1953

Statistical approach in planning production programs for interdependent activities

Madan Mohan Babbar

Iowa State College

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Economics Commons

Recommended Citation
Babbar, Madan Mohan, "Statistical approach in planning production programs for interdependent activities " (1953). Retrospective Theses and Dissertations. 12857.
https://lib.dr.iastate.edu/rtd/12857

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
STATISTICAL APPROACH IN PLANNING PRODUCTION PROGRAMS
FOR INTERDEPENDENT ACTIVITIES

by

Madan Mohan Babbar

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Economics

Approved:

Signature was redacted for privacy.

In Charge of Major Work /

Signature was redacted for privacy.

Head of Major Department /

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1953
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform DP11919
Copyright 2005 by ProQuest Information and Learning Company. All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>PART I THEORETICAL</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>II. PROGRAM PLANNING</strong></td>
<td>9</td>
</tr>
<tr>
<td>2.1 Interindustry analysis (basic model)</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Open model</td>
<td>15</td>
</tr>
<tr>
<td>2.3 Linear programming</td>
<td>19</td>
</tr>
<tr>
<td>2.4 Application at the firm level</td>
<td>22</td>
</tr>
<tr>
<td>2.5 The basic theorem</td>
<td>27</td>
</tr>
<tr>
<td>2.6 Model for the problem</td>
<td>31</td>
</tr>
<tr>
<td><strong>III. ANALYSIS BASED ON DISCREPANCIES</strong></td>
<td>33</td>
</tr>
<tr>
<td><strong>IV. PROBABILITY APPROACH</strong></td>
<td>39</td>
</tr>
<tr>
<td>4.1 Approximate reduction of the model</td>
<td>40</td>
</tr>
<tr>
<td>4.2 General procedures</td>
<td>46</td>
</tr>
<tr>
<td>4.3 Normal distributions</td>
<td>53</td>
</tr>
<tr>
<td>4.4 Confidence limits</td>
<td>56</td>
</tr>
<tr>
<td>4.5 Confidence limits for the results of programming</td>
<td>58</td>
</tr>
<tr>
<td>4.6 Cumulative distributions</td>
<td>62</td>
</tr>
<tr>
<td>4.7 Program efficiency control</td>
<td>65</td>
</tr>
<tr>
<td>4.8 A critical note</td>
<td>68</td>
</tr>
<tr>
<td><strong>V. SUMMARY</strong></td>
<td>70</td>
</tr>
<tr>
<td><strong>PART II APPLICATION TO FARM PRODUCTION PLANNING</strong></td>
<td>74</td>
</tr>
<tr>
<td><strong>VI. APPLICATION OF LINEAR PROGRAMMING</strong></td>
<td>75</td>
</tr>
<tr>
<td>6.1 Constant returns to scale</td>
<td>75</td>
</tr>
<tr>
<td>6.2 Uncertainty</td>
<td>79</td>
</tr>
<tr>
<td>6.3 Optimum production program</td>
<td>82</td>
</tr>
<tr>
<td><strong>VII. STATISTICAL PREDICTIONS ABOUT THE RESULTS OF THE PROGRAM</strong></td>
<td>102</td>
</tr>
<tr>
<td>7.1 A preliminary note</td>
<td>102</td>
</tr>
<tr>
<td>7.2 Examination of the input coefficients series</td>
<td>105</td>
</tr>
<tr>
<td>7.3 Preliminary analysis for predictions</td>
<td>107</td>
</tr>
<tr>
<td>7.4 Confidence limits for the activity levels</td>
<td>110</td>
</tr>
<tr>
<td>7.5 Confidence limits of the profit function</td>
<td>113</td>
</tr>
<tr>
<td>7.6 Statistical predictions based on the cumulative distributions</td>
<td>116</td>
</tr>
<tr>
<td>7.7 A critical note</td>
<td>129</td>
</tr>
<tr>
<td>VIII. SUMMARY</td>
<td>Page</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
</tr>
<tr>
<td>IX. BIBLIOGRAPHY</td>
<td>132</td>
</tr>
<tr>
<td>X. ACKNOWLEDGEMENTS</td>
<td>135</td>
</tr>
<tr>
<td>XI. APPENDIX</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td>142</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

The basic problem in production economics is the efficient allocation of resources, land, labor, capital, and management, with a view to achieve a predetermined end. This end may be a maximization of profit or a minimization of cost in the case of a firm or it may be maximization of the social welfare or national security and economic stability on the national level. An economic unit is concerned with choice and decision-making in the use of resources at its disposal. The problems involved in this role of choice-making are similar in nature whether these are faced by the management of a firm or the government of a country.

As a simple example of the decisions involved in production economics, we may consider a firm involved in the production of a single commodity using a single resource which is obtainable in competitive market. Also, any amount of the commodity produced can be sold at the competitive price. Then what amount of commodity should be produced? The production function computed to show the relations between the output and the corresponding amount of input and the prices of the input relative to that of the output will determine the amount of commodity to be produced. The amount of output for which the marginal productivity of the input is equal to the price
ratio of the resource and the output is the optimum amount to be produced by the firm. These kinds of decisions become complicated when the number of inputs used in the production of a commodity is more than one. Moreover, there may be alternative possibilities of production. Alternative resources could be used to produce the same commodity or the given amounts of resources could be used to produce alternative products. Some products may be complementary. The production of one accompanies the production of the other in varied proportions. The role of management is to make a selection of the outputs and the amounts to be produced. A farmer, for example, may grow corn, oats, or some other crop or a combination thereof, using the land and other resources he has. These various possibilities of production activities can be referred to as inter-dependent activities. On a national scale, the various productive activities in the form of industries, household and government services, international trade and so forth, are also inter-dependent. In fact, the output of any industry is an input for other sections of the economy. The planners of the national economy are interested in determining the levels of these activities which may yield the desired ends of the society. In war, for example, the government needs certain amounts of food, grains, war materials, and
consumer goods. The productive activities may be controlled so that after mutual interindustry consumption, the government may be able to procure the desired amounts for military use and also insure that the consumers will get planned amounts of goods. Even in peace time, a government in a position to regulate and control various national productive activities, is faced with a similar allocation problem.

We have used the word inter-dependent activities to cover production possibilities in the realm of activity of an economic unit which are related either as alternative consumers of resources as in the case of a firm or mutually related as inputs and outputs as national economy sectors as described above.

Two activities are inter-dependent when they must share limited amounts of a commodity which they use in common, when one produces a commodity which is used by the other, or when each produces a commodity used by a third activity.¹

Marginal analysis, the well-accepted criterion of choice-making in production economics, has been criticised recently by many economists. One reason is that it assumes

continuity of the production function. In simple, words, it means that it is supposed to be responsive to infinitesimally small changes in inputs. In practice, however, one hundredth fraction of an acre of land, added to the existing land of production may not show yield increase which is measurable. Similarly, the machinery used in production must be increased or decreased in whole units. Even if one hires machine services, it ordinarily cannot be hired on the basis of an infinitesimal unit. This practical consideration encouraged development of "linear models" in planning programs for inter-dependent activities.

Leontief's\(^1\) inter-industry analysis were probably the first linear models presented in a comprehensive form, particularly the open model which could be used for planning national productive activities. However, the mathematical model of linear programming with a view to maximize a linear function of the activity levels was presented by Dantzig\(^2\)\(^3\)\(^4\).


\(^3\)Dantzig, G. B. Maximization of a Linear Form Whose Variables are Subject to a System of Linear Inequalities, Mimeo., Washington, Headquarters, U. S. Air Force, Comptroller, 1949.

in an integrated form in 1948 and 1949. Since then, this field of programming has received contributions from various economists, mathematicians, and administrators alike; it has become an important tool in research. We will discuss these techniques in Chapter II assuming static situation.

The Problem

The problem we propose to take up is as follows:

Suppose a schedule of production is planned by an economic unit using linear coefficients of production. What statistical procedures can be formulated to predict the outcome of the plan? This is of particular interest in a situation in which the linear coefficients of production are liable to vary. These technological relations which will be referred to as the input coefficients are in the form of a particular input required to produce a unit of particular output. In agricultural production, for instance, it is very difficult to predict how much land will be used to produce a bushel of corn, due to weather variability, and so forth. Hence, we will attach a measure of variability to the input coefficients which we use in planning the production schedule and try to gauge the variability of the activity levels as they will be realized in production. Although we will discuss
discrepancies analysis in Chapter III, our main approach to solve the problem will be based on the concept of probability distributions. Assuming that the input coefficients are liable to vary according to normal distributions about the expected values which we used for planning the program and with corresponding assumed variances, our purpose will be to derive approximate probability distributions of the activity levels (which will be realized). An effort will be made to give simplified procedures by which statistical predictions, for example, to find the probability that the level of a particular activity would not fall below a certain limit, could be made without cumbersome computation of the distributions of the activities. We propose to accomplish this in Part I of the dissertation. Some general procedures, which can handle cases wherein the distributions of the input coefficients may be assumed to be other than normal, will be added in Chapter IV.

To exhibit the working of this statistical approach, an empirical example is included in Part II of the dissertation. A model family farm in Iowa with given amount of capital, land, and labor is considered. First, an optimum program of production is derived, and then, using the analysis of Part I, probability statements about the results of the program are made. A critical examination
of the results is included in the end of Part II.

It may be observed that, in Chapter II and in the empirical example, we have restricted ourselves to particular situations of production economics. The mathematical formulae derived in Chapter IV are quite general in nature. Those results may be applicable to many other situations of applied sciences wherein the mathematical model conforms with the one used.
PART I

THEORETICAL

Derivation of the mathematical model for the problem and its theoretical analysis.
II. PROGRAM PLANNING

Program planning or programming may be defined as the construction of a schedule of actions by means of which an economy, organization, or other complex of activities may move from one defined state to another or from a defined state to some specifically defined objective. The methodology refers to recently developed, associated techniques which have been called by various names in the literature: input-output analysis, programming of interdependent activities, or linear programming. These techniques may be used on a national level or on less aggregative levels of an organization or a firm. Moreover, as is apparent from the definition, the formulation of the program may be a simple analysis of inputs and outputs, for example, Leontief's inter-industry models, or a more general formulation of allocation of scarce resources with a view to maximize an "objective function." It may be proper to call the latter formulation "optimal programming," but we will use the word programming or program planning to describe the whole area of these techniques.

---

Recently, these techniques have been used in determining optimum feed mixtures, optimum storage shipment, and machine productivities. Various United States Government agencies have given a major impetus to the development of these techniques, particularly the United States Air Force and the Bureaus of the Budget and of Labor Statistics. Economy-wide mobilization and employment programs are some of the problems with which these agencies are mostly concerned. Most of the work done in these applications, as well as in the theoretical development of the techniques, is listed in the Bibliography.

In this chapter we will present, rather briefly, Leontief's "open model" and the theory of linear programming as applied to production economics, particularly at the level of the firm. These techniques are directly relevant to planning production in an economy or in a firm. The purpose of this presentation is to give a simple mathematical background of our problem, as enunciated in Chapter I, and to derive a specific mathematical model to solve it.

2.1 Interindustry analysis (basic model).

In Leontief's input-output analysis, the economy of a nation is considered as a combination of a large number of inter-dependent activities of production, transportation,
distribution, consumption, and so forth.

This partition of the whole national activity into a finite number of sectors is based on grouping the similar productive, distributive, and consuming agencies. These sectors or groups are referred to as industries or activities. Each one of these activities consumes products of some or all other activities as its input and contributes its output to other activities as their inputs. Household is considered as one of these activities, its inputs being the commodities consumed by households and its output being the services of households used in other activities. Similarly, international trade may be considered as another sector with exports as inputs and imports as outputs.

The mathematical model of the input-output relationships among these activities may be set up as below.

Suppose $X_1 \cdots X_n$ represent the outputs of n sectors and $x_{ij}$ represents the amount of output $X_j$ absorbed in the $i$th industry.

Then the distribution of inputs and outputs may be described by the following equations:

\[ -X_1 + x_{11} + x_{12} + \cdots + x_{1n} = 0 \]
\[ x_{1n} - X_2 + x_{22} + \cdots + x_{2n} = 0 \] (2.1.1)
\[ x_{1n} + x_{2n} + x_{3n} + \cdots - X_n = 0 \]
These equations state that the total output of each industry (measured in physical terms) equals the sum total of the amount of its product consumed by other industries.

Another set of equations describes the technical relationships between the output and the input of an industry. Instead of a general type of production function,

\[ x_i = f(x_{i1}, x_{i2}, \ldots, x_{in}) , \]

it is assumed that each input is proportional to the quantity of products, for instance,

\[ x_{ik} = a_{ik}x_i \quad i=1,2,\ldots,n \quad i \neq k \quad (2.1.2) \]

where the constants of proportionality \( a_{ik} \) (called the input coefficients) are assumed to be fixed and known.

The input coefficients can be empirically determined by constructing a square table with number of rows and columns equal to the number of industrial sectors, and by entering in it the physical inputs and outputs of all the sectors during some past period. Since everybody's output is somebody's input, each item in this table will be a double entry. Figures in any row will add to the total output of the corresponding industry, whereas the figures in any column will show the inputs consumed in the industry.
represented by that column. From this table showing the
distribution of the inputs and the outputs of a past period,
we can estimate the coefficients of production or the input
coefficients by dividing any item of this table by the
corresponding row total. For example, let us suppose $T_{ij}$
is the entry in that table in the $i$th row and the $j$th
column and $T_i$ is the row total of the $i$th row. $T_i$
represents the total output of the $i$th activity during
that period, and $T_{ij}$ represents the part of that output
of the $i$th industry which was consumed by $j$th industry.
Then $T_{ij}/T_i$ will give an estimate of $a_{ij}$, the input co­
efficient, that is, the amount of the output of the $i$th
industry required in producing one unit of the output of
the $j$th industry.

An assumption of this type (2.1.2) of production
function means a rejection of marginal productivity theory.
The output will not increase by increasing a single input
unless corresponding increases in all other inputs according
to their production coefficients are also effected.
Technical substitutability is not possible. This form of
production function is the basis of the linear programming
approach.

The above system is described on the assumption of a
stationary state, but the equations could be modified by
introduction of savings coefficients to include the
consideration of saving and investment. Mathematically, the modified system remains similar to the above (with necessary changes in the coefficients).

To find the solution of the levels of activities in this model, we substitute equations (2.1.2) in equations (2.2.1) to eliminate the small $x$'s. We get

\[-X_1 + a_{11}X_2 + a_{12}X_3 + \cdots + a_{1n}X_n = 0\]

\[a_{12}X_1 - X_2 + a_{22}X_3 + \cdots + a_{2n}X_n = 0\]  \hspace{1cm} (2.1.3)

\[a_{1n}X_1 + a_{2n}X_2 + a_{3n}X_3 + \cdots - X_n = 0\]

Now, these are $n$ homogeneous linear equations in $n$ variables $X_1, \ldots, X_n$, and they will be consistent if, and only if, the determinant of the coefficients vanishes. In that case, we can have a non-trivial solution. Suppose this condition is satisfied. Then we can omit any one of the equations, say the $n$th one, and rewrite the equations in the following form:

\[-\frac{X_1}{X_n} + a_{21} \frac{X_2}{X_n} + a_{31} \frac{X_3}{X_n} + \cdots + a_{n-1} \frac{X_{n-1}}{X_n} = -a_{n1}\]

\[a_{12} \frac{X_1}{X_n} - \frac{X_2}{X_n} + a_{32} \frac{X_3}{X_n} + \cdots + a_{n-1} \frac{X_{n-1}}{X_n} = -a_{n2}\]  \hspace{1cm} (2.1.4)

\[^1\text{Leontief, W. W. The Structure of American Economy, New York, Oxford Univ. Press, 1951.}\]
\[ a_{1n} \frac{X_1}{X_n} + a_{2n} \frac{X_2}{X_n} + \ldots + a_{n-1n} \frac{X_{n-1}}{X_n} = 1 \]

which gives a unique solution for quantities \( \frac{X_1}{X_n}, \frac{X_2}{X_n}, \ldots, \frac{X_{n-1}}{X_n} \).

