AN ABSTRACT OF THE THESIS OF

MILTON LEE HOLLOWAY for the DOCTOR OF PHILOSOPHY
(Name) (Degree)

in Agricultural Economics presented on June 9, 1971
(Major) (Date)

Title: A PRODUCTION FUNCTION ANALYSIS OF WATER RESOURCE PRODUCTIVITY IN PACIFIC NORTHWEST AGRICULTURE

Abstract approved: Redacted for Privacy
Joe B. Stevens

The competition or rivalry for the use of water resources among economic sectors of the Pacific Northwest and among geographical regions of the western United States has intensified in recent years. This rivalry and the long run prospects for water shortages have increased the demand for research concerning the productivity of this resource in alternative uses. This demand exists because the distribution and use of water resources require investment which typically comes from both public and private sources. Private and public planning groups seek answers to questions regarding future water resource development alternatives.

Agriculture has historically been a major user of water in the Pacific Northwest. A substantial portion of total investment in water resource development has also been in agriculture. As a result water use planners and decision making bodies are necessarily interested

in water use in agriculture. The success of water resource planning requires answers to questions regarding the value of the productivity of water in all its major uses, including various aspects of water use in agriculture.

Different aspects of water use in agriculture which are important to decision makers include (1) the value productivity of various kinds or types of water resource investments, (2) the value productivity of water in various kinds of agricultural production in different geographical areas, and (3) the returns to private and public investment in agricultural water resources. This study was directed to providing answers to these questions. Pacific Northwest agriculture was studied from this viewpoint.

Agricultural water resources were classified as irrigation, drainage, and water related Agricultural Conservation Program (ACP) practices. These are the major classifications of water resources in which investments are made in the Pacific Northwest.

Production function analysis was selected as a method of investigation. Production functions were estimated for five areas or subregions in the Pacific Northwest. These areas are composed of counties with similar patterns of production. The Agricultural Census was the primary data source, supplemented by related U. S. Department of Agriculture publications, and various state publications.

Ordinary least-squares regression (OLS) techniques were employed to derive the initial estimates of the parameters of the production function models. Tests for detecting interdependence within the independent variable set of the models revealed a considerable degree of instability in the OLS parameter estimates. This condition makes the OLS solutions (and various derivations) particularly vulnerable to change from measurement error, poor model specification, and equation form.

A prior information model was selected to explicitly include available prior knowledge in the estimation process. The model selected allows (1) tests of comparability of the two information sources (prior and sample), (2) over-all contribution of prior information to the new solution set, and (3) derivation of percentage contribution of the two information sources to individual parameter estimates.

The results of the study indicate that no reliable estimates of value of production from drainage and ACP were possible from the sample information. Returns to irrigation were considered lower than expected in two of the farming areas and higher than expected in another. Estimated returns were high in the area which produces primarily field crops (about nine dollars per acre foot). The area has a small level of current irrigation development. Indications are that irrigation development is probably beyond the optimum level in the area where most large projects have been developed in the past (less than four dollars per acre foot). Future development would be most profitable (assuming equal development cost) in the dryland field crop area.

Estimated returns to other factor inputs indicate (1) low returns to labor in two areas, (2) generally high returns to current operating expenditures, and (3) low returns to machinery capital. Returns to cropland were about as expected in two areas (five to seven percent) but low in two other areas (about two percent). Indications are that labor mobility should be increased in the area and that future land development should be in the livestock-field crop and the field crop areas rather than the coastal area or the west-central valley areas (primarily the Willamette Valley).

A Production Function Analysis of Water Resource Productivity in Pacific Northwest Agriculture

by

Milton Lee Holloway

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1972

APPROVED:

Redacted for Privacy

Associate Professor of Agriculture Economics
in charge of major

Redacted for Privacy

Head of Department of Agricultural Economics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented June 9, 1971

Typed by Susie Kozlik for Milton Lee Holloway

ACKNOWLEDGMENT

I wish to express appreciation to my major professor, Dr. Joe B. Stevens, for his assistance, guidance, and motivating influence throughout the course of this research; to Dr. George A. Pavelis, Economic Research Service for his guidance and advice during early phases of the project; to Dr. William G. Brown for his suggestions and direction during the analysis; and to other members of my graduate committee, Dr. John A. Edwards, Dr. Emery N. Castle, and Dr. David R. Thomas for their beneficial criticism. I would also like to extend to Dr. Stanley F. Miller my sincere appreciation for his counseling and advice.

I especially acknowledge my wife, Barbara, whose unlimited patience and encouragement helped make this study a reality.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDIX TABLES

LIST OF FIGURES

A PRODUCTION FUNCTION ANALYSIS OF WATER RESOURCE PRODUCTIVITY IN PACIFIC NORTHWEST AGRICULTURE

I. INTRODUCTION

Water is an important natural resource which occupies a unique position in the development and maintenance of the communities, cities, and states, and consequently, the entire Pacific Northwest region. The uniqueness of its role is apparent whether it is abundant or scarce; whether its forces are harnessed for power or leave a periodic path of destruction from flooding. Water supply and water quality problems, or the exposure of an area to floods or drought, are important factors in the intensity and location of economic activity of the region.

Water as a Natural Resource

Natural resources are defined or set apart from "unnatural" or "man-made" resources in that they exist as a source or supply in nature. Our sources of water may, in some sense, be thought of as man-made supplies, but in general, our water supplies are thought of as having their origin in nature, and are appropriately called a natural resource. The term "water resources" includes a wide variety of sub-classifications which are associated with particular locations or forms in which we find water. As such, snow packs in the mountains and moisture in the soil are as much water resources as streams, lakes, and estuaries.

Water Resource Development in the Northwest

Water resource development typically refers to changing the hydrology of water, thus making it usable, or more usable, by people. In some cases this may require the building of dams and canal structures, the digging of irrigation and drainage canals, and dredging harbors--or in the opposite vein--building access roads to high mountain lakes, planting trees to protect the soil from rapid run-off and thus, the quality of downstream water, or simply diverting flash flood run-off in the desert to form livestock watering ponds. A typical classification of water uses includes domestic and municipal, industrial, electric power, agricultural, navigation, recreation, and fish and wildlife.

The Pacific Northwest has perhaps one of the broadest ranges of water resource uses and the most diverse system of development of any comparable sized region in the United States. Rivers, streams and lakes are numerous in western Oregon and Washington where too much water (flooding and slow drainage) is often a problem in winter and drought consistently comes in the summer when stream flows are also low. Eastern Oregon and Washington and southern Idaho are semi-arid regions with low year-round average precipitation. Water shortage is almost always a problem. Major rivers, including the Snake and Columbia, flow through the area and considerable water diversion is practiced to supplement other sources. Most of the region's electrical power is generated on these two rivers.

Perhaps the most apparent example of development of the water resource in the Pacific Northwest involves streams and rivers. The development of streams and rivers began early in this century. The Corps of Engineers completed a navigation project on the Alsea River in western Oregon as early as 1898 (9, p. 3). Other projects completed by the Corps which are most apparent to the casual observer include dam sites on major rivers such as The Dalles, McNary, and John Day on the Columbia River between Washington and Oregon, and the Chief Joseph on the Columbia near Bridgeport, Washington. Total Federal costs of projects completed in the Columbia North Pacific District by the Corps of Engineers up to 1967 was approximately $1.5 billion (9, p. 3). These projects include water use for navigation, flood control, power, and recreation. Non-Federal costs of the projects total $10.8 million (9, p. 3). The Bureau of Reclamation, whose primary function is irrigation development, also has a long record of project construction in the region. Among the first projects completed were the Sunnyside portion of the Yakima project in north central Washington in 1907 and the Umatilla project in north central Oregon in 1908 (62, p. 754). Net Federal investment in Bureau of Reclamation projects (initial investment minus repayments) in the region up to 1965 totaled $715 million (63, p. 51).

Individual municipalities and small groups have done much to develop docks, access roads, irrigation outlets, etc., along the streams and rivers in the area. The total private investment is perhaps unmeasurable, but is a major source of water resource development in the region. Data from the Census of Irrigation (61, State Table 1 and 2 for Oregon, Washington, and Idaho) and a publication by the U. S. Water Resources Council (64, p. 6-16-5) indicate at least 54% of water use in the period 1959-1965 is from private development sources. These private sources include rural domestic, municipal, self-supplied industrial, individual farmers and farm mutuals in agriculture. Approximately 14.5% of the total water used in 1965 was from groundwater sources, less than one percent from saline sources, and the remaining 85.5% from surface sources (64, p. 6-16-5).

Strictly within agriculture, the Agricultural Stabilization and Conservation Service (ASCS) through the Agricultural Conservation Program has been instrumental in promoting investment in water resources. This program is administered with the cooperation of the Soil Conservation Service, Forest Service, Extension Service, Soil and Water Conservation District supervisors, and other agricultural agencies and includes approximately thirty-five water related practices including establishment and management of drainage systems, irrigation systems, water conserving cultural practices, livestock water facilities, and others (see Appendix Table III for a complete listing of the practices included in this study). ASCS has invested $76.7 million (49, p. 2) in land and water cost-sharing agreements in the state of Oregon alone from 1936 to 1964. The Soil Conservation Service also conducts the Small Watershed Program for assistance in the construction of small dam projects. This program was authorized under Public Law 83-566 and amended in 1966. Investments to date have been relatively small in this program, but the program provides a significant potential source of future investment.

Individual farm investment in water resource development and use are partly evidenced by the growth in acres irrigated and drained. The farmer's share of the cost-sharing program of ACP suggests that at least $76.7 million have been invested by the farmers in Oregon in land and water conservation programs from 1936 to 1964.[1] Additional investments by farmers have been made independently of these Federal programs.

Approximately 89% of the total regional use of water in 1965 was in agriculture (64, p. 6-16-5). An estimated 51% of the total agricultural water use for irrigation in 1959 came from private sources--individual and farmers' mutuals (61, State Table 1 and 2 for Oregon, Washington, and Idaho). Of the total agricultural use in 1965, about 13% came from underground sources, which is almost totally from private investments (64, p. 6-16-5).

---

[1] Cost-sharing agreements under ACP are usually one-half the per unit cost.

In general, the water supply of the region is abundant, although distribution varies widely and seasonal flows are low in many of the smaller streams. The available average annual natural runoff is approximately 289 million acre feet (maf/yr.) of which 19% originates in Canada. Total annual withdrawals average 33.3 maf/yr. of which 10.5 maf/yr. are consumed. About 95% of the consumption is due to irrigation (64, p. 6-16-1).

Water Resources in Agriculture

Irrigation is no doubt the most recognized aspect of water resource development in agriculture and usually the most important. Other aspects are usually present, however, and at times more important. These other aspects are classified in this study as drainage and water conservation practices. In many areas of western Oregon and Washington, irrigation cannot be developed without also developing a drainage system and/or flood protection. Sometimes the soils are such that natural water percolation downward is almost nonexistent and excess water must be taken off the land by surface drainage systems. In other cases, natural water supplies are sufficient and only drainage is necessary. Many areas of land along rivers are useless for agricultural purposes without flood protection. In the semi-arid regions of eastern Oregon and Washington, conservation practices increase the effectual water supply by making better use of natural precipitation.

Total investment in agricultural water resource development is difficult to assess since a substantial portion comes from private sources. In addition, public investments are often in the form of multiple purpose projects which serve both agricultural and nonagricultural sectors. Special reports from Census of Agriculture indicate an average capital investment of $137.00 per irrigated acre by irrigation organizations in 17 western states and Louisiana in 1959 (64, p. 4-4-6).

Problem Statement

A significant portion of the water resource investments in the Pacific Northwest has been allocated to water resource use and development in agriculture. The public portion of these investments takes many forms, administered under various programs by several agencies and includes the building of dams and other structures, as well as the promotion of various cost-sharing arrangements with individual farmers. The decisions to invest have been historically based upon a project by project or program by program evaluation. Various decision making units have been involved in these decisions and include individual farmers, farm groups (irrigation and drainage districts), municipalities, and state and Federal agencies. Recent planning efforts have been designed to coordinate many of these activities. An example is comprehensive river basin planning in which Federal, state and local groups have the opportunity to participate in the planning process. This approach will hopefully remove some of the piece-meal, sometimes contradictory, decisions.

The private decisions to invest are not independent of the public sector's investment decisions. Present period investments are influenced by the availability of present public funds and the expectation of future public investments. The development of an irrigation project requires some private investment. Cost-sharing agreements which are traditionally renewed year after year affect not only total investment but the timing of private investment.

The piece-meal decision process and its impact on private decision making point out the necessity of coordinated public water management policy. Growing demands (relative to supply) for water and water-related capital (due to increased population and increased public demand for water via recreation activities) increase the competition for water and the importance of making correct decisions regarding development. The recent awareness of ecological problems associated with misuse of water adds prudence to the development question and adds an additional note of urgency to implementing "good" decisions.

Comprehensive planning and the coordination of public and private water development decision making seem imperative in the determination of the best use of our water resources. The success of such an approach depends on a great many factors--not the least of which is reliable information concerning water use and productivity in the agricultural sector. This implies the necessity of several kinds of information including (1) the productivity of water in various uses in different geographical areas, and (2) the aggregate regional productivity of these water resources. Additional information requirements are the returns from both public and private investments in agricultural water resources. To these ends the following objectives are outlined.

Objectives of the Study

The overall objective of the study is to determine the contribution of agricultural water resource development to recent agricultural production. More specifically, the objectives are to:

1. Determine production response coefficients for irrigation, drainage, and water conservation practices in each of several farming areas in the Pacific Northwest.
2. Determine the public and private returns per dollar invested in agricultural water resources in the Pacific Northwest.

Justification

In summary of the above discussion, and in addition to it, it is sufficient to say that this study is justified on the basis of providing information for decision makers regarding an important problem of the day. It is designed specifically for public water management policy, including Federal, regional, state, and local decision making groups--though it may be of some value to individuals. The study is viewed as providing partial information to the input requirements for intelligent public decisions regarding an increasingly vital, publicly managed resource.

The kinds of information which this study is intended to provide are considered important by the United States Water Resources Council (64, p. 4-4-6) as evidenced by the following statement:

> Federal agricultural water management policy should include consideration of both the policy's overall effect on agricultural production, and the productivity of investment in irrigation relative to alternative investments such as drainage, clearing of land, and other technological developments.

II. THEORY AND CONCEPTUAL FRAMEWORK

The estimation of agricultural production functions was selected as the basic technique for the analysis of water resource productivity and returns to public and private water resource investment in Pacific Northwest agriculture. The analysis was accomplished by explicitly specifying important types of water resource investment as variables in the production functions. Information regarding the contribution of water resource investment to the value of farm production and the relationship to the other production inputs was obtained by statistically estimating these functions.

The Production Function Concept

The concept of a production function is essentially a physical or biological science concept of the relationships between inputs and outputs in a production process. As such the concept is crucial to, and has been predominantly used in the development of firm production theory in economics. Coupled with input and output prices, the production function determines the shape of the firm demand functions for factor inputs, and the firm supply function for the output.

The production function concept has also been extended to include the production responses of an aggregate of firms, of industries, and of regions. Many empirical studies have been concerned with the estimation of production responses at a level of aggregation above that of individual firms. These functions are typically referred to in the literature as aggregate production functions.[2]

The extension of the production function concept beyond the firm level of aggregation has been a response to the need for answers to a certain class of questions. Questions of intercommunity or interregional allocation of resources in agriculture, for example, are concerned with aggregate effects. Policy issues of farm organizations, counties, states, regions, and nations are necessarily concerned with the performance of groups of people, groups of firms, and perhaps groups of industries.

Policy implementation usually requires control or influence on a system at the aggregate level. This is not to say that individuals within the group are unimportant, only that it is usually an unworkable proposition to consider each individual separately. Even if this could

---

[2]The term "aggregate production function" is typically defined to mean a function which is at a higher level of aggregation than the firm level. The distinction was probably made at this level because of the traditional firm orientation of micro-economic theory. However, this definition is completely arbitrary since any function in the hierarchy of aggregation may be thought of as an aggregate of some lower level functions. This traditional definition is sometimes confusing, especially in connection with discussions on aggregation bias. Reference to an "aggregate production function" is also less descriptive than other terms such as firm function or industry function. Therefore, the term "aggregate production function" will not be used in this writing.

be done, it is not usually acceptable to disregard the aggregate effects.

The static pure competition model in micro-economics has no real need for a production function of a higher level of aggregation than the firm production function. Equilibrium conditions (which are considered the standard or usual case) require that the marginal productivity, and thus the marginal value product (MVP), of each production factor be the same for all firms since all price ratios are equal for each firm and each interpreneur is a profit maximizer. The theory does not necessarily require that each firm have the same production function -- only that each function exhibit diminishing marginal productivities of the factor inputs. Aggregation to the market level is accomplished through the aggregation of firm supply functions for the output and the aggregation of demand functions for the factor inputs. The theory is designed to conceptually explain the firm side of the market system and to provide a framework for predicting future market conditions. The system is always considered to be moving toward equilibrium. As a result the theory provides a static concept of how the market "tends" to function but provides very little guidance to conceptual measures of the "severity" and "causes" of a particular disequilibrium condition.

The existence of disequilibrium in the system is first evident in the market place where quantity supplied does not equal quantity demanded. But this evidence does not isolate the source of the disequilibrium or allow its analysis without an investigation of the individuals which make up the aggregate supply and demand functions. The theory assumes that the industry firms are attempting to maximize profits based on expected output prices and that appropriate industry adjustments will tend to be made in the next time period in the event that output prices were not as expected in the present period. Conceptually, the mechanics for tracing the disequilibrium to its source are contained in the theory -- we may simply analyze each firm in the industry. But the theory does not show how to analyze aggregate disequilibrium associated with particular production inputs.

An obvious alternative is to analyze the aggregate relationships, provided it is possible to do so without ambiguity. To insure the absence of ambiguous answers from the aggregates, the relationship between the individuals and the aggregate must be unique and identifiable. Given this realization and the set of existing prices, one will be able to determine, ex post, whether firms in the aggregate used the appropriate level and combination of inputs. To explain the full conceptual implications of unambiguous aggregate functions, the following discussion uses the simplified case where firms produce a single homogeneous product and use the same set of homogeneous inputs. The discussion is developed based on the relationship between firm production functions and the industry function. It should be recognized that the same principles hold for any aggregation level.

Consistent Aggregation

A question of concern as to the usefulness of aggregate functions is whether the aggregation is consistent. Aggregation will be said to be consistent when the definition of the aggregate function is such that solutions derived from it are not in conflict with the aggregation of individual function solutions; i. e., the aggregate results are not ambiguous. They are free of aggregation bias. Green (18, p. 35-44) has derived the necessary and sufficient conditions for consistent aggregation. The essentials of his derivations as related specifically to production functions are presented in Appendix I, along with three simple examples for illustrative purposes. Only the results and implications are presented here.

In general, any set of continuous functions can be aggregated consistently if the appropriate weights are used. For non-stochastic models, consistent aggregation depends completely upon the aggregation procedure used. Some important results of Appendix I are:

1) If individual functions are linear with the same slopes, the aggregate function will be consistent when aggregates are defined to be simple sums (see example A of Appendix I).

2) If individual functions are linear with different slopes, the aggregate function will be consistent if (a) input aggregates are weighted sums with weights equal to the firm marginal product

for that particular input and the output aggregate is the simple sum of outputs, or (b) input aggregates are simple sums and output aggregates are weighted sums (see example B of Appendix I).

3) If individual functions are of the Cobb-Douglas type and homogeneous of degree one, all having the same value for the exponent of each input, aggregation will be consistent for constant input ratios if (a) inputs and outputs are simple sums in the case where firms have identical functions, or (b) inputs or outputs are simple sums while the other is an appropriate weighted sum when firms have functions with the same exponents but different constant terms.

The requirements of consistent aggregation are slightly more complicated when the functions are stochastic instead of exact (see Appendix I). This added dimension makes consistent aggregation depend upon (1) the aggregation procedure, (2) the algebraic form of the equations, and (3) the statistical estimation method used.

Conceptually, the industry production function must meet some very specific requirements. (It should be clear, however, that these requirements are mathematically and statistically the same as is required for unambiguous aggregate supply and demand functions from our usual market equilibrium theory). Given that firms exist and that they each have a physical production function, an industry production or physical response function also exists which will provide the same aggregate information as the aggregation of the individual firm function responses, provided the aggregates are appropriately defined. Only in special cases, however, will the firm functions contribute in equal proportions to the aggregate; thus, simple sums data (summed over firms) are appropriate only in special cases.

In some other respects, however, the industry function is conceptually the same as conventional firm functions. The aggregate function is defined for a specific production unit (the industry) and for a specific unit of time (a production period of one year in most studies). Given the function and the existing input-output prices one could indicate the aggregate discrepancy (if any) from optimal levels of resource use. Given similar functions for other groups of firms, one could also indicate the desirable direction of the movement of resources between groups.

Consistent Aggregation in Perspective

The aggregation problem has, for the most part, been ignored in empirical aggregate studies. The obvious reason for neglecting the problem is that there seems to be no practical alternative. Correct aggregation of the data to provide a consistent aggregate function requires specific information about the individual functions which make up the aggregate. If we had such information, we would have no need for the aggregate function. Most available data (e.g., Census of Agriculture) are reported as simple sums and we usually have no good methods by which to disaggregate them. The only functional forms consistent with a simple sums data are linear functions with like slopes or functions homogeneous of the first degree, with the special restriction of identical functions with fixed input ratios. We may assume, however, that the wider the divergence from both similarity and linearity between the individual functions, the greater the probability of a large aggregation bias at the aggregate level.

From an empirical point of view, it is noteworthy that there is no guarantee that we would be more accurate in evaluating aggregate results if we first estimated the firm functions and then aggregated the results. The firm level function estimation is subject to the same kind (if not the same potential magnitude) of error as the industry function. These errors are from estimation, equation formulation, and aggregation.[3] This approach is, of course, much more costly in time and reserarch expenditures when the study involves large of firms.

Two implications of the above discussion on consistent aggregation are important to this study: (1) The aggregation problem is not

---

[3]From a mathematical point of view, a firm function is also an aggregate function -- the components being some subdivision of the farm; e.g., 200 one-acre production functions for a 200 acre farm.

peculiar to the traditionally defined aggregate functions (industry, regional, or a higher level of aggregation). The same problem exists at the firm level. The potential bias at the firm level may be less, but we have no such assurance. (2) A function (regardless of the level of aggregation) will be consistent only if the aggregates are appropriately defined in accordance with the form of the individual functions making up the aggregate. Thus, the use of simple sums data (which are usually the only data available for analysis) for the estimation of a nonlinear aggregate function is necessarily an inconsistent aggregate. (3) Conceptually, the most appropriate level of aggregation for a particular case depends on the kind of research question for which answers are being sought. If one is interested in results at a high level of aggregation, there is perhaps a trade-off between probable inaccuracy due to aggregation bias and the cost of doing the analysis at a lower level of aggregation. Limited research time and funds often prevent the analysis at the lower aggregation level.

Historical Development of Production Function Analysis

Advantages of the production function technique as compared to alternative techniques are its relative simplicity, the potential adaptability to low-cost secondary data sources, and the existence of numerous references to (apparently) successful past studies of a similar type. Available alternatives to provide similar information are limited and costly, given the present state of economic research technology. For example, farm survey data would provide the data for an aggregate regional analysis but would be very costly for so large a region as the Pacific Northwest. Studies for small areas would provide some information but would lack the universality of the aggregate approach.

The following sections are designed to give the reader an overview of how production functions have been used in the past. A cross-section of production function studies is given along with the criticisms which followed in the literature. The discussion includes agricultural and non-agricultural studies. Studies at various levels of aggregation are included.

Early Studies

The essential characteristics of the present day production function concept find their origin in early economic writings. The characteristic of eventually diminishing marginal productivities of the factor inputs did not have its beginning in contemporary firm theory but rather in early descriptions of agriculture as an industry. In particular, Ricardo described diminishing returns in his theory on rents.

Specific algebraic forms of the production function were not suggested until early in this century. Wicksell, as cited by Earl Heady (29, p. 15) suggested that agricultural output was a function of labor, land, and capital, and that the function was homogeneous of degree one.

Cobb and Douglas (7) were the first to try empirical estimation. They estimated a production function for American manufacturing industries by the use of time series data. The functional form used was

$$Y = aL^{\alpha}K^{1-\alpha} \tag{2.10}$$

where Y was the predicted index of manufacturing output over time, L was the index of employment in manufacturing industries, and K the index of fixed capital in the industry. Additional studies by Cobb and Douglas using this same basic functional form (with the sum of the exponents not necessarily equal to one) resulted in the common usage of the name Cobb-Douglas to describe the general form of (2.10).

Various formulations of the Cobb-Douglas function have since been designed in response to a number of criticisms which arose over the initial formulation and its implications. The function has been used for both national (or regional) and industry functions from both time series and cross-sectional data.

Criticisms of early empirical estimation . . . Criticisms of early attempts to estimate production functions included conceptual questions, measurement, and estimation questions. Reder (40) indicated that the empirical functions differ from the theoretical firm production function in three ways: (1) In theory, the production function shows the relationship between input quantities and the output of a firm and not the input-output relationship from an aggregate of firms. (2)

Theoretical production functions are in terms of physical quantities; not in terms of value of output added. (3) In firm theory the marginal value productivity (MVP) of a factor input is the first partial derivative of the total product function, times the marginal revenue. In the empirical function the marginal value product is assumed to be the first partial derivative of the total value function. Accordingly, an MVP of the empirical function should be called an inter-firm MVP while the theoretical concept is an intra-firm MVP. Only under conditions of static pure competition equilibrium would the two concepts be the same.

Reder's criticisms of a statistical nature pointed out weaknesses in the quality of data, inaccurate measurement, and the lack of real experiments to generate the data. The measurement of capital was criticized since it did not measure the annual flow of capital but measured either the capital stock or current investment. This may be of particular importance in the use of cross-sectional data to estimate firm functions where firms employ different technology because of fixed plants inherited from the past. Observations for the firm functions partially reflect differences in management skills over time in the case of time-series data, and differences in management skills between enterpreneurs in the case of cross-sectional data. Neither time-series data nor cross-sectional data provide true experimentation where capital and labor are combined at various levels to determine a corresponding output level.

More recent attempts to estimate production functions have taken essentially two routes; (1) inter-industry functions of the original Cobb-Douglas type which have been primarily concerned with estimating functional distributive shares between labor and capital, returns to scale, and technological change over time and, (2) inter-firm functions which have been primarily concerned with MVP estimates of particular (more specific) inputs (and their comparisons) and returns to scale for the industry. In addition to these two categories, and within agriculture, experimental data have been used to estimate physical production responses from various levels of fertilizer or other experimentally controllable inputs.

Recent Estimation of Agricultural Production Functions

Resource productivity questions of a very specific nature (e.g. marginal productivity of various kinds of fertilizer on a particular soil type) have been recently analyzed with experimental data from state experiment stations. Examples of these types of studies are Miller and Boersma (35) and Heady and Pesek (30). Studies such as these are numerous in agriculture and provide a great deal of specific information regarding production responses. Although these functions provide more specific information than the firm or industry functions, the results are usually less applicable to extension or policy issues since the control exercised in the experimental design (usually) necessarily requires the exposition of less variance in the important factors than will be found in the "real world".

Firm production functions estimated from a cross-section of farm records provide diagnostic information to the group of farms. They indicate, for example, whether equilibrium conditions exist, i.e., whether returns to labor and various forms of capital are different from their market prices. The nature of the information is more general than that from experimental data and usually applies to broad groups or aggregations of inputs. Heady and Dillon (29, p. 554-585) report several of these functions, each designed for a specific purpose.

A major criticism of this procedure (as in the case of early Cobb-Douglas functions) is that the data are non-experimental. But as explained in the above paragraph, the usual experimental data are not necessarily ideal for this type of analysis either, since full application to extension or policy issues would require an experimental design which would allow at least real-world magnitude changes in all the important input variables. Since the data for the firm functions <u>are</u> non-experimental, considerable care and judgement are required to select appropriate observations. The observed input levels are the results of resource owners' decisions to produce and are not subject to <u>ex ante</u> control by the researcher. Careful <u>ex post</u> selection by the researcher, however, may yield a set of real-world data comparable to experimental data in terms of the factor levels and production units to which the data relate (29, p. 187).

Other recent production functions for agriculture have been estimated from county data as opposed to firm data in the above case. The essential difference between the two formulations is simply that the latter is based on an aggregate (simple sums) of the firm input-output records. Historically, the data source for these functions has been the Census of Agriculture. A disadvantage over the firm level data above is that the researcher can not exercise as much selectivity in obtaining appropriate observations. It is difficult, if not impossible, to take account of differences in management skills. (This problem is also encountered in firm functions where cross-section farm survey data are used.) An additional difficulty is that it is more difficult to select aggregates of farms with similar products.

Recent studies based on county data include attempts by Griliches (20) to isolate the effects of labor quality differentials (measured by level of education) on agricultural production. Headley (28) attempted to measure the effects of agricultural pesticides and Ruttan (42) estimated regional agricultural production functions and the demand for irrigated acreage. These studies have typically made use of cross-sectional rather than time-series data.

Use of firm and county data in agriculture has some advantages over the early Cobb-Douglas functions for U. S. industry. An important difference is the nature of the industry with which the researcher is working. Since agriculture more nearly resembles the pure competition model, the required assumption of equal price ratios for the firms making up the aggregate is more plausible than for other sectors of the economy. Homogeneity of input and output mix may be approximately maintained with careful selection of observations. Another important difference is that the input set has been more completely specified. The specification usually includes labor, cropland, and various forms of fixed and operating capital. (Capital is specified according to the kind of capital which is most important in the area under study.) Nevertheless, essentially the same kinds of criticisms have been made against these agricultural production functions -- namely conceptual problems, measurement problems, and estimation problems. The conceptual problem is as follows. If there exists one function for the industry as we suppose, and if pure competition exists as required to make sense of the value function, then why would we expect to observe more than one point cross-sectionally? Questions have been raised, depending on the study in reference, as to the measurement and combination of inputs which are obviously not of a homogeneous nature. In addition, the requirements for mathematically consistent aggregation are not strictly followed in the combination of variables and in the specification of the industry function at the outset. Questions have also been raised as to the appropriate measure of fixed capital to use in the estimation process and the implications of different measures. Statistical questions have been raised regarding the lack of repeat observations associated with non-experimental data. The question is also asked why we would expect any more error associated with the value of output (taken as the dependent variable) than with any of the independent variable set. Attempts which have been made to deal with some of these problems are cited in the following section.

Typical Measurement Procedures for Agricultural Production Functions

A direct empirical correspondence to the conceptual production function is not available at the firm or higher aggregation level. In practice, firms do not produce a single product but several; all inputs are not clearly separable nor distinctly defined. Some inputs which appear to be variable and entirely "consumed" in the present production process may, in fact, leave a residual which is carried over into a future production period. Fixed inputs may exhibit an unobservable service flow in a particular production period and consequently, are difficult to quantify.

Typically, the problem of multiple outputs has been treated by using output prices as weights and summing over these value products to obtain a total value of output. (As pointed out earlier, these are not experimental data, but rather the values generated by economic decisions to produce.) The justification for this procedure is to group firms within the aggregate so that they reflect the production of the same set of products in relatively fixed proportions, implying that firms all face essentially the same set of output prices. If it can be assumed that all firms in the aggregate produce a fixed proportion of the various outputs, the problem is eliminated. If neither of these assumptions are realistic, then some correction should obviously be made. Griliches (22) attempted to adjust for differences in product mix by explicitly accounting for the differences in percent of output that is accounted for by livestock and livestock products. Mundlak (36) demonstrated the use of an implicit production function for the case of different product mix between firms, where he made use of regression and covariance analysis (as well as instrumental variables) applied to both cross-sectional and time-series data.

Measurement of the labor variable has been a source of criticism in estimating production functions. The typical approach is to estimate the total input in hours or man-year units, without accounting for quality differences. Lack of adequate data have prevented refinement of the specification. The problem is really two-fold; (1) the agricultural labor force is composed of hired, family, and operator labor, and (2) large discrepancies in productivity may exist both within and among the three components. Griliches (22) attempted to establish whether education was a significant factor in labor productivity. The difficulty with trying to adjust for quality differentials (labor or otherwise) is that the adjustment requires a priori information about the relative productivity of the different components. This is precisely the information we seek from the analysis in the beginning. Conceptually, we could include each component as a separate variable and derive the separate productivities. From a statistical standpoint, however, this is not usually a real alternative since the number of possible variables is restricted by the number of observations.

The measurement of the land variable has typically been strictly an acreage measurement without consideration of differences in land quality. Griliches (22) used the interest on value of land as a measure of the service flow from land -- assuming that land value reflects the quality differentials. Ruttan (26, p. 38) used two variables to represent the land input -- dryland acres and irrigated acres.

Other forms of capital (e.g., machinery and farm buildings) also present a measurement problem. Conceptually, only the service flow from "fixed" capital items should be entered as an input in the present production period. In practice, the value of the stock of capital has been used as a "proxy" for the service flow. In other cases, a simple, annual depreciation rate has been used to represent the flow from fixed capital.

