and nonparametric tests which are appropriate for determining the accuracy of alternative estimation methods. This chapter does not imply that one apply a number of parametric and/or nonparametric tests to the same set of county population estimates in determining the accuracy of various estimation methods. On the contrary, only one appropriate test should be selected and applied, and the conclusions should be based entirely upon that one specified test. Nevertheless, the one particular test selected for each analysis will be governed by the specific hypotheses one proposes to test and will, of course, vary from one problem to another.

TABLE 21

ILLUSTRATIVE EXAMPLE WHICH CONTRASTS THE CONVENTIONAL WITH THE FORMULATED PROCEDURE OF DETERMINING THE ACCURACY OF TWO ESTIMATION METHODS, BASED UPON FIVE HYPOTHETICAL COUNTIES

County	Federal Decennial Census Enumeration on April 1 (1)	Method A: Population Estimate for April 1 By Method A (2)	Method A: Difference (Col. 1—Col. 2) (3)	Method A: Absolute Percentage Deviation (Col. 3÷Col. 1) (4)	Method B: Population Estimate for April 1 By Method B (5)	Method B: Difference (Col. 1—Col. 5) (6)	Method B: Absolute Percentage Deviation (Col. 6÷Col. 1) (7)	Common Logarithms of the Absolute Differences: Method A (8)	Common Logarithms of the Absolute Differences: Method B (9)
1	1,000,000	900,000	100,000	10	1,200,000	—200,000	20	5.000	5.301
2	2,000,000	2,100,000	—100,000	5	2,200,000	—200,000	10	5.000	5.301
3	500,000	540,000	— 40,000	8	600,000	—100,000	20	4.602	5.000
4	5,000	3,000	2,000	40	6,000	— 1,000	20	3.301	3.000
5	1,000	3,000	— 2,000	200	2,000	— 1,000	100	3.301	3.000
Total	3,506,000	3,546,000	— 40,000	263	4,008,000	—501,500	170	21.204	21.602
Mean				52.6			34.0		

Correct Percentage Deviations of the Five Counties by Each Method:

$$\frac{-40{,}000}{3{,}506{,}000} = -1.1 \qquad \frac{-501{,}500}{3{,}506{,}000} = -14.3$$

	Method A	Method B
Means of the Common Logarithms of the Absolute Differences:	4.241	4.320
Variances of the Common Logarithms of the Absolute Differences:	.762	1.468

TABLE 22
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN OKLAHOMA, PREPARED BY THE VITAL RATES METHOD, COMPONENT METHOD II, AND THE BOGUE-DUNCAN COMPOSITE METHOD, USING THREE DIFFERENT PARAMETRIC MEASURES OF ESTIMATION ERRORS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Calculated Variance Ratio
Common Logarithms of the Absolute Differences				
Total	231	2,023.1398		
R (μ)	1	1,944.2182		
Corrected Total	230	78.9216		
Estimation Methods R($\alpha\|\mu$)	2	.2008	.1004	<1
Error	228	78.7208	.3453	
Absolute Differences				
Total	231	2,974,089,013		
R (μ)	1	710,363,018		
Corrected Total	230	2,263,725,995		
Estimation Methods R($\alpha\|\mu$)	2	1,064,413	532,207	<1
Error	228	2,262,661,582	9,923,954	
Numerical Differences (with signs)				
Total	231	2,974,089,013		
R (μ)	1	2,001,641		
Corrected Total	230	2,972,087,372		
Estimation Methods R($\alpha\|\mu$)	2	15,879,323	7,939,662	<1
Error	228	2,956,208,049	12,965,825	

TABLE 23
ANALYSIS OF VARIANCE OF ERRORS IN THE APRIL 1, 1960 POPULATION ESTIMATES OF COUNTIES IN OKLAHOMA, PREPARED BY THE VITAL RATES METHOD, COMPONENT METHOD II, AND THE BOGUE-DUNCAN COMPOSITE METHOD, USING THE NONPARAMETRIC KRUSKAL-WALLIS[1] ONE-WAY ANALYSIS OF VARIANCE BY RANKS OF THE ABSOLUTE DIFFERENCES

Estimation Method		
Component Method II	Vital Rates	Bogue-Duncan Composite Method
$\Sigma R_1 = 9{,}203.0$	$\Sigma R_2 = 8{,}863.0$	$\Sigma R_3 = 8{,}730.0$

$$H = \frac{\dfrac{12}{N\,(N+1)} \displaystyle\sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3\,(N+1)}{1 - \dfrac{\Sigma T}{N^3 - N}} = \frac{.346}{.99999853} = .346.\ p < .90,$$

which is not significant at the five percent level.

$N = 231$; all N_j's $= 77$; $T = t^3 - t$, where $t = 2$ tied in every case (4 cases); and H is distributed approximately as chi-square, with $df = k - 1$ $(k = 3)$.