We can write equations (2.1.4) in matrix form as

\[ BX = Q \quad (2.1.4') \]

and the solution of relative activity levels is given by this matrix equation. For detailed discussion of the closed model, one may refer to Leontief's work quoted above.

2.2 Open model.

In Section 2.2, linear technological coefficients \( a_{ij} \) were used to replace the classical production function. It was also suggested how these coefficients could be determined empirically using some known data. It may be noted that the same coefficients could be determined by consulting technical experts. Anyway, if we know these technical coefficients, we can reconstruct the whole set of entries in the input-output table from them, if the output of any one or more sectors is considered exogeneous, that is, determined outside the model. This approach becomes important if we are interested in planning an economic system. The National Economy, for instance, may be considered as an open system for many purposes, such as
evaluation of alternative policies in respect to allocation of primary resources. "A closed system becomes open as soon as one disregards (that is considers as being pliable at will or even entirely unknown) one or more of the basic structural relationships of which it is made."¹

Suppose we regard the output of households and government agencies as well as their purchases as exogenous. In other words, we may have a predetermined list of goods and services which the government plans to consume in a specified period and also a list of the goods and services which are expected to be utilized by the households during that period. Such a plan, for example, may be necessary for an emergency period, such as war. The list of demands is referred to as "bill of goods" in the literature. The problem is how the other national activities may be regulated to yield that "bill of goods" during the relevant period. We will now describe the mathematical model for this system.

As in the closed system, let $X_1, X_2, \ldots, X_m$ denote total outputs of industrial sectors which contribute the quantities $q_1, q_2, \ldots, q_m$ in the final "bill of goods" and

if $x_{ij}$ represents physical amount of output $X_j$ going into the $i$th industry as input, then we have:

$$
X_1 - x_{a1} - x_{s1} - \cdots - x_{m1} = q_1
$$

$$
-x_{12} + x_2 - x_{s2} - \cdots - x_{m2} = q_2
$$

$$
-x_{1m} - x_{sm} - x_{sm} - \cdots + x_m = q_m
$$

(2.2.1)

If $a_{ij}$, as before, represents the technical coefficients, the above system can be written as

$$
X_1 - a_{s1}X_2 - a_{s1}X_3 - \cdots - a_{m1}X_m = q_1
$$

$$
-a_{12}X_1 + x_2 - a_{s2}X_3 - \cdots - a_{m2}X_m = q_2
$$

$$
-a_{1m}X_1 - a_{sm}X_2 - a_{sm}X_3 - \cdots + X_m = q_m
$$

(2.2.2)

or

$$(I - A)X = Q$$

(2.2.2)

where $I$ is the identity matrix of order $m$ and

$$A = \text{matrix } (a_{ij})$$

(2.2.3)

where $a_{ij} = 0$ when $i = j$.

Let us write the matrix equation (2.2.2) as $BX = Q$.

Thus, given a final bill of goods and the input coefficients, in other words, given vector $Q$ and the matrix $B$, we get a unique solution for levels of production in the
industrial sectors by solving the above \( m \) linear equations if \( B \) is non-singular. We have in that case

\[
X = B^{-1} Q .
\]  \hspace{1cm} (2.2.4)

It may be noticed that neither the bill of goods nor the amount of services of the households and the government is determined by the system. If we are considering household and the government as one sector of the economy, its contribution to various other industries can be determined from the solution vector \( X \), if the corresponding input coefficients were known. Also, if only the household services and their demands are considered to be exogenous in the open model described above, then the total employment may be computed as follows. Assuming linear coefficients for labor inputs, that is

\[
x_{iq} = a_{iq}X_i \quad \text{for } i = 1, 2, \ldots, m
\]

where \( a_{iq} \) represents labor input (and government services) required in a unit output of the \( i \)th industry, we find that the total employment is

\[
a_{iq}X_i + a_{aq}X_a + \cdots + a_{mq}X_m,
\]

a linear function of \( X \)'s. For detailed discussion of the
assumptions and criticism, one may refer to Georgescu-Roegen and Goldsmith's articles.¹,²

2.3 Linear programming.

Mathematically, the general formulation of linear programming is to find the maximum of a linear function say

\[(C_1x_1 + C_2x_2 + \cdots + C_nx_n)\] (2.3.1)

subject to conditions

(1) \(x_j \geq 0\) for \(j = 1, 2, \cdots, n\) (2.3.2)

and

(11) \(\sum_{j=1}^{n} a_{ij}x_j \leq q_i, 1 = 1, 2, \cdots, m\) (2.3.3)

where

\(C_1, C_2, \cdots, C_n\) and \(a_{ij}\)'s are known constants.

¹Georgescu-Roegen, N. Leontief's System in the Light of Recent Results, Rev. of Econ. and Stat., 32:214-222, 1950.

In matrix notation, we could say that $C^\text{T}X$ is to be maximized subject to conditions:

1. $X \geq 0$ \hspace{1cm} (2.3.2')

and

2. $A \cdot X \leq Q$. \hspace{1cm} (2.3.3')

This model applied to production economics has the following interpretation.

The vector $X$ may represent the levels of $n$ alternative productive activities of an economy, organization or a firm. Elements $(q_i)$ of $m$ column vector $Q$ represent the $m$ fixed resources required in the production. Elements of the $(m \times n)$ matrix $A$, represent the input coefficients. Each of the $m$ inequalities (2.3.3) represents the fact that all productive activities can be carried on to the extent they do not require more than the given quantity of a resource. Therefore, the set (2.3.3) of the inequalities represent the resource limitations on the production capacities of the economic unit. $C^\text{T}X$ represents a linear objective function which the economic unit is interested to maximize. For example, if element $C_j$ of column vector $C$ represents the net value of one unit of output of the $j$th activity, then $C^\text{T}X$ will represent the net value of the whole production program. The problem is, to determine the levels of
activities, that is $X_1, X_2, \ldots, X_n$ which can be carried on with the given supplies of resources so as to maximize the objective function $C'X$. The following are three basic assumptions of this economic model.

1. The production opportunities of an economy or an economic unit are defined by its resources and a finite number of productive activities available to it. The quantities of at least some of the resources are limited.

2. Any productive activity may be carried on at any positive level consistent with the limitations on resources available. The consumption of resources and the output of products are proportional to the level at which the activity is carried on.

3. Several productive activities may be used simultaneously within the limits of available resources. Consumption of each resource is the sum of the consumptions of that resource in the activities carried on, and if the same product could possibly be produced by different activities, then the total output of that product is the sum of the outputs of the individual activities producing that product.

These assumptions and the discussion in the following
section are mainly based on R. Dorfman's\textsuperscript{1} work.

2.4 Application at the firm level.

We present the problem of production program planning on the level of a firm, assuming perfect competition, in an elaborate manner. A productive event may be defined by a column vector

\[ E_i = (a_{i1}, a_{i2}, \ldots, a_{im}, b_{i1}, b_{i2}, \ldots, b_{ip}) \]  \hspace{1cm} (2.4.1)

where a's refer to the amounts of inputs and b's the amounts of outputs. Another such event, say

\[ E_k = (a_{k1}, \ldots, a_{km}, b_{k1}, \ldots, b_{kp}) , \]

will be an event of the same productive activity as that of \( E_i \) if their corresponding elements are proportional, that is, if

\[ \frac{a_{kl}}{a_{i1}} = \frac{a_{k2}}{a_{i2}} = \ldots = \frac{b_{kl}}{b_{i1}} = \ldots = \frac{b_{kp}}{b_{ip}} = \lambda, \text{ say.} \]  \hspace{1cm} (2.4.2)

If event \( E_i \) defines a unit level of that activity, then the level of \( E_k \) will be \( \lambda \).

The corresponding productive activity, in fact, is characterized by \( E_i \), and there are \( n \) such alternative

\textsuperscript{1}Dorfman, R. Application of Linear Programming to the Theory of the Firm, Berkeley, Univ. of Calif., 1951.
production activities. Let the column vector \( X, (X_1, X_2, \ldots, X_n) \) denote levels of the \( n \) activities denoted by \( E_1, E_2, \ldots, E_n \). Thus, two column vectors \( (X_1, X_2, \ldots, X_n) \) and \( (E_1, E_2, \ldots, E_n) \) specify a production program completely. If \( a_{ij} \) denotes the amount of the \( j \)th scarce resource used in the \( i \)th activity and if the vector \( Q = (q_1, q_2, \ldots, q_m) \) represents the supply of these factors, then the inequalities

\[
A^T \mathbf{X} \leq Q
\]

represent the resource restrictions in the production program.

The next question is the evaluation of the production program. Society or any other economic unit concerned with the program will make its judgment on the basis of the inputs of resources consumed and the outputs of the products produced which are functions of the intensities of the production activities. Therefore, the measure of desirability is a function of vector \( X \), say \( f(X) \). More specifically, we assume it to be a linear function of the activity levels, say \( C^T \mathbf{X} \). Thus, the basic problem of linear programming is to maximize \( C^T \mathbf{X} \) (2.4.4) subject to the conditions:

\[
(1) \mathbf{X} \geq 0
\]

and

\[
(11) A \mathbf{X} \leq Q
\]
To solve this problem, Dr. G. B. Dantzig suggested a device of introducing formally disposal activities equal to the number of limited resources. These disposal activities do not involve costs and profits. That is, the unused resources are wasted and are disposed of without cost. By introducing \( n \) such activities, we can reduce the resource inequalities to equalities; for example, for the \( j \)th resource, we have

\[
s_1 x_1 + s_2 x_2 + \cdots + s_m x_n + x_{n+j} = 0
\]

for \( j = 1, 2, \ldots, m \) where \( x_{n+1}, x_{n+2}, \ldots, x_{n+m} \) represent the levels at which resources \( 1, 2, \ldots, m \) are disposed respectively. If

\[
\begin{pmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & 0 & \cdots & 0 \\
    a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & 0 & \cdots & 0 \\
    a_{31} & a_{32} & \cdots & a_{3n} & 0 & 0 & 1 & \cdots & 0 \\
    a_{m1} & a_{m2} & \cdots & a_{mn} & 0 & 0 & 0 & \cdots & 1 \\
\end{pmatrix}
\]

(2.4.7)

then the resource restrictive conditions may be written as:

\[ \begin{bmatrix} B & Y \end{bmatrix} \leq \begin{bmatrix} Q \end{bmatrix} \quad (2.4.8) \]

where

\[ Y = \text{vector} \left( X_1, X_2, \ldots, X_n, X_{n+1}, \ldots, X_{n+m} \right) \]

The linear objective function \( C^t X \) can also be modified. If \( C_j \) represents the net value (or net profit) or a unit level of the \( j \)th activity, we could re-define the unit levels of all the productive activities such that the new unit level of each activity has net value unity (one dollar if it is profit). Therefore, the objective function can be written as

\[ (X_1 + X_2 + X_3 + \ldots + X_n + OX_{n+1} + OX_{n+2} + \ldots + OX_{n+m}) \quad (2.4.9) \]

or \( V^t Y \) where \( V \) is the column vector

\[ \left( \frac{1, 1, \ldots, 1}{n \text{ times}} ; \frac{0, 0, \ldots, 0}{m \text{ times}} \right) \quad (2.4.10) \]

Thus the problem is to maximize \( V^t Y \) subject to the restrictions

\[ (1) \ Y \leq 0 \quad (2.4.11) \]

\[ (11) \ BY = Q. \quad (2.4.12) \]
It may be noticed that \( BY = Q \) represent in linear equations in \( m+n \) variables (activity levels) and, in general, will have many solutions.

**Definition:** A solution of these equations in which no more than \( m \) elements of \( Y \) enter with positive values (others being zeroes) is called a basic solution. The set of activities who have these positive levels is called the basis of that solution. For example, the set of the \( m \) disposal levels equal to the total supplies of the fixed resources gives a basic solution. That is,

\[
Y = (0, 0, \ldots, 0; q_1, q_2, \ldots, q_m) \tag{2.4.13}
\]
satisfies the requirements (2.4.11) and (2.4.12). To check (2.4.12), we may write (2.4.13) in vector notation as

\[
Y = \begin{pmatrix} 0 \\ q \end{pmatrix} \tag{2.4.14}
\]
by substitution, we get

\[
BY = (A|I)\begin{pmatrix} 0 \\ q \end{pmatrix} = IQ = Q . \tag{2.4.15}
\]

Therefore, \( Y = \begin{pmatrix} 0 \\ q \end{pmatrix} \) is a solution of \( BY = Q \), and it is basic because only \( m \) activities (in this case all disposal) have positive levels. However, the objective function \( V'Y \) will have value zero in this case, since no productive activity is carried out.
2.5 The basic theorem.

Robert Dorfman, in his book referred to earlier, has proved an important theorem making three additional assumptions, ultimately reducible to:

1. The rank of the matrix $B$ is $m$

2. The rank of the matrix $\begin{pmatrix} Y' \\ B \end{pmatrix}$ is $m + 1$

3. The vector $Q$ is linearly independent of every set of $m - 1$ column vectors of the matrix $B$.

(By the third assumption, a basic solution will involve precisely $m$ elements of $Y$ with positive values, otherwise $Q$ would be linearly dependent on fewer vectors of $B$ than $m$ in number.)

Conclusion: The solution which maximizes the linear function $V'Y$ is a basic solution. The proof of this theorem is given below.

Suppose $Y \geq 0$ is a solution such that $V'Y$ is maximum and $BY = Q$.

