NORTHWESTERN UNIVERSITY

THE CONSTRUCTION OF PERSONAL INCOME ESTIMATES FOR COUNTIES: A STUDY IN ECONOMIC STATISTICS

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

George Christian Ehemann

Evanston, Illinois

August 1968

PREFACE

In an increasing number of states, estimates of personal income are being prepared by county on a regular basis. Recent county estimates of personal income by government and non-profit agencies are now available for more than half the states. Within the past few years county income estimation in six Plains and South Central states--Iowa, Nebraska, Oklahoma, Kansas, Missouri, and Arkansas--has been greatly stimulated by financial support from the Midwest Research Institute of Kansas City to university bureaus of business and economic research in those states. In each state estimates of personal income by county were prepared annually for the years 1950 through 1962.

My own interest in county income estimation began in the fall of 1964 when Dr. Lewis E. Wagner, then Director of the Bureau of Business and Economic Research at the University of Iowa, asked me to look over the county income estimates for Iowa which had recently been completed. The approach to estimation was the same as that now used in making almost all county income estimates: for each component of personal income, a county series was selected as the best measure of that component; this series was converted to a set of county income estimates by allocating the U. S. Department of Commerce state level estimate of the income component to counties in the same proportion that the county values bore to their own state total. An analysis of the estimation procedures revealed that although the Iowa methodology compared favorably with that used in many other states, the estimates had serious shortcomings. No account was taken of the important fact that many data on components of personal income, notably those for components of wages and salaries, measure income in the county earned rather than in the county of residence of the recipient, as required by the definition of personal income. In addition, large year to year changes in the county estimates of some income components seemed to be caused by a shift from one source of county data to another.

The analysis also led to serious doubts as to whether, in view of the qualitative and quantitative limitations of existing county data, meaningful annual estimates of personal income by county could be constructed at all. Nevertheless, it was clear that there were challenging economic and statistical questions connected with the problem of obtaining the best possible estimates of county personal income. The economic basis of the choice of county data needed to be re-examined for almost every component of income. A practical means of resolving the situs problem had to be devised, at least for wage and salary income and income of non-farm proprietors. Methods had to be devised for combining information from different sources which had the contrasting virtues of reliability and frequency of observation. Finally, some conclusions had to be drawn with regard to what limitations on a set of county personal income accounts, in terms of meaningful frequency and detail, are in fact implied by existing sources of county data. In undertaking this study, it was hoped that progress in these areas might make possible a general improvement in the quality of personal income estimates by county, estimates which are assuming an increasing importance in decision-making by state and local government, in marketing, and in regional economic analysis.

The methods of county income estimation described in the following pages are being used to estimate personal income in Iowa counties for the years 1948, 1953, 1958, and 1963. This work is now in progress. A number of individuals have provided unpublished Iowa data in connection with the preparation of these estimates. In two instances these data were used in empirical work which forms part of the present study in methodology. I am indebted to David H. Johnston, Chief, Research and Statistics, Iowa Employment Security Commission, for unpublished county data on wages and salaries by industry, and to David E. Wortman, Director, Research and Statistics, Iowa State Tax Commission, for the use of computer tapes containing selected information from the Iowa state personal income tax returns for 1963. The income tax data made it possible to undertake an interesting, and perhaps unique empirical analysis of the determinants of the income of unincorporated business enterprises.

The unpublished data not used explicitly for analysis has nevertheless made an important contribution to the author's knowledge, and the existence of these data is frequently cited in the text.

The data include unpublished detail in the official state personal income estimates for Iowa. Data of this type were supplied by Robert E. Graham, Jr., Chief, and Edwin J. Coleman, Chief, Economic Measurement Section, Regional Economics Division, Office of Business Economics, U. S. Department of Commerce; and Albert R. Kendall, Agriculture Statistician, Economic and Statistical Analysis Division, Economic Research Service, U. S. Department of Agriculture.

I am indebted to the following individuals who also supplied unpublished county data: Lenore Adkisson, Auditor, Property Tax and Valuation Department, Iowa State Tax Commission; Alvin M. David, Assistant Director, Bureau of Old-Age and Survivors Insurance, U. S. Department of Health, Education and Welfare; Don E. Dyer, Acting State Executive Director, Agricultural Stabilization and Conservation Service, U. S. Department of Agriculture; R. H. Sutherland, Agricultural Statistician, Iowa Crop and Livestock Reporting Service, U. S. Department of Agriculture and Iowa Department of Agriculture; and James W. Tarver, Professor of Sociology, Oklahoma State University.

This study would not have been possible without the generous support of the Bureau of Business and Economic Research and the University Computer Center of the University of Iowa. In particular, I am indebted to the following present and past members of the staff of the Bureau of Business and Economic Research: Burton Gearhart and Lawrence Snyder, who did the computer programming; Carol Oliven who assisted with data collection and made helpful stylistic suggestions on the manuscript; Duane Oyen, who assisted with data collection and with the computations; and Dona Fae Park and Kathy Smith, who typed the many drafts. The University of Iowa Computer Center has been supported by a series of grants from the National Science Foundation.

Finally, I am indebted to Professor (Emeritus) Frank W. Fetter of Northwestern University for his advice and ecnouragement throughout the course of this study. Helpful comments were also made by Professor Walter D. Fisher.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

AN INTRODUCTION TO COUNTY INCOME ESTIMATION

Personal income has long been recognized as a basic measure of economic activity and economic well-being. The early efforts in the 1920's and 1930's to measure personal income in the United States were followed closely by attempts to measure personal income in smaller areas. Thus, the National Bureau of Economic Research and the National Industrial Conference Board quickly extended their work on national income estimation to the estimation of personal income by state,[1] and by 1940 the U. S. Department of Commerce had undertaken to provide estimates of personal income by state on a regular basis.[2] Attempts to measure personal income by county date from 1926,[3] and several efforts in the 1930's were followed by a larger volume of work in the 1950's and 1960's. However, county income estimation has met with much less success than has the estimation of personal income by state.

---

[1]Studies that should be noted are Oswald W. Knauth, Distribution of Income by State in 1919 (New York: National Bureau of Economic Research, 1922); Maurice Leven, Income in the Various States: Its Sources and Distribution in 1919, 1920, and 1921 (New York: National Bureau of Economic Research, 1925); and John A. Slaughter, Income Received in the Various States, 1929-1935 (New York: National Industrial Conference Board, 1937).

[2]John L. Martin, "Income Payments to Individuals by States, 1929-38," Survey of Current Business, 20 (April, 1940), 8-15.

[3]H. G. Weaver, "The Development of a Basic Purchasing Power Index by Counties," Harvard Business Review, IV (April, 1926), 275-89.

The U. S. Department of Commerce has not provided official estimates of personal income by county, although estimates for large SMSA's have appeared very recently, and other work on small area income estimation is in progress.[1] The present study examines the problems that have arisen in the estimation of personal income by county and suggests ways in which the quality of personal income estimates can be improved.

Personal income has been defined as "the current income received by persons from all sources, inclusive of transfers from government and business but exclusive of transfers among persons."[2] It is the sum of wage and salary disbursements, other labor income, earnings of proprietors of unincorporated business enterprises, rental income, interest and dividends, and transfer payments. Personal contributions to social insurance funds are subtracted from this total. The personal income of an area is the sum of the personal incomes of all individuals residing in the area. Conceptually, there are no differences between the measurement of personal income by county and the measurement of

---

[1]Robert E. Graham, Jr. and Edwin J. Coleman, "Personal Income in Metropolitan Areas: A New Series," Survey of Current Business, 47 (May, 1967), 18-44. A pilot study which formed the basis of this and forthcoming work at the Department of Commerce in the estimation of personal income for multi-county areas is Robert E. Graham, Jr., "Measuring Regional Market Growth: A Case Study of the Delaware River Area," Survey of Current Business, 39 (January, 1959), 10-19. Publication of personal income estimates for single counties is not planned.

[2]U. S. Department of Commerce, Office of Business Economics, National Income, 1954 Edition, A Supplement to the Survey of Current Business (Washington, D. C.: U. S. Government Printing Office, 1954), 58.

personal income in states or nations. In practice, the sources of primary data which can be used for county income estimation are much smaller in quantity and often less suited for income estimation than the data which form the basis of the state and national personal income estimates. The contrasts in data quality are important enough to necessitate significant differences in estimation procedures, and to make county income estimation a distinct topic in social accounting.

In the past few years there has been a growing interest in the estimation of personal income by county. This interest is reflected in the increasing number of states for which recent county income estimates have been published. In contrast to the six year period 1946-1951, when statewide county income estimates other than commercial estimates were published for only six states,[1] the period 1961-1966 saw the publication of county income estimates for twenty-seven states.[2] Recent noncommercial estimates of county personal income now exist for most of the larger states, and for states in all parts of the country except New England. These states, and the year of the most recent estimate, are shown in Figure 1.

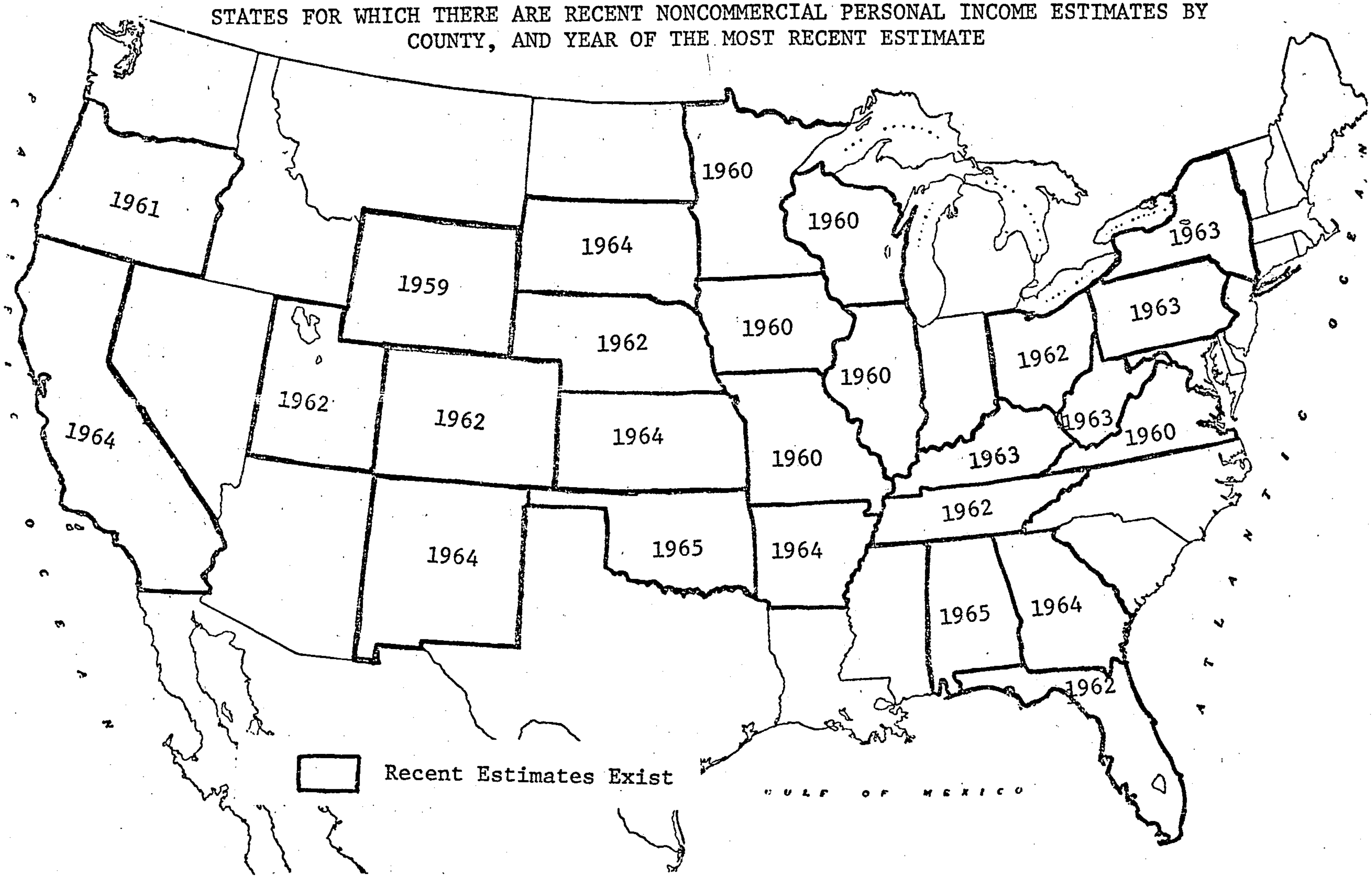
---

[1]Lewis C. Copeland, Methods for Estimating Income Payments in Counties. A report prepared by the Technical Committee for the Use of the Conference on the Measurement of County Income (Charlottesville: Bureau of Population and Economic Research, University of Virginia, 1952), 86-88.

[2]A listing of recent county income publications is provided in the bibliography.

Figure 1

STATES FOR WHICH THERE ARE RECENT NONCOMMERCIAL PERSONAL INCOME ESTIMATES BY COUNTY, AND YEAR OF THE MOST RECENT ESTIMATE



At least as notable as the larger number of states for which county income estimates are being made is the larger number of studies which provide a statement of methodology. It is now possible to form a clear picture of current practice in county income estimation. Unfortunately, these statements of methodology take us only a short way in an analysis of the problems of county income estimation. Although they indicate what was done, these statements rarely indicate why one choice was made rather than another. Typical evaluative comments are that "good cooperation was received from all persons supplying data," and "county income estimates must be interpreted with caution." Perhaps significantly, the postwar period has seen only one article on county income estimation in a major professional journal.[1] The recent studies show wide variations in methodology beyond those required by state variations in available data, and a professional concensus on methodology is needed before county income estimates are widely accepted. The considerable resources being devoted to county income estimation and the potential usefulness of good estimates in regional economic analysis indicate clearly the current need for further analytical discussion of income estimation methods at the county level.

Oskar Morgenstern has done much to make economists conscious of comparative unreliability of many of the statistics with which they deal. His book <u>On the Accuracy of Economic Observations</u> provides many

---

[1]A careful check of the <u>A.E.A. Index of Economic Journals</u> revealed only John L. Fulmer, "Regression Methods for Estimating Agricultural Income by County," <u>Review of Economics and Statistics</u>, <u>38</u> (February, 1956), 70-80.

examples of official statistics which appear to be built up from very small amounts of data, of widely different estimates for similar economic magnitudes, and of statistics whose reported frequency and accuracy cannot be justified by their informational content.[1] In view of Morgenstern's criticism, the publication, in some states, of annual estimates of county personal income in considerable detail by source raises immediate questions as to whether the volume of statistics reported are justified by the underlying data. In one state, for example, county personal income is reported in hundreds of dollars.[2]

The present study is, most broadly, a study in economic statistics, and it has been strongly influenced by the considerations of reliability raised by Morgenstern. The problems of county income estimation have been viewed as those of extracting the maximum amount of information with regard to the personal income of a county from an extremely heterogeneous body of primary data. The existing primary data have been carefully scrutinized for shortcomings in terms of their appropriateness for county income estimation, and adjustments to take account of perceived shortcomings have been suggested whereever possible. Equally important, the problems of the design of a set of county personal income accounts, which would present estimates of personal income for counties by major component and for various years, have been viewed as those of

---

[1]Oskar Morgenstern, On the Accuracy of Economic Observations (2nd ed.; Princeton: Princeton University Press, 1963).

[2]V. E. Montgomery, "Income in South Dakota in 1964," South Dakota Business Review, XXIV (November, 1965), 8.

determining the limits of the underlying data's informational content. The detail and frequency with which meaningful income statistics can be provided cannot be determined in advance, and questions in the design of county income accounts must be pursued hand in hand with questions of estimation methods for particular components of income. The guidelines which are suggested in this study for reportable county income statistics rest on a detailed analysis of the amount and quality of the relevant county data, and the methods by which these data can be processed into income estimates.

Because the method used to estimate a given component of personal income depends on the particular array of data that are available, a narrowing of the focus of the study is necessary for the discussion of some of the more specific problems of county income estimation. Many of the data which can be used for county income estimation come from federal government sources, and are thus available for all states. But other valuable sources of data cover only a single state. To restrict our attention to sources with national coverage would lend to a highly distorted picture of the extent to which data relevant for county income estimation are available. While some of the special state sources that have been used in previous work for county income estimation can be noted, it is not practical to survey the special sources that exist in each state. Hence, the present study will be especially concerned with the evaluation of alternative methods of income estimation for a particular state. Iowa has been chosen as convenient and representative. In addition, the study will be limited to consideration of methods of personal income estimation that can be used to obtain estimates for the postwar period.

The construction of social accounts is usually thought of as a task which is a prerequisite to quantitative analysis of an economy, but one which does not itself draw very heavily on economic and statistical analysis. It is apparent, however, that any processing of primary economic data into social accounts relies, implicitly or explicitly, on economic and statistical assumptions. Because the data which must serve as the basis of county income estimates are relatively weak, the realism of the economic and statistical assumptions underlying the estimation methodology has an especially important effect on the quality of the estimates. The explicit use of economics and statistics in designing a methodology for county income estimation is the underlying theme of the present study.

1. Approaches to County Income Estimation: A Historical Survey

It is useful to have, at the outset, an idea of approaches to county income estimation that have been adopted in the past, and some understanding of the degree of success which each approach has encountered. In surveying previous work, we hope to find a reference point upon which to build in refining methods of county income estimation, and to obtain some insight into the problems that have been most troublesome. The historical sketch of the present section will be followed by a comparison of some recent results of the estimation of personal income in Iowa counties using different procedures. The chapter concludes with a discussion of the need for analytical methods in the development of improved procedures for estimating income by county, and this discussion serves to introduce the problems that are the concern of the remainder of the study.

The four general approaches to county income estimation that have been applied in practice may be classified as index number methods, allocation methods, censuses and surveys, and regression methods. The nature of each of these approaches may be summarized as follows:

(1) Index numbers designed to measure personal income have been constructed by forming weighted averages of county series which are believed to be highly correlated with personal income or one or more of its major components. Both the selection of county series and the weights given to them reflect the judgment of the investigator. Index numbers measure relative levels of personal income by county; no attempt is made to translate these into a measure of the absolute level of income.

(2) Allocation methods of county income estimation are concerned with the distribution of state personal income estimates to counties. The motivation for allocation methods is that state data for personal income estimation are more extensive and more reliable than county data, and thus provides a superior guide to the magnitude of personal income components.

County data, however, may be taken as indicating the shares of particular components of personal income received by residents of the various counties. Thus, allocation methods normally treat personal income on a disaggregated basis, and associate each component of personal income with a county series. County values for an income component are obtained by assigning a share of state total to each county in proportion to the share that the county series is of its own state total.

(3) Censuses and surveys usually estimate personal income by household interviews. Information may be obtained on the distribution of income by source, although the method is not appropriate for non-cash components of personal income such as imputed rent and interest, and employer contributions to private pension and welfare funds.

(4) Regression methods of county income estimation are based on the assumption that a set of variables reported by state and by county can be specified which explain both state and county variations in personal income. Thus, a regression equation is estimated using state personal income as the dependent variable, and county income estimates are made by using the estimated regression coefficients and county values of the independent variables.

A fifth approach to county income estimation which has been suggested is to combine tabulations of gross income reported on personal income tax returns with estimates of unreported income.[1] Estimates would have to be made of types of income not subject to tax, earnings of persons with low incomes not filing returns, and under-reporting of income. Methods have not been proposed for dealing with the latter two problems, which are peculiar to this approach to county income estimation, and it will not be discussed further.

The various approaches to county income estimation could, of course, be used in combination. For example, regression methods might be used to estimate some components of personal income and a survey used to estimate others. The allocation approach subsumes the others, since county estimates of an income component obtained by any method can be scaled to a state control total. In particular, weighted sums of several county series (the index number approach) might be used to allocate certain components of personal income, rather than single series.

The first attempt to estimate income by county may be found in a 1926 article by H. G. Weaver.[2] His work combined the essential features of the index number and allocation approaches. Four economic indicators, all percentage shares of state totals, were summed with integral weights. The basic data were the sum of value added in manufacture and total values of mineral, fishery, and farm products; number of retail outlets;

---

[1]John H. Cumberland, "Suggested Improvements in Regional Income Accounting," Regional Science Association, Papers, 2 (1956), 259-71.

[2]H. G. Weaver, op. cit.

total population; and number of federal income tax returns. All but the last were always given a weight of unity, and tax returns were given a weight of one through five depending on which value resulted in the best predictions of income for states of a particular region or type. Estimates of county income were obtained by distributing the income of a state to counties in proportion to the value of the index. Weaver made estimates for all the counties in the United States, using five year averages of state income estimates.

A second example of index numbers as small area income estimates appeared in an article by Edward Thorndike in 1937.[1] Thorndike constructed an eleven component index (and two variants of it) with arbitrary weights, and computed the index for 117 small and medium sized cities. The index, which used data for the early 1930's, was expressed on a per capita basis, but no attempt was made to estimate dollar magnitudes. Thorndike attempted to choose indicators that reflected the well-being of different components of the population, but his inclusion of "sales of retail cigar stores" makes his index look strange to the modern reader. The index approach to county income estimation is no longer in current use.

Howard Bowen appears to have been the first to estimate county income by allocating individual components of state income to counties, although the allocation approach had already been applied in making

---

[1]Edward Lee Thorndike, "Variations Among Cities in Per Capita Income." Journal of the American Statistical Association, 32 (September, 1937), 471-79.

personal income estimates for states.[1,2] Bowen's estimates for Iowa counties, made in 1935, were of three years averages of income for 1927-29 and 1931-33. The definition of income used by Bowen, the returns from current production that accrue to individuals, was less inclusive than the modern definition of personal income, and income was measured on a "where earned" rather than a "where received" basis. Income was treated as the sum of income arising in each of ten industry groups, and each industry's income was allocated using a single county series. Examples of allocators used were, for agriculture, value of crops; for manufacturing, value added; for transportation, assessed value of railroad property. Bowen's work, which was intended to measure the effects of the depression on Iowa incomes, was interesting in that it showed a geographic pattern of income change for the two periods that was quite different from that of the level of per capita income.

A much more elaborate version of the allocation method was developed a few years later by W. M. Adamson, who made county income estimates for Alabama for the years 1929 and 1935.[3] Adamson made separate allocations for wages and salaries of major industry groups and farm and non-farm income from profits received by individuals. Wages

---

[1]Howard R. Bowen (coordinator), The Income of the Counties of Iowa, Iowa State Planning Board, Committee on Population and Social Trends, 1935.

[2]Knauth, *op. cit.*, p. 7 *et. seq.*

[3]W. M. Adamson, "Measurement of Income in Small Geographic Areas," *Southern Economic Journal*, *8* (April, 1942), 479-92.

and salaries were allocated in proportion to payrolls reported in industrial censuses or other sources where this was possible, and estimates for non-covered industries were based on the occupational distribution of employment reported in the 1930 Census of Population. Profits from manufacturing and mining were allocated to counties on the basis of value added. An allocator for profits from retail trade was constructed by multiplying estimated average ratios of gross returns to sales by county retail sales by type of store, and then subtracting wages and salaries from the estimate of gross return for all store types. Profits from agriculture were allocated on the basis of the value of farm products sold and consumed at home as reported by the Census of Agriculture, with adjustments for farms not reporting.

Adamson's definition of income was broader than Bowen's; he also included imputed rent on owner-occupied dwellings and, contrary to modern practice, profits from the sale of property. Rent and implicit rent from residential property was allocated on the basis of data on the value on owner-occupied dwellings and monthly rents reported in the 1930 Census of Population. Two allocators were used for rent from business property, the number of dwelling units with high value or high monthly rent, which was taken as an indicator of high income, and the number of personal income tax returns filed. An average of the two allocations was taken as an estimate of this component of income. Profits from the sale of property were also estimated on the basis of the presumed distribution of high incomes.

After 1945, efforts were made to estimate county income by the allocation method in a number of other states. In 1949 the Conference on the Estimation of County Income was organized for the purpose of developing a standard methodology for county income estimation. Although this group had a varied membership, most of the participants were associated with universities in the southern states. Two publications that resulted directly from the efforts of the Conference were <u>County Income in Seven Southeastern States</u>, by John Lancaster,[1] and a technical supplement, <u>Methods of County Income Estimation</u>, by Lewis Copeland.[2] Copeland's monograph gave the methods of allocation that had been adopted in the southeastern states, or which were under consideration at the time of writing.

County income estimation as practiced by the Conference participants showed a number of advances over the pre-war period. The definition of income was broadened to include transfer payments, and an attempt was made to measure a few components of income on a "where received" basis. Wages and salaries reported under unemployment compensation programs were tabulated, and the use of this source greatly increased portion of this type of income that could be measured directly. Several alternative methods of farm income estimation were suggested. They varied chiefly according to whether farm expenditure items in addition

---

[1]John Littlepage Lancaster (ed.), <u>County Income in Seven Southeastern States</u> (Charlottesville: Bureau of Population and Economic Research, University of Virginia, 1952).

[2]Copeland, <u>op. cit</u>.

to farm receipts were allocated to counties, and according to the level of disaggregation used in allocating receipts from crops. Non-farm proprietors' income was estimated, for most industries, by allocating the product of the number of establishments or the number of proprietors and wages and salaries per worker covered by unemployment insurance. Rent allocations were based on Census of Housing data or on property tax statistics. A variety of data were suggested for allocating interest and dividends: tabulations from state personal income tax returns, commercial bank deposits, sales of savings bonds, assessed value of intangible property, and others.

A large amount of later work in the estimation of personal income by county drew heavily on the procedures cataloged by Copeland, and the allocation approach now dominates the estimation procedures used by government and non-profit agencies. A publication of the U. S. Department of Commerce provides a bibliography of the studies that had appeared up to 1961.[1] The details of innovations made in these studies will not be followed here, since they have consisted primarily in the selection ofnew allocators for various income components, with little variation in basic approach. The detailed analysis of allocation procedures in the next chapter draws primarily on work since 1961, and is taken to represent current practice.

---

[1]U. S. Department of Commerce, Business and Defense Services Administration, Personal Income: A Key to Small-Area Market Analysis (Washington: U. S. Government Printing Office, 1961), 16-43.

The third approach to county income estimation to be discussed is use of the results of censuses and surveys. Income questions were first included in the Census of Population as part of the census of 1940. However, the amount of income received was asked only in the case of wages and salaries. Although the 1950 Census collected data on all monetary income, only an approximate indication of income by county was provided, since households were not required to report amounts of income received in excess of $10,000. Thus, the 1950 Census provided tabulations of the size distribution of income by county and reported median income, but mean or total income by county was not reported. The 1960 Census of Population, however, provides mean income and number of recipients by county for all monetary income, wages and salaries, and self-employment income. All of the censuses report income statistics for the year preceding the census.

One shortcoming of income data obtained by household interview is the tendency toward under-reporting. Because responses are based on memory rather than records, the extent of under-reporting will be greater the longer the period between the time the income was received and the time of the interview, and will also be greater for types of income received irregularly than, say, for wages and salaries. A rough guide to the reliability of the Census county income data may be obtained by comparing Census results for states and for the United States with the corresponding estimates prepared for 1949 and 1959 by the

Office of Business Economics of the Department of Commerce. An analysis of these estimates has been made by Herman Miller.[1] Miller obtained state and national estimates of monetary income in 1949 from Census data by assuming that the income recipients in each size class had incomes equal to the midpoint of the size class, while for 1959 these quantities could be obtained directly by multiplying mean income by number of recipients. Monetary income was derived from the OBE estimates of personal income by netting out non-cash components. The tendency toward under-reporting in the Census of Population was illustrated by the finding that estimated Census income for the United States was 91 per cent of adjusted OBE income in 1949, and 94 per cent of adjusted OBE income in 1959.

However, there was considerable variation in the extent of under-reporting for different income components, and in one case there was marked over-reporting. The correspondence for income from wages and salaries was quite good: the Census national estimate was 97 per cent of the OBE estimate for 1949 and 99 per cent of that estimate for 1959. At the other extreme, the Census covered only 54 per cent of income other than from earnings in 1949, and only 62 per cent in 1959. An anomolous result was that Census self-employment income was 99 per cent

---

[1]Herman P. Miller, Income Distribution in the United States, A 1960 Census Monograph (Washington: U. S. Government Printing Office, 1966), 172-181.

of the OBE national total in 1949, but 114 per cent of the national total in 1959. Apparently there was a tendency for self-employed persons to report gross rather than net income.

Similar results were obtained in comparisons of Census and OBE income estimates for states. In only four states was the 1959 Census income figure less than 90 per cent of the OBE figure. A comparison of the two estimates for self-employment income by state indicated that the over-reporting in 1959 tended to be more significant in non-farm states than in farm states, suggesting that it was the non-farm component of self-employment income that was over-reported. For Iowa in 1959, the Census estimate of wage and salary income was 103 per cent of OBE, self-employment income was 102 per cent of OBE, and income other than earnings was 58 per cent. These findings suggest good reliability for the wage and farm components of personal income, but a much lower reliability for the other components. However, because of reliance on sampling--a 20 per cent sample was used for the income question in 1950 and a 25 per cent sample in 1960--the reliability of the income statistics in the Census of Population should be expected to be lower for counties than for states.

Aside from the work of the Bureau of the Census, survey methods have not been used, except incidentally, in the construction of personal income estimates for counties. The high cost of surveys relative to other approaches is the primary reason for the neglect of this method. However, an interesting attempt to estimate personal income for an urbanized area smaller than a county was made by Charles Leven.[1] The Elgin-Dundee area, northwest of Chicago, was the area investigated. Leven's objective was to estimate a set of income and product accounts which included gross area product and personal income for the year 1956. Personal interviews and mail surveys were used to collect information for business establishments, and personal interviews were used to collect information from households. In the estimation of personal income, primary reliance was placed on the data received from businesses, which were asked the amounts of the various types of income payments which were made to households, and the shares of each type of income going to persons within the Elgin-Dundee area. The household survey focused on determining the amounts of income of various types received from outside the area, which was combined with the data from firms to obtain an estimate of the total personal income of area residents. However, respondents to the household survey were also asked to indicate a size class for their total personal income, which served as a check.

Lorin Thompson appears to have made the only attempt to estimate county personal income using the regression approach, although other investigators have used regression analysis to estimate the farm income

[1]Charles L. Leven, "Theory and Method of Income and Product Accounts for Metropolitan Areas, Including the Elgin-Dundee Area as a Case Study," doctoral dissertation, Department of Economics, Northwestern University, 1958. Leven's study was carried out under the auspices of the Center for Metropolitan Studies at Northwestern University and was supported by the Ford Foundation and the Social Science Research Council.

component.[1] Thompson developed an equation for predicting per capita income by state for the year 1950 using Department of Commerce personal income estimates and explanatory variables derived from the 1950 Census of Population. This equation was then used to estimate per capita income in 127 counties and cities in Virginia. Three independent variables were chosen: (1) the ratio of the white non-farm population to total population, (2) the ratio of military and civilian employment to population, and (3) an index of the favorableness of the industrial mix. The last variable was constructed by weighting employment by industry according to national averages of the value of total production per worker in the industry. The coefficient of determination using data for 48 states was 0.91, so that nine per cent of the variation in per capita income was unexplained.

Another measure of the reliability of the regression estimates suggest rather unfavorable results for this approach. The estimates obtained for Virginia counties and cities were scaled for consistency with the Department of Commerce estimate for the state and compared with estimates obtained by the allocation method. Almost half the counties and cities showed differences of 10 per cent or more, and in three cases the income estimates differed by more than 40 per cent. The estimates

---

[1]Lorin A. Thompson, "Appraisal of Alternative Methods of Estimating Local Area Income," in Conference on Research in Income and Wealth, Studies in Income and Wealth, XXI: Regional Income, A Report of the National Bureau of Economic Research (Princeton: Princeton University Press, 1957).

of per capita income cannot be considered true values, and in fact, they probably contained significant errors. Nevertheless, they are based on substantial amounts of direct data for income components. It is reasonable to conclude that Thompson's equation explains the distribution of income among Virginia counties rather less well than the distribution of income among states, and less well than does the allocation approach.

Two studies have used a regression approach in the estimation of farm income. Byron Johnson and Carl Nordquist used data for states to obtain an estimating equation for farm proprietors' income.[1] The dependent variable was the Department of Commerce estimate and the two explanatory variables were cash receipts from crops and cash receipts from livestock, both taken from the 1945 Census of Agriculture. A difficulty with using this equation to estimate farm income in counties is that it takes no account of intercounty differences in expenditures per dollar of sales, except to the extent that these reflect interstate averages associated with crops and livestock. A more elaborate version of the regression approach was developed by John Fulmer, who based his work on the assumption that interarea differences in wage rates for farm labor provide a good indication of interarea differences in

[1]Byron L. Johnson and Carl G. Nordquist, An Estimate of Personal Income Payments by Colorado County, 1948 (Denver: University of Denver Press, 1951), 22-25.

all agricultural income.[1] Fulmer used his model to estimate farm income by county in Virginia for 1953, and obtained results substantially different from estimates made by the allocation method.

The variables used by Fulmer are somewhat complex, and a series measuring imputed farm labor plays an important role. This series was constructed by combining state or county data for land in crops, livestock on farms, and livestock products produced with state estimates of rates of labor utilization in the corresponding farm activities. Fulmer's dependent variable was agricultural income per hour of imputed labor, where agricultural income was defined as the sum of farm proprietors' income and wages received by farm labor, less government payments to farm operators and rent on farm dwellings. The three independent variables were:

(1) the hourly farm wage rate, chosen on the ground that under competitive conditions it is a measure of the marginal productivity of both hired labor and labor supplied by the farm operator.

(2) the ratio of the farm wage rate in the following year to the farm wage rate in the current year, chosen on the ground that it provided a measure of disequilibrium in the market for farm labor. Given the wage rate, the change in the wage

---

[1] John L. Fulmer, "Measurement of Agricultural Income of Counties," in Conference on Research in Income and Wealth, Studies in Income and Wealth, XXI, op. cit., 343-58. See also Fulmer, "Regression Methods of Estimating Agricultural Income by Counties."

rate would be expected to be positively correlated with the level of agricultural income.

(3) The difference between farm receipts per hour and the hourly wage rate, divided by the wage rate. This variable was chosen on the ground that it reflected costs other than for labor.

The third variable might have been replaced by the ratio of total receipts to total labor costs, and equivalent regression results would have been obtained. Data for the farm wage rate was available for states and state economic areas, but not for counties, and the values for state economic areas had to be used in estimating agricultural income by county.

There are difficulties with all of the explanatory variables specified in Fulmer's model. If both sides of his equation are multiplied by number of hours worked, the dependent variable becomes total agricultural income and the first independent variable becomes the total return to labor. Thus, Fulmer has regressed farm income against one of its components, and in fact, Fulmer cites Gale Johnson's estimate that labor's share of agricultural income is about 60 per cent. Hence, we should expect substantial upward bias in estimates of F and R^2. However, spurious correlation may be largely offset by measurement error in this independent variable. Even though agriculture is a competitive industry, it is characterized by significant occupational immobility on the part of farm operators. Hence, the wage rate for farm labor will not accurately reflect the marginal productivity of labor which farm operators supply. Perhaps the most serious difficulty with this variable is the lack of theoretical grounds for believing, as does Fulmer, that wage rates are associated positively with rates of return to factors other than labor, so that the farm wage rate may be expected to explain part of area differences in income attributable to these factors. In equilibrium, the marginal productivity of labor (and of capital) will be equal in all areas, while differences in agricultural income persist as a result of differences in the return to land. And, short of equilibrium, no systematic association between rates of returns to different factors can be inferred.[1]

The difficulty with Fulmer's second independent variable, the ratio of wage rates in the following year to the current year, is a statistical one. Because year to year variation was small, the variable was not statistically significant at the 5 per cent level in three out of four regressions estimated by Fulmer. (The equation was estimated for four regions; the Northeastern states, the South, the Corn Belt, and the West.) Finally, there is Fulmer's third independent variable, which is equivalent to the ratio of farm receipts to labor costs. This variable is probably a poor indicator of non-labor returns to agriculture because it will be large either when returns to farm resources, including entrepreneurship, are large, or when high expenditures for intermediate goods must be covered. The relations between labor costs and expenditures

---

[1]Edward F. Denison, "Comment (on Fulmer's paper)," in Conference on Research in Income and Wealth, Studies in Income and Wealth, XXI, op. cit., 366.

for intermediate goods differs widely for different types of farming; for example, expenditures for intermediate goods are relatively higher for livestock farming than for crop farming.[1]

Fulmer compared the estimates of agricultural income obtained from his model using data for four sets of states with the U. S. Department of Commerce estimates of agricultural income. The mean discrepancies ranged from 3.3 per cent for Corn Belt states to 4.8 per cent for the Northeast. By other criteria, however, the results of the regressions with state data were less favorable. The regression for the West was not significant at the 5 per cent level. The coefficients of all statistically significant variables were positive. However, the magnitudes of the coefficients were such that, since the wage rate appeared in the denominator of two independent variables, an increase in the wage rate could lead to a decrease in predicted agricultural income.[2] This result appears to contradict the basic theoretical assumption of Fulmer's model.

Although the attempts by Thompson and Fulmer to develop a regression approach to county income estimation must be regarded as unsuccessful, it might be thought that more satisfactory regression models for estimating county income could be constructed. Walter Isard, for one, has recommended further work along these lines.[3] However, the same models which explain

---

[1]Robert H. Johnson, "Comment (on Fulmer's paper)," in Conference on Research in Income and Wealth, Studies in Income and Wealth, XXI, op. cit., 363-64.

[2]Ibid., 362.

[3]Walter Isard, Methods of Regional Analysis (Cambridge: MIT Press, 1960), 89.

interstate variations in income may not satisfactorily explain intercounty variations in income. In particular, the rural-urban dichotomy is much more important for variations in income among counties than among states. Even where the same model applies, structural differences in the economies of the states used to obtain estimates of the regression coefficients may be great enough to limit the usefulness of these coefficients for measuring the income of a particular state.

In the preceding discussion the allocation method emerges as the most promising of the approaches to county income estimation through a process of elimination, rather than through a demonstration of its effectiveness. Nevertheless, the allocation approach has a number of merits. Perhaps the most important of these are (1) the income estimates can be based on existing sources of primary data, thus avoiding the high costs of survey methods, although limited purpose surveys and special tabulations of data may be undertaken and incorporated; (2) the allocation approach, unlike survey methods, may be used to derive a set of historical personal income accounts for a county; (3) the allocation approach can provide a fair amount of detail on the components of county personal income; and (4) the allocation approach is very flexible in terms of its ability to utilize the varying arrays of primary data that exist for different states and for different years.

2. A Comparison of Existing County Income Estimates for Iowa

In addition to the early work by Bowen and the income statistics in the Census of Population, a number of other estimates have been made of income in Iowa counties. Three firms publish estimates of personal income or disposable income annually for all counties in the United States--Sales Management Magazine (disposable income),[1] Editor and Publisher (personal income),[2] and Standard Rate and Data, Inc. (disposable income).[3] A private research organization, the National Planning Association,[4] has published personal income estimates for the counties of a number of midwestern states including Iowa for the years 1950 and 1960. Other estimates have been made at the Bureau of Business and Economic Research at the University of Iowa. Robert Johnson estimated personal income for 1939 and 1947,[5] and Conrad Stucky estimated personal income

---

[1]Survey of Buying Power, Supplement to Sales Management: The Magazine of Marketing (New York: Sales Management, annual).

[2]Editor and Publisher Market Guide (New York: Editor and Publisher Company, annual).

[3]Identical estimates in Spot Television Rates and Data, Newspaper Rates and Data, and Spot Radio Rates and Data (Skokie, Illinois: Standard Rate and Data Service, annual).

[4]National Planning Association, Economic Base Study, Upper Mississippi River Basin Service Area, Technical Report No. 2, Personal Income Estimates by County for Minnesota, Wisconsin, Iowa, Illinois, and Missouri and by Selected Counties of South Dakota and Indiana, 1950 and 1960, for U. S. Army Engineer Division, North Central (Washington: National Planning Association, 1965).

[5]Robert H. Johnson, An Analysis of Iowa Income Payments by Counties (Studies in Business and Economics No. 1; Iowa City: Bureau of Business and Economic Research, State University of Iowa, 1950).

annually for the period 1950-1962.[1] Stucky revised and updated farm income estimates that had been made by Ethel Vatter.[2]

These estimates represent a variety of estimation methods and techniques. The estimates of the National Planning Association and the Bureau of Business and Economic Research are made using versions of the allocation approach primarily with establishment data, in the tradition of Adamson and Copeland. The Sales Management estimates rely primarily on projections of the income data in the Census of Population. However, other sources are used to estimate non-cash components. Raw estimates of personal income are scaled to state control totals. Because of the firm's choice of June or July publication dates for county estimates for the preceding year, Sales Management makes its own estimates of personal income by state. The final step in the Sales Management procedure is to deduct an estimate of taxes to arrive at disposable income by county.[3]

---

[1]Bureau of Business and Economic Research, State University of Iowa, "Personal Income by Major Component Annually, 1950-1962." Unpublished.

[2]Ethel G. Vatter, "The Composition and Distribution of County Farm Incomes, 1948-1957." Unpublished doctoral dissertation, Department of Economics, State University of Iowa, 1962.

[3]Although the Sales Management county income estimates are scaled to state control totals, Miller (op. cit., p. 191) seems incorrect in identifying the procedures used with the traditional allocation approach. Miller cites the following passage (Survey of Buying Power, July, 1960, p. 60): [A procedure is to] "segregate the state total into income derived from farming, manufacturing, trade, property, etc. Then the farm income would be distributed among all counties in accordance with the number of farm operators and labor . . . and so on until the sum of the income earned by the components of the county labor force would be the county income total." However, this passage occurs in the context of a discussion of estimation methods used previously, and is followed by the passage (p. 60): "These techniques were employed because prior to 1950 there had never been . . . a Census of Income to provide benchmarks as a base from which annual projections could be made."

The other two firms publishing annual county income estimates do not provide an outline of their methods.

Thus, for the years 1959 and 1960, five different estimates of Iowa income by county are available, and a comparison of these estimates could provide a guide to the precision with which income by county is currently known. Comparisons are reported below of the five estimates for 1960, to which the Census of Population estimates for 1959 have been added. The results of these comparisons are related to a similar analysis by Herman Miller. Further comparison of methods of estimation is deferred to the next chapter.

The choice of a measure for comparing county income estimates is worth considering carefully. Our interest is in the extent of agreement among the estimates in terms of the county distribution of a state's personal income. One measure of agreement, the coefficient of correlation, will be higher the more alike are the two series, but because this measure of association is a function of squared quantities, it is not readily interpreted as an index of reliability. For this purpose, an index based on absolute differences in estimated income should be preferred. A second consideration is that we should not be concerned if one estimate of county personal income leads to a different estimate of state personal income than another, since this discrepancy can be removed by a final allocation step which forces both county distributions to the same state total. Hence the index of reliability should be independent of the estimates of state personal income underlying or implied by the county estimates. Finally, the index of reliability should be independent of the size distribution of the counties being compared. Instead, the index should reflect the extent of agreement of the estimates "on the average."

A measure which satisfies these criteria is the sum of the absolute values of differences in county shares for two estimates. This measure will be called an index of dissimilarity. The index is given by the formula

$$\phi = \sum_i \left| \frac{X_i}{\Sigma X_j} - \frac{Y_i}{\Sigma Y_j} \right| ,$$

where X and Y are two income estimates, and the subscripts i and j run over counties. If n is the number of counties and S is any measure of state income, then multiplication and division of ϕ by S/n gives

$$\phi = \frac{\frac{1}{n} \Sigma \left| X_i \frac{S}{\Sigma X_j} - Y_i \frac{S}{\Sigma Y_j} \right|}{\frac{S}{n}} .$$

The terms $X_i \frac{S}{\Sigma X_j}$ and $Y_i \frac{S}{\Sigma Y_j}$ are the county income estimates scaled to a particular state control total, say the Department of Commerce personal income estimate. The index ϕ is thus the ratio of the mean absolute difference in the (scaled) county income estimates to S/n, which is mean county income.