Interpretation of the Agricultural Production Function Estimated from County Data

There exist many levels of aggregation and two types of data from which to estimate functions; thus, it is important that the reader understand the interpretation of the function estimated from cross-sectional data. A cross-sectional approach using county data is taken in this study. The underlying assumptions required to make "economic sense" of such aggregate functions are related to the use of replications in experimental design. A simple example will help convey the idea.

Assume that we want to estimate a production function for fertilizer in the production of corn on a particular farm. Two possibilities exist for obtaining repeat observations. We could produce several (say ten) crops on the same acreage under controlled greenhouse conditions, varying the levels of fertilizer over time. Alternatively, we could isolate ten "identical", one-acre tracts of land to provide observations and vary the levels of fertilizer between tracts. To the extent that the tracts are identical and other factors (e.g., plowing between tracts) are invariant between tracts, the difference in yields would measure only the response due to fertilizer. The basic assumption required is that everything not explicitly accounted for in the functional relationship is quantitatively fixed or unimportant. We must also require that units of both fertilizer and the other "important" factors are of a homogeneous quality.

Using the cross-sectional approach it is clear that we were thinking not of estimating an aggregate function for ten acres of corn in the sense of estimating total corn production from ten acres. We were thinking of estimating total corn production from one acre of corn from various levels of fertilizer, using ten tracts of land to provide observations -- and assuming that the ten acres were otherwise identical in all important respects.

In the present study, county cross-sectional data were used to estimate the production functions. It should be clear that these functions are county production functions. The reasoning for this conclusion is analogous to the corn example above. The functions are aggregate functions in that they represent input-output relationships for the aggregate of firm level input-output records. We assume that counties are homogeneous units[4] and that we are measuring output aggregates from different levels of aggregate input use, but that each county has the same production function and is operating at a unique position on it. The "homogeneous" county units provide cross-sectional observations from which to estimate a county production function.

---

[4]Considering counties as homogeneous units simply implies that for any two counties having the same quantity of homogeneous inputs (labor, machinery, cropland, etc.), output would be the same.

III. UNITS OF OBSERVATION, VARIABLE MEASUREMENT, AND FUNCTIONAL FORM

Three major components of the development and implementation of a model are the choice of the units of observation, decisions regarding variable measurement, and the selection of the functional equation form. The three major sections of this chapter are devoted to these topics.

Units of Observation

The choice of the units of observation partially depends upon the geographical, hydrologic, and climatic characteristics of the study area. The following description is given to enhance the reader's understanding of the area characteristics.

Description of the Study Area

The study area consists of the three states of Oregon, Washington and Idaho. The region is commonly known as the Pacific Northwest and includes most of the drainage area of the Columbia River basin within the United States, the portion of the Great Basin within Oregon, and the coastal areas of Oregon and Washington.

The region is physiographically diverse. Western Oregon and Washington are characterized by two parallel mountain ranges which extend from north to south through the two states. The coastal range parallels the ocean a few miles back from shore while 100 miles east the Cascade range extends the entire distance from northern Washington to southern Oregon. The Willamette-Puget Trough lies between the two ranges. East of the Cascades lie the basin and range area, including parts of the Columbia Basin, the Snake River Plains, and numerous intermountain valleys of the Rocky Mountain system.

The mountain system has a great impact on the region's climate. West of the Cascades the winters are wet and mild while summers are typically very dry. Annual rainfall varies from about 30 inches in the valleys to as high as 100 inches in areas along the coast. East of the Cascades, temperature extremes are greater and rainfall less. Although precipitation varies with elevation, annual averages are as low as eight inches in the central plains.

The average annual water runoff of the region is in excess of 200 million acre feet per year (64, p. 6-16-3). About 54 maf/yr. originates in Canada. Major ground water aquifers capable of providing supplies for irrigation, municipal, and industrial uses underlie about one-fourth of the region. Total irrigated acreage in the region was estimated to be over four million acres in 1966. Approximately 1.4 million acres were irrigated in Oregon, with 1.5 million and 1.3 million in Washington and Idaho, respectively. Both ground water and surface sources are important, but the major supply of irrigated water comes from surface sources. An average annual 5.4 million acre feet of stream flow depletion is estimated for the states of Washington and Oregon. An additional .5 million acre feet are depleted from ground water sources in the two states. An estimated 8.5 million acre feet of stream and ground water depletion is expected in a recent typical year in Idaho.

The study area consists of 157.2 million acres of land (58, 59, 60, state Table 1) of which 79.2 million acres (32, p. 60) are national forest lands. Approximately 19.2 million acres of land are cultivated in crop production (26, p. 70) and the remaining 58.8 million acres include range, forest and waste land which are important in livestock production, wildlife habitat, and in providing various forms of recreation. In general, the region has a very highly diversified output of agricultural products. Agricultural production west of the Cascade Mountain range in Oregon and Washington is predominantly dairy and livestock products near the coast and highly diversified (field crops, vegetables, fruits, and nuts) in the Willamette Valley and northward into Washington. Livestock production and field crops are important in eastern Oregon and Washington as well as most of Idaho.

Delineation of the Study Area by Homogeneous Subregions

The three-state study area was divided into five county groups or subregions. The delineation was based on the type of farm output which was most prevalent. The five subregions are designated Areas

A, B, C, D, and E, and are characterized by the dominant types of farm output as follows:

(1) Area A contains 41 counties which typically produce field crops, and livestock and livestock products;

(2) Area B is composed of 15 counties which produce primarily livestock and livestock products;

(3) Area C is composed of 20 counties which produce mostly field crops;

(4) Area D (27 counties) produces mostly livestock, and dairy and livestock products, and

(5) Area E (16 counties) is highly diversified in its production (see Figure 1).

The procedure for grouping the counties was based on the percent of the total value of farm products sold (TVFPS) from the various Census classifications of farm output. The Census classification includes the following:

1. All crops (AC)

   (a) field crops (FC)
   (b) vegetables (V)
   (c) fruits and nuts (FN)
   (d) forest products (FP)

2. All livestock and livestock products (ALLP)

   (a) poultry and poultry products (PPL)
   (b) dairy products (DP)
   (c) livestock and livestock products (LLP)

AREA A

Oregon

Benton (56)
Crook (64)
Gilliam (46)
Jefferson (58)
Klamath (70)
Malheur (73)
Morrow (47)
Umatilla (48)
Union (49)
Wallowa (50)
Wasco (44)

Washington

Grant (17)
Klickitat (33)
Yakima (29)

Idaho

Bannock (110)
Bear Lake (117)
Boundary (74)
Butte (92)
Camas (98)
Canyon (95)
Caribou (111)
Cassia (114)
Clark (93)
Clearwater (80)
Custer (91)
Elmore (97)
Fremont (94)
Gooding (103)
Idaho (83)
Jefferson (100)
Jerome (108)
Kootenai (76)
Lincoln (104)
Minidoka (105)
Oneida (115)
Owyhee (112)
Payette (88)
Teton (102)
Twin Falls (113)
Valley (85)
Washington (87)

AREA B

Oregon

Baker (61)
Douglas (66)
Grant (60)
Harney (72)
Lake (71)
Wheeler (59)

Washington

Asotin (37)
Ferry (4)
Kittitas (16)
Pend Oreille (6)
San Juan (0)

Idaho

Adams (84)
Blaine (99)
Boise (90)
Lemhi (86)

AREA C

Oregon

Linn (57)
Sherman (45)

Washington

Adams (24)
Benton (30)
Columbia (35)
Douglas (11)
Franklin (31)
Garfield (36)
Lincoln (18)
Spokane (19)
Walla Walla (34)
Whitman (25)

Idaho

Benewah (77)
Bingham (106)
Bonneville (107)
Latah (79)
Lewis (82)
Madison (101)
Nez Perce (81)
Power (109)

AREA D

Oregon

Clatsop (38)
Columbia (39)
Coos (65)
Curry (67)
Deschutes (63)
Josephine (68)
Lincoln (53)
Tillamook (40)

Washington

Clallam (7)
Clark (32)
Grays Harbor (12)
Island (9)
Jefferson (8)
King (15)
Lewis (23)
Mason (13)
Pacific (22)
Snohomish (9-1)
Stevens (5)
Thurston (20)
Wahkiakum (26)
Whatcom (1)

Idaho

Ada (96)
Bonner (74)
Franklin (116)
Gem (89)
Shoshone (78)

AREA E

Oregon

Clackamas (52)
Hood River (43)
Jackson (64)
Lane (62)
Marion (55)
Multnomah (42)
Polk (54)
Washington (41)
Yamhill (51)

Washington

Chelan (10)
Cowlitz (27)
Kitsap (14)
Okanogan (3)
Pierce (21)
Skagit (2)
Skamania (28)

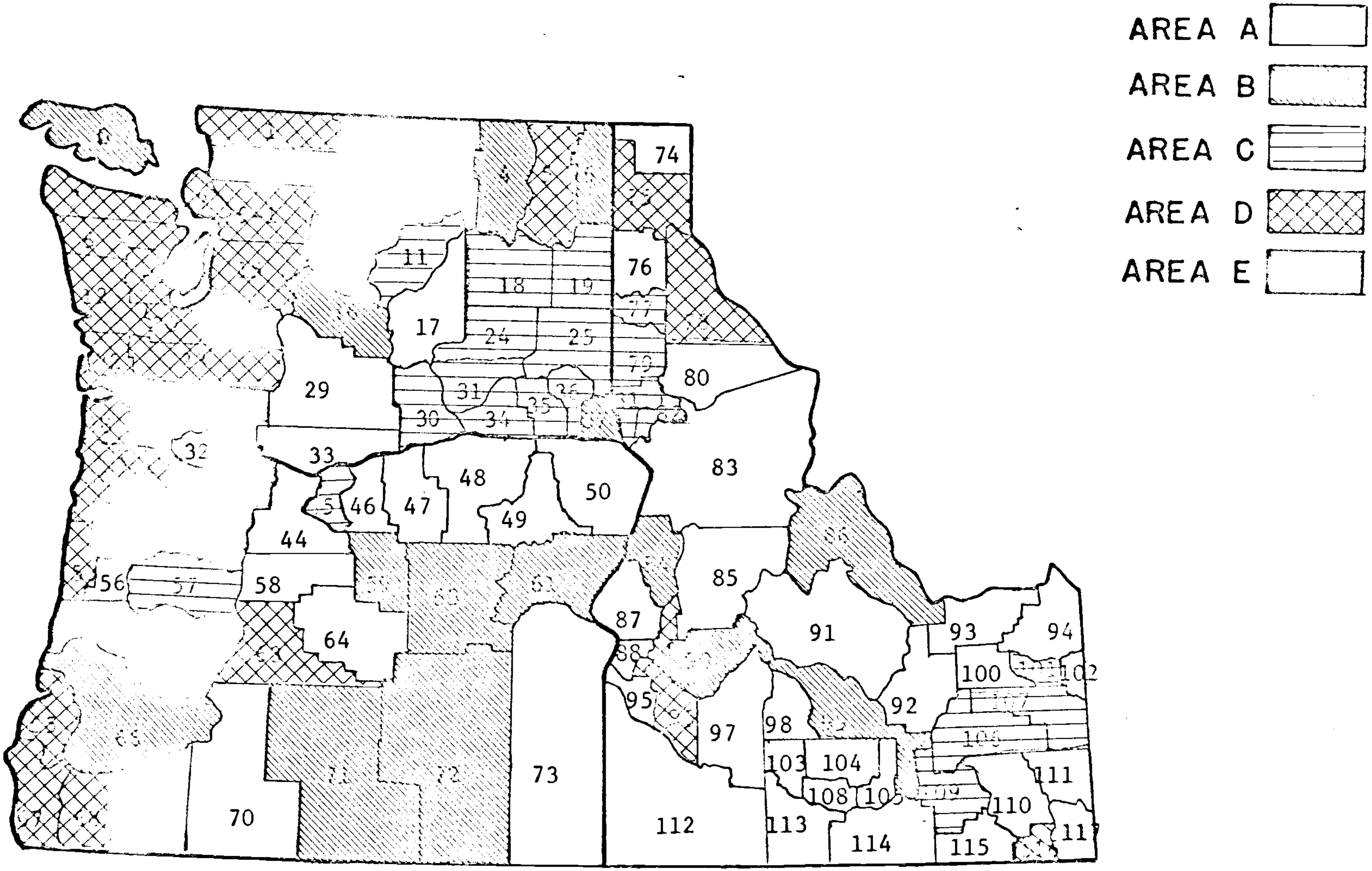

Figure 1. The Pacific Northwest and Five Homogeneous Farming Areas

Area A contains counties with greater than 50% of TVFPS from FC plus LLP, where the percent from FC and from LLP is greater than 20%. Area B contains counties with at least 50% of TVFPS from LLP and less than 20% from any other single source. Area C contains at least 50% of TVFPS from FC and less than 20% from any other single classification. Area D contains counties with at least 50% of TVFPS from ALLP and not less than 10% from DP and not less than 10% from LLP. Area E contains the remaining counties which exhibit a diversity of TVFPS between the seven classifications.

The rationale for this delineation is to group production units which have similar production relationships and input-output prices in order to reduce aggregation bias. The two important factors in aggregation bias are constant input and output prices among observations and proportional input and output combinations. By delineating homogeneous farming areas according to type of farm output, the input combinations and prices of inputs and outputs are expected to be very similar, or at least more similar than if the entire Pacific Northwest was included in one category. Some differentials in prices, no doubt, exist in cases where transportation costs for some counties would be substantially greater than others in the area.

Another purpose of the delineation is to hold constant a set of output-oriented agricultural policy variables with which this study is not concerned. Price supports and allotment programs have considerable impact on the value of certain classes of agricultural production -- especially in certain "unusual" years. Since this study is concerned with the effects of certain subsidized water resource inputs in agriculture, it is necessary to delete the output policy effects.

The use of political boundaries (counties) is not ideal from a conceptual point of view since other units would be more important in defining an internally homogeneous unit. Political boundaries do provide some measure of internal homogeneity, however, since various farm programs are administered by county delineation. As a practical matter, county observational units were required because of data limitations.

Variable Measurement

The aggregate production function for each of the five farming areas was specified to include eight input variables. This specification allows for the explicit recognition of the water resource inputs -- irrigation, drainage, and water conservation practices -- which are the focal points of the study. A complete specification and appropriate measures of all the inputs were considered essential to "good" estimation.

Definition of Variables

The production function for each of the five homogeneous farming areas was specified as:

$$Y = f(X_1, X_2, \ldots., X_8) \qquad (3.10)$$

where Y = value of farm products sold plus value of home consumption ($1000)

X_1 = man years of family, hired, and operator labor

X_2 = value of current operating expenses, including feed for livestock and poultry, seed, bulbs and plants, fertilizer, gas, fuel and oil, machine hire, repairs and maintenance, and pesticides ($1000)

X_3 = service flow of capital on farms, including most types of mechanical equipment and farm buildings ($1000)

X_4 = cropland: quantity adjusted by a quality index (1000 acres)

X_5 = AUMs (animal unit months) of available grazing (1000) units)

X_6 = irrigation water application (1000 acre feet)

X_7 = service flow of farm investment in drainage ($1000)

X_8 = service flow of farm investment in water conservation practices ($1000).

For a detailed explanation of the data sources and procedures used, see Appendix I. Particular attention is given here to the measurement of the service flow of capital and the importance of quality differentials in land and labor variables.

Flow vs. Stock Concepts of Input Measurement

The measurement of capital assets in the cross-sectional production function presents some conceptual and operational difficulties. As mentioned in Chapter II, a common practice found in the literature is to use the stock value of capital assets as a proxy variable for the actual portion of the input used in the present production period. This practice can legitimately be used only in a special case and is generally not satisfactory. Yotopoulas (65, p. 476) points out the fallacy of this approach and, at the same time, shows that the correct measurement can be calculated from information usually available. The proof and detailed explanation of his suggested procedures are presented in Appendix I. Griliches presents basically the same argument (19, p. 1417).

Capital is a multiperiod input of production and yields outputs in several time periods. The portion used in an early time period is small compared to the remainder to be allocated to future time periods and vice-versa in later years. In agriculture, capital usually constitutes a significantly large portion of the total input in the production process and thus should be measured properly if the analysis is to be useful.

Conceptually, it is clear that only the current service flow of capital inputs properly belongs in the input category of a production function estimated for the current time period. Ideally, in a perfect market situation, this amount would be equal to the rental price per unit of time, times the units of time the input is used in the production period. Data of this kind are not usually available. Data on the initial investment or survey data of current market value of the stock are usually the type of data available.

Use of the stock proxy as mentioned above is justified only on the basis of an assumption which requires that the stock be proportional to the flow. If this property holds, then no information is lost by the use of stocks instead of flows if the ratios are known. This can be seen from the following examples.

If the function is of the multiplicative form

$$Y = a\, S_1^{\alpha 1} S_2^{\alpha 2},$$

where Y = output

S_1 and S_2 are stock values of two inputs, and stocks are proportional to flows such that

$$S_1 = k_1 F_1 \text{ and } S_2 = k_2 F_2$$

where k_1 and k_2 are constants greater than zero and F_1 and F_2 are service flows corresponding to the above stocks, then

$$\begin{aligned} Y &= a(k_1 F_1)^{\alpha 1} (k_2 F_3)^{\alpha 2} \\ &= a(k_1)^{\alpha 1} (k_2)^{\alpha 2} F_1^{\alpha 1}\ F_2^{\alpha 2} = A\ F_1^{\alpha 1}\ F_2^{\alpha 2} \end{aligned}$$

where $A = a(k_1)^{\alpha 1} (k_2)^{\alpha 2}$

and the exponents α_1 and α_2 are unchanged -- the constant term absorbs the total effect.

In the case of an additive function of the form

$$Y = a + b_1 S_1 + b_2 S_2,$$

where $S_1 = k_1 F_1$ and $S_2 = k_2 F_2$ as before, a similar conclusion is drawn.

Substituting for S_1 and S_2 we have,

$$Y = a + b_1' F_1 + b_2' F_2$$

where $b_1' = k_1 b_1$, and $b_2' = k_2 b_2$.

The appropriate equation for the flow concept can be obtained from the stock equation without loss of information just by knowing k_1 and k_2.

The above use of stocks, then is valid when proportionately holds. This is not the usual case, since most capital items produce a variable flow of services over the life of the asset and the change in stocks (by deterioration), usually is at a different rate.

A summary of the results by Yotopoulos (Appendix I) is as follows:

1) When the service flow of an asset with a finite life span is constant over time (i. e., the sum of interest and depreciation is a constant, with the interest charges falling and depreciation charges rising as the asset ages), the use of stocks instead of flows places more weight on the more durable asset.

2) When the service flow of an asset with a finite life span deteriorates with time, the use of stocks instead of flows also places more weight on the more durable asset (proportionally more weight than in the case of a constant service flow).

3) A varying weight from stocks to flows may result (depending on their relative rates of change) in the case where assets (e.g., livestock) first appreciate with age and then depreciate.

4) When an asset has an infinite life span (e.g., land), stocks will remain proportional to flows and either measure will do.

The appropriate service flow for each of the four cases above were derived by Yotopoulos and are shown in Appendix I, equations 2a.12, 2a.15 and 2a.17. Equations 2a.12 and 2a.15 are the continuous form for the case where the service flow is constant, and the case where the service flow varies with time, respectively. Equation 2a.17 is the discrete, general case where market values are available at discrete points in time.

The above formulas were utilized in the calculation of the service flow for three of the fixed capital variables; capital service flow from machinery (X_3), drainage (X_7), and ACP (X_8). The data for X_7 and X_8 were available in a time series from 1940 through 1964. The service flow from any particular year's investment was assumed to deteriorate at a rate equal to the inverse of the expected life of the asset. Using formula 2a.15, the service flow for 1964 was calculated for each annual investment and individual service flows were summed to obtain a total 1964 flow for the asset. For example, if a particular aspect of drainage investment had an expected life of ten years and investments were made in each of the ten years preceeding 1964, then ten service flow estimates were made for 1964; the investment made in 1955 contributes its tenth and final flow increment, the 1956 investment contributes its ninth flow increment, and so on, through 1964. These components were summed to obtain a flow for 1964. This procedure should be a substantial improvement over an attempted current market valuation of the capital stock which is often used.

In the case of machinery capital, time-series data on investments by classes of machinery items were not available, so the discrete form, 2a.17, was used to calculate the 1964 service flow. That formula is

$$R_t = rV_t - (V_{t+1} - V_t)$$

where R_t = service flow in period t

V_t = market value at beginning of period t

V_{t+1} = market value at end of period t

r = discount rate.

The base data for this variable comes from Farm Income Situation estimates of capital consumption, which approximates the annual change in market value. $V_{t+1}-V_t$ was assumed to be equal to this change for 1964, for each county (see Appendix I for a detailed explanation).

Quality Dimensions of Input Variables

Quality differentials among units of an aggregate create essentially the same problem of aggregation bias as summing over firms which are not using proportionally the same input-output combinations. This problem has a certain potential of occurring on any of the aggregate variables but, given the present nature of the input classes in this study, it is most likely to be important in the labor and land variables. The other inputs, -- irrigation, AUMs, current operating expenditures, capital, drainage and ACP -- appear to be relatively homogeneous among units of the aggregates.

Labor quality differentials . . . The aggregation procedure used for labor assumes homogeneity of labor units within the three components -- family, hired, and operator labor. This variable has the potential of introducing considerable aggregation bias into the model. By delineating the farming areas according to type of farm output, however, much of the bias is expected to be reduced since we would expect proportionally the same amounts of hired, family, and operator labor among the counties producing nearly the same product.[5] Quality

[5] An analysis of variance of the labor components seems to support this conclusion. The ratio of family, operator, and hired to total labor is significantly greater between areas than within areas, at the 95% probability level.

differentials within each of the three components, however, is difficult to access. One indicator of quality differentials is education levels. Griliches (22, p. 419) found education to be an important factor in specifying agricultural production functions. His study, however, used regional data as observations for a national model. There would not be as much variation expected within the Northwest among counties as across the nation among regions. One could weight the labor components by an index of education, as did Griliches. The difficulty with this approach is that no real procedure exists to indicate the appropriate weights for different levels of education. For example, we need to know the productivity of twelve years of education as opposed to the productivity of six years.

To the extent that there exist significant differences in the productivity of hired labor between counties in a homogeneous farming area, the productivity of family labor between counties, and the productivity of operator labor between counties, then the coefficients for labor will be biased. Some bias will also enter because of the differences in the ratios of these components, but this is expected to be minimal. In general, the bias is not expected to be large because of the delineation of the homogeneous farming areas.

Land quality differentials . . . The variable which probably displays the greatest potential for a biased coefficient due to the neglect of quality differentials is land. Given the varied geography and all the various soil characteristics of the region, it is clear that the aggregation of units of land irrespective of their quality differences would be a meaningless measure.

A literature search on the subject suggests that no really satisfactory index is available. Griliches (22, p. 423) used the interest on the value of cropland as a measure for the cropland input. The procedure has the disadvantage of including "site" or "location" value and is not independent of other fixed assets on the land such as buildings, irrigation canals and underground pipe systems, and drainage facilities. Ruttan (42, p. 38) used quantity of cropland, distinguishing only between irrigated and nonirrigated land. Headley (28, p. 22) used the capital stock of land and buildings, which includes not only site value but is an inappropriate measure because of the influence of buildings and other improvements.

Ideally, what is needed is an index which measures the natural productivity of the unimproved soil. Such a measure is difficult to define and seemingly impossible to determine.

A different approach was taken in this study. To reduce the aggregation bias due to quality differentials, two procedures were followed: (1) A distinction was made between cropland and grazing land because of the unlikely possibility of defining a weighting scheme which would condense these two categories into homogeneous units -- thus, two variables were included; cropland and AUMs. (2) A cropland productivity index was constructed to transform the cropland acres to relatively homogeneous units. Thus, the cropland variable is the index-weighted quantity of cropland in the county.

The cropland quality index was constructed separately in each area in two steps: (1) A base county which grew crops most common to all other counties in the area was selected. Ratios of average county per acre yields for all common crops were calculated using the base county yields as the denominator. (2) The county land quality index was calculated by summing these ratios, weighted by the ratio of each county's acreage to the total area acreage for each common crop.

To illustrate the index construction, consider the following hypothetical example. Assume Area I contains two counties producing wheat and corn. Assume further that County 1 has average yields of four bushels of wheat and six bushels of corn; County 2 (the base county) has yields of six and eight, respectively. Acreages, of wheat and corn are; County 1 - ten and twenty acres, respectively, and County 2 - twenty and forty acres, respectively. The yield ratios are; County 1, 2/3 for wheat and 3/4 for corn, and County 2, one for both wheat and corn. Ratios of county acreages to the total are; County 1, 1/3 for both wheat and corn, and County 2, 2/3 for wheat and 2/3 for corn. The resulting indexes are .47222 for County 1 and 1.3333 for County 2. The ratio of these two indexes is invariant with the selection of the base county.[6] Appendix Table II lists the indexes by Area and County. Yields were taken from the Columbia-North Pacific Comprehensive Framework Study (39) and represented the normal yields for all counties in a recent average year.

The index used in this study is only a rough approximation of the ideal index but is expected to remove much of the problem of combining heterogeneous cropland units. This index is based on yield data which reflects, to some degree, the use of irrigation, fertilizer and the other inputs. Yields under strictly dryland conditions would perhaps have produced a "better" index; this procedure was not used since most crops are not produced at all under dryland conditions in some areas. The index based only on dryland yields then would have neglected the productivity advantage of much of the best cropland.

---

[6] In tabular form the example is shown as follows:

Land Quality Index Example

	Crop Wheat	Crop Corn	Index
Yield ratios			
county 1	2/3	3/4	
county 2	1	1	
Acreage ratios			
county 1	1/3	1/3	
county 2	2/3	2/3	
Index			
county 1			.4722
county 2			1.3333

Errors in the Variables and Specification Bias

Two major problems associated with variable measurement and inclusion of relevant variables are errors in the variables and specification bias. Both of these problems result in biased coefficients in the application of the ordinary least-squares (OLS) technique. The individual consequences and indications of the effects of these conditions are discussed in this section.

Errors in the variables. -- The underlying assumptions of OLS for the linear model $Y = X\beta + U$ (matrix notation) may be simply stated as follows (32, p. 107):

$E(U) = 0$

$E(UU') = \sigma^2 I_n$

X is a set of fixed numbers

X has rank $k < n$.

The first assumption states that the expected value of each U_i equals zero (this implies that U is a random variable with expectation zero, and we are not required to know its distribution). The second assumption implies that the variance of the U_i are all equal and that the U_i values are pairwise uncorrelated. The third assumption states that the X's are observed without error and are a set of fixed numbers. The fourth assumption states that the number of parameters to be estimated is less than the number of observations.

When the assumption of fixed X is violated due to measurement error, the observed values of X only approximates the underlying "true" values. Applicatoon of OLS in this case will yield biased estimates of β, even when the sample size is large (31, p. 149). The problem can be solved (i.e., unbiased coefficients may be obtained) only with some very special, additional information and some additional, more restrictive assumptions.

The data used in this study are almost certain to contain some measurement error since the data are from secondary sources designed for multiple uses. In the case of census data, the values used to derive the X's in this study are partially based on sample data which implies the existence of an error term.

One method of approaching the problem consists of using instrumental variables (proxy variables) which can be measured without error. The difficulty with this approach is that it may simply change the problem from one whose effects are rather obvious to one whose effects are mostly hidden. That is, the problem may simply be transformed to a problem of not knowing how good a representation of the true variable the proxy really is. In addition to this difficulty, it is not likely that even a proxy variable (in this study) could be measured without error. Another method for solving the problem requires assumptions about the distributions of the error terms for each X and for U. It also requires knowledge of the relationship between the variances of the error terms for each X and the variance of U (31, p. 151-175). Although this approach may hold good possibilities in some cases, it is not likely that such explicit information exists for the data in this study. About the best that can be done in this case is to recognize that, other things equal, the regression coefficients will be biased due to measurement error.

Specification bias. -- When unimportant variables are included in an OLS model, significance tests indicate that the coefficient is not significantly different from zero and we may drop the variable from the equation. When important variables are not included in the model, specification error is said to exist and the estimated coefficients of the variables in the equation will be biased. [For a proof of this statement, see Draper and Smith (11, p. 82).] This is the reason that the model must include all of the important variables and not just the ones of particular interest.

A related problem exists in this and other similar economic studies when no repeat observations for X exist. In this case, the residual mean square, $\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 / d.f.$, has expectation σ^2 only if the model is correct.[6] If the model is not correct, the expected value of the residual mean square contains a positive bias and will result in a tendency for smaller F and t values which are used in testing total

[6]The model is "correct" if $E(Y) = E(\hat{Y})$ for all values of X_i.

regression and individual coefficients for significance.[7] It is noteworthy that R^2 is not a reliable indicator of correct models since it is independent of the "lack of fit" sum of squares.[8] Without repeat observations on X (or perhaps prior information on "pure error") no precise method exists for evaluating model correctness. Consequently, the t and F tests have the distinct possibility of reflecting bias and must be used with the understanding that they are valid only if we assume the model is correct.

Functional Forms and Estimation Techniques

Several considerations are important in specifying the functional form of a production function. Some of these considerations are: (1) the compatibility with economic theory, (2) data limitations, (3) limitations in available statistical techniques, and (4) consistent aggregation. Several equation forms are discussed in the following sections. Each of these forms is evaluated with respect to the above considerations in the context of this study.

---

[7]The actual mean square values generated by a particular "incorrect" model may, by chance, not have a larger value since it has a random element.

[8]Given repeat observations on X, "pure error" can be estimated. The residual sum of squares used to calculate R^2 is the sum of, the sum of squares for pure error and the sum of squares for lack of fit and is invariant with any distribution of values between the two (11, p. 27).

Linear Functions

Functions which are linear in both the variables and the parameters were used in the study. These functions can be estimated with OLS and are compatible with the requirements for consistent aggregation (if we assume micro units have linear functions with the same slope). Thus, the available, simple sums data are applicable. Statistically, these functions are "efficient users" of degrees of freedom.[9] These functions, however, do not exhibit the usually expected diminishing returns and as a result, marginal products and marginal rates of technical substitution are constants. This functional form implies that the inputs are "independent" rather than substitutes or compliments in production. The elasticity of production depends on the constant and the level of output and is; (1) always equal to one when the constant equals zero, and (2) begins at zero when Y is equal to the constant and approaches one as Y increases without bound (provided the constant term of the equation is assumed positive). Although Cobb-Douglas functions have been the most widely used functions for the type of aggregate data used in this study, linear functions have been used with some success and considered superior by the researcher (33, p. 2).

Functions which are linear in the parameters but not in the

---

[9]Equations are said to be efficient users of degrees of freedom if the number of parameter estimates, relative to the number of variables is a minimum.

variables were tested on a portion of the data but were not given serious consideration for statistical reasons. These functions involving cross-product terms and powers of the X's (e.g., $Y = b_0 + b_1X_1 + b_3X_1X_2 + b_4X_1^2 + b_5X_2^2$) are highly inefficient users of degrees of freedom. The degrees of freedom were relatively important in this study since the total observztions for one area were as small as 15.

Cobb-Douglas Functions

Probably the best known functional form found in the literature on aggregate production functions is the Cobb-Douglas function which has the general form

$$Y = b_0X_1^{b_1}X_2^{b_2} \cdot \cdot \cdot X_n^{b_n}.$$

The parameters for this function were estimated by ordinary least-squares (OLS) techniques applied to the logs of the variables and by a non-linear technique in the real number form. The function is an efficient user of degrees of freedom. It also allows the possibility of consistent aggregation from simple sums data when the sum of the exponents equals one (in this case it is a homogeneous function of degree one). Diminishing factor returns are possible with the Cobb-Douglas function (they may also be continually increasing) but will be always negative or always positive. Marginal rates of technical substitution are symmetrical and vary directly with the ratio of the inputs. The elasticity of production is a constant over the entire function and may be $\gtreqless$ 1 but never any combination of the three. (Thus, it does not exhibit the qualities of the traditional textbook production function.) The elasticity of substitution is always equal to one and any two inputs are always, either substitutes or complements but never change over the function.

Cobb-Douglas functions were also estimated in their <u>real number form</u>. An iterative procedure was used to derive the parameters estimates for the functions. Some kind of an approximation technique is required since the function is nonlinear in the parameters and application of the usual linear regression techniques is not possible. The computer program used to estimate the parameters in this study makes use of the Standard Gauss-Newton mathematical method (11, ch. 10). Briefly explained, the method uses the first term of the Taylor Series expansion and a series of linear regressions to converge on the parameter values which minimize $\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$. Given the function

$$Y = F(X, P)$$

where Y is the dependent variable vector,

X the independent variable vector, and

P the parameter vector,

an initial estimate of P is selected and the relation

$$Y_i - F_i^o = a_1 \frac{\partial Y}{\partial P_1}_i + a_2 \frac{\partial Y}{\partial P_2}_i + \cdots + a_n \frac{\partial Y}{\partial P_n}_i + e_i$$

is estimated using OLS procedures to estimate a_i (i=1 . . . n). (The superscript on F indicates the function value for the i^{th} observation of X given the initial estimates of P.) The α_i estimate a change in the marginal contribution of P_i to the i^{th} deviation about Y, given the initial parameter set. The α_i are subtracted from the initial estimates of P and these values provide a new set of parameter estimates. The procedure repeats itself until the change in the P (i.e., α_i) is arbitrarily small and the last set of P is taken as the least-squares estimates.