[1] Sidney Siegel, *Nonparametric Statistics For the Behavioral Sciences* (New York: McGraw-Hill Book Company, Inc., 1956), pp. 185-193.

Chapter IX

SUMMARY AND CONCLUSIONS

The purpose of this concluding chapter is to synthesize as concisely as possible the pertinent materials presented in the preceding eight chapters.

Chapter I placed the population estimation work reported in this publication in its proper perspective. The April 1, 1960 six-state county population estimates and tests of accuracy presented in this study constitute only one phase of work of a major three-year research project undertaken early in 1962. It was a cooperative pilot project encompassing the six states of Arkansas, Iowa, Kansas, Missouri, Nebraska, and Oklahoma.

It was pointed out that a basic problem confronting the regional investigator and/or decision maker was inadequate, incomparable, or nonexistent data. Thus, it was concluded that one of the essential next steps in regional analysis is the generation and collection of reasonably uniform, comprehensive data in a systematic framework. The six-state project represents a modest effort along these lines. Our reference to the project as a pilot program reflects our conviction that such a framework and data collection system for regional analysis must eventually be nationwide. From the inception of the project, it has been our hope that the cooperative six-state pilot project will provide support for the emergence of a national program.

Chapter II described in detail the four most highly recommended methods of preparing postcensal county population estimates: Component Method II, the Vital Rates Method, the Bogue-Duncan Composite Method, and the Census Variation of the Composite Method. Less widely used methods were briefly covered.

Chapter III summarized the major findings of four of the most comprehensive published tests of accuracy of alternative methods of estimating county population: Bureau of the Census evaluation tests for selected metropolitan counties in 1950 and 1960; evaluations by the Study Group on Postcensal Population Estimates, Public Health Conference on Records and Statistics, for counties in the four states of Montana, Ohio, Oregon, and Pennsylvania in 1960; the Schmitt-Crosetti evaluations for counties in the State of Washington in 1950; and the Goldberg-Balakrishnan evaluations for counties in the State of Michigan in 1960. These four studies gave conflicting results; for, in eight separate comparisons, six different estimation methods were purported to be the most accurate.

Chapter IV formulated a statistical model for testing the accuracy of four different methods of estimating county population. The model was designed specifically for the purpose of establishing which one of the four most highly recommended methods of estimating the April 1, 1960 population of counties in the six-state region was actually the most precise estimation technique.

The procedure developed in Chapter IV for testing accuracy uses the common logarithms of the absolute differences between the April 1 postcensal county population estimates and April 1 decennial census enumerations and employs parametric tests (F- and t-tests) for significant differences. The statistical procedure developed in this chapter for testing the accuracy of alternative estimation methods is one of the major methodological contributions of this study, as it departs from the conventional procedures used prior to this time. Moreover, it overcomes the inherent deficiencies of the statistical procedures and assumptions employed in conventional tests of accuracy.

Chapter V applied five specific variations of the basic statistical model developed in Chapter IV to test the accuracy of various methods of estimating county population. Three preliminary tests of accuracy conducted in Chapter V were made on county population estimates prepared by the Bureau of the Census for metropolitan counties in 1950 and 1960 and two preliminary tests on estimates prepared by the Study Group on Postcensal Population Estimates, Public Health Conference on Records and Statistics, for counties in Montana, Ohio, Oregon, and Pennsylvania in 1960. In the five separate analyses undertaken in this chapter, no significant differences were found in the accuracy of the various methods of estimating county population.

The testing of these various estimation methods served as preliminary tests of the basic statistical model, and the tests were used specifically for the purpose of determining the complexity of the statistical model as well as for determining the basis for contradictory findings reported in previous studies about the accuracy of various estimation methods.

Chapter VI described the 1950 to 1962 population censuses and population estimates of counties in the six-state area existing at the time this research project was initiated in 1962. No April 1, 1960 population estimates for counties in all six states were available at the time the study began. Therefore, it was necessary to develop April 1, 1960 estimates for the 564 counties by selected estimation methods for the proposed evaluation test. The plan was to employ the four most highly recommended methods (Component Method II, the Vital Rates Method, the Bogue-Duncan Composite Method, and the Census Variation of the Composite Method) in estimating the April 1, 1960 population of all counties, and then to employ the most accurate estimation method in developing the annual July 1, 1950 to 1962 county population estimates for the six-state area. The original plan could not be achieved in its entirety, since it was impossible to obtain the required basic data to make the April 1, 1960 population estimates of Arkansas and Nebraska counties by the two composite methods. Therefore, it was necessary to limit the tests of accuracy in the entire six-state area to only Component Method II and the Vital Rates Method. The basic input data necessary for making the April 1, 1960 county population estimates were assembled, and the actual estimates were prepared.