It is easily seen that the index of dissimilarity satisfies the criteria for a measure of reliability suggested above. Because each county value is deflated by the corresponding state total, the index measures only differences in the estimated county distribution of income and not differences in the estimated total. Since ϕ has, in effect, the denominator S/n, the index provides a measure of average reliability over counties. Moreover, use of the index involves no implicit assumption that one of the estimates is the correct one, relative to which the discrepancies in the estimates are to be measured. Because the index of dissimilarity is based on absolute differences, it has a natural interpretation as a measure of reliability. If $\phi = .1$, then, after adjustment for differences in the state estimate, the absolute differences between the two county estimates will average 10 per cent of average county income.

Table 1 presents the indices of dissimilarity for the five 1960 Iowa county income estimates and the 1959 Census of Population estimate. No adjustments were made on the data for variations in the definition of income. The Sales Management estimates for 1960 are projections from the 1950 Census, since the 1959 data were not available at the time of their preparation. The table indicates average differences in the income estimates of from 6 per cent to 15 per cent. At least some of the lower values of index reflect similarities of method, since the National Planning Association--Bureau of Economic Research pairing and the comparisons involving commercial estimates account for all of the indicated average discrepancies of 6 or 7 per cent. Several high values associated with the Editor and Publisher estimates suggest that these estimates may be especially weak. Perhaps the most meaningful comparisons are those between the NPA and Bureau allocation method estimates and the Census of Population data. The indicated average discrepancies are 8 and 10 per cent, and because of differences in income definition and

TABLE 1

DISSIMILARITY INDICES FOR IOWA COUNTY INCOME ESTIMATES, 1959-1960

	CP[a]	NPA	BBER	SM	EP
National Planning Assn. (1960)	.08				
Bureau of Business and Economic Research (1960)	.10	.07			
Sales Management (1960)	.09	.11	.10		
Editor and Publisher (1960)	.13	.15	.14	.06	
Standard Rate and Data (1960)	.10	.11	.11	.06	.07

[a]*Census of Population* (1959).

in year, this range probably should be taken as an upper bound on the uncertainty associated with estimates of Iowa county income in the 1959-1960 period. But it should be noted that the average discrepancy for the NPA-Bureau estimates, which agree in income definition and year, is not very much lower.

The results may be compared with Herman Miller's analysis of 1959 county income estimates.[1] Miller's comparisons were restricted to *Census* and Sales Management data, but covered all 48 states. A somewhat different income concept was used: monetary disposable income received by households. County estimates of this quantity are published by

[1]Miller, *op. cit.*, pp. 190-197. For the underlying data see Herman P. Miller, *Comparison of 1960 Census Aggregates with Independent Estimates by State and County* (New York: Advertising Research Foundation, 1964).

Sales Management in addition to the more inclusive estimates of disposable income. Monetary income reported in the Census of Population was adjusted by Miller to exclude taxes and income not received by households, and a further adjustment was made for under-reporting. Thus two county income series were obtained that corresponded closely in definition. Miller found that for 28 per cent of all counties, differences in the two estimates were less than 5 per cent. In about a third of all counties, the series differed by more than 15 per cent. Counties which were SMSA's were found to have smaller discrepancies, on the average, but in about a fourth of these cases, the discrepancies were greater than 10 per cent. In Iowa, 61 of the 99 counties had discrepancies in the two series of 10 per cent or more.

These findings are consistent with the appraisal of county income estimates suggested by Table 1. The level of uncertainty is much larger than those associated with the national and state income estimates, and would be regarded by economists as not very satisfactory for many types of economic analysis. The comparative unreliability of county income estimates may provide an explanation of the small amount of published empirical work which utilizes such estimates, in spite of the strong upsurge of interest in regional economic analysis in the last decade. Nevertheless, the level of uncertainty is not so great as to suggest that the possibility of good county income estimates is hopeless.

3. The Need for Analytical Methods in County Income Estimation

If the allocation method is adopted as the basic approach to county income estimation, the question arises as to how this method may be refined and supplemented in such a way that the reliability of the resulting estimates is improved. A difficulty in the evaluation of alternative income estimation procedures is that since true income is unknown, there is no empirical standard against which the results of alternative procedures may be judged. The choice of procedures rests almost entirely on theoretical considerations. Nevertheless, economic and statistical theory can contribute significantly to resolving the problems of county income estimation, once these problems have been identified. This section introduces the problems that will be the concern of the remainder of the study.

The need for analytical methods in estimating income by county is a consequence of a number of important shortcomings in county data. It must be recognized that, in this context, theory is a highly imperfect substitute for a larger and more appropriate array of primary data. It is unlikely, however, that improvements in county primary data will result in the near future in data systems comparable to those which now exist at the state level. Thus it is necessary to consider the most serious shortcomings of the present data, and the extent to which they can be ameliorated through methods derived by theoretical analysis.

The most fundamental contrast for county income estimation between data availability at the county and at the state and national levels is the simple fact that many of the series that form the basis of the state and national income estimates are not tabulated by county. As a result, the corresponding components of personal income are not measured as satisfactorily, and less direct measures of these income components must be adopted. It should be the role of economic analysis, in this situation, to suggest which of several county series, or perhaps what combination of series, provides the most satisfactory basis for estimation of a particular income component. Clearly the more indirect the measures that one must resort to, the greater the burden that is placed on the economic arguments for their use.

The introduction of explicit economic arguments is crucial if progress is to be made in the estimation of these more difficult components of personal income. A substantial portion of the present study will be devoted to detailed economic evaluation of the available county series. Nevertheless, the smaller number of series tabulated by county suggests only the need to refine the allocation method, not to modify it. Attention still centers on "choosing the allocators." The same is not true of other major shortcomings of the county data, which require that the allocation approach be supplemented or modified. Adjustments on the data may be needed to strengthen their focus on the magnitude being estimated in several respects: with regard to the year for which estimates are to be prepared, the particular component of income which the data are being used to measure, or the county to which they refer. In addition, methods need to be found for supplying occasional missing values in series which are otherwise satisfactory.

The first two problems in strengthening the focus of the data--on the year of estimate and on the component of income--are somewhat interrelated. It often happens that the most appropriate and reliable data for estimating a component of personal income are available only at infrequent intervals, but that an economic relation can be specified between that series and other series which are observed more frequently. Hence, although there are two "conventional" resolutions of this dilemma--use of the less relevant contemporary data or use of the more relevant but more distant data--the possibility of exploiting the economic relationship among the variables suggests itself. This possibility leads to an interest in statistical methods for the interpolation of series with the aid of related series, or in other words, to an interest in the construction of economic models in which the variable explained (predicted) is the preferred measure of a component of personal income and the exogenous variables are variables which are observed more frequently.

Although classical linear regression models might be used to generate predicted (interpolated) values of the preferred measure of income, one assumption of such models, in particular, would seem to make them generally inappropriate for this purpose. The suspect assumption is that residuals in the regression relation for a given county are distributed independently over time. Except in cases where this assumption holds, the use of related data for interp-lation in county income estimation requires methods of estimation and prediction for a more general model in which (1) observations exist for successive cross sections, (2) disturbances associated with a single cross section are uncorrelated, and (3) the disturbances associated with a given county are autocorrelated over time. The properties of several models with these specifications will be analyzed in detail, and applications will be indicated to the estimation of farm income and wages and salaries. The values predicted by these models are recommended as allocators for the corresponding components of personal income.

With regard to improving the focus of the data on the county, the most important problem to be overcome is that series which measure components of personal income are often reported in terms of the county in which income is earned rather than the county in which it is received. Some of the cases in which this problem is most serious, including the estimation of wages and salaries and of the incomes of non-farm business proprietors, are linked directly or indirectly to intercounty commuting between places of residence and employment. In these cases economic theory is able to contribute to the development of procedures for the required adjustment of the data by providing an analysis of commuting. Most of the evidence on intercounty commuting is indirect, and it therefore appears necessary to rely on a very simple model with strong economic assumptions. In the model explored below, the principal assumption is that, taking the spatial distributions of employment and residence as given, the labor market operates in such a way that the total costs of commuting are minimized.

The assumption of minimal commuting costs given the distribution of locations is a necessary condition for the efficient allocation of economic resources in space, subject to possible constraints imposed by other factors in the location preferences of households and firms. It will be argued below that, in addition, low commuting costs are an attribute of economic equilibrium. However, the extent to which the minimum cost commuting pattern is approximated by the actual pattern depends, in part, on the levels of geographic and industrial disaggregation in terms of which two commuting patterns are compared. Computer simulation of least cost intercounty commuting patterns for a number of major industries for the census years 1950 and 1960 is permitted by the existing data. A complete model of intercounty commuting may be specified as an example of the transportation problem of linear programming, and several transportation algorithms are available for its solution. Estimates of wages and salaries, and of other components of personal income, may be redistributed from county where earned to county where received in accordance with the estimated amounts of commuting.

A final problem is that of missing values in a series for particular counties. Often _ad hoc_ methods must be adopted. However, when related information is available in a systematic way, it is sometimes possible to develop a more sophisticated method for dealing with a class of missing value problems of particular importance. In county income estimation the most important missing value problems arise in the county wage and salary data published by the U. S. Bureau of the Census. These missing values result from the need to avoid disclosure of information on the operations of individual firms. Hence, wage and salary data are withheld for an industry and county when the number of firms in a particular industry and county is small, or when one firm is sufficiently large relative to the others to dominate the county-industry statistics. Related data, which are always presented in two Census Bureau publications, County Business Patterns and the Census of Manufacturers, are the county distributions of firms by industry and employment size class. By obtaining estimates of the average size of firms in each class, these data can be used to obtain an estimate of employment, which can be used in turn to obtain an estimate of wages and salaries. In our consideration of this problem, it will be shown that the estimation of average employment of firms in each size class is greatly facilitated if one assumes that the size of firms in any industry has a lognormal distribution. The results of this discussion will apply, of course, to any problem in which class means are needed for grouped data drawn from a lognormal population.

In the following chapters these problems are discussed in turn. The choice of measures of personal income which can be used as allocators is discussed in the next chapter. Since most previous work on county income estimation has centered on this problem, the chapter presents a critical survey of recent efforts, in addition to a more complete analysis of the alternatives. Chapter Three develops a statistical model for the interpolation of mixed time-series cross-section data. Chapter Four covers the two remaining topics, a linear programming model for situs adjustment and the use of the lognormal distribution in supplying missing values in employment data. It will be seen, in the second chapter, that considerable improvement over current practice can be made in the choice of allocators for a large number of components of personal income. When this improved collection of data is used in a way that incorporates the developments in statistical methodology of the concluding chapters, the outlook appears good for reliable estimates of personal income by county.

CHAPTER TWO

ALLOCATION METHODS FOR COUNTY INCOME COMPONENTS

In Chapter One we concluded that, while the allocation method provides the most satisfactory approach to county income estimation, the choice of variables used in making the allocations needs to be reexamined. The present chapter makes the first comprehensive survey of possible county income allocators since Copeland's monograph of 1952 and goes considerably beyond that study in appraising the economic relevance and reliability of the various data. For concreteness, the discussion will focus on the problem of choosing allocators that could be used to construct a set of personal income accounts for the counties of Iowa. Data sources to be surveyed are those covering the period 1947 through 1965. The allocators selected for Iowa will be compared with the choices made in a number of county income studies published in the last few years.

Our survey and evaluation of the data available for county income estimation will center on four sets of questions: (1) What allocators are available for each component of personal income, and which allocators are best? (2) At what points does the allocation method need to be supplemented by other methods? (3) For what years and at what level of detail could good county income estimates be made? and (4) To what extent do the answers we have given to these questions differ from those that have been given in previous county income studies?

The number of county income studies that have provided a detailed statement of methodology is now sufficiently great that not all variations of the allocation method can be considered. In describing current estimation procedures, the present chapter will review the procedures of some studies in detail and cite others when they take a noteworthy alternative approach to a problem. Discussion of other studies must be omitted entirely. Of the seven studies that will be cited frequently, five appeared during 1965 and 1966. These very recent studies include four which developed county income estimates for single states--Arkansas,[1] Kansas,[2] Pennsylvania,[3] and Oklahoma[4]--and a study of the National Planning Association[5] that developed county income estimates for a number of states in the Midwest. The two other studies that will be cited frequently are earlier efforts which derived

---

[1]W. A. Heffelfinger, County Income Estimation Methods--1965 (Fayetteville: Bureau of Business and Economic Research, University of Arkansas, 1965). In the remainder of the chapter, single state county income studies will be identified by state.

[2]Darwin Daicoff, Kansas County Income: 1950-1964 (Topeka: Office of Economic Analysis, State of Kansas, 1966). The methodology employed is described on pages 39-58.

[3]Commonwealth of Pennsylvania, Department of Internal Affairs, Pennsylvania's Personal Income by County: Selected Years 1929-1963, (Report No. IP-1; Harrisburg, 1965), pp. 84-94.

[4]W. Nelson Peach, Richard W. Poole, and James D. Tarver, County Building Block Data for Regional Analysis: Oklahoma (Stillwater: Research Foundation, Oklahoma State University, 1965), pp. 4-11.

[5]National Planning Association, *op. cit.*, pp. 1-20.

county income estimates for Kentucky[1] and Illinois.[2] These seven studies, it is believed, present a fair cross section of recent work. Other studies which might have been substituted would not have added to or detracted greatly from their overall quality.

Each major component of personal income is itself composed of a number of smaller components, and one of the problems in county income estimation is to determine the level of disaggregation for allocation purposes that is feasible and worthwhile. The present chapter recommends specific allocation procedures for more than eighty components of personal income. This level of disaggregation is less fine than the 260 component breakdown used in estimating personal income by state, but it is about the average level of detail used in the more careful estimates of income by county. In order to guide the reader through this array of income components and the conclusions reached for their estimation, a set of summary tables has been prepared. The one or more tables for each category of personal income provide a list of the components into which income category should be subdivided, the allocators selected as best for each component, and the source of each allocator. Separate tables summarize

---

[1]John L. Johnson, Income in Kentucky: County Distributions by Amount, by Type, and by Size (Lexington: University of Kentucky Press, 1955), pp. 130-143.

[2]Scott Keyes, Felix C. Rodgers, and Wallace E. Reed, Personal Income in Illinois Counties: 1950, 1954, 1956, 1958, 1959 (Urbana: Bureau of Community Planning, University of Illinois, 1962), pp. 3-7 and 34-40.

the frequencies with which data from the various sources are available. It is convenient in the tables to indicate sources of data by mnemonic symbols. A key to the symbols is provided for future reference in Table 2.

The preparation of reliable estimates of personal income by county requires a finer degree of income detail for states than is provided by the published estimates of the Department of Commerce. State level detail is needed to serve as control totals for county allocators. Some additional detail for the farm sector is published for the farm sector by the Department of Agriculture.[1] State level estimates of many of the components of personal income to be discussed, however, exist only in the records of the Departments of Commerce and Agriculture. County income estimation has benefited greatly from the willingness of these departments to make detailed estimates of state personal income estimates available to qualified investigators.

The major components of personal income will be discussed in an order corresponding to the importance of their current contribution to personal income for the United States. Hence, considered in turn are the estimation of wages and salaries, property income, proprietors' income, transfer payments, other labor income, and personal contributions to social insurance funds. Separate sections are devoted to the estimation of incomes of farm and non-farm proprietors. It should be borne in mind that the importance of these components of personal income varies from one state to another, and in particular, from county to county.

---

[1] U.S. Department of Agriculture, Economic Research Service, Farm Income State Estimates 1949-1965, A Supplement to the Farm Income Situation (August, 1966).

TABLE 2

KEY TO SYMBOLS USED TO DENOTE SOURCES OF DATA

Symbol	Source[1]
BA	Iowa State Bar Association
CA	U. S. Census of Agriculture
CBP	County Business Patterns
CG	U. S. Census of Governments
CM	U. S. Census of Manufacturing
CP	U. S. Census of Population
CR	U. S. Census of Retail Trade
CS	U. S. Census of Selected Services
CW	U. S. Census of Wholesale Trade
DD	U. S. Department of Defense
FRB	Board of Governors, Federal Reserve System
HEW	U. S. Department of Health, Education, and Welfare
IBL	Iowa Bureau of Labor
ICLR	Iowa Crop and Livestock Reporting Service
IDA	Iowa Department of Agriculture
IDH	Iowa Department of Health
IDSW	Iowa Department of Social Welfare
IESC	Iowa Employment Security Commission
ISTC	Iowa State Tax Commission
JCRN	Joint Committee on Reduction of Non-essential Federal Expenditures, U. S. Congress
MHLD	Martindale-Hubbell Law Directory

Table 2 (continued)

Symbol	Source[1]
SBSI	Salary Book: State of Iowa
SCB	Survey of Current Business
TD	U. S. Treasury Department
TSC	Iowa Taxation Study Committee, Iowa Legislature
USDA	U. S. Department of Agriculture
VA	Veterans Administration

[1]Complete citations of sources are provided in the text.

A final section of the chapter draws together the findings on the frequency of observation of the many series and the quality of the estimates of income components that can be made. The implications are considered for the frequency and detail for which meaningful county personal income accounts can be prepared.

1. Wages and Salaries

Wage and salary disbursements make up by far the largest share of personal income. In 1965, for example, they contributed 66.8 per cent of U. S. personal income and 52.4 per cent of the income of Iowa.[1]

[1]"State Personal Income, 1948-65," Survey of Current Business, 46 (August, 1966), 14-15.

Fortunately, the county data sources for wages and salaries are relatively satisfactory. Most industries are covered by at least one source, and for some industries two or three sources exist. Employment must be relied upon where wage data are not available. To a large extent, wage and salary estimates for the private and government sectors must be based on different sources of data. For this reason, the problems of income estimation for the two sectors will be discussed in turn.

Private Sector Wages and Salaries

There are three major sources of county wage and salary data for industries in the private sector. The U. S. Bureau of the Census publishes wage and salary data by county for the calendar year in its periodic industrial censuses. A second source of wage and salary data is provided by County Business Patterns, another Bureau of the Census publication. These data are tabulated from taxable payrolls reported by firms under the old-age and survivors' insurance program (OASI). The third major source derives from state unemployment security programs (UI), under which covered firms report payrolls to state employment security agencies. County data from this source are usually not published, but special tabulations have been made in many states in connection with efforts to estimate county personal income. In particular, tabulations have been made for all states for the year 1962 in connection with the current area income estimation project of the U. S. Department of Commerce.[1] Each of these sources reports wages and salaries by county of work, so that the data require situs adjustment. Another common feature is that the sources are each based on information which firms are required to report by law. The legal basis of the underlying records is an important contributor to their reliability.

There is some variation among the sources in the breadth of industrial coverage, with the OASI (County Business Patterns) data most complete and the industrial censuses most selective. The industries covered by the various industrial censuses are farming, mining, manufacturing, wholesale trade, retail trade, and selected services. Coverage of services excludes, in particular, domestic and professional and related services. Within these industry groups, only the Census of Selected Services provides additional industrial detail. The published OASI data also cover mining, manufacturing, wholesale trade, retail trade, and services, but farming is not covered. Broader coverage of services is provided which includes most employment in professional and related services but excludes domestic workers. Prior to 1959, employment in religious, charitable, educational, and other non-profit organizations for which coverage by social security is elective was not reported, although by that date coverage extended to most employees in these categories.[2] In addition, the OASI data cover

---

[1]See page 2, footnote 1.

[2]U. S. Bureau of the Census, County Business Patterns, 1959, U. S. Summary, p. vii.

contract construction, transportation and public utilities (except railroads), and agricultural and related establishments other than farms (agricultural services, forestry, and fisheries). Industrial detail is provided within these categories depending on the size of the county, but because it is not presented systematically, only the nine major industrial categories may be used for county income estimation. Industrial coverage in the UI data is similar to that of County Business Patterns, except that coverage has not been extended to non-profit organizations with elective OASI coverage.

Employment security (UI) data are the dominant choice of recent county income studies for the estimation of wages and salaries in covered industries, and most rely on this source entirely.[1] The employment security data have three important advantages over both the OASI wage data and those of the industrial censuses. First, only the UI data can be obtained annually. County Business Patterns has appeared irregularly since 1946 at one- to three-year intervals, while the industrial censuses have appeared at five-year intervals on the average. Moreover, the Census of Agriculture reports data for a different set of years than do the other industrial censuses. Second, since employment security data are obtained from special

---

[1]A single example of a study relying primarily on OASI data is that of the National Planning Association, although the Oklahoma study used OASI data for small counties. None of the studies examined used the Census of Manufacturing, the Census of Wholesale Trade, or the Census of Retail Trade. Both the Census of Wholesale Trade and the Census of Retail Trade as well as the Census of Selected Services are contained in the Census of Business.

tabulations, there are no meaningful restrictions on the level of industrial disaggregation that can be obtained. The annual availability and greater industrial detail of the employment security data have important implications for the design of a set of personal income accounts for counties. These properties affect the form and frequency of the estimates that can be derived, and they allow the data limitations which exist in connection with other categories of income to play a greater role in determining the years for which income estimates are constructed. Finally, because the UI data are taken from special tabulations, they are not necessarily subject to the disclosure problems that plague the use of published data for small counties.

The employment security data have two further advantages over the OASI wage data, although these advantages are shared by the industrial census data. First, County Business Patterns reports wages and salaries for the first quarter of a year only, while the employment security data can be tabulated for all four quarters. Second, County Business Patterns reports only payrolls taxable under social security, while total payrolls of covered firms are reported under the employment security program. The second consideration is less serious than the first. The criterion for taxable payrolls has been modified several times in the postwar period. Originally, the first $3,000 of an employee's earnings were subject to tax. It has been estimated that in the first quarter of 1951, when the first $3,600 of an employee's earnings were taxable, 97.5 per cent of earnings in covered industries were taxable and reported.[1] For the years 1959 through 1965, County Business Patterns reported earnings up to the first $4,800 so that only the first quarter earnings of employees whose annual salaries exceeded $19,200 were underreported.

In spite of these advantages, employment security data appear to be less suitable for county income estimation than the leading alternatives, at least for some states. A minor disadvantage of employment security data is that, for 1959 and later years, special and less satisfactory sources must continue to be used for the estimation of wages and salaries paid by non-profit organizations. The two major disadvantages of employment security data are that (1) in some states, small firms are excluded from coverage, and (2) in some states, firms are not required to report payrolls by establishment. A consequence of the second point is that it becomes difficult or impossible to assign to counties the payrolls of firms with establishments in more than one county. The law also allows firms to report OASI taxable payrolls by firm rather than establishment. However, in connection with the publication of County Business Patterns, generally successful assignments to counties have been made as a result of intensive efforts to obtain voluntary reporting by

---

[1]U.S. Bureau of the Census, County Business Patterns, 1951, U.S. Summary, p. vi.

establishment, special surveys of multi-establishment firms, and reconciliation with the industrial censuses.[1]

Prior to 1956, coverage of firms with fewer than eight employees was at the option of the individual states,[2] and in 1951, for example, only 13 states provided coverage of firms with one or more employees.[3] Since 1956 coverage has been optional for firms with fewer than four employees, but by 1965 the number of states providing coverage of firms with one employee had increased only to 19.[4] Some evidence on the importance of firms with fewer than four employees may be obtained from the 1964 edition of County Business Patterns, which presents a two-way classification of employment--by industry and by size of establishment--for states and for the nation. For the United States, employment in covered establishments with one to three employees was 7.1 per cent of total covered employment; in Iowa, establishments in this size class accounted for 11.1 per cent of total covered employment. In retail trade, the share of employment contributed by establishments with fewer than four employees was 17.5 per cent for the United States and 17.6 per cent for Iowa.[5] Even more significant for county income estimation were the

---

[1]Strictly, County Business Patterns covers not establishments but "reporting units," where a reporting unit is defined as all the establishments of a firm within a single county. However for simplicity of exposition, the term establishment will be used for reporting unit in the text.

[2]Charles I. Schottland, The Social Security Program in the United States (New York: Appleton-Century-Crofts, 1963) p. 83.

[3]County Business Patterns, 1951, U. S. Summary, p. xiii.

[4]County Business Patterns, 1965, U. S. Summary, p. xii.

[5]County Business Patterns, 1964, U. S. Summary, p. 19, and Iowa, p. 11.

much larger shares of employment contributed by small firms in some counties.

Employment security data for Iowa are reported by firm rather than by establishment, and firms with fewer than four employees are excluded. The disastrous results from the standpoint of county income estimation are indicated in Table 3, which presents some relevant comparisons of OASI and UI data for the first quarter of 1962. The first two columns of the table show, in six industry detail, the shares of payrolls covered by OASI and UI programs that could not be allocated to Iowa counties.

Unallocable OASI-covered payrolls were less than 4 per cent of the state total in all industries, and averaged only 1.67 per cent. By contrast, unallocable UI-covered payrolls were almost 28 per cent of the total in transportation and public utilities, almost 16 per cent of the total in wholesale and retail trade, and averaged 11.76 per cent for all industries. The final column of the table shows allocable employment security payrolls as a per cent of allocable OASI payrolls. The superiority of the OASI data is indicated by the fact that in all six industries, allocable UI-covered payrolls were smaller than OASI-covered payrolls. Only in the case of manufacturing was the discrepancy within 5 per cent, while the relative rates of coverage ranged down to 72 per cent for transportation and public utilities and to less than 50 per cent for services. Hence, in Iowa, employment security data do not provide a satisfactory basis for county income estimation.

TABLE 3

COMPARISON OF OASI- AND UI-COVERED PAYROLL DATA FOR IOWA, FIRST QUARTER, 1962

	OASI-Covered Payrolls Not Allocable to Counties as Per Cent of Total	UI-Covered Payrolls Not Allocable to Counties as Per Cent of Total	Allocable UI-Covered as Per Cent of Allocable OASI-Covered Payrolls
Construction	3.92	7.32	92.0
Manufacturing	0.08	9.14	95.3
Transportation and public utilities	1.64	27.93	72.0
Wholesale and retail trade	3.63	15.75	80.8
Finance, insurance, and real estate	2.60	4.28	88.5
Services	0.73	0.98	47.7
Total[a]	1.67	11.76	83.8

[a]Includes agriculture and mining, not shown separately.

Source: Derived from County Business Patterns, 1962, and unpublished data provided by the Iowa Employment Security Commission.

The high cost of making special tabulations is yet another disadvantage of reliance on employment security data, especially for the earlier postwar years before employment security records were automated. Two recent county income studies which relied on employment security data, Illinois and Pennsylvania, used tabulations for the first quarter rather than the full year; apparently cost considerations led to giving up one of the most important advantages of UI over OASI data. The choice of UI data may reflect a failure to appreciate the weaknesses as well as the advantages of data from this source as compared to the OASI and industrial census data. The problem of missing observations for small counties in OASI data, however, has also been a significant obstacle to the choice of this source.

The seriousness of the missing value problem is suggested by examples from the Iowa data for 1948 and 1962. If we restrict our attention to the basic nine industry classification system used by County Business Patterns, the Iowa volume has 891 county-industry cells for reporting wages and salaries. Of these, 69 cells were empty in both the 1948 and 1962 editions, giving a frequency for non-reporting of 7.6 per cent. These summary statistics somewhat overstate the problem, since most of the missing values occur for two small industries--mining and agricultural services, forestry, and fisheries. However, missing values have occurred in the Iowa County Business Patterns data for all nine industries. A preference for OASI data over UI data must be qualified by the condition that the missing value problem is resolved satisfactorily. A satisfactory solution to this problem is particularly important for the selection of OASI data if the wage and salary estimates are to be adjusted to a place of work basis. When portions of wage and salary payments are assigned to neighboring counties, unknowns multiply quickly.

A middle ground in the choice of data sources for county wage and salary estimates is available when employment security reporting is by establishment and extends at least to firms with four or more employees: the employment security data might be retained and supplemented by independent estimates of payrolls of small firms. This approach has been used in making the Illinois county income estimates and was used by the Department of Commerce in making personal income estimates for a group of multicounty areas in the Middle Atlantic States.[1] The Department of Commerce secured special tabulations of OASI first-quarter wage data for small firms. These were converted to annual estimates and added to the employment security wage data. In Illinois, estimates of the number of employees in small firms were made based on the number of establishments with 1 to 3 employees, and for years prior to 1956, on the number of establishments with 4 to 7 employees, reported in County Business Patterns. It was assumed that average earnings per employee were the same for small firms as for firms covered by employment security. Assuming high quality for the county employment security data, the Commerce procedure must be given the highest marks for reliability of the alternatives for wage and salary

---

[1]Graham, op. cit.

estimation that we have considered. It is not clear that the Illinois procedure would result in significantly better wage and salary estimates than the much less laborious course of relying on County Business Patterns alone.

If employment security data are not used, then the optimal means of utilizing the remaining sources, County Business Patterns and the industrial censuses, must be considered. For industries covered by both sources, the censuses have the advantages of reporting by establishment and for the full year, but they have the disadvantage of infrequent appearance. The quality of wage and salary estimates in two important sectors, manufacturing and retail and wholesale trade, might be significantly improved if the information provided by the two sources could be combined. An interpolation procedure would be required that could use OASI data to adjust census data, the preferred source, to other years. Allocators obtained by such a procedure would be superior to allocators obtained by simple arithmetic interpolation and extrapolation of industrial census data. Aside from the missing value problem, designing an interpolation procedure is the most important problem that must be resolved in using the industrial census and OASI data for the estimation of county wages and salaries. A similar interpolation procedure to that suggested for manufacturing and wholesale and retail trade might also be used for mining. However, in many states, including Iowa, mining contributes so small a share of personal income that the extra work of combining two sources is not justified.

The remaining choices between OASI and industrial census data are straight forward. County Business Patterns rather than the Census of Selected Services must be used for wage data for the service industries, since only the former covers wages and salaries in professional and related services. For industries covered by only one of these data sources, that source must be used. The industries involved are farming (Census of Agriculture) and agricultural services, contract construction, and transportation and public utilities except railroads (County Business Patterns).

We have yet to consider estimation of wages and salaries in the industries not covered by these sources--railroads, domestic services, and prior to 1959, uncovered non-profit organizations. Wages and salaries earned by railroad and domestic service workers must be allocated to counties on the basis of employment, as reported in the 1950 and 1960 censuses of population. In most county income studies, intercounty differences in earnings are taken into account by weighting employment by the average earnings of employees in other industries. Average earnings per employee in industries covered by County Business Patterns may be used in place of the more common measure, average earnings in UI-covered industries.

A difficulty in allocating wages and salaries of non-profit organizations is that both covered and non-covered enterprises occur in several of the industries for which state control totals have been provided by the Department of Commerce. Thus, for example, medical services includes a component, non-profit private hospitals, which is not covered by UI data or by OASI prior to 1959. On the other hand, UI and OASI coverage has extended from the beginning to some non-profit membership organizations and educational services. Thus, an attempt to derive allocators for these industries using OASI or UI data in combination with data from other sources would result in double counting of employment in some industries and underrepresentation of workers in others. The National Planning Association and Pennsylvania studies avoided this problem by taking a rather different approach to the allocation of service industry wages and salaries. Earnings in all professional and related services were allocated according to industry employment as given by the censuses of population, while primary reliance was placed on the Census of Selected Services (National Planning Association) or UI data (Pennsylvania) in allocating wages and salaries in other service industries. This approach, however, fails to utilize the existing data on wages and salaries from professional and related services.

A preferred procedure would appear to be the following. Since the non-profit privately-owned hospitals in a state are a relatively small number of establishments, it is not difficult to tabulate their payrolls by county. Payrolls by hospital are reported in guide issues of Hospitals: Journal of the American Hospital Association for the years 1947 onward, although estimates of some missing values have to be made on the basis of size (number of beds). This payroll figure may be added to County Business Patterns figure for service payrolls (multiplied by four to obtain annual estimates) to obtain an allocator for all service wages and salaries except private education and non-profit membership organizations. Private education may be allocated on the basis of Census of Population employment data. Since average earnings in private education depend primarily on whether the county contains a private four-year college, the usual procedure of weighting employment by the average county wage does not seem appropriate. For non-profit membership organizations, the most appropriate employment figure provided by the Census of Population refers to professional services except education and medical. The average county wage may be used to weight this allocator.

This treatment of wages and salaries from professional services employment seems more satisfactory than other methods. The Illinois and Kentucky county income studies gave careful attention to the hospital and education components, but otherwise this treatment of wages and salaries from professional services employment seems more satisfactory than other methods that have been suggested. Both Illinois and Kentucky tabulated hospital payrolls by county. Illinois allocated private education on the basis of employment weighted by assumed earnings differentials at different types of institutions, while the Kentucky study relied on number of teachers in private elementary and secondary schools and a tabulation of payrolls in private colleges. However, in treating non-profit membership organizations, one study (Illinois) used wages and salaries in the UI covered portion and the other (Kentucky) used the number of members of religious organizations in 1936. Arkansas and Oklahoma allocated all service wages and salaries on the basis of the portion covered by employment security.

Table 4 summarizes this analysis by showing the preferred allocator for each industrial component of private sector wages and salaries. For this category of personal income, none of the preferred allocators rely on data which are peculiar to Iowa.

Government Sector Wages and Salaries

The principal components of government sector wages and salaries are those of federal civilian, military, state, and local government employees. Data for the allocation of wages and salaries originating in the government sector are much less plentiful than for the private sector, and the methods used, especially for state and local government, will depend on what data are available in a particular state. In the absence of suitable published data, a number of county income studies have made special tabulations of payrolls or employment. The most important sources of data and methods of estimation that have been used in earlier studies will be considered for the major components of the government sector. Then some alternative allocators for Iowa estimates will be compared using regression analysis.

Since 1957, employment security programs have covered federal civilian employees, and UI data provide the only source for wages and salaries by county for this sector. Because many federal government

TABLE 4

ALLOCATORS FOR PRIVATE SECTOR WAGES AND SALARIES

Income Component	Allocator and Sources[1]
Farming	Annual cash wages paid to farm labor (CA)
Agricultural services, forestry, and fisheries	First quarter industry payrolls (CBP)
Mining	First quarter industry payrolls (CB))
Contract construction	First quarter industry payrolls (CBP)
Manufacturing	Annual industry payrolls (CM)
Wholesale and retail trade	Annual industry payrolls (CW), (CR)
Finance, insurance, and real estate	First quarter industry payrolls (CBP)
Transportation and public utilities, except railroads	First quarter industry payrolls (CBP)
Railroads	Number of employees (CP) weighted by average earnings per employee in OASI-covered industries (CBP)
Services except domestic	First quarter industry payrolls (CBP)[2]
Domestic services	Number of employees (CP) weighted by average earnings per employee in OASI-covered industries (CBP)

[1]See Table 2 for key to symbols for source.

[2]See text for procedures for use prior to 1959.

departments have employees in more than one county, the allocation of reported payrolls to counties presents a difficulty similar to that encountered in the use of UI data in the private sector. Only Oklahoma, of the county income studies examined, used this source. The allocator used in almost all county income studies is federal civilian employment, a series reported by county of employment for 1950 and 1960 in a publication of a joint committee of Congress.[1] The Illinois county income study weighted employment as given in this source by average earnings in UI covered industries.

Employment data must also serve to allocate the military component of government wages and salaries. Number of military personnel by county of residende is reported in the 1960 Census of Population and is readily derived from the 1950 Census as the difference between total and civilian population. Some states have obtained unpublished data on military strength by county (a place of work series) on an annual basis from the Department of Defense. Illinois disaggregated the military component into wages and dependency allotments. Wages were allocated according to the number of military personnel, while dependency allotments were allocated according to male population of military age. This disaggregation, although appropriate, is probably not justified by the size of the components involved.

---

[1]U. S. Congress, Joint Committee on Reduction of Nonessential Federal Expenditures, Federal Civilian Employment by County, December 31, 1960 (Washington: U.S. Government Printing Office, 1961); and Federal Civilian Employment 1950 (Washington: U.S. Government Printing Office, 1950).

No sources of data with national coverage provide either payrolls or employment by county for state government. County income studies have used a number of procedures for this component. The Oklahoma and Kansas studies relied on payroll data supplied by state government agencies, and Illinois relied on a combination of tabulations of employment and payrolls for various state agencies. The Kentucky study tabulated the number of state employees by county to obtain an allocator. A defect of the latter choice is that employment does not reflect the higher earnings per worker that would be expected in counties containing a state university.

For local government, payroll data again exist in a federal census, the Census of Governments. Although this census is available for only two postwar years, 1957 and 1962, and unlike the other industrial censuses payrolls are reported only for the census month, the infrequent use of this source in county income studies is surprising. The Kansas and Illinois studies use the Census of Governments, with Illinois relying also on supplementary sources. Examples of other approaches are provided by the Kentucky and Oklahoma county income studies, which are among those that disaggregate the local government component into wages and salaries paid by school districts and by counties and municipalities. Both studies allocated the former according to salaries paid to teachers and school superintendents. Kentucky estimated wages and salaries paid by municipal and county governments using a formula based on city and county populations, while Oklahoma tabulated wages and salaries reported by counties and municipalities to the state auditor.

Other county income studies have used less disaggregation in estimating government sector wages and salaries. The Pennsylvania study allocated all the civilian component according to the number of persons employed in public administration as reported in the censuses of population, apparently overlooking the large number of government employees engaged in education. Arkansas allocated all state and local government wages and salaries on the basis of the difference between total government employment and federal civilian employment. This procedure is attractive in that it avoids any special tabulations of data, but it compounds employment statistics by place of residence (Census of Population) and place of work (Joint Committee). The National Planning Association also uses this difference for state and local government, but weights it according to average earning in local government, derived from the Census of Governments. Although the weighted allocator is preferable, both overlook a consideration noted above, the possibility of large earnings differentials in counties containing a state university.

Recent Iowa data on government sector wages and salaries include two special tabulations which may be used to supplement the published statistics. Both tabulations were prepared by the Bureau of Business and Economic Research at the University of Iowa. A tabulation by county of adjusted gross income of government employees reported on 1963 state personal income tax returns was made from computer tapes provided by the Iowa State Tax Commission. Before these data became available, a laborious tabulation by desk calculator of wages and salaries of state government employees had been completed for the fiscal year 1962, using a state source.[1] Regression analysis can be used to test the goodness of fit, or consistency, of the various data for the government sector. It can also be used to test the Illinois hypothesis that federal government employment should be weighted by a measure of the average county wage.

To test the consistency of the data sources, adjusted gross income of government employees in 1963, Y, was regressed against measures of federal civilian, state, and local government wages and salaries. The measure for the federal civilian component, X_1, was a straight line extrapolation of 1950 and 1960 civilian employment to 1963. The result differed from the 1963 state total by only 0.16 per cent.[2] The state government measure, X_2, was tabulated fiscal 1962 wages and salaries, and the local government measure, X_3, was payrolls for October, 1962, taken from the Census of Governments. The latter variable was converted to an annual basis by multiplying by 12. The regression model was thus

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + u,$$

where u was a disturbance term. The constant was specified equal to zero because its interpretation would have been ambiguous. An independent variable measuring earnings of military personnel was not included,

---

[1]State of Iowa, State Salary Book (Des Moines, annual).

[2]Obtained from U. S. Civil Service Commission, Annual Report (Washington: U. S. Government Printing Office, 1963).

since these were not included in the dependent variable. No situs adjustment was attempted on variables X_1 and X_3, which are measured by place of work, since the information necessary for such adjustment was not available. However, because of the large amount of federal employment in Rock Island, Illinois, the observation for the adjacent Iowa county, Scott, was deleted. This left 98 observations.

In addition to the value of R^2, interest attaches to the correspondence between actual and theoretical values of the coefficients. If we neglect the effect of income from sources other than wages and salaries included in Y, these are easily deduced. The coefficient b_1 is the average earnings per worker in federal employment, which in Iowa in 1963 was $7,225.[1] The coefficients b_2 and b_3 should be approximately one, since wages are regressed against wages. The regression results obtained showed a close fit ($R^2 = 0.996$),[2] but the coefficients differed substantially from their expected values. The estimated relation, together with standard errors for the coefficients, was

$$Y = \underset{(291.8)}{5472X_1} + \underset{(0.01420)}{0.7977X_2} + \underset{(0.02255)}{1.092X_3}.$$

[1]The Department of Commerce estimate of federal civilian wages and salaries was divided by federal civilian employment. The August, 1964 Survey of Current Business was used for the former, since subsequent revisions include a situs adjustment.

[2]When the intercept constant is suppressed, R^2 is still defined in the usual way, that is, as one minus the ratio of the mean square residual to the variance of the dependent variable. However, R^2 no longer corresponds to the percentage of variation explained and must be considered simply as an index of the goodness of fit. This index will not exceed one, but it may be less than zero if omitting the intercept is a serious specification error.

The coefficients for the federal and state government components are definitely too low, while the coefficient of the local government variable is somewhat too high. Because the residuals show some heteroscedasticity, the low standard errors are not sufficient in themselves to rule out the possibility of multi-collinearity as the source of the discrepancies. However, similar results, reported below, were obtained when the heteroscedasticity was removed by deflating the observations by population.

Further analysis suggests other possible sources of discrepancy in the coefficients, but these are not always in the direction observed. All three coefficients should be higher because of taxable income received other than wages and salaries. The coefficients of state government and local government wages and salaries should be higher to the extent that employment and average earnings increased between 1962 and 1963, although in fact they differ from their expected values in opposite directions. In a more subtle way, the absence of a situs adjustment could contribute jointly to the observed over-estimation of the coefficient for local government and underestimation of the coefficient for federal government. While federal government employment tends to be concentrated in the largest population centers, local government employment is more nearly proportional to population. Hence, intercounty commuting by federal employees would lead to a geographic distribution of residence of these employees (and hence of their reported income for tax purposes) which somewhat resembled the geographic distribution of local government employment. Intercounty commuting of local government employees would be less predominately away from the major employment centers. Thus the high coefficient for local government would reflect, spuriously, the earnings of federal employees residing outside their county of work.

A more mundane and probably more important source of discrepancy in the coefficients is revealed by comparison of the number of tax returns classified as government and an independent estimate of government employment. There were 100,783 Iowa tax returns in 1963 classified as from public employees.[1] Full-time employment in state and local government (October, 1962) plus federal civilian employment (June, 1963) was 103,313, a difference of 2.5 per cent. But in addition there were in October, 1962, 25,310 part-time state and local government employees in Iowa. Clearly, large numbers of government employees either did not file returns or were classified in other industries and occupations. The largest share of these employees may have been part time. Under these conditions, deviations in the estimated coefficients from their expected values is not surprising.

In spite of the shortcomings and discrepancies which have been noted, the close fit obtained from regression remains impressive, and it is not clear that either the income tax data or the payroll and employment data can be discarded as the less reliable. The tax data,

---

[1] Iowa State Tax Commission, Income Tax Division, Annual Statistical Report for the Fiscal Year Ended June 30, 1964, (Des Moines 1964), p. 1.

however, does not allow separate estimates of the federal and the state and local government sectors, and this consideration leads to a preference for the payroll and employment allocators.

The remaining question is whether federal employment should be weighted by a county series for average earnings. Although federal pay scales are set by Congress, it is plausible that earnings per employee would reflect local differentials, since higher paid administrative personnel are concentrated in the larger, and higher-wage, employment centers. To test this hypothesis, it is necessary only to compare the income-employment-payroll regressions obtained with and without weighting. Because the heteroscedastic residuals were obtained from the previous regression, a valid comparison requires that all observations be deflated by an appropriately chosen variable. It was found that the heteroscedasticity was removed when the observations were deflated by population.[1] The variable, X_4, used to weight federal civilian employment for earnings differentials, was the ratio of average earnings of employees covered in the 1962 edition of <u>County Business Patterns</u> to the county average of this variable. Hence, X_4 was defined

---

[1]Estimates of county civilian population in 1963 were prepared which adjusted estimates published by a state agency for consistency with the estimate for Iowa published by the U. S. Bureau of the Census. Data were taken from Iowa Department of Health, <u>Vital Statistics</u> (Des Moines, annual), and U. S. Bureau of the Census, <u>Current Population Reports</u> (Series P-25; Washington: U. S. Government Printing Office, 1964 and 1965), and the <u>Census of Population</u>: 1960. The discrepancies result from a higher estimate by the Bureau of the Census of the outmigration rate for Iowa since 1960.

to have a mean of one. The regression results, without and with weighting were

$$Y^* = \underset{(383.2)}{4010X_1^*} + \underset{(0.0230)}{0.8351X_2^*} + \underset{(0.02260)}{1.008X_3^*} ,$$

$$(R^2 = 0.943)$$

and

$$Y^* = \underset{(347.7)}{3953X_1X_4^*} + \underset{(0.2240)}{0.8330X_2^*} + \underset{(0.2105)}{1.012X_3^*} ,$$

$$(R^2 = 0.948)$$

where asterisks denote deflated variables. It should be noted that when observations are deflated, the interpretation given above of the coefficient of the first independent variable no longer strictly holds. Nevertheless, the greater discrepancies in all three coefficients from their expected value when X_1 is weighted, together with an increase of only half a percentage point in R^2, indicates that there is little basis for weighting X_1 if this variable is used as an allocator for federal civilian wages and salaries.