The advantage of this technique over the log transformation procedure is that it allows the assumption of an additive error term which may be more realistic than a multiplicative one. A disadvantage is that the usual "t" and "F" tests are no longer strictly valid. However, in the case where the Cobb-Douglas is essentially linear over the range of the data the parameters may be tested by assuming an approximate "t" distribution for $\frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}}$.

Also, individual or groups of parameters may be tested by the "extra sum of squares" principal as presented by Draper and Smith (11, p. 67).

IV. PRELIMINARY ESTIMATES

The preliminary results of attempts to estimate the parameters of the linear and Cobb-Douglas functions are given and discussed in this chapter. Ordinary least-squares (OLS) regression was used to estimate the parameters for the linear and log linear functions. The parameters for the Cobb-Douglas functions were also estimated by a nonlinear regression technique. The two sets of parameters for the Cobb-Douglas functions are compared and analyzed for consistency with the implied assumptions. MVP estimates were calculated and compared among the three equations and among the five areas.

Multicollinearity tests were conducted for the linear equations to test the reliability of the statistical tests from OLS. Procedures by Farrar and Glauber (14) were used to test the models for multicollinearity. The results of these tests are discussed and compared among the equations for the five areas.

The Linear Functions From Ordinary Least-Squares Estimation

In four of five cases, the linear functions have larger R^2s than the log linear functions (see Table I). Only in Area E is the reverse true. Given the assumptions of ordinary least-squares, the estimated parameters for X_1 through X_5 for the linear function in Area A are

Table 1. Aggregate Agricultural Production Functions for Five Homogeneous Farming Areas, Pacific Northwest, 1964.

Function Form and Homogeneous Farming Area **	Constant	Parameter Estimates for X_1 (labor in man years)	X_2 ($1000 current operating expenditures)	X_3* ($1000 capital service flow)	X_4 (1000 acres cropland)	X_5 (1000 AUMS)	X_6 (1000 acre feet irrigation water)	X_7* ($1000 drainage service flow)	X_8* ($1000 service flow of ACP)	R^2 Natural	R^2 Log	Sum of Exponents for Cobb-Douglas Functions	Critical Values of Student's "t" .90	.95
Area A													1.697	2.042
Linear	-1142.336	5.184	2.920	-2.069	15.184	5.454	5.513	-263.436	17.801	.9925				
		(.767)	(.343)	(.543)	(4.580)	(1.580)	(2.845)	(146.870)	(33.902)					
Log of C.D.	5.048	.2076	.7637	-.1024	.1074	-.0019	.0861	.0022	-.0054					
		(.159)	(.097)	(.146)	(.045)	(.035)	(.025)	(.021)	(.029)	.9904	.9861	1.057		
Nat. of C.D.	6.946	.4470	.7248	-.3697	.0755	.0747	.1075	-.0395	.0568	.9937		1.077		
	(3.727)	(.087)	(.096)	(.117)	(.055)	(.038)	(.059)	(.031)	(.032)					
Area B													1.943	2.447
Linear	-37.275	2.355	3.078	.112	-5.076	1.050	1.246	102.430	-80.709	.9993				
		(1.155)	(.273)	(.280)	(1.711)	(.418)	(1.144)	(30.506)	(24.130)					
Log. of C.D.	4.066	.5814	.6733	-.1840	.0260	.1103	.0194	.1084	-.3289					
		(.327)	(.187)	(.319)	(.081)	(.044)	(.024)	(.046)	(.089)	.9980	.9947	1.006		
Nat. of C.D.	1.3906	.4336	.7035	.1811	-.0895	.1142	-.0257	.1517	-.4573	.9977		1.012		
	(1.010)	(.217)	(.162)	(.204)	(.073)	(.069)	(.054)	(.057)	(.163)					
Area C													1.796	2.201
Linear	-402.438	-4.931	2.738	1.240	8.583	-1.925	9.324	16.100	-1.812	.9809				
		(4.127)	(.526)	(1.636)	(3.801)	(5.736)	(3.729)	(71.098)	(5.097)					
Log of C.D.	3.162	-.6168	.5450	.8789	.1863	-.0253	.0752	-.0137	-.0460	.9845	.9793	.984		
		(.277)	(.164)	(.213)	(.068)	(.050)	(.028)	(.037)	(.034)					
Nat. of C.D.	.4425	-.6624	.7383	.9881	.1919	-.0199	.0277	-.0803	-.0586	.9824		1.125		
	(.449)	(.276)	(.174)	(.216	(.058)	(.048)	(.044)	(.055)	(.022)					

Table I. Continued.

Function Form and Homogeneous Farming Area	Constant	Parameter Estimates for X_1 (labor in man years)	X_2 ($1000 current operatn expend-itures)	X_3^* ($1000 capital service flow)	X_4 (1000 acres crop-land)	X_5 (1000 AUMS)	X_6 (1000 acre feet irriga-tion water)	X_7^* ($1000 drain-age service flow)	X_8^* ($1000 service flow of ACP)	R^2 Natural	R^2 Log	Sum of Expon-ents for Cobb-Douglas Functions	Critical Values of Student's "t" .90	.95
Area D													1.734	2.101
Linear	-325.217	2.721	1.927	-.750	20.368	2.010	3.028	6.811	-6.570	.9854				
		(1.895)	(.292)	(.567)	(11.785)	(3.588)	(3.558)	(55.423)	(22.698)					
Log of C.D.	1.187	.2987	.5993	.2465	-.0643	.0243	.0339	-.0045	-.0331					
		(.227)	(.152)	(.280)	(.110)	(.057)	(.033)	(.042)	(.038)	.9833	.9821	1.101		
Nat. of C.D.	2.6235	.4658	.7000	-.1511	.0141	.0554	.0172	-.0072	-.0167					
	(2.663)	(.238)	(.156)	(.262)	(.094)	(.056)	(.042)	(.061)	(.044)	.9827		1.078		
Area E													1.895	2.365
Linear	-22.417	6.432	1.611	-2.104	26.402	.254	2.831	-29.754	-19.642	.9907				
		(1.913)	(.593)	(.917)	(28.305)	(.190)	(12.570)	(55.991)	(82.588)					
Log of C.D.	10.880	.8607	.7301	-.7472	.0872	.0437	-.0198	-.0553	.0884					
		(.200)	(.188)	(.244)	(.069)	(.033)	(.051)	(.039)	(.845)	.9942	.9952	.988		
Nat. of C.D.	12.2475	.8967	.6809	-.7391	.0878	.0374	-.0239	-.0460	.0838					
	(12.370)	(.274)	(.242)	(.324)	(.102)	(.039)	(.040)	(.035)	(.066)	.9874		.978		

* These inputs were calculated using a 0.05 discount rate. Discount rates of 0.075 and 0.10% were also used but the estimated coefficients did not change significantly. Results from these regressions are available upon request from the author.

** Log of C.D. refers to the log form of the Cobb-Douglas function; Nat. of C.D. refers to the natural form of the Cobb-Douglas function.

all significant at the 99% level. $\hat{\beta}_6$ and $\hat{\beta}_7$ are significant at the 90% level, while $\hat{\beta}_8$ is significant at only 50%. The variable parameters for Area B are all significant at the 95% level except X_3 and X_6 which are significant at 40% and 65%, respectively. Area C has parameter estimates which are significant at 95% for X_2, X_4 and X_6. The parameter for X_1 is significant at 70%. In Area D, the parameter for X_2 is significant at 99%; $\hat{\beta}_1$ and $\hat{\beta}_4$ are significant at 90%. $\hat{\beta}_6$ is significant at only 65%. Area E has parameters significant at 99% for X_2 and X_3. The parameter for X_1 is significant at 95% while the level is only 75% and 60% for X_5 and X_4, respectively. $\hat{\beta}_6$, $\hat{\beta}_7$, and $\hat{\beta}_8$ are significant only at very low probability levels.

The Cobb-Douglas Functions

Cobb-Douglas functions were fitted to the data under two different assumptions. The results of the two procedures are discussed in the following sections.

Log-linear Functions

Estimates of the parameters for the function $Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \ldots X_8^{\beta_8} U$ were derived by transforming the function to the log form:

$$\log Y = \log \beta_0 + \beta_1 \log X_1 + \ldots \beta_8 \log X_8 + \log U.$$

This procedure assumes a multiplicative error term for the function in its original form. Thus, the derivations of R^2 and the significance tests are strictly applicable only in the log form of the equation. By calculating $\hat{Y}$ in the real numbers, given the parameters from the log fit, R^2s in the real numbers were derived and are recorded in Table I under the column labeled R^2, Natural. Only in some cases are the R^2s in the logs greater than the R^2s in the real numbers.[6]

There were generally fewer parameters significant at the 95% level for the log-linear functions than for the linear functions. In area A only three parameters were significant at 99% as opposed to four in the linear form. Parameters for X_3, X_5 and X_7 are significant at a very low level while they were significant at 90% in the linear equation. Parameter signs were negative for X_3, X_5, and X_8 instead of X_3 and X_7 as in the linear form.

In Area B, only $\hat{\beta}_2$, $\hat{\beta}_5$, $\hat{\beta}_7$ and $\hat{\beta}_8$ were significant at the 95% level while all except β_2 and β_6 were significant at the 95% level in the linear model. X_3 and X_8 had negative parameters instead of X_4 and X_8 as in the linear model.

---

[6] R^2 is defined as:

$$R^2 = \frac{y'y - e'e}{y'y} = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y' - n\bar{Y}^2} \quad \text{(matrix notation except for } n\bar{Y}^2\text{)}$$

where the lower case and upper case letters refer to the mean corrected and uncorrected sums of squares, respectively. Since β' are constants, whether R^2 decreases or increases with the log transformation depends on the relative size of the elements of X and Y. The log transformation reduces the absolute value of large numbers proportionately more than the small numbers.

Greater "t" values were obtained for Area C in the log form than in the linear. Only parameters for X_5, X_7, and X_8 fell below the 95% level while X_1, X_3, X_5, X_7 and X_8 were all below 95% in the linear model. However, four parameters (for X_1, X_5, X_7 and X_8) had negative signs while the signs were negative for X_1, X_5 and X_8 in the linear model.

Similar results are evident for Areas D and E. Signs were negative for parameters of X_4, X_7 and X_8 in the log form for Area D while negative signs appeared for X_3 and X_8 in the linear form. In Area E, signs changed from negative for $\hat{\beta}_3$, $\hat{\beta}_6$, and $\hat{\beta}_7$ in the log form to negative for $\hat{\beta}_3$ and $\hat{\beta}_7$ in the linear form. Several differences in significance levels were also present. It is interesting to note that R^2 was higher in the real numbers than in log form except in Area E.

Nonlinear Functions

The signs and magnitudes of parameters for the Cobb-Douglas functions estimated by nonlinear techniques were considerably different from the parameters of functions estimated by the log transformation technique. Also, the estimates of R^2 were slightly different and higher in the nonlinear case than in the log form, except in Area E. In general, signs of the parameters conform more closely with the parameter signs in the linear functions than with the log functions.

Bias in predicting Y was always slightly downward for the functions fitted by nonlinear techniques. Use of the log form parameters in the natural equation form produced slightly more bias in three out of five cases than did the parameter set generated from nonlinear techniques. These biases, in percentage terms, are presented in Table II. No serious bias is encountered for either case.

Table II. Precent Bias in Cobb-Douglas Production Functions for Five Homogeneous Farming Areas, Pacific Northwest, 1964.

Area and Equation	Percent average bias in predicting Y	
	% Upward	% Downward
Area A		
Log of C. D.*		1.37
Nat. of C. D.*		.14
Area B		
Log of C. D.		.10
Nat. of C. D.		.001
Area C		
Log of C. D.		.73
Nat. of C. D.		1.05
Area D		
Log of C. D.	.14	
Nat. of C. D.		.42
Area E		
Log of C. D.	.30	
Nat. of C. D.		.07

* Log of C. D. refers to the log form of the Cobb-Douglas function; Nat. of C. D. refers to the natural form of the Cobb-Douglas function.

Marginal Value Product Estimates

The marginal value product (MVP) estimates for the five areas, based on the sample information, are shown in Table III. The estimates for the Cobb-Douglas functions are evaluated at both the geometric and arithmetic means of $\hat{Y}$ and X_r ($r = 1 \ldots . 8$). In areas A, B and D, the MVP estimates between equation forms were most consistent (of the three possible comparisons) between the linear estimates and the natural Cobb-Douglas estimates. Similar statements can not be made about Areas C and E.

It is significant to note the relative stability and instability of the various MVP estimates among the three equations in each area. The MVP for labor (X_1) is relatively stable (especially in Areas D and E) and indicates "reasonable" estimates of returns to labor, except for Area C where large negative returns are indicated. MVP estimates for current operating expenditures (X_2) are the most stable among the three equations. In general, the magnitude of the MVPs indicate under-investment in variable expenditures. [This finding is fairly consistent with those of Headley (28), Griliches (20), and Ruttan (42, p. 102-109).] Machinery capital (X_3) is relatively unstable. MVP estimates change signs in two out of the five cases. The negative signs indicate that too much capital is being used -- even to the point of reducing total value product; this conclusion seems very

Table III. Marginal Value Product Estimates for Three Equation Forms in Five Homogeneous Farming Areas, Pacific Northwest, 1964.

Area and equation *	Marginal Value Product of X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
	($1000/man yr.)	($/$)	($/$)	($/ac.)	($/AUM)	($/ac. ft.)	($/$)	($/$)
A) Linear	5.184	2.920	-2.069	15.184	5.454	5.513	-263.440	17.801
Log of C.D. **								
G. mean	2.421	2.644	-.355	8.799	-.087	5.485	20.295	-5.136
A. mean	2.714	2.711	-.461	11.643	-.105	4.219	12.828	-5.510
Nat. of C.D. **								
G. mean	5.202	2.504	-1.277	6.173	3.405	6.823	-363.677	53.913
A. mean	5.914	2.605	-1.686	8.287	4.179	5.323	-233.188	58.680
B) Linear	2.355	3.078	.112	-5.077	1.050	1.246	102.430	-80.709
Log of C.D.								
G. mean	5.264	2.156	-.403	1.830	1.545	1.262	305.146	-82.132
A. mean	6.014	2.299	-.476	1.856	1.234	.537	132.847	-84.624
Nat. of C.D.								
G. mean	3.926	2.253	.396	-6.299	1.600	-1.672	427.036	-114.196
A. mean	4.485	2.402	.468	-6.390	1.278	-.712	185.912	-117.660
C) Linear	-4.931	2.738	1.240	8.583	-1.925	9.234	16.100	-1.812
Log of C.D.								
G. mean	-8.862	1.751	4.021	7.668	-2.813	26.058	-44.376	-27.091
A. mean	-9.085	1.903	4.131	7.299	-2.567	6.094	-23.190	-9.952
Nat. of C.D.								
G. mean	-9.079	2.262	4.312	7.534	-2.111	9.156	-248.106	-32.920
A. mean	-9.726	2.569	4.629	7.495	-2.012	2.237	-135.492	-12.638

Table III. Continued.

Area and equation*	Marginal Value Product of X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
	($1000/man yr.)	($/$)	($/$)	($/ac.)	($/AUM)	($/ac. ft.)	($/$)	($/$)
D) Linear	2.721	1.927	-.750	20.368	2.100	3.028	6.811	-6.570
Log of C.D.								
G. mean	2.315	1.356	.607	-16.408	.863	10.987	-7.685	-13.122
A. mean	2.714	1.360	.694	-14.537	.834	4.384	-52.684	-11.496
Nat. of C.D.								
G. mean	3.621	1.589	-.374	3.610	1.974	5.593	-12.336	-6.642
A. mean	4.209	1.580	-.423	3.170	1.890	2.212	-8.439	-5.768
E) Linear	6.432	1.611	-2.104	26.402	.254	2.832	-29.754	19.642
Log of C.D.								
G. mean	6.705	1.949	-2.765	20.049	2.954	-7.422	-62.810	90.455
A. mean	7.248	2.003	-3.185	15.834	.758	-5.686	-31.874	74.609
Nat. of C.D.								
G. mean	6.576	1.711	-2.575	19.002	2.380	-8.433	-49.180	80.715
A. mean	7.523	1.862	-3.138	15.885	.668	-6.839	-26.417	70.469

* G. mean indicates the geometric mean of $\hat{Y}$ and X_r; A. mean indicates the arithmetic mean of $\hat{Y}$ and X_r.

** Log of C.D. refers to the log form of the Cobb-Douglas function; Nat. of C.D. refers to the natural form of the Cobb-Douglas function.

"unreasonable".

The MVP for cropland (X_4) is relatively stable in Areas A, C, and E, but very unstable for Areas B and D. "Reasonable" estimates of returns to cropland are shown in the three areas which show stability -- but negative returns are sometimes indicated in Areas B and D.

MVP estimates for AUMs (X_5) are relatively stable between equations within areas but not among areas. Negative signs appear in Areas A and C.

Stability for the MVP estimates of irrigation (X_6) are indicated in Areas A and D but not in Areas B, C, and E. The magnitudes of the estimates vary somewhat between areas where stability is indicated, but appear "reasonable" when compared with estimates derived from other studies. [Ruttan's study (42, p. 40) indicates returns to irrigated acreage in the Pacific Northwest at about $30 per acre above that for dryland, which is approximately equivalent to $10 per acre foot of water applied].

MVPs for drainage and ACP. (X_7 and X_8) are highly unstable (and "unreasonably" large) in Areas A, C, and D. They are also "unreasonably" large in Areas B and E.

Testing Ordinary Least-Squares Assumptions

The consequences of errors in the independent variable set and specification bias have already been discussed (see Chapter III).

Another problem often encountered in economic data (especially aggregate data) is the lack of assumed independence in the independent variable set. This problem is referred to as multicollinearity. The instability of the above MVP estimates, the instability of the estimates of parameter standard errors as other variables enter the equations in a step-wise fashion (not shown here but exhibited on computer printouts of the regressions), and the relatively high simple correlation coefficients between the X's (see Table IV and Appendix Table VII) are all indications of the existence of multicollinearity in the models. The following section explains the problem of multicollinearity and its relation to specification bias and errors in the variables. The section also provides tests for the existence of multicollinearity.

Multicollinearity

The specific purpose of regression analysis is to estimate the parameters of a hypothesized dependency relationship (at least in a single equation model) between a dependent variable and a set of "independent" variables. It is not designed to isolate the effects of an interdependency relationship within a set of variables; hence, the assumption in OLS of an independent set of X. When this assumption is violated to a significant degree, multicollinearity is said to exist and the expected results are not generated by the application of OLS. The following discussion related to the detection, measurement, location, and causes of multicollinearity is primarily due to Farrar and Glauber (14, p. 92-107).

The interdependency condition defined as multicollinearity ...

"can exist quite apart from the nature, or even the existence, of dependence between X and Y. It is both a facet and a symptom of poor experimental design. Multicollinearity constitutes a threat--and often a very serious threat--both to the proper specification and the effective estimation of the type of structural relationship commonly sought through the use of regression techniques" (14, p. 93).

The difficulty with a multicollinear set of X's is that as the interdependency becomes more severe, the correlation matrix (x'x) approaches singularity[7] and the elements of $(x'x)^{-1}$ explode; consequently, so do the estimates of the parameter variances corresponding to the linearly dependent members. Farrar and Glauber (14, p. 93) explain the results of the extreme case as follows:

---

[7] It should be acknowledged that the OLS model $Y = X\beta + U$ may be mean corrected and then transformed so that the least-squares estimates are

$$b' = (x'x)^{-1}x'y,$$

where (x'x) is now the simple correlation matrix of the original set of X and x'y the simple correlations of X and Y.

The b' has a variance-covariance matrix, var (b'), and the usual least-squares estimates, $\hat{\beta}_1$, are expressed by

$$\beta_i = b'_i \left(\frac{\Sigma\, Y^2}{\Sigma\, X_i^2} \right) \qquad (i=1, \; . \; . \; . \; k \text{ parameters})$$

where ΣY^2 and ΣX^2 refer to the mean corrected sums of squares (11, p. 147).

> "The mathematics, in its brute and tactless way, tells us that explained variance can be allocated completely arbitrarily between linearly dependent members of a completely singular set of variables, and almost arbitrarily between members of an almost singular set. Alternatively, the large variances on regression coefficients produced by multicollinear independent variables indicate, quite properly, the low information content of observed estimates. It emphasizes one's inability to distinquish the independent contribution to explained variance of an explanatory variable that exhibits little or no truly independent variation."

The application of OLS to a multicollinear variable set results in (1) parameters which are very unstable and as a result very sensitive to changes in model specification, and (2) "t" values that become small as multicollinearity increases.

The most important results of multicollinearity may be more far reaching and undermine the entire empirical research process--primarily because of its potential impact on model specification. The researcher brings to the research problem a pre-conceived idea of the functional relationship between a dependent variable and a set of "independent" variables (including which independent variables are important). This pre-conceived idea comes from a combination of theory, prior information, and possibly some intuition. This pre-conceived model is tested on the data -- the results of which are used to modify the prior information, and thus, the model specification. This procedure continues until an "acceptable" model is found which is some compromise between the initial prior information and the sample information as discrepancies between the two decrease to some tolerable level. As the sample information becomes increasingly interdependent, however, parameters estimated from the sample become increasingly unstable and sample significance decreases. If the researcher has little confidence in his prior beliefs and is unaware of the extent of the multicollinearity in the sample information, he will have a tendency to over-adjust his prior information. This often leads to the exclusion of important variables and oversimplified models since increasing the complexity of the model usually increases the multicollinearity problem and decreases the "t" values of more and more variables.

The existence of multicollinearity also presents an additional complication when measurement problems are present. In this case not only are the estimated parameters biased but the variances of the parameters (due to multicollinearity) are large and "t" tests are even more misleading.

To avoid these problems (or at least reduce their effects), it is essential that the researcher be able to detect the extent, magnitude and patterns of multicollinearity. Solutions to the problems, once these characteristics are known, are not well established at this point in time, but the consensus of opinion among economists indicates some approaches. Farrar and Glauber support this conclusion.

> "Economists are coming more and more to agree that . . . correction requires the generation of additional information. Just how this information is to be obtained depends largely

on the tastes of an investigator and on the specifics of a particular problem. It may involve additional primary data collection, the use of extraneous parameter estimates from secondary data sources, or the application of subjective information through constrained regression, or through Bayesian estimation procedures" (14, p. 92).

A series of three tests designed by Farrar and Glauber was used to determine (1) the degree of overall model multicollinearity, (2) the location of multicollinearity, i.e., which variables cause the problem, and (3) the patterns of collinearity between variables.

Assuming that x is distributed as a multivariate normal, the quantity

$$\chi^2_{(v)} = -[N-1-\frac{1}{6}(2n+5)] \log |x'x| \qquad (4.10)$$

is approximately distributed as a chi-square with $v = 1/2\, n(n-1)$ degrees of freedom (14, p. 101). (x represents the n independent variables, each of which is normalized, by sample size and standard deviation, to unit length--thus $|x'x|$ is the determinant of the simple correlation matrix for the independent variable set.) N is the sample size and n the number of variables. The statistic, $\chi^2_{(v)}$, indicates the extent to which x is interdependent and may be judged significantly different from zero for different percentage points in the chi-square distribution--provided, of course, we accept the assumption of multivariate normality. $\chi^2_{(v)}$ under this assumption provides a <u>cardinal</u>

measure of the departure of x from orthogonality.[8]

The quantity

$$\omega = (r^{ii}-1)\,\frac{N-n}{n-1} \qquad v_1 = N-n \qquad (4.11)$$
$$v_2 = n-1$$

where r^{ii} = i^{th} diagonal element of $(x'x)^{-1}$

N = sample size

n = number of variables

is distributed as an F distribution with v_1 and v_2 degrees of freedom (14, p. 102). This statistic is calculated for each of the n variables, and indicates a cardinal measure of whether each is significantly collinear when we assume multivariate normality as before.

The quantity

$$t_{ij\,(v)} = \frac{r_{ij}\sqrt{N-n}}{1-r_{ij}^2} \qquad (4.12)$$

where

$$r_{ij} = \frac{-r^{ij}}{\sqrt{r^{ii}\;r^{jj}}}\quad,\ i \neq j \text{ and}$$

r^{ii} and r^{jj} = the i^{th} and j^{th} diagonal element of $(x'x)^{-1}$, is distributed as student's "t" with v = N-n degrees of freedom. This statistic is calculated for each possible pair of variables in the independent variable set and indicates, pairwise, the location of the

---

[8] The independent variable set is said to be orthogonal if the sums of cross products for all pairs of variables is equal to zero.

multicollinearity problems (14, p. 104). Each of the three tests provides a cardinal measure of multicollinearity given the assumption of multivariate normality[9]. If one does not wish to make this assumption, ordinal measures for all three tests are still present.

Multicollinearity Tests

Both linear and log-linear equations were tested for multicollinearity by the three-step procedure outlined in the above section. The test results for the linear equation of Area A are presented in Table IV; the tests for the remaining functions are shown in Appendix Table VII.

With reference to Table IV, the general model test implies that there is a problem in the model even at the .995 probability level. The locational test shows relative stability in the coefficients of X_4, X_5, X_7, and X_8, but those for X_1, X_2, X_3 and X_6 are almost certain to be affected by multicollinearity since F_7^{33} (.995) = 7.53. At 90%, all the variables are significantly affected since F_7^{33} (.90) = 2.56. The pairwise patterns indicate the major source of the problem to be between X_1 and X_2, X_3 and X_5, X_3 and X_6, X_4 and X_8, and X_6 and X_7. A summary of the results of the multicollinearity tests for all

---

[9]It should be recognized, of course, that the choice of an appropriate probability level is itself subjective, as in any statistical test.

Table IV. Multicollinearity Tests for the Linear Function of Area A, Pacific Northwest, 1964.

Test					Value of Statistic					Critical Values for the χ^2, F, and t distributions
General $\chi^2_{(v)}$					352.414					$\chi^2_{(28,\ .95)} = 41.3$ $\chi^2_{(28,\ .995)} = 51.0$
Location $\omega_{(v_1,\ v_2)}$		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$F^{33}_{7,\ (.90)} = 2.56$ $F^{33}_{7,\ (.95)} = 3.38$
		71.4	154.0	122.9	8.1	4.6	39.6	9.2	6.7	
Patterns* $t_{(v)}$		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$t_{(33,\ .90)} = 1.697$ $t_{(33,\ .95)} = 2.042$
	X_1		4.97	1.40	-.81	-.28	-2.71	1.03	-2.32	$t_{(33,\ .999)} = 3.646$
	X_2	.945		2.30	1.51	-1.51	2.43	.76	1.48	
	X_3	.913	.964		.98	2.75	3.61	.88	-.20	
	X_4	.464	.573	.535		-.79	-1.46	-1.49	3.44	
	X_5	.433	.466	.559	.270		.02	1.12	1.39	
	X_6	.738	.842	.881	.371	.430		-2.76	-1.14	
	X_7	.651	.602	.575	.352	.471	.276		1.65	
	X_8	.066	.176	.145	.590	.287	-.039	.354		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t value.

five areas for both linear equation forms is given in Table V. (These are objective tests except, of course, in the choice of the probability level, which is subjective. In this particular case, however, we can be almost 100% confident that multicollinearity does exist, and if it does, we know which variables most likely cause the problem.[10]

Alternative Approaches to Multicollinearity Problems

One obvious solution to the multicollinearity problem is to discard the data, or portions of it, and select new data. The multicollinearity characteristic of aggregate data, however, is wide spread and the probability of obtaining "better" data from secondary sources (sources similar to the original data in this study) is small.

An alternative exists in the use of prior information. Several kinds of prior information models exist which have been developed to incorporate different types and degrees of completeness of the prior knowledge. The purpose and application of prior information models

---

[10] It is interesting to note that the use of factor analysis to locate the variables which are most closely related indicates an interdependency in the linear equations between X_1, X_2, X_3 and X_6 for for Area A; X_1, X_2, X_3 and X_8 for Area B; X_1, X_2, and X_3 for Area C; X_1, X_2, X_3 and X_7 for Area D; and X_1, X_2, X_3 and X_7 for Area E. These results are almost identical with those in Table V. For a short treatment of the topic of factor analysis in economics and business research see Ferber and Verdoorn (15, p. 101). Some of the usual "rules of thumb" for detecting multicollinearity provided little or no help in identifying the problem.

Table V. Summary of Multicollinearity Tests Results for Two Equation Forms in Five Homogeneous Farming Areas, Pacific Northwest, 1964.

Area and Equation	Significant Equations from the General Test (probability level .995)	Significant Variables from the Localization Test (probability level .995)	Significant Pairs of Variables from the Pattern Tests (probability level .95 to .995)
Area A			
Linear	Yes	X_1, X_2, X_3, X_6	X_1X_2, X_3X_5, X_3X_6, X_4X_8, X_6X_7
Log-linear	Yes	X_1, X_2, X_3, X_6	X_1X_2, X_1X_3, X_5X_6, X_7X_8
Area B			
Linear	Yes	X_1, X_2, X_3	X_2X_7, X_6X_7, X_8X_7, X_8X_1, X_8X_2, X_8X_6
Log-linear	Yes	X_1, X_2, X_3	X_1X_2, X_1X_3
Area C			
Linear	Yes	X_1, X_2, X_3, X_4, X_6	X_1X_3, X_3X_6, X_3X_8, X_4X_5, X_4X_6, X_4X_8, X_5X_6, X_5X_8, X_6X_8
Log-linear	Yes	X_1, X_2, X_3	X_1X_2, X_1X_3
Area D			
Linear	Yes	X_1, X_2, X_3, X_4, X_7	X_1X_2, X_1X_3, X_3X_4, X_5X_8, X_6X_7, X_6X_8
Log-linear	Yes	X_1, X_2, X_3, X_4	X_1X_2, X_1X_3, X_2X_3, X_2X_4, X_3X_4, X_4X_5, X_4X_6, X_5X_6, X_7X_8
Area E			
Linear	Yes	X_1, X_2, X_3, X_4, X_7	X_1X_4, X_1X_8, X_2X_4, X_7X_4, X_8X_4
Log-linear	Yes	X_1, X_2, X_3, X_4	X_1X_2, X_1X_4, X_1X_8, X_3X_8, X_6X_7

is discussed in the following chapter where the use of a model by H. Theil (45) is explained and incorporated with the regression equations derived in this chapter.

V. MODIFIED STATISTICAL RESULTS

Prior Information Models

Economists have traditionally thought of empirical econometric models as being based strictly on sample information. This has not been strictly true. The tendency has been to look upon the empirical estimation process as the application of completely objective criteria to a set of relevant, factually accurate data which in turn, reveal the nature of the structure or process. This may be an overstatement, but it emphasizes the failure to explicitly recognize the importance and the traditional, but implicit use of prior information.

The process of specifying a function with a certain set of input variables implies that the researcher has prior information which indicates that another set of input variables are not important, i. e., parameters for the omitted variables are not significantly different from zero. More explicitly, prior information is often used to impose certain constraints on the selected model (e. g., requiring that the sum of the estimated exponents for the Cobb-Douglas function equal unity). Prior information is also used to define the variables themselves (e. g., labor is defined as the sum of family, operator and hired labor rather than defining the three components as separate variables). As already stated in an earlier section, the traditional approach to finding an "acceptable" function has been to begin with a particular model, test it on the data, and revise the model until some tolerable compromise between the two is found. The results of the compromise are such that the final model is not only acceptable statistically, but the estimated values are economically "in the ball park". The end result is a model which contains an indeterminate amount of prior information.

More recently, statistical models have been developed to explicitly make use of both prior and sample information. In this way, a potentially valuable source of information is not accidentally discarded or given an unknown weight in the estimation process. This approach openly reveals the two sources of information for evaluation by both critics and clientele.

Several techniques for incorporating prior information have been developed. They differ primarily because of differences in the type of prior information available. These techniques are designed for situations where; (1) exact prior information for an individual or group of parameters is available, (2) the information on parameters is of a statistical nature with known finite means and variances, (3) the prior information is represented by *a priori* distribution for a set of parameters, and (4) prior information is less complete and information exists only in the form of inequality restraints. Models for situation (2) were used in this study where an upper and lower bound was placed on the parameter with the mean and variance specified such that there is a very small probability of values existing outside the range. A model by H. Theil (46, p. 401) which allows the use of prior information on any or all of the parameters was selected. The model was used in this study to (1) explicitly include prior information in the estimation process, and (2) as a tool for the analysis of the OLS models which exhibit multicollinearity in the independent variable set.