Suppose out of $m+n$ elements of vector $Y$, $p$ are positive (where $p > m$). We relabel $m$ activities in such a manner that the first $p$ elements of $Y$ are positive, and we rewrite $B$ accordingly. Then we partition matrix $B$ and vector $Y$ as follows,

$$B = \begin{pmatrix} B_1 & B_2 & B_3 \\ (m \cdot n+m) & (m \cdot m) & (m \cdot p-m) & (m \cdot n+m-p) \end{pmatrix}$$

(2.5.1)
and
\[
\begin{bmatrix}
Y_1' & Y_2' & Y_3' \\
(1 \cdot m) & (1 \cdot p-m) & (1 \cdot m+n-p)
\end{bmatrix}
\]
(2.5.2)

Then by hypothesis,
\[
(B_1|B_2|B_3)egin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = (B_1|B_2)egin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = Q
\]
(2.5.3)
or
\[
B_1Y_1 + B_2Y_2 = Q.
\]
(2.5.4)

By the first assumption of Dorfman, \(B_1\) is a non-
singular matrix. Therefore, we have

\[
Y_1 = B_1^{-1}Q - B_1^{-1}B_2Y_2 \\
(m \cdot 1) (m \cdot m) (m \cdot 1) (m \cdot m) (m \cdot p-m) (p-m \cdot 1)
\]

\[
= \alpha \cdot Y_2 \\
(m \cdot 1) (m \cdot p-m) (p-m \cdot 1)
\]
(2.5.5)

We partition \(V'Y\) the same way.

\[
V'Y = V_1' Y_1 + V_2' Y_2 \\
(1 \cdot m) (m \cdot 1) (1 \cdot p-m) (p-m \cdot 1)
\]
(2.5.6)
Substituting for \(Y_1\) from (2.5.5), we have

\[
V'Y = V_1' (\alpha \cdot Y_2) + V_2' Y_2
\]

\[
= V_1' \alpha + (V_2' - V_1' \beta)Y_2.
\]
We can show that $V'Y$ is maximum if, and only if, $Y_a = 0$. If $(V'_a - V'_1\beta) \neq 0$, we can vary the elements of $Y_a$ up and down a bit (without making any of them zero, and get a solution which gives a value of $V'Y$ greater than the original.

On the other hand, if

$$V'_a - V'_1\beta = 0 \quad \text{(that is, if } V'_a = V'_1\beta) ,$$

$$\begin{pmatrix} V'_1 \\ V'_a \end{pmatrix} \begin{pmatrix} V'_1 \\ V'_a \end{pmatrix}$$

will have rank $m$, since

$$\beta = B_1 B_a, \quad (m+1) \times p \text{ matrix} .$$

But this contradicts the second assumption of Dorfman.

Therefore, $Y_a = 0$ and the solution must be basic. Also, by the third assumption, a basic solution will have $m$ and only $m$ activities (productive or disposal) with positive levels. Therefore, the schedule of levels which maximizes the linear objective function, will involve $m$ and only $m$ activities.

It may be noticed that for simplicity of proof, unit activity levels were defined so as to give unit profit or net values, yet this is not necessary. That is, going back
to the original units we can say that \( C'Y \) will be maximized by the corresponding solution involving the same basis, only measured in the original units.

The simplex criterion and the practical method for computing an optimum solution is very well presented by Charnes, Cooper and Henderson.\(^1\) We will use the same while dealing with an empirical example. Now suppose we relabel the activities so that the first \( m \) activities correspond to those which enter in the optimum solution with positive levels; and accordingly, readjust the columns of matrix \( B \).

That is, \( Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \) where \( Y_1 \) is the \( m \)-column vector representing the optimum solution, \( Y_2 \) is the \( n \)-column vector (all zeroes), and \( B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \) with the corresponding arrangement.

The optimum solution can be written (after obtaining it) as

\[
Y_1 = B_1^{-1} Q
\]

where \( Q \) is an \( m \) vector representing the factor supplies.

---

\(^1\)Charnes, A., Cooper, W. W., and Henderson, A. An Introduction to Linear Programming, New York, John Wiley and Sons, Inc., 1952.
2.6 Model for the Problem.

It may be noticed in the preceding sections that a non-degenerate solution of a programming problem obtained by the use of linear models occurs as a solution of a set of linear equations equal in number to the variables $X$. That is,

$$BX = Q.$$  \hspace{1cm} (2.6.1)

The matrix $B(m \times m)$ consists of elements representing known input-output relations and $Q$ is a $m$-column vector with known elements. The plan of production is drawn, using these constant quantities. At the time of operation, however, they may vary a little, with the result that the activity levels realized will be different from those planned. We may write the model at operation as

$$(B + b)X = (Q + \varepsilon)$$ \hspace{1cm} (2.6.2)

where $b$ is a matrix and $\varepsilon$ a column vector of errors corresponding to the elements of $B$ and $Q$ respectively.

Our problem can thus be defined in terms of this model. If the elements of $b$ and $\varepsilon$ are assumed to be random variables, can the statistical distributions for the activity levels to be realized be derived? These activity levels (realized, and not the planned values) are variables
whose distributions depend on those of the errors involved, and also on the inter-relation given by model (2.6.2). We are also interested in deriving distribution of a linear function of the activity levels

\[(C + c)'X\] (2.6.3)

where \(c\) is a \(m\)-column vector of errors in the corresponding elements of \(C\), (known constants). Obviously, very useful statistical predictions can be made from these distributions. We will attempt to solve this problem in Chapter IV.
III. ANALYSIS BASED ON DISCREPANCIES

It has been shown in Chapter II that the activity levels planned for production by the use of the programming techniques occur as a solution to a set of equations:\(^1\)

\[ ax = f \]  
\[(3.1.1)\]

where \(\vec{a}\) is a matrix of the input coefficients and \(f\) is the vector of fixed constants. In practice, it is quite likely that at the time, when the program is put in operation, the operating values of the input coefficients may be different from those planned. Suppose, at the time of operation of the program, we have

\[ [a + \varepsilon(a)] [x + \varepsilon(x)] = [f + \varepsilon(f)] \]  
\[(3.1.2)\]

elements of \(\varepsilon(a), \varepsilon(x)\) and \(\varepsilon(f)\) will be called discrepancies between the corresponding values at the time of planning and the values realized. The problem is to estimate approximately \(\varepsilon(x)\), if \(\varepsilon(a)\) and \(\varepsilon(f)\) are known.

Paul S. Dwyer has dealt with this problem in his book "Linear Computations."\(^2\) We will briefly describe his

---

\(^1\)We have preferred to use different notation in this chapter to distinguish the analysis of this chapter from the one relevant to our problem.

\(^2\)Dwyer, Paul S. Linear Computations, New York, John Wiley and Sons, Inc., 261-284, 1951.
procedure below. First of all, we assume that $a$ and $a + \varepsilon(a)$ are both non-singular so that the equations (3.1.1) and (3.1.2) have unique solutions. Now expanding (3.1.2) we get

$$ax + a\varepsilon(x) + \varepsilon(a)x + \varepsilon(a)\varepsilon(x) = f + \varepsilon(f). \quad (3.1.3)$$

Subtracting (3.2.1) from it we get

$$a\varepsilon(x) + \varepsilon(a)x + \varepsilon(a)\varepsilon(x) = \varepsilon(f) \quad (3.1.4)$$

limiting the analysis to first order errors, we neglect $\varepsilon(a)\varepsilon(x)$ and have

$$a\varepsilon(x) + \varepsilon(a)x = \varepsilon(f) \quad (3.1.5)$$

or

$$a\varepsilon(x) = \varepsilon(f) - \varepsilon(a)x. \quad (3.1.6)$$

Therefore

$$\varepsilon(x) = a^{-1} \left[ \varepsilon(f) - \varepsilon(a)x \right] \quad (3.1.7)$$

or

$$\varepsilon(x) = a^{-1} \left[ \varepsilon(f) - \varepsilon(a)a^{-1}f \right] \quad (3.1.8)$$

$$= c \left[ \varepsilon(f) - \varepsilon(a)c \right] \quad (3.1.9)$$

where

$$c = a^{-1}. \quad (3.1.9)$$

This formula gives the first order approximation to $x$ when $a$, $\varepsilon(a)$, $f$ and $\varepsilon(f)$ are specified. It is the basic
In practical cases, however, values of the discrepancies \( \varepsilon(a) \) and \( \varepsilon(f) \) are generally unknown but the maximum extent of each discrepancy may be known. These may be called bounds of the discrepancies. These are taken to be positive but may be added to or subtracted from the elements of \( a \) and \( f \). Denoting these bounds by \( \eta(a) \) and \( \eta(f) \), we can get the bound for discrepancies in \( x \), that is, \( \eta(x) \) from (3.1.9), as below

\[
\eta(x) = |c| \left[ \eta(f) + (a) \right] \quad (3.1.10)
\]

where \(|c|\) and \(|x|\) denote the matrices of the absolute values of the elements of the matrices \( c \) and \( x \).

Further, if \( \eta \) is the maximum element of \( \eta(f) \) and \( \eta(a) \), then

\[
\eta(x) = \eta |c| \left[ (1_f) + (1_a) \right] \quad |x|
\]

where \((1_f)\) and \((1_a)\) are matrices of unit elements with numbers of columns and rows corresponding to \( f \) and \( a \).

The same result (3.1.10) can be simplified for discrepancy of an individual element of \( x \) to the following form.

\[
\eta(x_k) = \frac{\eta (1 + \sum |x_k|)}{\Delta} \quad (3.1.11)
\]

\( ^1 \)Ibid.
where $\Delta$ is the determinant of $A$ and the $A_{kj}$ the co-factor of the corresponding element of $\Delta$ in the $k$th row and the $j$th column; and $\gamma$ is the common bound of elements of $\epsilon(a)$ and $\epsilon(f)$.

Further, Dwyer\(^1\) deals with some special cases when some and not all the coefficients of the linear equations considered, are subject to discrepancies. We will not go into these details.

In a recently published paper\(^2\), P. S. Dwyer and F. V. Waugh have dealt with the effects of the discrepancies in the elements of a matrix, on the elements of its inverse. It is a very detailed and extensive presentation but since we are not interested in the discrepancies of the elements of the inverse matrix as such, we will not present any part of the discussion here. To measure the discrepancies in the solution vector, we need to consider linear functions of the elements of an inverse matrix, and for that purpose, the results of that paper need the necessary extension so as to be applicable in this case. We do not propose to derive that extension. It will appear that a linear function of a large number of variables, whose maximum discrepancy may be known, might have a very big

\(^{1}\)Ibid.

discrepancy if the discrepancies in the variables are cumulative or maybe too small if they are mutually cancelling.

One needs the discrepancies in the constants to apply this analysis. In the case of the programming problem, the discrepancies in the input coefficients need to be determined. While dealing with Leontief's input-output analysis in a static situation, one needs data concerning the portions of the outputs of one industry going as inputs into the other industries. Oskar Morgenstern\(^1\) has presented estimates of discrepancies in the reporting of this kind of data. He has dealt with the sources of errors in economic data and has derived the discrepancies in the data concerning foreign trade, mining, agriculture, national income, employment and prices. One important omission is that of discrepancies in the input-output data concerning manufacturing industries. However, using these discrepancies, one could apply the analysis of Dwyer and Waugh to estimate approximately the discrepancies in the variables \(X\).

This analysis, it seems to us, will not be very useful, if one is interested in planning production for some future date. The problem in that case becomes that of

prediction. It becomes difficult to determine the discrepancies between the input coefficients used in planning and the values which will be realised. It is particularly so in those fields of production which involve risk and uncertainty. It seems to us that the probability approach discussed in the next chapter will be more appropriate in such cases.
IV. PROBABILITY APPROACH

The input coefficients which are needed to draw a production program are, in general, based on future expectations. The planning unit has to decide on the expected values of these constants. It may be done with the help of past experience or investigating technological relationships with the help of new experiments. These input coefficients are used in planning and are represented by the elements of matrix $B$, in the model

$$BX = Q.$$  

$Q$ is a vector of constants and $X$ represents the levels of activities. We now assume that at the time of operation of the program, the elements of $B$ and $Q$ are subject to prediction errors. It is assumed that these errors are random. Our aim is to derive approximate probability distributions of the activity levels if the distributions of the errors are known. We will be interested in making predictions as to the variability of the outcome of the plan. We will deal with the problem in a pure mathematical form so that the results can be used in any other applied field, too. In the latter sections of this chapter, we will come back to linear programming.
4.1 Approximate reduction of the model.

The mathematical structure of the following discussion is based on an unpublished paper by Dr. Gerhard Tintner.¹

A. Preliminary reduction

We consider m linear equations in m variables

\[(B + b)x = (q + \varepsilon) \quad (4.1.1)\]

where \(B\) is \(m \times m\) non-singular matrix \((B_{ij})\) whose elements are known constants and \(Q\) is an \(m\)-column vector \((q_i)\) whose elements are also known. These are the constants used in planning the program.

\(b\) is an \(m \times m\) matrix \((b_{ij})\) of random errors whose elements have known probability distributions, such that

\[E(b_{ij}) = 0 \quad \text{for} \quad i = 1, 2, \ldots, m\]

and

\[E(b_{ij}^2) = \sigma_{ij}^2 \quad \text{for} \quad j = 1, 2, \ldots, m.\]

Also \(\varepsilon\) is an \(m\) column vector of errors \((\varepsilon_i)\) whose elements have known probability distributions such that

---

¹Tintner, Gerhard. The Distribution of Solutions of Linear Equations whose Coefficients are Subject to Error, unpublished, Department of Economics, Iowa State College.

²Notation: \(E\) represents the mathematical expectation of the corresponding function.
Therefore, the equations under consideration are

\[
\sum_{j=1}^{m} (B_{ij} + b_{ij}) x_j = q_i + \epsilon_i
\]

for \( i = 1, 2, \cdots, m \).

We notice that the distributions of the coefficients of the variables \( x_1, x_2, \cdots, x_m \) as well as those of the quantities on the right hand sides of the equations are known. For example, the distribution of \((B_{ij} + b_{ij})\) will be the same as that of \(b_{ij}\) except that the mean value will be \(B_{ij}\) instead of zero since

\[
E(B_{ij} + b_{ij}) = B_{ij}
\]

and the variance of

\[
(B_{ij} + b_{ij}) = \sigma_{ij}^2.
\]

Our aim is to derive approximate distributions for the variables \( x_1, x_2, \cdots, x_m \) which occur in the model as a solution of the \( m \) linear equations \((4.1.1')\).
Let

\[ |B| \] denote the determinant of the matrix \((B_{ij})\).

\[ |B+b| \] denote the determinant of the matrix \((B_{ij} + b_{ij})\).

\(B_{ij}\) denote the co-factor of the element \(B_{ij}\) in \(|B|\).

\[ |D^k| \] denote the determinant of matrix \((B_{ij})\) when its

kth column is replaced by the column \((q_j)\).

\(D_{ij}^k\) denote the co-factor of the element in ith row

and jth column of \(|D^k|\).

\[ |D^k + d^k| \] denote the determinant of the matrix

\((B_{ij} + b_{ij})\) when its kth column is replaced by column vector \((q_j + \varepsilon_j)\).

We know from the theory of equations that the solution of

equations (4.2.1) is given by

\[ x_k = \frac{|D^k + d^k|}{|B+b|} \quad \text{for } k = 1, 2, \ldots, m. \] (4.1.2)

Also, we will be interested to find the distribution of a

linear function of the solutions \(x_i\), \(i = 1, 2, \ldots, m\). For

this purpose, let \(C\) be an m-column vector of fixed con­

stants \((C_1, C_2, \ldots, C_m)\), and \(c\) be a column vector of

errors \((c_1, c_2, \ldots, c_m)\), where \(E(c_i) = 0, i = 1, 2, \ldots, m,\)

and \(E(c_i^2) = w_i^2, i = 1, 2, \ldots, m\). We consider the linear

function

\[ Y = \sum_{r=1}^{m} (C_r + c_r) X_r \]
= \frac{1}{|B+b|} \sum_{r=1}^{m} (C_r + c_r)(|D_r + d_r|) \quad (4.1.3)

B. Rule of procedure for an approximation.

Now the determinants involved in expressions (4.1.2) and (4.1.3) will be reduced to approximate expressions using the following rule of procedure.

We will ignore all cross products of errors of second and higher order in the expansion of a determinant whose elements involve errors.