Table 5 presents the preferred allocators for Iowa government sector wages and salaries. The difficult data problems for the state and local government components in the earlier years are resolved by additional tabulations of state government salaries and by the use of two unpublished tabulations of local government wages and salaries by the Iowa Employment Security Commission.

TABLE 5

ALLOCATORS FOR GOVERNMENT SECTOR WAGES AND SALARIES

Income Component	Allocator and Source[1]
Federal civilian government	Number of federal civilian employees (JCRN)
Federal military	Number of military personnel (CP), (DD)
State government	Fiscal year payrolls (SBSI)
Local government	Census month payrolls (CG) or first quarter payrolls (IESC)

[1]See Table 2 for key to symbols for source.

---

2. Property Income

Turning from wage and salary income to property income, we move from the area in which county data are most plentiful to one of those for which they are least satisfactory. Property income, the second largest category of personal income, made up 14.3 per cent of U. S. personal income in 1965, and 15.2 per cent of personal income in Iowa. The major components of property income are imputed rent on owner-occupied non-farm dwellings, monetary rent on non-farm dwellings and commercial property, rent on farm property received by non-farm landlords, dividends, monetary interest, and imputed interest. Royalties are so small a component of personal income that they are properly neglected in county income studies, and monetary and imputed rent on farm property received by farm landlords is considered to be part of farm proprietors' income, according to Department of Agriculture and Department of Commerce definitions.

There is little agreement among the various county income studies as to the choice of allocators for property income. Several county income studies have used a single allocator. The Oklahoma study allocated all property income in proportion to deposits at Federal Reserve member banks, a procedure that does not seem adequate in view of the importance of this category of income. A different approach was followed in making county income estimates for Virginia[1] and in early estimates of county income in Iowa.[2] Both studies assumed that property income was concentrated among persons with relatively high incomes, and consequently, they chose allocators reflecting the size distribution of income derived from state personal income tax returns. In Virginia, the allocator was the proportion of all income reported by persons with incomes of $7,000 or over. The Iowa study used the proportion of all tax receipts paid by persons whose tax payment was $100 or more, and corresponded to a roughly equivalent level of income.

---

[1]John Littlepage Lancaster, Personal Income Estimates for Virginia Counties and Cities (Charlottesville: Bureau of Population and Economic Research, University of Virginia, 1963), p. 20.

[2]Robert H. Johnson, An Analysis of Iowa Income Payments by Counties (Studies in Business and Economics, New Series No. 1; Iowa City: Bureau of Business and Economic Research, State University of Iowa, 1950), p. 46.

Cross section data for states were used in the Virginia study to support the $7,000 criterion. The correlation between the amount of income received by persons with incomes over $7,000 and the amount of income from investments reported on federal income tax returns by persons above that income level was found to be 0.9939. While there is little doubt that an economic relation exists between these variables, the very high correlation is partly spurious, since no account was taken of differences in size among states.

In evaluating the size distribution of income as an indicator of property income, a relevant consideration is that some components appear to be more closely associated with differences in income size than do others. Statistics of Income reports dividends and interest reported on federal personal income tax returns by income size class annually for the United States.[1] In 1959, for example, returns showing adjusted gross income over $10,000 accounted for 76.2 per cent of dividends after exclusions and 43.1 per cent of reported interest. Nevertheless, the lower figure indicates a substantial degree of concentration, since returns over $10,000 were only 7.9 per cent of the total. It is the absence of corresponding evidence for other components of property income, and the importance of this income category, which suggests that property income should be estimated by component.

---

[1]U. S. Treasury Department, Internal Revenue Service, Statistics of Income: Individual Income Tax Returns (Washington: U. S. Government Printing Office, annual).

In the discussion that follows, we shall first consider the estimation of rental income, and then the estimation of income from interest and dividends. The conclusions which emerge from this discussion are a preferred set of allocators for the components of property income. These are shown in Table 6.

Rental Income

As noted above, the rental income of persons consists of three components: imputed rent on owner-occupied non-farm dwellings, monetary rent on non-farm dwellings and commercial property, and rent on farm property received by non-farm landlords. Although there is considerable variety among county income studies in the method of treating rental income, most studies begin from one of two types of data. One indicator of rental income is the assessed value of real property, which may be obtained from state tax commissions. For many states, county estimates of the ratio of assessed to market value of real property are also available, so that, by dividing one series by the other, an estimate of the market value of real property can be obtained. Kansas uses the market value of real property to allocate all real and imputed rental income. An alternative source of data is the U. S. Census of Housing, which has appeared for 1950 and 1960. Users of this source usually treat monetary rent and imputed rent separately, and Pennsylvania provides an example of the procedures followed. Imputed rent on owner-occupied non-farm dwellings is allocated according to the number of owner-occupied non-farm dwellings, weighted by median

TABLE 6

ALLOCATORS FOR PROPERTY INCOME

Income Component	Allocator and Source[1]
Monetary and imputed rent on non-farm residential property	Assessed value of non-farm residential property (ISTC) divided by the ratio of assessed to market value of urban residential property (ISTC)[2]
Monetary rent on commercial property	Number of establishments with 1-3 employees (CBP)
Rent on farm property received by non-farm landlords	Cash receipts from farm marketings (CA) weighted by the ratio of estimated net acres rented by farmers to acres in farms (CA), (IDA)[3]
Dividends	Weighted sum: Number of tax returns by income size class (ISTC) times share of dividend income reported by income size class, for the U. S. (TD)[4]
Monetary interest	Weighted sum: number of tax returns by income size class (ISTC) times share of interest income reported by income size class, for the U. S. (TD)[4]
Imputed interest	Demand deposits at federal reserve member banks except government and interbank (FRB)

[1]See Table 2 for key to symbols for source.

[2]For 1952-54 the assessment ratios for all urban property are reported (TSC) rather than residential.

[3]Net acres rented in 1954 may be derived from (CA) and extrapolated by acres rented (IDA) for other years.

[4]Number of families by income size class (CP) provides another measure of the size distribution of income by county.

value; monetary rent is allocated according to the number of renter-occupied non-farm dwelling units, weighted by median gross rent. A third source of data on rental income, tabulations from state personal income tax returns, provides the most desirable allocator for the monetary component, but this source has been developed only in Kentucky.[1]

The treatment of the rent component of personal income in county income studies seems in most cases to reflect misunderstanding of the definition of this income component and/or insufficient awareness of the relative magnitudes of its constituent parts.[2] In 1950, for example, approximately one-fifth of the rental income of persons in the United States was rent on business and industrial property, and about one-eighth was rent on farm property.[3] In the Pennsylvania

---

[1] John L. Johnson, Income in Kentucky: County Distributions by Amount, by Type, and by Size (Lexington: University of Kentucky Press, 1955).

[2] A portion of the definitional misunderstanding, having to do with the treatment of monetary rent on farm property received by other farmers, may have arisen from a reading of the summary definition of the rental income of persons given in National Income, 1954 Edition, A Supplement to the Survey of Current Business. On page 59 it is stated that the rental income of persons includes "the monetary earnings of persons from the rental income of real property, except those of persons primarily engaged in the real estate business." The coverage of monetary rental income is further qualified on page 91, however, with the statement: "In conformity with the Department of Agriculture treatment, all farm net rents received by or imputed to landlords living on farms are regarded as incident to the business of farming, and hence are included in national income under the heading of net income of unincorporated (farm) business rather than under the heading of rental income of persons."

[3] U. S. Department of Commerce, National Income, 1954 Edition, A Supplement, page 86.

procedure, both of these components are allocated according to an estimate of monetary rent on non-farm dwelling units paid by consumers, and it seems unlikely that rent paid on residential property would be a good indicator of rents received from other types of real property. In the Kansas procedure, on the other hand, too little weight is given to rental income from residential property. Since the Kansas allocator is an estimate of the total value of all real estate, it should be roughly proportional to all monetary and imputed rent. However, most monetary and imputed rent on business and industrial property is received by corporations, and most monetary and imputed rent on farm property is received by farmers. The share of the rental value of non-farm residential property going to corporations and unincorporated real estate firms is much smaller.

The inadequacy of the treatment of rental income extends to other studies of county income. The National Planning Association constructed an allocator which was intended as a rough measure of monetary rental income. This allocator was the sum of (1) acres of land rented by farm operators, weighted by average cash rent per acre, and (2) the number of rented farm and non-farm dwelling units, weighted by median gross rent on rented dwellings. Like the Kansas procedure, this approach gives too much weight to the farm component of rental income. In the Illinois study, the allocator for imputed rent is an estimate of the mean value of owner-occupied non-farm dwellings multiplied by the number of 1-4 unit residential structures. Since the latter series contains rented as well as owner-occupied structures, the product of the two series is an inappropriate allocator for this component of income. This confusion is compounded in the Illinois treatment of monetary rent, since the allocator for the monetary component is the sum of three series, one of which is, illogically, the estimated imputed rent on owner-occupied non-farm dwellings obtained from the foregoing allocation procedure.

The estimation procedure adopted by Arkansas differs from those in other states in that an attempt is made to compensate for the errors of single allocators by taking the average of two allocations, but the same conceptual problems persist. The allocators, applied to all rental income, are (1) the number of non-farm dwelling units, weighted by the median gross rent on renter-occupied units, and (2) the assessed value of all urban real property weighted by assessment ratios. The farm component is not represented in the allocators, and one cannot tell what weight has been given to income from business and industrial property. In addition, the reliability of the first allocator requires that the ratio of the average rental value of owner-occupied dwellings to average rental value of rented dwellings is the same in all counties. If the size distribution of income varies between counties, there does not seem to be any economic reason why this should be so.

With this background regarding current practice, consideration may now be given to the selection of optimal allocation methods for the components of rental income.

For the estimation of income from non-farm residential property, two approaches appear to be acceptable: (1) allocation on the basis of the estimated value of residential property, not total property, based on assessed values and ratios of assessed to market value, and (2) allocation of imputed rent on the basis of Census of Housing data on market value and allocation of monetary rent from estimates of rent payments from the same source. Assessment to market value ratios are available in Iowa only for an average of the years 1952-1954,[1] and for the years since 1962.[2] By taking three-year averages of assessment ratios for recent years, county statistics based on rather large samples may be obtained. In terms of the frequency of data required to implement the two approaches directly, neither approach has the advantage, since ten-year intervals separate observations in each case. However, the assessed value of residential property in Iowa can be obtained annually. If assessment ratios are interpolated, these values can be applied to current year values for assessed residential property. This advantage in data availability leads to a preference for the assessed value approach in estimating county incomes in Iowa.

No satisfactory allocators exist for rental income from non-farm business property. Estimates of the value of business property from property tax sources are unsatisfactory because of the relative unreliability of assessment ratios for such property and because only a small portion of imputed and monetary rent on business property accrues

---

[1]Iowa Tax Study Committee, Report of the Iowa Taxation Study Committee to the Governor and the General Assembly of Iowa, Part I, Iowa's Tax System--A Factual Study (Des Moines, 1956), p. 94.

[2]Iowa State Tax Commission, Summary of Real Estate Assessment Ratio Study (Des Moines, annual).

to individuals. It seems plausible, however, that most of the rental income on business property received by individuals is paid by very small firms. Hence, the allocator selected for this component of income is the number of establishments with one to three employees, as reported in <u>County Business Patterns</u>.

Another serious difficulty in the allocation of monetary non-farm rents is that state control totals are not available which distinguish rents from residential property from those received from non-farm businesses. Instead, the Department of Commerce estimates these amounts jointly from rental income reported on tax returns. One procedure, to distribute non-farm monetary rent equally between residential and business property, is suggested by the fact that for the United States as a whole, these quantities were approximately equal in 1950. An alternative approach to this problem is to base the disaggregation of the state total on a regression analysis of the data for rental income from residential property.

In states in which data on assessed values and assessment ratios exist for the census years of 1950 or 1960, property tax and <u>Census of Housing</u> data might be combined to estimate the total monetary and imputed rent attributable to residential property. Rent arising from non-farm business property in the state can then be computed as a residual. Consider the regression model utilizing county data

$$Y = b_1X_1 + b_2X_2 + u$$

where

Y = estimated market value of residential property, from property tax sources,

X_1 = number of owner-occupied dwellings times median value (<u>Census of Housing</u>)

X_2 = number of renter-occupied dwelling units times median gross rent (<u>Census of Housing</u>)

The estimated coefficient $\hat{b}_1$ will be one except as it scales for measurement error, but measurement error will be present, in particular, from the use of the median rather than the mean in constructing the variables X_1 and X_2. The coefficient $\hat{b}_2$ also scales for measurement error, but it may be thought of as reflecting primarily the reciprocal of the rate at which monetary rents are capitalized. Consider the sums $\Sigma\hat{b}_1X_1$ and $\Sigma\hat{b}_2X_2$ taken over counties. If monetary and imputed rents are capitalized at the same rate, the ratio of $\Sigma\hat{b}_1X_1$ to the state total for imputed rent will equal the ratio of $\Sigma\hat{b}_2X_2$ to the state total for monetary rent on dwellings. Since the former state total is known, this equality can be used to obtain the latter, and monetary rent on business property can be obtained as the difference between monetary rent on dwellings and total monetary rent. Because of shortcomings in the Iowa data, however, the required regression analysis has not been attempted.

The remaining component of rental income is that arising from farm property. Our earlier discussion noted one statistic, acres of land rented times average rent per acre, that could be used to allocate the farm component of monetary rent. There are two difficulties with this statistic as an allocator. First, rent per acre as reported by the Census of Agriculture may be a poor measure of rent per acre actually paid, if only a small portion of agricultural rents are determined by cash contracts. In Iowa in 1959, only 10.2 per cent of the land rented by tenants was for cash, while for the remainder rents were specified by product shares or a share-cash formula; 25.3 per cent of all tenants paid rents determined entirely by livestock shares. Second, the suggested statistic uses acres of land rented rather than net acres of land rented from owners outside the agricultural sector. It is net acres rented by the farm sector that is needed to estimate farm rents received by non-farm landlords.

The suggested allocator for this component of income is farm receipts per acre times net acres rented. The prevalence of cash-share agreements makes it impossible to use separate allocators for land rented for cash and for share, and separate data on farm receipts are not available for rented farms. Hence, it is necessary to assume that receipts per acre are the equal for owned and rented farms (or at least that their ratio is the same in all counties), and that contract cash rents, the less important component of rent on farm property, reflect cash receipts per acre. A difficulty in constructing this allocator is that the data needed to obtain net acres rented by farmers is reported only in the 1954 Census of Agriculture. To estimate this value for other years, it is necessary to multiply total acres rented by the ratio of net to total acres in 1954.

A shortcoming of all the allocators associated with monetary rent is that these variables reflect amounts of the rental income of persons originating by county more accurately than they reflect rental income received. In the absence of tax data, however, there seems to be no solution to this problem.

Interest and Dividends

The allocators that have been used in county income studies to estimate interest and dividends, like those used for rental income, seem largely unsatisfactory. Most studies have treated interest and dividends as a single income category. Illinois allocated dividends and interest according to estimated rental income, obtained previously, and Kansas and the National Planning Association chose deposits at federal reserve member banks as an allocator. Arkansas averaged allocations obtained from bank deposits and the sum of estimated wage and salary and proprietor's income obtained previously. Pennsylvania, on the other hand, made separate allocations for dividends, imputed interest, and monetary interest originating in the private and government sectors. Allocators utilized were total bank deposits, bank time deposits, dividends reported on 1934 federal income tax returns, and savings bonds sold in selected years. Kentucky made tabulations of dividends and monetary interest from state personal income tax returns, by far the best source for these components.

In spite of the shortcomings of the available data, it would appear that considerable improvement in the estimation of this category of income can be made. Separate allocations are needed for imputed interest, monetary interest, and dividends. Imputed interest arises primarily through the ownership by households of demand deposits, which pay no interest but yield income in kind in the form of banking services. Recent changes in the definition of imputed interest have reduced the amounts coming from other sources.[1] Demand deposits at federal reserve member banks suggests itself as an allocator. These data have been published, except for minor variation, at two-year intervals during the postwar period.[2] This series has the serious shortcoming that it includes the deposits of firms, although government and interbank deposits can be excluded. It also presents a situs problem, since deposits are measured for the county in which they are held, rather than the county of ownership. It seems likely, however, that there is a close linkage between where people bank and where they work. Hence if intercounty commuting patterns can be established for the redistribution of employment income to counties of residence, these commuting patterns could also be used to redistribute bank deposits to counties of ownership. Given that a situs adjustment is made, no variable appears preferable to bank deposits for the allocation of imputed interest.

---

[1]"The National Income and Product Accounts of the United States: Revised Estimates, 1929-64," Survey of Current Business, 45 (August, 1965), pp. 7-12.

[2]Board of Governors, Federal Reserve System, Distribution of Bank Deposits by Counties and Standard Metropolitan Areas. (Washington: Board of Governors of the Federal Reserve System, biennial).

In the allocation of monetary interest and dividends, it seems best to return to the notion, discussed at the beginning of this section, that property income is highly correlated with differences in income size. Although only national data exist for the amounts of dividends and interest received by income size class, separate tabulations have been made from federal personal income returns annually beginning with 1951. County data on the size distribution of income by family is provided in the 1950 and 1960 censuses of population for the preceding years. The value of the 1949 data is limited because all incomes over $10,000 are assigned to the same size classes. The 1959 data, however, recognize the size classes $10,000-$14,999, $15,000-$25,000, and over $25,000. In Iowa, county size distributions of income can be obtained for a third year, 1963, from tabulation of state personal income tax returns.

The procedure of allocating all property income according to the amount of income contributed by incomes above a critical size may be refined by constructing county indices for dividends and monetary interest which can be used as allocators. Each index would weight the share of dividend or interest income received nationally by persons in an income size class by the number of persons in that size class in the county. There are discrepancies between the two definitions of income employed and the number of tax returns as opposed to the number of families, but adjustments for these discrepancies are probably not worthwhile. Another shortcoming of the index approach is that for the early postwar years a hybrid "1950" index must be used which combines data for 1949 and 1951. Nevertheless, dividends and interest indices appear to provide the best measures of county distribution of these components of personal income.

3. Income of Non-farm Proprietors

Income of non-farm proprietors, the third largest category of personal income nationally, contributed 7.6 per cent of personal income in the U. S. in 1965, and 9.6 per cent of the income of Iowa. This category of income includes the earnings of self-employed professional workers and of business proprietors. In most states the data for estimating the income of non-farm proprietors by county is very weak. Data from 1963 state personal income tax returns are of great help in estimating this type of income in Iowa. In addition to providing a basis for allocating the major components of non-farm proprietors' income to counties in that year, the tax data can provide dependent variables for regression analysis of the determinants of non-farm proprietors' income. The regression results can then be used as a basis for forming allocators for other years. Thus, a number of the allocators that will be recommended in this section are linear combinations of county variables, where the weights are least-squares regression coefficients.

It will be convenient in this section to discuss professional income and business income separately, and in each case to precede the regression analysis with a survey of the allocation methods used in other states.

Professional Income

Allocators which have served as a basis for estimating professional income include number of employees and self-employed workers in professional services (Illinois); state personal income tax paid except farm returns (Oklahoma); and sales tax collections (Kansas). Other studies have used employment data which are better focused than Illinois' on the relevant groups of workers. The National Planning Association allocator was the sum of non-federal physicians and dentists,[1] and self-employed professional workers other than medical.[2] Arkansas used the average obtained from two allocations: (1) the number of self-employed professional workers, and (2) the number of OASI-covered establishments in five professional service industries, obtained from a special tabulation. The Pennsylvania study made separate allocations based on the numbers of physicians, dentists, and lawyers, allocating the remainder of professional income on the basis of the estimated incomes of these groups. Kentucky followed a similar procedure, but combined the remainder with business service income for purposes of allocation.

The most appealing of these allocators is the first of the Arkansas series, the number of self-employed professional workers. (The NPA allocator includes some physicians and dentists not in private practice.) Nevertheless, this series has three important shortcomings.

---

[1] U. S. Department of Health, Education, and Welfare, Public Health Service, Health Manpower Source Book, Section 10, Location of Manpower in 8 Health Occupations (Washington: U. S. Government Printing Office, 1965). (Source supplied. Reference given in National Planning Association, op. cit., p. 20 is in error if the series named are correct.)

[2] U. S. Census of Population, 1960.

First, the employment data do not reflect earning differentials among counties, which are likely to be large for self-employed professionals. Second, the aggregate employment series does not take account of earnings differentials among occupations, in particular, the higher earnings of physicians. Third, this series is available only from the 1960 Census of Population.

Income of professional workers as reported on 1963 Iowa state income tax returns includes both employee and self-employment income, and also income of professional workers from other sources. Professional workers in government are excluded. In spite of the shortcomings of this series, the Iowa tax returns do provide a reference point for evaluating the influence of other county variables on professional income, and thus assist in the estimation of the income of self-employed professional workers in other years. In constructing a model for regression analysis, it is useful, from a theoretical standpoint, to consider two dependent variables: Y_1, the income of professional employees; and Y_2, the income of self-employed professionals. Independent variables may be suggested to explain each of these variables, although the equation that can actually be estimated is the regression of the sum of Y_1 and Y_2 against the full list of independent variables. While the objective of unbiased estimates of the regression coefficients in the equation for Y_2 cannot be realized, the use of regression coefficients from the aggregated equation to predict Y_2 in years other than 1963 leads to an allocator which is probably superior to any of the alternatives.

Interest attaches to the ability to predict variations in professional income per capita. The view taken here is that population is an "allocator of last resort," and that other allocators should be judged on the basis of the extent to which they improve upon population as an allocator--that is, on the basis of percentage of variation in income per capita which is explained. It is natural to attempt to explain the variable Y_1/P, where P denotes population, on the basis of wages and salaries paid to employees in professional service industries, also deflated by population. Denoting this variable by X/P, and as in the government sector model of Section 1 omitting the intercept, we have a homogeneous equation in one independent variable,

$$\frac{Y_1}{P} = b_1 \frac{X}{P} + u_1.$$

Direct measurements of professional wages and salaries by county do not exist, but a proxy may be constructed as the difference between two other variables. These variables are (1) first quarter payrolls in all service industries, as reported in <u>County Business Patterns</u>, scaled to the annual state total estimated by the Department of Commerce, and (2) annual payrolls in business service industries, as reported in the <u>Census of Selected Services</u>, also scaled to the corresponding state personal income total. Using 1962 data for the first variable[1] but 1963 state control totals, implicit values for professional wages and salaries in Iowa counties were constructed. Although the series contained one negative value, this county was not excluded in the regression analysis reported below since it was found to have little effect on the results.

In the explanation of Y_2/P, income of self-employed workers per capita, it seems likely that the following variables would be relevant: the average earnings per employee in OASI industries, W, which serves as a proxy for per capita income; and the numbers per thousand population of physicians, H/P; dentists, D/P; and lawyers, L/P.[2] Including a constant term, the proposed equation is

$$\frac{Y_2}{P} = b_0 + b_2W + b_3\frac{H}{P} + b_4\frac{D}{P} + b_5\frac{L}{P} + u_2.$$

The sum of the two equations--the equation to be estimated--is thus

$$(1) \qquad \frac{Y}{P} = b_0 + b_1X + b_2W + b_3\frac{H}{P} + b_4\frac{D}{P} + b_5\frac{L}{P} + u.$$

---

[1]For reasons of convenience, data from the 1962 edition of County Business Patterns were used in all the regression results presented in this section. An alternative would have been to use averages from the 1962 and 1964 editions, but it is believed that the findings would not have been appreciably changed. The Census of Selected Services data and the population estimates were for 1963.

[2]The number of physicians in private practice in 1964 was taken from American Medical Association, Department of Survey Research, Distribution of Physicians, Hospitals, and Hospital Beds in the U. S. by Census Region, State, County, and Metropolitan Area, (Chicago: American Medical Association, 1964). The number of non-federal dentists in 1962 was taken from U.S. Department of Health, Education and Welfare, op. cit. The number of lawyers in 1961 was tabulated from the Martindale-Hubbell Law Directory (Summit, New Jersey: Martindale-Hubbell, Inc., annual).

It might be expected that multicollinearity would exist among the numbers of physicians, dentists, and lawyers per capita, so that not all of these variables would contribute appreciably to explanation of the dependent variable. In the work to be reported here, the view is adopted that for purposes of prediction, the model that should be chosen is the one that explains the greatest proportion of the variance in the dependent variable. But if two models have equal explanatory power, the simpler model--i.e., the one with the smaller number of independent variables--should be chosen. This position is essentially that taken by Henri Theil.[1]

Observing that the residual variance from regression obtained by least squares must be adjusted for degrees of freedom to obtain an unbiased estimate of the disturbance variance, Theil has introduced a corresponding adjustment in the coefficient of determination. Substitution of the adjusted for the unadjusted sample variance in R^2 leads to an identity for adjusted R^2, $\bar{R}^2$:

$$\bar{R}^2 = 1 - \frac{n}{n-k}(1-R^2)$$

where n is the number of observations and k is the number of independent variables, including the vector of ones. Hence $\bar{R}^2$, which (when the regression equation contains a constant) may be thought of as an estimate

[1]Henri Theil, Economic Forecasts and Policy, 2nd ed. revised. (Amsterdam: North-Holland Publishing Company, 1961), pp. 205-215.

of the true proportion of the variance explained by the equation,[1] is always smaller than R^2, the sample proportion of variance explained. Adding an additional variable to an equation will always raise R^2, but it will not necessarily raise $\bar{R}^2$. If the addition of a variable does not raise $\bar{R}^2$, we shall prefer a version of the model in which that variable is excluded. If the cause is in fact multicollinearity, then no harm is done if the multicollinearity persists over time. This seems at least as likely as does another assumption underlying our analysis: that the coefficients of the model are stable over time.

There is an alternative means of dealing with multicollinearity, however, that needs to be considered, and that is aggregation of the independent variables. Theil has shown that if two independent

---

[1]Theil does not speak of $\bar{R}^2$ in this way. His criterion, more exactly, is that regressors should be chosen which minimize the sample variance (adjusted for degrees of freedom). Values of the dependent variable are considered fixed for a particular sample, so that this criterion is equivalent to choosing regressors which maximize $\bar{R}^2$. If $\bar{R}^2$ is to be considered as an estimate of the true proportion of the variance explained, one must evaluate

$$E(\bar{R}^2) = E(1 - \frac{n\bar{s}^2}{y'y}),$$

where $\bar{s}^2$ is the adjusted residual variance and y is the column vector of deviations of observed values of the dependent variable from their mean. Both the numerator and denominator of the fraction $n\bar{s}^2/y'y$ are stochastic, and the estimator may be shown to be slightly biased. An upper bound on the (absolute value of the) bias has been found to be .096/n, if the disturbances are assumed to be normal and independently distributed. For the sample sizes of this section the bias is thus no more than .001, and we adopt $\bar{R}^2$ as an approximate measure of the proportion of the variance explained. See Carl F. Christ, Econometric Models and Methods (New York: John Wiley and Sons, 1966), p. 510.

variables are added together, the estimated coefficients of the revised model will be unbiased if it happens that the true coefficients of the combined variables are equal. In our model, a reasonable conjecture is the coefficients for dentists and lawyers should be about the same, while the coefficient for physicians should be somewhat higher. Hence, if two variables are combined, they should be the number of lawyers and dentists per thousand population. If equation (1) is the correct specification and if $b_4 = b_5$, then the expected values of R^2 and $\bar{R}^2$ are higher when $(D + L)/P$ is included among the independent variables than when D/P or L/P are included and the other is excluded. Hence, the running of this regression provides a check on any decision to exclude one of these variables.

One more question must be considered before proceding to the results, and this is the problem of outliers. A characteristic of the 1963 Iowa income tax data, which we shall meet again in our discussion of the incomes of business proprietors, is that the series for adjusted gross income of the various occupational groups tend to have one or more values much larger than the series mean when the series are expressed on a per capita basis. This situation might be discussed from the standpoint of possible heteroscedasticity in the disturbances of equation (1). One alternative is that although the value of the dependent variable is large, the values of one or more independent variables are correspondingly large, so that the estimated residual for the observation is not large. In this case no problem arises;

one would not want to disregard the additional information that the outlying observation provides. A second alternative is that the independent variables are large, but not so large that the residual is not also large. This alternative may be treated as a standard case of heteroscedasticity. It is likely that the heteroscedasticity can be removed by dividing through each observation by one of the independent variables.

A third alternative which might be said to produce outliers is that the large value of the dependent variable is accompanied by "typical" values of the independent variables. The estimated residual will in this case be especially large. There are four ways in which this alternative can arise: (1) the model is misspecified, and an additional independent variable should be included; (2) the model is misspecified, and an additional variable should be used to deflate observations for heteroscedasticity; (3) there is measurement error in the dependent variable; (4) an unlikely event (the large disturbance) has occurred. The situation is troublesome, of course, when there is no obvious candidate for an additional variable and when there is no particular reason to suspect measurement error. Since further adjustment of the data is ruled out by hypothesis, the strategies which are open are to delete the observation or to retain it.

On our assumption of typical values of the independent variables, the estimated coefficients of the variables will not be much different whether the outlying observation is included or not. But when it is included, the constant term will be larger. We recall that our objective is to make good predictions of the dependent variable. The position taken here is that whichever of the four possible states of the world is the correct one, the observation ought to be deleted.[1] This course is clearly correct in the case of measurement error. If the large disturbance is an unlikely event, it should not be allowed to color the results as if, in another drawing of the same size, a similarly large disturbance were likely to occur. If the origin of the large residual were ordinary heteroscedasticity, then the correct but non-implementable procedure for obtaining minimum-variance linear regression coefficients would be to divide each observation by a certain variable that was large only for the troublesome observation. But the result of this procedure, which would (sharply) reduce the weight given to the observation, would be about the same as that from deleting the observation entirely. Finally, if the large residual results from the specification error of omitting a variable, then the result of including the outlier is clearly undesirable. There is no reason, in this case, why the addition of a constant (the increment to the intercept caused by including the outlier) should be added to

[1] Provided, of course, that the evidence that an outlier is present is sufficiently strong. This is always in the last analysis a subjective matter, although rules may be adopted to routinize rejection. One procedure that has been used is to place an arbitrary upper bound on the acceptable size of computed residuals measured in standard deviations. An alternative rule for rejection of outliers is proposed in the next paragraph.

the predicted values of the dependent variable for all the other observations.[1]

It is necessary, for empirical work, to adopt some rule of thumb for the rejection of outliers. Such a rule should lead to a high probability of rejecting an outlier when any of the four conditions associated with our third alternative occur, but a low probability of rejection otherwise. The third alternative was marked, it will be recalled, by the joint occurrence of a large value of the dependent variable and a large residual. When the number of observations is small, it will generally be impossible to distinguish the presence of sampling variation from that of "true outliers." However, when the sample size is fairly large, conditions become more favorable. If the dependent variable and the computed residual in a regression are uncorrelated, this is evidence against the third alternative. On the other hand, with 99 observations, there is, in the absence of such correlation, only about a one per cent chance that the observation with largest value of the dependent variable will also have the largest residual. (We assume that the possibility of "ordinary" heteroscedasticity, our second alternative, has already been considered and rejected.) Hence, a reasonable rule for the treatment of outliers--applicable with a sufficient number of observations--would

---

[1]The regressions reported in Section 1 involving income of government employees per capita may be reconsidered in light of these comments on outliers. For two counties, Johnson and Story, the value of the dependent variable was more than three standard deviations larger than its mean. Both counties contain a state university. In each case, the estimated residual was small, so that the problems of concern in this section did not arise. Hence, the two observations were retained.

appear to be to reject the observation with the largest value of the dependent variable if, in the estimation of a regression equation believed to be correctly specified, that observation also has the largest computed residual.[1] This criterion could be applied more than once, and result in the rejection of more than one observation, so long as ordering of observations according to magnitude of the dependent variable and according to magnitude of the residual were identical.

We turn, finally, to the results of least-squares estimation of equation (1). Both of the problems discussed in the preceding paragraphs were encountered. When both dentists per thousand population and lawyers per thousand population were included in the regression, $\bar{R}^2$ was lower than when only L/P was included. When the aggregate variable (D + L)/P was used instead, the regression coefficients, standard errors, and coefficient of determination were:

$$(2) \qquad \frac{Y}{P} = \underset{(11.40)}{-79.93} + \underset{(0.0547)}{0.3446}\frac{X}{P} + \underset{(8.05)}{27.03}\frac{H}{P} + \underset{(5.070)^{*}}{8.767}\frac{D+L}{P} + \underset{(14.4)}{140.4}W$$

$$(R^2 = .8434)$$

---

[1]That our criterion is conditional upon a theoretical model is one dimension of its subjectivity, but a less important one than might be imagined. If there are rival hypotheses to be evaluated, a possible outcome, in the situations considered here, is that the largest residual would occur with the same observation under each hypothesis. On the other hand, if under one hypothesis the largest value of the dependent variable and the largest residual occurred on different observations, this would be good reason to favor that hypothesis.

where

D = number of dentists

H = number of physicians

L = number of lawyers

P = population, in thousands

W = average wage in OASI industries

X = wages in professional services

Y = adjusted gross income of professional persons, in thousands of dollars,

and the asterisk indicates that the coefficient of (D + L)/P is not significant at the 5 per cent level. The largest values of the dependent variable and the residual both occurred for Linn County (the Cedar Rapids SMSA), and both were more than four standard deviations from their respective means. Hence, this observation was deleted.

Two versions of equation (1), estimated with 98 observations, are presented as equations (3) and (4)

$$\text{(3)} \qquad \frac{Y}{P} = \underset{(8.24)}{-65.48} + \underset{(0.0391)}{0.3293}\frac{X}{P} + \underset{(5.56)}{29.81}\frac{H}{P} + \underset{(4.85)}{12.15}\frac{L}{P} + \underset{(10.4)}{123.0}W \qquad (R^2 = .8962)$$

$$\text{(4)} \qquad \frac{Y}{P} = \underset{(8.23)}{-65.85} + \underset{(0.0389)}{0.3306}\frac{X}{P} + \underset{(5.72)}{27.20}\frac{H}{P} + \underset{(3.605)}{9.374}\frac{D+L}{P} + \underset{(10.4)}{122.2}W \qquad (R^2 = .8967).$$

Both equations explain almost 90 per cent of the variation in professional income per capita, and there is no meaningful difference in explanatory power between them.[1] All coefficients are significant at the 5 per cent level. The variable D/P is omitted from equation (3) because it did not raise $\bar{R}^2$. As explained above, the purpose of estimating equation (4), in which (D + L)/P replaces L/P, is to serve as a check on the decision to exclude D/P. Since the explanatory power of (4) is not greater, the specification of equation (3) is accepted as correct.

For Iowa in 1963, income from independent professional practice may be estimated as the algebraic sum of three allocations. Formally, we want the difference between total professional income and professional income from wages and salaries. County estimates of the former may be obtained by allocating state professional income to counties according to the adjusted gross income of professional persons (Y). Professional income from wages and salaries may be estimated as the difference between county allocations of all service wages and salaries and of business services wages and salaries. This difference (X) has already been derived. For years other than 1963, an allocator may be derived as a weighted sum from the right hand side of equation (3). The term in X/P should be omitted, and the remaining terms should each be multiplied by population. These conclusions are indicated below in Table 8.

---

[1]It will be noted that, contrary to expectations, the constant term in equations (3) and (4) is larger algebraically than the constant in equation (2). This occurs because the deleted observation contains one of the largest values of W. When W is replaced by its mean, equations (3) and (4) show a fall in sum of the term in W and the constant. The decision to reject the observation is retained because the variation in W is not great enough to satisfactorily deflate for heteroscedasticity.

Business Proprietors' Income

The earlier work on the allocation of business proprietors' income which needs to be considered is that of Kentucky, Arkansas, Pennsylvania, and Illinois. The most satisfactory allocations of business proprietors' income were made by Kentucky for the components arising in contract construction, manufacturing, and finance; the allocator used was adjusted gross income reported on state tax returns for proprietors in these industries. Another allocator, also a reasonable one, was used by Kentucky for wholesale and retail trade and for business services. This was the number of proprietors in these industries (taken from the Census of Business) weighted by average earnings in all employment covered by the state unemployment compensation (UC) program. For mining, an allocator was constructed from state and federal sources which measured the value of coal production by unincorporated enterprises, while no county estimate was made of income originating in several smaller industries.

Employment data from the 1960 Census of Population were the basis of the county estimates of business proprietors' income for Arkansas. This Census was the only census of population to report the number of self-employed workers in retail trade and the number of self-employed workers except in retail trade, the professions, and agriculture. The two series for number of proprietors, each weighted by average earnings in all UC-covered employment, were used to allocate to counties the two corresponding components of business proprietors' income. The Pennsylvania study allocated business proprietors' income according to the number of establishments (as reported in County Business Patterns), taking this variable as a proxy for the number of proprietors. Separate allocations were made for nine industries; the allocators were not weighted by a measure of average earnings. Disadvantages of establishments as a proxy for proprietors are that incorporated establishments are counted and that self-employed workers with no employees are not counted.

The allocators selected in the Illinois county income study were also used by the National Planning Association. The retail and wholesale trade allocator, as in Kentucky, was the number of proprietors, but the weight was average earnings in UC-covered trade employment rather than average earnings in all industries. For other industries the number of establishments with 1 to 3 employees was used as an allocator, on the view that establishments in this size class were most likely to be unincorporated. Industry average earnings were used to weight the business service allocator, but for other industries, which were treated as an aggregate, average earnings in all industries was used as a weight.

Except for the industries transportation and agricultural services-forestry-fisheries, the 1963 adjusted gross income of business proprietors in Iowa can be tabulated by county from state personal income tax returns. Again we shall rely on regression analysis to provide predicted values that can be used as allocators for earlier years. The philosophy underlying the selection of variables and observations will be the same as that discussed in connection with professional income; our interest is still in explaining county variations in income per capita. However, we need to consider the industries for which regression analysis can be performed and the independent variables that should be introduced.

The largest components of business proprietors' income in Iowa arise in retail trade, wholesale trade, contract construction, and business services. Adjusted gross income is available for each of these industries, although tax returns from business services are combined with insurance and real estate. These are the industries for which regression analysis has been attempted. The transportation returns appear unusable because they include wage and salary employees. Only a small amount of proprietors' income originates in mining and manufacturing, and there is a difficulty with independent variables in these industries because of the dominance of incorporated firms. No proprietors are classified in the tax data in the industries finance or public utilities, and proprietors in agricultural services, forestry, and fisheries do not seem to be classified consistently.

Explanatory variables must be taken either from the industrial censuses or <u>County Business Patterns</u>. Most of these statistics include both incorporated and unincorporated firms. Situs adjustments needed for comparability with the dependent variable are not attempted.

Slightly different models must be used for different industries, and we begin by considering variables that could be used to explain per capita proprietors' income from retail and wholesale trade. Four independent variables suggested by theoretical considerations are:

1) Proprietors per thousand population (B/P): The coefficient of this variable should be positive.
2) Number of employees per establishment with employees (N/E): In retail and wholesale trade, the coefficient of this variable should be expected to be negative, since large firms are more likely to be incorporated, and their presence would lower the income of proprietors. The sign could be positive, however, if this effect were more than outweighed by economies of scale experienced by the unincorporated firms.
3) The ratio of number of proprietors to establishments with employees (B/E): This variable is intended to measure the economies of agglomeration experienced in counties that are major trading centers. If the economies of agglomeration are positive, as expected, the coefficient of this variable would be negative, since in that case, an increase in the number of establishments, the number of proprietors being constant, would increase proprietors' income.
4) Average earnings per employee in OASI-covered employment, excluding government (W): This variable measures the opportunity cost for proprietors of wage and salary employment. The proprietor's alternative is assumed to be wage and salary

employment in general and not employment in his present industry. Hence, average industry earnings of employees is not used as a variable. Moreover, it is held that rural-urban earnings differentials, although a symptom of disequilibrium, may be expected to have a fairly stable effect on proprietors' income geographically and over time, because of the long term and persistent nature of urban growth. (Thus, the definition and rationale of the wage variable in the present model is somewhat different from that of Fulmer's model of agricultural income, discussed in Chapter One, Section 1.)

The regression results for business proprietors' income are presented in Table 7. Good results are obtained in the case of retail trade. All four of the suggested variables contribute to the explanation of retail proprietors' income in the sense that their exclusion would lower $\bar{R}^2$, although the coefficient of establishment size (N/E) is not significant at the 5 per cent level. All coefficients have the expected sign. In the case of the economies-of-agglomeration variable B/E this result is of some theoretical interest, apart from our current objectives, in view of the unsatisfactory nature of previous attempts to measure agglomeration economies. Almost 60 per cent of the variation in retail proprietors' income is explained. There were no outliers.

TABLE 7

ESTIMATING EQUATIONS FOR BUSINESS PROPRIETORS' INCOME: PARAMETER ESTIMATES FOR ALTERNATIVE VERSIONS BY INDUSTRY[a]

Retail Trade (99 observations):

I.
$$\frac{Y_r}{P} = \underset{(13.56)}{58.11} + \underset{(0.625)}{5.653}\,\frac{B_r}{P} - \underset{(1.163)}{2.831}\,\frac{N_r}{E_r} - \underset{(7.28)}{50.60}\,\frac{B_r}{E_r} + \underset{(10.37)}{25.38}W \qquad (R^2 = .5958)$$

II.
$$\frac{Y_r}{P} = \underset{(0.094)}{4.991}\,\frac{B_r}{P} \qquad (R^2 = .2612)$$

III.
$$\frac{Y_r}{P} = \underset{(0.114)}{6.027}\,\frac{B_r \cdot W}{P} \qquad (R^2 = .2576)$$

Wholesale Trade (99 obs.)

I.
$$\frac{Y_q}{P} = \underset{(0.986)}{3.709} + \underset{(0.952)}{2.824}\,\frac{B_q}{P} - \underset{(1.853)^*}{3.218}\,\frac{B_q}{E} \qquad (R^2 = .0866)$$

II.
$$\frac{Y_q}{P} = \underset{(0.249)}{3.728}\,\frac{B_q}{P} \qquad (R^2 = -.0485)$$

III.
$$\frac{Y_q}{P} = \underset{(0.284)}{4.654}\,\frac{B_q \cdot W}{P} \qquad (R^2 = .0767)$$

IV.
$$\frac{Y_q}{P} = \underset{(1.023)}{2.674} + \underset{(1.053)}{4.150}\,\frac{B_q \cdot W}{P} - \underset{(1.542)^*}{2.808}\,\frac{B_q}{E_q} \qquad (R^2 = .1416)$$

[a]A key to symbols occurs at the end of this table.