Chapter VII evaluated the accuracy of Component Method II and the Vital Rates Method in estimating the April 1, 1960 population of counties in Arkansas, Iowa, Kansas, Missouri, Nebraska, and Oklahoma. Also, this chapter evaluated the accuracy of Component Method II, the Vital Rates Method, and the Bogue-Duncan Composite Method in estimating the April 1, 1960 population of counties in Oklahoma. The tests uniformly indicated that there were no significant differences in the accuracy of the estimation methods tested. On the basis of these findings, Component Method II was used to prepare annual July 1, 1950 to 1962 population estimates for all 564 counties in the six-state area because it provided detailed components of population changes. These annual county estimates were then proportionately adjusted to sum to the official annual state population estimates prepared by the Bureau of the Census. The last section of Chapter VII described some of the data limitations affecting annual 1950 to 1962 county population estimates in the six states.

Finally, Chapter VIII examined the assumptions inherent in the evaluation tests for determining the accuracy of different estimation methods. It demonstrated rather conclusively that the mean absolute percentage deviations employed in published evaluation tests gave erroneous findings when counties have highly divergent total populations. Likewise, it showed that the absolute percentage estimation errors declined proportionately as the population size of counties increased. In contrast, every empirical test conducted in Chapters V and VII except one showed that the absolute estimation errors increased directly as the population size of the county (population density) increased.

Also this chapter presented various parametric and nonparametric tests which are appropriate for establishing the accuracy of two or more estimation methods for the same group of counties. Finally, it gave an illustrative example of one of the most efficient nonparametric tests for determining estimation accuracy—that of the Kruskal-Wallis one-way analysis of variance based upon ranks.

Recapitulating, Chapters V and VII examined errors in April 1, 1950 and April 1, 1960 county population estimates prepared by six different methods. Eight different multiple covariance models were used in the analysis of covariance of the common logarithms of the absolute differences between the county population estimates and census enumerations. The stated objectives of the proposed tests of accuracy were accomplished by applying one of the most precise analytical techniques known—that of the least squares method of solving simultaneous equations. From the findings of the eight different evaluation tests conducted in this study, the following six specific major conclusions were drawn:

First, there were no significant differences in the accuracy of the different estimation methods. In each of the tests, the differences were too small to be significant at the five percent level. Parametric and nonparametric tests, using different measures of estimation errors, gave identical conclusions.

Second, errors in county population estimates increased directly as the population density of the counties in the previous census decade increased. Therefore, the larger the population size of the county, the greater was the estimation error.

Third, the population estimates of nonmetropolitan counties in the six-state region had significantly smaller errors than the population estimates for metropolitan counties with central cities and the suburban metropolitan counties. However, the differences in estimation errors for the two types of metropolitan counties were not significant. Moreover, the interaction of estimation methods and metropolitan classification was not significant.

Fourth, the 1960 population estimates of metropolitan counties in the United States by the Vital Rates Method and Component Method II were more accurate than the 1950 estimates by the same two methods.

Fifth, there were significant differences among the states of Montana, Ohio, Oregon, and Pennsylvania in the accuracy of county population estimates: errors in county estimates for Montana were significantly smaller than for counties in Oregon, Ohio, and Pennsylvania; and errors in the county estimates for Oregon were significantly smaller than for counties in Ohio. Apparently two major factors account for these significant differences: **one,** the school data used in the Component Method II county population estimates in Ohio and Pennsylvania were unreliable; and **two,** the number of people per county was much higher in Ohio and Pennsylvania than in Montana and Oregon, thus the estimation errors were larger for Ohio and Pennsylvania.

Sixth, errors in county population estimates became progressively larger as the ratios of 1960 to 1950 births and deaths increased.

The failure to detect significant differences in the accuracy of the various methods in estimating county populations was an unanticipated finding, since published evaluation tests indicated actual differences. The Bogue-Duncan Composite Method was found to give smaller estimation errors than either the Vital Rates Method or Component Method II for the Oklahoma nonmetropolitan counties, even though the differences were not significant. Component Method II was more accurate than the Vital Rates Method in estimating the population of metropolitan counties with central cities in the six-state region, although the differences between the two methods were not significant. Moreover, the Vital Rates Method gave smaller errors than Component Method II in estimating the population of suburban metropolitan and of nonmetropolitan counties in the six-state area, although the differences were not significant.

The major findings of the evaluation tests carried out in this study appear rather consistent from one test to another. Finally, the statistical procedure formulated in this study for testing the accuracy of different methods of estimating population is the major methodological contribution of this pubilcation. The technique developed for determining accuracy, as well as other recommended parametric and nonparametric tests which are appropriate for testing accuracy, depart from the conventional procedures employed prior to this time (which use average absolute percentage errors to gauge accuracy) and overcome the deficiences inherent in the conventional tests. This study has shown that the percentage errors in population estimates depend to a greater extent upon the population size of counties than upon actual estimation error. Consequently, absolute percentage errors tend to decline as the population size of counties increases, irrespective of the estimation method used.