It may be noticed that although our rule of procedure does not assume the square and higher powers of the errors to be zero, they do not occur in the expansion of the determinants anyway, since no element is multiplied by itself in the expansion of a determinant.

Also, the procedure of replacing second order cross product of errors by zero will, in fact, imply that we are assuming that all errors are mutually uncorrelated. However, there is another possibility. In the expansion of a determinant whose elements involve random errors, we will have a linear function of cross products of order two. This linear function may be zero or small, although separately the products may not be zero. Anyway, we will follow the rule of procedure as given above.
C. Approximate Expressions.

Using the rule of procedure given above, we get expansions of the determinants involved in expressions (4.1.2) and (4.1.3) as follows.

\[ |D^k + d^k| = |D^k| + \sum_{i=1}^{m} \beta_{ik} \epsilon_i + \sum_{j=1}^{m} \sum_{j \neq k} D_{ij} b_{ij} \]

\[ = N(X_k) \text{ say.} \]  \hspace{1cm} (4.1.4)

Similarly

\[ |B+b| = |B| + \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_{ij} b_{ij} = D(x) \text{ say.} \]  \hspace{1cm} (4.1.5)

Now

\[ E(|D^k + d^k|) = |D^k| = \sigma_k \text{ say} \]  \hspace{1cm} (4.1.6)

\[ V(|D^k+d^k|) = \sum_{i=1}^{m} (\beta_{ik})^2 \Gamma_i^2 + \sum_{j=1}^{m} \sum_{j \neq k} (D_{ij})^2 \sigma_{ij}^2 \]

\[ = \sigma_k^2 \]  \hspace{1cm} (4.1.7)

and

\[ E(|B+b|) = |B| = \beta \text{ say} \]  \hspace{1cm} (4.1.8)
and

\[ V(|B+b|) = \sum_{i=1}^{m} \sum_{j=1}^{m} (\beta_{ij})^2 \sigma_{ij}^2 = \sigma B^2 \text{ say. (4.1.9)} \]

Also

\[ C^k(|D^k+d^k| |B+b|) = \sum_{i=1}^{m} \sum_{j=1}^{m} D_{ij} \beta_{ij} \sigma_{ij}^2 \]

\[ = \sigma_{Bk} \text{ say.} \tag{4.1.10} \]

We are interested in finding the distribution of the quotient corresponding to (4.1.2)

\[ X_k = \frac{N(X)}{D(x)} \tag{4.1.11} \]

For the reduction of expression (4.1.3) we have

\[ \sum_{r=1}^{m} (C_r+c_r)(|D^r+d^r|) = \sum_{r=1}^{m} C_r |D^r| \]

\[ + \sum_{r=1}^{m} C_r \left( \sum_{i=1}^{m} \beta_{ir} \epsilon_i \right) + \sum_{r=1}^{m} C_r \sum_{j=1}^{m} D_{ij} b_{ij} \]

\[ + \sum_{r=1}^{m} |D^r| c_r = N(y) \text{ say,} \tag{4.1.12} \]

\[ ^1 \text{In this chapter, } V \text{ represents the variance and } C, \text{ the co-variance of the corresponding functions.} \]
\[ E(N(y)) = \sum_{r=1}^{m} c_r |D^r| = \delta y \text{ say} \quad (4.1.13) \]

\[ V(N(y)) = \sum_{r=1}^{m} \left( \sum_{i=1}^{m} c_i \beta_{ri} \right)^2 r^4 + \sum_{r=1}^{m} c_r \sum_{i=1}^{m} \sum_{j=1}^{m} (D^r_{ij})^2 \sigma_{ij}^2 \]

\[ + \sum_{r=1}^{m} (D^r)^2 w_r^2 = \sigma^2_{N(y)} \text{ say.} \quad (4.1.14) \]

Similarly \[ |B + b| = D(x) = D(y) \text{ say (4.1.5)} \] where \[ E(D(y)) = 0 \]
and \[ V|D(y)| = \sigma^2_B \quad (4.1.8) \text{ and (4.1.9)} \text{ and } \]

\[ C(N(y)D(y)) = \sum_{r=1}^{m} c_r \sum_{i=1}^{m} \sum_{j=1}^{m} D^r_{ij} \beta_{ij} \sigma^2 \]

\[ = \sigma_B \cdot N(y) \text{ say} \quad (4.1.15) \]

and the problem again is to find the distribution of the quotient,

\[ y = \frac{N(y)}{D(y)} \quad (4.1.16) \]

4.2 General procedure.

Suppose that we know the probability distributions of the elements of \((b_{ij})\) and other errors \((c_k)\). We have seen in Section 4.1 that we want the distribution of a quotient of two linear functions of those errors derived. Therefore,
our first step will be to find separately the distributions of the numerator and the denominator, which are linear functions of random variables with known distributions.

A. Distribution of a linear function of variates (general procedures).

In statistics we have, in general, two ways of finding the distribution of a linear function of random variables. First, we write down the joint distribution of the variables involved and then make the necessary transformations. In our case, we have linear functions of variables. Assuming mutual independence we can write down the joint distribution of these variables as a product of the individual probability distributions. Suppose we have two random variables, \( x \) and \( y \), with probability distributions, \( f_1(x) \) and \( f_2(y) \), then the joint distribution with the assumption of independence, is \( f(x, y) = f_1(x) \cdot f_2(y) \). If we are interested in finding a distribution of a linear function, say

\[
ax + by = w, \quad (4.2.1)
\]

using the transformation \( x = \frac{w - by}{a} \), we will have

\[
f_3(y, w) \, dy \, dw = f_1 \left( \frac{w - by}{a} \right) f_2(y) \, dy \, dw \quad (4.2.2)
\]
The ranges of variation of $y$ and $w$ can be determined from the ranges of $x$ and $y$.

To get the distribution of $w$, we integrate (4.2.2) with respect to the other variable or variables (in this case $y$) over its whole for its range. This is one way to get the distribution of a linear function of random variables.

An alternative way of doing the same is with the help of characteristic functions. Given the probability distributions of a number of random variables, and hence, given the corresponding characteristic functions, the characteristic function of a linear function of these can be obtained by the following device. If $x$ and $y$ are independent variables, then the characteristic functions of

$$
(ax + by) = E(e^{i(ax + by)t}) = E(e^{ix(at)} E(e^{iy(bt)}).
$$

Thus, if $\phi_x(t)$ and $\phi_y(t)$ are the characteristic functions of $x$ and $y$, respectively, the characteristic function of $(ax + by)$ is $\phi_x(at) * \phi_y(bt)$. It may be remarked here that the characteristic function of a random variable always exists. Therefore, if we knew the probability distribution of the errors involved, we can get the characteristic function of a linear function of these errors. Now we can get the cumulative distribution for the linear function.
is known by using the following inversion formula.

**Inversion formula** 1. Let \( F(x) \) be a cumulative distribution function and \( \phi(t) \) be the corresponding characteristic function. Then

\[
F(x) + F(x-0) = 1 - \frac{1}{\pi i} \oint e^{-it} \phi(t) \frac{dt}{t}
\]

(4.2.4)

where the integral on the right is Cauchy’s principal value.

**B. Distribution of the quotient of two variates.**

Now we list the procedures to find the distribution of the quotient of two random variables, whose distributions are known. H. Cramer\(^a\) presented the following theorem in his paper "Random Variables and Probably Distributions."

**Theorem 1.** Let \( x_1 \) and \( x_2 \) be independent variables with finite mean values, the corresponding cumulative distributions being \( F_1(x) \) and \( F_2(x) \) with characteristic functions \( f_1(t) \) and \( f_2(t) \). If \( F_2(0) = 0 \) and if the integral

\[
\int_{-\infty}^{\infty} \frac{|f_2(t)|}{t} dt
\]

converges the cumulative distribution function of the quotient \( x_1/x_2 \) is given by the relation


\(^{a}\text{Cramer, H. Random Variables and Probable Distributions, Cambridge, Cambridge Univ. Press, 1937.}\)
If the integral obtained by formal differentiation of this relation with respect to \( x \) is uniformly convergent in a certain interval, we will have in that interval for the frequency function

\[
G(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{f_2(t) - f_1(t)}{t} f_2(-tx) \, dt. \tag{4.2.3}
\]

where \( f'_2 \) is the derivative of \( f_2(t) \) with respect to \( t \).

R. C. Geary\(^1\) gave an extension of this result for the case where \( x_1 \) and \( x_2 \) are not necessarily independent, but he himself doubted if that form can be of much practical use.

We may also use the following theorem by J. H. Curtiss\(^2\):

**Theorem 2.** If \( x \) and \( y \) are independent chance variables with respective cumulative distribution functions \( F(x) \) and \( G(y) \), the cumulative distribution function of \( Z = x/y \) is given by

\[
H(z) = G(0) + \int_0^\infty F(zw) \, dG(w) - \int_0^\infty F(zw-0) \, dG(w) \tag{4.2.5}
\]

for all values of \( z \).

---


These theorems may have particular advantage in certain particular cases, but the fact that \( x \) and \( y \) have to be independent makes it sufficiently restrictive for application.

The above methods of finding distributions of a quotient may be very cumbersome in many cases. Another way which might help in those cases will consist of calculating the moments of such a distribution. Karl Pearson\(^1\) made the first attempt in that direction. He gave general formulas for first four moments of a quotient \( x/y \), but added that they were practically unworkable if \( x \) and \( y \) were correlated, since these involved calculations of third and fourth order product moments. He gives formulae in case \( x \) and \( y \) were uncorrelated. He attributed these to Dr. M. Greenwood but quotes that the latter was unable to make them work easily in practice.

Dr. C. C. Craig\(^2\) has worked on the same lines and has also derived the important extension to the case when the numerator and the denominator are correlated, but all these procedures are very tedious, since after deriving the constants, one has to work out approximate distributions. However, the following inversion formulae can directly give the distribution.

---

\(^1\)Pearson, Karl. On the Constants of Index Distributions, etc., Biometrika, 7:531–541, 1910.

Inversion formulae. If \( \phi(t_1 t_2) \) is the characteristic function corresponding to the joint distribution of \( x \) and \( y \), then the cumulative distribution function of \( z = x/y \) is given by

\[
G(z) + G(z-0) = \frac{1}{\pi^1} \int \phi \left( \frac{t_2-t_1}{t} \right) dt
\]

where the integral is Cauchy's principle value.

The advantage of this formula is that in its derivation the independence of \( x \) and \( y \) is not assumed. We find the joint distribution of \( x \) and \( y \), that is, the numerator and the denominator of the quotient and then get the characteristic function corresponding to that joint distribution and apply this result to get the distribution of the quotient. However, to find the joint distribution of the numerator and the denominator may be very difficult in certain cases.

The procedures suggested in this section are added mainly for the completeness of the theory. It may be noticed, however, that when we have a quotient of two functions which are linear in a large number of variables, it will be extremely difficult to find its distribution. Further, even if these distributions may be derived, tremendous computational work will be necessary before they can be used for

---

the practical purposes of making statistical predictions as previously suggested.

4.3 Normal distributions.

In the preceding sections, we have suggested general procedures to handle the problem of finding distributions of the solutions of linear equations when the constants involved are distributed according to given probability distributions. However, in practice, if the distributions involved are assumed other than normal, the handling of the problem becomes extremely difficult. Therefore, we assume that the errors involved in our analysis are normally distributed with means zero and with corresponding variances, and then we deal with the expressions (4.1.11) and (4.1.16). We notice that in both the expressions, the numerators $N(x_k)$ and $N(y)$ and the denominators $D(x)$ and $D(y)$, are linear functions of the normally distributed errors. Since such functions are themselves normally distributed with the corresponding parameters, $N(x_k)$ will be normally distributed with expected value $\bar{\epsilon}_k$ and variance $\sigma^2_k$. $N(y)$ will be normally distributed with expected value $\bar{\epsilon}_y$ and variance $\sigma^2_N(y)$. Also, $D(x)$ and $D(y)$ will be normally distributed with expected value $\beta$ and variance $\sigma^2_D$. In this manner, the problem reduces itself to finding the distribution of a quotient of two
normal variates.

We will use the following theorem proved by

R. C. Geary.¹

Theorem. If N and D are normally distributed variables

with \( E(N) \) and \( E(D) \) as the mean values, \( \sigma_N^2 \), \( \sigma_D^2 \) as

variances and \( \sigma_{ND} \) as covariance, and \( Z = N/D \) is the

quotient, then the expression

\[
\frac{E(D) Z - E(N)}{(\sigma_D^2 Z^2 - 2 Z \sigma_{ND} + \sigma_N^2)^{1/2}}
\]

is approximately normally distributed with mean zero and

variance one, provided \( E(D) > 3 \sigma_D \). (4.3.1)

This formula will be invalid if the denominator is

zero but the proviso (4.3.1) ensures that this is very

unlikely to happen.

Therefore, the frequency function of \( Z \) is

\[
f(Z)dz = \frac{\left[ E(D) \sigma_N^2 - E(N) \sigma_{ND} \right] + Z \left[ E(N) \sigma_D^2 - E(D) \sigma_{ND} \right]}{\left[ \sigma_N^2 - 2 \sigma_{ND} Z + Z^2 \sigma_D^2 \right]^{3/2}}
\]

¹Geary, R. C. The Frequency Distribution of the

Quotient of Two Normal Variables, Royal Stat. Soc. Journ.,

93:442-446, 1930.
\[
\frac{1}{\sqrt{2\pi}} e^{-1/s} \left\{ \frac{z \ E(D) - E(N) \ #}{\sigma_N^2 - 2\sigma_{ND} + \sigma_D^2 \ z^2} \right\}_{z}
\]

In particular, applying this to the quotients of Section 4.1, we get the distributions of elements of the vector \( X \) and the linear function thereof.

The frequency distribution of \( X_k = \frac{N(X_k)}{D(X)} \), the \( k \)th element of the solution vector \( X \) of the equations

\[(B+b)X = (Q+\varepsilon),\]

is given by

\[
f(X_k) = \frac{\beta \sigma_k^2 - \beta \sigma_B^2 + X_k \beta \sigma_B^2 - \beta \sigma_{Bk}^2}{\left( \sigma_k^2 - 2\sigma_{Bk} X_k + X_k^2 \sigma_B^2 \right)^{3/2}}
\]

for \( k = 1, 2, \ldots, m \).

Also, the distribution of \( Y = \frac{N(Y)}{D(Y)} \), (4.1.16) is given by

\[
f(Y) = \frac{\beta \sigma_N^2 - \beta \sigma_{BN}^2 + Y \beta \sigma_{B(N)}^2 - \beta \sigma_{BN(Y)}^2}{\left( \sigma_N^2 - 2 \sigma_{BN} Y + Y^2 \sigma_B^2 \right)^{3/2}}
\]

and so on.
which gives frequency function of $C'X$, a linear function of the solution $(X_i)$, \( i = 1, 2, \ldots, m \).

It should be clearly noted that the derivation of the above distributions are subject to the restriction that the denominator of the quotient is unlikely to have negative values, and if

$$E(D) > 3\sigma_D$$

it may be reasonably so assumed. Since the probability that a normal variate deviates towards the left from its mean value by more than three times its standard derivation is .00135. Therefore, it is very unlikely that the variable in the denominator will assume zero value and the above derivation is thus permissible.