Wholesale Trade (96 obs.):

I. $$\frac{Y_q}{P} = \underset{(0.574)}{3.298} + \underset{(0.4547)}{0.9253}\frac{B_q}{P} \qquad (R^2 = .0422)$$

II $$\frac{Y_q}{P} = \underset{(0.188)}{3.366}\frac{B_q}{P} \qquad (R^2 = -.2941)$$

III. $$\frac{Y_q}{P} = \underset{(0.218)}{4.197}\frac{B_q \cdot W}{P} \qquad (R^2 = -.1558)$$

IV. $$\frac{Y_q}{P} = \underset{(0.635)}{2.936} + \underset{(0.617)}{1.492}\frac{B_q \cdot W}{P} \qquad (R^2 = .0585)$$

Contract Construction (99 obs.):

I. $$\frac{Y_c}{P} = \underset{(1.583)}{3.505} + \underset{(0.809)}{5.596}\frac{E_c}{P} + \underset{(0.2782)^*}{0.5143}\frac{N_c}{E_c} \qquad (R^2 = .3728)$$

II. $$\frac{Y_c}{P} = \underset{(0.215)}{8.389}\frac{E_c}{P} \qquad (R^2 = .2793)$$

III. $$\frac{Y_c}{P} = \underset{(0.278)}{9.454}\frac{E_c \cdot W}{P} \qquad (R^2 = .0655)$$

Contract Construction (98 obs.):

I. $$\frac{Y_c}{P} = \underset{(1.398)^*}{2.540} + \underset{(0.710)}{5.858}\frac{E_c}{P} + \underset{(0.2444)}{0.6168}\frac{N_c}{E_c} \qquad (R^2 = .4667)$$

II. $$\frac{Y_c}{P} = \underset{(1.191)}{8.312}\frac{E_c}{P} \qquad (R^2 = .3735)$$

Table 7 (continued)

III. $$\frac{Y_c}{P} = \underset{(0.250)}{9.371}\ \frac{E_c \cdot W}{P} \qquad (R^2 = .1686)$$

Business Services, Finance, Insurance, and Real Estate (99 obs.):

I $$\frac{Y_{sf}}{P} = \underset{(4.78)}{27.11} + \underset{(0.510)}{1.397}\ \frac{B_s}{P} + \underset{(2.828)}{7.191}\ \frac{E_f}{P}$$

$$- \underset{(1.305)^*}{1.906}\ \frac{N_s}{E_s} \qquad (R^2 = .1619)$$

II. $$\frac{Y_{sf}}{P} = \underset{(0.116)}{6.119}\ \frac{B_s}{P} \qquad (R^2 = -.7580)$$

III. $$\frac{Y_{sf}}{P} = \underset{(0.198)}{7.336}\ \frac{B_s \cdot W}{P} \qquad (R^2 = -.7838)$$

IV. $$\frac{Y_{sf}}{P} = \underset{(4.82)}{27.52} + \underset{(0.518)}{1.336}\ \frac{B_s}{P} + \underset{(2.957)}{7.508}\ \frac{E_f}{P}$$

$$- \underset{(1.112)^*}{1.623}\ \frac{N_{sf}}{E_{sf}} \qquad (R^2 = .1619)$$

Key to symbols:

<u>variables</u>

B = number of proprietors

E = number of establishments

N = employment

P = population, in thousands

W = average wage in OASI industries (first quarter)

Y = adjusted gross income of proprietors, in thousands of dollars

Table 7 (continued)

industry subscripts

c = construction

f = finance, insurance, and real estate

q = wholesale trade

r = retail trade

s = business services

sf = business services, finance, insurance, and real estate

*indicates coefficient not significant at 5 per cent level

Much less satisfactory results are obtained from the wholesale trade regression. Table 7 shows that only two independent variables, proprietors per capita and the agglomeration variable, raised $\bar{R}^2$ and thus entered the forecasting equation. Both variables have the correct sign, although only proprietors per capita is significant at the 5 per cent level. Unadjusted R^2 is less than .09. An examination of the residuals shows that our criterion for the deletion of outliers--that the largest values of the dependent variable and the residual occur on the same observation--is not satisfied. However, it is nearly satisfied: the largest value of the dependent variable is 19.6 and the second largest is 19.2. Some investigators would no doubt exclude the three largest observations, including Woodbury County which contains Sioux City, a regional trading center. The regression results with 96 observations are presented in Table 7 for comparison. The only independent variable which enters on our criterion is proprietors per capita, and the value of R^2 falls by half. The model estimated with 99 observations is the one preferred.

Table 7 contains some additional results that may be used to compare the reliability of income estimates made using the regression equations with those made using the allocation method. Since the allocation method essentially adopts a one variable model with no intercept, such a model has been estimated for each industry. Two versions of the model have been estimated, corresponding to two "reasonable" choices of allocators: the number of proprietors and the number of proprietors times average earnings in OASI industries. In the regression, however, we continue to deflate by population. For retail trade, both variants of the model give an R^2 of about .26, less than half the value obtained from the complete model. Hence, the use of the regression model to estimate proprietors' income in the years prior to 1963 should yield a considerable gain in reliability.

When number of proprietors (deflated by population) is used as a single variable for wholesale trade, the value of R^2 is negative.[1] This result may be interpreted as indicating that income estimates using number of proprietors as an allocator would be less reliable than those obtained using population as an allocator.[2] On the other hand,

---

[1]On negative R^2 when there is no intercept, see p. 68, footnote 2.

[2]A moments reflection will indicate why this is so. If the independent variable were replaced by a vector of ones, none of the variation in the dependent variable would be explained, but the regression coefficient would still exist--it is simply the mean of the dependent variable. But a zero R^2 is better than a negative one. On multiplying both sides of the latter equation by P, one is led to the forecasts $\hat{Y} = bP$. If $\hat{Y}$ is used as an allocator, this is a better allocator than P times the original independent variable.

when BW/P is used as the independent variable, R^2 is greater than .07. Hence, one might argue that in estimating wholesale proprietors' income by county one would not do much better with the complete regression equation than in using BW as an allocator. The danger in the allocation approach is seen by comparing the regression result with 96 observations when BW/P is the only independent variable. Again R^2 is negative, so that the apparently good performance may be attributed to three observations. However, the variable BW/P was investigated further by substituting it for B/P in the complete regression model. In both the 99 and 96 observation cases, R^2 was larger when BW/P was used. With 99 observations, R^2 was .14--still low, but an appreciable gain. Hence this version of the model should be used to forecast wholesale proprietors' income per capita.

In specifying an equation for contractors' income, we must take account of the fact that there are no county data on the number of proprietors in contract construction. The number of establishments (per thousand population) must be used in place of the number of proprietors, and a term designed to pick up economies of agglomeration can no longer be included. But employees per establishment and average earnings in OASI industries can be retained as variables.

The attempt to explain variations in contractors' income per capita was moderately successful. Average earnings did not contribute to explanation and was excluded; R^2 equaled .37. One outlier was present, occurring in a small county with a large and apparently unincorporated construction firm. When the observation was excluded, R^2 rose to .47 and the coefficient of employees per establishment changed from not significant to significant at the 5 per cent level. This coefficient was positive, however, suggesting that economies of scale outweigh the competition from large incorporated firms in this industry. When these results are compared with the results from regression models with one independent variable, a distinctly different pattern appears than that found in the case of wholesale trade. The values of R^2 (with 98 observations) are .37 and .17, well below those of the complete model. Further, the higher value of R^2 occurs when the explanatory variable is not weighted by average earnings.

In selecting a model for proprietors' income from business services, insurance, and real estate, we must deal with most of the problems encountered thus far, and in addition cope with an unfortunate choice of industry definition. On the one hand, the data for the insurance-real estate variables are contaminated by the inclusion of financial establishments, none of which are proprietorships. On the other hand, the existence of data on number of proprietors for business services but not the remaining components necessitates a hybrid of the models used for retail and wholesale trade and for contract construction. Fortunately, we are not troubled in this model with problem of outliers.

Variables to be considered in the explanation of business service-insurance-real estate proprietors' income per capita include business service proprietors per capita, finance-insurance-real estate establishments per capita, employees per establishment (business services),[1] employees per establishment (finance-insurance-real estate), proprietors per establishment (business services), and average earnings in OASI employment. Another variable that was tried was the number of finance-insurance-real estate establishments with 1 to 3 employees. It was conjectured that since the finance component of this industry tended to have establishments that were larger than the other components, establishments with 1 to 3 employees might be a better measure of establishments in insurance and real estate.

Only three of the seven proposed independent variables contributed to explanation of proprietors' income per capita on the $\bar{R}^2$ criterion; the regression results with these variables are shown in Table 7. The three variables entering were business service proprietors per capita, all finance-insurance-real estate establishments per capita, and employment per establishment (business services). The coefficient of the last variable was negative in sign but not significant at the 5 per cent level. Only about 16 per cent of the variation in the dependent variable was explained. However, this is much better than the results that follow from the allocation method. Single independent variable regressions without an intercept were run using business proprietors per capita and business proprietors per capita weighted by average OASI earnings.

---

[1]The definition of an establishment for business services is not the same as that used for other industries in this section. Here we mean total establishments in a county, with and without a payroll. Elsewhere we mean the number of OASI reporting units in a county, a series that exists for all services but not business services. See page 53, footnote 1, on reporting units.

In both cases large negative values of $\bar{R}^2$ were obtained, indicating the inappropriateness of these models.

Using an argument similar to that used in investigating the determinants of professional income, a check on the decision to exclude one variable, employees per establishment in finance-insurance-real estate, can be made by aggregation. Employees per establishment in the composite industry business services-finance-insurance-real estate was included in the regression in place of employees per establishment for its two components. The regression results were very similar and R^2 was the same to four places. Thus, no reason was found to reject the earlier model.

In summary, regression models have been found which explain the county variation in the four major components of business proprietors' income. The amount of variation per capita explained, however, varies from high in the case of retail trade to low in cases of wholesale trade and business services-insurance-real estate. In all four industries, on the other hand, the regression models provide better predictions of proprietors' income than do allocation methods. This finding is illustrated by coefficients of determination based on allocation models which are often less than half as large as those obtained from more complete models, and by negative coefficients of determination for some allocators, indicating that they perform less well than would population. Thus, for these four industries, the regression models should be used to generate allocators for proprietors income in years prior to 1963.

The question which remains is the selection of allocators for business proprietors income in other industries. The single variable regressions do not suggest any consistent pattern of performance for the three leading alternatives--number of establishments, number of establishments weighted by average earnings in OASI-covered employment, and population. Since a choice must be made, we follow the results obtained for contract construction, which is the only industry where the single variable regressions used establishments rather than proprietors. For this industry, the results point to the number of establishments (unweighted) as an allocator. Moreover, since in the complete model (98 observations), employees per capita entered positively and significantly, the measure of establishments should not be restricted to those with small numbers of employees. We thus return, perhaps unexpectedly, to the allocator chosen by Pennsylvania. An outline of these conclusions is provided in Table 8.

4. Proprietary Income From Farming

The income of farm proprietors contributed only 2.8 per cent of U. S. personal income in 1965. In Iowa, one of the leading agricultural states, its share was much larger--15.0 per cent--almost as large as that of property income, the state's second most important category of personal income. Historically, the share of farm proprietors' income in Iowa has been even greater: in 1948 it equaled 37.2 per cent of total personal income. Hence, our interest in designing a methodology for constructing a series of personal income accounts for Iowa counties in the postwar period should lead us to give careful attention to this component. A broader reason for giving special attention to farm income in county income estimation follows from the fact that income from

TABLE 8

ALLOCATORS FOR NON-FARM PROPRIETORS' INCOME

Income Component	Allocator and Source[1]
Agricultural services, forestry, and fisheries	Industry establishments (CBP)
Mining and manufacturing	Industry adjusted gross income of proprietors (ISTC)[2]
Contract construction	Industry adjusted gross income of proprietors (ISTC)[3]
Wholesale and retail trade	Industry adjusted gross income of proprietors (ISTC)[3]
Transportation and public utilities	Industry establishments (CBP)
Business services, finance, insurance, and real estate	Industry adjusted gross income of proprietors (ISTC)[3]
Professional services	Algebraic sum of the following allocations: all professional earnings, according to adjusted gross income of professional workers (ISTC); less all service wages and salaries, according to first quarter service payrolls (CBP); plus business service wages and salaries, according to annual selected services payrolls (CS)[3]

[1]See Table 2 for key to symbols for source.

[2]For years other than 1963, industry establishments (CBP).

[3]For years other than 1963, value predicted from equation described in text.

farming shows the smallest degree of geographic concentration of any major income category. As a result, a state in which the overall share of farm income is small may at the same time have many counties in which the share of farm income is large or even dominant. If the incomes of these counties are to be estimated reliably, good estimates of the farm income component must be obtained.

The national and state level estimates of farm proprietors' income are constructed by the U. S. Department of Agriculture using methods which are quite different from those used in estimating other categories of personal income. In the absence of direct information on farm income, an indirect procedure has been adopted, in which farm income is estimated as the difference between receipts and expenditures, with an adjustment for the value of changes in farm inventories. The state and national farm income estimates are based on detailed estimates of the respective components of these quantities. It is natural to consider this approach in the estimation of farm income by county. Coupled with the allocation method, this approach leads to the distribution to counties of the various state receipt and expenditure items in proportion to available county data.

Although this approach has been widely used in estimating county farm income, there are persuasive grounds for rejecting it. Estimation of net income as a residual places a heavy burden on the accuracy of the allocation procedure, since the errors in the estimates of receipts and expenditures obtained by allocation will not necessarily be in the same direction. The percentage error in the difference between receipts and expenditures will, on the average, be substantially greater than the error in either receipts or expenditures taken alone. Because less data are available at the county level than at the state or national levels, the problem of the accuracy of a residual is much more serious.

Alternative methods of farm income estimation that have been employed in other county income studies are Fulmer's regression analysis with state data, described in Chapter One, and estimates based primarily on the allocation of farm receipts. Examples of the latter approach are the Oklahoma study, which simply allocated net farm income on the basis of cash receipts from farm marketings, and the Kansas study, which subtracted government payments to farmers from net farm income, allocating the former according to tabulated disbursements by county and the remainder according to cash receipts. Illinois used a highly modified version of the allocation method, in which the allocator for production expenses was constructed by multiplying estimated gross receipts (defined to include cash receipts from marketings, government payments, and change in the value of inventories) by a ratio of production expenses to gross receipts, obtained for multi-county areas from farm management records.

For areas of any size the assumption underlying the Oklahoma and Kansas studies, that farm income or farm income less government payments is proportional to cash receipts from marketings, does not seem very plausible. The Illinois approach is more attractive, although questions must be raised with regard to the size and representativeness of the sample of farms used to construct the required ratio, which may have included disproportionate numbers of large and well-managed farms. In any event, the Illinois methodology depends on data peculiar to that state. Past county income studies thus have failed to develop a theoretically attractive and practical alternative to the residual approach to county estimation of farm income.

It is significant that Oklahoma, Kansas, and Illinois each made estimates of farm income by the residual approach and rejected these estimates in favor of the alternatives of the preceding paragraph. The basis for rejecting these estimates was not the difficulty of making reliable estimates, however, but an unfavorable appraisal of the farm income estimates that resulted when this method was applied. None of these studies base their conclusions on an examination of the reliability of the county data or on a comparison with other estimates of farm income known to be more reliable. Rather, their authors contend that estimates obtained by this approach do not appear plausible, and they express concern that wide differences in net income per farm are found in adjacent counties, that negative values are sometimes found for net income, and that there is too wide a dispersion in the estimates by counties of net income per farm.

There are two difficulties in accepting this line of reasoning. In the first place, the precise methods used in obtaining the farm income estimates which were judged unacceptable have not been published. As we shall see below, there are substantial differences in the way in which the USDA approach has been adapted for income estimation by county in different states. One cannot tell whether the methodological choices underlying the rejected estimates compare favorably or unfavorably with the best methodologies that have been employed. Second, it is not at all clear what are sufficient grounds--in terms of results, not methodology--for rejecting a set of estimates. The variation in per farm income by county in some states might be large. There might also be wide differences in per farm earnings in adjacent counties. In fact, casual observation suggests that the relative proportions of flat and hilly land, presumably an important factor in per farm earnings, may vary widely between adjacent counties.

Finally, there should be no difficulty in accepting negative values for net farm income in a county when it is remembered that depreciation on capital is counted as expenditure as well as operating costs. Losses during some years are common among firms in other industries, and farming is not noted for high or sustained rates of return. Moreover, farms sustaining losses may well cluster geographically because of similarities in weather, terrain, etc.

In view of these considerations, our discussion will center on the task of refining the methods used to allocate farm receipts, expenditures, and inventory changes to counties. One fact that stands out in the detailed analysis presented below is the superiority of data presented in the U. S. Census of Agriculture, which appears only at five-year intervals, for the allocation of almost every component of farm receipts and expenditures. Only the value of home consumption, government payments, taxes on farm property, and the value of changes in farm inventories are not associated with allocators based or partly based on this data source in the estimation procedure selected for Iowa. It does not seem possible to obtain reasonable estimates of net farm income as a residual in noncensus years. Farm receipt and expenditure variables are highly volatile over time, and when Census of Agriculture allocators are applied to a different year than that to which they refer, much larger proportional adjustments are required to produce consistency with state totals for the corresponding receipt and expenditure items.

Although farm income cannot be estimated reliably by the allocation approach in intercensal years, a large volume of county data exists, in Iowa and other major farm states, which is related to levels of farm income. Annual data which may be obtained for Iowa include production of major crops and livestock products, numbers of major types of livestock on farms, and number on farms of tractors, trucks, and persons.[1] Thus the problem of making farm income estimates for intercensal years is one of designing an interpolation procedure which will incorporate this information, and in which farm income estimates made by procedures described in this section function as benchmarks.

Farm Receipts

Farm receipts are defined to include a number of types of monetary and imputed farm income: cash receipts from farm marketings, government

[1]The relevant sources are Iowa Department of Agriculture, Iowa Assessors Annual Farm Census, (Des Moines, annual) and Iowa Crop and Livestock Reporting Service, County Estimates of Cattle and Hog Numbers, (Des Moines, annual, mimeographed).

payments to farmers under various programs, the value of home consumption, and the gross rental value of farm dwellings. The gross rental value of farm dwellings is partly offset by an item that enters on the expenditure side, farm rent paid to non-farm landlords. The item government payments excludes payments under Commodity Credit Corporation price support programs, since these appear as part of cash receipts from marketings.

Farm marketings and government payments are the receipts components which are estimated most easily. The Census of Agriculture permits the disaggregation of receipts from crops into four components: vegetables, fruits and nuts, other field crops, and forest and horticultural specialty products. Dollar receipts for each crop category are reported in the Census year. However, since all these components except other field crops are small in Iowa, a satisfactory procedure is to use a single allocator--total receipts--for receipts from crops. A more useful disaggregation is presented for livestock and livestock products. Because of the importance of livestock receipts in Iowa agriculture, separate allocations should be made for cattle and calves, hogs, dairy products, and poultry and poultry products. An allocator for receipts from other livestock and livestock products may be obtained as the difference between total reported receipts from livestock, and the sum of reported receipts from the preceding items.

The allocation of government payments to counties is facilitated by annual county tabulations of payments by program made by the Agricultural Stabilization and Conservation Service of the Department of Agriculture. The importance of various programs varies from one state to another. In Iowa, payments under the agricultural conservation, conservation reserve (soil bank), and feed grains programs account for most of the total and provide an adequate basis for allocation. Payments under wheat, sugar, and wool subsidy programs are small in Iowa but significant in other states.

Of the county income studies surveyed, only Kentucky reports the procedure suggested. Several other studies do not report precise methods used. Arkansas relies on production data to allocate receipts from leading crops, and number of livestock on farms to allocate receipts from leading livestock activities. Total receipts from farm marketings (Census of Agriculture) is used to allocate the remainder. In addition to the double counting involved, the Arkansas procedure for crops is inappropriate to the extent that there is variation among counties in the share of feed crops marketed and the share used as input for livestock production.

There are no satisfactory data for allocating the value of home consumption to counties. Although the value of farm products consumed on the farm can be derived from the 1945 Census of Agriculture, 1945 was a war year, and the pattern of home consumption in that year is probably not a good indicator for the postwar period. Nevertheless, this series is used by Kentucky, and Pennsylvania extrapolated this series to 1959 on the basis of change in the number of farm operators. The Maryland study relies on the number of farms reporting vegetables harvested for home use, as reported in the 1954 Census of Agriculture. Other allocators that have been used are number of farms (Illinois), number of farm operators (National Planning Association), and number of persons on farms (Arkansas). Persons on farms is probably the best of these, since it is the most general measure of the potential demand for farm products by farm families. In addition to being reported in the U. S. Census of Population, the number of persons on farms in Iowa is reported annually in a state publication.

Several methods have been used to allocate the gross rental value of farm dwellings. Illinois used the value of farm land and buildings, derived from the U. S. Census of Agriculture. The value of farm buildings alone would be a more closely related economic variable, but the Census of Agriculture last reported this quantity in 1940. Pennsylvania extrapolated the 1940 data to 1959 on the basis of change in the value of farm land and buildings. Estimates of the value of farm buildings for Iowa can be made by multiplying the reported value of farm land and buildings by the ratio of the assessed value of farm buildings to the assessed value of all farm real estate, obtainable from a state source.[1] Although the relation between assessed and market value varies among counties, the ratio of the assessed value of farm buildings to assessed value of all farm real estate is independent of this variation. Thus, if the Census series for the value of land and buildings is multiplied by this ratio, the result is an allocator that depends only on current data.

The variables preferred for the allocation of farm receipts to Iowa counties are summarized in Table 9.

---

[1]Iowa State Tax Commission, Annual Report (Des Moines, annual).

TABLE 9

ALLOCATORS FOR FARM RECEIPTS

Income Component	Allocator and Source[1]
Cash receipts from cattle and calves	Cattle and calves sold, dollars (CA)
Cash receipts from hogs	Hogs sold, dollars (CA)
Cash receipts from dairy products	Dairy products sold, dollars (CA)
Cash receipts from poultry and products	Poultry and products sold (CA)
Cash receipts from other livestock and products	All livestock and products sold, dollars (CA), less sum of preceding allocators
Cash receipts from crops	All crops sold, dollars (CA)
Government payments	Payments under agricultural conservation, conservation reserve (soil bank) and feed grain programs (USDA)
Value of home consumption	Number of persons living on farms (IDA)
Gross rental value of farm dwellings	Estimated value of farm buildings (CA), (ISTC)[2]

[1]See Table 2 for key to symbols for sources.

[2]Derived as follows: The value of farm land and buildings is obtained as the product of value of farm land and buildings per farm (CA) and number of farms (CA). This quantity may be multiplied by the ratio of the assessed value of farm buildings (ISTC) to the assessed value of all farm realty (ISTC) to give the value of farm buildings.

Farm Expenditures

Farm expenditures include a variety of operating expenses, depreciation on capital, interest on farm mortgage debt, and net rent to non-farm landlords. The most important classes of operating expenditures are repair and maintenance of farm capital, and purchases of feed, seed, fertilizer and lime, petroleum, and hired labor. It is not possible to associate some components of farm expenditures with reasonable allocators, and these must be grouped together as a residual for separate treatment. The items in this category are primarily those classified by the Department of Agriculture as miscellaneous operating expenses, which in 1965 made up about 12 per cent of farm expenditures in the United States, and about 8 per cent in Iowa.

It is convenient to consider farm repair, maintenance, and depreciation expenditures jointly, and to disaggregate these expenditures into those associated with farm buildings, tractors, trucks, automobiles, and other farm machinery. The only data related to these items reported in the Census of Agriculture are tractor repairs and other farm machinery repairs in 1949. Indirect indicators are available, however, for categories other than other farm machinery. Expenditures associated with farm buildings may be allocated according to the estimated value of farm buildings as described above, while those associated with tractors, trucks, and automobiles may be distributed in accordance with the number of these items on farms. Except for 1949, it seems best to group all expenses associated with other farm machinery with miscellaneous operating expenses for treatment as a residual.

The major categories of farm operating expenses other than repair and maintenance may be allocated to counties primarily on the basis of expenditures in these categories reported in the Census of Agriculture. The Census reports dollar expenditures for feed, livestock, seed, fertilizer, lime, petroleum fuel and oil, and hired labor, but with omissions for certain Census years. Dollars spent for lime is reported only for 1954. This item, which was significant in the early postwar years, has more recently fallen to minor importance. Dollars spent for fertilizer is reported only in 1954 and 1964. In all years, tons of fertilizer and lime applied are reported. These quantities can be used as allocators for the years in which expenditure data are missing. The other omissions are livestock and seed purchases in 1954. For each of these variables, a good deal of related Iowa data are available for 1949, 1954, and 1959. The desirability of using this information provides another illustration of the need for a statistical method of interpolation.

Four categories of farm expenditures remain: net rent to non-farm landlords, interest on farm mortgage debt, taxes on farm property, and the residual. The allocation of net rent to non-farm landlords was discussed in Section 2. Interest on farm mortgage debt may be allocated to counties on the basis of the value of farm land and buildings. State sources must be used to allocate taxes on farm property. In Iowa, farm real estate taxes may be estimated by multiplying the assessed value of farm land and buildings by the average net millage rate levied in rural districts. Both series are given by county in a state source.[1] Although in Iowa about 10 per cent of farm property taxes are personal property taxes--in effect, taxes on livestock--their county distribution is not readily determined, and estimated real estate taxes are taken as a suitable allocator.

One approach to the allocation of the residual, miscellaneous operating expenses and expenses associated with other farm machinery, would be to distribute them in proportion to the estimated sum of all other farm expenditures. For this purpose, estimated taxes on farm property should be excluded, since they do not reflect costs of production. To some extent, however, intercounty differences in residual costs will be reflected more accurately by differences in output. This consideration is important in view of the sensitivity of estimated net farm income to errors in estimated costs. Thus, a preferred treatment of residual costs is the average of two allocations: one, based on other expenditures, the other based on the value of farm production. The latter may be estimated as the sum of estimated farm receipts and the change in the value of farm inventories, to be discussed below.

The allocation of other farm expenditures is summarized for Iowa in Table 10. This selection of allocators is essentially a refinement of the treatment of farm expenditures in the Kentucky county income

---

[1]Iowa State Tax Commission, Annual Report.

TABLE 10

ALLOCATORS FOR FARM EXPENDITURES

Income Component[1]	Allocator and Source[2]
Feed	Feed for livestock and poultry, dollars (CA)
Livestock	Purchase of livestock and poultry, dollars (CA)
Seed	Seed purchased, dollars (CA)
Fertilizer and lime	Fertilizer and fertilizing materials, dollars (CA)[3]
Petroleum fuel and oil	Purchase of gasoline and other petroleum fuel and oil, dollars (CA)
Hired labor	Hired labor, dollars (CA)
Farm building repair and depreciation	Estimated value of farm buildings (CA), (ISTC)[4]
Tractor repair and depreciation	Tractors on farms, number (CA), (IDA)
Truck repair and depreciation	Trucks on farms, number (CA), (IDA)
Automobile repair and depreciation	Automobiles on farms, number (CA)
Other operating and depreciation expenditures	Average of estimates from two allocators: 1) sum of estimated preceding expenditures, and 2) estimated value of production.[5]
Taxes on farm property	Assessed value of farm land and buildings (ISTC) weighted by average net millage levied in rural districts (ISTC)

Table 10 (continued)

Income Component[1]	Allocator and Source[2]
Interest on farm mortgage debt	Value of farm land and buildings per farm (CA) weighted by number of farms (CA)
Net rent to non-farm landlords	Cash receipts from marketings (CA) weighted by ratio of net acres rented to acres in farms (CA), (IDA)[6]

[1]These components enter negatively in the definition of personal income.

[2]See Table 2 for key to symbols for source.

[3]Other allocators, needed for some years, include fertilizer applied, tons (IDA) and lime applied, tons (IDA).

[4]See Table 9, footnote 2.

[5]The value of production may be derived as the sum of receipts from marketings and change in the value of inventories. For the estimation of these quantities see Tables 9 and 11.

[6]See Table 6, footnote 3.

---

study. Differences in the treatment of operating expenses are that Kentucky used the value of production of five leading crops to allocate fertilizer and lime expenditures, and a single allocation was made for operating costs of motor vehicles using the sum of its relevant expenditure series in the 1950 Census of Agriculture. The residual, which includes farm mortgage interest in addition to miscellaneous operating expenses, but does not include depreciation on other farm machinery, is allocated according to the sum of all other expenses. The greatest differences are in the treatment of depreciation. Kentucky used the value of farm land and buildings as the allocator for buildings, other farm machinery, repairs for other farm machinery, and an unweighted sum of the number of trucks, tractors, and automobiles for depreciation on motor vehicles.

The allocation procedures used to allocate farm expenditures by the Pennsylvania and Arkansas studies are much less satisfactory than those of Kentucky, and although the sample is small, they may help to explain the "implausibility" of county farm income estimates obtained as a residual. Pennsylvania allocates only three categories of expenditure. The first is fertilizer and lime; the second is mortgage interest, taxes, and building depreciation, allocated as the value of land and buildings; the third, the residual, is allocated according to the sum of selected operating expenditures reported in the *Census of Agriculture*. The Arkansas study largely avoids the expenditure data provided in the *Census of Agriculture*. Feed and livestock purchases are allocated according to livestock receipts, while seed costs are allocated according to crop receipts. The allocation of repair and operation of capital items neglects the data for equipment on farms and relies on the sum of expenditures for fuel and machinery rentals. Depreciation on farm buildings is allocated according to the number of farms.

Inventory Adjustment

The difference between the receipt and expenditure items discussed above has been termed "realized net farm income" by the Department of Agriculture. In order to obtain net farm income, a measure of the income actually earned during the year, it is necessary to make an adjustment for the value of changes in farm inventories. The size of the adjustment required is often large and shows high variation from year to year. In order to obtain net farm income for the United States (48 states) for 1964, realized net farm income must be reduced 6.4 per cent. To obtain net farm income for 1965, adjustment for change in the value of farm inventories increases realized net farm income by 6.7 per cent. The relative magnitude of inventory adjustment may be expected to be greater for states than for the nation as a whole. In Iowa, the inventory adjustment increased farm income by 36.7 per cent in 1952 and by 49.7 per cent in 1957.[1]

The value of the change in farm inventories and not the change in value is a component of farm income, since here as elsewhere in the definition of personal income capital gains resulting from changes in price are excluded. The Department of Agriculture estimates farm inventory change for states and for the nation by weighting changes in physical inventories by calendar year average prices. All livestock on farms are counted as inventory, while crop inventories exclude

[1] Farm Income Situation, July, 1966, pp. 21, 43.

quantities held under CCC loan. Separate estimates of inventory change are made for six livestock items and nineteen crops.[1] Two county income studies that report the method used to estimate the value of change in farm inventories--Kentucky and Pennsylvania--allocate this quantity, whether positive or negative, in proportion to total cash receipts from crops and livestock in Census years.

A consideration of changes in farm inventories in Iowa over the postwar years shows that for that state, attention can be restricted to three livestock items, four crops, and a residual. The livestock items are cattle, hogs, and chickens; and the crops are corn, soybeans, oats, and hay. County data on the numbers of cattle, hogs, and chickens on farms as of January first are available from state sources, and changes during the year can be used to allocate changes in the corresponding components of state farm inventories. One possible allocator for the crop components would be the difference between the quantity of a crop sold during the year and the quantity produced.

Both quantities are reported for major crops in the Census of Agriculture, and since prices may be assumed to be virtually constant throughout the state, the difference between quantity sold and quantity produced would appear to be proportional to the value of inventory change. However, important quantities of each of the four major

---

[1] U. S. Department of Agriculture, Agricultural Marketing Service, Gross and Net Farm Income, vol. 3 of Major Statistical Series of the U. S. Department of Agriculture: How They are Constructed and Used (Agriculture Handbook No. 118; Washington: U. S. Government Printing Office, 1957), pp. 16-17.

Iowa crops are used on the farm as feed, and there are wide inter-county differences in amounts retained which relate to regional differences in farm specialization. An alternative approach would follow from the assumption that change in crop inventories in a given year is proportional to the change in production from the preceding year. When production increases for the state as a whole and inventories for the state also increase, this assumption implies that the increase in inventories should be distributed to counties in proportion to their increase in production. Conversely, when both inventories and production decrease, the amount of the decrease becomes the basis for allocation to counties. It is possible, however, that crop inventories would decrease when production increases, or that the inventories would increase when production decreases. In these cases the change in production does not provide a reasonable basis for allocating inventory change. Current year production should be the allocator in these cases, since it provides a measure of the importance of the crop in the county.

The problem remains of how to deal with the change in the value of inventories for those livestock and crop items which are not large enough to merit separate treatment. The sum of estimated inventory change for the separately treated items is suggested as an allocator. For years in which this sum and the residual change in the same direction, the estimated change in the value of farm inventories in each county is increased proportionately, while a change in the opposite direction leads to a scaling down in absolute value of the estimated change in farm inventories. It does not seem unreasonable that in those years in which inventories of most types are increased or decreased, this adjustment would be general across counties but that in those years when inventory changes were mixed, farmers in a particular county would increase inventories of some items while reducing them for others.

The recommended allocators for the components of farm inventory adjustment are shown in Table 11.

5. Transfer Payments, Other Labor Income, and Contributions for Social Insurance

Either because the data are good (transfer payments) or poor (other labor income and contributions for social insurance), there is less to be said about their estimation than about other categories of personal income. For this reason, these components are grouped together in a single section. These categories of income, although smaller, nevertheless merit careful attention. Transfer payments, nationally the fourth largest income category, were 7.5 per cent of U. S. income in 1965, and 7.3 per cent of the personal income of Iowa.

Transfer Payments

In terms of reliability, the estimation of transfer payments of personal income by county is one of the most satisfactory of the major components of income. This is true particularly because of the rapid increase, over recent years, in old-age and survivors insurance benefits, which by 1959 made up two-fifths of the national total. Transfer payments to persons are made by the federal government,

TABLE 11

ALLOCATORS FOR VALUE OF CHANGE IN FARM INVENTORIES

Income Component	Allocator and Source[1]
Cattle	Change during year in number of cattle on farms (ICLR)[2]
Hogs	Change during year in number of hogs on farms (ICLR)[2]
Chickens	Change during year in number of hens on farms (IDA)[2]
Corn	Change from preceding year in corn harvested for grain, bushels (IDA)[3]
Soybeans	Change from preceding year in soybeans harvested for beans, bushels (IDA)[3]
Oats	Change from preceding year in oats production, bushels (IDA)[3]
Hay	Change from preceding year in hay harvested, acres (IDA)[3]
Other livestock and crops	Sum of the preceding components of the value of change in farm inventories

[1]See Table 2 for key to symbols for source.

[2]When this change differs in sign from change in value of inventories, average number on farms during the year should be the allocator.

[3]When this change differs in sign from change in value of inventories, current year production should be the allocator.

state and local government, and businesses. The allocation of transfer payments to counties presents few conceptual difficulties, although the number of components making up this income category is especially large. The reader may wish to turn at once to Table 12, which lists these components at the level of disaggregation that will concern us in this section.

Most transfer payments from the federal government consist of benefits from social insurance funds and payments to veterans. Old-age and survivors insurance benefits may be allocated to counties on the basis of the dollar amount of monthly benefits in current payment status, which has been tabulated annually by the Social Security Administration except for 1950. Lump sum death benefits are not tabulated by county but are of much smaller magnitude. Benefits under the state unemployment insurance program may be distributed to counties on the basis of county tabulations of benefits received by state unemployment security agencies. In states where these data are not available for 1950, unemployment, as given by the Census of Population, may be used as a substitute. The other major components of benefits from social insurance funds are unemployment and retirement benefits paid to railroad and federal government civilian employees. No county data are available for these components, and they must be allocated according to the number of employees in the respective industries.

There are no tabulations by county of the amounts paid to veterans, but these amounts may be allocated to counties according to the number

TABLE 12

ALLOCATORS FOR TRANSFER PAYMENTS

Income Component	Allocator and Source[1]
Old-age and survivors insurance benefits	Dollar amount of monthly OASI benefits in current payment status (HEW)[2]
State unemployment insurance benefits	State unemployment insurance benefits paid (IESC)
Railroad unemployment and retirement benefits	Number of railroad employees (CP)
Federal civilian unemployment benefits and pensions	Number of federal civilian employees (CP)
Government life insurance benefits	Number of World War II and Korean War veterans (CP), (VA)
Veterans pensions and compensation	1950 and earlier: number of WW II veterans (VA), 1951 and later: number of Korean War veterans (CP), (VA)
Other payments to veterans and military retirement	Number of World War II veterans (CP), (VA)
State and local government pensions	Estimated state and local government wages and salaries[3]
Direct relief	Benefits paid under state social security programs, average of fiscal years (IDSW)
Other government transfers	Population (CP)
Business transfers	Average of estimates from two alloca-allocators: 1) estimated private sector wages and salaries, and 2) retail sales (CR)[4]

[1]See Table 2 for key to symbols for sources.

[2]For 1952 and later, year-end to year-end averages.

[3]See Table 5.

[4]See Table 4.

of veterans in the categories primarily receiving payments of the various types. Government life insurance benefits, an important component of personal income in the early 1950's, are best allocated on the basis of the number of World War II and Korean War veterans. After 1950, veterans pensions and compensation were paid primarily to Korean War veterans. Hence the number of Korean War veterans should be used to allocate this component from 1951 onward, and the number of World War II veterans should be used for the earlier period. Other payments to veterans and military retirement should be allocated according to the number of World War II veterans. Although military retirement might be broken out and allocated in part on the basis of the number of World War I veterans, this component is very small, and reliable data for World War I veterans are available only for 1960. Other payments to veterans were an important component of personal income only in the late 1940's, when they included terminal leave pay, veterans readjustment allowances, and similar items.[1] By 1965, Viet Nam War veterans were not yet an important group.

The distribution of veterans by county is reported in the Census of Population for 1960 but not for 1950. However, the Veterans Administration estimated the number of veterans on the basis of a one per cent sample of the recipients of the first Government Life Insurance

---

[1]For a discussion of transfer payments to veterans, see Charles F. Schwartz and Robert E. Graham, Jr., Personal Income by States Since 1929: A Supplement to The Survey of Current Business. U. S. Department of Commerce, Office of Business Economics (Washington: U. S. Government Printing Office, 1956), pp. 132-134.

dividend, paid in June 1950, and this estimate may be used for the earlier year. The Veterans Administration also tabulated the county distribution of Korean War veterans on the basis of address immediately after discharge, and these tabulations exist for 1953, 1955, and 1958. Although no adjustments are made in the data for subsequent change in address, changes in the county of residence of Korean War veterans were probably much smaller over this period than the number of new veterans leaving the service.

Several small components of federal transfer payments remain, chief among them federal payments to non-profit organizations and federal scholarship payments, which are not sufficiently large, at least in Iowa, to merit individual treatment. These components may be allocated according to civilian population.

The preferred allocators for the components of federal transfer payments, summarized in Table 12, may be compared with those selected in other county income studies. One difference among various past studies is the number of components of transfer payments which are distributed to counties according to population or to the size of some segment of the population. Pennsylvania uses population to allocate federal civilian pensions. Arkansas uses white male population 65 years of age and over to allocate railroad unemployment and retirement benefits, federal civilian pensions, government life insurance benefits, state and local government pensions, and other transfer payments from federal, state, and local governments. These alternatives are less satisfactory than those indicated above because (1) the geographic distribution of retired government and railroad employees is probably closer to the distribution of present employees than to the distribution of population or to the aged male population in general; (2) railroad unemployment benefits are not received by persons over 65; (3) most government life insurance benefits are received by veterans or their survivors under 65 years of age; and (4) "other" federal transfer payments do not accrue primarily to older persons.

Another source of variation in the treatment of federal transfer payments occurs in the allocation of payments other than those associated with social insurance funds or veterans. In the Oklahoma study "other" transfer payments from governments, federal, state, and local, are allocated according to monthly OASI benefits. Illinois allocates federal payments to non-profit organizations to counties in proportion to estimated total wages and salaries in education. Since "other" federal transfer payments consist primarily of federal grants for hospital construction, grants to private colleges, and scholarships,[1] these choices seem to be poorly focused.

Finally, there are a number of differences in the treatment of payments to veterans from that suggested above. Most of these differences arise from greater disaggregation of the components and a less careful appraisal of the principal recipients of each type of benefit. Illinois, Oklahoma, Pennsylvania, and the National Planning

[1] *Ibid.*, p. 134.

Association include World War I veterans in the allocator for government life insurance benefits; and Maryland also includes military personnel stationed in the state, although the share of benefits received by both groups are small compared to their size.

It remains to consider transfer payments originating in state and local government and in the business sector. The largest components of state and local government transfers are employees' pensions and direct relief payments. Pensions may be allocated to counties in proportion to state and local government wages and salaries, as estimated previously. Direct relief payments should be allocated according to benefits paid under the state public assistance programs which are eligible for federal support under the Social Security Act. These programs, which make up most of direct relief, are old-age assistance, medical assistance for the aged, aid to families with dependent children, aid to the blind, and aid to the permanently and totally disabled. Total payments under these programs exceed direct relief payments, however, because they include payments to vendors as well as payments to persons. Other state and local government payments which are quite small, may be allocated according to population. An exception is that in the early postwar period, some states provided bonuses and aid to veterans, which should be allocated according to the number of World War II veterans. Business transfer payments include a number of miscellaneous items, including corporate gifts to non-profit organizations and consumer bad debts, with regard to which existing county data are not very relevant. One means of handling this component would be to average allocations based on wages and salaries, as estimated previously, and retail sales.

Only minor variations in the treatment of state and local government and business transfers occur in the surveyed studies. Illinois was able to utilize some direct data on state and local government pensions while Oklahoma allocated other state and local government transfers according to public assistance payments. Single allocators were used in all but one of the studies to allocate business transfers, and included estimated wages and salaries, estimated wages and salaries plus non-farm proprietors' income, retail sales, and population. The National Planning Association used retail sales to allocate consumer bad debts and estimated wages and salaries and non-farm proprietors' income to allocate the remainder.

Other Labor Income

Other labor income is the name given to a heterogeneous group of income components whose combined magnitude has been until recent years quite small. One component, employer contributions to pension and welfare funds, has, however, grown rapidly; and in 1965 other labor income was 3.5 per cent of personal income of the United States and 2.7 per cent of that of Iowa. In addition to employer contributions to pension and welfare funds, other labor income includes compensation to employees for industrial injuries and pay of military reservists. Fees received by directors of corporations, supplemental unemployment benefits, and a handful of minute components, such as fees received by justices of the peace, contribute a small residual.

In spite of the fact that other labor income is now an important category of personal income, virtually no county data exist which indicate its magnitude, either directly or indirectly. Hence county income studies must adopt allocators which are more appropriate as indicators of other components of income. The most common practice is to allocate all components of other labor income in proportion to its share of estimated wages and salaries; Arkansas, Oklahoma, Pennsylvania, Kansas, and the National Planning Association take this approach. Illinois allocates employer contribution to private pension and welfare funds according to wages and salaries in mining, manufacturing, and public utilities only, while Kentucky allocates compensation for injuries by weighting wages and salaries by industry in proportion to injury benefits paid in the state. Illinois uses allocators other than wages and salaries in several instances. Compensation for injuries is allocated by a weighted sum of industrial deaths and other work injuries; pay of military reservists, by male population in selected age groups; and the remainder, by population. Kentucky allocates pay of military reservists by the number of veterans.

With the data situation so unsatisfactory, none of these alternatives can be considered unreasonable. Pay of military reservists is probably best allocated by number of veterans in the early postwar years, but from about 1950 onward, male population of military age is probably the better allocator. The residual components of other labor income are probably associated more closely with wages and salaries than with population. Iowa data permit the allocation of compensation for work injuries by the number of injuries since 1959; and before that date, a weighted sum of wages and salaries by industry, the weights based on state totals of injury benefits by industry, may be used.[1] However, rather than weight wages and salaries by industry in proportion to an industry's share of injury compensation payments, it seems preferable to weight them by the state ratio of industry injury payments to wages and salaries. The resulting index would provide an exact measure of injury compensation by county if the ratio of compensation payments to wages in each industry were the same in all counties.

A more important refinement in the treatment of other labor income follows from the fact that a similar index can be constructed as an allocator for employer contributions to private pension and welfare funds. The weights, however, must be derived from national data. Contributions by corporations to private pension and welfare funds are reported on federal tax returns, and are tabulated by industry in Statistics of Income annually except for 1951.[2] Weights may be formed

---

[1]Iowa Bureau of Labor, Statistical Department, Iowa Work Injuries (Des Moines, quarterly).

[2]U. S. Treasury Department, Internal Revenue Service, Statistics of Income: Corporation Income Tax Returns (Washington: U. S. Government Printing Office, annual).

by dividing these values by industry wages and salaries, reported in the annual national income issues of the Survey of Current Business.

The preferred allocators for components of other labor income are summarized in Table 13.

Contributions for Social Insurance

Just as benefits received from social insurance funds are treated as components of personal income, contributions by individuals to social

TABLE 13

ALLOCATORS FOR OTHER LABOR INCOME

Income Component	Allocator and Source[1]
Employer contributions to private pension and welfare funds	Weighted sum: Estimated wages and salaries by industry times ratio of employer contributions to wages and salaries, for the U.S. (TD), (SCB)[2]
Compensation for injuries	Number of work injuries (IBL)[3]
Pay of military reservists	Male population of military age (CP)[4]
Residual components	Estimated wages and salaries[5]

[1]See Table 2 for key to symbols for source.