Theil's Prior Information Model

Theil's model combines the usual sample information and prior information in a particular form. The usual OLS model is specified as

$$Y = X\beta + u \tag{5.10}$$

where Y is a T element vector of the dependent variable

X is a (T by Λ) matrix of independent variables

β is Λ element vector of parameters

u is a T element vector of the disturbances.

The usual OLS assumptions apply (see Chapter III).

In addition to the sample information, we also have prior information assumed to be of the following form:

$$r = R\beta + v \tag{5.11}$$

where r is a k element vector of estimates for $R\beta$ (where $k \leq \Lambda$).

R is a (k by Λ) matrix of known, nonstochastic elements and determines which parameters have prior information and how they are weighted.*

β is the Λ element vector of fixed, unknown, true parameters.

v is the k element vector of prior information errors.

The error vector v is distributed independently of the u vector and has the following (known) nonsingular matrix of second-order moments:

$$E(vv') = \psi . \qquad (5.12)$$

An example of the R matrix is $R = [I, O]$,

where I is a (k by k) unit matrix and O the $[k \text{ by } (\Lambda-k)]$ zero matrix. This matrix provides prior estimates for the first k elements of β, with equal weights.

The derivations for the model which combines the two sources of information makes use of Aitkin's generalized least-squares solutions.[11] When the true error variance is known, the estimates of the parameters for the combined model are best linear unbiased estimates when it is assumed that $E(v) = 0$. However, when the true

[11] Aitken's generalized least-squares is only slightly more complex than OLS. It consists of the following. $Y = X\beta + u$ as usual; where $E(uu') = \Omega$. The estimates for β are:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y; \text{ where } (X'\Omega^{-1}X)^{-1} \text{ is the matrix}$$

of second-order moments.

error variance (σ^2) is not known and we use in its place the estimate of the variance (S^2) from OLS, the parameter estimates are asymptotically unbiased when E(v) = 0. Also, the matrix of second-order sampling moments is asymptotically efficient. These two results indicate that the parameter estimates approach best linear unbiased estimates as the sample size becomes large (36, p. 406).

Theil's prior information model contains the following tests and estimates:

(1) $$\hat{\gamma} = (r-Rb)' [S^2R(X'X)^{-1}R' + \psi]^{-1} (r-Rb). \qquad (5.13)$$

$\hat{\gamma}$ is a test for comparability of the two information sources, where $\hat{\gamma}$ is distributed as a $\chi^2_{(k)}$ (k = number of prior parameter estimates).

(2) $$\hat{\beta} = (S^{-2}X'X + R'\psi^{-1}R)^{-1}(S^{-2}X'Y + R'\psi^{-1}r). \qquad (5.14)$$

$\hat{\beta}$ is the vector of asymptotically, best linear unbiased estimates of β derived simultaneously from sample and prior information.

(3) $$V(\hat{\beta}) = (S^{-2}X'X + R'\psi^{-1}R)^{-1}. \qquad (5.15)$$

$V(\hat{\beta})$ is the matrix of asymptotically efficient second-order moments (assuming E(v) = 0 and S^2 is random and unbiased).

(4) $$\theta_s = \frac{1}{\Lambda} \operatorname{tr} S^{-2}X'X(S^{-2}X'X + R'\psi^{-1}R)^{-1}. \qquad (5.16)$$

(5) $$\theta_p = \frac{1}{\Lambda} \operatorname{tr} R'\psi^{-1}R(S^{-2}X'X + R'\psi^{-1}R)^{-1}. \qquad (5.17)$$

θ_s and θ_p are the relative shares of the sample and prior information, respectively, assuming that the prior and sample information are independent.

(6) $$\hat{\sigma}^2 = \frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{T-\theta_s \Lambda}. \qquad (5.18)$$

$\hat{\sigma}^2$ is a new estimate of the model variance.

A partial explanation of equations (5.13 through 5.18) is given in the same order below:

(1) The comparability statistic, $\hat{\gamma}$ (when the covariances between the prior estimates are all zero) is expressed as

$$\hat{\gamma} = \sum_{i=1}^{k} \left[\frac{r_i - b_i}{\text{Var}(r_i) + \text{Var}(b_i)} \right]^2$$

Or, $\hat{\gamma}$ is the differences in the estimates from the two sources (divided by the sum of their variances), squared and summed.

(2) The new estimate, $\hat{\beta}_i$ (in the case where X is orthogonal) is equal to:

$$\hat{\beta}_0 = \left[\frac{1}{\text{Var}(r_0)} + \frac{1}{\text{Var}(b_0)} \right]^{-1} \left[\frac{\Sigma Y}{S^2} + \frac{r_0}{\text{Var}(r_0)} \right]$$

$$\hat{\beta}_1 = \left[\frac{1}{\text{Var}(r_1)} + \frac{1}{\text{Var}(b_1)} \right]^{-1} \left[\frac{\Sigma X_1 Y}{S^2} + \frac{r_1}{\text{Var}(r_1)} \right]$$

$$\vdots$$

$$\hat{\beta}_\Delta = \left[\frac{1}{\text{Var}(r_\Delta)} + \frac{1}{\text{Var}(b_\Delta)} \right]^{-1} \left[\frac{\Sigma X_\Delta Y}{S^2} + \frac{r_\Delta}{\text{Var}(r_\Delta)} \right]$$

(3) The new variance-covariance matrix for $\hat{\beta}$, $V(\hat{\beta})$, is the inverse of the sums of the inverted sample and prior, variance-

covariance matrices. In the case where the covariances for all the sample parameter estimates are zero, then $V(\hat{\beta})$ would reduce to

$$V(\hat{\beta}_1) = \left(Var(r_i) + Var(b_i)\right).$$

(4) and (5) Since $\theta_s + \theta_p$ necessarily equals unity, it will suffice to explain only one of them -- θ_p is easier. The proportional contribution of prior information to the new estimates is simply shown as:

$$\theta_p = \frac{\sum_{i=1}^{k} \frac{V(\hat{\beta}_i)}{Var(r_i)}}{\Delta}$$

And $\theta_s = 1 - \theta_p$. It is also true that for each parameter in $\hat{\beta}$, 100% of the information comes from two sources -- sample and prior information. Further, the percentage from prior information plus the percentage from sample information (for each β_i) equals $\frac{1}{\Delta}$ (100%).

(6) The new estimate of the variance is simply the deviation sums of squares for the model with the newly estimated set of parameters, $\hat{\beta}$, divided by its degrees of freedom. The degrees of freedom are equal to the total number of sample observations, minus a positive number between zero and Δ-- the number of OLS parameters. As the prior information approaches the sample information, θ_s approaches unity, the error sums of squares approaches that of the sample, and $\hat{\sigma}^2$ approaches the model variance derived from OLS.

Estimates From Prior Information Models

Simultaneous solutions for the parameters based on prior and sample information were obtained for the five homogeneous areas. Linear and log-linear models for the sample were used with Theil's model which is designed for linear models. Exact prior information was specified for the nonlinear models. Two sets of prior information were used with Theil's model and the results compared. First, prior information was specified as expected market equilibrium values of the parameters. Secondly, parameter estimates from other studies were used as prior information estimates for the parameters.

Prior information was obtained for all the parameters except β_0. The multicollinearity tests (Table IV and Appendix Table XI) indicate a high interdependency between X_1, X_3 and usually X_4, X_6 or X_8 at the .995 probability level. At the .90 level, however, the F tests are significant for all the variables except X_7 in Area C and X_5 in Area E; therefore, prior information was specified for β_1 through β_8.

Prior Information From Expected Market Equilibrium Conditions[12]

Prior information for expected equilibrium conditions requires input prices for the input sub-set measured in physical terms. The price of labor was determined for the area in general. Average yearly earnings per hired worker were estimated at $3, 000 (54, p. 15) while the average family and operator returns were estimated at $5, 600 per year (50, p. 86-87). A weighted average price for labor was then estimated for the Pacific Northwest using the percent labor used in each of these categories in 1964. The weighted average price was $4, 560 per man year. Land values were estimated using Statistical Reporting Service unpublished work sheets for Oregon (56) which contain estimates of value per acre of nonirrigated cropland by reporting units (eight units in Oregon). The value per acre of the unit corresponding most closely to the homogeneous areas defined in this study were used as an estimate of value per acre (it should be remembered that the cropland variable in this study is a proxy for a variable which would measure land quantity weighted by its "natural" productive capacity). The expected value of the parameter for land

---

[12] Whether observed market prices and quantities in any particular year are in fact market clearing values is a question for research. The quantities used here as prior information are values which the investigator judges (*ex post*) would have been the correct values for the aggregate of farmers to use to obtain an optimum solution.

was estimated as the market value per acre times five percent. The price per AUM was taken to be equal to the average price charged on Forest Service grazing lands plus associated permit costs (38). The price per acre foot of irrigation water was estimated to be $6.00. This estimate was based on unpublished research data (8) showing the pricing structures of three major Bureau of Reclamation projects in Oregon. It is difficult to estimate an average water price since irrigation districts use different pricing schemes, some of which are composed of a fixed charge plus a variable charge based on use rates. The range was always between $4 and $8 total cost per acre foot; thus, a uniform $6 price was assumed. The other variables were measured in terms of value of the service flow. The expected market equilibrium price would be equal to a dollar return per dollar of investment service flow. Thus, $1.00 was the prior estimate for X_3, X_7, and X_8. (The service flows include an assumed 5% return on undepreciated investment.)

Prior Information From Other Studies

The second set of prior information estimates was taken mostly from Ruttan's study (42, p. 40, 109). Ruttan's estimates for the Pacific Northwest region were utilized. Direct correspondence between Ruttan's variables and all the variables of this study was not possible. MVP estimates for labor, current operating expenditures, and nonirrigated cropland were taken as prior estimates for the parameters of the same variables in this study. Returns to three forms of capital in this study [X_3 (machinery), X_7 (drainage), X_8 (ACP)] were assumed to be equal to returns for current operating expenditures in Ruttan's study. He omitted the capital variable because of multicollinearity and assumed (42, p. 37) that current operating expenditures and capital were combined in fixed proportions.

The returns per AUM of grazing was taken to be slightly above the average fee and associated cost. Returns per acre foot of irrigation water were calculated by taking the difference between the returns per acre of irrigated and nonirrigated cropland [from Ruttan's study] divided by the average water application rate for the region.

Standard Errors for the Prior Information

Standard errors for the prior information estimates were determined by taking a range sufficient to include most of the probability for the parameter. This was accomplished by finding a standard error which would just satisfy the requirements for significance in Student's "t" test, given the degree of freedom for the particular equation. For example, the "expected equilibrium value" for labor (PI-1 for $\hat{\beta}_1$) in Area A was 4.560; the standard error for the labor coefficient was determined as

$$\hat{\sigma}_{r_1} = \frac{r_1}{t_{(v)\,.95}} = 2.685 \qquad (5.19)$$

where v = 32 and t refers to the appropriate point on the t distribution. The same procedure was used to determine prior information standard errors for the other variables and in the other equations with the appropriate degrees of freedom. The other components of ψ (the correlation coefficients between r_i and r_j, for $i \neq j$) were all assumed to be .01. This procedure is equivalent to saying that prior estimates of the parameters exist which, a priori, are expected to also have variances small enough to provide a significant statistical test at the 95% probability level.

Prior Information Model Solutions

The parameter solutions, measures of comparability, prediction bias, and percentage contribution from the two information sources are summarized in Tables VI, VII, and VIII. With reference to Table VI, the comparability tests fail at the 99.5% level for Areas A, B, and PM-2 of Area E [$\chi^2_{(8,\ .995)} = 22.0$]. Areas C and D, and PM-1 of Area E, however, show comparability between the two information sources. The $\hat{\gamma}$ values are not significant at $\chi^2_{(8,\ .90)} = 13.4$ for Areas C and D and not significant at $\chi^2_{(8,\ .990)} = 20.1$ for PM-1 of Area E. When the standard errors ($\hat{\sigma}_{r_i}$) were estimated using $t_{(v,\ .90)}$ in area A, comparability tests hold at the 99% level

Table VI. Prior Information Models for Five Homogeneous Farming Areas, With Two Sources of Prior Information for Linear Functions, Pacific Northwest, 1964.

Farming Area and Equation	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	R^2	$\hat{\gamma}$	$\hat{\sigma}^2$	θ_s	θ_p
				Parameter Estimates and Standard Errors										
Area A														
Linear Equation														
OLS $\hat{\beta}_i$	-1142.336	5.184	2.920	-2.069	15.184	5.454	5.513	-263.436	17.801	.9925		4,583,173		
S.E.		(.767)	(.343)	(.543)	(4.579)	(1.580)	(2.845)	(146.870)	(33.902)					
PI-1* $\hat{\beta}_i$		4.560	1.000	1.000	9.550	1.650	6.000	1.000	1.000					
S.E.		(2.685)	(.589)	(.589)	(5.630)	(.972)	(3.536)	(.589)	(.589)					
PI-2* $\hat{\beta}_i$		1.081	2.120	2.120	36.000	2.250	6.570	2.120	2.120					
S.E.		(.637)	(1.249)	(1.249)	(21.213)	(1.327)	(3.873)	(1.249)	(1.249)					
PM-1** $\hat{\beta}_i$	-2528.403	4.620	2.061	-.068	14.869	1.763	3.897	.964	.977	.9856	32.09	7,849,537	.5566	.4434
S.E.	(614.817)	(.600)	(.249)	(.300)	(3.074)	(.787)	(1.755)	(.588)	(.588)					
PM-2** $\hat{\beta}_i$	-1606.662	2.825	3.171	-1.116	13.047	2.851	3.736	2.153	2.107	.9854	31.22	8,056,756	.6118	.3882
S.E.	(692.820)	(.460)	(.263)	(.437)	(3.564)	(.974)	(1.876)	(1.249)	(1.249)					
Area B														
Linear Equation														
OLS $\hat{\beta}_i$	-37.275	2.355	3.078	.112	-5.076	1.050	1.246	102.43	-80.709	.9993		30,908		
S.E.		(1.155)	(.273)	(.280)	(1.711)	(.418)	(1.144)	(30.506)	(24.130)					
PI-1 $\hat{\beta}_i$		4.560	1.000	1.000	5.000	1.650	6.000	1.000	1.000					
S.E.		(2.347)	(.515)	(.515)	(2.573)	(.849)	(3.089)	(.515)	(.515)					
PI-2 $\hat{\beta}_i$		1.081	2.120	2.120	36.000	2.250	6.570	2.120	2.120					
S.E.		(.556)	(1.091)	(1.091)	(18.520)	(1.158)	(3.376)	(1.091)	(1.091)					
PM-1 $\hat{\beta}_i$	-107.840	-.845	3.794	-.341	-2.601	-.235	4.986	.978	.878	.9913	87.147	269,731	.7044	.2956
S.E.	(86.063)	(.469)	(.093)	(.151)	(1.353)	(.208)	(.418)	(.514)	(.512)					
PM-2 $\hat{\beta}_i$	-24.777	-.146	3.837	-.663	-3.657	.075	5.248	2.092	1.666	.9965	46.884	109,866	.7134	.2866
S.E.	(87.977)	(.383)	(.090)	(.141)	(1.530)	(.225)	(.411)	(1.086)	(1.084)					

Table VI. Continued.

Farming Area and Equation														
Area C Linear Equation	β_o	β_1	β_2	β_3	β_4 (Parameter Estimates and Standard Errors)	β_5	β_6	β_7	β_8	R^2	$\hat{\gamma}$	$\hat{\sigma}^2$	θ_s	θ_p
OLS $\hat{\beta}_i$	-402.438	-4.931	2.738	1.240	8.583	-1.925	9.324	16.100	-1.812	.9809		4,760,308		
S.E.		(4.127)	(.526)	(1.636)	(3.801)	(5.736)	(3.729)	(71.098)	(5.097)					
PI-1 $\hat{\beta}_i$		4.560	1.000	1.000	10.350	1.650	6.000	1.000	1.000					
S.E.		(2.540)	(.557)	(.557)	(5.762)	(.919)	(3.347)	(.557)	(.557)					
PI-2 $\hat{\beta}_i$		1.081	2.120	2.120	36.000	2.250	6.570	2.120	2.120					
S.E.		(.602)	(1.179)	(1.179)	(20.050)	(1.253)	(3.661)	(1.179)	(1.179)					
PM-1 $\hat{\beta}_i$	-1367.389	.917	1.562	.785	10.904	1.539	8.674	.994	.894	.9727	10.394	6,079,471	.4729	.5271
S.E.	(1,018.200)	(.628)	(.329)	(.318)	(1.637)	(.892)	(1.634)	(.556)	(.541)					
PM-2 $\hat{\beta}_i$	-1474.536	.739	1.762	.532	11.011	1.946	9.031	2.076	1.549	.9762	11.579	4,261,391	.5146	.4854
S.E.	(984.886)	(.586)	(.341)	(.473)	(1.800)	(1.192)	(1.723)	(1.179)	(1.082)					
Area D Linear Equation														
OLS $\hat{\beta}_i$	-325.217	2.721	1.927	-.750	20.368	2.010	3.028	6.811	-6.570	.9854		951,596		
S.E.		(1.895)	(.292)	(.567)	(11.785)	(1.482)	(3.558)	(55.423)	(22,698)					
PI-1 $\hat{\beta}_i$		4.560	1.000	1.000	14.400	1.650	6.000	1.000	1.000					
S.E.		(2.631)	(.577)	(.577)	(8.307)	(.952)	(3.464)	(.577)	(.577)					
PI-2 $\hat{\beta}_i$		1.081	2.120	2.120	36.000	2.250	6.570	2.120	2.120					
S.E.		(6.23)	(1.221)	(1.221)	(20.761)	(1.300)	(3.795)	(1.221)	(1.221)					
PM-1 $\hat{\beta}_i$	-545.756	1.750	1.628	.167	6.397	1.609	3.451	.963	.943	.9817	12.654	958,077	.5213	.4787
S.E.	(343.511)	(1.300)	(.211)	(.282)	(6.025)	(.754)	(1.729)	(.576)	(.576)					
PM-2 $\hat{\beta}_i$	-391.504	1.023	1.905	-.139	12.137	2.008	3.959	2.098	2.036	.9850	11.137	797,407	.5556	.4444
S.E.	(331.663)	(.584)	(.628)	(.117)	(9.192)	(.282)	(1.792)	(1.221)	(1.217)					

Table VI. Continued.

Farming Area and Equation	β_o	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	R^2	$\hat{\gamma}$	$\hat{\sigma}^2$	θ_s	θ_p
	Parameter Estimates and Standard Errors													
Area E														
Linear Equation														
OLS $\hat{\beta}_i$	-22.417	6.432	1.611	-2.104	26.402	.254	2.831	-29.754	19.642	.9907		2,432,095		
S.E.		(1.913)	(.593)	(.917)	(28.305)	(.190)	(12.570)	(55.991)	(82.588)					
PI-1 $\hat{\beta}_i$		4.560	1.000	1.000	18.000	1.650	6.000	1.000	1.000					
S.E.		(2.406)	(.528)	(.528)	(9.497)	(.871)	(3.162)	(5.28)	(.528)					
PI-2 $\hat{\beta}_i$		1.081	2.120	2.120	36.000	2.250	6.570	2.120	2.120					
S.E.		(.570)	(1.118)	(1.118)	(19.000)	(1.187)	(3.464)	(1.118)	(1.118)					
PM-1 $\hat{\beta}_i$	-1092.031	5.646	.654	.338	8.092	.141	5.720	.911	.900	.9806	18.525	3,119,799	.5177	.4823
S.E.	(775.242)	(.745)	(.209)	(.283)	(6.481)	(.146)	(2.927)	(.527)	(.527)					
PM-2 $\hat{\beta}_i$	-54.090	2.502	1.406	.038	26.439	-.027	9.179	2.040	1.982	.9421	45.082	9,449,018	.5355	.4645
S.E.	(791.202)	(.475)	(.241)	(4.66)	(7.804)	(.145)	(3.134)	(1.118)	(1.118)					

* PI-1 and PI-2 refer to prior estimates based on (1) expected equilibrium conditions and (2) previous studies, respectively.

** PM-1 and PM-2 refer to Thiel's prior information model solutions using PI-1 and PI-2, respectively.

Table VII. Mean Predicted Values of Y for Two Prior Information Models vs. Actual Sample Mean Values for Five Homogeneous Farming Areas, Pacific Northwest, 1964.

Area and Equation	Actual Mean Values of Y	Predicted Mean Values of Y	Average Precision (%)
Area A	16,553.00		
PM-1		16,551.20	99.99
PM-2		16,553.10	100.00
Area B	4,661.70		
PM-1		4,662.00	99.98
PM-2		18,417.50	99.99
Area C	18,416.10		
PM-1		18,411.60	99.98
PM-2		18,417.50	99.99
Area D	7,179.60		
PM-1		7,180.20	99.99
PM-2		7,179.20	99.99
Area E	16,492.80		
PM-1		16,493.20	100.00
PM-2		16,494.20	99.99

Table VIII. Percentage Contribution of Prior and Sample Information to PM-1 and PM-2 Parameter Estimates For Five Homogeneous Farming Areas, Pacific Northwest, 1964.

Area and Information Source	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	θ_p and θ_s
PM-1										
A										
Prior	0.0	0.55	2.00	6.00	3.32	7.28	2.74	11.08	11.08	44.05
Sample	11.11	10.56	9.11	5.11	7.79	3.83	8.37	.03	.03	55.95
B										
Prior	0.0	.44	.37	2.52	3.07	.67	.20	11.05	11.04	29.36
Sample	11.11	10.67	10.74	8.59	8.04	10.44	10.91	.06	.07	70.64
C										
Prior	0.0	4.56	3.89	8.31	.90	10.46	2.65	11.09	10.51	52.37
Sample	11.11	6.55	7.22	2.80	10.21	.65	8.46	.02	.60	47.63
D										
Prior	0.0	2.71	1.48	5.74	5.86	6.97	2.77	11.08	11.07	47.68
Sample	11.11	8.40	9.63	5.37	5.25	4.14	8.34	.03	.04	52.32
E										
Prior	0.0	1.06	1.74	7.96	5.18	.31	9.51	11.07	11.07	47.90
Sample	11.11	10.05	9.37	3.15	5.93	10.80	1.60	.04	.04	52.10
PM-2										
A										
Prior	0.0	5.80	.49	1.36	.31	6.00	2.61	11.10	11.09	38.76
Sample	11.11	5.31	10.62	9.75	10.80	5.11	8.50	.01	.02	61.24
B										
Prior	0.0	5.30	.08	.49	.08	.42	.16	11.06	10.96	28.55
Sample	11.11	5.81	11.03	10.62	11.03	10.69	10.95	.05	.15	71.45
C										
Prior	0.0	10.54	.93	3.88	.09	10.05	2.47	11.10	9.34	48.40
Sample	11.11	.57	10.18	7.23	11.02	1.06	8.64	.01	1.77	51.60
D										
Prior	0.0	9.75	.29	2.21	2.18	5.26	2.48	11.10	11.06	44.33
Sample	11.11	1.36	10.82	8.90	8.93	5.85	8.63	.01	.05	55.67
E										
Prior	0.0	7.72	.51	4.82	1.88	.17	9.08	11.09	11.10	46.37
Sample	11.11	3.39	10.60	6.29	9.23	10.94	2.03	.02	.01	53.63

without significant changes in coefficient estimates or in θ_s and θ_p. Due to the large values for $\hat{\gamma}$ in Area B (for both PI-1 and PI-2), the prior information models were rejected. Better prior information is evidently required for this area (the reader will recall that Area B is the livestock area).

Both prior model solutions for all five areas predict the average Y very precisely, as shown in Table VII. It should also be observed that R^2 (in terms of the sample) for both models in each area decreased less than 1% except for PM-2 of Area E (see Table VI).

Table VIII shows the relative contribution (from the two sources of information) to each new parameter estimate. (These results were obtained by disaggregating θ_s and θ_p from (5.16) and (5.17)). As one might expect from examining the results of Table VI, the major contribution from the prior information is concentrated in $\hat{\beta}_7$ and $\hat{\beta}_8$ in each area -- ranging from 41% in Area C[13] to 75% in Area B. The new parameter estimates for labor in PM-1 draw mostly on the sample information, except for Area C where prior information is about two-thirds as important as the sample information. The new machinery-capital parameter contains varying proportions of the two information sources among the areas. The new parameter for land is based

---

[13] With reference to Table VIII, Area C, and line "Prior", the sum of columns $\hat{\beta}_7$ and $\hat{\beta}_8$ equals to 21.60. Dividing 21.60 by 52.37 and multiplying by 100 equals 41%.

mostly on sample information in all except Area E where prior information is very influential.

The t values for the prior information models show a marked increase in cases where they were small in the OLS solutions. This is especially true for $\hat{\beta}_7$ and $\hat{\beta}_8$ where the new t values are usually significant at the 90% level. This result occurs for $\hat{\beta}_7$ and $\hat{\beta}_8$, of course, because of the dominant influence of the prior information. This is not always true however. For example, the t value for PM-1 of Area D (for X_6) changed from 0.85 to 2.00 while the prior information contributes only 25% to the value of the new parameter. The t value for $\hat{\beta}_4$ in Area E increased from .93 in the OLS solution to 1.25 in PM-1 and the prior information contribution was about 47%.

Comparison of the Two Prior Information Models

The results of the two models were not greatly different except in Area E where R^2 was significantly less for PM-2, and the new variance estimate was three times larger. The differences in the results in the other areas are primarily due to the differences in the prior information for $\hat{\beta}_7$ and $\hat{\beta}_8$ -- the PM-2 information being slightly nearer the sample information.

The PM-1 model was selected over the PM-2 model for several reasons. The prior estimates for PM-2 were determined from another study where variable definition is not really comparable (and is non-existent in the case of $\hat{\beta}_7$ and $\hat{\beta}_8$, in which case they were determined by implication only). In short, the prior information for PM-2 is considered inferior to that for PM-1 because, (1) the variable definitions in the study from which PI-2 information was taken are not comparable, (2) the study from which the PI-2 information was taken had problems of specification bias, errors in the variables, and multicollinearity, and (3) the results from using the PI-2 information was not greatly different from the results of using PI-1 information.

Prior Information Models and Multicollinearity in the Sample Models

The measures from Tables VI, VII, and VIII help point out the seriousness of the multicollinearity problem in the sample data. In some cases (Area C, for example) the PM-1 solutions are markedly different from the OLS solutions in that the parameters for X_1, X_5 and X_8 changed from relatively large negative values to relatively large positive values; also, $\hat{\beta}_7$ changed from a relatively large positive value (16. 1) to relatively small positive value (.994). Despite these changes, R^2 decreased less than 1% for both PM-1 and PM-2 and both predict Y with at least 99.90% accuracy on the average. Comparability tests also hold with probabilities as low as 80%. The sensitivity of the OLS solutions are now obvious; it is apparent that there may exist many sets of $\hat{\beta}$ in the neighborhood of the OLS solutions which do almost as well as the "BLUE" estimators.[14] Given this indication, the advisability of incorporating prior information -- in which case the relative contribution of each source is known and given -- becomes clearer and undue reliance is not placed on the OLS solutions as might otherwise be. The capability of the sample data to "reveal" information is more adequately assessed and utilized without forcing erroneous solutions.

Marginal Value Product Estimates from PM-1

The MVP estimates for the prior information models in four areas are shown in Table IX. (Prior models for Area B were rejected because comparability tests fail at very high levels of probability even when prior variances were specified relatively large.) Several very large changes from the estimates of the linear equations in Table III of Chapter IV are evident. MVP estimates for X_7 and X_8 generally changed from very large values (positive or negative)

[14]It should be pointed out that this "discovery" is by no means uniquely attributable to the prior information model itself. Even a "trial and error" procedure of trying slight variations from b would reveal the same thing, i.e., that although the OLS solutions do find the global minimum, in this case it is not a very "deep" minimum, which points up the potential magnitude of bias from even a small specification or measurement error.

Table IX. Marginal Value Product Estimates for Linear Prior Information Models for Four Homogeneous Farming Areas, Pacific Northwest, 1964.

Area and Equation	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
	(1000 $/man yr.)	($/$)	($/$)	($/ac.)	($/AUM)	($/ac. Ft.)	($/$)	($/$)
Area A								
PM-1	4.620	2.061	-.068	14.869	1.764	3.897	.969	.977
Area C								
PM-1	.917	1.562	.785	10.904	1.539	8.674	.994	.894
Area D								
PM-1	1.750	1.628	.167	6.397	1.609	3.451	.963	.943
Area E								
PM-1	5.646	.654	.338	8.092	.141	5.720	.911	.900

to estimates very close to the specified prior information.[15] X_1 and X_4 were relatively sensitive to prior information but X_2 and X_3 continue to show consistently high returns to current operating expenditures (X_2) and low returns to capital (X_3). Although the signs for the capital variable changes from negative to positive in Areas D and E, the irrigation (X_6) MVP did not change drastically from the estimates based only on the sample.

Prior information models were also tried in the log-linear functions. Parameter estimates from these prior models are not shown since the inverses required in (5. 13) and (5. 15) were so near singular as to render further calculations unreliable.

Exact Prior Information for Cobb-Douglas Functions

Exact prior information for the Cobb-Douglas functions was specified and the results examined. The rationale for fitting such functions is related to the relevant factors for selecting an appropriate functional form of the aggregate production function. These factors (as discussed in Chapter III) are economic usefulness, available statistical techniques, and consistent aggregation. These three factors are discussed in relation to the three functions used thus far.

[15] This is a natural consequence of solving the set of simultaneous equations where one equation has large coefficients and variances and another small coefficients and variances. See Appendix V for an example problem.

This establishes the reason for fitting exact prior information to the Cobb-Douglas functions.

From strictly a statistical point of view, there is little difference between the linear and log-linear functions if we make the usual OLS assumptions -- both functions yield BLUE estimators of the parameters. Both functions would have to be considered superior over the natural form of the Cobb-Douglas function. The natural form predicts quite accurately on the average as well as having R^2s which are about equal to those of the log form. In Areas A and D, R^2 was slightly greater than in the log linear function.

From strictly a mathematical point of view (given the simple sums data) the linear functions and the natural form of the Cobb-Douglas (given fixed input ratios at the firm level) probably contain less aggregation bias than the log-linear functions. The log linear functions would require geometric sums data to be consistent in the log form, but could not then be consistent with retransformation.

The Cobb-Douglas function in its natural form (whether parameters are estimated directly or by the log transformation technique) are more consistent with economic theory than the linear functions. This is not because economic theory disallows linear production functions, per se. Rather, it is because no data on purely technical input-output relationships existed for the study and the sums of economic decisions to produce, given the price system, were relied upon for data to estimate the aggregate production function. In a strict sense, the linear aggregate function is economically unrealistic since profit maximizing behavior was assumed for the firm operator. Given fixed input prices for the firm and the required assumption of firm functions with the same slope (for consistency), only corner solutions would exist; thus, economically speaking, the linear equation is untenable. As a practical matter, however, the linear function may be considered a "good" approximation over the range of the data.

An additional consideration is the use of the prior information models. Theil's procedure is applicable only to linear models. Attempts to use the model on the log-linear functions failed because of the near singularity of some of the matrices required in the calculations. Thus, the only models usable for prior information by Theil's procedure were the linear models.

In consideration of the above factors and due to the similarity of MVP estimates from the natural Cobb-Douglas and the linear functions (see Table III), exact prior information (based on PM-1 solutions for the linear functions) was specified for the Cobb-Douglas functions. Since prior information by Theil's procedure could not be incorporated directly into the Cobb-Douglas functions, another procedure was used. The prior information estimates of MVPs from the linear functions (Table IX) were used to specify exact prior estimates for all the parameters except β_0 in the Cobb-Douglas function.[16] The results of imposing exact, prior β values on the natural form of the Cobb-Douglas function are presented in Table X.

Comparing the functions with the natural form of the Cobb-Douglas functions (Table II, Chapter III), it is clear that the set of exact prior parameters (Table X) do almost as well as those of Table I. R^2 values are almost as high, and bias in predicting Y, on the average, is not serious. In fact, the bias for Area C, Table X is actually smaller at the mean than it's counterpart in Table I.

MVP estimates calculated at the arithmetic mean of $\hat{Y}$ from the functions in Table X were changed slightly from those of Table IX due to a small bias at the mean of $\hat{Y}$. If the equations had been completely unbiased at the mean of Y, the MVP estimates would be exactly the same as those from PM-1. The MVPs are presented in Table XI.

Summary

Theil's prior information model was selected as a method of introducing additional information into the regression analysis.

---

[16] Prior information was determined from the estimates in PM-1 of Table VI and calculated as $r_i = (MVP_i) \frac{X_i}{Y}$ where MVP_i is the i^{th} parameter from Table IV, PM-1. $\frac{X_i}{Y}$ is the ratio of the i^{th} input to total output, evaluated at the arithmetic means. Parameters with "t" values less than 1.0 in PM-1 were specified as zero.