4.4 Confidence limits.

In practice, we may not be interested in the theoretical distributions as derived in Section 4.3. We may be interested in making statistical inferences about our solutions. In this section, we consider confidence limits of the quotient $Z = N/D$. We have seen that

$$t = \frac{E(D) Z - E(N)}{\left[ \frac{\sigma_D^2 Z^2 - 2Z\sigma_{ND} + Z^2 \sigma_N^2}{\sigma_D^2} \right]^{1/2}}$$
is approximately normally distributed with mean zero and variance unity. If we want limits with 100 per cent confidence, then consulting the standard normal tables, we find a positive number \( \gamma \) such that the probability of the absolute value of the standard normal deviate to be greater than \( \gamma \) is \((1-a)\). In other words

\[
P \left[ \left| \frac{E(D)Z - E(N)}{\sigma_D^2 Z^2 - 2Z\sigma_{ND} + \sigma_N^2} \right| \leq \gamma \right] = a, \quad (4.4.1)
\]

or

\[
P \left[ \left\{ E(D)Z - E(N) \right\}^2 \leq \gamma^2 \left( \sigma_D^2 Z^2 - 2Z\sigma_{ND} + \sigma_N^2 \right) \right] = a,
\]

or

\[
P \left[ \left\{ E(D)^2 - \gamma^2 \sigma_D^2 \right\} Z^2 - 2\left\{ E(D)E(N) - \gamma^2 \sigma_{ND} \right\} Z + \left\{ E(N)^2 - \gamma^2 \sigma_N^2 \right\} \leq 0 \right] = a. \quad (4.4.2)
\]

Thus the roots of the quadratic equation in \( Z \)

\[
\left\{ E(D)^2 - \gamma^2 \sigma_D^2 \right\} Z^2 - 2\left\{ E(D)E(N) - \gamma^2 \sigma_{ND} \right\} Z + \left\{ E(N)^2 - \gamma^2 \sigma_N^2 \right\} = 0 \quad (4.4.3)
\]

\(^1P\) in the discussion will stand for a probability statement.
will give the two numbers and we will be 100% per cent sure that \( Z \) lies between those values.

Applying this general formula to find confidence limits of \( x_k \) and \( y \) of Section 4.1, we see that the roots of the quadratic equation in \( Z \)

\[
\left\{ \beta^2 - \gamma^2 \sigma^2_B \right\} Z^2 - 2 \left\{ \beta \cdot \bar{z} - \gamma \sigma_B \right\} Z + \left\{ \bar{z}^2 - \gamma^2 \sigma_k^2 \right\} = 0
\]

(4.4.4)

will give confidence limits for solution \( x_k \) (for \( k=1,2, \cdots, m \)) and the roots of equation

\[
\left\{ \beta^2 - \gamma^2 \sigma^2_B \right\} Z^2 - 2 \left\{ \beta \bar{y} - \gamma \sigma_{BN}(y) \right\} Z
\]

\( + \left\{ \bar{y}^2 - \gamma^2 \sigma_N^2(y) \right\} = 0 \)

(4.4.5)

will give confidence limits for the linear function of the solution. In both cases, the level of confidence is 100% per cent.

4.5 Confidence limits for the results of programming.

The above results when applied to programming models yield valuable economic interpretations.

If, in the case of a programming problem, we know the expected values and the variances of the input coefficients, which specify the linear production relations between all outputs and inputs, then we can get the approximate
frequency distributions of the levels of activities by the use of the results of Section 4.3. In practice, if we are making a program schedule to realize those levels, we will be faced with the problem of finding the expected values and variances of the technical coefficients. We need very extensive data and will try to get the best possible values with the known econometric methods. We plan the schedule according to the expected values of those input coefficients. Now, when the whole program is in process, there is no reason to believe that the values of the input coefficients with which we planned will actually be in operation at that time. We assume that the input coefficients are random variables which are normally distributed about the values used in planning, and with known variances. The assumption of normality may not be justified in many cases. But we confine ourselves to the cases in which this assumption is fairly justifiable to demonstrate the working of this approach. If the actual operating input coefficients at the time of operation of the program were the same as our estimates used, the results expected in our program analysis will certainly be realized, but in all practical cases, they are bound to be actually different. If we know the variance (at least its best estimate) we have an idea of the extent of the variability. Therefore, when we make a production program, we may be interested not only to know
the activity levels expected to prevail in our program but also certain limits between which those levels may be realized. The application of Section 4.4 will give such limits with any preassigned confidence coefficients. For instance, the probability that the level of an activity realized will fall between limits determined by the method of Section 4.4, is .95 if a is taken to be .95.

The tables of standard normal distribution will give $\gamma = 1.96$ since the probability is .05 that the absolute values of a standardized normal variate may be greater than 1.96. The values of $\beta$, $\delta_k$, $\sigma_k^2$, $\sigma_{Bk}^2$, and $\sigma_B^2$ are obtained from Section 4.1. Using these values in equation (4.4.4), we get the confidence limits for kth activity, by solving that equation for $Z$. Also, we can get confidence limits for the realization of final bill of goods or the employment function in the Leontief models or profit function in linear programming analysis by using the equation (4.4.5), since these are linear functions of the activity levels.

In Section 4.1 we considered the coefficients of the linear function also as random variables. In Section 4.3 we derived approximate frequency distribution of a linear function considering the coefficients of the linear function to be normally distributed. This may be very useful in certain economic analyses. For example, if the
linear function represents the profit function and the coefficients are prices or the net values of unit activity levels. Then from equation (4.3.4) we get the approximate frequency distribution of the profit function when there is a fair reason to believe that those net values may have normal distributions. In case we propose to consider prices fixed, we just replace the corresponding variances by zeroes and get the frequency distribution of the profit function. Similarly to get the confidence limits for the profit function (or any other linear function of the activity levels) we compute $\beta$, $\partial y$, $\sigma_B^2$, $\sigma_N(y)$, and $\sigma_{BN}(y)$ from Section 4.1.

Substituting these values in the equation (4.4.5) and taking $\gamma = 1.96$ for 95 per cent confidence, we get a quadratic equation in $Z$ whose roots give the confidence limits for the profit function.

Thus we notice that, even in the areas of production economics, in which it is difficult to predict with certainty the technological input-output relationships, we can employ programming techniques for planning purposes. We plan with the expected values of the input coefficients and the variability of the results using the analysis of this chapter.
4.6 Cumulative distributions.

If the frequency distribution of a continuous variable is known, its cumulative distribution is obtained by integration from the lower limit of the range of variation to another variable. For instance, in Section 4.3 we obtained distributions of $X_k$ and $y$, that is an activity level and a linear function thereof. The variable $Z = N/D$ where $N$ and $D$ are normal variates, can represent both functions. If $f(Z)$ represents the frequency function of $z$, then $F(v)$, the corresponding cumulative distributions function, is obtained as follows:

$$ F(v) = \int_{-\infty}^{v} f(Z) \, dZ \quad (4.6.1) $$

In practice, however, it may be too tedious to evaluate the integral. Therefore, we suggest the following procedure, based on a paper by E. C. Fieller. He has shown that the chance of obtaining a value of the variable $Z = N/D$ not less than $v$, that is $[1 - F(v)]$ can be computed as below.

---

\[ 1 - F(v) = \int_{h}^{\infty} \int_{k}^{\infty} N(p) + \int_{-h}^{0} \int_{-k}^{\infty} N(p) \, dx \, dy. \]  \hspace{1cm} (4.6.3)

Where \( N(p) \) is bivariate normal distribution with means zeroes, variances each equal to unity and correlation \( \rho \).

And where \( h, k \) and \( \rho \) are computed as below

\[ h = \frac{E(D)}{(D)} \]  \hspace{1cm} (4.6.3)

\[ k = \frac{E(N) - v E(D)}{\left[ \sigma_N^2 - 2 \sigma_{ND} v + v^2 \sigma_D^2 \right]^{1/2}} \]  \hspace{1cm} (4.6.4)

and

\[ \rho = \frac{(\sigma_{ND} - v \sigma_D^2)}{\sigma_D \left[ \sigma_N^2 - 2 \sigma_{ND} v + v^2 \sigma_D^2 \right]^{1/2}}. \]  \hspace{1cm} (4.6.5)

The values of integrals in (4.6.3) can be obtained from Karl Pearson's\(^1\) tables. The tables are available only

\(^1\)Pearson, Karl. Tables for Statisticians and Biometricians, part II, Tables VII and IX, Cambridge Univ. Press, 1924.
for positive values of $h$ and $k$. However, the following relations regarding the bivariate normal distribution with means zeroes and variances unity and correlation coefficient $\rho$, can be used in case either or both of $h$ and $k$ are negative.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N(\rho) \, dx \, dy = \int_{-\infty}^{\infty} N(0,1) \, dy - \int_{-\infty}^{\infty} N(-\rho) \, dx \, dy \quad (4.6.6)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N(\rho) \, dx \, dy = \int_{-\infty}^{\infty} N(0,1) \, dx - \int_{-\infty}^{\infty} N(-\rho) \, dy \, dx \quad (4.6.7)$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N(\rho) \, dx \, dy = 1 - \int_{-\infty}^{\infty} N(0,1) \, dx - \int_{-\infty}^{\infty} N(0,1) \, dy$$

$$(4.6.8)$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N(\rho) \, dx \, dy$$

$h \, k$

where $N(\rho)$ refers to the bivariate normal distribution and $N(0,1)$ refers to normal distribution with mean zero and variance one.

By the use of these tables and the formulae (4.6.2) to (4.6.5) we can get the probability that $Z$ may not be less
than any pre-assigned number v.

To get the probability that \( Z \) may not be greater than \( v \) we compute \( 1-F(v) \) by the above procedure and subtract it from unity.

Identifying \( Z \) with \( X_k \), the kth activity level, we can find the probability that in the final result of our program, the level of the kth activity realized will not be less than or greater than a pre-assigned number. Similarly, identifying \( Z \) with \( y \), a linear function of the activities, we can make a similar statement. The importance of these results can hardly be over-emphasized. For example, if our optimum plan is expected to bring us a profit of five thousand, we can apply the above analysis to get the approximate probability that the profit will not be less than four thousand and if the probability is fairly low, we have increased confidence in our program. Possible losses cannot be too great. Therefore, given \( v \), we find the values of \( \rho \), \( h \) and \( k \) and then consult "Tables for Statisticians and Biometricians," and use the results of this section.

4.7 Program efficiency control.

In a practical programming schedule, we use the input coefficients based on future expectations. In actual operation, these input coefficients may be different, with the result that the values of levels of activities will
differ from what occurs in our schedule. For instance, if the kth input coefficients for the jth activity is more than the one used in planning, our program will realize less output of jth activity, since the amount of kth input was limited. But a large deviation may lead us to think that something is seriously wrong with the values of coefficients used in planning. Suppose an economic unit is to apply that program every month (or any unit of time) and has limited monthly supply of the inputs. Further, it makes not only the program schedule, but also computes the limits of variation of the outcome variables or their linear functions with a certain confidence probability say 75 percent. Of course, this probability will depend on the value a firm attaches on its efficiency as well as on the cost of re-check and re-planning. If the final quantities of the outcome variables violate these limits, the economic unit may decide to re-evaluate the technical coefficients and plan according to the new values for the next production period. This will be a mechanical way of checking the efficiency of our planning.

However, it may be noticed that we can't get the probability level of rejecting the planning as a whole unless we investigate the joint distribution of the activity levels, such as the X's. That seems too involved,
since the x's are not independently distributed. This, in fact, suggests that setting up an efficiency control chart on the basis of activity levels may not be a very sensitive and efficient procedure. Therefore, we suggest an alternative way to gauge the efficiency of the planning. We plan production with certain values of the input coefficients which we regard as expected values.

Now when the production is carried on according to that program we have observed values of those input coefficients. Let us denote these by $O(B_{ij})$ where the $B_{ij}$ are expected values, that is, the ones used in planning. Then the quantity

$$
\sum_{i=1}^{m} \sum_{j=1}^{m} \frac{(B_{ij} - O(B_{ij}))^2}{\sigma_{ij}^2}
$$

is distributed as $x^2$ with $m^2$ degrees of freedom, since the input coefficients are assumed normally distributed and also our rule of procedure for approximation implies their independence.

Comparing this value with the tabulated one, we can perform tests of significance. With a pre-determined probability $a$ we find the value $x^2$ with $m^2$ degrees of freedom from the table, say $c$, such that $[x_m^2 > c] = a$ and if $> this$ number $c$, we doubt the validity of the planning of the program. We reject the input coefficients with
which we planned and try to get better estimates.

However, the efficiency of this criterion will depend on the extent, to which the assumptions of normality and independence of the input coefficients can be justified in a particular situation.

4.3 A Critical note.

Before closing the chapter we must acknowledge some limitations of our approach. Firstly, the assumption of normality is too restrictive. The analysis will give useful results only in situations, in which the assumption is valid to a sufficient extent. Even if it is assumed as an approximation, the results should be carefully checked. We may be able to illustrate this point in our empirical example.

Secondly, the rule of procedure for approximate reductions could be modified so as to avoid the implication of independence of the errors. We could have retained second order products ignoring the higher order ones. Theoretically speaking, it would have given us better results. However, the product of two normal variates is not normal. First of all, we would have to find the distributions of two linear functions of normal and non-normal variates. Next, we would have to find the distribution of
quotient of such two linear functions. Finally, that
distribution, even when derived, would not be useful for
purposes unless it is tabulated.

Therefore, we feel justified to suggest approximate
procedures to deal with that situation. We have derived
useful formulae which can be applied easily, at the cost
of some precision.
V. SUMMARY

The problem of finding statistical distributions of the activity levels, when the input coefficients vary according to known distributions, can be reduced to the following mathematical model

\[(B + b)x = (Q + \varepsilon)\]

where \((B + b)\) is a \(m \times m\) matrix whose elements are considered as random variables, with the corresponding elements of matrix \(B\) as expected values. \((Q + \varepsilon)\) is an \(m\) column vector of \(m\) variables with the corresponding elements of the column vector \(Q\) as expected values.

Elements of the matrix \(b\) and the vector \(\varepsilon\) are errors whose distributions are assumed to be known. The problem is to find the distributions of the elements of \(m\)-column vector \(X\).

Also it was proposed to find the distribution of a linear function of the elements of \(X\), that is

\[\sum_{r=1}^{m} (C_r + c_r)X_r \quad \text{or} \quad (C + c)'X\]

in matrix notation. Elements of the column vector \((C + c)\) are also random variables with the corresponding elements of \(C\) as expected values. Elements of \(c\) again represent
corresponding errors.

This model is quite general. When identified with a non-degenerate solution of a linear programming problem, it will have the following interpretation.

B is the matrix of input coefficients used in planning the program. Q is the column vector of fixed resources. Expected values of the elements of X represent the levels of activities (active or disposal) which were planned for production with a view to maximize the profit function $(C + c)^T X$. The elements of $(C + c)$ represent the value of a unit level of activity. These may be prices of the products or net profit per unit of activity. If these values are assumed to be constants, $c = 0$.

Similarly, the model can be identified with Leontief's inter-industry models.