[2]See Table 4.

[3]For years prior to 1959, the allocator should be a weighted sum: estimated wages and salaries by industry times ratio of compensation for injuries (IBL) to wages and salaries, for the state. See Table 4.

[4]For 1950 and earlier, the allocator should be number of World War II veterans (VA).

[5]See Tables 4 and 5.

insurance funds are counted as deductions from personal income. Hence, to arrive at estimates of personal income by county, these contributions must also be allocated. Nationally, personal contributions to social insurance funds were equal to 2.5 per cent of personal income in 1965, while in Iowa these contributions were equal to 2.2 per cent of personal income. There are virtually no county data on the components of this category, and the allocations must be made on the basis of estimates of other income components. However, contributions to social insurance funds are closely related to wage and salary income and to proprietors' income, so that contributions to social insurance funds may be estimated almost as reliably as are these income components.

Six categories of contributions to social insurance funds may be distinguished. These are self-employed persons' OASI contributions, premiums for government life insurance, and four types of contributions by employees: OASI, railroad retirement, federal civilian government retirement, and state and local government retirement. The contributions by employees to the various funds should be allocated to counties in proportion to wages and salaries in covered industries, except that, since federal employees are excluded from OASI coverage if they are covered by the federal civilian retirement program, the OASI and federal civilian components should be combined for purposes of allocation. The wage and salary total that should be used in the allocation of the OASI-federal civilian employees' contributions varies in industrial coverage from year to year in accordance with amendments to the Social Security Act. Railroads should be excluded in all years, the military before 1957, and farms, domestic services, and non-profit organizations should be excluded before 1952.[1] The extent of OASI coverage of state and local government employees varies by state, and wages and salaries in this sector should be weighted to reflect the extent of coverage. Data suitable for this purpose are the estimates of the extent of coverage prepared regularly by the Bureau of Old-Age and Survivors Insurance.[2] Coverage of state and local government employees in Iowa, however, is virtually complete.

Amendments to the Social Security Act also indicate a varied treatment of self-employed persons' OASI contributions, which began in 1952. For the years 1952-55, the allocator should be non-farm proprietors' income, and from 1956 onward all proprietors' income should be used. Since self-employed physicians were not covered until 1966, a refinement to exclude them would be appropriate, but not justified by the magnitudes involved. The largest share of government life insurance payments are made by World War II and Korean War veterans, so their number provides a suitable allocator for this component.

The preferred allocators are shown in Table 14. Most of the county income studies surveyed used less disaggregation than the procedures selected. Kentucky, Arkansas, and Pennsylvania each used a

---

[1]Charles I. Schottland, The Social Security Program in the United States, pp. 42-48.

[2]Department of Health, Education and Welfare, Social Security Administration, Bureau of Old-Age and Survivors Insurance, State and Local Government Employment Covered by OASI (Washington: U. S. Government Printing Office, quarterly).

TABLE 14

ALLOCATORS FOR CONTRIBUTIONS TO SOCIAL INSURANCE FUNDS

Income Component	Allocator and Source[1]
OASI employees	Estimated wages and salaries except railroads and federal civilian government[2]
OASI self-employed persons	1951-1954: Estimated non-farm proprietors' income. 1955 and later: Estimated farm and non-farm proprietors' income.[3]
Unemployment insurance, except railroad and government	Estimated wages and salaries except railroads and government[4]
Railroad retirement and unemployment insurance	Estimated railroad wages and salaries[3]
Federal civilian retirement and unemployment insurance	Estimated federal civilian wages and salaries[5]
State and local government retirement	Estimated state and local government wages and salaries[5]
Government life insurance	Number of World War II and Korean War Veterans (VA), (CP)

[1]See Table 2 for key to symbols for source.

[2]See Tables 4 and 5.

[3]See Tables 7, 8, 9, 10, and 11.

[4]See Table 4.

[5]See Table 5.

single allocator for contributions to social insurance funds, while Kansas provided separate treatment only for OASI contributions by self-employed persons. More complete disaggregation would appear worth-while because of the variations in coverage noted above and because, since no new data collection is involved, the cost of disaggregation is low. Detailed allocations for these contributions are made only by the Illinois and National Planning Association county income studies, and the procedures of these studies are similar to those given above. Two errors made by the National Planning Association are the omission of government sector wages and salaries from the allocator for OASI contributions by employees and the omission of farm proprietors' income from the allocator for OASI contributions by self-employed persons. Both studies make separate allocations for OASI employees and federal civilian employees. Illinois gives a more detailed treatment contribution by employees of state and local government, and the National Planning Association gives a more detailed treatment for government life insurance.

6. Synthesis and Summary

In the preceding pages, the selection of county allocators for components of personal income has been considered in detail. For a majority of the components of personal income, the allocator which was chosen as most satisfactory differed from the typical choices in recent estimates of county personal income. If the components of personal income are weighted according to their magnitude, the proportion of personal income for which significant innovations have been suggested is quite high. Clearly there is considerable room for improvement in the selection of county allocators.

Yet to be considered are the implications of this newly selected group of allocators for the detail and frequency with which meaningful sets of personal income accounts can be prepared. There would appear to be opportunity for substitution between detail by component and frequency in the reporting of personal income. Meaningful and useful estimates of total personal income might be reportable at short intervals, say annually, if no significance were attached to the estimates of the underlying components. The greater frequency of reporting for total personal income would be justified if errors in the components cancelled, at least to some extent, so that the error in personal income (reported) was smaller than the errors in income components (not reported). While the conjecture of offsetting errors is plausible, the probability mechanism insuring randomness is not completely specified. This question is most serious when the more frequent county income estimates take the form of current estimates based largely on extrapolations of recent trends. A further difficulty in reliance on offsetting errors is that, in preparing county income estimates for additional years, primary interest attaches not to the magnitude of personal income, but to the change that occurred over the time interval separating the two estimates. The errors in income components must be offsetting to such a degree that the change in income, and not just its absolute magnitude, is a meaningful statistic.

The question of frequency and detail can also be discussed by considering, in turn, the maximum amount of detail by component with which county income estimates can be provided for any year, and the frequency with which maximum detail can be obtained. The maximum level of detail for county income estimates may be compared to the level of detail provided in the state personal income accounts of the Department of Commerce. Totals only are reported for property income, non-farm proprietors' income, farm proprietors' income, transfer payments, other labor income, and contributions to social insurance. Enough data are available at the county level to make separate estimates of each of these quantities. With regard to wages and salaries, however, county wage and salary data do not permit the high degree of disaggregation of the state estimates. The maximum industrial detail at the county level would appear to consist of the nine major industries of County Business Patterns, plus farming, railroads, federal civilian, military, and state and local government. The small magnitude of wages and salaries in some of these industries suggests that they might be combined with others. The following composite industries might be adopted: (1) farming, agricultural services, forestry, fisheries, and mining; (2) railroads and other transportation and public utilities; and (3) federal civilian government and the military. If all of these aggregates were employed, wages and salaries would be reported in ten component detail.

As we have seen, there is no year in which county allocators are available for all components of personal income. In choosing any year for county income estimation, it is not sufficient to attempt to find a year for which many allocators are available. Instead, one must look for a year around which allocators tend to group. In this way, years can be chosen such that measures of most components of personal income are drawn either from the same year or a year that is close. A further consideration in the choice of a set of years for county income estimates is that a set of income accounts is more convenient for analytical purposes if the estimates are made at regular intervals.

Tables 15, 16, 17, and 18 show the years for which each of the preferred allocators selected in preceding sections are available. Only the allocators for net farm income and contributions to social insurance funds are omitted. Similar tables could be constructed for the frequencies of county income allocators for other states. The tables provide a visual summary which can be extremely helpful in evaluating the possibility of county income estimation in each of the years 1947-1965, although it must be remembered that some allocators apply to larger income components than do others. An examination of Table 15, which covers wage and salary allocators, indicates that for the most part, County Business Patterns data are available at three-year intervals beginning in 1947. The only exception in the appearance of the publication in 1951 rather than 1950. An alternative set of years suggested by Table 15 is that obtained by taking five-year intervals from 1948. The advantage of these years is that in all cases both County Business Patterns and all industrial censuses appear either for the designated year or the year immediately

TABLE 15

FREQUENCY TABLEAU FOR WAGE AND SALARY ALLOCATORS

Allocator and Source[1]	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
First quarter payrolls (CBP)	x	x			x		x			x			x			x		x	x
Farm wages (CA)			x					x					x					x	
Manufacturing payrolls (CM)	x							x				x					x		
Wholesale trade payrolls (CW)		x						x				x					x		
Retail trade payrolls (CR)		x						x				x					x		
Railroad employment (CP)				x										x					
Domestic services employment (CP)				x										x					
Federal civilian employment (JCRN)				x										x					
Military personnel (CP), (DD)[2]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
State government payrolls (SBSI)[3]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Local government payrolls (CB), (IESC)	x							x	x		x					x			

[1]See Table 2 for key to symbols for source.

[2]Collected by the author for 1950-1962 only.

[3]Tabulated only for fiscal 1947, 1949, 1954, 1959, and 1962.

TABLE 16

FREQUENCY TABLEAU FOR PROPERTY INCOME ALLOCATORS

Allocator and Source[1]	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
Assessed value of non-farm residential property (ISTC)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Assessment ratios (ISTC), (TSC)							x									x	x	x	x
Number of establishments with 1-3 employees (CBP)	x	x			x		x			x			x			x		x	x
Cash receipts from farm marketings (CA)			x					x					x					x	
Share of farmland rented, (CA), (IDA)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Size distribution of income (CP), (ISTC)	x		x										x				x		
Share of dividend (interest) income reported by income size class (TD)					x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Demand deposits at federal reserve member banks (FRB)	x		x	x		x		x		x		x		x		x		x	

[1]See Table 2 for key to symbols for source.

TABLE 17

FREQUENCY TABLEAU OF ALLOCATORS FOR NON-FARM PROPRIETORS' INCOME

Allocator and Source[1]	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
Number of establishments and number of employees, by industry, first quarter payrolls (CBP)	x	x			x		x			x			x			x		x	x
Number of proprietors: retail trade (CR), wholesale trade (CW), business services (CS)		x						x				x					x		
Adjusted gross income of business proprietors, selected industries (ISTC)																	x		
Adjusted gross income of professional workers (ISTC)																	x		
Annual payrolls, business services (CS)		x						x				x					x		
Number of physicians (HEW)[2]				x									x			x		x	
Number of lawyers (BA)[2]				x							x		x						

[1]See Table 2 for key to symbols for source.

[2]Can be tabulated annually from professional directories.

TABLE 18

FREQUENCY TABLEAU FOR TRANSFER PAYMENTS AND OTHER LABOR INCOME ALLOCATORS

Allocator and Source[1]	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
Monthly OASI benefits in current payment status (HEW)	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
State unemployment insurance benefits paid (IESC)			x					x	x	x			x				x		
Number of railroad employees (CP)				x										x					
Number of federal civilian employees (JCRN)				x										x					
Number of veterans (CP), (VA)				x			x		x			x		x					
Benefits paid under state social security program (IDSW)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Population, and population of military age (CP)				x										x					
Number of work injuries													x	x	x	x	x	x	x

[1]See Table 2 for key to symbols for source.

preceding and following. The only exception, the absence of the Census of Governments in the early postwar period, can be remedied with unpublished data available at the same frequency. The presence of three components for which allocators exist only decennially does not weigh heavily against the choice of five-year intervals.

Table 17, which gives the observation frequencies for allocators for non-farm proprietors' income, shows a similar picture, in part because the same data sources are important for both categories of income. The principal differences are the greater importance of the state personal income tax data, which exist only for 1963, and the irregularity of other data related to the distribution of professional income. A somewhat different data situation is revealed in Table 16, which refers to allocators for property income. The allocators for monetary rent from business property and imputed rent appear at intervals which never exceed three years. The three remaining components of property income, however, are each allocated by combining two series, one of which is observed frequently and the other observed infrequently. The more frequently observed series should be given only moderate weight in appraising the frequency with which these property income components can be estimated.

The sharpest contrasts in data availability are seen in Table 18, which gives the frequency of observation for the transfer payments and other labor income allocators. The allocators for two of the largest components are available annually, while the allocators for other components are available decennially or irregularly. The data situations for allocators not shown in the tables are readily summarized. Sufficient data for the estimation of farm income by the allocation method exist only for the years 1949, 1954, 1959, and 1964, although related data, available for interpolation, appear annually. Employer contributions to private pension and welfare funds and employee contributions to social insurance funds are associated primarily with allocators for other income components, especially wages and salaries, whose frequency of observation has been noted already.

It may be concluded from this analysis that the estimation of county personal income by major component is both feasible and meaningful in view of the existing data at five-year intervals over the postwar period, and that the years 1948, 1953, 1958, and 1963 are years particularly favorable for estimation. Some increase in frequency, perhaps to three-year intervals, could be adopted without impairing the quality of the county income estimates seriously, although the reduction in the reliability of estimated changes in income would be substantial.

The possibility of frequent estimates of county personal income in Iowa for the years since 1963 is somewhat brighter. The Bureau of the Census plans to continue the publication of County Business Patterns on an annual basis, thus providing a regular source of wage and salary data. The Iowa State Tax Commission intends to prepare magnetic tapes containing selected information from state personal income tax returns, and these could provide a continuing source of information for the allocation of wages and salaries of government employees, income of non-farm proprietors, and the size distribution of income. The annual data on the farm sector and major components of transfer payments may be added to this list. These considerations suggest that in the future the problem of frequency of observation of county income allocators should largely disappear, although there will still be lags in availability.

An important result of the analysis presented in this chapter is the greater precision with which the need to supplement allocation methods with more powerful techniques can be defined. Situs adjustments are needed for many of the components of personal income. In the estimation of wages and salaries only a few industrial components do not require adjustment for residence: farm wages, where intercounty commuting may be assumed low; state government; and the several industries where wages and salaries are allocated on the basis of employment. All the components of non-farm proprietors' income require situs adjustment, except where 1963 state income tax data are used, as do most components of other labor income and contributions to social insurance funds. In addition, situs adjustments should be made, if possible, for the monetary components of rental income, imputed interest, and business transfer payments.

A common aspect of almost all these components of personal income is that the allocators which have been selected for their estimation by county may be expected to differ in their geographic distribution from the components they are intended to measure in a way closely related to intercounty commuting. The relationship is most immediate in the cases of non-farm wages and salaries and proprietors' income, and is only slightly less direct for other labor income and contributions to social insurance funds, where many of the same allocators are used. But the connection should also be expected to exist for those components of property income and transfer payments for which the allocators used are bank deposits and retail sales, since a household can reduce transportation costs by choosing the same community for banking and shopping as for employment. Thus, while the adjustments that must be made on county income estimates to place them on a "where received" basis are extensive, the task is greatly simplified, conceptually, if all of these adjustments are related to commuting.

Two other respects in which allocation methods are insufficient to yield reliable income estimates by county require developments in statistical theory. First, in order to use the OASI and industrial census payroll data for the estimation of wages and salaries, a means must be found to supply missing values for small counties satisfactorily. The same allocators reappear in the estimation of business transfers payments and of some components of other labor income and contributions to social insurance funds. The same problem in another guise--the estimation of employment in small counties--must be resolved in order to implement the approach to situs adjustment developed below. The second problem that requires statistical analysis is that of improving the reliability and time-focus of county income estimates by using related series to interpolate the most reliable allocators. Such an interpolation procedure is needed for the estimation of manufacturing and wholesale and retail trade wages and salaries, and particularly, for the estimation of farm proprietors' income. It is to this last problem that we now turn.

CHAPTER THREE

MATHEMATICAL METHODS FOR COUNTY INCOME ESTIMATION I: INTERPOLATION OF TIME SERIES-CROSS SECTION DATA USING RELATED SERIES

One of the most pervasive problems in the statistical adjustment of data, both in county income estimation and other areas of economic statistics, is the estimation of a quantity for a particular year when observations for that quantity are available only for some other year or years. Sometimes the statistician's only alternative is to use the data for the years they are available; other times the standard formula for arithmetic or geometric interpolation can be used to center the data on the desired years. A much more satisfactory situation exists when there are data for variables related to the variable of interest both for the years that it is available and the years for which an estimate is desired. The problem is then to choose an appropriate way to utilize the data for the related variables.

Economists concerned with the production of economic statistics have usually approached the problem of utilizing related data by choosing, as an estimate of the desired variable, an *ad hoc* function of the observed values of desired and related variables. Specifically, the attempt is made to use the related variables to adjust, in some plausible way, the results obtained by simple arithmetic or geometric interpolation. The absence of statistical theory and methods in this work, either in choosing or evaluating the function used as an estimator is somewhat surprising, since the problem of estimating the values of desired variables is so clearly a problem in statistical forecasting, a familiar topic to most economists. Nevertheless, the alternative approach of specifying and estimating a linear statistical model, and using this model to generate the desired values, appears to have been little used in the construction of economic statistics.

In this chapter we take up the question of making estimates of quantities for desired dates in the context that is most important for county income estimation. We assume that both time series and cross section observations exist for the relevant variables; however, the time series is short, while the number of elements in the cross section (the counties) is fairly large. In addition to data for which an estimate is to be made, there are assumed to be time series-cross section data for a number of related variables that might be utilized in making the estimate. The outline of the chapter is as follows: we first consider previous work dealing with methods for the use of related series in interpolation; second, we introduce a simplified version of the time series-cross section model that will be recommended for this purpose; third, we derive some results that are useful in evaluating the estimates generated by the model; and finally, we present some generalizations of the model which are needed in applications.

1. Current Methods of Interpolation

Milton Friedman has discussed the nonstatistical methods for using a related variable x_k to interpolate a variable y_t that dominate research practice, and has suggested a statistical method of interpolation.[1] Consider the simplest case, in which x_t is observed at the equally spaced dates $t = 0, 1, 2$, while y_t, the variable to be interpolated, is observed only at $t = 0$ and $t = 2$. The non-statistical methods all assume some exact relation between the deviation of x_t from its trend at $t = 1$ and the deviation of y_t from its trend at $t = 1$. On this assumption, the movement in x_t except for trend is assigned to y_t; the values of x_t are used to make an adjustment on the trend of y_t. Different researchers have made alternative assumptions about the relation between the deviations of x_t and y_t from their respective trends--for example that the deviations were of equal magnitude, or that the ratios of the deviations to trend values were equal.[2] The trend value at $t = 1$ has been computed as the arithmetic mean of the values at $t = 0$ and $t = 2$ by some, and as the geometric mean of these

---

[1]Milton Friedman, "The Interpolation of Time Series by Related Series," Journal of the American Statistical Association, 57 (December, 1962), 729-57. Our notation x_t, y_t reverses Friedman's usage.

[2]Three examples cited by Friedman of non-statistical interpolation by means of related series occur in: Simon Kuznets, National Income and its Composition, 1919-1958, II (New York: National Bureau of Economic Research, 1941); U. S. Department of Commerce, Office of Business Economics, National Income, 1954 Edition, A Supplement; and Allyn A. Young, "An Analysis of Bank Statistics for the United States, Part IV, The National Banks," Review of Economic Statistics, IX (July 1927), 121-41.

values by others. Friedman shows that all the variants of the non-statistical method have a common defect: unless the correlation between the deviations from trend for the two variables is unity, too much of the movement in x_t is imputed to y_t. The non-statistical procedures cannot be optimal in the sense of minimizing the mean square error of the estimates y_t.

Friedman's suggestion turns on the well-known property of least-squares regression that (under suitable conditions) it leads to predicted values of the dependent variable that have minimum mean square error of forecast. He argues that it would be desirable to know the regression coefficient obtained by regressing whatever measure of the deviation from trend for y_t is felt to be appropriate against the corresponding measure for the deviation involving x_t. To estimate y_1, the predicted value of the deviation of y_1 from its trend would be used to adjust the trend value computed from y_0 and y_2. For example, Friedman would replace the procedure in which the deviations from trend of x_t and y_t were assumed to be equal and trend was computed as an arithmetic mean with the following procedure: first, find the least-squares regression coefficient b for the relation

$$y_{t-1} - \frac{1}{2}(y_t + y_{t-2}) = b\ [x_{t-1} - \frac{1}{2}(x_t + x_{t-2})] + w'_t,$$

then estimate y_1 as

$$\text{(1)} \qquad \hat{y}_1 = \frac{1}{2}(y_0 + y_2) + \hat{b}[x_1 - \frac{1}{2}(x_0 + x_2)].$$

A practical difficulty with this approach is that except in the treatment of isolated missing values, y_t will be a variable observed less frequently than x_t, and thus the regression cannot be carried out in the form specified. Friedman contends that it is frequently possible to find a substitute regression involving other data or other time periods that can be used to derive a value for b, and that use of such a value would be preferable to either disregarding the related series or using it in one of the non-statistical ways. The regression method has been used by Friedman to interpolate historical monetary and banking statistics.[1]

Friedman's method may be compared with another statistical method which is occasionally used for interpolation: simply run the regression of y_t on x_t and use the predicted values of y_t in place of the missing values. This method gives

$$y_1^* = a^* + b^* x_1$$

as the predicted value of y_1. It is clear that the methods give different results, since Friedman's prediction involves a linear combination of x_0, x_2, y_0, and y_2, while the alternative does not. Both $\hat{y}_1$ and y_1^* are least-squares estimates, but both cannot have minimum variance. Which model is appropriate depends on the structure of the disturbances in the two cases.

---

[1]Milton Friedman and Anna J. Schwartz, A Monetary History of the United States (National Bureau of Economic Research, Studies in Business Cycles, No. 12, Princeton: Princeton University Press, 1963).

We know that if the disturbance u'_t in

(2) $$y_t = a + bx_t + u'_t$$

is independently distributed with zero mean and constant variance, the estimate y^*_t should be preferred. Friedman does not raise the question of the distribution of the disturbances explicitly, but it is interesting to ask the following question: Suppose that y_t and x_t are connected by the linear relation given by equation (2); for what distribution of the disturbances would Friedman's estimate $\hat{y}_1$ be optimal? If we form second differences of both sides of equation (2), we obtain

(3) $$y_{t-1} - \frac{1}{2}(y_t + y_{t-2}) = b[x_{t-1} - \frac{1}{2}(x_t + x_{t-2})]$$
$$= [u'_{t-1} - \frac{1}{2}(u'_t + u'_{t-2})],$$

which involves the same functions of x_t and y_t as used by Friedman. Least-squares estimation of equation (3) will be optimal if the disturbance

$$w'_t = u'_{t-1} - \frac{1}{2}(u'_t + u'_{t-2})$$

is independently distributed with zero mean and constant variance. Rearranging terms, we see that if these conditions hold for w'_t, u'_t is generated by the second-order autoregressive process

$$u'_t = 2u'_{t-1} - u'_{t-2} - 2w'_t$$

Thus the two statistical interpolation methods may be contrasted in terms of the pattern of disturbances that is assumed for the regression equation. One method assumes that these disturbances satisfy the conditions of the classical linear model, while Friedman's method assumes a second-order autoregressive process with the rather unusual coefficients 2, -1 chosen _a priori_.

Another way in which these two interpolation methods may be contrasted can be seen as follows. Rearranging terms in equation (1), Friedman's forecasted y_1 becomes

$$\hat{y}_1 = \hat{b}x_1 + \frac{1}{2}[(y_0 - \hat{b}x_0) + (y_2 - \hat{b}x_2)],$$

which has as its expected value

$$E(\hat{y}_1) = bx + \frac{1}{2}[(y_0 - bx_0) + (y_2 - bx_2)].$$

Now assuming equation (2) holds, we may write

$$E(\hat{y}_1) = bx + \frac{1}{2}(a + u_0' + a + u_2') = a + bx + \frac{1}{2}(u_0' + u_2').$$

An implication of this result is that if equation (2) with some pattern of disturbances is taken to be the basic model, then Friedman's procedure differs from the simple regression model not only in leading to different estimates of the parameter b (and implicitly, of a), but in including a term involving the disturbances u_0' and u_2'. More precisely, Friedman's method implicitly involves the average of the computed residuals corresponding to u_0' and u_2', which functions as a proxy for the noncomputable residual at $t = 1$. Nevertheless, because of the pattern of disturbances which Friedman implicitly assumes, this average may not be a "good" estimate of the missing residual.

On the basis of this discussion it is natural to ask whether less restrictive assumptions can be made about the disturbances when regression methods are considered in interpolation tasks, and whether, in view of the missing observations, these methods can be implemented computationally.

2. A Time Series-Cross Section Model for Interpolation

The model which we shall consider now is somewhat more general than those considered in the preceding paragraphs. We still assume three equally spaced observation dates $t = 0, 1, 2$ with the dependent variable missing for $t = 1$, but now we assume that for each value of t we have observations for a fairly large cross section, so that we can reasonably invoke arguments about asymptotic sampling distributions. We assume further that there are several independent variables, so that y_t is given by the relation

$$(4) \qquad y_t = X_t' B + u_t.$$

X_t is a vector of independent variables and B a vector of coefficients; y_t and u_t are scalars as before. (We shall continue with the convention of letting capitals denote vectors and lower case letters denote scalars for the remainder of this chapter.) The disturbance is assumed to be autocorrelated for each element in the cross section and to be generated by the first-order autoregressive process

$$(5) \qquad u_t = \lambda\, u_{t-1} + w_t. \qquad\qquad \begin{array}{c} -1 < \lambda < 1 \\ t = \ldots -1, 0, 1, 2, \ldots \end{array}$$

We assume that the sign of λ is known a priori but that the magnitude of λ is not known except for the condition $|\lambda| < 1$, which is necessary if the variance of u_t is to be finite. The values of λ and of σ_w^2, the variance of the disturbance in the autoregressive process, are assumed to be constant over time and for all elements in the cross section. Further, w_t is assumed to have mean zero and to be distributed independently of each of the explanatory variables in equation (4). All covariances of disturbances between different elements in the cross section are taken to be zero. Several of these assumptions will be relaxed below.

A brief comment may be appropriate with regard to the assumption that the sign of λ is known, since this assumption plays a crucial role in the estimation of the basic model. It will become clear below, when some generalizations of the model are investigated, that the assumption is needed only when, as in the present case, the number of time periods separating complete observations is even. In any event, the additional information required should not seriously limit applications of the model, since the economist will usually have a good idea what the sign of λ should be. The prevailing view is that a positive λ is the typical case in autoregressive economic models,[1] and

---

[1]See, for example, Arthur S. Goldberger, Econometric Theory (New York: John Wiley and Sons, 1964), p. 153, or Christ, op. cit., p. 529.

a specification that λ is positive would probably be appropriate for most models explaining components of personal income.

The assumption of autocorrelated disturbances makes an important contribution to the utility of the time series-cross section model. Because of the omission from equation (4) of relevant independent variables that cannot be measured or identified, at least some autocorrelation between residuals from the same cross section element will normally be present. Moreover, the fact that independent variables in economic cross section relations frequently explain only a low proportion of variation in y_t gives additional importance to taking the autoregressive structure of the disturbances, if one exists, into account.

Some manipulation of equations (4) and (5) will allow an explicit derivation of an expression for y_1, and suggest the way that estimation should proceed. Substituting equation (4) taken at time t - 1 into (5) to remove u_{t-1}, we obtain

$$u_t = \lambda \ (y_{t-1} - X'_{t-1} B) + w_t.$$

Multiplying each side by - λ and rearranging terms gives

$$\lambda^2 \, y_{t-1} = \lambda^2 \, X'_{t-1} \, B + \lambda u_t - \lambda w_t. \tag{6}$$

Now substituting (4) at time t - 1 into (5) taken at t - 1 to remove u_{t-1}, we obtain, on rearranging terms,

$$y_{t-1} = X'_{t-1} \, B + \lambda u_{t-2} + w_{t-1}. \tag{7}$$

Adding (6) and (7) gives

$$(1 + \lambda^2)y_{t-1} = (1 + \lambda^2)X'_{t-1}\,B + \lambda(u_t + u_{t-2}) - \lambda w_t + w_{t-1},$$

or

$$(8) \qquad y_{t-1} = X'_{t-1}\,B + \frac{2\lambda}{1 + \lambda^2} \cdot \frac{u_t + u_{t-2}}{2} - \frac{\lambda w_t - w_{t-1}}{1 + \lambda^2} \;.$$

Equation (8) may be interpreted as stating that if the variable $(u_t + u_{t-2})/2$ could be added to the list of regressors specified in equation (4), the coefficients of the original regressors would be unchanged, the coefficient of the new variable would be $2\,\lambda/(1 + \lambda^2)$, and the variance of the disturbance term would be reduced from σ_u^2 to

$$\text{var}\left[- \frac{\lambda w_t - w_{t-1}}{1 + \lambda^2}\right] = \left(\frac{1}{1 + \lambda^2}\right)^2 (\lambda^2\,\sigma_w^2 + \sigma_w^2) = \frac{1}{1 + \lambda^2}\;\sigma_w^2$$

That $\sigma_w^2\,/(1 + \lambda^2)$ is in fact smaller than σ_u^2 when $\lambda \neq 0$ may be shown by first taking the variance of both sides of equation (5). We have

$$\text{var}\,(u_t) = \lambda^2\;\text{var}\,(u_{t-1}) + \text{var}\,(w_t),$$

so that, using the fact that the autoregressive process with $|\,\lambda\,| < 1$ is stationary,

$$\sigma_w^2 = (1 - \lambda^2)\;\sigma_u^2.$$

Hence

$$(9) \qquad \text{var}\left[-\frac{\lambda w_t - w_{t-1}}{1+\lambda^2}\right] = \frac{(\lambda^2+1)}{(1+\lambda^2)^2}\,\sigma_w^2 = \frac{1-\lambda^2}{1+\lambda^2}\,\sigma_u^2 \,.$$

If $0 < \lambda^2 < 1$, the coefficient $(1 - \lambda^2)/(1 + \lambda^2)$ is a positive number less than one, and introduction of the term $(u_t + u_{t-2})/2$ in equation (4) reduces the variance of the disturbance. If $\lambda^2 = 0$, however, the variance of the disturbance is unchanged.

Equation (8) may be used to estimate y_1 if we set the disturbance in that equation equal to its expected value zero, and if we have estimates of B, $2\lambda/(1 + \lambda^2)$, and $(u_t + u_{t-2})/2$. To obtain these estimates, two statistical properties of equation (8) must be established. First, in order to have consistent estimates of the parameters, it is necessary that the new variable $(u_t + u_{t-2})/2$, in addition to X_t, be uncorrelated with the disturbance $-(\lambda w_t - w_{t-1})/(1 + \lambda^2)$. That these magnitudes are uncorrelated follows if we expand $(u_t + u_{t-2})$ into an infinite series by repeated application of (5). One obtains

$$\text{cov}\left(\frac{u_t + u_{t-2}}{2}, \frac{\lambda w_t - w_{t-1}}{1+\lambda^2}\right) = \frac{1}{2(1+\lambda^2)}\,\text{cov}\,[w_t + \lambda w_{t-1}$$

$$+ (\lambda^2 + 1)w_{t-2} + \ldots, \lambda w_t - w_{t-1}]$$

$$= \frac{1}{2(1+\lambda^2)}\ \text{cov}\,[w_t, \lambda w_t] + \text{cov}\,[\lambda w_{t-1}, -w_{t-1}]$$

$$= \frac{1}{2(1+\lambda^2)}\,(\lambda\sigma_w^2 - \lambda\sigma_w^2) = 0\,,$$

since w_t and $w_{t'}$ are uncorrelated for $t \neq t'$.

Second, in order to carry out the estimation of B, it is necessary to show that $(u_t + u_{t-2})/2$ is uncorrelated with each of the variables included in the vector X_{t-1}. This specification follows at once from our assumption that w_t and X_t' are independent, since again considering the expansion of $(u_t + u_{t-2})/2$, each cov $(x_{i,t-1}, \frac{u_t + u_{t-2}}{2})$ may be expressed as an infinite sum of covariances that vanish. The significance of this result is that unbiased estimates of B can be obtained by least-squares estimation of equation (8) when $(u_t + u_{t-2})/2$ is omitted from the list of regressors.[1] Hence the first step in the estimation of (8) should be to estimate equation (4) by direct least squares. All of the complete observations should be used--that is, the regression should be based on the pooled cross-sections for $t = 0$ and $t = 2$.[2]

---

[1]Goldberger, op. cit., pp. 200-201.

[2]It may be helpful, at this point, to compare the model and estimation procedure considered here with those considered by Arnold Zellner in his paper "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Journal of the American Statistical Association, 57 (June, 1962), pp. 348-368. Zellner treats the case in which the model contains several equations having the form of our equation (4), and these equations may be interpreted as corresponding to successive cross sections. (This is one of several interpretations given by Zellner.) As in our model, disturbances corresponding to different cross section elements are uncorrelated, but disturbances corresponding to the same cross section are interdependent over time, perhaps satisfying the relation given by our equation (5). Under these conditions, Zellner shows that there is a two stage estimation procedure that, by taking advantage of zero restrictions in the complete model, improves the efficiency of the estimation of B. Our model, in contrast to Zellner's, assumes that B is the same for each cross section. Hence, there are no zero restrictions in our model, and no gain in efficiency from using Zellner's estimating procedure. (footnote continued)

In time series analysis, when autocorrelation is present, a recommended procedure is to follow the direct least-squares estimation with a second pass that takes account of the autocorrelation by appropriate transformation of the original data. In the case of a first-order autoregressive process, the procedure is to obtain the estimated residuals from the direct least-squares regression, use these to estimate the autoregressive coefficient $\hat{\lambda}$, and to reestimate the first equation using the transformed variables $y_t^* = y_t - \hat{\lambda} y_{t-1}$, $x_{1t}^* = x_{1t} - \hat{\lambda} x_{1,t-1}$, etc. The same transformation, applied twice, could be used in estimating the present time series-cross section model. The variables for the second pass of the two pass estimation procedure would be $y_t^{**} = y_t - \hat{\lambda}^2 y_{t-2}$, $x_{1t}^{**} = x_{1t} - \hat{\lambda}^2 x_{1,t-2}$, etc. The difficulty with this procedure is that instead of giving up two observations, as would be the case if we had a pure time series model, application of this transformation to the two pooled cross sections would reduce the number of observations by half. Moreover, the extent to which the covariance matrix of the disturbances

---

Footnote 2, p. 176, continued.

Zellner does consider a model formally identical to ours in connection with the question of "aggregation bias" resulting from estimation with pooled cross sections. In our context there is no aggregation bias by assumption, but there is a question of the stability of the coefficients over time. Zellner's test for the former is also (under our hypotheses) a valid test for the latter. To make this test, which is desirable in any application of the interpolation procedure recommended here, one begins by estimating (4) using individual cross sections (_a la_ Zellner) as well as with pooled data. An F-test is then used to test the significance of the difference in total sample variance of the residuals under the two model specifications.

departs from classical assumptions is much smaller in the time series-cross section case than in the pure time series case. After striking rows and columns for the alternate (unobserved) time periods, the covariance matrix in the pure time series case is

$$\sigma^2 \begin{bmatrix} 1 & \lambda^2 & \lambda^4 & \cdots & \lambda^{2(n-1)} \\ \lambda^2 & 1 & \lambda^2 & & \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \lambda^{2(n-1)} & & \cdots & & 1 \end{bmatrix},$$

while our time series-cross section case yields the block diagonal covariance matrix

$$\sigma^2 \begin{bmatrix} 1 & \lambda^2 & & & \bigcirc & \\ \lambda^2 & 1 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & \bigcirc & & & 1 & \lambda^2 \\ & & & & \lambda^2 & 1 \end{bmatrix},$$

in which each 2 x 2 non-zero block represents the variances and covariance of the disturbances for an element of the cross section. In view of the substantial reduction in precision that might result from a halving of the number of observations, it would appear preferable to estimate B by direct least squares, and to make no allowance in the estimating procedure for autocorrelation of disturbances.

When B has been estimated, the computed residuals e_t may be obtained for the two cross sections, and $(u_0 + u_2)/2$ may be estimated as $(e_0 + e_2)/2$. The computed residuals may also be used to estimate $2\lambda/(1 + \lambda^2)$. Substituting (5) into itself recursively, we obtain

$$u_t = \lambda^2 u_{t-2} + (\lambda w_{t-1} + w_t),$$

and using e_0 and e_2 as proxies for u_0 and u_2, this relation becomes

$$e_2 = \lambda^2 e_0 + w^*. \tag{10}$$

Least-squares estimation of (10) yields an estimate $\hat{\lambda}^2$ of λ^2, and $\hat{\lambda}$ may be obtained by taking the square root of the regression coefficient and affixing the appropriate sign. These considerations suggest that one might take, as an estimator of $2\lambda/(1 + \lambda^2)$, the quantity $\pm 2\hat{\lambda}/(1 + \hat{\lambda}^2)$. We shall investigate some properties of this estimator in the next section.

It is possible that the estimated $\hat{\lambda}^2$ will be negative, in which case this procedure cannot be carried out. When this occurs, one possibility is that the true value of λ^2 is in fact positive or zero, but that λ^2 has been estimated as negative because of sampling variation. The other possibility is that the autoregressive process has been misspecified. For example, suppose that the residuals could be described by a first-order autoregressive process if the interval between observations were two-thirds as great, and that λ would then be negative. The equation actually estimated would then be

$$e_3 = \lambda^3 e_0 + w^{**},$$

in which the true value of the regression coefficient λ^3 would be negative. In either of these cases an appropriate procedure would be to disregard equation (5) in the model, and to estimate y_1 directly from equation (4). The principle of Occam's razor suggests that equation (5) should also be disregarded when $\hat{\lambda}^2$ is estimated to be positive but not statistically significant. A final outcome from estimation not consistent with the model would occur if λ^2 were estimated to be statistically significant and greater than one. This case is considered below in connection with heteroscedastic disturbances.

Although $2\,\hat{\lambda}/(1 + \hat{\lambda}^2)$ may be evaluated directly as a function of $\hat{\lambda}^2$, an alternative procedure is available which is computationally a bit shorter. Since

$$\hat{\lambda}^2 = \frac{\text{cov }(e_0,\ e_2)}{\text{var }(e_0)},$$

we have

$$(11) \qquad \frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} = \frac{\pm\left(\dfrac{\text{cov }(e_0,\ e_2)}{\text{var }(e_0)}\right)^{\frac{1}{2}}}{1 + \dfrac{\text{cov }(e_0,\ e_2)}{\text{var }(e_0)}} = \frac{\pm\sqrt{\text{var}(e_0) \cdot \text{cov}(e_0,\ e_2)}}{\frac{1}{2}[\text{var}(e_0) + \text{cov}(e_0,\ e_2)]}\ .$$

Thus the coefficient of $(u_0 + u_2)/2$ is seen to be equal except possibly for sign to the ratio of the geometric mean and the arithmetic mean of two second-order moments. The author has found it computationally convenient to use this form of the regression coefficient. A test for the significance of the coefficient may be based on the sample correlation between e_0 and e_2, rather than the sample regression coefficient.

So far as the treatment of computed residuals is concerned, the method of this section is easily related to the two methods described previously. When $\hat{\lambda}^2$ is zero, the present procedure is equivalent to the simple regression method. As $\hat{\lambda}^2$ approaches one, with λ positive, $\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2}$ approaches one, and our procedure becomes equivalent to Friedman's.

3. The Reliability of Estimates from the Model

What is the reliability of estimates of y_1 obtained from the time series-cross section model? Useful measures of reliability are the forecast bias and the variance of the error of forecast for specified values of the exogenous variables. One might wish to evaluate these quantities for given e_0, e_2 and observed exogenous variables for a particular element of the cross section, or alternatively for the cross section mean. Both of these tasks are easier in the latter case, since, with the mean of $(e_0 + e_2)/2$ equal to zero, the results turn out to be independent of the sampling distribution of $2\hat{\lambda}/(1 + \hat{\lambda}^2)$. After obtaining results for the simpler case, we shall investigate this sampling distribution, and apply our findings to the case in which a non-zero value of $(e_0 + e_2)/2$ is specified. In this case we will have to be content with approximate results that hold asymptotically.

Let us express the true value of y_1 by

$$y_1 = X_1' B + \frac{2\lambda}{1 + \lambda^2} \cdot \frac{e_0 + e_2}{2} + v_1. \tag{12}$$

Comparison with equation (8) shows that the disturbance v_1 is algebraically greater than $-(\lambda\ w_t - w_{t-1})/(1 + \lambda^2)$, the disturbance in that equation, by the amount

$$\frac{2\lambda}{1 + \lambda^2}\left(\frac{u_0 + u_2}{2} - \frac{e_0 + e_2}{2}\right).$$

However, the expected value of this expression is zero, and the contribution of measurement errors in e_0 and e_2 to the variance of v_1 will be neglected.

The predicted value y_1 is

$$\text{(13)} \qquad \hat{y}_1 = X_1' \hat{B} + \frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} \cdot \frac{e_0 + e_2}{2},$$

which is unbiased when $(e_0 + e_2)/2 = 0$. The error of forecast is

$$\hat{y}_1 - y_1 = X_1' (\hat{B} - B) + \left(\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} - \frac{2\lambda}{1 + \lambda^2}\right) \frac{e_0 + e_2}{2} - v_1,$$

and since e_0 and e_2 are distributed independently of X_1, B and $\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2}$ are also distributed independently, and the mean square error of forecast is

$$\text{(14)} \qquad E(\hat{y}_1 - y_1)^2 = X_1' E(\hat{B} - B)\ (\hat{B} - B)'X_1 + \frac{1}{4}(e_0 + e_2)^2$$
$$E\left(\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} - \frac{2\hat{\lambda}}{1 + \hat{\lambda}^2}\right)^2 + \text{var}(v_1).$$

Again taking the case in which $(e_0 + e_2)/2 = 0$, we see that, since $\hat{B}$ is unbiased, the mean square error of forecast is equal to the forecast variance, and may be expressed as the sum of the variance of the disturbance $\mathrm{var}(v_1)$ and a quadratic form in the observed independent variables, the coefficients of which are the variances and covariances of the corresponding regression coefficients.

All of these quantities are readily evaluated. For example, if we let f_1 denote the "residual" corresponding to the disturbance term v_1, equation (9) suggests as an estimator of $\mathrm{var}(v_1)$:

$$\mathrm{var}(f_1) = \frac{1 - \hat{\lambda}^2}{1 + \hat{\lambda}^2} s_e^2 .$$

Here we have substituted into equation (9) $\hat{\lambda}_2^2$ for λ^2 and s_e^2 for σ_u^2, where s_e^2 is the sample variance of the pooled residuals obtained from the estimation of equation (4). With regard to the covariance matrix for $\hat{B}$, we need only note that the covariance matrix associated with least-squares estimation of (4) will be too large by a constant factor, since each element of $\hat{B}$ will contain the factor s_e^2 rather than the factor $\mathrm{var}(f_1)$. Thus to obtain the covariance matrix for $\hat{B}$ associated with equation (12), the variances and covariances associated with (4) should each be multiplied by $(1 - \hat{\lambda}^2)/(1 + \hat{\lambda}^2)$.

We now investigate the sampling properties of $2\hat{\lambda}/(1 + \hat{\lambda}^2)$ to the extent necessary to evaluate the term $E[2\hat{\lambda}/(1 + \hat{\lambda}^2) - 2\lambda/(1 + \lambda^2)]^2$ in equation (14). Since it is a continuous function of the consistent estimator $\hat{\lambda}^2$, $2\hat{\lambda}/(1 + \hat{\lambda}^2)$ is itself a consistent estimator,[1] and we have

---

[1] S. S. Wilks, <u>Mathematical Statistics</u> (New York: John Wiley and Sons, 1962), pp. 102-103.

$$\operatorname{plim} \frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} = \frac{2\lambda}{1 + \lambda^2} .$$

The variance of $2\hat{\lambda}/(1 + \hat{\lambda}^2)$ is equal, to a first approximation, to[1]

$$\operatorname{var}(\hat{\lambda}^2) \left(\frac{d}{d\hat{\lambda}^2} \quad \frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} \right)^2 ,$$

and evaluating the derivative we have

$$\operatorname{var}\left(\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} \right) = \operatorname{var}(\hat{\lambda}^2) \cdot \left(2 \frac{(1 + \hat{\lambda}^2) \frac{d\hat{\lambda}}{d\hat{\lambda}^2} - \hat{\lambda} \frac{d}{d\hat{\lambda}^2} (1 + \hat{\lambda}^2)}{(1 + \lambda^2)^2} \right)^2$$

$$= \operatorname{var}(\hat{\lambda}^2) \left(2 \frac{\frac{1 + \hat{\lambda}^2}{2\hat{\lambda}} - \hat{\lambda}}{(1 + \hat{\lambda}^2)^2} \right)^2$$

$$= \operatorname{var}(\hat{\lambda}^2) \frac{(1 - \hat{\lambda}^2)^2}{\hat{\lambda}^2 (1 + \hat{\lambda}^2)^4} .$$

The first factor is the variance of the least-squares estimate of $\hat{\lambda}^2$ that one would obtain from equation (10).