Table X. Cobb-Douglas Prior Information Models for Four Homogeneous Farming Areas, Pacific Northwest, 1964.*

Area	Constant	Exponent of								$\hat{\sigma}^2$	R^2	Average Bias (%)
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8			
A	2.8179 (.039)	.3487	.5726	.0000	.1353	.0315	.0784	.0002	.0009	5,855,139	.9905	-3.14
C	7.4245 (.221)	.0618	.4441	.1658	.2763	.0151	.1063	.0006	.0041	8,302,004	.9668	.40
D	4.0612 (.080)	.1929	.7183	.0000	.0283	.0470	.0267	.0008	.0027	978,515	.9850	1.67
E	4.9300 (.123)	.6725	.2390	.0795	.0447	.0000	.0200	.0016	.0011	3,788,712	.9855	-.44

* Prior information is specified exactly for all the parameters except β_0.

Table XI. Marginal Value Product Estimates for the Cobb-Doublas Prior Information Models for Four Homogeneous Farming Areas, Pacific Northwest, 1964.

	MVP of							
Area	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
	($1000/man yr.)	($/$)	($/$)	($/ac.)	($/AUM)	($/ac.ft.)	($/$)	($/$)
A	4.475	1.996	.000	14.404	1.709	3.772	1.145	.902
C	.921	1.569	.788	10.942	1.551	8.699	1.027	.897
D	1.779	1.655	.000	6.496	1.637	3.505	.951	.952
E	5.621	.651	.336	8.056	.000	5.701	.915	.921

* Marginal value product estimates are different from those in Table 9 accordingly, as the mean of Y is different from the mean of the observed Y.

Additional information was needed since a considerable degree of multicollinearity was evidenced in the OLS models of Chapter IV.

Theil's model allows the specification of prior parameter estimates with variances. This was considered an improvement over specifying exact information as traditional approaches to the problem have required. In this study, for example, the variance for each prior parameter estimate was specified such that there is only a small probability (approximately .05) that the parameter would fall outside a range defined by two standard deviations about the mean (the prior parameter estimate). This is equivalent to saying, for example, that our prior information for labor in this study (Table VI) consists of the parameter estimate of $4,560 per man year, and further, that there is only a small probability (approximately 0.05) that the parameter value is less than zero or greater than $10,000 per man year; hence a variance of $15,370. Variances were similarly constructed for the other parameters.

Traditional approaches to the multicollinearity problem have been to (1) delete variables, or (2) combine variables. Both of these approaches involve the use of prior information -- exact prior information. When variables are deleted, the assumption is that the appropriate parameter estimate for the deleted variable is the same as the estimated parameter for the variable with which it was interdependent. When the deleted variable is not an exact linear combination of another included variable, however, a problem of specification bias occurs when the new model is estimated. When variables are combined, the specification problem associated with the deletion of important variables is eliminated but individual information (parameter estimates and variances) is lost. The only implication which can be made is that the parameters for the individual variables are the same (in the case where the combination is the simple sum of the variables). This implication is the result of exact prior information; i.e., that parameters are the same. No real information about individual parameters is gained from the sample.

Theil's model was used to incorporate two alternative sources of prior information; (1) expected equilibrium values (the parameter values which would have been optimum for the aggregate of farmers, given the observed set of prices), and (2) parameter estimates from other sources. Expected equilibrium values were judged "best" of the two prior information sets. In general, the overall contribution of prior information was about 50% except in Area B where the comparability test failed almost with certainty. Most of the aggregate contribution from prior information was due to the influence on $\hat{\beta}_7$ and $\hat{\beta}_8$ which contain about 99% prior information. Exact prior information was specified for the Cobb-Douglas functions (for all except Area B) based on the results of the linear prior information models. The reasons for specifying exact prior information in this case were:

(1) the Cobb-Douglas functional form was considered more appropriate than the linear equations from an economic theory viewpoint, and (2) the initial MVP estimates from the two equation forms were near the same values. The results in terms of MVP estimates were very near those for the prior linear model results, and the equation performance was almost equal the initial regression results for the Cobb-Douglas functions. R^2s were near the same values and bias in predicting Y on the average was not significantly different (compare Table X with Table I, Chapter IV).

The overall performance of both the linear and the Cobb-Douglas prior information models (Tables VI and X) has some important implications about the usefulness of the prior information. As a set, the new parameters performed (in terms of the data) almost as well as the OLS set from the sample. This indicates the instability of the OLS parameter set. The results indicate that the new parameter set could be the "correct" set and not conflict greatly with the derived sample information.

The individual contribution of prior information to the new parameters varies considerably between parameters. Given these relative contributions, some parameters may be judged statistically significant (from the sample statistics) since they are based primarily on sample information. Others may be judged significant only subjectively since they were based almost entirely on subjective prior information -- the variances of which were assigned such that the parameters would be significant if a valid stastical test existed. This applies particularily to $\hat{\beta}_7$ and $\hat{\beta}_8$. As a result, the estimates based almost exclusively on prior information will not be given attention in later sections on economic interpretation. To compare the results for $\hat{\beta}_7$ and $\hat{\beta}_8$ among areas, for example, would be futile since 99% of the estimates came from prior information which was specified to be equal in each area.

VI. ECONOMIC INTERPRETATION

Three major implications can be drawn from the MVP results of Table IX and Table XI. First, statements can be made about intraregional (within homogeneous farming areas) market equilibrium; i. e., whether (*ex post*) resources were allocated efficiently within the area. Secondly, since homogeneous farming areas are delineated by type of farm output, statements can be made about resource allocation between outputs, and more generally, allocation between farming areas. Thirdly, statements can be made about private and public returns per dollar investment in water resource development.

Intraregional Productivity Comparisons

The reader should be cautioned that any statements regarding market disequilibrium (or aggregate firm misallocation of resources) in this study necessarily refers to an *ex post* condition. As such, one should be careful in drawing inferences about the testing of static economic theory. We have only the *results* of (1) aggregate decisions to produce, (2) weather conditions, (3) interregional *price* effects, and (4) other factors that actually occurred. We do not have *ex ante* information on the decision-makers' set of alternatives, the constraints under which he operated, or his expectations of future values of the relevant variables. Simply because equilibrium was not observed does not necessarily imply that efficiency criterion was not used.

Farms in Area A produce mostly a mixture of crop and livestock products and occupies much of the area immediately east of the Cascade mountain range in Oregon and Washington, as well as some mountainous areas of central Idaho. Within Area A there is an indication (from Table IX and XI) of over-investment in (machinery) capital and a significant under-investment in current operating expenditures. Labor returns are very near the estimated price of labor for the region at large. Returns to land are approximately equal to a 7.5% return on an estimated $190 per acre land.[17] (This is an estimate of an exclusive return to land and does not include buildings, irrigation, drainage, and conservation improvements.) Returns per acre foot to irrigation are slightly less than the estimated price per unit. This estimate may reflect the relatively high level of irrigation development in the area. The area contains most of the major irrigation projects in the Pacific Northwest which have been developed by the Bureau of Reclamation. Included in the area are counties served by irrigation water from the Columbia Basin and Yakima Projects in Central Washington, the Crooked River and Deschutes projects in Central Oregon, the Vale, Burnt River, and Boise projects near the

[17] This land value is the estimated value of nonirrigated cropland from Statistical Reporting Service data (62) for a reporting area in Oregon roughly equivalent to Area A.

Oregon-Idaho boundary and portions of the Minidoka and American Falls projects in southeastern Idaho.

The prior information model for Area B was rejected because of the lack of comparability of the two sources of information. More precise prior information (specifically for this livestock region) would have been desirable, but better information was not available. Some information is available from the OLS solutions and the multicollinearity tests (see Table I and Appendix Table VI). Without consideration of the multicollinearity tests, coefficients for X_3, X_4 and X_6 would probably be rejected. Multicollinearity tests indicate that X_3 and X_6 are highly interrelated with X_1, X_2, X_7, and X_8 which tends to yield small "t" tests. So we have no real basis for rejection of the parameters resulting from multicollinear variables, since the t values would be larger in absence of the multicollinearity. Returns to current operating expenditures appear to be relatively stable, both among the functional forms and within the equations as other variables enter the regression in a stepwise fashion. That is, the coefficient and its standard error seem relatively independent of the absence or presence of other variables in the equation, and the first partial derivatives (evaluated at the means) are also relatively stable among the three equation forms. Estimates of returns to current operating expenditures (evaluated at the arithmetic mean in the Cobb-Douglas functions) range from \$2.29 to \$3.08 per dollar invested, indicating that expenditures were too small in this variable. Estimates for returns to AUM's and labor were slightly lower than expected returns, indicating that expenditures should be adjusted away from these variables.

Area C is composed primarily of southeastern Washington and a portion of eastern Idaho. These are semi-arid counties with concentrated irrigation development in limited areas. Referring again to Tables IX and XI, low returns to labor, high returns to current operating expenditures, and slightly low returns to machinery capital are indicated. Sample information contributes about 25% of the information for capital, 65% for current operating expenditures and 60% to labor. Returns to cropland are equivalent to 5.3% on land valued at $207 per acre. About 92% of this estimate is from sample information. The sample contribution to the MVP of irrigation is about 76%. The MVP seems somewhat above the likely cost per acre foot (approximately $8 per acre foot is a maximum charge by irrigation districts in the region). The relatively high returns to irrigation most likely reflect the high productivity of the limited water supply in this area. The production of field crops in the area evidently respond well to irrigation. Franklin County, Washington, for example, had irrigated wheat yields of about 80 bushels per acre while the overall average was only 37 bushels in 1964 (60, State Tables 13 and 14). In general, the implication regarding Area C is that too much labor and machinery capital are being used and not enough of current operating expenditures and irrigation.

The MVP estimates for Area D (the coastal area of Oregon and Washington) lead to about the same conclusions as in Area C. Too much labor and machinery capital are being used and too little current operating expenses. Cropland is only earning 2.3% return on land valued at $288 per acre. The MVP of irrigation water is a little below the minimum estimated cost of $4 per acre foot. This low MVP probably reflects the low productivity of irrigation water in an area where rainfall is mostly adequate. The average yield for irrigated hay (clover, timothy, mistures of cloves, and grass cut for hay) in Coos County, for example was 2.34 tons per acre while the overall average yield was almost as high (about 2.16 tons per acre). The MVP estimate of $3.45 per acre foot may indicate over development. The cost of irrigation in the area, however, is not doubt less than the estimated $6 per acre foot for the regional average since the development is almost entirely private stream use with low capital investment.

The MVP estimates for Area E (primarily the Willamette Valley of Oregon) indicate "reasonable" returns for labor and irrigation. The estimate for β_6, however, is about 85% prior information. The sample information estimate was considerably lower. This is an area where summer irrigation is crucial to many crops. The area is mixed with both private (stream and underground water) and public irrigation development. Development has evidently obtained, or slightly exceeded, the optimal level under 1964 prices and yields. Returns to machinery capital and current operating expenditures are low and returns to cropland indicate a 2.2% return on $360 per acre land. Prior information contributes less than 15% to $\hat{\beta}_1$ and $\hat{\beta}_2$ and less than 3% to $\hat{\beta}_5$. About 50% of $\hat{\beta}_4$ is due to prior information, while the estimate for $\hat{\beta}_3$ contains 71% prior information. In general, the MVPs indicate that investment should move slightly away from current operating expenditures, machinery capital, cropland, and AUMs to labor.

Interregional and Interproduct Comparisons

The results of Tables IX and XI indicate some discrepancies of resource allocation among regions and thus among types of farms. Labor returns are considerably higher in Areas A and E than in Areas B, C, and D. This could indicate that labor is "trapped" in Areas B, C, and D relative to Areas A and E. Another possibly explanation is that the diversity of types of production in Area E allows more efficient use of indivisible man years of family and operator labor. Many factors may be important in explaining these discrepancies and would require further study to isolate the causes. Without consideration of the causes or possible barriers to labor mobility we can only say that the discrepancy exists and efficiency could be increased by improving labor mobility between the areas (assuming, of course, that the cost of increasing the mobility is less than the benefits).

Interregional differences in returns to current operating expenses and (machinery) capital are not so different as to warrant discussion or draw inferences about misallocation.[18]

Returns to cropland indicate that future development should take place in Areas A and C rather than Areas D and E (assuming of course, that new land is available for development). Whether lands in Areas D and E should actually be taken out of production depends upon the opportunity costs for the investment and the "salvage" value of the land.

Irrigation returns indicate that future investment should probably take place in Areas C and E rather than the other areas (depending, of course, on the relative cost of development). Area C produces field crops under semi-arid conditions and contains a considerable quantity of irrigable, presently cultivated, cropland. Present irrigation is diverse throughout the area -- concentrated near large

---

[18]Statistical tests are available to test the differences in parameter values among the regions. The application of such tests would likely show no significant difference between the parameters among areas for current operating expenses and capital. The reliance of these tests (due to the influence of prior information) would be difficult to access and as such could be misleading; therefore, the tests were not conducted.

development projects in Area A. Area E contains counties where high valued fruit and vegetable crops are grown. The area has high annual rainfall but the dry summers make dryland farming very hazardous. Further project development in Areas A and D would appear unprofitable (at present private cost per unit), especially since the best alternative development sites are already taken.

Public and Private Returns From Water Resource Investments

Private Returns

The sample provided very little information regarding returns from drainage and ACP. These investments (as measured in this study) are a relatively small portion of the total agricultural capital investment in each of the areas, and as such, the data would not allow the isolation of the "independent" effects of these variables. One possible reason is that the output effects of drainage may be mostly "hidden" in the irrigation variable. The "t" values for X_6 and X_7 were significantly greater than zero at the 95% probability level in the multicollinearity tests in Areas A, B, D, and E (see Appendix Table VI). This indicates a substantial interrelationship. Drainage systems are required in many cases before irrigation is possible; thus, the two systems are often installed simultaneously. The ACP variable contains a group of different practices, some of which are no doubt output-increasing while others are output-decreasing. Returns to irrigation appear adequate to cover private cost (as indicated by irrigation district charges) at least in Areas C and E. The returns may be slightly below the per unit cost on Areas A and D. No reliable information was obtained for Area B concerning irrigation.

Public Returns

In an aggregate sense, we might expect public or private investments in any capital item anywhere in the economy to earn a return approximately equal to the return on investment at the margin in any other sector or area, if the marginal conditions for optimization are met. As a practical matter, however, we observe some parts of private industry earning 20% returns or greater, while returns on government investment may be quite low (perhaps almost zero in some cases). Some arguments in favor of this kind of allocation have been supported by suggestions that government policy may have a companion (or even primary) purpose of making income transfers, either from one sector to another, or from the present to a future generation. The latter argument is that government is the guardian of our resources and as such, may legitimately invest funds with higher opportunity costs in projects designed to retard resource use rates (and therefore current returns) to the benefit of future generations. Arguments have also been made for investments by government only if they earn a return equal to or greater than the opportunity cost in the private sector (see Baumol, 52 p. 489).

Given the MVP estimates of Table IX and XI, it is not likely that public investments in irrigation are earning an "acceptable" rate of return. The total current annual cost of irrigation in the mid-1950's (private plus public) as estimated by Ruttan (26, p. 46) was approximately $44 per acre of irrigated land in the Pacific Northwest. This is approximately equivalent to $10 per acre foot of water delivered. MVP estimates on the basis of per acre foot of water delivered were considerably below this figure in each of the areas.

Two additional factors may influence the total impact (in agriculture) of public investment in agricultural water resources. These factors are; (1) the effect on the value of the productivity of other factor inputs, and (2) the overall total revenue effect.

Related Input Effects

If we accept the linear functions as the appropriate functional form then related input effects do not exist. Inputs are classified as substitutes, compliments, or independents as the change in the marginal product of one input from a change in the quantity of another input is negative, positive, or zero, respectively. Mathematically, this is expressed as

$$\frac{d\,\dfrac{\partial y}{\partial x_r}}{dx_s} \gtreqless 0. \tag{6.10}$$

The quantity (6. 10) is, of course, zero for the linear equations. If we accept the Cobb-Douglas function as appropriate, then (6. 10) for all r and s, is always positive since,

$$\frac{\partial \hat{y}}{\partial x_r} = \frac{\hat{\beta}_r \hat{y}}{x_r}, \text{ for all r} \tag{6.11}$$

and,

$$\frac{d\dfrac{(\partial \hat{y}\ )}{(\partial x_r)}}{d\ x_s} = \frac{\hat{\beta}_r \hat{\beta}_s \hat{Y}}{x_r x_s} \quad , \text{ for all r and s.} \tag{6.12}$$

The quantity (6. 12) may also be written:

$$\frac{d\dfrac{(\partial \hat{y}\ )}{(\partial x_r)}}{d\ x_s} = \frac{MP_r}{AP_r}\,\frac{MP_s}{AP_s}\,\hat{Y} \tag{6.13}$$

where, MP_r = marginal product of input r.

AP_r = average product of input r.

Since $MP_r < AP_r$ for all r and all values x_r, and since we require both MP_r and AP_r to always be positive for the Cobb-Douglas function, then (6. 12) is always positive and the pair of inputs are complements over the entire range of x_r. Values of (6. 12) for irrigation are presented in Table XII.

Table XII. Estimates of Changes in the Value of the Marginal Product of Other Factor Inputs from a One Unit Change in the Use of Irrigation, Four Homogeneous Areas, Pacific Northwest, 1964.

Area and Variable	Change in MVP of x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
I. 1.0 Unit change in irrigation								
Area A	.0011	.0005	.0000	.0034	.0004	--	.0003	.0002
Area C	.0004	.0007	.0004	.0051	.0007	--	.0005	.0004
Area D	.0009	.0008	.0001	.0031	.0008	--	.0005	.0005
Area E	.0037	.0004	.0002	.0054	.0001	--	.0006	.0006

With reference to Table XII, we may infer that a marginal increase in irrigation has the greatest impact on the marginal productivity of land. The incremental change in land productivity is greater in Area C and E where crop production is most prevalent. Increased irrigation has its next greatest impact on labor productivity in Areas A, D and E.

The estimates from Table XII indicate the effects of incremental changes in irrigation on the other factor inputs. The magnitudes of these cross-effects, however, are limited by the equation form. The ratios of marginal to average products are not allowed to be greatly different due to the nature of the equation. Further investigation would be required to accurately assess these cross effects.

Value of Output Effects

An offer by the federal government to share the cost of agricultural water resource development is an immediate inducement for the individual firm to use more of the resource, produce more output, and use relatively more of the input than other inputs which are relatively more costly. The effects may be considered the same as a price decrease for the input in question. Table XII shows estimates of what happens to the MVPs of related inputs when irrigation is increased, given fixed output prices. How much prices of the inputs themselves change, and how much the resulting increase in output will effect output prices depends upon the elasticities of supply and demand in both factor and product markets. Although estimates of these elasticities is beyond the scope of this study, an article by Brandow makes possible some general comments (2, p. 898).

Given an aggregate production function which is homogeneous of degree one, and the assumption that the firm functions making up the aggregates are all the same Cobb-Douglas functions, homogeneous of degree one, Brandow shows some interesting results. In addition to the functional form of the production function, we make the usual assumptions that input supply functions to the industry of agriculture are positively sloped and that the aggregate demand for agricultural output is a negatively sloped function. We also assume equilibrium in both markets before and after a price change. The generalization may then be made that

> demand for an input is elastic, unit elastic, or inelastic accordingly as demand for output is elastic, unit elastic, or inelastic, regardless of what happens to the prices and quantities of other inputs, so long as the parameters of the system do not change (2, p. 898).

If the aggregate demand for output in a region is inelastic, as we usually suppose, and the output market is cleared, then a subsidy-induced increase in irrigation, drainage, or ACP may cause a decline in total revenue in <u>both</u> the agricultural input supply industry, <u>and</u> agriculture itself.[19] Whether net revenues in agriculture increase or decrease depends upon the relative size of the price changes in relation to the quantities of inputs used and outputs sold. It is conceivable, however, that the results of such a price subsidy may be to lower aggregate incomes in agriculture while inducing greater output and lower consumer costs such that the consumer of agricultural products may be receiving the total benefits from the water resource development in the form of consumer surpluses. Whether this would, in fact, happen depends upon the rate of growth in demand for agricultural products in relation to the rate of development of projects, and upon the industry production function.

---

[19]Increases in output under conditions of inelastic demand lead to smaller total revenues since the percent increase in output sold is less than the percent decrease in prices received.

VII. SUMMARY AND CONCLUSIONS

Summary

The competition for water resources among regional economic sectors has intensified in recent years. This rivalry and the long run prospects for water shortages have intensified the demand for research concerning the productivity of this resource in alternative uses. The distribution and use of water resources require investment which typically comes from both public and private sources. The primary purpose of this study was to evaluate the productivity of this investment (from both sources) in agriculture in the Pacific Northwest.

Estimation of production functions for agriculture in the area was selected as a method of analysis. Past studies seemed to indicate "acceptable" results from the use of models at a relatively high level of aggregation (macro level), as opposed to the micro approach in which implications (from selected micro units) are extrapolated to the aggregates. Precautions were taken in the measurement of variables, model formulation, and statistical estimation because of (valid) criticisms in the literature regarding various methodological aspects of production function analysis. An attempt was made to isolate the productivity of water resources in agriculture through separate productivity estimates for irrigation, drainage, and water-oriented conservation practices.

A major conceptual and methodological problem with aggregate functions is whether the data to be used are consistent with the functional form of the micro equations. Micro-level data were not available for aggregation by a selected procedure; neither were the micro-level functional forms known. The data available for analysis were already aggregated by a specific aggregation procedure (the simple sums of farm level data). The inability to select an aggregation procedure and the lack of knowledge of the functional forms of the micro functions implies a strong possibility that the aggregate functions are somewhat biased due to aggregation (i. e., the macro equation may not give the same results as the simple sum of the micro results). Although this possibility exists, there is no reason to believe that the resulting bias is any greater than the alternative of estimating a large number of micro equations and aggregating the results.

An attempt was made to measure the input from long term capital items in terms of the annual service flow rather than using a "proxy" variable (such as the stock of capital) as has been done in past studies. It has been shown that the use of stocks instead of flows is an inappropriate measure except in special cases such as land(65).

The aggregation of micro units of an input which does not exhibit homogeneity is a problem in some input variables. Variables which typically have this characteristic are labor and land. The delineation of the study area into homogeneous farming areas seemed to reduce this problem in the labor variable. A relative productivity index was constructed for the cropland variable; observations were defined as the quantity of cropland weighted by the index of cropland quality.

Ordinary least-squares (OLS) regression procedures were used to obtain the first estimates of the parameters. Multicollinearity tests indicated a strong probability of significant interdependence between several input variables, signifying that estimated standard errors were not minimum (compared to the standard errors if the variables were independent) and that the usual "t" tests for significance of the parameters were not reliable indicators of significance. This problem, combined with probable measurement error, prompted the conclusion that at least part of the parameter estimates and their standard errors were unreliable.

Prior information was selected as an alternative to problems associated with OLS solutions. Other alternatives include (1) gathering new data, (2) selecting new methodology, and (3) abandoning the project. Considering the restrictions on research time and expenditures, as well as the existence of techniques for incorporating prior information into the analysis, the prior information alternative was selected.

A prior information model developed by H. Theil was selected for use in the analysis because of its relatively low information requirements and ease of computations. The results of using this model helped to identify the seriousness of the multicollinearity problem of the OLS models. The new results exhibited a set of parameters which are economically more realistic, and statistically almost as "good" (by several measures) in explaining the variation in the dependent variable.

In general, the results indicated that any reallocation (or at least a redirection of future investment) in irrigation investment should move in the direction of Areas C and E (Area C is primarily a field crop producing area and Area E produces a high proportion of high valued fruits, nuts, and truck crops). The returns to irrigation were sufficient to cover all or most of the private cost, but were not sufficient to cover the public investment.

Estimated returns to other factor inputs generally differed from equilibrium values. Returns to labor were evidently low in three of the five areas; returns to current operating expenditures were usually high. Returns to machinery capital were generally low or insignificant. The attempt to separate the productivity of irrigation from the cropland was at least partially successful -- this has not been done in past studies. Returns to cropland were somewhat low in Area D and E, but were about as expected in Areas A and C.

Limitations and Implications for Further Research

The greatest shortcoming of this study is possibly the quality of the data available for analysis. The sample information alone does not exhibit reliable estimates of some of the parameters (particularly $\hat{\beta}_7$ and $\hat{\beta}_8$) generated from ordinary least-squares regression. Use of prior information models indicate that the new set of parameter values for drainage ($\hat{\beta}_7$) and ACP ($\hat{\beta}_8$) could be near the expected values without conflicting greatly with OLS solutions based only on the sample. Reliance on the estimates of $\hat{\beta}_7$ and $\hat{\beta}_8$ from the prior information model solutions is limited to the reliability of the prior information itself.

In addition to the interdependent nature of the county data, some measurement problems were encountered. The problem of an obviously heterogeneous labor input was treated by stratification of the study area. A land quality index was constructed to adjust the land input to a homogeneous unit. The problem of homogeneous specification of all the variables needs additional research.

The measurement of the drainage and ACP variables needs additional research effort. These variables are no doubt quite important in some areas but they are difficult to quantify. A "proxy" service flow (the assumption that farmer participation in ACP represented the private investment in these variables) was used in this study. Other physical measures such as acres drained or feet of drainage pipe, for example, are poor measures. A more complete accounting of investment in these variables is perhaps the best possibility. Another possibility is that the analysis could be done at a lower level of aggregation, considering only areas where drainage and ACP are a more significant portion of farm capital.

A further implication of the study is that the reliability of estimates from other studies of a similar nature, which have attempted to isolate the parameters of relatively unimportant variables perhaps, suffer from similar problems. Improved and alternative sources of data are no doubt required for the long run solution to the problem.

The use of prior information models seems to be a promising methodological tool to explicitly set out and utilize the researcher's prior information. Additional research in the use of prior information models would be helpful in determining the advantages and disadvantages of particular models.

BIBLIOGRAPHY

1. Baumol, William J. On the discount rate for public projects. In: The analysis and evaluation of public expenditures: the PPB System, comp. by U. S. Congress. Subcommittee on Economy in Government. Joint Economic Committee. Vol. 1, Washington, D. C., U. S. Government Printing Office, 1969. p. 489-503.

2. Brandow, G. E. Demand for factors and supply of output in a perfectly competitive industry. Journal of Farm Economics 44: 895-899. 1962.

3. Bronfenbrenner, M. Production functions: Cobb-Douglas, interfirm, intrafirm. Econometrica 12: 35-44. 1944.

4. Bryant, W. Keith and J. Frank O'Connor. Industrial-Urban development and rural farm income levels. Journal of Farm Economics 50: 414-426. 1968.

5. Burns, James H. Measuring the effect of irrigation on rate of technology change. Washington, D. C., 1967. 23 p. (Economic Research Service, U. S. Department of Agriculture. Agricultural Economic Report No. 125)

6. Ciriacy-Wantrup, S. V. Water Policy. In: Handbook of applied hydrology, ed. by Vente Chow, New York, McGraw-Hill, 1964. p. 28-3 to 28-9.

7. Cobb, Charles W. and Paul H. Douglas. A theory of production. American Economic Review 18:139-165. (March supplement), 1928.

8. Conklin, Frank S. Interview. Department of Agricultural Economics, Corvallis, Oregon. December, 1970.

9. Department of the Army. North Pacific Division. Corps of Engineers. Unpublished project data. Portland, Oregon. 1968.

10. Doll, John P., Emil H. Jebe and Robert D. Munson. Computation of variance estimates for marginal physical products and marginal rates of substitution. Journal of Farm Economics 42: 596-608. 1960.

11. Draper, N. R. and H. Smith. Applied regression analysis. New York, John Wiley and Son, Inc., 1966. 407 p.

12. Edwards, Clark. Demand elasticity in the factor market as implied by Cobb-Douglas production functions. Journal of Farm Economics 43: 142-143. 1961.

13. _______________ Resource fixity and farm organization. Journal of Farm Economics 41: 747-760. 1959.

14. Farrar, Donald E. and Robert R. Glauber. Multicollinearity in regression analysis: the problem revisited. Review of Economics and Statistics 49: 92-108. 1967.

15. Ferber, Robert and P. J. Verdoorn. Research methods in economics and business. New York, The Macmillan Co., 1962. 573 p.

16. Gardner, Delworth B. and Herbert H. Fullerton. Transfer restrictions and misallocations of irrigation water. American Journal of Agricultural Economics 50: 556-572. 1968.

17. Gisser, Micha. On benefit-cost analysis of investment in schooling in rural farm areas. American Journal of Agricultural Economics 50: 621-630. 1968.

18. Green, John H. A. Aggregation in economic analysis, an introductory survey. Princeton, Princeton University Press, 1964. 129 p.

19. Griliches, Zvi. Measuring inputs in agriculture. Journal of Farm Economics 42: 1411-1427. 1960.

20. _______________ Research expenditures, education, and the aggregate agricultural production function. American Economic Review 54: 961-975. 1964.

21. _______________ Sources of measured productivity growth; U.S. agriculture 1940-1960. Journal of Political Economy 71: 331-347. 1963.

22. _______________ Specification and estimation of agricultural production functions. Journal of Farm Economics 45: 419-429. 1963.

23. Griliches, Zvi. Specification bias in the estimation of production functions. Journal of Farm Economics 39: 8-20. 1957.

24. ________________ The demand for inputs in agriculture and a derived supply elasticity. Journal of Farm Economics 41: 309-323. 1959.

25. Griliches, Zvi and Yehuda Grunfeld. Is aggregation necessarily bad? Review of Economics and Statistics 42: 1-14. 1960.

26. Halter, A. N. Croplands. In: The atlas of the Pacific Northwest, ed. by Richard M. Highsmith Jr., Corvallis, Oregon, Oregon State University Press, 1968. p. 67-71.

27. Hardin, E. S., E. O. Heady, and G. L. Johnson (eds.). Resource productivity, returns to scale and farm size. Ames, Iowa, Iowa State College Press, 1956. 208 p.

28. Headley, J. C. Estimating the productivity of agricultural pesticides. American Journal of Agricultural Economics 50: 13-23. 1968.

29. Heady, Earl O. and J. L. Dillon. Agricultural production functions. Ames, Iowa, Iowa State University Press, 1961. 667 p.

30. Heady, Earl O. and J. Pesek. A fertilizer production surface with specification of economic optima for corn grown on Calcareous Ida Silt Loam. Journal of Farm Economics 36: 466-482. 1954.

31. Johnston, J. Econometric methods. New York, McGraw-Hill Book Co., 1963. 300 p.

32. Keniston, Robert F. Forest resources. In: The atlas of the Pacific Northwest, ed. by Richard M. Highsmith Jr., Corvallis, Oregon, Oregon State University Press, 1968. p. 53-61.

33. Knight, D. A. Farm size and resource efficiency on eastern Kansas farms. Manhattan, Kansas, 1967. 43 p. (Kansas Agricultural Experiment Station, Technical Bulletin 155)

34. Knox, Ellis G. Soils. In: The atlas of the Pacific Northwest, ed. by Richard M. Highsmith Jr., Corvallis, Oregon, Oregon State University Press, 1968. p. 43-47.

35. Miller, Stanley F. and Larry L. Boersma. Economic analysis of water, nitrogen, and seeding rate relationships in corn production on Woodburn soils. Corvallis, 1966. 40 p. (Oregon Agricultural Experiment Station. Technical Bulletin 98)

36. Mundlak, Yair. Specification and estimation of multiproduct production functions. Journal of Farm Economics 45: 433-443. 1963.

37. Nerlove, Marc. Estimation and identification of Cobb-Douglas production functions. Chicago, Rand McNally and Co., 1965. 193 p.

38. Nielsen, Darwin B. Public policy and grazing fees on federal lands: some unresolved issues. Land and Water Law Review 5 (Number 2): 293-321. 1970.

39. Pacific Northwest River Basins Commission. Columbia-North Pacific region comprehensive framework study of water and related lands. Unpublished research on crop yields by county. Vancouver, Washington. 1969.

40. Reder, M. W. An alternative interpretation of the Cobb-Douglas production function. Econometrica 11: 259-264. 1943.

41. Reinsel, Edward I. Labor movements between farm and non-farm jobs: comment. American Journal of Agricultural Economics 50: 745-747. 1968.

42. Ruttan, V. W. The economic demand for irrigated acreage, Baltimore, John Hopkins Press, 1965. 139 p.

43. Ruttan, V. W. and Thomas T. Stout. Regional differences in factor shares in American agriculture. Journal of Farm Economics 42: 52-69. 1960.

44. Solow, R. M. The production function and the theory of capital. Review of Economic Studies 23: 101-109. 1955-56.

45. Theil, H. Linear aggregation of economic relations. Amsterdam, North-Holland Publishing Co., 1954. 205 p.

46. _______________ On the use of prior information in regression analysis. American Statistical Association Journal 58: 401-414. 1963.

47. U. S. Department of Agriculture. Agricultural Marketing Service. Agricultural Research Service. U. S. Department of Commerce. Bureau of the Census. Farmers' expenditures for motor vehicles and machinery with related data, 1955. Washington, D. C., March, 1959. 97 p. (U. S. Dept. of Agriculture Statistical Bulletin No. 243)

48. U. S. Department of Agriculture. Agricultural Stabilization and Conservation Service. Annual statistical report: Idaho. 1945-1964. (Boise), 1946-1965.