In dealing with the general model we have seen in Chapter IV that useful conclusions, which are easy to apply in practice, can only be drawn if we assume the distributions of the random errors involved to be normal. Section 4.2 gives general procedures which may be helpful in dealing with non-normal distributions, but it is realized that it will be very difficult, if not impossible, to draw results easily applicable in practical problems.

Assuming normality, however, one can find at once, the approximate confidence limits for the activity levels and
the linear function by solving for \( Z \), the equations (4.4.4) and (4.4.5). The constants involved in these equations are evaluated by formulae in Section 4.1 and are, in fact, functions of the expected values and the known variances of the distributions of the parameters of our model. The procedure of Section 4.6 enables us to make probability statements with the help of cumulative distributions. For example, if one is interested in finding the probability that a particular activity level will realize more than or less than a given amount, one can get an approximate answer by using the easy computational procedure given in Section 4.6. The confidence limits, of course, can also be obtained by this procedure. Moreover, since the analysis of Sections 4.4 and 4.6 are based on different computational procedures, the results obtained by one can be checked against those arrived at by the other.

Apart from the assumption of normality, another disturbing factor in this analysis is the implication that the probable errors are uncorrelated. Our rule of procedure for approximation equates second order product terms to zero. This is equivalent to assuming that the probable errors are uncorrelated with each other. Hence, if in any practical problem the covariances of errors are not sufficiently small relative to the variances, the approximate results given by using the Sections 4.3, 4.4, and 4.6
may be poor.

In Chapter III, a review of another approach is given briefly. If the possible discrepancies between the constants used for planning and their actual values are considered to be known, the possible discrepancies between the planned levels of activities and the levels realized can be approximated by that procedure. It may be noticed that the variation in the constants in this case is not considered as random.
PART II
APPLICATION TO FARM PRODUCTION PLANNING

Application of the theory of linear programming in planning production program for a family farm of model size in Iowa and application of the probability approach to the results of the program.
VI. APPLICATION OF LINEAR PROGRAMMING

6.1 Constant returns to scale.

The application of programming models to agricultural production has not been considered practicable mainly for two reasons. First of all, there is the classical notion of the diminishing returns to scale in agricultural production. To apply the techniques of linear programming we need strict linearity of the input coefficients. That is, the output increases proportionately as all the inputs are increased in the same proportion. In other words, strictly constant returns to scale are to be assumed. In problems like optimum mixture of certain ingredients constant returns to scale are justifiable. There has been a prevalent belief that in agricultural production, constant returns to scale cannot be assumed. We think that this belief in diminishing returns refers to proportionality relationship rather than scale relationship. As Dr. E. O. Heady\(^1\) says,

Decreasing physical returns to scale in agriculture are likely to be explained mainly in managerial limitations. However, the relationship falls in the realm of proportionality when a single stock of management is limited, etc.

If the variable $y$ represents the yield of an agricultural commodity and $x_1, x_2, \ldots, x_n$ are all different inputs, then a scale relationship will be of the form

$$y = f(x_1, x_2, \ldots, x_n).$$

All the inputs are variable. We will have constant returns to scale if

$$f(kx_1, kx_2, \ldots, kx_n) = kf(x_1, x_2, \ldots, x_n),$$

where $k$ is an arbitrary number. However, if we keep management constant or land constant, then the production function will be of the form

$$y = f(x_1, x_2, \ldots, x_{n-1}|x_n)$$

where $x_n$ represents that fixed factor. This is a proportionality relationship and doubling the amounts of $x_1, x_2, \ldots, x_{n-1}$, we may not have doubled the yield. However, it cannot be concluded that we have diminishing return to scale. Similarly, if more than one factor, say $x_{n-1}$ and $x_n$ are held constant, and the production function

$$y = f(x_1, x_2, \ldots, x_{n-2}|x_{n-1}, x_n)$$

is derived which indicates diminishing returns for the proportionate increase in the variable inputs $x_1, x_2, \ldots, x_{n-2}$ it would not be legitimate to say that diminishing returns
to scale is indicated. These are not scale relationships. We observe that most of the statements regarding the diminishing returns to scale in agriculture are based on the idea of limited land and managerial difficulties to control large farms. Although decreasing returns hold true for the individual resources for all farm types, there seems reason to believe that if all factors are increased proportionately, the total production would increase proportionately. However, the exact nature of returns to scale in agriculture is still a matter to be thoroughly investigated. As Dr. E. O. Heady¹ points out, "Until more conclusive data is derived, the exact nature of returns to scale for any segment of agriculture will remain unknown."

The following examples may be quoted in regard to investigations for returns to scale. A production function² derived for southern Iowa farms in 1950 gave the following results, where $y$ refers to the value of farm production, $R$ refers to real estate (land and buildings) $L$ refers to labor, $M$ refers to machine services, $F$ refers to live stock and feed services, and $Z$ refers to miscellaneous resources services.

¹Ibid. p. 359.
²Ibid. pp. 359-360.
\[ Y = 0.31 R^{23} L^{0.03} M^{1.10} F^{0.50} Z^{0.25}, \]

where the exponents of \( R, M, L, F, \) and \( Z \) represent the corresponding coefficients of elasticity. Since the sum of the coefficients of elasticities is 1.11, increasing returns to scale should be indicated. However, a test of significance indicated that, at a 5 per cent probability level, the results were not significantly different from constant returns. Another example is a production function\(^1\) estimated by means of simultaneous equations for cash-crop farms in Central Iowa for 1948. It resulted in the following regression equation with input aggregated into labor (\( L \)) all capital services (\( C \)) and land (\( R \)).

\[ Y = 1.46 L^{0.06} C^{0.47} R^{0.57}. \]

As before, while the sum of the coefficients of elasticities is 1.1, thus indicating increasing returns to scale, it was not statistically significantly different from constant returns. Thus, if we assume constant returns to scale, particularly in a short run situation, as we need for the application of linear programming, we will not be violating some established fact to the contrary.

\(^1\)Ibid. pp. 359-360.
We will deal with a family farm which has fixed land capital and labor resources at its disposal. It is interested to draw up a production plan for the coming season with a view to maximize its profit. The continuation of the present state of technology is also assumed. In this short run situation, constant returns to scale implicit in the assumption of strict linearity of input coefficients can be fairly well justified. It amounts to saying that if to produce 100 bushels of corn, \( a_1, a_2, \) and \( a_3 \) are the amounts of land labor and capital required, then to produce 200 bushels, we need \( 2a_1, 2a_2, \) and \( 2a_3 \) of those factors respectively. Apart from the uncertainty inherent in agricultural production, this can be reasonably assumed.

6.2 Uncertainty.

Another reason why programming techniques are considered to be more suitable in industrial production than in agriculture problems, is the relatively more certain technological relationships between inputs and outputs existing in that field.

In problems of optimum mixture of certain ingredients, say different kinds of nuts, the input coefficients are certain. In agriculture production, it is almost impossible to determine exact input or technological coefficients pertaining to next year's production. It is very important to
consider the element of risk and uncertainty in agriculture.

As we have suggested in previous chapters, in spite of risk and uncertainty in agriculture, there is no reason to give up all attempts of efficient planning. Under the circumstances, we endeavor to extend our theory and do the best we can. Obviously, we will have to make some sacrifices. If the exact technological relationship cannot be predicted, the exact prediction of outputs is not possible.

In the literature, a distinction is made, sometimes, between risk and uncertainty. The variability of outcomes which can be measured in probability, empirically or "apriori" is referred to as risk. It is a variability which is insurable in an actuarial sense since the parameters of the probability distribution or distributions can be estimated. For instance, normal wear and tear and depreciation on farm machinery can be predicted with sufficient accuracy, whereas losses due to fire and such hazards cannot be predicted. As Dr. E. O. Heady\(^1\) says,

\[\text{The year to year "variability" in crop yields associated with fluctuations in the weather may be classed as risks on farms (a) where climate is highly stable (b) where small random variations occur from year to year and (c) where the complete}\]

\(^1\)Ibid. p. 441.
range of yield outcomes is repeated frequently enough that the farmer operator can establish mean or model outcome and the range (variance of outcomes).

On the other hand, the probability of an outcome may not be estimable in an empirical sense. A serious drought resulting in utter failure of the crops may occur once in fifty years or once in two hundred years. Such happenings are classed as pure uncertainty.

Bearing in mind, this distinction between risk and uncertainty, we will assume that no uncertain hazard is going to happen in the year for which the farm is going to plan. In spite of this simplification, there is year to year yield variation, due to small weather fluctuations and so on. It only allows us to measure that variability in a probability sense, thus making it possible for us to apply the analysis of Chapter IV. The general scheme is to get the best possible values of the input coefficients and make a production-program which will maximize the profit if these values actually prevail next year. Then, having some idea about the variation of the input coefficients, we may predict the limits of the variation of the outcomes. This procedure is definitely better than no planning at all.
6.3 Optimum production program.

The first problem to be faced now, is the evaluation of the input coefficients. In the case of a farm, it involves prediction, and expectations, mainly concerning next year's yields.

There are many naive procedures to predict these yields:

1) "random prediction." For example the yields and corresponding information regarding the inputs for the past ten or twenty years may be written on different pieces of paper and may be drawn at random. The technical coefficients (that is each input divided by the corresponding output) may be used to draw the production program for the next year.

2) "projection of current yields." That is to say we draw the production program on the information regarding the most recent past year and calculating the input coefficients therefrom. On the other hand, if yields for the current year are very high, the corresponding yields for next year may be expected to be low.

3) "model yields." The most frequent yield figure in the data of the past twenty years or so, and the corresponding information regarding inputs, may be used to compute input coefficients for planning next year's production.
4) "extension of the trend." If the time series of yields shows certain trend, the trend extrapolated may be to predict next year's yield.

5) "average yields." Yields or any other statistics (in our case, input coefficients) to be predicted may be estimated by their mean values corresponding to a number of past years. For detailed discussion of these procedures, one may refer to Dr. E. O. Heady's book.\(^1\)

In fact, the relative efficiency of these procedures will depend on different problems at hand. One may employ more sophisticated econometric techniques to predict those parameters, provided sufficient funds are available to invest in complex research, but unless very detailed and thorough care is taken to include the significant variables, such methods may not be much more efficient than naive procedures. In our example, we compute all relevant input coefficients for the last twenty-five years, from 1928 to 1952, and took their averages to draw the optimum production plan for next year. This was considered appropriate since we propose to use the estimated variances from these series as variances of the distributions of these input coefficients.

\(^{1}\)Ibid. pp. 478-496.
## Table 1
### Yield data

**Hancock County - Ellsworth Township**

**CRD No. 2**

<table>
<thead>
<tr>
<th>Year</th>
<th>Corn</th>
<th>Oats</th>
<th>Soybeans</th>
<th>Flax</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1928</td>
<td>35</td>
<td>36</td>
<td>11.3</td>
<td>11.2</td>
<td>13.6</td>
</tr>
<tr>
<td>1929</td>
<td>36</td>
<td>36</td>
<td>10</td>
<td>12</td>
<td>13.4</td>
</tr>
<tr>
<td>1930</td>
<td>40</td>
<td>37</td>
<td>21.1</td>
<td>16</td>
<td>18.4</td>
</tr>
<tr>
<td>1931</td>
<td>29</td>
<td>37</td>
<td>14.4</td>
<td>10</td>
<td>20.3</td>
</tr>
<tr>
<td>1932</td>
<td>28</td>
<td>22</td>
<td>20.4</td>
<td>10.6</td>
<td>7</td>
</tr>
<tr>
<td>1933</td>
<td>51</td>
<td>33</td>
<td>23</td>
<td>7.7</td>
<td>12.8</td>
</tr>
<tr>
<td>1934</td>
<td>38</td>
<td>14</td>
<td>10</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>1935</td>
<td>44</td>
<td>45</td>
<td>20</td>
<td>9.6</td>
<td>8.1</td>
</tr>
<tr>
<td>1936</td>
<td>25</td>
<td>40</td>
<td>20</td>
<td>9.2</td>
<td>10</td>
</tr>
<tr>
<td>1937</td>
<td>46</td>
<td>50</td>
<td>11</td>
<td>13.1</td>
<td>10</td>
</tr>
<tr>
<td>1938</td>
<td>53</td>
<td>42</td>
<td>22</td>
<td>14.8</td>
<td>6.2</td>
</tr>
<tr>
<td>1939</td>
<td>56</td>
<td>43</td>
<td>22</td>
<td>8</td>
<td>11.3</td>
</tr>
<tr>
<td>1940</td>
<td>56</td>
<td>46</td>
<td>20</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>1941</td>
<td>55</td>
<td>32</td>
<td>16</td>
<td>11</td>
<td>0.4</td>
</tr>
<tr>
<td>1942</td>
<td>63</td>
<td>44</td>
<td>17</td>
<td>11</td>
<td>4.0</td>
</tr>
<tr>
<td>1943</td>
<td>59</td>
<td>45</td>
<td>20</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>1944</td>
<td>44</td>
<td>34</td>
<td>19</td>
<td>7.4</td>
<td>4.6</td>
</tr>
<tr>
<td>1945</td>
<td>46</td>
<td>43</td>
<td>17</td>
<td>18</td>
<td>13.6</td>
</tr>
<tr>
<td>1946</td>
<td>60</td>
<td>45</td>
<td>22</td>
<td>20</td>
<td>26.1</td>
</tr>
<tr>
<td>1947</td>
<td>38</td>
<td>36</td>
<td>16</td>
<td>14.7</td>
<td>10.5</td>
</tr>
<tr>
<td>1948</td>
<td>60.3</td>
<td>44.5</td>
<td>18.4</td>
<td>17.1</td>
<td>26.1</td>
</tr>
<tr>
<td>1949</td>
<td>47.1</td>
<td>45.8</td>
<td>21.7</td>
<td>11.4</td>
<td>8.1</td>
</tr>
<tr>
<td>1950</td>
<td>50.7</td>
<td>48.9</td>
<td>21</td>
<td>9.6</td>
<td>16.4</td>
</tr>
<tr>
<td>1951</td>
<td>47.3</td>
<td>39</td>
<td>18.2</td>
<td>7.7</td>
<td>13.6</td>
</tr>
<tr>
<td>1952</td>
<td>67.2</td>
<td>41.3</td>
<td>23.7</td>
<td>11</td>
<td>4.7</td>
</tr>
</tbody>
</table>

$^+$County average yield

Source: Iowa Crop Reporting Service.
Table 2

Capital expense per acre (in dollars)