A computational advantage of this approximation to $\operatorname{var}[2\hat{\lambda}/(1 + \hat{\lambda}^2)]$ is that the second factor on the right depends only on $\hat{\lambda}^2$. A better small sample approximation to $\operatorname{var}[2\hat{\lambda}/(1 + \hat{\lambda}^2)]$ may be obtained if one is willing to compute some third and fourth order moments. It was shown above in equation (11) that $2\hat{\lambda}/(1 + \hat{\lambda}^2)$ is a simple function of

---

[1]Maurice G. Kendall and Alan Stuart, The Advanced Theory of Statistics, Vol. I, Distribution Theory (London: Charles Griffin and Company, 1958), p. 232.

the sample moments var (e_0) and cov (e_0, e_2). The population variance $\mathrm{var}[2\hat{\lambda}/(1 + \hat{\lambda}^2)]$ may be expressed as a function of the corresponding population moments. It is convenient to introduce the notation m_{rs} for sample moments, so that

$$m_{20} = \mathrm{var}(e_0)$$

$$m_{11} = \mathrm{cov}(e_0, e_2).$$

Similarly, μ_{rs} will denote the corresponding population moments. Then we have, to an approximation,[1]

$$\begin{aligned}
\mathrm{var}\left(\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2}\right) &= \mathrm{var}\left(\frac{\sqrt{m_{20}m_{11}}}{\frac{1}{2}(m_{20} + m_{11})}\right) \\
&= \mathrm{var}(m_{20}) \cdot \left(\frac{\partial}{\partial m_{20}} \frac{\sqrt{m_{20}m_{11}}}{\frac{1}{2}(m_{20} + m_{11})}\right)^2 \\
&\quad + 2\,\mathrm{cov}(m_{20}, m_{11}) \cdot \left(\frac{\partial}{\partial m_{20}} \frac{\sqrt{m_{20}m_{11}}}{\frac{1}{2}(m_{20} + m_{11})}\right) \\
&\quad \cdot \left(\frac{\partial}{\partial m_{11}} \frac{\sqrt{m_{20}m_{11}}}{\frac{1}{2}(m_{20} + m_{11})}\right) + \mathrm{var}(m_{11}) \cdot \left(\frac{\partial}{\partial m_{11}} \frac{\sqrt{m_{20}m_{11}}}{\frac{1}{2}(m_{20} + m_{11})}\right)^2,
\end{aligned}$$

in which the partial derivatives are evaluated at the values $m_{20} = \mu_{20}$ and $m_{11} = \mu_{11}$. Forming the partial derivatives, we have

---

[1] *Ibid.*, pp. 231-232.

$$(16)\qquad \frac{\partial}{\partial m_{20}} \frac{\sqrt{m_{20}m_{11}}}{\frac{1}{2}(m_{20}+m_{11})} = \frac{2}{1} \cdot \frac{(m_{20}+m_{11})\dfrac{m_{11}}{2\sqrt{m_{20}m_{11}}} - \sqrt{m_{20}m_{11}} \cdot 1}{(m_{20}+m_{11})^2}$$

$$= \frac{2}{1} \quad \frac{(m_{20}+m_{11})m_{11} - 2m_{20}m_{11}}{2\sqrt{m_{20}m_{11}} \quad (m_{20}+m_{11})^2}$$

$$= \sqrt{\frac{m_{11}}{m_{20}}} \; \frac{m_{11} - m_{20}}{(m_{20}+m_{11})^2}$$

and

$$(17)\qquad \frac{\partial}{\partial m_{11}} \frac{\sqrt{m_{20}m_{11}}}{\frac{1}{2}(m_{20}+m_{11})} = \sqrt{\frac{m_{20}}{m_{11}}} \; \frac{m_{20} - m_{11}}{(m_{20}+m_{11})^2} \quad .$$

The variances and covariance of the sample moments m_{20} and m_{11} may be expressed as follows. For m_{20},

$$(18)\qquad \text{var}(m_{20}) = \frac{1}{n}(\mu_{40} - \mu_{20}^2) - \frac{2}{n^2}(\mu_{40} - 2\mu_{20}^2) - \frac{1}{n^3}(\mu_{40} - 3\mu_{20}^2),$$

where n is the number of observations in the sample. Following Kendall and Stuart,[1] we find

$$(19)\qquad \text{cov}(m_{11}, m_{20}) = \frac{1}{n}(\mu_{31} - \mu_{11}\,\mu_{20} + 2\mu_{20}\,\mu_{01}\,\mu_{10} + 2\,\mu_{11}\,\mu_{10}^2$$

$$- 2\mu_{21}\,\mu_{10} - \mu_{01}\,\mu_{30} - \mu_{10}\,\mu_{21})$$

and

$$(20)\qquad \text{var}(m_{11}) = \frac{1}{n}(\mu_{22} - \mu_{11}^2 + \mu_{20}\,\mu_{01}^2 + \mu_{02}\,\mu_{10}^2$$

$$+ 2\mu_{11}\,\mu_{01}\,\mu_{10} - 2\mu_{21}\,\mu_{01} - 2\mu_{12}\,\mu_{10}).$$

---

[1] Ibid., p. 235.

These formulas simplify somewhat on the assumption that the autoregressive process is strongly stationary, since then the population moments all satisfy a symmetry condition $\mu_{rs} = \mu_{sr}$.

Equation (15) may be evaluated by substituting into it the results of equations (16)-(20), and then replacing the population moments with unbiased sample moments. By the rule for functions of consistent estimators, the consistency of the right-hand side of (15) as an estimate of var $\left(\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2}\right)$ is not affected by the substitution of sample moments. Cramér has shown that an expression of the form of the right-hand side of (15) is a consistent estimate of the variance of a function of sample moments, and further that the distribution of a function of sample moments is asymptotically normal with the variance approaching the form given in (15).[1]

4. Generalizations of the Model: Heteroscedastic Disturbances

So far, in discussing the estimation of y_1, we have not been concerned with the possibility of heteroscedastic disturbances. In studies based on cross section data, heteroscedasticity almost always occurs, and it can cause serious difficulty if not taken into account explicitly by the model. Specifically, since heteroscedastic disturbances give rise to biased estimates of the sample variances and covariances of least-squares regression coefficients (such as our $\hat{B}$), they can undermine attempts to evaluate the reliability of predicted values of y_t. For the

[1]Harold Cramér, Mathematical Methods of Statistics (Princeton: Princeton University Press, 1946), pp. 366-367.

same reason, they interfere with any attempt to choose among alternative versions of equation (8) which contain different sets of observed explanatory variables. A disturbing result peculiar to our model and developed below is that heteroscedastic disturbances can lead to an inconsistent estimate of $2\lambda/(1 + \lambda^2)$.

Since ours is a combined time series-cross section model, it is useful to distinguish two ways in which heteroscedasticity may arise. One form, which might be called within-year heteroscedasticity, occurs if the variances of the disturbances within a single cross section are unequal. The other form, which might be called between-year heteroscedasticity, occurs if the variances of the disturbances vary over time. One or both types of heteroscedasticity may be present. In terms of our two equation model, within-year heteroscedasticity may be thought of as heteroscedasticity in first equation. Thus in the case of a single cross section one might have in place of equation (4),

$$(21) \qquad y_i = X_i' B + z_i u_i^* ,$$

which satisfied the condition that $\text{var}(u_i^*)$ was a constant over the cross section. One might be able to make the additional assumptions that z_i was an observable variable, and that, in the case of several cross sections corresponding to different years, the same variable z_i, observed in those years, could be used to define

$$\text{var}(u_i) = z_i^2 \text{ var}(u_i^*),$$

where $\text{var}(u_i^*)$ was a constant within the cross section.

The variance of u_i^* might still vary between cross sections, however, and this variation may be thought of as arising from heteroscedasticity in the second equation of the model. There might be an observed variable s_t, the same at a given date for every element in the cross section, such that the unobserved portion of the disturbances, w_{it}^*, satisfied

$$u_{it}^* = \lambda u_{i,t-1}^* + s_t \, w_{it}^*,$$

with $\text{var}(w_{it}^*)$ constant over time as well as over the cross section. On the other hand, since the assumed sample contains many observations for each date t, one might replace s_t with a function of time containing a parameter not known a priori. An interesting example of such a function is $e^{\alpha t}$, which leads, on dropping the subscript i, to the autoregressive relation

$$u_t^* = \lambda u_{t-1}^* + e^{\alpha t} \, w_t^*. \tag{22}$$

It may be shown that $\text{var}(u_t^*)$ increases or decreases over time according to whether α is positive or negative. Substituting (22) into itself repeatedly for successively smaller values of t, one obtains the infinite expansion

$$\begin{aligned} u_t^* &= e^{\alpha t} \, w_t^* + \lambda e^{\alpha(t-1)} \, w_{t-1}^* + \lambda^2 \, e^{\alpha(t-2)} \, w_{t-2}^* + \ldots \\ &= (w_t^* + \lambda e^{-\alpha} \, w_{t-1}^* + \lambda^2 \, e^{-2\alpha} \, w_{t-2}^* + \ldots)e^{\alpha t}. \end{aligned}$$

Since $var(w_t^*)$ is constant over time,

$$var(u_t^*) = (1 + \lambda^2 e^{-2\alpha} + \lambda^4 e^{-4\alpha} + \ldots)e^{2\alpha t} var(w^*).$$

The expression in parentheses is a geometric series which converges when $\left| \lambda^2 e^{-2\alpha} \right| < 1$, and hence when $\left| \lambda e^{-\alpha} \right| < 1$. If we assume that this condition holds,

$$(23) \qquad var(u_t^*) = \frac{e^{2\alpha t}}{1 - \lambda^2 e^{-2\alpha}} var(w^*).$$

The rate of change in $var(u_t^*)$ over time is then

$$\frac{d\, var(u_t^*)}{dt} = \frac{2\ \alpha\ e^{2\alpha t}}{1 - \lambda^2 e^{-2\alpha}} var(w^*),$$

which will have the same sign as α.

Equation (22) can thus be used to explain either increases or decreases in the variance of u_t^*, and since the condition $\left| \lambda e^{-\alpha} \right| < 1$ for this variance to be finite is more general than the condition $\left| \lambda \right| < 1$, the present model may be able to accommodate a case in which λ^2 is estimated as greater than one. Moreover, a consistent estimator of α, when there are two cross sections, is readily derived. From equation (23),

$$\frac{var(u_2^*)}{var(u_0^*)} \ \frac{e^{4\alpha}}{e^0} = e^{4\alpha}$$

and taking logs,

$$\alpha = \frac{1}{4} \log_e \frac{\text{var}(u_2)}{\text{var}(u_0)} \quad .$$

Hence by the theorem on functions of consistent estimators, a consistent estimator $\hat{\alpha}$ of α is obtained if $\text{var}(u_2)$ and $\text{var}(u_0)$ are replaced by consistent estimates of these quantities.

We shall take equations (21) and (22) as the interpolation model generalized for heteroscedastic disturbances. In order to relate the implications of this model to our results for the homoscedastic case, we need to express the required value y_1 as a function of $u_0 = z_0 u_0^*$ and $u_2 = z_2 u_2^*$. The derivation parallels that of equation (8) and is given in a footnote.[1] The result is

---

[1] Substitution of (21) into (22), after adopting the notation $w_t = e^{\alpha t} w_t^*$, gives

$$u_t^* = \lambda \quad \frac{y_{t-1}}{z_{t-1}} - \frac{1}{z_{t-1}} \quad X'_{t-1} B \quad + w_t.$$

Multiplying each side by $- \lambda z_{t-1}$ and rearranging terms,

$$\text{(f.1)} \qquad \lambda^2 y_{t-1} = \lambda^2 X'_{t-1} B + \lambda u_t^* z_{t-1} - \lambda w_t z_{t-1}.$$

Substitution of (21) into (22) for u_t^* rather than for u_{t-1}^*, and then reducing time subscripts by one, gives

$$\frac{y_{t-1}}{z_{t-1}} - \frac{1}{z_{t-1}} X'_{t-1} B = \lambda u_{t-1}^* + w_{t-1},$$

or

$$\text{(f.2)} \qquad y_{t-1} = X'_{t-1} B + \lambda u_{t-2}^* z_{t-1} + w_{t-1} z_{t-1}.$$

(footnote continued)

$$(24) \qquad y_1 = X_1' B + \frac{2\lambda}{1+\lambda^2} \frac{u_0^* + u_2^*}{2} z_1 - \frac{1}{1+\lambda^2} z_1 (\lambda e^{2\alpha} w_2^* - w_0^*).$$

Equation (24) is similar to (21) in that the variance of its disturbance term is proportional to z_1^2. It differs from equation (8) in that the term in the average of the disturbances, $(u_0^* + u_2^*)/2$, now also involves z_1. However, the coefficient of this term is the familiar $2\lambda(1 + \lambda^2)$.

Because of the greater complexity of the heteroscedastic model, and because the details of the model may not be fully specified in advance, it may be helpful to think of the estimation of the model as a sequence of smaller computational tasks. Our approach is to estimate the model given by (21) and (22) by carrying out those transformations of the data required for the application of generalized least squares. However, different routes must be followed depending on whether $\hat{\lambda}^2$ and/or $\hat{\alpha}$ turn out to be significantly different from zero. In general, the following sequence of computations should be used:

1) The data for y_t and X_t at $t = 0$ and $t = 2$ should be deflated by the variable selected as z_t, and B and the computed residuals should be estimated by the least-squares relation

$$(25) \qquad \frac{y_t}{z_t} = \frac{1}{z_t} X_t' \hat{B} + e_t^* .$$

Footnote 1, p. 191, continued.

Addition of (f.1) and (f.2), and division by $1 + \lambda^2$ yields

$$y_{t-1} = X_1' B + \frac{2\lambda}{1+\lambda^2} \frac{u_t^* + u_{t-2}}{2} z_{t-1} - \frac{z_{t-1}}{1+\lambda^2} (\lambda w_t - w_{t-1}).$$

2) The computed residuals should be partitioned into those for $t = 0$ and $t = 2$, and the sample variances of the residuals should be computed for each year. The sample variances should be tested for equality. If the hypothesis of equality is not rejected, the ratio of the variances will not differ significantly from one, and equation (24) suggests that the model should be simplified by setting α equal to zero. If the sample variances are found to be unequal, however, it follows that the residuals associated with the transformed variables of equation (25) are still heteroscedastic, and that further transformation of the data is required. This further transformation is needed, in the first instance, because, with the variance of the residuals in equation (25) biased downward, the residuals themselves, which are the basis of any estimates of λ and α, are not estimated reliability. Unequal variances in the computed residuals can be removed by dividing each of the observations by the sample standard deviation of the residuals in the corresponding year and re-estimating the equation. The equation to be estimated is then

$$(26) \qquad \frac{y_t}{z_t \, s_{e_t^*}} = \frac{1}{z_t \, s_{e_t^*}} X_t'B + e_t^{**}.$$

If re-estimation is necessary, the e_t^* can be obtained afterwards from the relation

$$e_t^* = s_{e_t^*} \; e_t^{**}$$

Because the estimated value of B will differ as between equations (25) and (26), the values of e_t^* will also differ.

3) Using the latest values of e_t^*, and again with these values partitioned by year, compute the correlation coefficient between residuals in the two cross sections. The correlation coefficient should be tested for statistical significance, and if it is not found significant, the model should be simplified by assuming λ equal to zero. At this point the form of equation (22) is completely specified, and if both α and λ are found not to differ significantly from zero, this equation may be dropped from the model. The remainder of our discussion will focus, however, on the general case in which both α and λ have been found to be non-zero.

4) Recursive substitution of (22) into itself gives

$$u_t^* = \lambda^2 u_{t-2}^* + (e^{\alpha t} w_t^* + \lambda e^{\alpha(t-1)} w_{t-1}^*), \tag{27}$$

in which the disturbance term is enclosed in parentheses. Clearly, if this equation is estimated from the residuals from two cross sections, the heteroscedasticity of the equation is "unobservable," and has no effect on the least-squares estimate of λ^2. Thus the estimation of λ^2 and $2\lambda/(1 + \lambda^2)$ may proceed exactly as in the homoscedastic case, and, using (11), we write down as an estimate for the coefficient of $z_t(u_0^* + u_2^*)/2$ in equation (24),

$$\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} = \frac{\sqrt{\operatorname{var}(e_0^*) \quad \operatorname{cov}(e_0^*, e_2^*)}}{\frac{1}{2}[\operatorname{var}(e_0^*) + \operatorname{cov}(e_0^*, e_2^*)]} .$$

This completes the estimation of the coefficients of the model.

Substituting $2\hat{\lambda}/(1 + \hat{\lambda}^2)$ for $2\lambda(1 + \lambda^2)$, $(e_0^* + e_2^*)/2$ for $(u_0^* + u_2^*)/2$, and $\hat{B}$ for B in equation (24), and setting $-z_1(\lambda e^{2\alpha} w_2^* - w_0^*)/(1 + \lambda^2)$ equal to zero, predicted values of y_1 may be computed from the relation

$$\hat{y}_1 = X_1' \hat{B} + \frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} \frac{e_0^* + e_2^*}{2} z_1 .$$

Estimation of the mean square error of forecast for the predicted values $\hat{y}_1$ is also similar to the homoscedastic case. In place of equation (14), we now have

$$(28) \qquad E(\hat{y}_1 - y_1)^2 = X_1' E(\hat{B} - B)(\hat{B} - B)' X_1 + \frac{1}{4} z_1^2 (e_0^* + e_2^*) \cdot E\left(\frac{2\hat{\lambda}}{1 + \hat{\lambda}^2} - \frac{2\lambda}{1 + \lambda^2}\right)^2 + \mathrm{var}(v_1^*) ,$$

$$(29) \qquad \text{where } v_1^* = - \frac{z_1}{1 + \lambda^2} (\lambda e^{2\alpha} w_2^* - w_0^*),$$

the disturbance in equation (24). As before it is necessary to evaluate the variances of $2\hat{\lambda}/(1 + \hat{\lambda}^2)$ and v_1^*, and the covariance matrix for $\hat{B}$.

Let us begin with the evaluation of $\mathrm{var}(v_1^*)$. Taking variances in equation (29), we have

$$(30) \qquad \mathrm{var}(v_1^*) = \left(\frac{z_1}{1 + \lambda^2}\right)^2 (\lambda^2 e^{4\alpha} + 1) \cdot \mathrm{var}(w^*),$$

since the variance of w_t^* is assumed constant over time. Similarly, forming the variance of each side of equation (27),

$$(31) \qquad \mathrm{var}(u_2^*) = \lambda^4 \mathrm{var}(u_0^*) + (e^{4\alpha} + \lambda^2) \mathrm{var}(w^*).$$

Solving equation (31) for var (w^*) and substituting into (30), we obtain

$$(32) \qquad \mathrm{var}(v_1^*) = \left(\frac{z_1}{1+\lambda^2}\right)^2 (\lambda^2 e^{4\alpha} + 1) \quad \frac{\mathrm{var}(u_2^*) - \lambda^4 \mathrm{var}(u_0^*)}{e^{4\alpha} + \lambda^2} .$$

Thus $\mathrm{var}(v_1^*)$ may be estimated by substituting into equation (32) the sample variances $\mathrm{var}(e_0^*)$ and $\mathrm{var}(e_2^*)$ for $\mathrm{var}(u_0^*)$, $\mathrm{var}(u_2^*)$, the least-squares estimate $\mathrm{cov}(e_0^*, e_2^*)/\mathrm{var}(e_0^*)$ for λ^2, and $\frac{1}{4}\log_e[\mathrm{var}(e_2^*)/\mathrm{var}(e_0^*)]$ for α (using (24)). Since each of these functions are consistant estimates, $\mathrm{var}(v_1^*)$ is estimated consistently. The estimation of $\mathrm{var}[2\hat{\lambda}/(1+\hat{\lambda}^2)]$ and the covariance matrix for $\hat{B}$ are now straightforward. To obtain the latter, each element in the moment matrix $X'X$ (pooled observations for $t = 0,2$) is divided by the estimate of $\mathrm{var}(v_1^*)$, while $\mathrm{var}[2\hat{\lambda}/\ 1 + \hat{\lambda}^2]$ may be approximated exactly as in the homoscedastic case.

In single-equation models with heteroscedastic disturbances, least-squares estimates of the regression coefficients are consistent, and in fact unbiased. In our model, similar results do not hold for our estimator of $2\lambda/(1+\lambda^2)$ if heteroscedasticity exists in the first equation of the model, and is neglected. For suppose that the disturbance in the first equation was in fact proportional to z_t, but that this fact was neglected. One would then be led to base the estimation of λ^2 not on equation (27) but on

$$z_t\, u_t^* = \lambda^2\, z_{t-2}\, u_{t-2}^* + (e^{\alpha t}\, w_t^* + \lambda e^{\alpha(t-1)}\, w_{t-1}),$$

or equivalently, on

$$(33) \qquad u_t^* = \lambda^2 \frac{z_{t-2}}{z_t} u_{t-2} + \frac{1}{z_t} (e^{\alpha t} w_t^* + \lambda e^{\alpha(t-1)} w_{t-1}).$$

Since equation (33) involves an error in specifying the independent variable, the resulting estimate of λ^2 will be biased, and in fact not consistent. Hence our estimate of $2\lambda/(1 + \lambda^2)$ will not be consistent, unless, as may sometimes happen, z_t is constant over time. This result is serious because of the difficulty in practice of choosing the "right" variable z_t by which to deflate y_t and X_t. However, Goldfeld and Quandt, in a recent article, have suggested an F-test for choosing among alternative deflators, and their results would be useful in empirically implementing the present model.[1]

5. Generalizations of the Model: Other Patterns of Observation Dates

We turn now to generalization of the model to the important case in which the intervals of time between observations are no longer equal. For ease of exposition we return to the case of homoscedastic disturbances, but there are no difficulties in combining the two types of extension. The model is thus that given by equations (4) and (5), with the only difference being the dates at which observations occur.

Suppose that complete observations are made at dates t and t-m-n, and that observations are made on X_t but not on y_t at date t-n.

---

[1]Stephen Goldfeld and Richard E. Quandt, "Some Tests for Homoscedasticity," Journal of the American Statistical Association, 60 (June, 1965), pp. 539-59.

The observations are thus separated by time intervals m and n units in length, and we assume that m and n take on integral values. Our objective is to estimate y_{t-n}, taking into account the underlying autoregressive process. Earlier results lead to ask whether the estimation of y_{t-n} can be improved by adding to the estimating equation some linear combination of the computed residuals for t and t-m-n. One might conjecture, in analogy to these results, that a weighted average of the computed residuals should be formed with weights which were proportional to the lengths of the two intervals; that is, one should form

$$\frac{n}{m+n} e_{t-m-n} + \frac{m}{m+n} e_t.$$

This strategy, however, does not in general minimize the residual variance in the estimating equation for y_{t-n}. The optimal weights for the residuals may be shown to depend on the sign and magnitude of the autocorrelation coefficient λ, or equivalently, on the correlation between e_{t-m-n} and e_t. For the "standard" case in which λ is positive, the relative weight that should be given to the residual less distant in time from y_{t-n} is greater, the greater the correlation between the two residuals.

In order to demonstrate these results, we shall let k be the unknown weight that should be given to the residual e_t and 1-k the weight that should be given to e_{t-m-n}. Our argument will be that the weights that should be given to the computed residuals are those which if assigned u_t and u_{t-m-n}, and the resulting variable added to the list of regressors in equation (4), lead to a new equation for y_{t-n} such that the variance of the disturbance in this equation is smaller than for any other pair of weights. Thus we proceed very much as in the derivation of equation (8) above, except that the parameter k is found by minimizing a variance.

If we substitute equation (5) into itself n-1 times, we obtain

$$(34) \qquad u_t = \lambda^n u_{t-n} + \lambda^{n-1} w_{t-n+1} + \lambda^{n-2} w_{t-n+2} + \ldots + w_t .$$

Substituting (5) into itself m-1 times, and shifting the time index n units, gives

$$(35) \qquad u_{t-n} = \lambda^m u_{t-m-n} + \lambda^{m-1} w_{t-m-n+1} + \lambda^{m-2} w_{t-m-n+2} + \ldots + w_{t-n} .$$

To obtain an equation for y_{t-n} which includes a weighted sum of u_t and u_{t-m-n} on the right hand side, we must next use equation (4) to substitute out u_{t-n} in (34) and (35); we obtain

$$(36) \qquad u_t = \lambda^n (y_{t-n} - X'_{t-n} B) + \lambda^{n-1} w_{t-n+1} + \ldots + w_t$$

and

$$(37) \qquad y_{t-n} - X'_{t-n} B = \lambda^m u_{t-m-n} + \lambda^{m-1} w_{t-m-n+1} + \ldots + w_{t-n} .$$

Multiplying both sides of (36) by $k\lambda^m$ and both sides of (37) by 1-k, and then subtracting (36) from (37), one derives

$$(1-k+k\lambda^{m+n})y_{t-n} = (1-k+k\lambda^{m+n})X'_{t-n}B + \lambda^m[ku_t + (1-k)u_{t-m-n}]$$

$$+ (1-k)(\lambda^{m-1}w_{t-m-n+1} + \ldots + w_t)$$

$$- k\lambda^m(\lambda^{n-1}w_{t-n+1} + \ldots + w_t).$$

Division of both sides by $1-k+k\lambda^{m+n}$ yields

$$(38) \qquad y_{t-n} = X'_{t-n}B + \frac{\lambda^m}{1-k+k\lambda^{m+n}}[ku_t + (1-k)u_{t-m-n}]$$

$$+ \frac{1}{1-k+k\lambda^{m+n}}\ [(1-k)(\lambda^{m-1}w_{t-m-n+1} + \ldots + w_{t-n})$$

$$- k\lambda^m(\lambda^{n-1}w_{t-n+1} + \ldots + w_t)],$$

which corresponds to equation (8) for the case m=n=1, if we put k and 1-k equal to 1/2.

A moderately simple expression for the variance of the residual disturbance, say ε, in this equation may be derived. Directly from (38) we have

$$\mathrm{var}(\varepsilon) = (1-k+k\lambda^{m+n})^{-2}\ [(1-k)^2(\lambda^{2(m-1)} + \lambda^{2(m-2)} + \ldots + 1)$$

$$+ k^2\lambda^{2m}\ (\lambda^{2(n-1)} + \lambda^{2(n-2)} + \ldots + 1)]\sigma_w^2,$$

and recalling that the sum of the first n terms of a geometric progression is $a(1-r^n)/(1-r)$, this simplifies to

(39) $$\mathrm{var}(\varepsilon) = \frac{\sigma_w^2}{(1-k+k\lambda^{m+n})^2}\left[(1-k)^2 \frac{1-\lambda^{2m}}{1-\lambda^2} + k^2\lambda^{2m} \frac{1-\lambda^{2n}}{1-\lambda^2}\right]$$

$$= \frac{\sigma_w^2}{1-\lambda^2} \frac{1-\lambda^{2m} - 2k\ (1-\lambda^{2m}) + k^2 - k^2\lambda^{2(m+n)}}{(1-k+k\lambda^{m+n})^2}$$

$$= \frac{\sigma_w^2}{1-\lambda^2} \frac{(1-2k)\ (1-\lambda^{2m}) + k^2\ (1-\lambda^{2(m+n)})}{(1-k+k\lambda^{m+n})^2} .$$

We need the partial derivative of var (ε) with respect to k, and the value of k for which the derivative is zero. The differentiation and subsequent collecting of terms are lengthy and are omitted, but the result is found to be

$$\frac{\partial \mathrm{var}(\varepsilon)}{\partial k} = \frac{-2\lambda^{2m+n}\sigma_w^2}{(1-\lambda^2)\ (1-k+k\lambda^{m+n})^3}[k(\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n})$$

$$- (\lambda^m - \lambda^{-m})].$$

The value of the partial derivative will be zero only when the expression in brackets is zero, so that the value of k satisfying the first order condition for a minimum is readily seen to be

(40) $$k = \frac{\lambda^m - \lambda^{-m}}{\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n}} .$$

For 1-k one obtains

(41) $$1-k = \frac{\lambda^n - \lambda^{-n}}{\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n}} .$$

If m=n, both k and 1-k reduce to 1/2, and all our previous results hold as a special case.

We have obtained the first of the results indicated at the beginning of this section: the optimal choice of weights k and 1-k depends on the value of λ. An equivalent statement of this result is that the weights depend on the correlation between u_t and u_{t-m-n}, since for a stationary stochastic process the population correlation coefficient will be λ^{m+n}.

We now ask, how do the weights vary when the autoregressive coefficient changes? The answer may be found by the differentiation of k. We have

$$\frac{\partial k}{\partial \lambda} = \frac{1}{\lambda} [(\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n}) (m\lambda^m + m\lambda^{-m}) - (\lambda^m - \lambda^{-m})$$

$$\cdot (m\lambda^m + m\lambda^{-m} + n\lambda^n + n\lambda^{-n})] [\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n}]^{-2}$$

$$= \frac{m(\lambda^n - \lambda^{-n}) (\lambda^m + \lambda^{-m}) - n(\lambda^m - \lambda^{-m}) (\lambda^n + \lambda^{-n})}{\lambda(\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n})^2} .$$

which after some further manipulation yields

$$(42) \qquad \frac{\partial k}{\partial \lambda} \quad \frac{(m-n) (\lambda^{m+n} - \lambda^{-m-n}) + (m+n) (\lambda^{-m+n} - \lambda^{m-n})}{\lambda(\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n})^2} .$$

The sign of $\partial k/\partial\lambda$ is not obvious, and it is useful to have the derivative of the numerator N of the right hand side of equation (42) with respect to λ. Differentiation gives

(43) $$\frac{\partial N}{\partial \lambda} = \frac{1}{\lambda} [(m-n)(m+n)(\lambda^{m+n} + \lambda^{-m-n}) + (m+n)(n-m)(\lambda^{-m+n} + \lambda^{m-n})]$$

$$= \frac{m^2-n^2}{\lambda} (\lambda^{m+n} + \lambda^{-m-n} - \lambda^{-m+n} - \lambda^{m-n})$$

$$= \frac{m^2-n^2}{\lambda} (\lambda^m - \lambda^{-m})(\lambda^n - \lambda^{-n}).$$

Let us consider first the sign of $\partial k/\partial\lambda$ when λ is positive. Since the denominator in equation (42) is positive, the sign of $\partial k/\partial\lambda$ is the sign of the numerator. Now when λ is zero, the numerator is zero, so that its sign when λ is positive is the sign of its derivative with respect to λ, provided that this derivative does not change sign over the relevant range. Thus we consider the terms in parentheses in the right hand side of equation (43). Since λ is less than one and m and n are each at least one, both $\lambda^m - \lambda^{-m}$ and $\lambda^n - \lambda^{-n}$ are negative, and their product is positive. Hence $\partial N/\partial\lambda$ and therefore $\partial k/\partial\lambda$ have a positive sign if m>n and a negative sign if n>m. The case of m>n is that in which y_{t-n} is more distant in time from u_{t-m-n} than from u_t. Since k is the weight given to u_t, a positive value of $\partial k/\partial\lambda$ when m>n implies that as λ rises in value, more weight should be given to the closer residual u_t and less weight should be given to the more distant residual u_{t-m-n}. Conversely, when n>m, the negative value of $\partial k/\partial\lambda$ means that the more distant residual u_t should be given less weight as λ increases.

These results for the sign of $\partial k/\partial\lambda$ carry over when λ is negative, provided that either m or n is an even integer, and the other is odd. With λ negative the denominator in (42) is negative, and equation (43) may again be used to determine the sign of the numerator. Now $\lambda^m - \lambda^{-m}$ will be negative when m is even and positive when m is odd; similar results hold for $\lambda^n - \lambda^{-n}$ when n is even and odd. Hence $(\lambda^m - \lambda^{-m})(\lambda^n - \lambda^{-n})$ will be negative, and since λ is negative, $\partial N/\partial\lambda$ will be positive if m>n and negative if n<m. But this conclusion, the same that was reached about the sign of $\partial N/\partial\lambda$ when λ is positive, has opposite implication for the numerator of $\partial k/\partial\lambda$. For if, as λ increases toward zero, $N(\lambda)$ increases toward $N(0)=0$, $N(\lambda)$ must be negative for $\lambda<0$, and if $N(\lambda)$ decreases toward $N(0)=0$, $N(\lambda)$ must be positive. Reversing these signs because of the negative denominator in (42) we again conclude that $\partial k/\partial\lambda$ is positive or negative according as m>n or n>m. Peculiar results follow, however, if m and n are both even or both odd. The product $(\lambda^m - \lambda^{-m})(\lambda^n - \lambda^{-n})$ is then positive and the conditions determining the sign of $\partial k/\partial\lambda$ are the opposite of those previously obtained.

Let us now turn to the question of estimating the coefficient of the weighted sum of disturbances in (38), the estimating equation for the unknown value y_{t-n}. Using (40) and (41) to remove k and 1-k from the coefficient of $ku_t + (1-k)u_{t-m-n}$ found in equation (38), we obtain

$$(44) \qquad \frac{\lambda m}{1-k+k\lambda^{m+n}} = \frac{\lambda^m(\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n})}{\lambda^n - \lambda^{-n} + (\lambda^m - \lambda^{-m})\lambda^{m+n}}$$

$$= \frac{\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n}}{(\lambda^n - \lambda^{-n})\lambda^{-m} + (\lambda^m - \lambda^{-m})\lambda^n}$$

$$= \frac{\lambda^m - \lambda^{-m} + \lambda^n - \lambda^{-n}}{\lambda^{m+n} - \lambda^{-m-n}} .$$

Substituting the autoregressive relation (5) into itself m+n-1 times and replacing disturbances with residuals, we obtain

$$(45) \qquad e_t = \lambda^{m+n} e_{t-m-n} + w_t^{\circ},$$

where w_t° denotes the new disturbance. This equation may be used to obtain a least-squares estimate of λ^{m+n} from which estimates of the values λ^{-m-n}, λ^{m}, λ^{-m}, λ^{n}, and λ^{-n} required in equation (44) may be obtained. As was the case for unit time intervals, the estimated coefficient will be consistent but not unbiased.

Discussions of the sample variance of this coefficient and of the forecast variance of y_{t-m-n} would lead to very complicated algebraic expressions but to no new concepts. The procedures used for the case m=n=1 remain valid. The largest task in evaluating the forecast variance is that of estimating the residual variance var(ε) as given by equation (39). Equations (40), (41), and (45) may be used to estimate k, 1-k, and the various functions of λ. The remaining factor needed is an estimate of σ_w^2. The disturbance in equation (45) may be written

$$w_t^{\circ} = w_t + \lambda w_{t-1} + \lambda^2 w_{t-2} + \ldots + \lambda^{m+n-1} w_{t-m-n+1}$$

and has the variance

$$\mathrm{var}(w_t^{\circ}) = \frac{1 - \lambda^{2(m+n)}}{1 - \lambda^2} \sigma_w^2 .$$

Hence an estimate of σ_w^2 may be constructed as a function of estimated values of $\mathrm{var}(w_t^\circ)$ and λ^{m+n}, but it may be noted that the factor $1 - \lambda^2$ cancels if this expression is substituted into equation (39).

We conclude our discussion of the case of interpolation with $m\geq 1$ and $n\geq 1$ by noting a significant result that follows from equation (45). If $m+n$ is an odd integer, then the sign of λ^{m+n} will be the sign of λ. Hence, in this case, estimates of y_{t-n} may be obtained without making an *a priori* assumption about that sign. The time interval $m+n$ will usually be given to the user of data, however, and both even and odd values will occur in practice.

6. Summary

The results of this chapter may be summarized as follows. Constructors of economic statistics have generally not availed themselves of regression methods that might have been used to solve problems of interpolation. Those applications of regression methods that have been made involve generally inappropriate assumptions about the autocorrelation of disturbances. We have considered the interpolation problem that arises when vector observations are available for a cross section at three successive points in time, except that y_t is missing for all elements of the cross section at the intermediate point in time, and must be estimated. Discussion of this problem has been restricted to the case in which y_t could be treated as the dependent variable in a regression analysis with first-order autoregressive disturbances.

It was found that the variance of the disturbance in the estimating equation for y_t could be reduced by adding a new variable--a weighted average of the disturbances associated with the complete observations--to the list of independent regressors. The coefficient of the new variable was determined, and found to be a function of the autoregressive coefficient λ and, except when m=n, of the time intervals m and n separating the observations. In addition, the weights for the disturbances that minimized the residual variance in the estimating equation for y_t were determined, and these weights were also found, in general, to be functions of λ, m, and n. These findings led to the recommendations that the regression coefficients for the observed variables should be obtained by least-squares estimation of $y_t = X_t'B + u_t$ using the pooled complete observations, and that a weighted average of the residuals should be used as a proxy for the required weighted average of disturbances in estimating the missing values of y_t. By regressing the computed residuals from the later cross section on those for the earlier cross section, a least-squares estimate of a function of λ was obtained that could be used in evaluating the weight functions and the coefficient of the weighted residuals. All of the estimators so obtained were shown to be consistent, but no other desirable properties for the estimators were established.

Even the property of consistency may break down if the disturbances are heteroscedastic, but detailed analysis of the ways in which heteroscedastic disturbances arise and the modifications in computational procedures that would preserve consistency was provided. Finally, it was shown that, although the calculations might be extensive, the asymptotic sampling variance of the coefficient of weighted residuals and forecast variance for y_t may be determined, thus providing measures of the reliability of the procedures we have described.

CHAPTER FOUR

MATHEMATICAL METHODS FOR COUNTY INCOME ESTIMATION II: SITUS ADJUSTMENT AND MISSING VALUES

Whenever the data which are available to a statistician differ in important respects from his needs, questions arise as to the best way to use that data. The statistician wants his procedures for utilizing data to be reasonable ones, but this requires that he have some notion of the relation between the observations he has been able to obtain and the observations he would like to have obtained. The stronger the theoretical bridge that can be constructed between the two sets of variables, the easier and more satisfying the task of designing reasonable statistical procedures.

In the preceding chapter we discussed the important case in which the proper variables are observed, but the observation dates differ from the desired dates. It was suggested that a certain two-equation linear stochastic model would often be appropriate when this type of problem arose in county income estimation, and the estimation of this model was investigated. A different class of problems arises when the desired observations are missing for all years, and the data which must be used either reflect a significant difference in definition or measure an entirely different but presumably related variable. Two such problems are considered in the present chapter, and a satisfactory solution to each is necessary if reliable county income estimates are to be obtained. One of these problems arises from the definitional discrepancy between earnings reported "where earned," which characterizes much of the primary data related to county income, and earnings "where received," which is the appropriate concept for a definition of the income of persons. This is the problem of situs adjustment. The other problem to be considered in this chapter arises because of frequent missing values in important sources of county wage and salary data. The missing value problem is that of supplying suitable values in these cases. The results of empirical work relating to each of these problems will be presented. Although further work remains to be done, these results lend support to the procedures developed in this chapter.

1. A Linear Programming Approach to Situs Adjustment

Of all the problems connected with county income estimation, the problem of situs adjustment has had the most serious effects on the quality and usefulness of previous work. The measurement of income originating in the county rather than income received by residents of the county is a common defect of county income data, occurring in the major sources for wages and salaries, non-farm proprietors' income, contributions to social insurance funds, and various components of other labor income and property income. To the extent that previous workers have attempted at all to make an adjustment for place of residence, these attempts almost always have been limited to wages and salaries. The method of situs adjustment described in this section can be applied to all situs problems that reflect commuting between counties--in effect, the situs problems of all income components except those derived from real property. For concreteness, however, our discussion will be in terms of the adjustment of wages and salaries.

A rough indicator of the importance of the situs problem in estimating wages and salaries by county is the number of employed persons working outside their county of residence, as reported in the 1960 Census of Population. For the United States as a whole, 14 per cent of the work force was reported to work outside its county of residence. For individual counties the percentage was often much higher: Du Page County in Illinois (near Chicago) reported 56 per cent of its work force employed outside the county, and a Nebraska county (Dakota) near Sioux City, Iowa, reported 45 per cent. In Iowa, Warren County (near Des Moines) reported a commuting rate of 34 per cent.[1] While commuting across a county line in one direction is always partially offset by commuting in the other, these magnitudes reflect sufficient net commuting between primarily residential areas and employment centers to make unadjusted place of work wages in many instances an unacceptable indicator of the wages of county residents.

In all cases in which situs adjustments have been made on county income estimates, these adjustments have been based on either direct or indirect evidence of the amount of commuting between counties. A method

[1]U. S. Bureau of the Census, County and City Data Book, 1962 (Washington: U. S. Government Printing Office), passim.

of situs adjustment will be recommended in this section which uses indirect evidence generated by a simple economic model of the commuting process. We shall begin with a brief survey of the methods of situs adjustment that have been used in earlier county income studies. Next the model of intercounty commuting and its underlying assumptions will be presented. Finally, we discuss the use of the model for situs adjustments.

Methods of Situs Adjustment Used in County Income Studies

Only a few county income studies have attempted to convert wage and salary estimates to a place of residence basis. We shall first review methods used in adjusting estimates for Kansas, Georgia, Kentucky, and Maryland, which were based on direct measurements of the extent of commuting, and then consider the adjustments made in the county income studies for Pennsylvania, New York, and Illinois, which rely on indirect evidence of commuting. All of these efforts leave much to be desired. When costs of the estimating procedures are taken into account, refinement of the indirect methods of New York and Illinois appears to be the most promising approach to the problem.

The 1960 Census of Population contains the only body of data on intercounty commuting which are national in coverage. Published tables report, for each county, the number of residents who work elsewhere, but not the county to which they commute or their distribution by industry or occupation. These data are thus inadequate for making a situs adjustment. However, tabulations of Census commuting statistics which provide county of employment by county of residence may be obtained from the Census Bureau on a contract basis. Darwin Daicoff used such a tabulation to make an adjustment to county of residence in his estimates of county income for Kansas. Daicoff distributed place of work wages and salaries to counties of residence in proportion to the share of employed persons working in a given county who resided in those counties. There was no disaggregation by industry, and the same commuting factors were used for each year from 1950 through 1964.[1]

An ambitious and able attempt to obtain the county data needed to adjust county wage and salary estimates for residence on an industry basis was made by John Fulmer for the state of Georgia.[2] A random sample of Georgia firms was selected that was stratified by county and industry. Mailed questionnaires and an intensive follow-up resulted in responses from approximately 6,000 firms, and these accounted for 63 per cent of nonagricultural employment in the state. In spite of the enormous amount of data collected, however, the sampling variability of many county-industry cells remained high. The large sample size had been chosen in part because of the large number of Georgia counties--159, more than any other state except Texas. Using a six industry-group classification, the standard errors of estimate of county-industry cells

---

[1]Daicoff, op. cit., p. 53.

[2]John L. Fulmer, Analysis of Intercounty Commuting of Workers in Georgia (Atlanta: Engineering Experiment Station, Georgia Institute of Technology, 1958).

were ten per cent or less only for the 5 largest population centers. Manufacturing and non-manufacturing commuters could be distinguished at this level of reliability for 53 counties, and for 44 counties, a ten per cent standard error was exceeded with no disaggregation by industry. Consequently, the procedure used for adjusting wage and salary estimates to county of residence incorporated varying levels of industrial detail.[1] It might be argued that a ten per cent standard error is too stringent a level of precision, or that for a state with fewer counties, the results of a comparable survey would be more favorable. Nevertheless, the Georgia experience suggests that enormous costs may be required to obtain useful data on intercounty commuting by industry.