49. ______________ Annual statistical report: Oregon. 1945-1964. (Portland), 1946-1965.

50. ______________ Annual statistical report: Washington. 1945-1964. (Spokane), 1946-1965.

51. U. S. Department of Agriculture. Economic Research Service. Farm costs and returns. Washington, D. C., 1967. 101 p. (Agricultural Information Bulletin No. 230)

52. ______________ Farmers' expenditures for custom pesticide service in 1964. Washington, D. C., October 1968. 24 p. (Agricultural Economic Report No. 146)

53. ______________ Farm Income: state estimates 1949-1964. Washington, D. C., Aug., 1965. 135 p. (A supplement to the July 1965 Farm Income Situation)

54. ______________ The hired farm working force in 1967: a statistical report. Washington, D. C., Sept., 1968. 31 p. (Agricultural Economic Report No. 148)

55. U. S. Department of Agriculture. Soil Conservation Service. Unpublished data on life span of ACP practices. Washington, D. C., April, 1957.

56. U. S. Department of Agriculture. Statistical Reporting Service. Unpublished worksheets by reporting districts in Oregon (base data for ERS series, Farm Real Estate Market Developments). Oregon, 1965.

57. ______________ Crop Reporting Board. Farm labor. Washington, D. C., Jan., 1965.

58. U. S. Department of Commerce. Bureau of the Census. United States census of agriculture, 1964 (Idaho). Vol. 1, part 39, Idaho. Washington, D. C., 1967. 331 p.

59. ______________ United States census of agriculture, 1964 (Oregon). Vol. 1, part 47, Oregon. Washington, D. C., 1967. 367 p.

60. ______________ United States census of agriculture, 1964 (Washington). Vol. 1, part 46, Washington. Washington, D. C., 1967. 389 p.

61. ______________ United States census of agriculture, 1959. The United States irrigation of agricultural lands. Vol. III. Washington, D. C., 1961. 400 p.

62. U. S. Department of the Interior. Bureau of Reclamation. Reclamation project data. Washington, D. C., U. S. Government Printing Office, 1961. 890 p.

63. ______________ Report of the Commissioner of the Bureau of Reclamation to the Secretary of the Interior for the year ended June 30, 1965. Statistical Appendix. Parts I, II, and III. Washington, D. C., 1965. 199 p.

64. U. S. Water Resources Council. The Nation's water resources. Washington, D. C., U. S. Government Printing Office, 1968. 7-3-13 p.

65. Yotopoulos, Pan A. From stock to flow capital inputs for agricultural production functions; a micro-analytic approach, Journal of Farm Economics 49: 476-491. 1967.

APPENDICES

APPENDIX I

DERIVATION AND PROOF OF NECESSARY AND SUFFICIENT CONDITIONS FOR CONSISTENT AGGREGATION

Chapter II aludes to the aggregation problem in the context of this study. This Appendix specifies the requirements for consistent aggregation of production functions. The discussion is first concerned with aggregation problems of exact models and then with aggregation problems associated with stochastic models.

Following Green (18, p. 99-107) we first define the necessary and sufficient conditions for consistent aggregation of production functions. Consider the firm production function

$$y_s = f_s(x_{1s}, \ldots, x_{rs}, \ldots, x_{ms}) \qquad (2a.1)$$

where; y_s = output for the s^{th} firm $(s = 1, \ldots, n)$

x_{rs} = the rth input $(r = 1, \ldots, m)$.

The only restriction is that y_s and x_{rs} be non-negative and continuous. We wish to be able to write

$$Y = F(x_1, \ldots, x_m) \qquad (2a.2)$$

where Y and x_r are defined by the aggregate function:

$$Y = y(y_1, \ldots., y_s, \ldots., y_n) \qquad (2a.3)$$

$$x_r = x_r(x_{r1}, \ldots., x_{rs}, \ldots., x_{rn}) \qquad (2a.4)$$

The necessary conditions for the function (2a.1) to be aggregated to the function (2a.2) are that, for all $r = 1, \ldots., m$ and $s = 1, \ldots., n$

$$\frac{\partial F}{\partial x_r} \frac{\partial x_r}{\partial x_{rs}} = \frac{\partial Y}{\partial y_s} \frac{\partial f_s}{\partial x_{rs}} . \tag{2a.5}$$

To prove this relationship we write the total differential for the two expressions for Y ((2a.2) and (2a.3)). From (2a.2) we have

$$dY = \frac{\partial F}{\partial x_1} dx_1 + \frac{\partial F}{\partial x_2} dx_2 \ldots + \frac{\partial F}{\partial x_m} dx_m = \sum_{r=1}^{m} \frac{\partial F}{\partial x_r} dx_r .$$

Finding dx_r from (2a.4) we have

$$dY = \sum_{r=1}^{m} \frac{\partial F}{\partial x_r} dx_r = \sum_{r=1}^{m} \sum_{s=1}^{n} \frac{\partial F}{\partial x_r} \frac{\partial x_r}{\partial x_{rs}} dx_{rs} . \tag{2a.6}$$

Then from (2a.3) we have

$$dY = \frac{\partial Y}{\partial y_1} dy_1 + \ldots + \frac{\partial Y}{\partial y_n} dy_n = \sum_{s=1}^{n} \frac{\partial Y}{\partial y_s} dy_s$$

and finding dy_s from (2a.1)

$$dY = \sum_{s=1}^{n} \frac{\partial Y}{\partial y_s} dy_s = \sum_{s=1}^{n} \sum_{r=1}^{m} \frac{\partial Y}{\partial y_s} \frac{\partial f_s}{\partial x_{rs}} dx_{rs} . \tag{2a.7}$$

Thus, the coefficients of dx_{rs} found in the two expressions (2a.6) and (2a.7) must be equal for each pair of values of r and s. This is what (2a.5) states.

The necessary and sufficient conditions for aggregation of the function (2a.1) to the function (2a.2) are that there exists functions G, H, g_r, h_s, G_r, H_s, g_{rs} and h_{rs} such that

$$Y = H(h_1(y_1) + \ldots + h_n(y_n)) \tag{2a.8}$$

$$= G(g_1(x_1) + \ldots + g_m(x_m)) \tag{2a.9}$$

where

$$y_s = H_s(h_{1s}(x_{1s}) + \dots\dots + h_{ms}(x_{ms})) \quad (s=1,\dots\dots,n) \quad (2a.10)$$

and

$$x_r = G_r(g_{r1}(x_{r1}) + \dots\dots + g_{rn}(x_{rn})) \quad (r=1,\dots,m). \quad (2a.11)$$

It is obvious that the existence of these functions provides a sufficient condition for (2a.5) since the functions are additive with no interaction terms. Equation (2a.5) holds for each pair of values of r and s.

It is more difficult to show that the equations (2a.8) through (2a.11) are necessary conditions for (2a.5). This will be shown in several parts:

(1) If we solve (2a.5) for $\dfrac{\partial Y}{\partial y_s}$

and take the ratio $\dfrac{\partial Y/\partial y_s}{\partial Y/\partial y_t}$

(where t is any of the firms $s=1,\dots,n$); i.e., take the ratio of the partial derivatives from (2a.3) for any two firms s and t. We obtain

$$\frac{\partial Y/\partial y_s}{\partial Y/\partial y_t} = \frac{\partial x_1/\partial x_{1s}}{\partial f_s/\partial x_{1s}} \div \frac{\partial x_1/\partial x_{1t}}{\partial f_t/\partial x_{1t}} = \frac{\partial x_2/\partial x_{2s}}{\partial f_s/\partial x_{2s}} \div \frac{\partial x_2/\partial x_{2t}}{\partial f_t/\partial x_{2t}}$$

$$= ,\dots\dots, = \frac{\partial x_m/\partial x_{ms}}{\partial f_s/\partial x_{ms}} \div \frac{\partial x_m/\partial x_{mt}}{\partial f_t/\partial x_{mt}} .$$

The ratio of the marginal contribution to aggregate output of any two firms s and t must be equal to the quotient of two ratios. The numerator of the quotient is the ratio of the marginal contribution of the s^{th}

firm input to the aggregate input, to the firm level marginal product of the same input. The denominator is the same except it is for the firm t. This quotient must then be equal for all inputs $(r=1,\ldots,m)$ for the two firms s and t. The partial derivatives in this expression (excluding the left-hand side) depend only on the values $x_{1s},\ldots,x_{ms}$ and $x_{1t},\ldots,x_{mt}$ as can be seen by examining equations (2a.1) and (2a.4). These variables determine the values of y_s and y_t but do not influence the value of any other firm's output. Thus $\dfrac{\partial y/\partial y_s}{\partial y/\partial y_t}$ is a function only of y_s and y_t. This is true for all s and t. With these relationships we can obtain the functions H, $h_1,\ldots,h_n$. It must be that $Y = y(y_1,\ldots,y_n) = H(h_1(y_1) + ,\ldots, + h_n(y_n))$.

(2) If we solve (2a.5) for $\dfrac{\partial F}{\partial x_r}$ and $\dfrac{\partial F}{\partial x_q}$ (where q is any of the aggregate inputs $r=1,\ldots,m$) and take the ratio of the two, we have

$$\frac{\partial F/\partial x_q}{\partial F/\partial x_r} = \frac{\partial f_1/\partial x_{q1}}{\partial x_q/\partial x_{q1}} \div \frac{\partial f_1/\partial x_{r1}}{\partial x_r/\partial x_{r1}} = \frac{\partial f_2/\partial x_{q2}}{\partial x_q/\partial x_{q2}} \div \frac{\partial f_2/\partial x_{r2}}{\partial x_r/\partial x_{r2}} =$$

$$,\ldots, = \frac{\partial f_n/\partial x_{qn}}{\partial x_q/\partial x_{qn}} \div \frac{\partial f_n/\partial_{rn}}{\partial x_r/\partial x_{rn}} .$$

This is true for all pairs of aggregate inputs $r=1,\ldots,m$. These partial derivatives (except the left-hand side) depend only on the values $x_{q1},\ldots,x_{qn}$ which determine x_q and $x_{r1},\ldots,x_{rn}$ which determine x_r.

Therefore, the left-hand side $(\partial F/\partial x_q \div \partial F/\partial x_r)$ depends only on x_q and x_r. This is true for all q and r so that we may write $Y = F(x_1, \ldots, x_m) = G(g_1(x_1) + \ldots + g_m(x_m))$, which shows the existence of the functions G, $g_1, \ldots, g_m$.

(3) Solving (2a.5) for $\partial f_s/\partial x_{qs}$ and $\partial f_s/\partial x_{rs}$ and taking the ratio of the two, we have

$$\frac{\partial f_s/\partial x_{qs}}{\partial f_s/\partial x_{rs}} = \frac{\partial F/\partial x_q}{\partial F/\partial x_r} \; \frac{\partial x_q/\partial x_{qs}}{\partial x_r/\partial x_{rs}} \; .$$

The left-hand side depends only on $x_{1s}, \ldots, x_{ms}$. $\partial F/\partial x_q \div \partial F/\partial x_r$ was shown above to depend only on $x_{q1}, \ldots, x_{qn}$ and $x_{r1}, \ldots, x_{rn}$. The second part of the right-hand side depends on the same variables; thus, the only variables on which both sides depend are x_{qs} and x_{rs}. So, the left-hand side is a function of only x_{qs} and x_{rs}. We can now write

$$y_s = f_s(x_{1s}, \ldots, x_{ms}) = H_s(h_{1s}(x_{1s}) + \ldots + h_{ms}(x_{ms}))$$

showing the existence of the functions H_s, and $h_{1s}, \ldots, h_{ms}$.

(4) Solving (2a.5) for $\partial X_r/\partial x_{rs}$ and $\partial x_r/\partial x_{rt}$ and taking the ratio of the two, we can show by the same argument as above that this ratio depends only on x_{rs} and x_{rt} and we may write

$$x_r = x_r(x_{r1}, \ldots, x_{rn}) = G_r(g_{r1}(x_{r1}) + \ldots + g_{rn}(x_{rn})).$$

This completes the statement and the proof of the necessary and sufficient conditions for consistent aggregation. We will now consider some interpretations and applications of these results.

We first consider the case where the aggregates are the simple sums

$$Y = \sum_{s=1}^{n} y_s \quad \text{and} \quad x_r = \sum_{s=1}^{n} x_{rs} \quad (r=1,\dots,m).$$

From the necessary condition (2a.5) the two partials $\partial x_r/\partial x_{rs}$ and $\partial Y/\partial y_s$ both equal unity and the equation becomes

$$\partial F/\partial x_r = \partial f_s/\partial x_{rs}$$

for all r and s. For any input r, $\partial F/\partial x_r$ depends only on the totals $x_1,\dots,x_m$ which are invariant with any finite number of distributions of inputs among the n firms. The firm level marginal products $\partial f_s/\partial x_{rs}$ must be equal for all firms s. The marginal products must also be constant for all values of x_{rs}. This implies that the firm production must be linear with identical slopes

$$y_s = a_s + \sum_{r=1}^{m} b_r x_{rs} .$$

This condition is also seen to be sufficient since

$$Y = \sum_{s=1}^{n} y_s = \sum_{s=1}^{n} a_s + \sum_{s=1}^{n} \sum_{r=1}^{m} b_r x_{rs}$$

$$= \sum_{s=1}^{n} a_s + \sum_{r=1}^{m} b_r \left(\sum_{s=1}^{n} x_{rs} \right) = a + \sum_{r=1}^{n} b_r x_r .$$

The restriction of identical slopes may be relaxed if we use weighted sums of the inputs

$$x_r = \sum_{s=1}^{n} w_{rs}\, x_{rs} ,$$

Then from (2a.5) we have

$$\partial F/\partial x_r = 1/w_{rs}\ \partial f_s/\partial x_{rs}$$

for each r and s. The firm functions must still be linear but they may have different slopes. The weights w_{rs} must be such that $\frac{b_{rs}}{w_{rs}}$ is the same for each firm s. Thus if we select w_{rs} to be the marginal products of each firm input, aggregation will be consistent.

A weighting procedure may also be applied to the dependent variable to obtain consistent aggregation. If we allow $Y = \sum_{s=1}^{n} w_s\, y_s$ and require $x_r = \sum_{s=1}^{n} x_{rs}$ as before then (2a.5) becomes

$$\partial F/\partial x_r = w_s\ \ \partial F_s/\partial x_{rs}$$

for each r and s. The firm functions must again be linear, and if

$$y_s = a_s + \sum_{r=1}^{m} b_{rs}\, x_{rs}$$

for all s, then we must have

$$\frac{b_{rs}}{b_{rt}} = \frac{c_t}{c_s} \quad \text{or} \quad c_s b_{rs} = c_t b_{rt} = k_r$$

for all pairs of firms s and t. In that case

$$y = \sum_{s=1}^{n} c_s y_s = \sum_{s=1}^{n} c_s a_s + \sum_{r=1}^{m} \sum_{s=1}^{n} c_s b_{rs} x_{rs}.$$

An obvious extension of the above results to non-linear functions is the case of the exponential function. If the functions are written in log form the three above results can be applied directly. The aggregates in this case are products instead of sums. The difficulty with this formulation is that the aggregation is consistent only in the log form which presents a problem of economic interpretation.

A further extension of consistent aggregation may be made if we include economics and take a special case of the exponential function. Consider n firms with production functions which are homogeneous of degree one.[18] If the optimal conditions for pure competition exist[19] and firms have production functions, which are homogeneous of the first degree, the expansion paths are all straight lines out of the origin with the same slope and aggregation will be consistent. This is true for the aggregation procedure

$$\sum_{s=1}^{n} x_{rs} = x_r \text{ and } \sum_{s=1}^{n} c_s Y_s$$

---

[18] Functions are homogeneous of the first degree if for the function $Y = f(x_1, x_2, \ldots x_n)$, $tY = f(tx_1, tx_2, \ldots tx_n) = t^m f(x_1, x_2, \ldots x_n)$ where t is an arbitrary constant and m indicates the degree of homogeneity--in this case m=1).

[19] The optimal conditions are that the marginal rate of substitution between any two inputs must be the same for any two firms.

on the previous page where the x's are allowed to take on any positive number.

Since profit maximization requires the firm to be on the expansion path, the production function may be written in terms of the constant input ratios and one input alone. Consider the Cobb-Douglas function

$$Y_s = a_s \prod_{r=1}^{m} x_{rs}^{b_{rs}}, \quad \sum_{r=1}^{m} b_{rs} = 1.$$

The optimal conditions and the pure competition assumptions require, in general that the marginal rates of substitution are equal for all pairs of inputs for all firms. The nature of the Cobb-Douglas functions assures that the expansion paths will be straight lines and the assumption that expansion paths all come from the origin and have the same slopes implies that $b_{r1}=b_{r2}=b_r$ for all r; i.e., the firms all have functions with the same exponents for x_r. This is true since

$$MRS_{rq,\ t} = MRS_{rq,\ s}$$

which implies that

$$\frac{MPx_{rt}}{MPx_{qt}} = \frac{b_{rt}x_{qt}}{b_{qt}x_{rt}} = \frac{MPx_{rs}}{MPx_{qs}} = \frac{b_{rs}\,x_{qs}}{b_{qs}\,x_{rs}},$$

and the ratios $\frac{x_{qt}}{x_{rt}}$ and $\frac{x_{qs}}{x_{rs}}$ are constants and equal, which implies that the only way the product of ratios can be equal for all s and t is the case where

$$b_{rs} = b_{rt} = b_r$$

for all r, s and t.

By substituting the general for the specific exponents we may write

$$Y_s = a_s \, x_{1s}^{b_1} \; . \; . \; . \; x_{ms}^{b_m} \text{ and}$$

by substituting

$$x_{qs} = m_r \, x_{rs} \; (\text{or } \frac{x_{qs}}{x_{rs}} = m_r \text{ for all s})$$

we have

$$Y_s = a_s \; x_{1s}\left(\frac{x_{2s}}{x_{1s}}\right)^{b2} \; . \; . \; . \left(\frac{x_{ms}}{x_{1s}}\right)^{bm}$$

Or since $\frac{x_{rs}}{x_{1s}}$ is equal to $\frac{\sum_{s=1}^{n} x_{rs}}{\sum_{s=1}^{n} x_{1s}}$ we may write

$$Y_s = a_s \; x_{1s}\left(\frac{x_2}{x_1}\right)^{b_2} \; . \; . \; . \left(\frac{x_m}{x_1}\right)^{b_m} \; .$$

Aggregation yields,

$$Y = \sum_{s=1}^{n} \frac{Y_s}{a_s} = \sum_{s=1}^{n} x_{1s}\left(\frac{x_2}{x_1}\right)^{b_2} \; . \; . \; . \left(\frac{x_m}{x_1}\right)^{b_m}$$

$$= x_1^{b_1} \, x_2^{b_2} \; . \; . \; . \; x_m^{b_m} \text{, which is the aggregate function --}$$

also homogeneous of degree one (9, p. 49-51). The weights for aggregating Y_s are equal to unity over a_s.[20]

Green further shows (18, p. 43) that aggregation will be consistent for polynomials by using weighted moments of the first and second orders. This approach requires knowledge of the probability distributions of the variables to be aggregated.

The above discussion is concerned with exact models. The results, as shown by Green (18, p. 100-103) are slightly more complex for stochastic models. We assume the firm functions to be of the following form where the aggregation procedure is simple sums

$$Y_s(t) = A_s + \sum_{r=1}^{m} b_{rs}\, x_{rs}(t) + U_s(t)$$

where t is the time subscript and U_s the error term. If we also assume that the micro inputs x_{rs} are related to the macro inputs in a liner fashion then,

$$x_{rs}(t) = \alpha_{rs} + \sum_{q=1}^{m} \beta_{rs.q} \cdot x_q(t) + v_{rs}(t)$$

and

$$\sum_{s=1}^{n} \alpha_{rs} = \sum_{s=1}^{n} v_{rs}(t) = 0; \quad \sum_{s=1}^{n} \beta_{rs.q} = 1$$

[20] This presents a slightly different concept of the aggregate production function. It tends to measure the existing institutional effects on Y_s and not strictly a technical relationship -- it is generated by a group of economic decisions, assumed to be optimal.

In a special case of this procedure where the firm functions are all identical, the aggregate function will be consistent when the aggregation procedure is to use simple sums. The aggregate function is

$$Y = \sum_{s=1}^{n} Y_s = nax_1^{b_1} x_2^{b_2} \ldots x_m^{b_m}.$$

for $r \neq q$ and zero if $r=q$, since

$$\sum_{s=1}^{n} x_{rs} = x_r .$$

The parameters and the error term for the aggregate function,

$$Y(t) = A + \Sigma\, b_q\, x_q(t) + U\,(t)$$

may be expressed as

$$A = \sum_{s=1}^{n} a_s + n \sum_{r=1}^{m} \operatorname{cov}\,(b_{rs}, \alpha_{rs})$$

$$b_q = \frac{1}{n} \sum_{s=1}^{n} b_{rs} + n \sum_{r=1}^{m} \operatorname{cov}\,(b_{rs}, \beta_{rs\,.\,q}$$

$$U(t) = \sum_{s=1}^{n} U_s(t) + n \sum_{r=1}^{m} \operatorname{cov}\,(b_{rs}, U_{rs}(t).$$

Three possibilities now exist for consistent aggregation: (1) If the parameters b_{rs} are the same for all firms, the covariance terms are all zero and we have consistent aggregation as before. (2) If the x_{rs} are related to only x_r and in an exact linear function, aggregation will be consistent but b_r is not equal to the arithmetic mean of the b_{rs}. (3) If x_{rs} is related only to x_r, but stochastically with cov (b_{rs}, U_{rs}) = 0, then aggregation is again consistent.

In general, however, the b_{rs} will not be equal for all firms and the functional form of x_{rs} may not be so "well behaved". In this case, the extent of the aggregation bias depends upon the equation forms, the aggregation procedure, and the statistical estimation method used.

EXAMPLES OF CONSISTENT AGGREGATION

A. Linear firm functions with equal slopes.

Consider two firm functions,

$$y_1 = 2 + 4x_1$$

$$y_2 = 4 + 4x_2 .$$

We use simple sums for aggregates and the aggregate function is

$$Y = y_1 + y_2 = 2 + 4 + 4x_1 + 4x_2$$

$$= 6 + 4(x_1 + x_2)$$

$$= 6 + 4X \quad , \quad \text{where } X = \sum_{s=1}^{2} x_s .$$

The aggregate function consistently gives us the same value of Y from any value of X, regardless of the distribution of the input among the two firms; all we need know is the sums of the two firm inputs.

B. Linear firm functions with unequal slopes.

Consider two firm functions,

$$y_1 = 2 + 4x_1$$

$$y_2 = 4 + 2x_2 .$$

We may use a weighted sum for the input, where the weights equal the marginal products. Thus, the aggregate function is

$$Y = 6 + (4x_1 + 2x_2)$$

$$= 6 + X \qquad \text{where } X = (4x_1 + 2x_2)$$

We must know the slope coefficients, 4 and 2. Given this information we will consistently get the same aggregate output whether we use the two firm functions (and sum the results) or use only the aggregate function.

Alternatively we may use the weighting procedure for y_i $(i = 1, 2)$. We have

$$c_s\, b_{rs} = c_t\, b_{rt} = k_r$$

or in this example,

$$(1/4)4 = 1/2(2) = 1 \qquad \text{(any value of } k_r \text{ will do).}$$

Then the aggregate function is,

$$Y = 1/4 y_1 + 1/2 y_2 = (1/4)(2) + 1/2(4) + (x_1 + x_2)$$
$$= 2(1/2) + X, \text{ where } X = \sum_{s=1}^{2} x_s \,.$$

C. Cobb-Douglas functions with inputs in fixed proportions and constant returns to scale.

Consider the two functions

$$y_1 = 2x_{11}^{.5} x_{12}^{.5}$$

$$y_2 = 4x_{21}^{.5} x_{22}^{.5} \,.$$

Given fixed proportions, $x_{12} = 2x_{11}$ and $x_{21} = 2x_{22}$ the function may be written

$$y_1 = 2x_{11} \left(\frac{x_{12}}{x_{11}}\right)^{.5} = 2\sqrt{2}x_{11}$$

$$y_2 = 4x_{21}\left(\frac{x_{22}}{x_{21}}\right)^{.5} = 4\sqrt{2}\,x_{21} .$$

The aggregate function may be written

$$Y = \frac{1}{2\sqrt{2}} \bullet y_1 + \frac{1}{4\sqrt{2}} \bullet y_2 = X \quad , \text{ where } X = (x_{11} + x_{21}).$$

In the special case where the constant term is equal to unity in each firm function the aggregate function is

$$Y = 1/2y_1 + 1/2y_2 \quad = \quad X \quad , \text{ where } X = x_{11} + x_{21}$$

or, since $y_1 + y_2 = x_{11}^{.5}x_{12}^{.5} + x_{21}^{.5}x_{22}^{.5}$

$$= 2x_1^{.5}x_2^{.5}$$

and we let $x_{12} = 2x_{11}$ and $x_{21} = 2x_{22}$ as before, then

$$y_1 + y_2 = 2x_{11} + 2x_{21} = 2(x_{11} + x_{21})$$

$$= 2X$$

So if we define the aggregate output as the simple sums of the firm outputs, then aggregation is consistent if we also use simple sums for the inputs. This also holds when firms have identical functions, e.g.

$$y_1 = 2x_{11}^{.5}x_{12}^{.5}$$

$$y_2 = 2x_{21}^{.5}x_{22}^{.5} .$$

The aggregate function

$$Y = y_1 + y_2 = 4X_1^{.5}X_2^{.5} \quad , \text{ where } X_1 = x_{11} + x_{21}$$

is consistent.

FLOW vs. STOCK CONCEPTS OF CAPITAL INPUTS

The relationship between stocks and flows over time depends on the shape of the service flow stream and the salvage value at the end of the useful life of the asset. First consider the simple case where the service flow is constant over the finite life of the asset and there is no salvage value. In the continuous form, the present value of the service flow (stock value) is [21]

$$V_0^T = \bar{R}\int_0^T e^{-rt}dt = \frac{\bar{R}}{r}\,\frac{e^{rT}-1}{e^{rT}} \tag{2a.12}$$

where:

V_0^T = present value of a new asset (indicated by subscript zero) with a useful life of T years

$\bar{R}$ = constant service flow

r = discount rate.

Thus, the proportionality of stock to flow becomes

$$\frac{V_0^T}{\bar{R}} = \frac{1}{r}\,\frac{e^{rT}-1}{e^{rT}}\ . \tag{2a.13}$$

After k years of depreciation ($1 \leq k \leq T$) the present value is

[21] The usual form of the discrete case of present value of a service flow is $V_0^T = \frac{R_1}{(1+r)} + \ldots + \frac{R_T}{(1+r)^T}$. When $R_1 = R_2 = \ldots. = R_T = \bar{R}$, the equation becomes $V_0^T = \bar{R}\sum_{t=1}^{T}\left(\frac{1}{(1+r)^t}\right)$. With continuous compounding $(1+r)^t$ becomes e^{rt}, so that the continuous form is $\bar{R}\int_0^T e^{-rt}dt$.

$$V_0^T = \bar{R}\int_1^T e^{-rt}dt = \frac{\bar{R}}{r}\,\frac{e^{r(T-k)}-1}{e^{rT}} \qquad (2a.14)$$

and $\dfrac{V_1^T}{\bar{R}} = \dfrac{1}{r}\,\dfrac{e^{r(T-k)}-1}{e^{rT}}$ which is less than $\dfrac{V_0^T}{\bar{R}}$ for positive

T and r. As a result the proportionality decreases as the asset ages. And the use of stocks instead of flows places more weight on the more durable asset. For example, consider two assets with identical service flows equal to \$100, no salvage value, and equal original, useful lives of 10 years. Let one asset be new and the other five years old, then

$$V_{0,1}^{10} = \bar{R}\int_0^{10} e^{-rt}dt = \frac{100}{r}\,\frac{e^{10r}-1}{e^{10r}}$$

and $$V_{5,2}^{10} = \frac{100}{r}\,\frac{e^{r(10-5)}-1}{e^{10r}} = \frac{100}{r}\,\frac{e^{5r}-1}{e^{10r}}$$

where the subscripts 1 and 2 indicate the new and the five-year old assets, respectively. With a discount rate equal to .10,

$$V_{0,1}^{10} = \frac{100}{.10}\,\frac{e-1}{e} = 1000(1-\frac{1}{e}) = \$632$$

and $$V_{5,2}^{10} = 1000\left(\frac{e^{.5}-1}{e}\right) = \$239\ .$$

The use of the stock value would place an unwarranted weight of 2.6 on the more durable asset.

If the service flow deteriorates with the age of the asset then the relationship between the value of the capital stock and the service flow is slightly changed. For example, if the rate of deterioration, s, is a constant over time such that $R(t) = R\, e^{-st}$, then the relationship becomes,

$$W_0^T = R\int_0^T e^{-(r+s)t}dt = \frac{R}{r+s}\ \frac{e^{(r+s)T}-1}{e^{(r+s)T}} \tag{2a.15}$$

where W_0^T = present value of the asset

s = a constant deterioration rate

R = the service flow in $t=0$.

After one year the value of the capital stock becomes,

$$W_1^T = \frac{R}{r+s}\ \frac{e^{(r+s)(T-1)}-1}{e^{(r+s)T}} \tag{2a.16}$$

and $\frac{W_0^T}{R} > \frac{W_0^T}{R}$ (the ratio of stock to flow decreases with age).

The result is the same as above; i.e., use of the stocks instead of flows places more weight on more durable assets.

A varying weight from stocks to flows may result (depending on the form of the two functions) in the case where assets (e.g. livestock) first appreciate with age and then depreciate. Only in the case of an asset with an infinite life span (e.g. land) will the proportionally remain constant. In this case

$$V_0^T = \lim_{T\to\infty}\left[\frac{\bar{R}}{r}\ \frac{e^{rT}-1}{e^{rT}}\right] = \frac{\bar{R}}{r}$$

and the proportionality factor is $\frac{\bar{R}}{r}\left(\frac{1}{\bar{R}}\right) = \frac{1}{r}$.

Having established the appropriate measure of the production input from capital assets, a question now remains as to how to measure the flow. If it can be assumed that the service flow is constant over time, with no salvage value, and the useful life, original market value of the asset, and r are known we may simply solve (2a. 12) for $\overline{R}$ to obtain the annual service flow

$$\overline{R} = \frac{rV_0^T}{1-e^{-rT}} .$$

Or, if the service flow is not constant over time and the function R(t) is known, (along with useful life, original market value and r), we may solve an equation similar to (2a. 15) for the year of interest, t. But R(t) may be difficult to establish. In this case there is an alternative available using only current market values. It may first appear that this could be done simply by calculating the relevant flow as $V_{t+1} - V_t$ (the change in the market value of the asset during the production period, t). But this calculation involves factors other than the depreciation (change in current flow of productive services) that we wish to measure. The current market devaluation of a used asset is likely to reflect not only a change in depreciation but also an obsolescence factor. The obsolescence factor is the penalty attached to old assets because of the probability of better assets becoming available. This factor can be removed simply, as Yotopoulas demonstrates.

If we have survey data available showing the market value of the asset at the beginning and ending of the production period, we can find the service flow as

$$R_t = rV_t = (V_{t+1} - V_t) \; . \qquad (2a.17)$$

This is the discrete form which would normally be required since market values will be available only for discrete points on the market value function. The proof of (2a.17) is as follows:

Using the discrete form of the present value formula, the value, V from t=1 through T will be, (where $d = \frac{1}{1+r}$)

$$V_1 = R_1 d + R_2 d^2 + R_3 d^3 + \ldots + R_T d^T$$

$$V_2 = 0 + R_2 d + R_3 d^2 + \ldots + R_T d^{T-1}$$

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

$$V_T = 0 + 0 + 0 + \ldots + R_T d^1 \; .$$

Or in general,

$$V_t = R_t d + R_{t+1} d^2 + \ldots + R_T d^{T-(t-1)}$$

or

$$V_t = R_t d + d(R_{t+1} d + \ldots + R_T d^{(T-1)})$$

and solving for R_t ,

$$R_t = \frac{V_t}{d} - V_{t+1} = rV_t - (V_{t+1} - V_t) \; .$$

The service flow in any time period, t, will be equal to the discount factor times the market value in time period, t, adjusted by the difference in market values between (t+1) and t.

All the data needed then are the market values for two years and the discount rate r.

In summary, it has been shown that stocks can be used as a proxy for the flows of capital services only in the case of an infinite life expectancy of the asset. Further, it has been shown that the service flow can be calculated from present market values or from original market values when the service flow function is known.