<table>
<thead>
<tr>
<th>Year</th>
<th>Corn</th>
<th>Oats</th>
<th>Soybeans</th>
<th>Flax</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1928</td>
<td>12.31</td>
<td>9.37</td>
<td>10.60</td>
<td>5.58</td>
<td>9.82</td>
</tr>
<tr>
<td>1929</td>
<td>12.41</td>
<td>9.26</td>
<td>10.27</td>
<td>9.02</td>
<td>9.61</td>
</tr>
<tr>
<td>1930</td>
<td>12.44</td>
<td>9.65</td>
<td>10.05</td>
<td>9.84</td>
<td>9.30</td>
</tr>
<tr>
<td>1931</td>
<td>11.02</td>
<td>7.98</td>
<td>8.61</td>
<td>7.63</td>
<td>8.38</td>
</tr>
<tr>
<td>1932</td>
<td>10.42</td>
<td>7.19</td>
<td>7.97</td>
<td>7.11</td>
<td>7.46</td>
</tr>
<tr>
<td>1933</td>
<td>11.17</td>
<td>7.38</td>
<td>8.27</td>
<td>7.19</td>
<td>7.64</td>
</tr>
<tr>
<td>1934</td>
<td>10.83</td>
<td>7.63</td>
<td>8.94</td>
<td>8.04</td>
<td>7.86</td>
</tr>
<tr>
<td>1935</td>
<td>11.41</td>
<td>8.39</td>
<td>8.99</td>
<td>7.11</td>
<td>8.30</td>
</tr>
<tr>
<td>1936</td>
<td>10.66</td>
<td>8.38</td>
<td>9.03</td>
<td>8.24</td>
<td>8.63</td>
</tr>
<tr>
<td>1937</td>
<td>12.34</td>
<td>8.93</td>
<td>9.20</td>
<td>7.81</td>
<td>8.89</td>
</tr>
<tr>
<td>1938</td>
<td>12.47</td>
<td>8.31</td>
<td>9.15</td>
<td>8.09</td>
<td>8.29</td>
</tr>
<tr>
<td>1939</td>
<td>12.27</td>
<td>8.59</td>
<td>9.30</td>
<td>8.31</td>
<td>8.63</td>
</tr>
<tr>
<td>1940</td>
<td>12.12</td>
<td>8.56</td>
<td>9.72</td>
<td>7.88</td>
<td>8.71</td>
</tr>
<tr>
<td>1941</td>
<td>12.42</td>
<td>8.60</td>
<td>9.40</td>
<td>8.18</td>
<td>8.54</td>
</tr>
<tr>
<td>1942</td>
<td>13.54</td>
<td>9.58</td>
<td>11.50</td>
<td>8.76</td>
<td>10.41</td>
</tr>
<tr>
<td>1943</td>
<td>13.90</td>
<td>10.45</td>
<td>11.89</td>
<td>9.88</td>
<td>10.34</td>
</tr>
<tr>
<td>1944</td>
<td>13.57</td>
<td>10.48</td>
<td>12.23</td>
<td>10.04</td>
<td>10.41</td>
</tr>
<tr>
<td>1945</td>
<td>14.06</td>
<td>10.93</td>
<td>12.59</td>
<td>10.73</td>
<td>11.10</td>
</tr>
<tr>
<td>1946</td>
<td>15.56</td>
<td>11.83</td>
<td>13.71</td>
<td>11.33</td>
<td>12.40</td>
</tr>
<tr>
<td>1948</td>
<td>19.90</td>
<td>15.21</td>
<td>18.42</td>
<td>16.73</td>
<td>16.69</td>
</tr>
<tr>
<td>1949</td>
<td>21.57</td>
<td>14.65</td>
<td>20.07</td>
<td>16.17</td>
<td>15.83</td>
</tr>
<tr>
<td>1950</td>
<td>22.13</td>
<td>15.50</td>
<td>21.63</td>
<td>17.02</td>
<td>16.68</td>
</tr>
<tr>
<td>1951</td>
<td>23.52</td>
<td>16.59</td>
<td>22.02</td>
<td>18.11</td>
<td>17.77</td>
</tr>
<tr>
<td>1952</td>
<td>23.79</td>
<td>15.09</td>
<td>22.29</td>
<td>16.61</td>
<td>16.27</td>
</tr>
</tbody>
</table>

Table 3

Land input coefficients (acreage per bushel)

<table>
<thead>
<tr>
<th>Year</th>
<th>Corn</th>
<th>Oats</th>
<th>Soybean</th>
<th>Flax</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1928</td>
<td>0.02857</td>
<td>0.02778</td>
<td>0.08849</td>
<td>0.08928</td>
<td>0.07353</td>
</tr>
<tr>
<td>1929</td>
<td>0.02778</td>
<td>0.02778</td>
<td>0.10000</td>
<td>0.08333</td>
<td>0.07463</td>
</tr>
<tr>
<td>1930</td>
<td>0.02500</td>
<td>0.02703</td>
<td>0.04739</td>
<td>0.06250</td>
<td>0.05435</td>
</tr>
<tr>
<td>1931</td>
<td>0.03448</td>
<td>0.03333</td>
<td>0.06944</td>
<td>0.10000</td>
<td>0.04926</td>
</tr>
<tr>
<td>1932</td>
<td>0.03571</td>
<td>0.04545</td>
<td>0.04902</td>
<td>0.09434</td>
<td>0.14286</td>
</tr>
<tr>
<td>1933</td>
<td>0.01961</td>
<td>0.03030</td>
<td>0.01348</td>
<td>0.12987</td>
<td>0.07812</td>
</tr>
<tr>
<td>1934</td>
<td>0.02631</td>
<td>0.07143</td>
<td>0.10000</td>
<td>0.14286</td>
<td>0.05988</td>
</tr>
<tr>
<td>1935</td>
<td>0.02273</td>
<td>0.02222</td>
<td>0.05000</td>
<td>0.10417</td>
<td>0.12346</td>
</tr>
<tr>
<td>1936</td>
<td>0.04000</td>
<td>0.02500</td>
<td>0.05000</td>
<td>0.10869</td>
<td>0.10000</td>
</tr>
<tr>
<td>1937</td>
<td>0.02174</td>
<td>0.02000</td>
<td>0.09091</td>
<td>0.07633</td>
<td>0.10000</td>
</tr>
<tr>
<td>1938</td>
<td>0.01887</td>
<td>0.02381</td>
<td>0.04545</td>
<td>0.06757</td>
<td>0.16129</td>
</tr>
<tr>
<td>1939</td>
<td>0.01786</td>
<td>0.02325</td>
<td>0.04545</td>
<td>0.12500</td>
<td>0.08849</td>
</tr>
<tr>
<td>1940</td>
<td>0.01786</td>
<td>0.02174</td>
<td>0.05000</td>
<td>0.08333</td>
<td>0.04167</td>
</tr>
<tr>
<td>1941</td>
<td>0.01818</td>
<td>0.03125</td>
<td>0.06250</td>
<td>0.09091</td>
<td>0.10869</td>
</tr>
<tr>
<td>1942</td>
<td>0.01587</td>
<td>0.02273</td>
<td>0.05882</td>
<td>0.09091</td>
<td>0.02500</td>
</tr>
<tr>
<td>1943</td>
<td>0.01695</td>
<td>0.02222</td>
<td>0.05000</td>
<td>0.08333</td>
<td>0.05556</td>
</tr>
<tr>
<td>1944</td>
<td>0.02273</td>
<td>0.02941</td>
<td>0.05263</td>
<td>0.13513</td>
<td>0.21739</td>
</tr>
<tr>
<td>1945</td>
<td>0.02174</td>
<td>0.02325</td>
<td>0.05882</td>
<td>0.05555</td>
<td>0.07353</td>
</tr>
<tr>
<td>1946</td>
<td>0.01666</td>
<td>0.02222</td>
<td>0.04545</td>
<td>0.05000</td>
<td>0.03831</td>
</tr>
<tr>
<td>1947</td>
<td>0.02631</td>
<td>0.02778</td>
<td>0.06250</td>
<td>0.06803</td>
<td>0.09524</td>
</tr>
<tr>
<td>1948</td>
<td>0.01658</td>
<td>0.02247</td>
<td>0.05135</td>
<td>0.05848</td>
<td>0.03831</td>
</tr>
<tr>
<td>1949</td>
<td>0.02127</td>
<td>0.02183</td>
<td>0.04608</td>
<td>0.08772</td>
<td>0.12315</td>
</tr>
<tr>
<td>1950</td>
<td>0.01972</td>
<td>0.02045</td>
<td>0.04762</td>
<td>0.10116</td>
<td>0.06097</td>
</tr>
<tr>
<td>1951</td>
<td>0.02114</td>
<td>0.02564</td>
<td>0.05494</td>
<td>0.12987</td>
<td>0.07353</td>
</tr>
<tr>
<td>1952</td>
<td>0.01488</td>
<td>0.02415</td>
<td>0.04219</td>
<td>0.09091</td>
<td>0.21276</td>
</tr>
<tr>
<td>Mean</td>
<td>0.02274</td>
<td>0.02770</td>
<td>0.05862</td>
<td>0.09244</td>
<td>0.09081</td>
</tr>
</tbody>
</table>
Table 4

Capital input coefficients (Dollars per bushel)

<table>
<thead>
<tr>
<th>Year</th>
<th>Corn</th>
<th>Oats</th>
<th>Soybeans</th>
<th>Flax</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1928</td>
<td>.35457</td>
<td>.26028</td>
<td>.93805</td>
<td>.76607</td>
<td>.72206</td>
</tr>
<tr>
<td>1929</td>
<td>.34672</td>
<td>.25722</td>
<td>1.02700</td>
<td>.75167</td>
<td>.71716</td>
</tr>
<tr>
<td>1930</td>
<td>.34100</td>
<td>.26081</td>
<td>.47630</td>
<td>.61500</td>
<td>.50543</td>
</tr>
<tr>
<td>1931</td>
<td>.38000</td>
<td>.26600</td>
<td>.59792</td>
<td>.76300</td>
<td>.41281</td>
</tr>
<tr>
<td>1932</td>
<td>.39114</td>
<td>.32682</td>
<td>.29068</td>
<td>.67075</td>
<td>1.06571</td>
</tr>
<tr>
<td>1933</td>
<td>.39002</td>
<td>.22361</td>
<td>.35956</td>
<td>.93377</td>
<td>.59687</td>
</tr>
<tr>
<td>1934</td>
<td>.28500</td>
<td>.54500</td>
<td>.89400</td>
<td>1.14857</td>
<td>1.70666</td>
</tr>
<tr>
<td>1935</td>
<td>.25932</td>
<td>.18644</td>
<td>.44950</td>
<td>.74662</td>
<td>1.02169</td>
</tr>
<tr>
<td>1936</td>
<td>.42640</td>
<td>.18450</td>
<td>.45150</td>
<td>.39565</td>
<td>.86300</td>
</tr>
<tr>
<td>1937</td>
<td>.26826</td>
<td>.17800</td>
<td>.35636</td>
<td>.59618</td>
<td>.89000</td>
</tr>
<tr>
<td>1938</td>
<td>.23528</td>
<td>.19786</td>
<td>.41591</td>
<td>.54662</td>
<td>1.33709</td>
</tr>
<tr>
<td>1939</td>
<td>.21911</td>
<td>.19977</td>
<td>.42772</td>
<td>1.03875</td>
<td>.76372</td>
</tr>
<tr>
<td>1940</td>
<td>.21643</td>
<td>.13609</td>
<td>.48600</td>
<td>.65667</td>
<td>.35583</td>
</tr>
<tr>
<td>1941</td>
<td>.22522</td>
<td>.26875</td>
<td>.58750</td>
<td>.74361</td>
<td>.92826</td>
</tr>
<tr>
<td>1942</td>
<td>.21492</td>
<td>.21773</td>
<td>.67647</td>
<td>.79636</td>
<td>.26025</td>
</tr>
<tr>
<td>1943</td>
<td>.23559</td>
<td>.23222</td>
<td>.57500</td>
<td>.82333</td>
<td>.57444</td>
</tr>
<tr>
<td>1944</td>
<td>.30841</td>
<td>.30823</td>
<td>.62579</td>
<td>1.35676</td>
<td>2.56404</td>
</tr>
<tr>
<td>1945</td>
<td>.30565</td>
<td>.25119</td>
<td>.71941</td>
<td>.59611</td>
<td>.81618</td>
</tr>
<tr>
<td>1946</td>
<td>.25933</td>
<td>.26289</td>
<td>.62318</td>
<td>.56650</td>
<td>.47509</td>
</tr>
<tr>
<td>1947</td>
<td>.42631</td>
<td>.38111</td>
<td>1.04375</td>
<td>1.20136</td>
<td>1.36762</td>
</tr>
<tr>
<td>1948</td>
<td>.33002</td>
<td>.31180</td>
<td>1.00108</td>
<td>.97836</td>
<td>.63946</td>
</tr>
<tr>
<td>1949</td>
<td>.45796</td>
<td>.31987</td>
<td>.92488</td>
<td>1.14382</td>
<td>1.95132</td>
</tr>
<tr>
<td>1950</td>
<td>.43649</td>
<td>.31697</td>
<td>1.03000</td>
<td>1.77292</td>
<td>1.01707</td>
</tr>
<tr>
<td>1951</td>
<td>.49725</td>
<td>.42538</td>
<td>1.20989</td>
<td>2.35195</td>
<td>1.30662</td>
</tr>
<tr>
<td>1952</td>
<td>.35402</td>
<td>.36537</td>
<td>.94051</td>
<td>1.51000</td>
<td>3.46170</td>
</tr>
</tbody>
</table>

Mean | .31772| .27870| .70812   | .96950| 1.00356
Table 5

Labor (yearly) input coefficients
(hours per bushel)

<table>
<thead>
<tr>
<th>Year</th>
<th>Corn</th>
<th>Oats</th>
<th>Soybeans</th>
<th>Flax</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1928</td>
<td>.29998</td>
<td>.19446</td>
<td>.72562</td>
<td>.65374</td>
<td>.55147</td>
</tr>
<tr>
<td>1929</td>
<td>.29169</td>
<td>.19446</td>
<td>.82000</td>
<td>.60831</td>
<td>.55973</td>
</tr>
<tr>
<td>1930</td>
<td>.26250</td>
<td>.18921</td>
<td>.38860</td>
<td>.45625</td>
<td>.40762</td>
</tr>
<tr>
<td>1931</td>
<td>.36204</td>
<td>.23331</td>
<td>.56941</td>
<td>.73000</td>
<td>.36945</td>
</tr>
<tr>
<td>1932</td>
<td>.37495</td>
<td>.31815</td>
<td>.40196</td>
<td>.68868</td>
<td>1.07145</td>
</tr>
<tr>
<td>1933</td>
<td>.20590</td>
<td>.21210</td>
<td>.35654</td>
<td>.94805</td>
<td>.58590</td>
</tr>
<tr>
<td>1934</td>
<td>.27625</td>
<td>.50001</td>
<td>.82000</td>
<td>1.04288</td>
<td>.44910</td>
</tr>
<tr>
<td>1935</td>
<td>.23866</td>
<td>.15554</td>
<td>.41000</td>
<td>.76044</td>
<td>.92595</td>
</tr>
<tr>
<td>1936</td>
<td>.42000</td>
<td>.17500</td>
<td>.41000</td>
<td>.79344</td>
<td>.75000</td>
</tr>
<tr>
<td>1937</td>
<td>.22827</td>
<td>.14000</td>
<td>.74546</td>
<td>.55721</td>
<td>.75000</td>
</tr>
<tr>
<td>1938</td>
<td>.19811</td>
<td>.16667</td>
<td>.37269</td>
<td>.49326</td>
<td>1.20967</td>
</tr>
<tr>
<td>1939</td>
<td>.18753</td>
<td>.16275</td>
<td>.37269</td>
<td>.91250</td>
<td>.66367</td>
</tr>
<tr>
<td>1940</td>
<td>.18753</td>
<td>.15218</td>
<td>.41000</td>
<td>.60831</td>
<td>.31253</td>
</tr>
<tr>
<td>1941</td>
<td>.19089</td>
<td>.21875</td>
<td>.51250</td>
<td>.66364</td>
<td>.81517</td>
</tr>
<tr>
<td>1942</td>
<td>.16663</td>
<td>.15911</td>
<td>.48232</td>
<td>.66364</td>
<td>.18750</td>
</tr>
<tr>
<td>1943</td>
<td>.17798</td>
<td>.15554</td>
<td>.41000</td>
<td>.60831</td>
<td>.41670</td>
</tr>
<tr>
<td>1944</td>
<td>.23866</td>
<td>.20587</td>
<td>.43157</td>
<td>.98645</td>
<td>1.63010</td>
</tr>
<tr>
<td>1945</td>
<td>.22827</td>
<td>.16275</td>
<td>.48232</td>
<td>.40551</td>
<td>.55148</td>
</tr>
<tr>
<td>1946</td>
<td>.17493</td>
<td>.15554</td>
<td>.37269</td>
<td>.36500</td>
<td>.28732</td>
</tr>
<tr>
<td>1947</td>
<td>.27625</td>
<td>.19446</td>
<td>.51250</td>
<td>.49662</td>
<td>.71430</td>
</tr>
<tr>
<td>1948</td>
<td>.17409</td>
<td>.15729</td>
<td>.44567</td>
<td>.42690</td>
<td>.28733</td>
</tr>
<tr>
<td>1949</td>
<td>.22334</td>
<td>.15261</td>
<td>.37786</td>
<td>.64036</td>
<td>.92587</td>
</tr>
<tr>
<td>1950</td>
<td>.20706</td>
<td>.11315</td>
<td>.39048</td>
<td>.76037</td>
<td>.45728</td>
</tr>
<tr>
<td>1951</td>
<td>.22197</td>
<td>.17948</td>
<td>.45051</td>
<td>.94805</td>
<td>.55147</td>
</tr>
<tr>
<td>1952</td>
<td>.15624</td>
<td>.16905</td>
<td>.34596</td>
<td>.66364</td>
<td>1.59570</td>
</tr>
</tbody>
</table>

Mean   | .23879| .19390| .48069   | .67518| .68108
The crops considered are: corn, oats, soybeans, flax, and wheat. Land, labor and capital are the three main categories of inputs.