John Johnson's earlier income estimates for Kentucky also made situs adjustments on the basis of survey results. However, in that study, the sample chosen was a nonrandom one based on judgment, and the data were collected by personal interviews with employers.[2] A more resourceful methodology was adopted by the Bureau of Business and Economic Research at the University of Maryland, which combined limited surveys with other types of data. Situs adjustment for two counties was based on a government report on the residences of

---

[1]John L. Fulmer, Analysis of Georgia Personal Income Payments, by Counties (Atlanta: Engineering Experiment Station, Georgia Institute of Technology, 1959), p. 7.

[2]John L. Johnson, op. cit., p. 146.

federal employees working in Washington, D. C. For two employment centers, including Baltimore, the situs adjustment was based on traffic survey data collected by the state highway commission. For two remaining major employment centers, situs adjustments were made on the basis of a specially conducted sample survey of manufacturing firms.[1] This strategy probably strikes a good balance between cost and reliability in the Maryland case, but the investigators were more fortunate than most in the data available to them. Even so, it falls far short of the goal of situs adjustments for all counties by industry.

Given these disappointing results from expensive direct data on commuting, the question arises as to how much might be inferred about commuting using already available data on employment alone. One alternative is the procedure adopted by the Department of Internal Affairs of the Commonwealth of Pennsylvania.[2] Place of residence employment as reported in the 1960 Census of Population was scaled so that the state total equaled state employment covered by unemployment insurance (a place of work figure) for each of eight major industrial groups. Estimated place of work wages and salaries for counties were then multiplied by county ratios of scaled census employment to covered employment. The 1960 employment ratios were applied to county wage and salary estimates for all years back to 1929. In addition to

---

[1] Personal Income in Maryland Counties, 1951-1955. (Studies in Business and Economics, Vol. X, No. 4; College Park: Bureau of Business and Economic Research, University of Maryland, 1957), pp. 10-11.

[2] Pennsylvania, Department of Internal Affairs, op. cit., p. 84-85.

neglecting changes in commuting patterns over time, this procedure neglects differences in wage rates between counties which are employment centers and those which are predominantly residential. It does, however, lead to a situs adjustment for each county on an industry basis.

A method of situs adjustment developed by the New York State Department of Commerce[1] and adopted in estimating county income in Illinois[2] attempted to meet the problem of county differences in average annual earnings that arises in the Pennsylvania method. The New York method grouped the counties of the state into a set of multi-county regions that each contained a major employment center and a hinterland. Within each region, employment by place of work and place of residence were compared. For the employment center, place of work wages and salaries were multiplied by the ratio of place of residence to place of work employment just as in the Pennsylvania method. However, the New York method completed the adjustment process by assigning the remaining wages and salaries originating in the employment center to the other counties of the region in proportion to each county's excess of place of residence over place of work employment. The chief drawback of this procedure is the arbitrariness with which the multi-county regions must be defined in

---

[1] Personal Income in Counties of New York State, 1948-1957, A supplement to the December, 1958, issue of the New York State Commerce Review.

[2] Scott Keyes, Felix C. Rogers, and Wallace E. Reed, op. cit., p. 5.

practice. It must be assumed that there is no commuting between regions, and thus, that all of the commuting workers in a predominately residential county can be assigned to a single employment center. Moreover, differences in place of work and place of residence employment for the region as a whole are ignored. The county is too large a geographic unit for the assumptions underlying the New York procedure to hold in general.

Nevertheless, the New York method is in several respects the most attractive procedure for situs adjustment that we have discussed. It uses data that are readily available, and the adjustments can be carried out separately for each industry. The higher wage rates of the major employment centers may be taken into account, and perhaps most important, the method focuses on the pattern of commuting between an employment center and its hinterland. We need, however, a procedure for defining the relevant multi-county labor market areas which is both more flexible and less arbitrary. It should be possible to divide the commuting work force residing in a county among more than one employment center, and it should be possible to find an economic basis for linking counties of residence and employment.

A Model of Inter-County Commuting

There does not seem to be any precedent in the literature for discussing commuting patterns over large areas which include a number of centers of employment. Commuting within a single metropolitan area has been investigated from several different points of view, and two studies which have an explicitly economic focus are John Kain's study using Detroit data[1] and a study by John Hamburg and others on Buffalo.[2] Although both studies are concerned with determinants of commuting, Kain's discussion is carried out in terms of an eight-equation structural model, while Hamburg and his associates use a single behavioral assumption. This assumption, from which the authors obtain a computer simulation of a commuting pattern that can be compared with the observed patterns, is that workers and firms are matched in such a way that total travel time is minimized. The locations of residences and plants are grouped into zones, so that the task of finding the commuting pattern which minimizes total commuting time for all workers may be treated as an example of the well known "transportation problem" of linear programming. It is this approach to the commuting problem that will be of concern here.

The assumption that time is minimized (or that cost is minimized) in commuting between counties provides an economic criterion for choosing one among all possible commuting patterns, and one that is simple enough to be used as a basis of situs adjustment. At the same time, arbitrary assumptions that commuting does not take place between certain pairs of counties are largely avoided. In principle,

---

[1]John F. Kain, "A Contribution to the Urban Transportation Debate: An Econometric Model of Urban Residential and Travel Behavior," The Review of Economics and Statistics, XLVI (February, 1964), 55-64.

[2]John R. Hamburg, et. al., "Linear Programming Test of Journey-To-Work Minimization," Highway Research Board Record, No. 102 (1965), 67-75.

the cost minimization criterion could be used to find the least-cost inter-county commuting pattern for a state or an even larger area without introducing a priori restrictions on the commuting pattern, although in practice some restrictions might be necessary because of the size of the computer to be used and other considerations. The realism of the least-cost solution is increased if workers are stratified into meaningful groups. In the Buffalo study, separate commuting patterns were found for white and nonwhite workers, and for drivers and nondrivers. For simulation of intercounty commuting, workers should be stratified by industry, and the finest industry classification available should be used.

We will first set out the bare bones of the inter-county commuting model, and then provide the additional interpretation that is necessary to evaluate and implement it. Our basic assumptions are:

1) There exists a given distribution of households providing employees (to a certain industry) and a given distribution of firms providing employment (in the industry).
2) These two distributions may be represented by a collection of points, each associated with a particular supply of labor and a particular demand for labor. This assumption is equivalent to supposing that all economic activity in a county takes place at a single point.
3) The total supply of labor and the total demand for labor, taken over all points, are equal.

4) There is a cost of commuting between each pair of points.

5) The commuting pattern for the industry is the one that minimizes the total cost of commuting. This assumption states in effect, that the labor market matches jobs and men in a way that is economically efficient, locational decisions of households and firms being given.

On the basis of these assumptions we can provide a mathematical formulation of the problem: find the intercounty commuting pattern that minimizes the total cost of travel between counties. The total travel cost, Z, between counties is equal to the number of workers x_{ij} who live in county i and work in county j, times c_{ij}, the travel cost from county i to county j, summed over all counties. In symbols

$$(1) \qquad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}x_{ij},$$

where n is taken to be the number of counties. The travel costs c_{ij} between counties are assumed to be known. The numbers of commuters x_{ij} between counties are unknown and are to be chosen in such a way as to minimize Z. The x_{ij}, however, must satisfy several sets of side conditions, or constraints. First, the number of workers who live in a county, say the ith, is known, and the sum of the workers residing in county i and working in the various counties (including those who work in their county of residence) must equal that number. We can express the corresponding restrictions for each county on the x_{ij} by a set of constraints

$$(2)\qquad \sum_{j=1}^{n} x_{ij} = a_i, \qquad i = 1, 2, \ldots, n$$

where $a_1, a_2, \ldots, a_n$ are the number of workers living in each of the counties. Similarly, the number of persons employed in a given county, say the jth, is assumed known, and the sum of the persons from each county (including j) who work there must equal the number employed in county j. This type of constraint on the x_{ij} may be expressed by the equations

$$(3)\qquad \sum_{i=1}^{n} x_{ij} = b_j \qquad j = 1, 2, \ldots, n,$$

where b_j is employment in county j by place of work. Finally, the x_{ij} must satisfy the obvious but mathematically necessary conditions that the number of persons commuting from county i to county j not be negative. In symbols, we require

$$(4)\qquad x_{ij} \geq 0 \qquad \begin{array}{l} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, n. \end{array}$$

Altogether, the expression Z of equation (1) whose minimum value is to be found is a function of n^2 unknowns, where these unknowns are to be chosen subject to 2n equality constraints and n^2 inequality constraints.[1,2]

[1]A number of computational methods have been proposed for solving the transportation problem and several of these are discussed, for example, in G. Hadley, Linear Programming, (Reading, Mass.: Addison-Wesley, 1962) Chapter 9.

[2]Only 2n-1 of the 2n equalities (2) and (3) are linearly independent, however, since they must satisfy $\sum_i a_i = \sum_j b_j$ (assumption 3 on page 219).

We can now consider the assumptions of the model, one at a time. The first assumption, that there are given distributions of households and firms that interact in the labor market of a particular industry, serves mainly to distinguish what is explained from what is not explained by the theory. Specifically, we do not seek to explain the locational choices of either households or firms. These choices might be thought of as reflecting long-run decision making, and choices of current employment as reflecting short-run decision making. On the other hand, the locational decisions made by households over time will reflect employment opportunities, and the locational decisions made by firms will take account of the availability of labor. In this way "long run" aspects of labor market behavior will tend to reinforce "short run" considerations and, if decision-making is rational, tend to reduce the economic resources devoted to commuting.

The second assumption, that counties can be treated as if they were concentrated at single points, raises questions which are essentially questions of measurement. The notion that all the population or all the employment is concentrated at a single point in a county should not lead to difficulty if the "center of gravity for the county" is appropriately identified. Various measures of the center of a county, based on the location and size of households and firms associated with an industry could be suggested on the theoretical grounds, but only very rough indicators of the county center can be derived from available data. The most reasonable choice appears to be the geographic center of the largest city or town in the county, with this point used as the center of both residence and employment in all industries. This measure has the advantage of being especially easy to obtain, and as a measure of central tendency it corresponds roughly to the mode.

Intuitively, one would expect two counties to have closer economic ties the closer were the largest towns in each, and it is these ties which the proposed measure of the county center takes into account. The highway distance between county centers may introduce considerable measurement error into the linear programming problem, however, and it would be important to determine the extent to which a "least cost" commuting pattern was affected by these errors. Sensitivity analysis, which examines changes in the solution of the linear programming problem when parameter values change, provides a means of investigating this question.[1] Efficient computational methods for solving a programming problem with a new cost vector after an initial optimizing solution is obtained have been explored. In the commuting model, the sensitivity of the optimizing solution to measurement errors in the costs could be evaluated by choosing a new cost vector in which further errors of measurement have been introduced. This cost vector could be obtained, for example, by adding to each element in the original cost vector a random value chosen from a probability distribution which was felt to reflect the types and magnitudes of measurement errors that were of interest.

---

[1]*Ibid.*, pp. 379-384.

The third assumption, that the number of persons in households who have jobs is the same as the number of persons employed by firms, is an essential feature of the programming model, and one which surely, in a relevant sense, characterizes the real world. For several reasons, however, measured employment in an industry by place of residence and by place of work for a given set of counties should not be expected to correspond. First, differences in the definition of employment, in the date of enumeration, and the method of measurement will produce systematic differences in the measured number employed which do not average out over the entire country and which may lead to larger discrepancies in some areas than in others. Errors of industry classification in the various data sources may be partly systematic. Differences in the definition of industries are a further but much less important source of discrepancy at the level of aggregation for which both place of residence and place of work employment data are available. A further reason for differences in employment totals is that no set of counties form a completely closed labor market. The amount of discrepancy from this source can be reduced by adding to (or possibly deleting from) the list of counties that one would otherwise select for inclusion in the linear programming problem, when a particular county boundary can be identified as a source of difficulty.

A consequence of the third assumption is that, when a least-cost commuting pattern is to be computed, there must first be a careful choice of the set of counties to include. When this choice has been made, the county values of one of the employment measures (place of work or place of residence) must be scaled so that its total matches that of the other. It should be clear that although for a scale factor to be close to one is desirable, this condition would provide a poor criterion for selecting the counties to be included in the transportation problem. The errors involved in choosing a collection of counties will be one of the two types: either cutting off too much of the hinterland around employment centers, so that "true" employment by place of work exceeds "true" employment by place of residence; or including more hinterland around an employment center than it actually draws from so that "true" place of work employment falls short of "true" place of residence employment. Clearly, one does not want to compensate for measurement errors in the data by making errors in the choice of hinterland around major employment centers. Rather, the set of criteria used to choose the counties for the transportation problem should include the stipulation that the discrepancies between employment totals should be of an order of size that could reasonably be attributed to measurement errors alone. A good practice would be to compare percentage discrepancy between the two employment measures for the counties under consideration with the percentage discrepancy in the two measures for the United States as a whole.

The fourth assumption requires that the cost of commuting between pairs of counties is measurable. Driving times between county centers would be a good indicator of this cost, but these are not generally available. Highway miles between county centers is an alternative indicator of commuting cost, and one which can be obtained easily. One might ask whether the subjective cost of commuting is proportional to highway miles, or whether it increases faster than proportionately as the length of trip increases, at least after some point. In the absence of any evidence on this question, the best strategy would appear to be that of assuming proportionality between length of trip and the cost of commuting in determining the commuting patterns used for situs adjustment. However, sensitivity analysis can again be used to determine how critical this assumption is to the linear programming results. In this case one would want to compare the original optimal solution with the solution when, for example, all costs are squared.

The fifth and final assumption of the model is that the commuting pattern determined in the labor market minimizes the total cost of commuting. This assumption implies that the labor market is in equilibrium, in at least one respect, and is subject to all the qualifications that are normally placed on equilibrium assumptions. At most points in time, the equilibrium commuting pattern should provide a tolerable approximation of the actual pattern, but it will involve an underestimate of the total amount of commuting. Fortunately for the estimation of county incomes, the effects of some of the discrepancy between actual and least cost commuting will cancel. Actual commuting can be thought of as the sum of a systematic factor, explained by the linear programming model, and a random factor. The random factor results in (say) some additional commuting from county A to county B, but also some commuting from B to A. (By the nature of a least cost solution, if there is commuting from A to B there will be none from B to A.) It is the difference in these two random quantities, and not their sum, which determines the extent to which the least cost solution is an inaccurate basis for situs adjustment. For situs adjustment one needs to know only the net commuting between pairs of counties, and all flows in the least-cost solution are net.

Although the context and interpretation of the linear programming model are somewhat different when applied to commuting between counties than when applied to intracity commuting, it is interesting to consider the empirical results obtained by Hamburg and his co-workers. Comparison of the time minimizing and average actual commuting times for several classifications of Buffalo workers leads the writers to conclude that their study "does not demonstrate that commuters minimize aggregate travel time," but that "minimization is a potent influence."[1] Nonwhite drivers in the sample had a minimum average commuting time of 7.6 minutes and an actual average commuting time of 10.3 minutes. A much less favorable result was that commuters with incomes under $5,000 had a minimum average travel time of 3.8 minutes, but an actual average travel time of 11.7 minutes. There are two reasons for expecting that the results of our intercounty commuting model

[1]Hamburg _et al._, _op. cit._, p. 74.

would be more realistic. In the first place, a stratification of workers by industry provides a finer classification of workers than that used in the Buffalo study, and this factor would reduce the number of incorrectly identified very short trips. Secondly, one would expect distance to be a greater deterrent to commuting, the larger the geographic area considered. While at least a few workers in any part of a city may work in any other part, cost considerations limit commuting over distances greater than fifty miles virtually to zero. Fulmer found that 95.3 per cent of the workers employed in the six largest Georgia cities lived within 30 miles of these cities.[1]

Use of the Model for Situs Adjustment

To estimate wages and salaries by place of residence, it is necessary to distribute estimated wages and salaries by place of work for each county to other counties in proportion to the work force assignments made by the cost minimizing commuting pattern. This approach to situs adjustment has the advantages that it can be implemented to give place of residence wages and salaries in more than one year, and that the situs adjustments can be carried out by industry. With regard to both of these aspects of the approach, however, there are details of procedure that must be specified.

A difficulty with the choice of years for this method of situs adjustment is that place of work and place of residence employment data

---

[1]Fulmer, Analysis of Intercounty Commuting of Workers in Georgia, p. 11.

are generally available for different years. Place of residence employment by industry is reported in the Census of Population for 1950 and 1960. Place of work employment is reported in County Business Patterns, the industrial censuses, and for some states, tabulations from unemployment insurance records. Questions which arise are: (1) How should the employment data be adjusted for comparability before solving the linear programming problem? (2) Should the linear programming problem be solved for the same years as those for which adjusted wage and salary estimates are desired (our 1948, 1953, 1958, and 1963), or do data considerations make it advisable to solve the programming problem for other years and then adjust the estimated least-cost commuting patterns? (3) Finally, if the programming problem is solved for years different from those for which situs adjustments are required, what modifications should be made in the solutions to the programming problem?

The interpolation methods of the preceding chapter might be used to adjust place of work and/or place of residence employment data, but it would generally be difficult to find good related variables which were available at the required frequency.[1] Thus, adjustment of the data for discrepancies in observation dates would normally be made by simple arithmetic or geometric interpolation. If the observation dates for the two types of employment data are distant, then a moderate amount of measurement error will be present in the

[1]Exceptions would be the use of OASI data to interpolate employment as reported in the industrial censuses for manufacturing, wholesale trade, and retail trade, but the County Business Patterns and Census of Population dates match only approximately.

interpolated series. The difference berween the two series, which is taken to indicate the net amount of commuting, will have an average percentage error greater than the percentage errors in either series. As a consequence, serious distortion may be introduced in the cost minimizing commuting pattern. For this reason a policy of solving the linear programming for the years for which comparable employment series can be derived most reliably would be preferable to obtaining least cost commuting patterns for the years for which final wage and salary estimates are to be made. Since place of work employment is reported more frequently, it can be interpolated with less distortion. Hence, 1950 and 1960, the years for which place of residence employment is reported, would be the best years for determining least-cost commuting patterns.

If this procedure is followed, the task remains of translating commuting patterns in the benchmark years into commuting patterns for the years for which they are required. A natural first step is to obtain simple interpolations of benchmark year commuting between all pairs of counties. Suppose that these have been obtained, and that they have been summed by county of work so that interpolated county of work employment estimates are obtained. The resulting employment estimates will differ from reported place of work employment for the same year, or from the best interpolation that can be made for that year, using data that are closer in time. The positive or negative discrepancy may be removed by distributing it among all of the counties which are connected to the given county by commuting, and to that county itself, in proportion to the interpolated commuting flows. This procedure is not the only way in which the discrepancy between the two employment figures could be resolved, but more complicated procedures would probably not make much contribution to the reliability of the estimated commuting pattern.

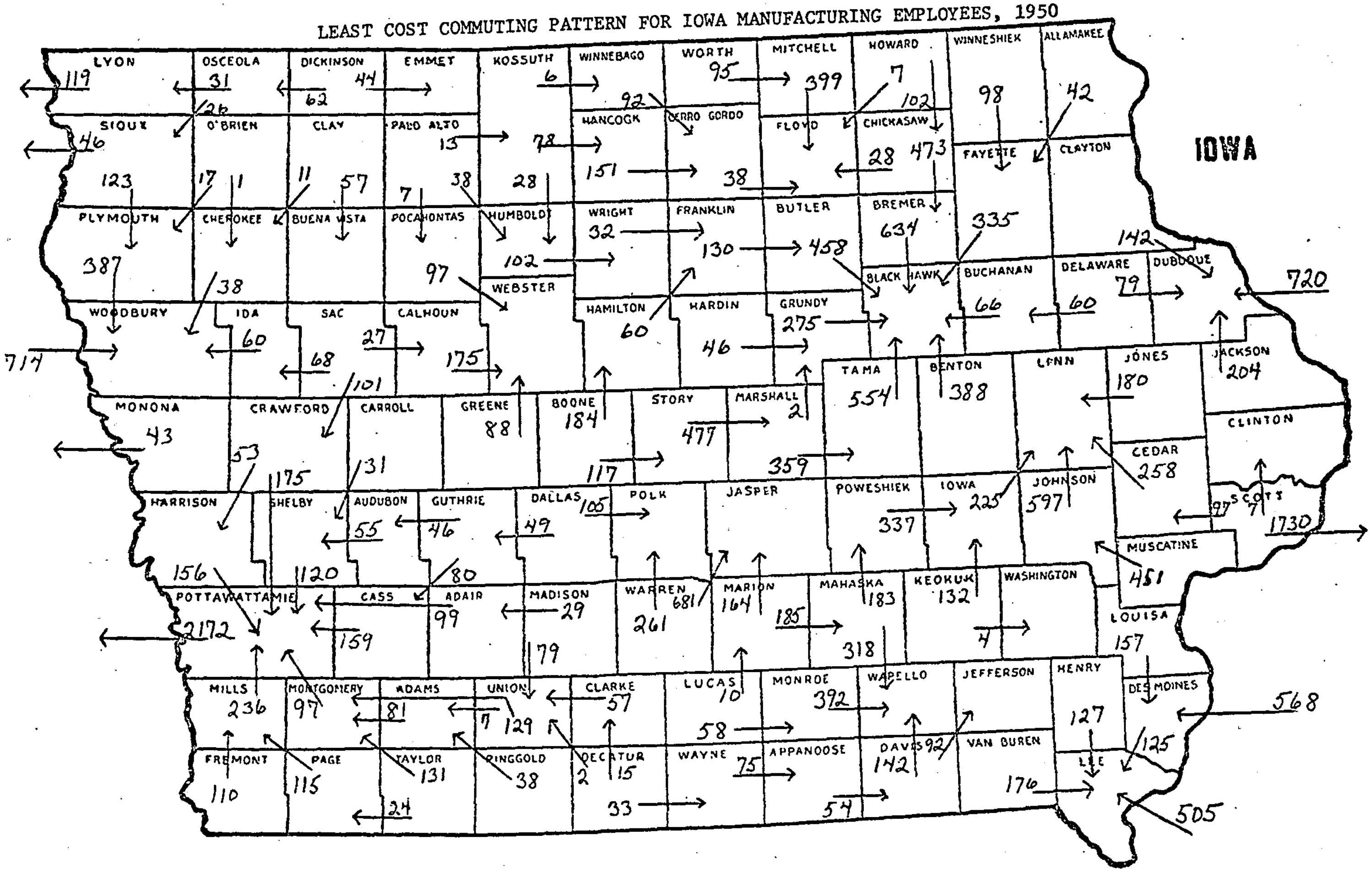
The industrial detail for which commuting patterns can be estimated is limited to the very broad classifications of County Business Patterns: mining, construction, manufacturing, public utilities and transportation (except railroads), wholesale trade, retail trade, finance-insurance-real estate, and services. A commuting pattern cannot be estimated for agricultural services-forestry-fisheries because this classification is not used in the Census of Population, which is the source of place of residence employment data. Wage and salary estimates for this industry and any others, such as the federal civilian sector, which are on a place of work basis, should be distributed to counties of residence in proportion to the sum of the least-cost commuting flows for the eight industries that can be treated explicitly. The combined commuting flows also provide a basis for distributing to county of residence other components of personal income, such as the income of non-farm proprietors, which may be based on data collected on a place of work basis. No adjustments need to be made for industries for which wages and salaries are initially on a residence basis, or for farming, in which intercounty commuting is low and unstructured.

It should be noted that in the industries wholesale trade, retail trade, and services, proprietors make up a significant portion of the labor force. To obtain the labor force on a place of work basis, an estimate of the number of proprietors, based on the number of proprietors reported in the industrial census for each of these industries, should be added to the estimated number of employees. Measurement errors may cause special problems in service employment, and if the discrepancy between total measured place of work and place of residence employment were very great, alternative methods of estimating service wages and salaries by place of work should be considered. These include distributing wages and salaries by place of work to counties on the basis of total estimated commuting for other industries, and estimating wages and salaries in the industry partly on the basis of place of residence employment, perhaps weighted by a measure of the average county wage.

An example of a least-cost intercounty commuting pattern obtained as the solution to a transportation problem is presented in Figure 2. Iowa data were used, and the industry and year chosen was manufacturing in 1950. Since an edition of County Business Patterns covering manufacturing only appeared for that year, the problems of preliminary data adjustment in this case were relatively minor--the OASI data were simply scaled to the Census of Population total. There is no significant interstate commuting across the northern or southern boundaries of Iowa, but major employment centers lie on the eastern and western boundaries. These include Clinton, Burlington, and Keokuk,

Figure 2

LEAST COST COMMUTING PATTERN FOR IOWA MANUFACTURING EMPLOYEES, 1950



Iowa and Rock Island, Illinois, on the east; and Sioux City, Iowa; Sioux Falls, South Dakota; and Omaha, Nebraska, on the west. In order to obtain a relatively self contained labor market area, 18 out-of-state counties were selected along the eastern and western borders of the state, making 117 counties in all.[1]

Data on the highway miles between the largest city or town in each county were taken from state and national sources.[2] Originally it was hoped that allowed commuting could be restricted to contiguous counties. Distances between counties that did not touch were assigned an arbitrarily high value. It was found, however, that no feasible solution existed that satisfied this constraint, because of the very large number of workers required by Omaha. Adding more counties in Nebraska was considered as a solution to this problem, but that type of adjustment was found inadequate and was rejected. Three commuting routes between non-contiguous counties were then introduced: from Crawford to Pottawattamie, from Adair to Pottawattamie, and from Union to Montgomery. With these modifications feasible solutions existed, but because of the highway structure, the optimal solution was not unique. A further modification was made by arbitrarily adding

---

[1]The counties selected were, in Illinois: Carroll, Hancock, Henderson, Henry, Jo Davies, Mercer, Rock Island, and Whiteside; in Nebraska: Burt, Cass, Dakota, Douglas, Sarpy, Thurston, and Washington; and in South Dakota: Lincoln, Minnehaha, and Union.

[2]Mileage between Iowa cities was taken from Mileage Guide of Iowa, (Emmetsburg, Iowa: McNamara's Moving and Storage, 1954). Mileage involving other cities was taken from the Rand McNally Road Atlas.

two miles to the cost of commuting between non-contiguous counties. With these specifications there was a unique optimal solution, and this is the solution shown in Figure 2. Numbers of workers commuting from counties of residence to counties of employment are indicated alongside the arrows, which correspond to selected routes.

Perhaps the most striking feature of Figure 2 is the drift of workers toward Omaha in the southwest portion of the state. Four other 1950 SMSA's show large amounts of in-commuting--Sioux City (Woodbury County), Des Moines (Polk), Waterloo (Black Hawk), and Cedar Rapids (Linn). The remaining 1950 SMSA, Davenport-Rock Island, does not show in-commuting from the Iowa side, but it does show a large number of manufacturing workers living in Iowa (Scott County) and commuting to the Illinois portion. The recently published Department of Commerce estimates of personal income for SMSA's (which include none in Iowa) make no situs adjustment other than those incorporated in the personal income estimates for states. In defending the absence of a further residence adjustment, Graham and Coleman state that "when the counties of the various SMSA's are combined, the differences between place of work and place of residence are eliminated, or at least minimized, and the income aggregate, therefore, measures the total income received by persons in the area, SMSA on either a 'residence' or a 'where-worked' basis."[1] The commuting pattern in Figure 2 does not support this assertion. In addition, large amounts of in-commuting are shown for other manufacturing centers in the state.

[1]Graham and Coleman, op. cit., p. 44.

An unfavorable result from this commuting pattern experiment was the large amount of computer time required for solution. The calculations were carried out on an IBM 7044 computer using a computer program believed to be highly efficient.[1] However, running time for the 117 county problem as finally specified was 3 hours and 45 minutes. By contrast, running times for a number of 30 county trial problems were in the neighborhood of one minute. This suggests that for a state with as many counties as Iowa, considerable savings would result from a preliminary regionalization of the state into three or four areas. The accuracy costs of such a procedure have not yet been investigated.

2. A Method for Employment Estimates Based on the Lognormal Distribution

The Department of Commerce publication County Business Patterns is a major source of wage and salary data and is potentially of considerable value in county income estimation. In addition to first-quarter payrolls, this publication reports employment and the size distribution of firms by employment size class, all by industry. We saw in Chapter Two that, unfortunately, values of many county-industry cells are not reported, in order to avoid disclosure of data for individual

---

[1]The algorithm selected is given in James Munkres, "Algorithms for the Assignment and Transportation Problems," Journal of the Society for Industrial and Applied Mathematics, 5 (March, 1957), 32-38. Munkres' algorithm is simpler than that of Dantzig, Ford, and Fulkerson (see Hadley, op. cit., pp. 257-266) but similar to it in that solution begins by solving of the dual. The computer programming for the IBM 7044 was done by Burton Gearhart.

firms operating in a county where the number of firms in a given industry is small (or where one large firm dominates the county statistics for the industry). The size distribution of firms in an industry and county, however, is always reported. In this section we develop a method for supplying the missing employment and payroll values based on the size distribution of firms.

Similar missing value problems occur in the Census of Manufacturing and the Census of Business. Missing values in the Census of Manufacturing could be supplied in the way described for County Business Patterns, since this source also provides the distribution of firms by size class. A more economic procedure, when both industrial census and OASI data are used, would be to use OASI values consistently to estimate missing values in the industrial censuses. Since the same disclosure rules are followed in both sources, missing values in the industrial censuses will tend to correspond to missing values in County Business Patterns. Thus for the industrial census values, the methods of this section will apply indirectly.

Because the data relied upon--the employment size distribution of firms--are more closely related to employment than to payrolls, our primary concern will be the estimation of missing employment values. These may be readily converted to estimates of missing values for payrolls by a simple procedure involving two proportional adjustments as follows: First, multiply the estimate of employment by the ratio of county payrolls to county employment, obtaining the payroll that would result if the estimated employees were paid at the average rate for the county. Second, multiply the resulting quantity by the ratio of payrolls in the industry to employment in the industry for the state as a whole. This step makes an adjustment for differences in earnings rates in different industries. Payrolls are thus estimated by the relation

$$\text{Payrolls (county, industry)} = \frac{\text{payrolls (state, industry)}}{\text{employment (state, industry)}} \times \frac{\text{total payrolls (county)}}{\text{total employment (county)}} \times \text{employment (county, industry).}$$

Estimates of missing employment values have an important application in county income estimation in addition to their use in making estimates of payrolls: they are needed in order to have a complete set of employment data for situs adjustment. With missing values in the employment data, the situs adjustment procedures described in the preceding section could not be carried out.

Alternative Uses of Employment Size Data

The idea of using the size distribution of firms to supply missing values has been applied by McCarty, Hook, and Knos to the *County Business Patterns* data for 1950 and 1953.[1] They make employment

[1] Harold H. McCarty, John C. Hook, and Duane S. Knos, *The Measurement of Association in Industrial Geography* (Iowa City: Department of Geography, State University of Iowa, 1956).

estimates by assigning to the firms in each size class an estimate of the number of employees in firms of that size class. For all size classes, except the largest, this number is simply the midpoint of the size class; for the largest size class, all of the employees in the county total for all industries are assigned who are not otherwise accounted for. Discrepancies between total county employment (always reported) and the sum of reported and estimated employment by industry are resolved by making adjustments in the larger size classes.

The difficulty with this procedure is the arbitrariness of the estimates of employment of firms by size class. Not only are these estimates made without reference to any related data, but the midpoint of a size class should be expected to overestimate the average employment in firms in the size class if the employment size distribution of firms, like most other size distributions in economics, is skewed to the right.

Regression analysis could be used to obtain an alternative estimate of average employment in each size class. Let N_k denote employment in county k and E_{ik} the number of firms in the _i_th size class. Then if the coefficients of

$$N_k = a_1E_{1k} + a_2E_{2k} + \dots + a_nE_{nk} + u_k \tag{5}$$

were estimated by least squares, using counties of a state as observations, these coefficients would be estimates of average employment in each of the n size classes. The estimated coefficients would be unbiased if there was no systematic difference between the size distribution in those counties in which employment was reported and in those counties in which it was not. A separate equation could be estimated for each industry and year in which missing values need to be supplied, and in this way, geographic, industrial, and temporal differences in size distribution of firms could be taken into account.

It could be argued that equation (5) is an inappropriate model for the prediction of employment because no meaning can be given to the disturbance term u_k, and that instead of including a disturbance term, the model should specify that the coefficients are random. A practical difficulty with the model is that it is expensive to apply because of the amount of data which must be transcribed and processed.[1] Since County Business Patterns reports 8 employment size classes, there are, again in the Iowa case, 792 observations of independent variables per industry per year, and most of these data are of little further relevance for county income estimation. For this reason it is desirable to have a method which relies on a smaller amount of data.

The requirement of smaller data input can be met if we assume that the size distribution of firms is lognormal, and this leads to the approach to the missing value problem to be recommended here. This approach, while incorporating the lognormality assumption, will be similar in some respects to the method based on linear regression.

---

[1]Data from recent edition of County Business Patterns may be purchased from the Bureau of the Census on magnetic tapes, but even for these years the data must be sorted by industry and for complete observations.

Estimated county employment will continue to be a function of state average employment in firms of each size class and county numbers of firms in each size class. Now, however, the additional assumption that the number of employees a firm has is a random variable distributed lognormally allows one to estimate average employment of firms by size class using state totals only. Further, the use of the lognormal distribution leads to computational methods that can be carried out graphically, thus reducing the burden of calculation still more. These graphical methods also can provide a visual test of the assumption of the lognormality of employment, and hence they permit evaluation of the suitability of the methods developed here in particular applications. We shall first indicate the theoretical basis for using the lognormal distribution in the missing value problem, and then show how graphical methods can be used to facilitate the computations.

The Lognormal Distribution and Employment

The lognormal distribution has the form:

$$F(x) = \int_0^x \frac{1}{x\sigma\sqrt{2\pi}} \quad \exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\, dx\right]$$

The variate x--which in our application denotes employment for an individual firm--takes on only positive values and the distribution is skewed to the right. The notation μ, σ^2 for the parameters is suggested by the fact that if a new variable is defined by $y = \ln x$, y is distributed normally with mean μ and variance σ^2. The lognormal distribution was used to estimate the size distribution of firms as measured by number of employees by the French economist R. Gibrat in 1931.[1] More recent work on the size distribution of firms has typically been based on other measures of size--in particular, dollar value of net assets and measures of capacity--but has often used the lognormal distribution. Richard Quandt has shown that the lognormal distribution provides a good description of the size distribution of firms, and fits data for most industries at least as well as other widely used distribution functions.[2] Hart and Prais have argued in favor of using the lognormal distribution to describe the size distribution of firms on the grounds that it provides good fits to the data, and that this distribution has properties that make it mathematically tractable.[3]

Simon and Bonini have argued that the Yule distribution, defined as $KB(N, \rho + 1)$, where $B(N, \rho + 1)$ is the beta function, K is a normalizing constant, and ρ is a parameter, should be preferred on

---

[1]R. Gibrat, Les inegalités economiques (Paris: Libraire du Recueil Sirey, 1931). Cited in J. Aitchison and J. A. C. Brown, The Lognormal Distribution (Cambridge: Cambridge University Press, 1957), p. 101.

[2]Richard E. Quandt, "On the Size Distribution of Firms," American Economic Review, LVI (June, 1966), 416-432.

[3]P. E. Hart and S. J. Prais, "An Analysis of Business Concentration," Journal of the Royal Statistical Society, Series A, 119 (1956, Part 2), 150-191.

theoretical grounds to the lognormal distribution.[1] They point out that the good fits obtained from the lognormal distribution are often rationalized as support for the "law of proportional effect," which states that over time the members of the population in question (here firms) experience random changes in size which are proportional in magnitude to their size at the moment change occurs. It may be shown that if the population experiences size changes of this type, its size distribution approaches the lognormal form as time passes regardless of the shape of the original size distribution. This result, however, does not allow for the entry of new firms. If new firms are created at a constant rate over time, then the result at the limit of the law of proportional effect is given by the Yule distribution.

Our choice of the lognormal rather than the Yule distribution is based on (1) the good fits obtained with size data for firms in previous studies, and (2) its mathematical convenience. The latter consideration is of significance because of the availability of certain theorems which facilitate evaluation of the coefficients a_i in equation (5) when a lognormal size distribution of firms is assumed. These theorems are well known, and some of the geometric implications are widely appreciated. For example, it is known that by suitable transformations of both the cumulated density and the random variate,

---

[1] H. A. Simon and C. P. Bonini, "The Size Distribution of Firms," American Economic Review, XLVIII (September, 1958), 610-611.

the distribution function may be transformed into a straight line, and this fact has been utilized in the design of commercially available lognormal probability paper to facilitate the graphing and analysis of empirical distributions. However, the geometrical implications of some results needed for our problem do not seem to have been fully realized.

Our first task is to relate the lognormal distribution to the coefficient a_i, which has been interpreted as the average number of employees of firms in the ith size class. This may be done as follows: Let the equation (5) be summed over counties. Setting $\sum_k N_k = N$ and $\sum_k E_{ik} = E_i$, and omitting the disturbance terms, one obtains

$$(6) \qquad N = a_1E_1 + a_2E_2 + \ldots + a_nE_n,$$

an equation for total state employment (in some industry). Let f(x) be the lognormal density function for the employment size of firms and let $\sum_i E_i = E$, the total number of firms in the state. We then have, as an identity,

$$N = E \int_0^x x \cdot f(x)dx.$$

Denoting by x_i the employment level forming the upper bound of the ith size class, the right hand side of this identity may be expanded so that it contains a term for each of n classes:

$$(7)\qquad N = E_1 \frac{1}{E_1/E} \int_0^{x_1} x \cdot f(x)\, dx + E_2 \frac{1}{E_2/E} \int_{x_1}^{x_2} x \cdot f(x)\, dx + \ldots$$

$$+ E_n \frac{1}{E_n/E} \int_{x_{n-1}}^{\infty} x \cdot f(x)\, dx.$$

From comparison of equations (6) and (7), a_i is found to be given by

$$a_i = \frac{1}{E_i/E} \int_{x_{i-1}}^{x_i} x \cdot f(x)\, dx$$

$$= \int_{x_{i-1}}^{x_i} xf(x)\, dx \Big/ \int_{x_{i-1}}^{x_i} f(x)\, dx.$$

This equation may be expanded as

$$(8)\qquad a_i = \frac{\int_0^{x_i} xf(x)dx - \int_0^{x_{i-1}} xf(x)dx}{\int_0^{x_i} f(x)dx - \int_0^{x_{i-1}} f(x)dx} = \frac{F^*(x_i) - F^*(x_{i-1})}{F(x_i) - F(x_{i-1})},$$

where $F(x_q)$ denotes the distribution function of x, the function $F^*(x_q)$ is called the first moment function. Equation (8) shows that an estimate of a_i can be obtained if $F(x_q)$ and $F^*(x_q)$ can be evaluated at $q = i$ and $q = i-1$. It is the evaluation of these four quantities which will be considered graphically.

Computational Method

At this point it is necessary to summarize some of the theory underlying the use of lognormal probability paper, and in particular the possibility of transforming a lognormal distribution into a straight line. The latter possibility follows from a theorem which is expressed in terms of the notion of the quantile of order q of a distribution, defined as the value of a variate such that q times 100 per cent of the distribution is to the left of that value. (Thus, the median is the quantile of order one-half.) The theorem asserts that if x_q and z_q are quantiles of order q of lognormal and standard normal distributions respectively, then[1]

$$\ln x_q = \mu + \sigma z_q.$$

Converting the logarithm to base 10 gives

$$(L_1) \qquad \log_{10} x_q = 0.434\,\mu + 0.434\,\sigma z_q,$$

the form of the relation usually plotted on commercially available lognormal probability paper. The significance of equation (L_1) is that, for properly labeled axes, it provides a linear representation of the lognormal distribution.

---

[1]Proof: Since the natural logarithm of x has the standard normal distribution with mean μ and variance σ^2, $(\ln x - \mu)/\sigma = z$ is standard normal.

Equation (L_1), for a particular choice of μ and σ, is shown in Figure 3, which also provides an illustration of lognormal graph paper. Along the horizontal axis, equal distances correspond to equal units of $\log_{10}x$.[1] However, a scale has been added along the horizontal axis, just as on the standard logarithmic graph paper, that is derived from the transformation

$$(9) \qquad x_q = 10^{(\log_{10}x_q)}.$$

The alternate scale allows one to plot values of x as values of $\log_{10}x$ directly without making the computations that the transformation requires. Along the vertical axis, equal distances correspond to equal units of z. Again, however, an alternate scale has been provided, in this case based on the transformation

$$(10) \qquad q = \int_{-\infty}^{z_q} \frac{1}{\sqrt{2\pi}} \exp\,(-\,t^2)dt,$$

which is the standard normal distribution. Thus, cumulative frequencies can be plotted directly using the scale provided on the vertical axis, without first converting to z. It will be noted that, as the normal distribution requires, a unit change in z leads to a large change in q near z = 0 (q = .5) and to small changes in q when z takes on large positive or negative values. Plotting values of the pair (x,q) on axes

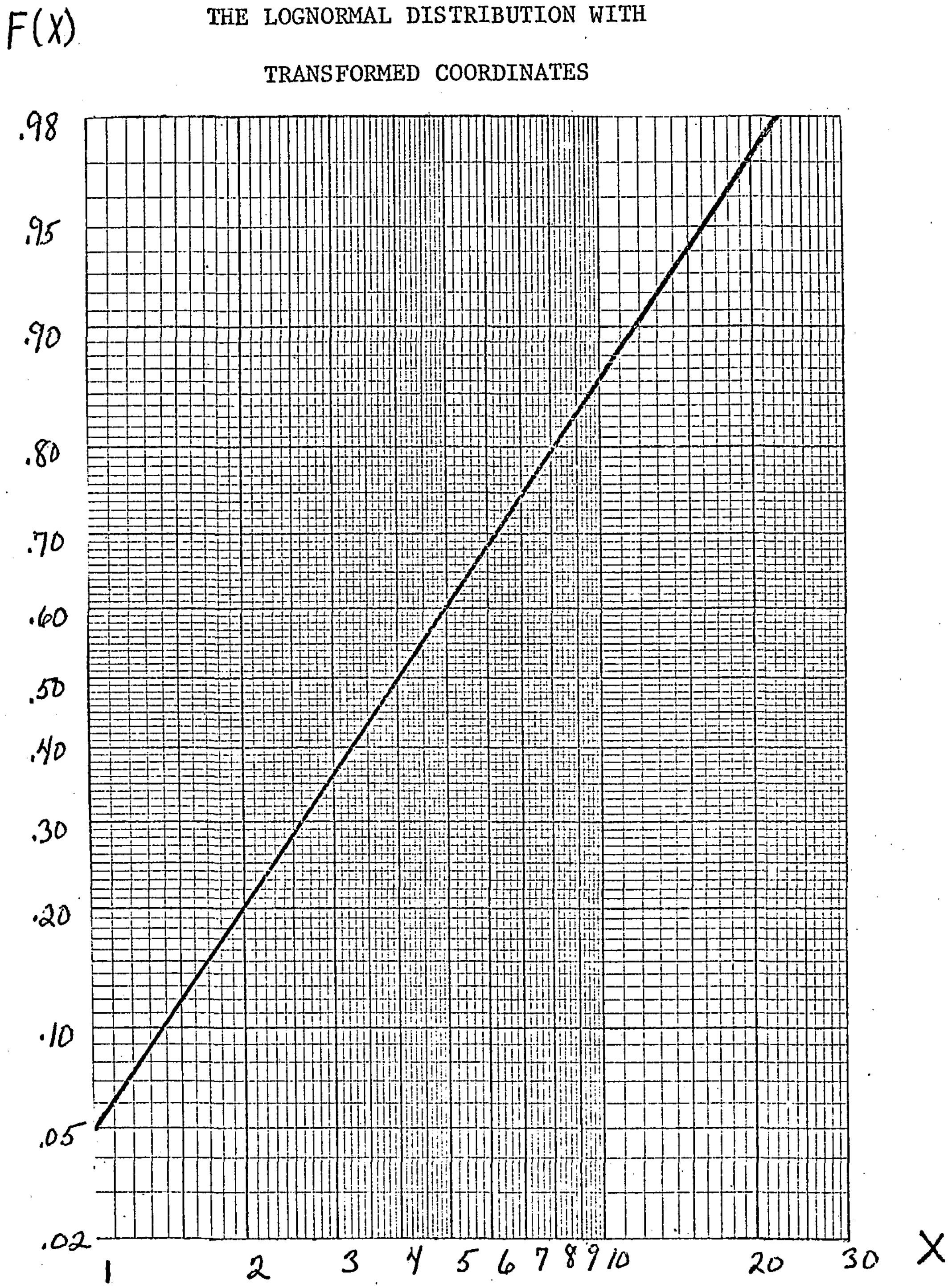
---

[1]It is usual, in working with lognormal probability paper, to reverse the convention that puts the random variate on the x-axis and the cumulative density on the y-axis. (Compare equation (L_1).) In this paragraph, however, we label axes in the familiar way.