HOMOGENEOUS FARMING AREAS

Table I is a percentage disaggregation of the total value of farm output into broad census categories of types of farm output. Area A contains counties with greater than 50% of total value of farm products sold (TVFPS) from field crops (FC) plus livestock and livestock products (LLP), where the % from FC and from LLP is greater than 20%. Area B contains counties with at least 50% of TVFPS from LLP and less than 20% from any other single source. Area C contains at least 50% of TVFPS from FC and less than 20% from any other single classification. Area D contains counties with at least 50% of TVFPS from all livestock and livestock products (ALLP) and not less than 10% from LLP. Area E contains the remaining counties which exhibit a diversity of TVFPS between the seven classifications. The percentages are based on data from the Agricultural Census for 1964 (57, 58, 59, Table 6, lines 63, 67, 69, 71, 73, 77, 79, 81, 83).

APPENDIX TABLE I. Value of major crop and livestock classifications as a percent of total value of farm products sold, Pacific Northwest, 1964.

Area, State and County	All Crops: Field Crops	Vege- tables	Fruits & Nuts	Forest Prod.	All Livestock and Liv. Prod.: Poultry P. Prod.	Dairy Prod.	Livestock Liv. Prod.
	Percent of Values of Farm Products Sold						
Area A							
Oregon							
Benton	39.3	10.8	.5	5.5	2.8	12.2	24.3
Crook	30.7	--	--	.1	.6	2.7	66.0
Gilliam	69.1	--	--	.1	.1	.1	30.6
Jefferson	67.2	.2	--	--	.1	.7	31.7
Klamath	44.4	--	--	.2	.9	2.9	51.5
Malheur	53.5	6.7	.3	.8	.5	7.9	30.6
Morrow	64.3	.3	.1	.5	.9	2.0	32.0
Umatilla	42.1	12.1	2.8	1.5	1.5	1.9	37.9
Union	40.2	5.1	5.3	1.2	.4	2.6	45.1
Wallowa	30.1	--	--	1.4	.2	3.5	64.0
Wasco	38.5	.8	28.3	.8	1.2	1.1	29.1
Washington							
Grant	57.3	1.4	.6	.5	.2	1.9	38.1
Klickitat	36.2	5.0	8.8	1.3	.8	6.6	40.9
Yakima	28.0	4.2	35.8	1.0	1.6	3.0	26.5
Idaho							
Bannock	54.8	--	.1	.6	2.9	8.9	32.7
Bear Lake	30.9	--	.1	--	.6	19.7	48.3
Boundary	61.1	--	--	2.7	.7	14.1	21.4
Butte	60.2	--	--	.1	11.0	2.9	35.8
Camas	66.8	--	--	.2	.1	3.8	29.0
Canyon	44.3	2.6	3.1	.9	1.9	10.8	36.4
Caribou	62.2	.5	--	--	.2	8.3	28.7
Cassia	56.3	.1	--	--	.6	4.5	38.4
Clark	21.2	--	--	--	.1	.7	78.1
Clearwater	57.6	--	.2	7.3	1.4	3.5	30.0
Custer	22.6	--	--	--	.1	2.2	74.4
Elmore	48.3	.6	--	.1	.3	1.9	48.8

APPENDIX TABLE I. (CON'T.)

Area, State and County	All Crops: Field Crops	All Crops: Vege-tables	All Crops: Fruits & Nuts	All Crops: Forest Prod.	All Livestock & Liv. Prod.: Poultry P. Prod.	All Livestock & Liv. Prod.: Dairy Prod.	All Livestock & Liv. Prod.: Livestock Liv. Prod.
	Percent of Value of Farm Products Sold						
Fremont	57.1	--	--	--	.1	4.5	38.2
Gooding	22.7	.4	.2	.3	5.2	15.2	56.2
Idaho	58.1	.1	.1	1.2	1.0	3.5	35.5
Jefferson	65.1	--	--	--	.4	8.2	26.0
Jerome	51.1	.2	.1	.1	.4	7.6	40.5
Kootenai	48.5	.7	.3	4.7	5.3	17.1	22.7
Lincoln	36.4	--	--	.1	3.7	18.8	41.1
Minidoka	68.7	--	--	.2	.8	5.3	24.9
Oneida	62.1	--	--	--	.1	5.8	32.0
Owyhee	48.2	.7	.9	.6	.2	7.5	41.9
Payette	22.1	6.3	12.3	1.3	1.4	15.3	41.3
Teton	49.8	.2	--	--	.3	16.0	33.3
Twin Falls	54.5	1.3	.4	.1	.8	8.3	34.6
Valley	23.3	--	--	.8	.7	2.7	70.4
Washington	29.6	8.1	1.5	.3	.6	10.7	49.2
Area B							
Oregon							
Baker	15.5	*	.1	.1	.5	5.5	78.4
Douglas	7.5	3.3	10.9	8.8	6.8	10.4	52.2
Grant	3.2	--	1.1	2.3	2.0	.7	90.5
Harney	8.0	--	--	.1	.2	.1	91.6
Lake	17.5	--	--	.1	.2	2.2	80.0
Wheeler	12.8	--	--	1.7	.3	.6	84.4
Washington							
Asotin	26.3	--	1.8	1.3	.3	4.0	65.9
Ferry	14.2	--	.9	7.1	.4	1.7	75.7
Kittitas	17.6	2.0	.2	.1	.6	5.4	73.6
Pend Oreille	19.3	--	--	10.4	.7	16.6	52.8
San Juan	9.7	11.8	.2	2.2	9.4	7.2	59.6

APPENDIX TABLE I. (CON'T.)

Area, State and County	All Crops: Field Crops	Vege-tables	Fruits & Nuts	Forest Prod.	All Livestock & Liv. Prod.: Poultry P. Prod.	Dairy Prod.	Livestock Liv. Prod.
	Percent of Value of Farm Products Sold						
Idaho							
Adams	8.5	--	.1	1.8	.6	3.2	85.8
Blaine	17.4	--	--	.1	1.1	13.6	67.7
Boise	12.8	--	--	.2	.4	.6	85.9
Lemhi	15.0	--	.2	.1	.8	4.5	78.6
Area C							
Oregon							
Linn	51.0	9.5	5.2	2.6	7.0	9.6	15.2
Sherman	83.0	--	.1	--	.1	--	16.6
Washington							
Adams	73.7	1.3	.3	--	.1	.4	24.3
Benton	53.6	4.2	13.8	.3	2.1	4.0	22.0
Columbia	64.6	20.8	1.8	.8	.1	.1	11.7
Douglas	52.8	--	36.1	.2	.3	.1	10.3
Franklin	75.1	1.3	.6	.1	.2	1.9	20.8
Garfield	83.0	--	.3	--	.1	.1	16.5
Lincoln	84.0	--	--	--	.5	.7	14.7
Spokane	57.9	1.2	1.1	5.1	11.0	11.1	12.4
Walla Walla	68.3	12.4	.4	2.1	2.2	2.9	11.6
Whitman	81.0	.6	.4	.1	.5	.7	16.7
Idaho							
Benewah	81.3	--	--	1.7	.5	1.2	15.2
Bingham	67.4	--	--	.1	.4	7.2	24.9
Bonneville	72.8	--	--	.5	.3	5.6	20.6
Latah	84.3	.2	.1	.6	.9	2.7	11.0
Lewis	91.7	.2	--	.5	.1	.4	7.1
Madison	76.6	--	--	.1	3.0	7.6	12.7
Nez Perce	70.4	5.2	.1	1.0	.9	3.3	19.0
Power	85.1	.3	--	--	.1	2.3	12.3

APPENDIX TABLE I. (CON'T.)

Area, State and County	All Crops: Field Crops	Vege-tables	Fruits & Nuts	Forest Prod.	All Livestock & Liv. Prod.: Poultry P. Prod.	Dairy Prod.	Livestock Liv. Prod.
	- - - -	Percent of Value of Farm Products Sold	- - - -				
Area D							
Oregon							
Clatsop	2.3	.2	.6	4.5	6.7	25.4	60.3
Columbia	6.0	1.6	11.7	18.5	11.2	20.3	39.5
Coos	2.4	*	6.6	7.7	2.1	54.1	27.0
Curry	.2	--	.6	34.6	2.6	19.8	42.1
Deschutes	17.3	--	--	.2	4.6	23.7	53.5
Josephine	20.8	.2	1.9	8.7	8.7	38.4	20.8
Lincoln	2.0	.5	1.9	14.5	6.2	37.6	36.5
Tillamook	.2	--	.1	1.6	.5	82.8	14.8
Washington							
Clallam	11.5	1.9	1.8	6.8	3.0	51.0	23.9
Clark	7.9	5.3	8.8	3.2	20.8	37.5	16.3
Grays Harbor	3.4	3.4	4.2	3.3	5.2	65.0	15.4
Island	6.6	1.6	2.7	5.7	49.7	18.3	15.1
Jefferson	2.8	.2	.1	6.8	10.2	45.6	34.0
King	.4	7.9	3.4	17.9	17.1	38.5	14.7
Lewis	5.3	2.1	3.0	3.8	34.9	30.1	20.6
Mason	2.3	2.0	2.2	22.6	2.7	38.0	30.0
Pacific	1.4	--	28.4	6.9	2.5	36.2	24.3
Snohomish	1.2	3.6	5.5	5.3	14.0	53.1	17.2
Stevens	22.7	--	.5	7.4	.9	33.4	35.0
Thurston	1.3	1.4	3.4	13.3	36.1	25.2	18.7
Wahkiakum	.5	.5	.5	1.9	.1	79.3	17.0
Whatcom	5.9	2.5	7.4	2.1	16.8	54.8	10.4
Idaho							
Ada	15.7	.7	.3	1.2	14.4	38.5	29.0
Bonner	13.0	--	.2	6.1	1.5	37.4	41.5
Franklin	29.6	3.2	--	,1	7.4	27.7	31.9
Gem	8.3	1.4	26.6	.1	1.9	18.7	43.0

APPENDIX TABLE I. (CON'T.)

Area, State and County	All Crops: Field Crops	Vege-tables	Fruits & Nuts	Forest Prod.	All Livestock & Liv. Prod.: Poultry P. Prod.	Dairy Prod.	Livestock Liv. Prod.
- - - - - - - -	Percent of Value of Farm Products Sold						
Shoshone	6.5	--	.3	7.6	25.7	11.3	48.5
Area E							
Oregon							
Clackamas	11.9	5.8	17.6	13.4	30.2	7.6	13.3
Hood River	1.0	--	85.4	.7	4.4	4.4	4.0
Jackson	6.7	1.0	44.9	1.6	11.0	14.7	18.7
Lane	16.8	15.6	10.6	10.9	15.6	14.9	15.9
Marion	27.4	19.6	19.4	7.7	8.7	7.6	9.5
Multnomah	6.9	16.9	15.9	36.8	4.2	10.3	8.7
Polk	40.8	8.6	17.5	1.1	6.4	11.0	14.4
Washington	19.1	4.7	28.0	12.0	10.3	17.3	8.5
Yamhill	23.6	12.7	17.8	3.9	18.7	9.4	13.8
Washington							
Chelan	.8	--	92.8	2.2	.1	.8	3.1
Cowlitz	7.3	6.6	3.5	25.4	14.5	24.2	18.4
Kitsap	1.1	.5	19.8	19.5	16.6	24.5	17.8
Okanogan	5.5	--	62.0	10.8	.1	1.3	20.2
Pierce	.6	6.7	10.0	11.8	33.7	24.1	13.0
Skagit	5.2	18.2	11.9	9.3	7.9	36.0	11.4
Skamania	2.8	.3	48.1	5.2	3.3	12.8	26.4

* Less than .05%.

EXPLANATION OF VARIABLE MEASUREMENT

The variables for the production function models were based primarily on data from the Agricultural Census of 1964. Other sources were used when census data was inadequate, but whenever possible the "other sources" were tied to the census data. In the case of drainage (X_7) and ACP (X_8), however, this was not possible.

The value of farm products sold was taken directly from the Agricultural Census for 1964 (57, 58, 59 Table 6, line 63). The value of home consumption was estimated by using the state estimates of value of home consumption from Farm Income Situation (53) and allocating this estimate among the counties according to the number of people on farms (57, 58, 59, Table 7, line 2).

Total man years of labor was estimated mostly from the Agricultural Census in three components: (a) Hired labor; (b) Family labor; and (c) Operator labor. Hired labor was estimated as expenditures for hired labor (57, 58, 59, Table 9, line 92) divided by average monthly farm wage rates for all farm laborers in the state times 12 (54). Family labor was estimated from the 1964 Agricultural Census by counting one man year of labor for each male person living on farms between the ages of 19 and 65 who was not a farm operator; plus 40% of a man year for each male person on farms between the ages of 15 and 19; plus 60% of a man year for each male person on farms over 65 years of age (who was not a farm operator); minus man years of work off farms by family members (assuming 300 days of off farm work equal to one man year). Operator labor was estimated from the 1964 Agricultural Census by counting one man-year per operator under 65 plus 60 percent of a man-year for operators over 65.

Current operating expenditures include expenditures for feed for livestock and poultry, seed, bulbs, and plants, fertilizer, gasoline, fuel and oil, and machine hire. The source for these items was the 1964 Agricultural Census (57, 58, 59, Table 9, lines 57, 73, 75, 78, and 89). Repairs and maintenance (R and M) were estimated using tractor, auto, truck and machinery repair and maintenance cost per unit by type of farm from a U. S. Department of Agriculture national survey (47, p. 25, 46, 52, 76). A weighted average cost per unit was obtained by taking R and M cost per unit (adjusted to 1964 price levels) times the appropriate percent of the corresponding type of farm in the county.

Pesticide expenditures were estimated using 1964 Agricultural Census and ERS Pesticide Uses Survey for 1964 (52). The latter was used to determine percentage of total acreage treated by crop, by state and expenditure per acre treated. The total expenditure (estimate by state and by crop) was allocated among the counties by the number of treated acres in each county. Pesticide expenditures on animals were estimated using the Pesticide Use Survey estimates of average cost per farm that treated any livestock, times the number of farms treating any animals (57, 58, 59, Table 8, lines 75 and 77).

The service flow of capital includes durable machinery items such as tractors, combines, trucks, etc., and was estimated by allocating Farm Income Situation Reports state estimates of 1964 capital consumption (53) among counties by; (1) dividing this state estimate among categories for major machinery items based on the ratios of one year's total depreciation for all major machinery items to one year's depreciation for each major item (based on new machinery prices), (2) calculating the service flow for each major item at the state level using equation (2a. 17) of Appendix I, and (3) allocating these service flows among counties according to the number of machinery items in each county contained in the major item category (57, 58, 59, Table 8, lines 5, 9, 12, 17, 20, 22, 26, 28, and 31).

Acres of cropland were taken directly from 1964 Agricultural Census (57, 58, 59, Table 1, line 17) and excludes timber land, range land, and waste land on farms and national or state forest and range lands. This quantity was adjusted by an index of land quality. The construction of this index was discussed in Chapter III with an example in footnote No. 6. The county indexes by Homogeneous Farming Areas are given in Appendix Table II.

The number of AUMs per county was taken from the Columbia

North Pacific Region Comprehensive Framework Studies (39) compiled under the direction of Economic Research Service, USDA.

Acre feet of irrigation water per county was estimated using average application rates from the 1957 Census of Irrigation (60, State Table 2, Idaho, Oregon, Washington, line 43). These rates were calculated using river basins irrigation rates (as reported in the Irrigation Census). A weighted average irrigation rate per county was estimated using the percentage contribution from each river basin to total irrigated acres in the county. Dot maps from (60) which show location of irrigated acres were used to establish these percentages. Total acre feet per county was estimated by multiplying this weighted average rate by the number of irrigated acres reported in the 1964 Agricultural Census. This procedure uses 1964 irrigated acres and assumes 1959 application rates. This variable was not measured in value of service flow terms since adequate private investment data was not available.

The service flow of drainage investment was based on Agricultural Stablization and Conservation Service (ASCS) historical records of farmer participation in Agricultural Conservation Program (ACP) cost sharing arrangements in drainage practices. It was assumed that most drainage investment was made under ACP and that the farmer's investment was equal to the Federal governments' share under ACP. (The farmers' share on drainage practices, as well as most other practices, is 50% of the total cost.) Time series data was obtained from ASCS Annual Reports (48, 49, 50) for each state and the service flow for 1964 was calculated using equation (2a. 15) of Appendix I solved for R and assuming a constant deterioration factor equal to the inverse of the expected life of the drainage practice. Life expectancies for all ACP practices were obtained from the Soil Conservation Service (55).

Water conservation practices include some 36 different practices under ACP (see Appendix Table III). The service flow was calculated for each practice using the same assumptions and data sources as for the drainage variable. Each practice included in this group is classified as water related conservation practices. Again, equation (2a. 15) solved for R was used to determine the 1964 service flow from each prior year's investment. And these flows were summed over years and practices to obtain the value for X_8.

APPENDIX TABLE II. Land Quality Index by Homogeneous Farming Areas, Pacific Northwest, 1964.

Area and County	Index of Cropland Quality
Area I	
<u>Oregon</u>	
Benton	154.6
Crook	95.7
Gilliam	84.1
Jefferson	75.1
Klamath	75.7
Malheur	74.7
Morrow	78.0
Union	95.2
Wallowa	100.4
Wasco	98.1
Umatilla	74.1
<u>Washington</u>	
Grant	106.7
Klickitat	97.2
Yakima	103.0
<u>Idaho</u>	
Bannock	86.6
Bear Lake	65.2
Boundary	273.4
Butte	81.7
Camas	109.1
Canyon	77.5
Caribou	100.5
Cassia	88.6
Clark	94.9
Clearwater	157.8
Custer	82.9
Elmore	97.7
Fremont	89.6
Gooding	75.2
Idaho	150.3
Jefferson	89.2
Jerome	91.6
Kootenai	140.4
Lincoln	76.2
Minidoka	89.7
Oneida	103.6
Owyhee	109.6

APPENDIX TABLE II Con't.

Area and County	Index of Crop-land Quality	Area and County	Index of Crop-land Quality
Area I Con't.		Idaho	
Payette	103.2	Adams	103.0
Teton	64.9	Blaine	110.3
Twin Falls	99.8	Boise	145.7
Valley	74.5	Lemhi	106.2
Washington	94.8	Area III	
Area II		Oregon	
Oregon		Sherman	105.1
Baker	99.7	Linn	141.3
Lake	78.1	Washington	
Wheeler	88.9	Adams	73.1
Grant	97.1	Benton	100.0
Harney	142.8	Columbia	167.6
Douglas	99.3	Douglas	117.9
Washington		Franklin	97.4
Asotin	101.6	Garfield	143.5
Ferry	114.9	Lincoln	135.8
Kittitas	155.2	Spokane	136.1
Pend Oreille	83.4	Walla Walla	126.9
San Juan	114.5	Whitman	190.9

APPENDIX TABLE II Con't

Area and County	Index of Cropland Quality	Area and County	Index of Cropland Quality
Area III Con't.		Washington	
		Clallam	102.3
Idaho			
Benewah	209.7	Clark	99.5
Bingham	58.7	Grays Harbor	97.7
Bonneville	66.9	Island	93.0
Latah	206.8	Jefferson	99.2
Lewis	198.2	King	94.8
Madison	85.5	Lewis	96.6
Nez Perce	170.2	Mason	99.0
Power	86.8	Pacific	100.1
Area IV		Snohomish	97.6
Oregon		Stevens	113.6
Josephine	84.5	Thurston	100.0
Lincoln	107.7	Whatcom	108.5
Coos	90.3	Wahkiakum	89.9
Tillamook	97.1	Idaho	
Columbia	88.1	Ada	109.9
Curry	108.5	Bonner	111.6
Deschutes	82.4	Franklin	78.9
Clatsop	89.7	Gem	82.8
		Shoshone	96.0

APPENDIX TABLE II Con't.

Area and County	Index of Cropland Quality	Area and County	Index of Cropland Quality
Area V		Hood River	88.7
Oregon		Jackson	113.9
Clackamas	113.3	Washington	
Lane	123.1	Cowlitz	109.5
Marion	129.9	Kitsap	93.1
Multnomah	103.2	Pierce	114.7
Polk	146.3	Skagit	72.6
Washington	157.4	Chelan	240.9
Yamhill	100.0	Okanogan	132.3
		Skamania	109.3

APPENDIX TABLE III. Selected Agricultural Conservation Practices Defined as Water Oriented Conservation Practices (ACP), Pacific Northwest, 1964.

Practice		Estimated Life Span*
A - ESTABLISHMENT OF PERMANENT PROTECTIVE COVER		
A-1	: Permanent Cover - Other	5
A-2	: Permanent Cover for Soil Protection	5
A-5	: Contour Stripcropping	20
A-6	: Field Stripcropping	20
A-8	: Tree Planting - Erosion	20
B - IMPROVEMENT AND PROTECTION OF ESTABLISHED VEGETATIVE COVER		
B-1	: Improving Established Forage Cover	5
B-5	: Wells for Livestock Water	20
B-5A	: Storage Tanks at Wells	20
B-6	: Springs or Seeps for Livestock Water	20
B-6A	: Additional Storage - Springs or Seeps	20
B-7	: Reservoirs for Livestock Water	20
B-8	: Pipelines for Livestock Water	20
B-8A	: Storage Tanks in Connection with Pipelines	20
B-8B	: Supplemental Livestock Water Storage	20
C - CONSERVATION AND DISPOSAL OF WATER		
C-1	: Sod Waterways to Dispose of Excess Water	12
C-2	: Permanent Cover Dams, Dikes, Ditchbanks, etc.	17
C-5	: Diversion Terraces, Ditches, Etc.	15
C-6	: Storage Type Erosion Dams	17
C-6A	: Non-Storage Type -- Dams, Ditches, Etc.	17
C-7	: Inlet or Outlet Protection	20
C-8	: Stream or Shore Protection	17
C-9	: Open Drainage System	7
C-9A	: Spreading--Spoil - Old Banks	10
C-10	: Underground Drainage System	20
C-11	: Leveling for Drainage	22
C-12	: Reorganizing Irrigation Systems - Siphons & Pipes	15
C-12P	: Reorganizing Irrigation Systems - Pools Only	15
C-13	: Leveling For Irrigation	22

APPENDIX TABLE III Con't.

Practice		Estimated Life Span*
C-14	: Reservoirs for Irrigation Water	20
C-15	: Lining Irrigation Ditches	15
C-16	: Spreader Ditches or Dikes	17
C-17	: Regular Subsoiling	5
C-17A	: Rotary Subsoiling	5
F-2	: Contour Farming	20
F-2B	: Deep Plowing	5

* In some cases these figures are midpoints of a range of expected life spans.

BASIC DATA

APPENDIX TABLE IV contains the data for the five homogeneous farming areas where X_3, X_7, and X_8 were calculated using a discount rate of five percent (R = 0.05). The data for each area where a discount rate of 0.075 and 0.10 was used to calculate X_3, X_7, and X_8 are available from the author upon request.

APPENDIX TABLE IV. BASIC DATA FOR FIVE HOMOGENEOUS FARMING AREAS, PACIFIC NORTHWEST, 1964.

AREA A R = .050

X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	Y
670.00	2599.98	1960.78	104.08	258.59	48.10	10.00	27.34	6373.10
514.00	2021.86	2220.89	67.87	267.00	237.54	.55	12.26	7379.30
411.00	1561.20	1339.77	223.26	230.55	23.97	1.47	17.90	5933.60
752.00	4807.22	2215.83	68.75	352.78	187.04	.81	11.73	18099.30
1510.00	7921.71	5823.92	146.67	723.78	740.97	4.01	41.35	25837.80
2747.00	8992.57	6373.38	157.35	1797.42	670.34	8.64	23.77	38246.60
624.00	2061.33	1941.13	286.58	201.35	111.30	1.42	23.00	8042.60
2081.00	10573.60	6155.78	466.64	559.26	302.84	13.63	60.26	34544.90
740.00	3325.52	2921.30	141.76	497.92	122.24	6.93	37.47	9085.80
621.00	1371.15	2541.02	91.42	383.73	186.65	2.16	12.76	5994.90
973.00	2151.63	2391.50	186.78	513.03	59.47	1.47	21.63	7836.00
2819.00	15010.94	6032.71	622.55	434.51	920.63	1.99	62.77	63616.90
695.00	2069.87	2578.67	182.06	157.71	66.63	1.31	39.72	6239.90
10944.00	29364.55	15955.71	345.61	690.60	1261.45	17.25	11.58	119948.90
766.00	2202.79	2574.67	152.25	48.91	188.75	.27	3.86	7370.20
469.00	1049.08	1930.57	74.02	77.57	119.38	.38	43.43	3742.00
344.00	807.63	1257.86	128.83	5.86	1.02	.47	14.36	2570.70
301.00	817.21	594.65	41.02	132.91	122.99	.04	1.91	3242.20
154.00	498.55	757.86	104.21	135.24	34.56	1.20	8.67	1963.90
4195.00	18464.42	11956.68	153.65	410.56	1195.84	6.05	9.74	57987.90
638.00	1924.11	2567.88	224.58	266.49	135.08	1.47	5.27	8548.80
1816.00	9793.83	5491.47	242.27	344.73	625.89	.54	5.16	33106.80
143.00	530.14	418.37	25.85	109.05	45.51	.05	3.04	1484.90
224.00	551.42	1060.85	51.40	54.00	.78	.72	11.41	1421.00
369.00	797.27	1349.39	33.85	184.62	257.12	.02	6.92	3072.70
510.00	2047.62	1539.52	40.32	134.62	197.46	2.61	7.43	8691.70
940.00	3339.91	3096.53	142.95	170.04	459.05	.11	5.92	14290.70
1061.00	3832.02	4036.36	57.17	164.87	413.96	.29	1.55	12474.50
937.00	2637.02	3513.06	293.53	168.31	8.53	4.27	24.93	9765.40
1239.00	4425.79	4551.78	140.47	356.84	964.73	.16	3.97	18656.90
1462.00	5833.10	4695.72	108.71	117.70	601.44	.34	.37	22067.50
659.00	2940.84	2299.72	130.00	225.49	14.52	2.31	19.16	4758.00
532.00	1693.81	1725.08	34.40	112.27	220.48	.14	1.91	5156.80
1618.00	7475.47	4588.03	139.08	320.28	812.27	.27	1.53	24743.80
521.00	1295.90	2106.55	227.43	108.25	91.25	.89	9.56	4938.20
975.00	3606.83	3582.90	86.56	624.50	432.68	2.98	9.82	11300.60
1016.00	3364.49	3021.72	47.10	44.58	305.82	10.46	3.90	10486.80
366.00	915.99	1467.45	68.14	113.43	81.71	.69	3.95	3561.50
2854.00	11694.57	9734.28	241.14	293.11	1212.27	.56	5.77	36833.60
157.00	343.74	639.27	15.54	120.19	74.93	1.77	25.19	1379.80
722.00	2096.90	2324.16	80.22	114.18	104.31	3.30	14.04	7878.20

APPENDIX TABLE IV (Con't.)

AREA B R = .0500

X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	Y
840.00	3129.34	3900.89	111.45	677.94	379.53	3.23	23.99	11029.90
1280.00	2361.78	3574.51	44.08	507.18	47.45	.91	56.75	6521.10
402.00	1184.98	2005.56	60.98	641.73	197.34	2.82	16.81	4306.30
533.00	1802.19	2975.20	245.76	1402.11	210.28	10.95	42.61	5220.20
521.00	1464.86	2594.56	99.42	1198.16	445.60	4.95	26.88	5673.30
167.00	481.33	767.27	28.03	322.63	38.88	.30	8.57	1641.50
312.00	1469.75	1028.47	86.57	100.13	5.12	1.40	8.10	4600.80
216.00	437.65	841.83	30.10	132.76	21.90	.11	9.74	1251.00
1024.00	4410.61	3265.61	105.18	742.54	482.60	26.82	36.07	17007.90
223.00	667.82	963.86	22.42	24.60	12.49	1.52	14.12	1410.30
85.00	196.30	356.01	6.79	51.25	.01	.46	6.76	458.00
247.00	650.17	1191.51	25.38	88.52	60.77	1.52	8.52	2265.80
325.00	1033.93	1178.03	42.67	206.21	177.36	.23	3.70	3771.30
99.00	230.81	406.16	7.80	69.85	37.34	.03	1.95	690.60
479.00	932.16	1969.85	61.11	76.96	406.96	1.00	6.72	4078.00

APPENDIX TABLE IV (Con't.)

AREA C R = .0500

X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	Y
2123.00	8931.63	6178.60	327.75	755.70	68.34	22.76	55.21	23039.90
393.00	1474.29	1252.08	288.27	154.93	6.84	25.85	10.65	6627.70
1410.00	8416.13	4378.14	565.95	220.98	304.46	2.78	70.88	31247.00
1474.00	5329.49	3610.45	263.71	116.36	247.96	2.29	5.61	16709.30
819.00	4471.13	1536.22	322.02	65.72	25.34	1.18	10.47	8666.10
1342.00	3268.15	3384.24	611.91	173.38	61.06	1.52	29.73	13520.00
1153.00	5949.23	3465.74	342.43	343.30	398.10	1.01	4.61	20744.00
419.00	1663.95	1305.72	265.40	121.65	7.46	1.68	15.76	6414.20
1285.00	5400.75	5139.85	1155.67	200.90	50.51	19.39	80.30	21442.30
2109.00	9014.91	7780.43	577.88	96.58	60.18	24.88	1098.75	27169.40
1752.00	8126.37	4323.10	691.53	68.19	178.36	4.19	25.31	24640.10
2356.00	12441.46	8359.76	1948.97	209.78	45.21	57.43	56.21	51233.50
253.00	914.34	1059.88	137.72	72.28	.27	10.89	22.93	2831.20
2551.00	9508.35	7995.36	145.64	389.67	1588.63	1.72	7.40	37040.30
1587.00	5994.35	5220.87	193.66	155.39	789.96	.57	90.43	24253.70
908.00	3389.22	3619.70	483.93	100.16	.87	21.53	37.90	11834.20
391.00	1654.93	1550.10	289.75	19.42	.62	4.64	27.11	7341.40
965.00	3635.52	2956.02	147.01	112.58	436.20	.44	12.55	12729.20
764.00	2962.78	2689.55	330.52	173.42	6.23	5.78	23.94	10713.60
668.00	2178.54	1931.02	242.83	53.42	235.62	4.54	3.61	10124.50

APPENDIX TABLE IV (Con't.)

AREA D R = .0500

X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	Y
465.00	1444.83	946.79	7.35	15.58	.23	.52	4.39	3118.40
745.00	1901.30	2244.05	18.03	421.04	2.79	7.79	7.88	4546.40
767.00	2437.74	1645.01	14.36	695.32	28.90	3.61	46.26	5926.80
264.00	723.32	626.24	3.25	213.83	5.82	.58	15.99	1369.40
550.00	2486.33	2046.26	31.00	518.23	164.58	.26	5.66	6216.70
698.00	1903.82	1628.85	12.31	151.50	62.42	3.37	19.61	4426.50
279.00	725.00	778.56	5.82	180.62	2.84	1.13	19.15	1236.30
596.00	3095.39	2651.76	7.46	336.03	9.29	4.03	37.58	7413.30
414.00	1234.26	1624.96	16.02	81.41	75.97	2.42	14.65	3378.10
1684.00	5861.21	5453.80	55.73	328.32	32.44	10.80	38.05	13036.90
531.00	2134.33	1808.89	17.37	116.86	15.03	1.27	27.04	4715.60
311.00	1827.02	950.85	9.99	44.72	3.52	2.60	8.63	3001.10
145.00	476.01	473.94	3.93	46.27	4.81	2.62	6.53	834.10
1789.00	9496.02	4247.52	18.83	333.25	12.19	26.25	28.18	21528.20
1273.00	6175.84	4926.54	58.50	316.26	31.09	13.14	38.92	11076.90
199.00	770.06	632.35	3.84	54.15	3.53	1.76	4.69	1065.70
313.00	720.69	958.92	8.97	115.14	10.89	4.54	9.92	2020.30
1761.00	8516.95	5432.85	31.58	229.65	20.72	32.65	39.60	17162.70
1113.00	2859.70	4686.74	146.77	141.15	64.31	6.24	59.56	7970.90
931.00	4463.64	2629.88	20.63	139.41	33.59	3.67	26.54	8409.80
189.00	633.70	712.69	4.79	67.92	1.03	4.46	5.46	1686.30
2322.00	11417.28	7570.79	77.24	605.85	28.53	20.84	61.34	26467.00
1651.00	7424.96	6092.76	102.13	154.87	492.91	2.52	1.57	17095.20
569.00	1208.48	1908.78	41.68	30.54	7.12	.53	17.94	2460.50
795.00	2904.47	3174.69	103.40	33.75	99.28	5.24	7.17	8329.70
961.00	2618.27	2941.50	35.16	285.69	286.52	1.62	6.07	8561.10
49.00	67.03	146.48	2.22	2.40	.14	.01	.01	174.20

APPENDIX TABLE IV (Con't.)

AREA E R = .0500

X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	Y
3400.00	11843.31	7464.33	115.70	277.93	13.83	57.25	26.13	28043.90
1127.00	2303.42	1277.43	15.29	44.99	60.02	5.19	.16	6734.60
1811.00	5007.02	3674.63	59.45	244.28	157.47	5.51	15.12	13507.40
2401.00	7754.58	6023.45	125.23	553.04	108.08	9.20	50.86	17280.40
5192.00	15565.49	8387.08	265.35	212.11	83.23	118.55	23.53	44657.60
1539.00	3463.93	1940.78	24.71	175.85	22.69	21.99	7.17	12110.50
1237.00	4223.34	3325.78	178.05	521.18	16.87	64.03	31.14	10856.10
2619.00	7745.56	5735.61	175.60	396.49	17.53	54.76	34.85	21360.60
1824.00	7331.06	4810.88	124.52	11596.82	21.14	59.57	14.17	18255.20
2953.00	4514.34	2899.68	112.57	74.11	116.66	.71	3.43	23565.30
516.00	1703.22	1611.18	13.34	69.07	13.48	3.29	11.35	4177.10
452.00	1190.23	1395.80	4.49	20.66	6.37	4.92	2.40	1980.50
2394.00	4349.27	4545.52	171.12	334.87	230.18	1.54	40.98	20787.30
1936.00	9309.31	4094.92	24.95	77.65	26.66	26.02	12.38	19335.70
1943.00	9916.77	4565.25	44.63	165.81	25.60	27.03	36.73	20765.60
86.00	227.43	351.36	2.92	5.70	1.92	.10	3.55	466.60

MEANS OF THE VARIABLES

The arithmetic and geometric means of the variables used in the production functions for the five homogeneous farming areas are given in APPENDIX TABLE V. The means of X_3, X_7, X_8, and the corresponding predicted values of Y as the discount rate was changed from 0.05 to 0.075 and 0.10 are available from the author upon request.

APPENDIX TABLE V. Arithmetic and geometric means of the variables, Pacific Northwest, 1964.

Area	Variable								Log Equation	Actual	Natural Equation
	X_1	X_2	X_3 *	X_4	X_5	X_6	X_7 *	X_8 *	$\hat{Y}$	Y	$\hat{Y}$
A											
Arith	1249.4	4598.8	3624.5	150.6	295.5	333.2	2.8	16.0	16,326.4	16,553.0	16,529.8
Geom.	791.1	2664.5	2665.0	112.6	202.0	144.8	1.0	9.7	9,259.9		9,207.0
B											
Arith	450.2	1363.6	1801.3	65.2	416.2	168.2	3.8	18.1	4,657.0	4,661.7	4,661.5
Geom	342.0	967.0	1414.1	44.0	221.0	47.6	1.1	12.4	3,095.1		3,125.9
C											
Arith	1241.1	5236.3	3889.3	466.6	180.2	225.6	10.8	84.5	18,416.1	18,416.1	18,223.1
Geom	1037.0	4638.0	3257.0	362.0	134.0	43.0	4.6	25.3	14,808.6		14,212.8
D											
Arith	791.3	3167.7	2553.4	31.8	209.6	55.6	6.1	20.7	7,189.5	7,179.6	7,149.3
Geom	573.0	1962.0	1802.0	17.4	125.0	13.7	2.6	11.2	4,439.60		4,454.6
E											
Arith	1964.4	6028.0	3881.5	91.1	923.2	57.6	28.7	19.6	16,542.2	16,492.8	16,482.0
Geom	1458.0	4254.0	3069.0	49.4	168.0	30.3	10.0	11.1	11,357.3		10,691.4

* Calculated using a discount rate of 0.05.

APPENDIX TABLE VI. Multicollinearity tests for the linear models in five homogeneous farming areas, Pacific Northwest, 1964.

Area	Test	Value of Statistic								Critical Values of χ^2, F, and t Distributions
B	General									
	$\chi^2_{(v)}$	136.829								$\chi^2_{(28, .95)} = 41.3$ $\chi^2_{(28, .995)} = 51.0$
	Location									$F^7_{7(.90)} = 2.78$ $F^7_{7(.95)} = 3.79$
	$\omega_{(v_1, v_2)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$F^7_{7(.995)} = 8.89$
		72.7	45.4	49.5	3.9	14.2	17.8	19.6	67.7	
Patterns*										$t_{(7, .90)} = 1.895$
$t_{(v)}$		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$t_{(7, .95)} = 2.365$
	X_1		2.87	-.17	-.75	-2.93	2.33	-2.01	4.59	$t_{(7, .999)} = 5.405$
	X_2	.861		2.10	.09	1.37	-3.08	6.41	-3.78	
	X_3	.913	.852		1.15	-.26	2.39	-2.90	1.47	
	X_4	.400	.536	.642		.60	-.42	.38	.02	
	X_5	.493	.539	.718	.835		2.43	-1.31	3.27	
	X_6	.517	.656	.683	.500	.586		4.06	-5.58	
	X_7	.511	.785	.522	.538	.535	.601		3.60	
	X_8	.860	.688	.833	.574	.704	.314	.501		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VI. (CON'T.)

Area	Test					Value of Statistic					Critical Values of χ^2, F, and t Distributions
C	General										$\chi^2_{(28,\ .95)} = 41.3$
	$\chi^2_{(v)}$					99.758					$\chi^2_{(28,\ .995)} = 51.0$
	Location										$F^{12}_{7(.90)} = 2.67$ $F^{12}_{7(.95)} = 3.57$
	$\omega_{(v_1,\ v_2)}$		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$F^{12}_{7(.995)} = 8.18$
			44.4	18.4	75.4	15.8	4.3	10.4	1.4	8.3	
	Patterns*										$t_{(12,\ .90)} = 1.782$
	$t_{(v)}$		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$t_{(13,\ .95)} = 2.179$
		X_1		2.85	3.48	-1.92	-.87	-.92	-.32	-1.58	$t_{(12,\ .999)} = 4.318$
		X_2	.936		-.67	1.45	.84	.53	.03	.72	
		X_3	.955	.914		4.69	2.75	3.68	.36	4.94	
		X_4	.422	.561	.522		-3.80	-5.80	.92	-5.07	
		X_5	.527	.486	.476	.019		-3.19	.05	-3.83	
		X_6	.499	.356	.436	-.277	.277		-.26	-4.93	
		X_7	.326	.438	.483	.758	.179	-.336		-.11	
		X_8	.332	.314	.444	.111	-.095	-.106	.266		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VI. (CON'T.)

Area	Test	Value of Statistic								Critical Values of χ^2, F, and t Distributions
D	General $\chi^2_{(v)}$	237.708								$\chi^2_{(28,\ .95)} = 41.3$ $\chi^2_{(28,\ .995)} = 51.0$
	Location $\omega_{(v_1,\ v_2)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$F^{19}_{7,\ (.90)} = 2.59$ $F^{19}_{7,\ (.995)} = 7.75$ $F^{19}_{7,\ (.95)} = 3.44$
		92.9	55.1	92.2	11.2	2.9	8.1	12.5	9.2	

Patterns* $t_{(v)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Critical Values
									$t_{(19,\ .90)} = 1.729$ $t_{(19,\ .95)} = 2.093$
X_1		2.39	2.72	-.19	.92	.49	.94	.27	$t_{(19,\ .999)} = 3.883$
X_2	.961		1.37	-2.04	-.29	.22	1.39	-.32	
X_3	.963	.912		3.44	-.22	.96	-.03	1.14	
X_4	.581	.453	.721		-1.63	.71	-.37	.82	
X_5	.511	.488	.448	.095		1.09	-.48	2.27	
X_6	.310	.246	.384	.470	.074		-2.71	-4.39	
X_7	.768	.819	.678	.204	.361	-.146		-.77	
X_8	.622	.573	.625	.413	.543	-.244	.542		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VI. (CON'T)

Area	Test	Value of Statistic	Critical Values of χ^2, F, and t Distributions
E	<u>General</u> $\chi^2_{(v)}$	122.776	$\chi^2_{(28, .95)} = 41.3$ $\chi^2_{(28, .995)} = 51.0$
	<u>Location</u> $\omega_{(v_1, v_2)}$		$F^8_{7, (.90)} = 2.75$ $F^8_{7, (.995)} = 8.68$ $F^8_{7, (.95)} = 3.73$

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
39.7	42.6	30.0	34.3	.9	3.6	23.7	10.7

<u>Patterns</u>*

$t_{(v)}$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Critical Values
									$t_{(8, .90)} = 1.860$ $t_{(8, .95)} = 2.306$
X_1		2.82	1.00	3.26	-2.01	.27	-1.94	-4.41	$t_{(8, .999)} = 5.041$
X_2	.869		1.08	-4.06	.69	.42	2.99	1.84	
X_3	.884	.929		.10	1.27	-.26	-.01	1.58	
X_4	.782	.620	.776		1.10	1.31	5.78	3.21	
X_5	-.009	.102	.142	.152		.22	-.57	-1.87	
X_6	.349	.041	.212	.353	-.133		-2.21	-.26	
X_7	.670	.760	.719	.729	.266	-.243		-2.43	
X_8	.413	.488	.668	.596	-.042	.327	.274		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VII. Multicollinearity tests for the log-linear models in five homogeneous farming areas, Pacific Northwest, 1964.

Area	Test	Value of Statistic								Critical Values of χ^2, F, and t Distributions
A	General $\chi^2_{(v)}$	345.378								$\chi^2_{(28,\ .95)} = 41.3$ $\chi^2_{(28,\ .995)} = 51.0$
	Location $\omega_{(v_1,\ v_2)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$F^{32}_{7,(.90)} = 2.56$ $F^{32}_{7,(.995)} = 7.53$ $F^{32}_{7,(.95)} = 3.38$
		187.2	93.1	120.5	7.7	5.8	12.5	5.8	5.5	
Patterns* $t_{(v)}$		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$t_{(32,\ .90)} = 1.697$ $t_{(32,\ .95)} = 2.042$
	X_1		3.93	6.04	.16	-.26	.07	.58	-.61	$t_{(32,\ .999)} = 3.646$
	X_2	.972		.24	.89	.92	1.17	.65	-.57	
	X_3	.980	.955		.86	-.42	.62	.00	.51	
	X_4	.648	.636	.661		.31	-2.35	-1.22	2.02	
	X_5	.547	.585	.539	.353		3.13	.62	1.92	
	X_6	.667	.693	.654	.185	.602		-1.90	-.84	
	X_7	.440	.431	.439	.423	.349	.035		3.38	
	X_8	.063	.058	.087	.376	.255	-.223	.578		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VII. (CON'T)

Area	Test	Value of Statistic								Critical Values of χ^2, F, and t Distributions
B	General $\chi^2_{(v)}$	134.867								$\chi^2_{(28,\ .95)} = 41.3$ $\chi^2_{(28,\ .995)} = 51.0$
	Location (v_1, v_2)	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$F^7_{7,\ (.90)} = 2.78$ $F^7_{7,\ (.995)} = 8.89$ $F^7_{7,\ (.95)} = 3.79$
		74.2	30.9	65.6	6.2	2.4	3.8	6.5	6.8	

Patterns* $t_{(v)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	$t_{(7,\ .90)} = 1.895$ $t_{(7,\ .95)} = 2.365$
X_1		3.89	3.89	-1.19	-.94	-.67	-2.46	.62	$t_{(7,\ .999)} = 5.405$
X_2	.954		-1.46	1.45	.70	.24	2.25	-.84	
X_3	.971	.934		1.10	.68	1.81	1.67	.47	
X_4	.801	.868	.860		.42	.06	-.06	-.15	
X_5	.684	.695	.752	.744		.38	-1.01	1.28	
X_6	.667	.646	.725	.675	.589		-.65	-1.65	
X_7	.692	.780	.757	.760	.577	.370		1.22	
X_8	.778	.752	.813	.673	.689	.320	.786		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VII. (CON'T)

Area	Test	Value of Statistic	Critical Values of χ^2, F, and t Distributions
C	General $\chi^2_{(v)}$	137.299	$\chi^2_{(28, .95)} = 41.3$ $\chi^2_{(28, .995)} = 51.0$
	Location $\omega_{(v_1, v_2)}$	X_1 60.4, X_2 24.4, X_3 32.8, X_4 2.2, X_5 1.2, X_6 7.1, X_7 3.3, X_8 1.9	$F^{12}_{7,(.90)} = 2.67$ $F^{12}_{7,(.995)} = 8.18$ $F^{12}_{7,(.95)} = 3.57$

Patterns* $t_{(v)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Critical Values
									$t_{(12, .90)} = 1.782$ $t_{(12, .95)} = 2.179$
X_1		2.75	3.69	.34	-.02	.91	-.67	-.54	$t_{(12, .999)} = 4.318$
X_2	.963		-.28	.96	.17	.25	.03	.23	
X_3	.954	.917		-.61	.47	-.01	1.31	1.56	
X_4	.418	.477	.457		-.64	-.32	1.69	.40	
X_5	.563	.539	.563	.124		.45	.81	-.67	
X_6	.730	.689	.605	-.009	.469		-1.94	-1.13	
X_7	.009	.045	.170	.545	.073	-.474		-.21	
X_8	.346	.358	.480	.465	.063	-1.08	.432		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VII. (CON'T)

Area	Test		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Critical Values of χ^2, F, and t Distributions
D	General $\chi^2_{(v)}$										$\chi^2_{(28,\ .95)} = 41.3$ $\chi^2_{(28,\ .995)} = 51.0$
	Value of Statistic: 271.844										
	Location $\omega_{(v_1,\ v_2)}$		86.6	59.2	143.5	31.6	8.4	6.0	7.4	5.9	$F^{19}_{7,(.90)} = 2.59$ $F^{19}_{7,(.95)} = 3.44$
Patterns* $t_{(v)}$		X_1		2.44	2.16	.45	.80	-1.01	-.63	.17	$t_{(19,\ .95)} = 1.729$ $t_{(19,\ .95)} = 2.093$ $t_{(19,\ .999)} = 3.883$
		X_2	.967		2.01	-2.25	-.76	.69	.44	.24	
		X_3	.976	.953		4.12	1.10	.04	1.33	-.39	
		X_4	.830	.759	.885		-2.78	2.08	-.97	.34	
		X_5	.688	.699	.665	.401		2.49	-.22	1.82	
		X_6	.669	.646	.717	.736	.589		-.56	-.27	
		X_7	.728	.765	.727	.485	.643	.409		2.23	
		X_8	.659	.683	.641	.417	.728	.429	.764		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

APPENDIX TABLE VII. (CON'T)

Area	Test	Value of Statistic	Critical Values of χ^2, F, and t Distributions
E	General $\chi^2_{(v)}$	142.451	$\chi^2_{(28,\ .95)} = 41.3$ $\chi^2_{(28,\ .995)} = 51.0$

	Location $\omega_{(v_1,\ v_2)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Critical Values
		48.7	49.4	41.8	10.6	2.9	4.1	6.0	4.5	$F^8_{7,\ (.90)} = 2.75$ $F^8_{7,\ (.995)} = 8.68$ $F^8_{7,\ (.95)} = 3.73$

Patterns* $t_{(v)}$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Critical Values
									$t_{(8,\ .90)} = 1.860$ $t_{(8,\ .95)} = 2.306$
X_1		2.21	.74	2.15	-1.12	1.22	-.28	-2.09	$t_{(8,\ .999)} = 5.041$
X_2	.950		1.45	-1.51	.32	-.06	1.36	.47	
X_3	.926	.964		.36	.36	-.06	.11	2.14	
X_4	.861	.816	.878		1.69	-.08	.12	1.28	
X_5	.629	.681	.726	.764		.76	.95	-.41	
X_6	.723	.570	.569	.630	.382		-2.05	-.22	
X_7	.654	.788	.741	.573	.650	.103		-.76	
X_8	.430	.545	.676	.626	.545	.172	.425		

* The lower diagonal is the simple correlation matrix and the upper diagonal the t values.

EXAMPLE OF THEIL'S PRIOR INFORMATION MODEL

The following example of Theil's prior information model is designed to show the steps involved in the calculations. Also, it shows the reason that combinations of large sample coefficients (and large standard errors) with small values from the prior information, yield small posterior estimates.

Consider a simple example of Theil's model. The data for this example were designed to indicate why the coefficients for X_7 and X_8 in Table VI (see Ch. IV, p. 92) changed drastically.

Data for the example are given in raw form as sums of squares and products:

Y	X
3	.001
8	.002
10	.003

$$\begin{bmatrix} X'X & X'Y \\ & \\ Y'X & Y'Y \end{bmatrix} = \begin{bmatrix} 3.000000 & 0.006000 & 21.000000 \\ & 0.000014 & 0.049000 \\ & & 173.000000 \end{bmatrix}.$$

We first calculate the OLS solutions which are:

$$b = \begin{bmatrix} 0.000007 \\ \\ 3500.000000 \end{bmatrix} ; \; s^2 = 1.500 \; ; \; (X'X)^{-1} = \begin{bmatrix} 2.333333 & -1,000.000000 \\ \\ & 500,000.000000 \end{bmatrix}.$$

also, S. E. of b_o = 1.871, S. E. b_1 = 866.025, and R^2 = .9423.

Assume that we also have prior information on only the slope coefficient and its variance -- nothing is known about the constant term. Assume,

$$r = [\,1.00\,]\ ; \quad R = [\,0 \quad 1\,]\ ; \quad = [\,0.04\,]\ .$$

First we calculate the comparability statistic, $\hat{\gamma}$ as shown in (3.17). Since we have only one prior estimate, $r - R\hat{\beta}$ is a scalar, viz., -3499.00 and $R(X'X)^{-1}R'$ reduces to the lower right hand term of $(X'X)^{-1}$, 500,000. Thus,

$$= 16.28.$$

This quantity indicates that the two information sources are not compatible since it is significantly greater than even a $\chi^2_{(1)}$ at the 99.5% level ($\chi^2_{(1)}$.995 = 7.88). If this were an actual case, it would indicate that we need go no further. However, the purpose of this example is to show all the solution parts. New estimates of the parameters are (from 3.18):

$$\hat{\beta} = \begin{bmatrix} 6.9954 \\ 2.3056 \end{bmatrix}; \quad \sigma_{\beta_0} = 0.7071 \text{ and } \sigma_{\beta_1} = 0.2$$

Notice that the coefficient for X changed from a sample estimate of 3500.0 to a posterior estimate of 2.3056. The new estimate of the model variance is $\hat{\sigma}^2 = 12.992$. The new R^2 (in terms of the sample data) is .00061, which we would expect due to the large difference in the estimates of the slope coefficient.

The results of this example were designed to demonstrate the general magnitude of change in the sample estimate compared to the posterior estimate. Similar results were shown in Chapter V, Table VI. The results in terms of R^2 and regression significance were not so drastically affected as in this example since the large changes were associated with only a small portion of the total explanation of the variance in Y, where as in the above example, the change in one coefficient had a direct impact on explained variation.

PRIOR INFORMATION FOR THE LINEAR MODELS

APPENDIX TABLE VIII contains prior information for the linear models of Areas A through E, including the coefficient estimates (designated SMALL R) and the variance-covariance matrix (designated VAR-COV MATRIX FOR SMALL R). The parameter estimates in this case are expected market equilibrium values. The variances are specified such that a "t" test (given the parameter estimate and the degrees of freedom in each area) would be significant if a statistical "t" test was actually appropriate. TABLE IX shows similar information where the parameter estimates are taken from former studies. Statistics from the ordinary least-squares regression calculations are included (e.g., X'X, X'Y, S^2, Y'Y and N).

APPENDIX TABLE VIII. PRIOR INFORMATION BASED ON EXPECTED MARKET EQUILIBRIUM VALUES IN FIVE HOMOGENEOUS FARMING AREAS, PACIFIC NORTHWEST, 1964.

AREA A

SMALL R

4.56	1.00	1.00	9.55	1.65	6.00	1.00	1.00

VAR-COV MATRIX FOR SMALL R

7.212	.01	.01	.01	.01	.01	.01	.01
.01	.347	.01	.01	.01	.01	.01	.01
.01	.01	.166	.01	.01	.01	.01	.01
.01	.01	.01	31.674	.01	.01	.01	.01
.01	.01	.01	.01	.945	.01	.01	.01
.01	.01	.01	.01	.01	12.503	.01	.01
.01	.01	.01	.01	.01	.01	.347	.01
.01	.01	.01	.01	.01	.01	.01	.347

AREA B

SMALL R

4.56	1.00	1.00	5.00	1.65	6.00	1.00	1.00

VAR-COV MATRIX FOR SMALL R

5.508	.01	.01	.01	.01	.01	.01	.01
.01	.265	.01	.01	.01	.01	.01	.01
.01	.01	.101	.01	.01	.01	.01	.01
.01	.01	.01	6.620	.01	.01	.01	.01
.01	.01	.01	.01	.721	.01	.01	.01
.01	.01	.01	.01	.01	9.536	.01	.01
.01	.01	.01	.01	.01	.01	.265	.01
.01	.01	.01	.01	.01	.01	.01	.265

AREA C

SMALL R

4.56	1.00	1.00	10.35	1.65	6.00	1.00	1.00

VAR-COV MATRIX FOR SMALL R

6.446	.01	.01	.01	.01	.01	.01	.01
.01	.310	.01	.01	.01	.01	.01	.01
.01	.01	.135	.01	.01	.01	.01	.01
.01	.01	.01	33.212	.01	.01	.01	.01
.01	.01	.01	.01	.845	.01	.01	.01
.01	.01	.01	.01	.01	11.162	.01	.01
.01	.01	.01	.01	.01	.01	.310	.01
.01	.01	.01	.01	.01	.01	.01	.310

SMALL R

APPENDIX TABLE VIII (Con't.)

AREA D

4.56	1.00	1.00	14.40	1.65	6.00	1.00	1.00

VAR-COV MATRIX FOR SMALL R

6.917	.01	.01	.01	.01	.01	.01	.01
.01	.333	.01	.01	.01	.01	.01	.01
.01	.01	.154	.01	.01	.01	.01	.01
.01	.01	.01	68.956	.01	.01	.01	.01
.01	.01	.01	.01	.906	.01	.01	.01
.01	.01	.01	.01	.01	11.972	.01	.01
.01	.01	.01	.01	.01	.01	.333	.01
.01	.01	.01	.01	.01	.01	.01	.333

AREA E

SMALL R

4.56	1.00	1.00	18.00	1.65	6.00	1.00	1.00

VAR-COV MATRIX FOR SMALL R

5.789	.01	.01	.01	.01	.01	.01	.01
.01	.279	.01	.01	.01	.01	.01	.01
.01	.01	.112	.01	.01	.01	.01	.01
.01	.01	.01	90.231	.01	.01	.01	.01
.01	.01	.01	.01	.759	.01	.01	.01
.01	.01	.01	.01	.01	10.024	.01	.01
.01	.01	.01	.01	.01	.01	.279	.01
.01	.01	.01	.01	.01	.01	.01	.279

APPENDIX TABLE IX. PRIOR INFORMATION BASED ON PREVIOUS STUDIES IN FIVE HOMOGENEOUS FARMING AREAS, PACIFIC NORTHWEST, 1964.

```
AREA  A

(8F10.0)
  SMALL B   K     S SQUARED                          N               YTY
     9      8     4583173.2127                       41         30857390900.

                                  XTY
   6786752359014990813660186051734474440 166198918 328321353 489616708   3986102
  13317052

                                 SMALL R
   1.081       2.12        2.12      36.00       2.25       6.57       2.12       2.12

                                 SMALL B
-1142.3364 5.183764  2.9200819 -2.068772 15.184418 5.4538322 5.5131765 -263.4358
 17.801240

                                  XTX
        41      51226     188550     148605       6176      12117      13662       114
       655
     51226  190064150  619952863  395725990   11733538   24384263   36172139    325743
    891655
    188550  619952863217887646013991286 30   44428108   87812018  133150167   1070970
   3640451
    148605  3957259901399128630  958882346   30847059   65672497   91170359    708668
   2667303
      6176   11733538   44428108   30847059    1526026    2221329    2718560     23977
    143465
     12117   24384263   87812018   65672497    2221329    7191795    5920613     56156
    247357
     13662   36172139  133150167   91170359    2718560    5920613    9871789     53933
    209503
       114     325743    1070970     708668      23977      56156      53933       946
      2695
       655     891655    3640451    2667303     143465     247357     209503      2695
     20150

                                   R
   0.0         1.0        0.0        0.0        0.0        0.0        0.0        0.0
   0.0
   0.0         0.0        1.0        0.0        0.0        0.0        0.0        0.0
   0.0
   0.0         0.0        0.0        1.0        0.0        0.0        0.0        0.0
   0.0
   0.0         0.0        0.0        0.0        1.0        0.0        0.0        0.0
   0.0
   0.0         0.0        0.0        0.0        0.0        1.0        0.0        0.0
   0.0
   0.0         0.0        0.0        0.0        0.0        0.0        1.0        0.0
   0.0
   0.0         0.0        0.0        0.0        0.0        0.0        0.0        1.0
   0.0
   0.0         0.0        0.0        0.0        0.0        0.0        0.0        0.0
   1.0

                        VAR-COV MATRIX FOR SMALL R
```

APPENDIX TABLE IX (Con't.)

.406	.01	.01	.01	.01	.01	.01	.01
.01	1.560	.01	.01	.01	.01	.01	.01
.01	.01	1.560	.01	.01	.01	.01	.01
.01	.01	.01	450.034	.01	.01	.01	.01
.01	.01	.01	.01	1.758	.01	.01	.01
.01	.01	.01	.01	.01	14.992	.01	.01
.01	.01	.01	.01	.01	.01	1.560	.01
.01	.01	.01	.01	.01	.01	.01	1.560

```
AREA  B

(8F10.0)
  SMALL B   K     S SQUARED                    N              YTY
     9      8      30907.6195                 15          595522302.0

                                   XTY
    69926  48637782 166218709 184787474   6404995  42849179  19803482    612936
  1844012

                                  SMALL R
   1.081      2.12      2.12     36.00      2.25      6.57      2.12      2.12

                                  SMALL B
 -37.27492   2.35479   3.07784    .11215  -5.07647   1.04980   1.24571 102.43136
 -80.70904

                                   XTX
  15.0     6753.0     20453.68  27019.32  977.75    6242.57   2523.63   56.25
 271.29
6753.0     4744429.0014133369.917476871.9559004.7403864862.071584180.7342711.5000
189704.970
20453.68  14133369.947103961.653477387.11867204.4612382518.85348881.07166396.951
551409.325
27019.32  17476871.953477387.168527727.22411229.8216482225.96567371.98161969.079
712116.501
977.75     559004.7401867204.462411229.82115425.047717757.003239910.8366856.07630
25535.6897
6242.57    3864862.0712382518.816482225.9717757.0035277650.371687241.1946233.4537
182301.015
2523.63    1584180.735348881.076567371.98239910.8361687241.19865251.15619858.6520
58206.1210
56.25      42711.5000166396.951161969.0796856.0763046233.453719858.6520890.879100
1803.70160
271.29     189704.970551409.325712116.50125535.6897182301.01558206.12101803.70160
8532.08590
```

APPENDIX TABLE IX (Con't.)

R

	1.						
		1.					
			1.				
				1.			
					1.		
						1.	
							1.
1.							

VAR-COV MATRIX FOR SMALL R

.309	.01	.01	.01	.01	.01	.01	.01
.01	1.190	.01	.01	.01	.01	.01	.01
.01	.01	.456	.01	.01	.01	.01	.01
.01	.01	.01	343.287	.01	.01	.01	.01
.01	.01	.01	.01	1.341	.01	.01	.01
.01	.01	.01	.01	.01	11.431	.01	.01
.01	.01	.01	.01	.01	.01	1.190	.01
.01	.01	.01	.01	.01	.01	.01	1.190

AREA C

(8F10.0)

SMALL B	K	S SQUARED	N	YTY
9	8	4760307.5306064	20	9531146060.0

XTY

368322 59775219726355143301910034400 235001472 81029081 120080080 5531382
43050822

SMALL R

1.081	2.12	2.12	36.00	2.25	6.57	2.12	2.12

SMALL B

-402.43849-4.9310 2.7377 1.2396 8.5828 -1.9247 9.3243 16.0995
-1.8122

APPENDIX TABLE IX (Con't.)

XTX

20	24822	104727	77787	9333	3604	4512	215	1689
24822	39874464	169922766	125325819	13936245	5607020	8087658	328024	3142280
104727	169922766	749436655	537034416	63592197	23797034	31988508	1512616	13507849
77787	125325819	537034416	402711324	45960266	17420748	24774183	1136769	11222522
9333	13936245	63592197	45960266	7778063	1706925	1258890	187505	1003464
3604	5607020	23797034	17420748	1706925	1160062	1140956	46696	233471
4512	8087658	31988508	24774183	1258890	1140956	3755445	14009	197807
215	328024	1512616	1136769	187505	46696.	14009	6179	35475
1689	3142280	13507849	11222522	1003464	233471	197807	35475	1238653

R

1.								
	1.							
		1.						
			1.					
				1.				
					1.			
						1.		
							1.	
								1.

VAR-COV MATRIX FOR SMALL R

.362	.01	.01	.01	.01	.01	.01	.01
.01	1.392	.01	.01	.01	.01	.01	.01
.01	.01	.642	.01	.01	.01	.01	.01
.01	.01	.01	401.802	.01	.01	.01	.01
.01	.01	.01	.01	1.570	.01	.01	.01
.01	.01	.01	.01	.01	13.381	.01	.01
.01	.01	.01	.01	.01	.01	1.392	.01
.01	.01	.01	.01	.01	.01	.01	1.392

APPENDIX TABLE IX (Con't.)

```
AREA   D
(8F10.0)
  SMALL B   K    S SQUARED                        N               YTY
     9      8 951596.40653388                    27 2562085720.0

                                 XTY
   193848 2547384321134699230 816520453   9427597  57197696  16597076   2296017
  5739963

                                SMALL R
   1.081     2.12      2.12     36.00      2.25      6.57      2.12      2.12

                                SMALL B
-325.21719   2.7211    1.9273   -.7500    20.3680    2.0999    3.0279    6.8112
  -6.5704

                                 XTX
       27     21364     85528     68942       858      5660      1500       164
      558
    21364  26259140 113015617  84526501   1012456   5952473   1705560    227969
   612816
    85528 113015617 508789083 361571967   4030218  25016154   6825876   1047062
  2563599
    68942  84526501 361571967 279665738   3568685  18753847   5967573    707407
  1998279
      858   1012456   4030218   3568685     62505    196774     95914      6826
    24718
     5660   5952473  25016154  18753847    196774   2074695    352845     48641
   163093
     1500   1705560   6825876   5967573     95914    352845    381755      5813
    19156
      164    227969   1047062    707407      6826     48641      5813      2736
     5431
      558    612816   2563599   1998279     24718    163093     19156      5431
    19637

                                  R
          1.
                    1.
                              1.
                                        1.
                                                  1.
                                                            1.
                                                                      1.

     1.
```

APPENDIX TABLE IX (Con't.)

VAR-COV MATRIX FOR SMALL R

.388	.01	.01	.01	.01	.01	.01	.01
.01	1.493	.01	.01	.01	.01	.01	.01
.01	.01	.691	.01	.01	.01	.01	.01
.01	.01	.01	431.019	.01	.01	.01	.01
.01	.01	.01	.01	1.685	.01	.01	.01
.01	.01	.01	.01	.01	14.357	.01	.01
.01	.01	.01	.01	.01	.01	1.493	.01
.01	.01	.01	.01	.01	.01	.01	1.493

AREA E

(8F10.0)

SMALL B	K	S SQUARED	N	YTY
9	8	2432095.4566133	16	6175328060.0

XTY

263884 72228680722237764901362607160 34106057 272369159 18372738 11523311
6284364

SMALL R

1.081	2.12	2.12	36.00	2.25	6.57	2.12	2.12

SMALL B

-22.416639 6.4317592 1.6111985-2.1037646 26.401909 .2542178 2.8314669-29.753834
19.642101

XTX

16	31430	96448	62104	1458	14771	922	460	314
31430	85369088	258049200	160113885	4028231	28512014	2238634	1325806	738176
96448	258049200	845303401	508204349	11871229	107408552	5725526	4373324	2372568
62104	160113885	508204349	319692630	7766765	71233614	4051898	2611418	1577527
1458	4028231	11871229	7766765	226676	1859723	111317	70883	39663
14771	28512014	107408552	71233614	1859723	135592159	480441	804897	261599
922	2238634	5725526	4051898	111317	480441	116937	18529	23084
460	1325806	4373324	2611418	70883	804897	18529	30050	11172
314	738176	2372568	1577527	39663	261599	23084	11172	9829

APPENDIX TABLE IX (Con't.)

R

	1.						
		1.					
			1.				
				1.			
					1.		
						1.	
							1.
1.							

VAR-COV MATRIX FOR SMALL R

.325	.01	.01	.01	.01	.01	.01	.01
.01	1.250	.01	.01	.01	.01	.01	.01
.01	.01	.500	.01	.01	.01	.01	.01
.01	.01	.01	360.886	.01	.01	.01	.01
.01	.01	.01	.01	1.409	.01	.01	.01
.01	.01	.01	.01	.01	12.020	.01	.01
.01	.01	.01	.01	.01	.01	1.250	.01
.01	.01	.01	.01	.01	.01	.01	1.250