Figure 3

THE LOGNORMAL DISTRIBUTION WITH

TRANSFORMED COORDINATES



labeled according to the transformations (9) and (10) is equivalent to plotting the pair $(\log x_q, z_q)$. Thus, by equation (L_1), if x has the lognormal distribution, the set of pairs (x,q) trace out a straight line.

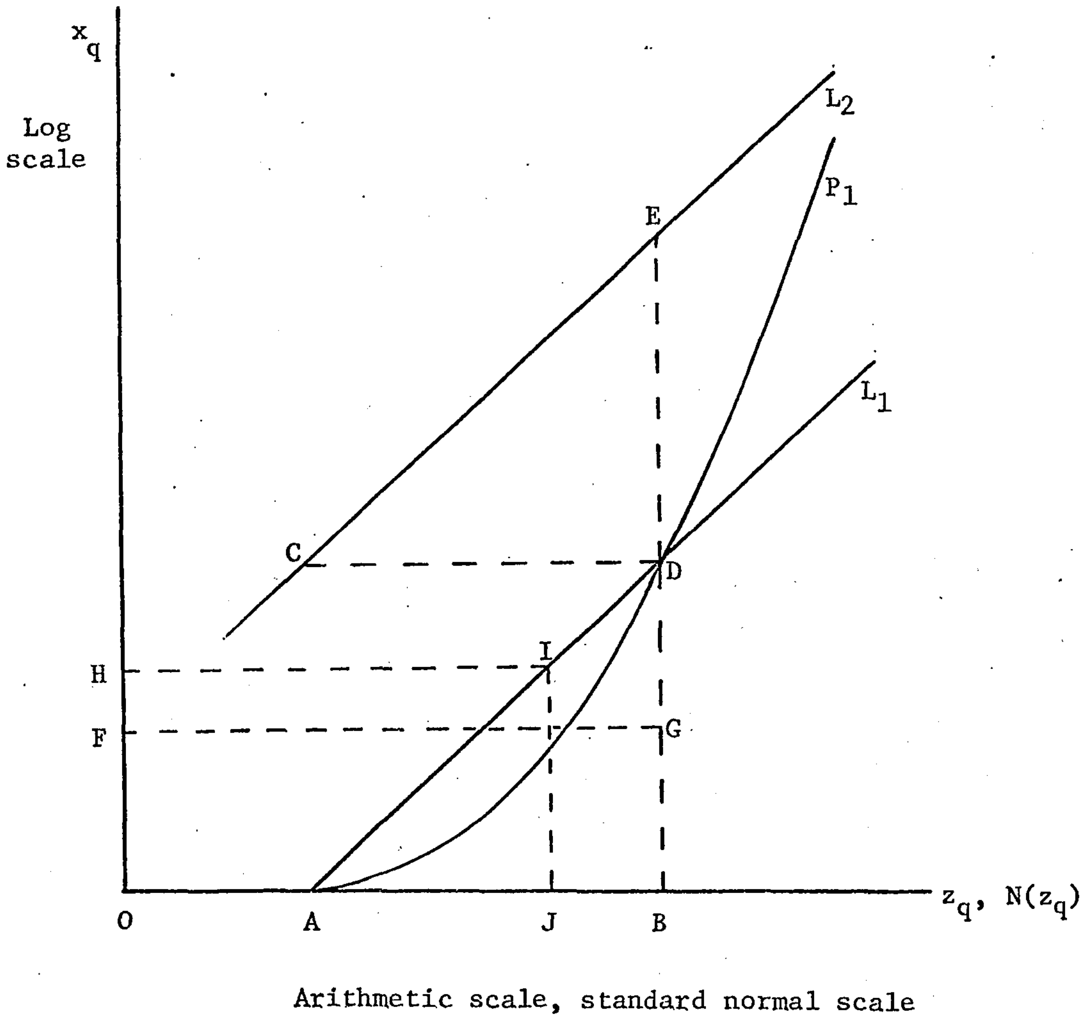
Empirical distribution functions may be plotted, using grouped data, as a series of points (x_i, q_i) where x_i is the upper bound of the ith group, and q_i is the proportion of observations which fall in the ith and all previous groups, that is, the observed cumulative frequency. If the observations are believed to be drawings from a lognormal distribution, a straight line can be fitted freehand to the plotted points, and the extent to which the lognormal distribution approximates the empirical distribution may be judged by how close the points fall to the line. In general, if the data are tabulated in r groups, r - 1 points, corresponding to the upper bound defining each group but the largest, can be plotted. Not as many points can be plotted on commercially available paper, however, if groups are chosen that fall in one of the tails of the distribution.

The graphical evaluation of the parameters a_i will be presented in terms of Figure 4, which gives the constructions that need to be carried out on lognormal probability paper.[1] The axes are the reverse of those in Figure 3, in conformity with equation (L_1). Thus the vertical axis in Figure 4 measures x_q on a log scale with base 10. At point 0, $\log_{10} x_q$ is zero. Two scales are provided on the horizontal axis, an

[1]To see how these constructions appear when drawn on lognormal probability paper, compare Figures 5-8. Graph paper background is omitted from Figure 4 in order to emphasize the geometric argument.

FIGURE 4

CONSTRUCTIONS FOR THE LOGNORMAL
FIRST MOMENT FUNCTION



arithmetic scale and a standard normal scale, which are used to measure the variables z_q and $N(z_q)$ respectively. No significance attaches to the values of the horizontal scale which occur at point 0. We let J be the point at which $N(z_q)$ takes the value .5. Commercial graph paper assigns the value 5 to this point on the arithmetic scale, but exposition is simplified if we give it the value zero. The variable z may then be interpreted a standard normal variate (instead of a variable with mean 5). Using the arithmetic scale in Figure 4, the line L_1 can be used to read off the relation between x_q and z_q. Because of the relation between z_q and $N(z_q)$, the line can also be used to read off the relation between x_q and $N(z_q)$. But the latter is equivalent to reading $F(x_q)$, since $N(z_q) = F(x_q)$ for any q. In this way one may obtain $F(x_i)$ and $F(x_{i-1})$, two of the four terms, according to equation (8), needed two evaluate a_i.

We thus come to the problem of evaluating the lognormal first moment function $F^*(x_q)$, which will lead to values for $F^*(x_i)$ and $F^*(x_{i-1})$ in equation (8). This function is known to satisfy the relation[1]

$$(11) \qquad F^*(x_q \; \mu,\sigma^2) = \int_0^{x_q} x \cdot f(\mu,\sigma^2)dx$$

$$= \exp\,(\mu + \tfrac{1}{2}\sigma^2) \cdot F(x_q \;\; \mu + \sigma^2,\ \sigma^2).$$

That is, the lognormal first moment function is itself a lognormal

[1]Aitchison and Brown, op. cit., p. 12.

distribution function, except for a constant factor $\exp(\mu + \frac{1}{2}\sigma^2)$. The mean of ln x for the new distribution is greater by σ^2 than for the function $F(x_q\ \mu, \sigma^2)$, but the variance of ln x is the same. Hence the linear representation of the new distribution function,

$$(L_2) \qquad \log_{10} x_q = 0.434(\mu + \sigma^2) + 0.434\ \sigma^2 z_q,$$

has the same slope as (L_1) but a different intercept. In Figure 4, L_2 will lie parallel to L_1 $0.434\ \sigma^2$ units above it (log scale). For a graphical evaluation of $F^*(x_q)$ it is necessary to construct L_2, and to derive from Figure 4 the constant factor of equation (11).

To construct L_2 we need to find a line segment in Figure 4 of length $0.434\ \sigma^2$. We introduce the "parabolic" function

$$(P_1) \qquad \log_{10} x_q = 0.434(z_q - \frac{\mu}{\sigma})^2,$$

which intersects (L_1) at the points $(-\frac{\mu}{\sigma}, 0)$ and $(-\frac{\mu}{\sigma} + \sigma, 0.434\ \sigma^2)$. These points are labeled A and B respectively. Hence if a perpendicular is dropped from D to the horizontal axis, the line segment DB has the required length $0.434\ \sigma^2$. If DB is extended an equal distance to E, the construction of L_2 is immediate.

It may appear that in order to draw P_1, one needs to know the value of $-\mu/\sigma$. This is not correct, since from equation (L_1), L_1 has $-\mu/\sigma$ as its horizontal intercept. Once scales are established for the two axes, the simpler function

$$\log_{10} x_q = 0.434 z_q^{\ 2}$$

can be sketched once and for all, and the pattern moved along the horizontal axis to accommodate a particular distribution function L_1. This is a great saving if group means for many distribution functions are to be evaluated.

The constant of equation (11) may be determined if it is factored into

$$\exp\ (\mu + \frac{1}{2}\sigma^2) = 10^{0.434\mu} \cdot 10^{0.217\sigma^2}.$$

Since point J indicates the mean of ln x, $10^{0.434\mu}$ may be found by moving vertically from J to point I on the distribution function L_1, and then reading the corresponding value of x_q (not $\log_{10} x_q$) at H. To obtain $10^{0.217\sigma^2}$, we find the point G which bisects DB (so that GB has length $\log_{10} 0.217\sigma^2$), and read the corresponding value of x_q at F.

The procedure for estimating the a_i may now be summarized. We assume that an empirical cumulative frequency table has been prepared, that the points have been plotted on lognormal probability paper, and that a satisfactory freehand fit has been obtained as the estimated distribution function L_1. Using the probability scale on the horizontal axis, tabulate the estimated distribution function for each value of x which defines a group upper bound. (If the sample data are discrete, as they are in the case of employment data, more accurate results are obtained by taking the upper bound plus one-half. The table should include the endpoints $F(0) = 0$ and $F(\infty) = 1$, although these cannot be graphed.) Next draw in the function P_1 and construct L_2. Tabulate the values of L_2 for each group upper bound. Convert the latter to values of $F^*(x_q)$ by multiplying each by the product of the values of x_q at points F and H. Finally, form the ratios $[F^*(x_i) - F^*(x_{i-1})]/[F(x_i) - F(x_{i-1})]$, as indicated by equation (8).

An estimate of employment in a county may then be obtained from the relation

$$N_k = \sum_i a_i E_{ik},$$

where E_{ik} is the number of firms in the _i_th size class in county k. It is seen that while the argument for the use of the lognormal distribution in supplying missing values for county employment is not always simple, the method is quite easy to apply in practice. After a few graphs have been made the procedure becomes routine. In connection with the empirical work presented above,[1] graphs were made for four industries using Iowa _County Business Patterns_ data for the first quarter of 1962. These graphs--for agriculture--forestry--fisheries, mining, contracting construction, and manufacturing--are shown as Figures 5-8. Except for mining, which has too many employees in the 4-7 employee size class, the fits obtained with the lognormal distribution are surprisingly close.

---

[1]Chapter Two, Section 3. Above, pp. 88-115.

Figure 5

EMPLOYMENT SIZE DISTRIBUTION OF NON-FARM

AGRICULTURAL ESTABLISHMENTS, IOWA, 1962

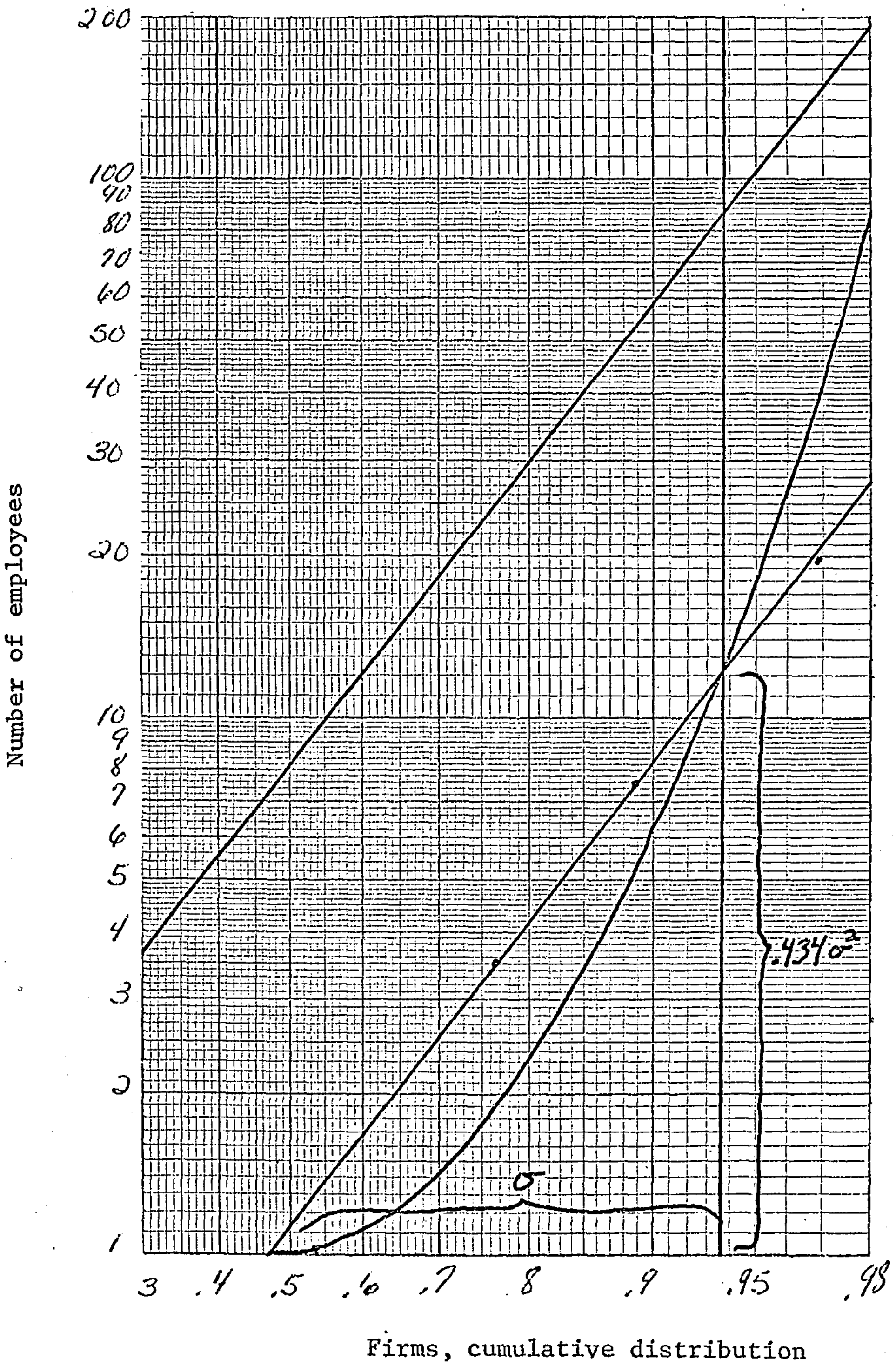


Figure 6

EMPLOYMENT SIZE DISTRIBUTION OF

MINING ESTABLISHMENTS, IOWA, 1962

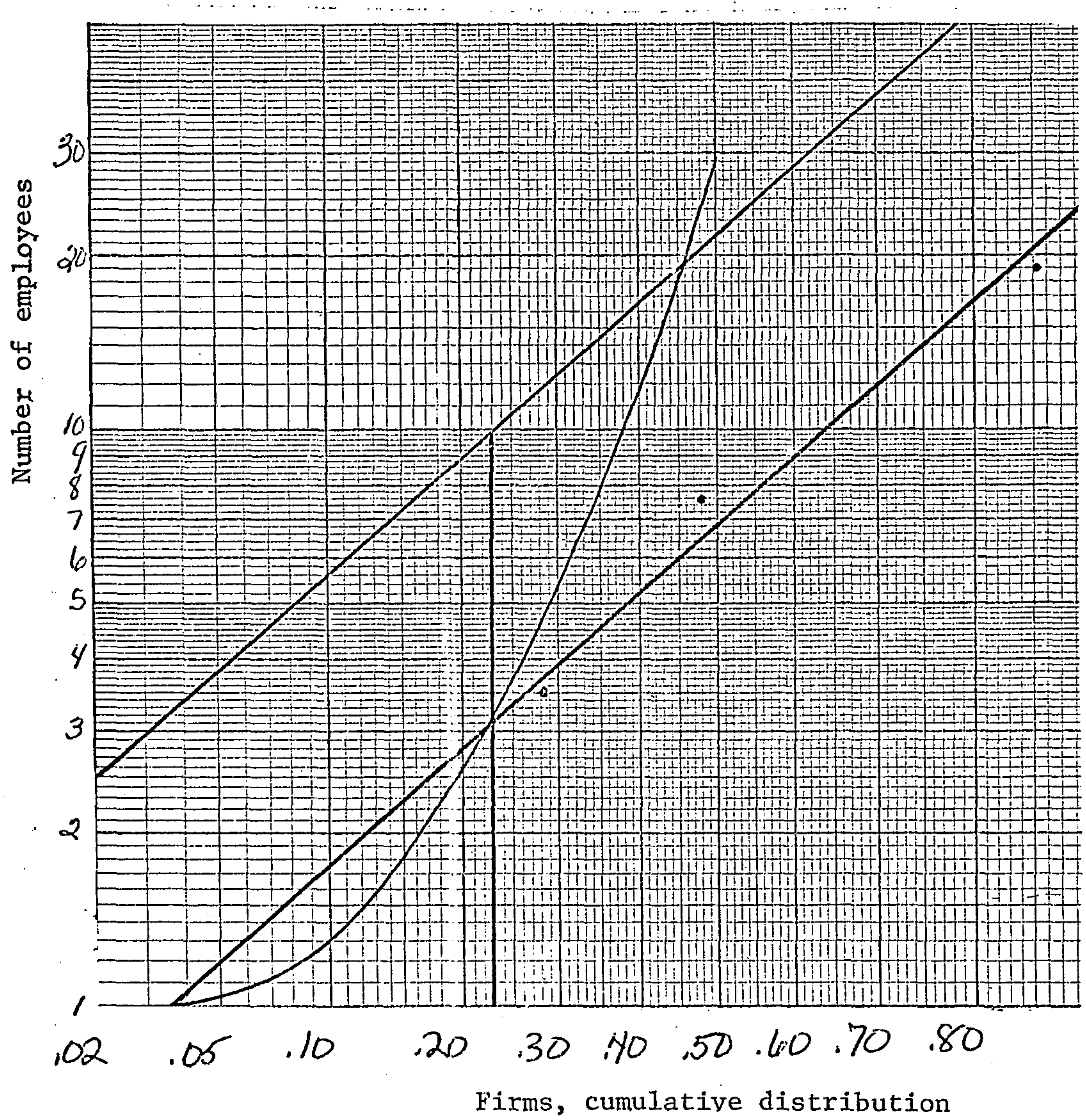


Figure 7

EMPLOYMENT SIZE DISTRIBUTION OF CONTRACT

CONSTRUCTION ESTABLISHMENTS, IOWA, 1962

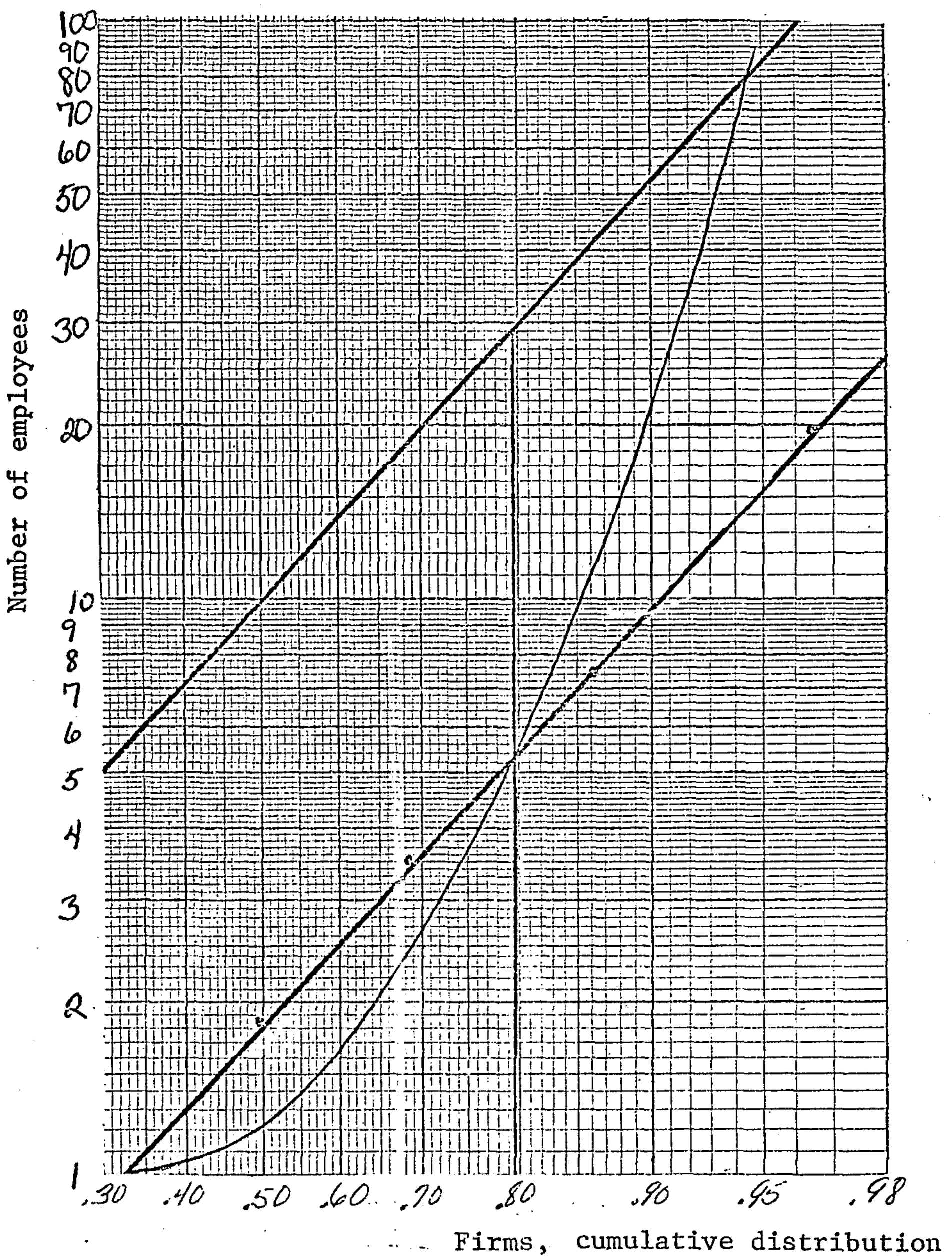
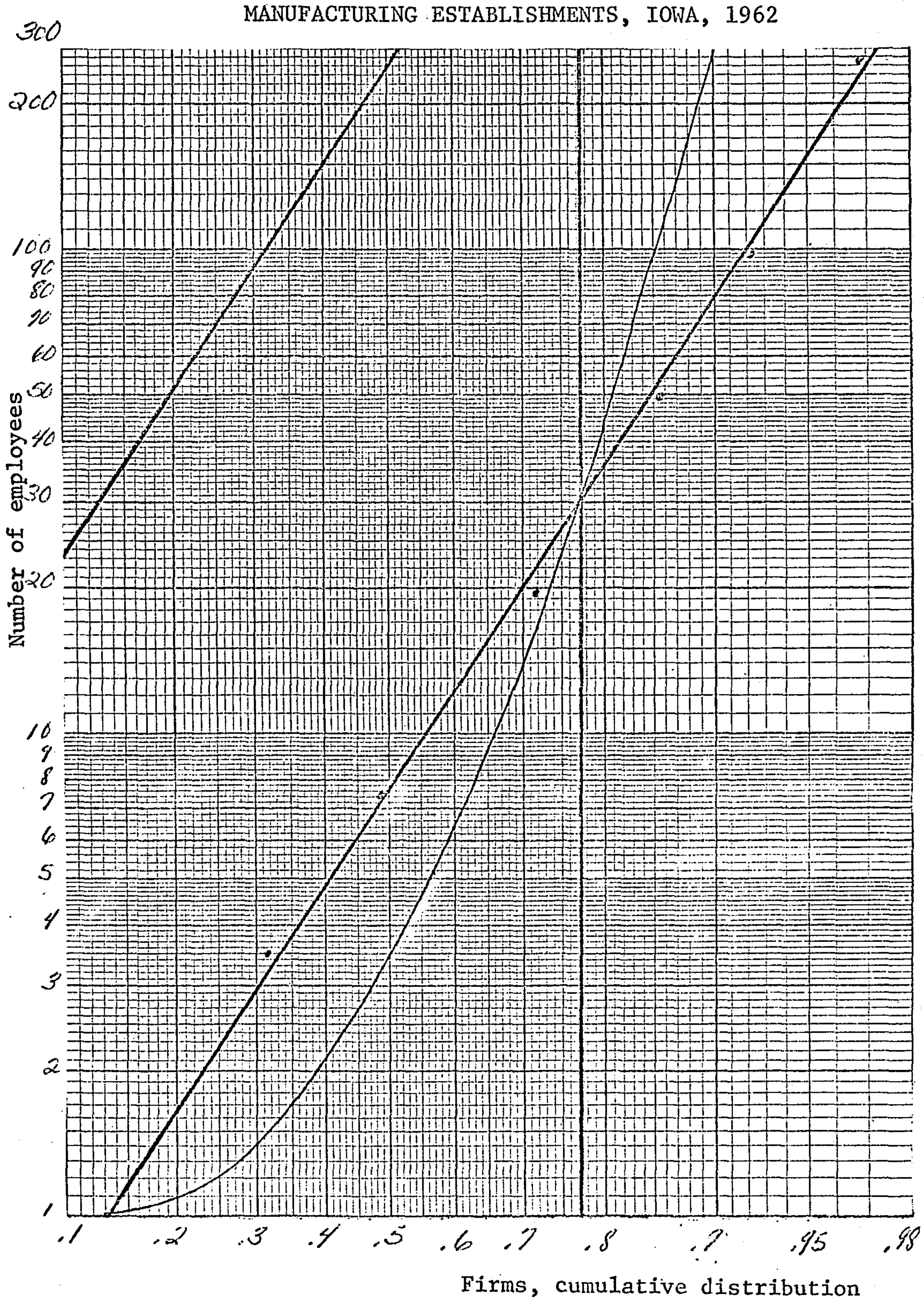


Figure 8

EMPLOYMENT SIZE DISTRIBUTION OF

MANUFACTURING ESTABLISHMENTS, IOWA, 1962



BIBLIOGRAPHY

1. General

Aitchison, J., and J. A. C. Brown. The Lognormal Distribution. Cambridge: Cambridge University Press, 1957.

Christ, Carl F. Econometric Models and Methods. New York: John Wiley and Sons, 1966.

Cramèr, Harold. Mathematical Methods of Statistics. Princeton: Princeton University Press, 1946.

Friedman, Milton. "The Interpolation of Time Series by Related Series," Journal of the American Statistical Association, 57 (December 1962), 729-757.

Goldberger, Arthur S. Econometric Theory. New York: John Wiley and Sons, 1964.

Goldfeld, Stephen, and Richard E. Quandt. "Some Tests for Homoscedasticity," Journal of the American Statistical Association, 60 (June, 1965), 539-559.

Hamburg, John R., et. al. "Linear Programming Test of Journey-To-Work Minimization," Highway Research Board Record, No. 102 (1965), 67-75.

Hart, P. E., and S. J. Prais. "An Analysis of Business Concentration," Journal of the Royal Statistical Society, Series A, 119 (1956, Part 2), 150-191.

Hadley, G. Linear Programming. Reading, Mass.: Addison-Wesley, 1962.

Isard, Walter. Methods of Regional Analysis. Cambridge: MIT Press, 1960.

Kain, John F. "A Contribution to the Urban Transportation Debate: An Econometric Model of Urban Residential Behavior," The Review of Economics and Statistics, 46 (February, 1964), 55-64.

Kendall, Maurice G., and Alan Stuart. The Advanced Theory of Statistics. Vol. I: Distribution Theory. London: Charles Griffin and Company, 1958.

Knauth, Oswald W. Distribution of Income by State in 1919. New York: National Bureau of Economic Research, 1922.

Leven, Maurice. Income in the Various States: Its Sources and Distribution in 1919, 1920 and 1921. New York: National Bureau of Economic Research, 1925.

Martin, John L. "Income Payments to Individuals by States, 1929-38," Survey of Current Business, 20 (April, 1940), 8-15.

McCarty, Harold H., John C. Hook, and Duane S. Knos. The Measurement of Association in Industrial Geography. Iowa City: Department of Geography, State University of Iowa, 1956.

Miller, Herman P. Income Distribution in the United States, A 1960 Census Monograph. Washington: U. S. Government Printing Office, 1966.

Morgenstern, Oskar. On the Accuracy of Economic Observations. 2nd ed. Princeton: Princeton University Press, 1963.

Munkres, James. "Algorithms for the Assignment and Transportation Problems," Journal of the Society for Industrial and Applied Mathematics, 5 (March, 1957), 32-37.

"National Income and Product in 1959," Survey of Current Business, 40 (July, 1960), 7-34.

Quandt, Richard E. "On the Distribution of Firms," American Economic Review, LVI (June, 1966), 416-432.

Schottland, Charles I. The Social Security Program in the United States. New York: Appleton-Century-Crofts, 1963.

Schwartz, Charles F., and Robert E. Graham, Jr. Personal Income by States Since 1929: A Supplement to the Survey of Current Business. U. S. Department of Commerce, Office of Business Economics. Washington: U. S. Government Printing Office, 1956.

Simon, H. A., and C. P. Bonini. "The Size Distribution of Firms," American Economic Review, XLVIII (September, 1958), 610-611.

Slaughter, John A. Income Received in the Various States: 1929-1935. New York: National Industrial Conference Board, 1937.

"State Personal Income 1948-1965," Survey of Current Business, 46 (August, 1966), 11-18.

"The National Income and Product Accounts of the United States: Revised Estimates, 1929-64," Survey of Current Business, 45 (August, 1965), 7-12.

Theil, H. Economic Forecasts and Policy. 2nd ed. revised. Amsterdam: North-Holland Publishing Company, 1961.

U. S. Civil Service Commission. Annual Report. Washington: U. S. Government Printing Office, 1963.

U. S. Department of Agriculture, Agricultural Marketing Service. Major Statistical Series of the U. S. Department of Agriculture: How They Are Constructed and Used. Vol. 3: Gross and Net Farm Income. (Agriculture Handbook No. 118) Washington: U. S. Government Printing Office, 1957.

U. S. Department of Agriculture, Economic Research Service. Farm Income: State Estimates 1949-1965. A Supplement to the July 1966 Farm Income Situation. Washington: U. S. Government Printing Office, August, 1966.

U. S. Department of Commerce, Bureau of Census. Current Population Reports, Series P-25. Nos. 294 and 304. Washington: U. S. Government Printing Office, 1964 and 1965.

U. S. Department of Commerce, Office of Business Economics. National Income, 1954 Edition. A Supplement to the Survey of Current Business. Washington: U. S. Government Printing Office, 1954.

__________, U. S. Income and Output. A Supplement to the Survey of Current Business. Washington: U. S. Government Printing Office, November, 1958.

U. S. Treasury Department, Internal Revenue Service, Statistics of Income: Corporation Income Tax Returns. Washington: U. S. Government Printing Office, annual.

__________. Statistics of Income: Individual Income Tax Returns. Washington: U. S. Government Printing Office, annual.

Wilks, S. S. Mathematical Statistics. New York: John Wiley and Sons, 1962.

Zellner, Arnold. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, 57 (June, 1962), 348-368.

2. County Income Estimates and Methods

(*Indicates publication contains recent non-commercial estimates of personal income by county.)

Adamson, W. M. "Measurement of Income in Small Geographic Areas," Southern Economic Journal, 8 (April, 1942), 479-92.

*"Arkansas County Personal Income and Retail Sales--1964," Arkansas Business Bulletin. Fayetteville: Bureau of Business and Economic Research, University of Arkansas, November, 1965, 4-5.

*Bishop, William D., Kentucky County Income, 1963 with Comparative Data, 1956-63. Lexington: Bureau of Business Research, University of Kentucky, 1965.

Bowen, Howard, coordinator. The Income of the Counties of Iowa. Iowa State Planning Board. Committee on Population and Social Trends, 1935.

*California State Chamber of Commerce, Research Department, Personal Income by Counties, 1964. (Economic Survey Series, No. 3) Sacramento: February, 1967.

Copeland, Lewis C. Methods for Estimating Income Payments in Counties. A Report Prepared by the Technical Committee for the Use of the Conference on the Measurement of County Income. Charlottesville: Bureau of Population and Economic Research, University of Virginia, 1952.

*Corry, Ormond C., and Patricia Ann Price. Comparative Economic Growth Measures--Population and Personal Income Estimates for Tennessee Counties 1950 Through 1962 (Study No. 30), Knoxville: Bureau of Business and Economic Research, University of Tennessee, April, 1964.

Cumberland, John H. "Suggested Improvements in Regional Income Accounting," Regional Science Association, Papers, 2 (1956), 259-271.

*Daicoff, Darwin. Kansas County Income: 1950-1964. Topeka: Office of Economic Analysis, State of Kansas, 1966.

*Edgel, Ralph L., and Peter J. La Londe. Income and Employment in New Mexico 1960-64. (Studies in Business and Economics, No. 15) Albuquerque: Bureau of Business Research, University of New Mexico, 1964.

Editor and Publisher Market Guide. New York: Editor and Publisher Company, annual.

Fulmer, John L. Analysis of Georgia Personal Income Payments, by Counties. Atlanta: Engineering Experiment Station, Georgia Institute of Technology, 1959.

__________. Analysis of Intercounty Commuting of Workers in Georgia. Atlanta: Engineering Experiment Station, Georgia Institute of Technology, 1959.

__________. "Regression Methods for Estimating Agricultural Income by County," Review of Economics and Statistics, 38 (February, 1956), 70-80.

Graham, Robert E., Jr. "Measuring Regional Market Growth: A Case Study of the Delaware River Area," Survey of Current Business, 34 (January, 1959), 10-19.

__________, and Edwin J. Coleman. "Personal Income in Metropolitan Areas: A New Series," Survey of Current Business, 47 (May, 1967), 18-44.

*Halley, Richard. "County Shares of Oregon's Personal Income: 1950-1961," Oregon Business Review. Eugene: Bureau of Business and Economic Research, University of Oregon (June, 1965).

*Hanks, J. Whitney. Personal Income in Utah Counties, 1958-1962. (Studies in Business and Economics, Vol. 24, No. 2) Salt Lake City: Bureau of Economic and Business Research, University of Utah, May, 1964.

Heffelfinger, W. A. County Income Estimation Methods--1965. Fayetteville: Bureau of Business and Economic Research, University of Arkansas, 1965.

*"Income and Population Estimates for Alabama Counties, 1965," Alabama Business. University, Alabama: Bureau of Business Research, University of Alabama (February, 1967).

Keyes, Scott, Felix C. Rogers, and Wallace E. Reed. Personal Income in Illinois Counties: 1950, 1954, 1956, 1958, 1959. Urbana: Bureau of Community Planning, University of Illinois, 1962.

Johnson, Byron L., and Carl G. Nordquist. An Estimate of Personal Income Payments by Colorado County: 1948. Denver: University of Denver Press, 1951.

Johnson, John L. Income in Kentucky: County Distributions by Amount, by Type, and by Size. Lexington: University of Kentucky Press, 1955.

Johnson, Robert H. An Analysis of Iowa Income Payments by Counties. ("Studies in Business and Economics," No. 1) Iowa City: Bureau of Business and Economic Research, State University of Iowa, 1950.

Lancaster, John Littlepage (ed.). County Income in Seven Southeastern States. Charlottesville: Bureau of Population and Economic Research, University of Virginia, 1952.

*________. Personal Income Estimates for Virginia Counties and Cities. Charlottesville: Bureau of Population and Economic Research, University of Virginia, 1963.

Leven, Charles L. "Theory and Method of Income and Product Accounts for Metropolitan Areas, Including the Elgin-Dundee Area as a Case Study." Doctoral dissertation, Department of Economics, Northwestern University, 1958.

*Leyden, Dennis R., and Albert Luther. Personal Income in West Virginia by Type and Magnitude, 1960-1963, a Pilot Study. (West Virginia University Business and Economic Studies, Vol. 10, No. 1) Morgantown: Bureau of Business Research and Office of Research and Development of the West Virginia Center for Appalachian Studies and Development, West Virginia University, June, 1966.

*Lund, Richard E. Personal Income in Wyoming Counties, 1929-1959. Laramie: Division of Business and Economic Research, 1963.

*Montgomery, V. E. "Income in South Dakota in 1964," South Dakota Business Review (November, 1965), 1-8.

*National Planning Association. Economic Base Study, Upper Mississippi River Basin Service Area. Technical Report No. 2. Personal Income Estimates by County for Minnesota, Wisconsin, Iowa, Illinois and Missouri and by Selected Counties of South Dakota and Indiana. For U. S. Army Engineer Division, North Central. Washington: National Planning Association, 1965.

*Norris, Clio Crosby. "Personal Income and Population in Georgia Counties 1962, 1964," Georgia Business. Athens: Bureau of Business and Economic Research, University of Georgia, June, 1966.

*Oklahoma Data Book, 1966. Norman: Bureau of Business Research, University of Oklahoma, 1966.

Peach, W. Nelson, Richard W. Poole, and James D. Tarver. County Building Block Data for Regional Analysis: Oklahoma. Stillwater: Research Foundation, Oklahoma State University, 1965.

*Pennsylvania, Commonwealth of, Department of Internal Affairs. Pennsylvania's Personal Income by County: Selected Years 1929-1963. Report No. IP-1, Harrisburg, 1965.

Personal Income in Counties of New York State, 1948-1957. A Supplement to the December, 1958 New York State Commerce Review.

Personal Income in Maryland Counties, 1951-1955. (Studies in Business and Economics, Vol. X, No. 4) College Park: Bureau of Business and Economic Research, University of Maryland, 1957.

*Population, Personal Income, Per Capita Income in Colorado, 1929 to 1962. Boulder: University of Colorado, 1963.

Spot Television Rates and Data. Skokie, Illinois: Standard Rate and Data Service, annual.

*Statistics of Personal Income, Population, Employment and Construction for Florida Counties. (State Economic Studies No. 16) Gainsville: Bureau of Economic and Business Research, University of Florida, 1964.

Survey of Buying Power. Supplement to Sales Management: The Magazine of Marketing. New York: Sales Management, annual.

*"The Trend of Business," Bulletin of Business Research, Vol. XXXVIII, No. 8, Columbus: Bureau of Business Research, Ohio State University, August, 1963.

Thorndike, Edward Lee. "Variations Among Cities in Per Capita Income," Journal of the American Statistical Association, 32 (September, 1937), 471-79.

U. S. Department of Commerce, Business and Defense Services Administration. Personal Income: A Key to Small-Area Market Analysis. Washington: U. S. Government Printing Office, 1961.

Vatter, Ethel G. "The Composition and Distribution of County Farm Incomes, 1948-1957." Doctoral dissertation, Department of Economics, State University of Iowa, 1962.

Weaver, H. G. "The Development of a Basic Purchasing Power Index by Counties," Harvard Business Review, IV (April, 1926), 275-89.

3. County Data for Income Estimation

American Medical Association, Department of Survey Research. Distribution of Physicians, Hospitals and Hospital Beds in the U. S. by Census Region, State, County, and Metropolitan Area. Chicago: American Medical Association, 1964.

Federal Reserve System, Board of Governors. Distribution of Bank Deposits by Counties and Standard Metropolitan Areas. Washington: Board of Governors of the Federal Reserve System, biennial.

Hospitals: Journal of the American Hospital Association. Annual.

Iowa Bureau of Labor, Statistical Department. Iowa Work Injuries. Des Moines, quarterly.

Iowa Crop and Livestock Reporting Service. "County Estimates of Cattle and Hog Numbers." Des Moines, annual (Mimeographed).

Iowa Department of Agriculture. Iowa Assessors Annual Farm Census. Des Moines, annual.

Iowa Department of Health. Vital Statistics. Des Moines, annual.

Iowa State Bar Association, Junior Bar Section. Report of 1957 Survey of Certain Factors Relating to the Practice of Law in the State of Iowa. No date.

__________. Report of 1959 Survey of Lawyer Population in Iowa, by County. March, 1960.

Iowa State Tax Commission. Annual Report. Des Moines, annual.

__________. Summary of Real Estate Assessment Ratio Study. Des Moines, annual.

Iowa State Tax Commission, Income Tax Division. Annual Statistical Report for the Fiscal Year Ended June 30, 1964. Des Moines, 1964.

Iowa Tax Study Committee. Report of the Iowa Taxation Study Committee to the Governor and the General Assembly of Iowa, Part 1. Iowa's Tax System--A Factual Study. Des Moines, 1956.

Martindale-Hubbell Law Directory. Summit, N. J.: Martindale-Hubbell, Inc., annual.

Mileage Guide of Iowa. Emmetsburg, Iowa: McNamara's Moving and Storage, 1954.

Rand McNally Road Atlas. Chicago: Rand McNally, annual.

State of Iowa. Report of the Department of Social Welfare. Des Moines, annual.

__________. State Salary Book. Des Moines, annual.

"Unemployment Benefits Paid to Iowans by County in 1963," Iowa Employment and Earnings (February, 1964), 6.

U. S. Congress, Joint Committee on Reduction of Nonessential Federal Expenditure. Federal Civilian Employment by County, December 31, 1960. Washington: U. S. Government Printing Office, 1961.

__________. Federal Civilian Employment 1950. Washington: U. S. Government Printing Office, 1950.

U. S. Department of Commerce, Bureau of Census. U. S. Census of Agriculture. Washington: U. S. Government Printing Office, 1945, 1950, 1954, 1959.

__________. 1964 U. S. Census of Agriculture, Preliminary Reports. Washington: U. S. Government Printing Office, 1966.

__________. U. S. Census of Business. Washington: U. S. Government Printing Office, 1948, 1954, 1958, 1963.

__________. U. S. Census of Governments. Washington: U. S. Government Printing Office, 1957, 1962.

__________. U. S. Census of Manufactures. Washington: U. S. Government Printing Office, 1947, 1954, 1958, 1963.

__________. U. S. Census of Mineral Industries. Washington: U. S. Government Printing Office, 1954, 1958, 1963.

__________. U. S. Census of Population. Washington: U. S. Government Printing Office, 1950, 1960.

__________. County and City Data Book, 1962. Washington: U. S. Government Printing Office, 1963.

__________. County Business Patterns. Washington: U. S. Government Printing Office, irregular.

U. S. Department of Health, Education, and Welfare, Public Health Service. Health Manpower Source Book. (Public Health Service Publication No. 263.) Section 4, County Data from 1950 Census of Population (1954); Section 10, Physicians' Age, Type of Practice, and Location (1960); Section 19, Location of Manpower in 8 Health Occupations (1965). Washington, D. C.: U. S. Government Printing Office.

U. S. Department of Health, Education and Welfare. Social Security Administration, Bureau of Old-Age Survivors Insurance. State and Local Government Employment Covered by OASI. Washington: U. S. Government Printing Office, quarterly.

Vita

Name: George Christian Ehemann.

Born: Pittsburgh, Pa., March 3, 1937.

Education:

College of Wooster, 1954-1959. B.A. with honors in economics, 1959. Junior year at the University of Edinburgh.

Northwestern University, 1959-1968. M.A. in economics and Ph.D. candidacy, 1964.

Publication:

"The Iowa Economy in 1965," Iowa Business Digest, April, 1966 (with C. Oliven).

Positions held:

Instructor in Economics, Middlebury College, second semester, 1963-64

Research Associate, Bureau of Business and Economic Research, The University of Iowa, 1964 to date.