The yield data for the period 1928-52 was taken from the reports of Iowa crop reporting service relevant to Ellsworth Township, Hancock County. Two very extreme entries for wheat yields for the years 1934 and 1941 were replaced by corresponding entries for the county. This data is shown in Table 1.

The capital expense per acre on the five crops for the same time period is shown in Table 2.

To compute the land input coefficient for a particular year and a particular crop, the corresponding information regarding yield per acre is divided into one. For example, if \( a \) bushels per acre is the yield for corn in a particular year, than \( 1/a \) is the land input coefficient for corn in that year. To get the capital input coefficient the figure for capital expense per acre is divided by the corresponding yield per acre. Similarly to get the yearly labor input coefficient for a particular crop in a particular year, the yearly labor requirement for that particular crop is divided by yield for acre for the corresponding year. The yearly labor requirements per acre were assumed to be as follows:
### Table 6

**Monthly Percentage Distributions of Yearly Labor Requirement**

<table>
<thead>
<tr>
<th>Month</th>
<th>Corn</th>
<th>Oats</th>
<th>Soybeans</th>
<th>Flax</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March</td>
<td>11.1</td>
<td></td>
<td>12.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>11.8</td>
<td>13.9</td>
<td>9.8</td>
<td>12.9</td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>22</td>
<td></td>
<td></td>
<td>24.3</td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>13.1</td>
<td></td>
<td>14.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>July</td>
<td>10.7</td>
<td>35.5</td>
<td>11</td>
<td>30.5</td>
<td>59.5</td>
</tr>
<tr>
<td>Aug.</td>
<td>39.5</td>
<td></td>
<td>44.5</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td>Sept.</td>
<td>2</td>
<td></td>
<td>2.9</td>
<td></td>
<td>27.8</td>
</tr>
<tr>
<td>Oct.</td>
<td>14.8</td>
<td></td>
<td>37.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov.</td>
<td>20.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec.</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Corn 10.5 hours per acre
Oats 7.0 hours per acre
Soybeans 8.2 hours per acre
Flax 7.3 hours per acre
Wheat 7.5 hours per acre

The land, capital, and yearly labor input coefficients thus computed are shown in Tables 3, 4 and 5.

Further, it was assumed that the yearly labor inputs have percentage distributions over the twelve months as shown in Table 6.

The following additional factors (labor per bushel) will be added to the corresponding monthly labor input coefficient for the respective months to take care of the additional burden due to a heavy harvest.

Table 7
Additional labor input coefficients

<table>
<thead>
<tr>
<th>Month</th>
<th>Corn</th>
<th>Oats</th>
<th>Soybeans</th>
<th>Flax</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>July</td>
<td>0</td>
<td>.00639</td>
<td>.00198</td>
<td>.00593</td>
<td>.01800</td>
</tr>
<tr>
<td>Aug.</td>
<td>0</td>
<td>.00711</td>
<td>0</td>
<td>.00865</td>
<td>0</td>
</tr>
<tr>
<td>Oct.</td>
<td>.00266</td>
<td>0</td>
<td>.00673</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nov.</td>
<td>.00367</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
It should be noticed, however, that this is not the only labor used in harvest. The corresponding percentage of the yearly labor during the harvest months is also meant for harvest. For example, in the case of corn, 20.4 percent of 10.5 hours per acre will be used in November mainly for harvest.

This information regarding capital and labor inputs was supplied by Dr. E. O. Heady and is based mainly on "Economic instability and choices involving income and risk in crop production."

The distributions of Table 6 are also supplied by Dr. E. O. Heady and are based on Iowa Agricultural Production Capacity Report, mimeograph, 1947, Ames.

A farm which has the following amounts of these input factors, is considered for purposes of drawing the production plan.

### Table 8

**Fixed resources**

<table>
<thead>
<tr>
<th>Land</th>
<th>Acres 148</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>$1800</td>
</tr>
</tbody>
</table>

Labor -- distributed over year as below.

<table>
<thead>
<tr>
<th>Month</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>182</td>
</tr>
<tr>
<td>February</td>
<td>182</td>
</tr>
<tr>
<td>March</td>
<td>182</td>
</tr>
<tr>
<td>April</td>
<td>182</td>
</tr>
<tr>
<td>May</td>
<td>182</td>
</tr>
<tr>
<td>June</td>
<td>23(\frac{1}{4})</td>
</tr>
<tr>
<td>July</td>
<td>23(\frac{1}{4})</td>
</tr>
<tr>
<td>August</td>
<td>23(\frac{1}{4})</td>
</tr>
<tr>
<td>September</td>
<td>182</td>
</tr>
<tr>
<td>October</td>
<td>182</td>
</tr>
<tr>
<td>November</td>
<td>182</td>
</tr>
<tr>
<td>December</td>
<td>182</td>
</tr>
</tbody>
</table>
This, again, was suggested by Dr. E. O. Head as a model family farm in Iowa.

It appeared in the first instant that one would have to consider these fixed inputs separately, but by a close examination of the percentage distributions given in Table 4, it is noticed that July and August are the months which require maximum labor. As a first trial, it was decided to draw the optimum program with the help of the simplex method with only four limitational factors: land, capital, and labor in the month of July and in the month of August. It was hoped that the plan of production, which will not need an amount of labor more than the available one in these months which require heaviest labor, will not require more labor than what is available in the other months. We assume, of course, that the monthly proportions are constant for the scale of production. The maximum labor requirement for corn falls in the month of May and that for soybeans in the month of October. Since the other three crops require no labor in these two months, it was hoped that if the plan contains some of these three crops, the labor requirements for these months will not exceed the available labor. As we will see later, that this hope was not realized.

The matrix of average input coefficients is as below:
The prices used for the profit function are assumed to be constant. That is, the average Iowa prices for the five crops during 1952 are assumed to prevail for the year for which the production is being planned. This was assumed for simplicity. However, the theoretical treatment of Chapter IV can take into account the price variability, if prices could be justifiably assumed to be normally distributed. The following fixed prices were used.  

---

1Iowa Cooperative Crop and Livestock Reporting Service, Crop and Livestock News, 12, No. 1-12, 1952.
Crop | Price (in dollars per bu.)
--- | ---
Corn | 1.56
Oats | .84
Soybeans | 2.79
Flax | 3.81
Wheat | 2.14

It should be noticed that by using this projection of 1952 average prices for the year under consideration, we do not mean to imply that this is the best way to do it. The prices for the relevant year may be predicted separately by the naive procedures referred to earlier or by other econometric techniques.

Thus the problem for solution was to find the amounts $x_1, x_2, x_3, x_4,$ and $x_5$ of the corn, oats, soybeans, flax, and wheat respectively which maximize the linear function:

$$(1.56)x_1 + (.84)x_2 + (2.79)x_3 + (3.81)x_4 + (2.14)x_5$$

subject to the inequalities

$$(.02274)x_1 + (.02770)x_2 + (.05862)x_3 + (.09249)x_4 + (.09081)x_5 \leq 148$$
\[
\begin{align*}
(.31772)x_1 + (.27870)x_2 + (.70812)x_3 \\
+ (.96956)x_4 + (1.00356)x_5 & \leq 1800 \\
(.02555)x_1 + (.07523)x_2 + (.05485)x_3 \\
+ (.21186)x_4 + (.42324)x_5 & \leq 234 \\
0 \times_1 + (.03370)x_2 + 0 \times_3 \\
+ (.30910)x_4 + (.08650)x_5 & \leq 234,
\end{align*}
\]

These were changed to equalities by the introduction of four disposal activities \(x_6, x_7, x_8, \) and \(x_9\) corresponding to the four functions. Using the simplex method as explained by Charnes, Cooper, and Henderson, the following optimum solution was obtained.

\[
\begin{align*}
x_1 & = 5666.7 \\
x_6 & = 19.13812 \\
x_8 & = 99.184 \\
x_9 & = 234
\end{align*}
\]

But when this solution is checked for the labor requirements in other months, it is found that labor requirements for the months of May is 298 hours whereas the labor at the disposal of the firm is 182 hours only. Thus the program is not workable. Therefore, it was essential to consider the labor equation for the month of May. We introduce another restriction

\[(.05253)x_1 + 0x_2 + (.11681)x_3 + 0x_4 + 0x_5 \leq 182\]

still anticipating that the optimum plan thus obtained will not require labor more than the amount available in the month of October. In fact, if we had considered May, July, August and October labor requirements, there was no need of this anticipation, yet to save labor and for the simplification of the model minimum restrictions were brought in.

The new solution\(^1\), thus obtained, was:

\[x_1 = 3464.49718\]
\[x_4 = 686.73761\]
\[x_6 = 5.70277\]
\[x_7 = 33.44223\]
\[x_9 = 21.73124\]

Thus we should plan to produce 3464.5 bushels of corn and 687 bushels of flax which will require the inputs of Table 10a.

\(^1\)The full process is exhibited in the Appendix.
Table 10a
Schedule of inputs

<table>
<thead>
<tr>
<th></th>
<th>Corn</th>
<th>Flax</th>
<th>Disposal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Land in Acres</strong></td>
<td>78.78</td>
<td>63.52</td>
<td>5.70</td>
<td>148</td>
</tr>
<tr>
<td><strong>Capital in $</strong></td>
<td>1100.73</td>
<td>665.83</td>
<td>33.44</td>
<td>1800</td>
</tr>
<tr>
<td><strong>Labor in</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>182</td>
<td></td>
<td></td>
<td>182</td>
</tr>
<tr>
<td>July</td>
<td>88.51</td>
<td>145.49</td>
<td></td>
<td>234</td>
</tr>
<tr>
<td>Aug.</td>
<td>212.27</td>
<td>21.73</td>
<td></td>
<td>234</td>
</tr>
</tbody>
</table>

Now we check for the requirements of labor for the rest of the year for this program. The following are the average labor input coefficients for corn and flax for the rest of the months of the year. These are computed from early input coefficients given in Table 5, in proportion to the percentage distributions given in Table 6 with additional amount of .00266 per bushel for the month of October and .00367 per bushel for the month of November, for corn input coefficients since these are the harvest months.
Table 10b
Labor input coefficients
(not previously considered)

<table>
<thead>
<tr>
<th>Month</th>
<th>Corn</th>
<th>Flax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March</td>
<td></td>
<td>0.08169</td>
</tr>
<tr>
<td>April</td>
<td>0.02818</td>
<td>0.08710</td>
</tr>
<tr>
<td>June</td>
<td>0.03128</td>
<td></td>
</tr>
<tr>
<td>Sept.</td>
<td>0.00417</td>
<td></td>
</tr>
<tr>
<td>Oct.</td>
<td>0.03800</td>
<td></td>
</tr>
<tr>
<td>Nov.</td>
<td>0.05238</td>
<td></td>
</tr>
<tr>
<td>Dec.</td>
<td>0.01242</td>
<td></td>
</tr>
</tbody>
</table>

We can get the labor requirements (Table 10c) for our program during the corresponding months with the help of the above input coefficients.

It is apparent from Table 10c that January and February are the most idle months whereas November may be the busiest.
Table 10c
Schedule of inputs

<table>
<thead>
<tr>
<th>Month</th>
<th>Corn</th>
<th>Flax</th>
<th>Disposal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan.</td>
<td></td>
<td></td>
<td>182</td>
<td>182</td>
</tr>
<tr>
<td>Feb.</td>
<td></td>
<td></td>
<td>182</td>
<td>182</td>
</tr>
<tr>
<td>March</td>
<td>97.63</td>
<td>56.10</td>
<td>125.90</td>
<td>182</td>
</tr>
<tr>
<td>April</td>
<td>97.63</td>
<td>59.81</td>
<td>24.56</td>
<td>182</td>
</tr>
<tr>
<td>June</td>
<td>108.37</td>
<td></td>
<td>125.63</td>
<td>234</td>
</tr>
<tr>
<td>Sept.</td>
<td>16.53</td>
<td></td>
<td>165.47</td>
<td>182</td>
</tr>
<tr>
<td>Oct.</td>
<td>131.65</td>
<td></td>
<td>50.35</td>
<td>182</td>
</tr>
<tr>
<td>Nov.</td>
<td>181.47</td>
<td></td>
<td>.53</td>
<td>182</td>
</tr>
<tr>
<td>Dec.</td>
<td>43.03</td>
<td></td>
<td>138.97</td>
<td>182</td>
</tr>
</tbody>
</table>

It may be noticed that we have made the production plan which is expected to maximize the gross monetary returns. However, since the cost on inputs is assumed to be fixed, therefore, the program is expected to maximize the net profit also.
VII. STATISTICAL PREDICTIONS ABOUT THE RESULTS OF THE PROGRAM

7.1 A preliminary note.

We have outlined the production program in the preceding chapter. We should notice that quantities of $x_1$, $x_4$, $x_6$, $x_7$, and $x_9$, which are shown in the schedule of the program, are a solution of the following 5 linear equations.

\[
\begin{align*}
(0.02774)x_1 + (0.09249)x_4 + (1)x_6 + (0)x_7 + (0)x_9 &= 148 \\
(0.31772)x_1 + (0.96956)x_4 + (0)x_6 + (1)x_7 + (0)x_9 &= 1800 \\
(0.02555)x_1 + (0.21186)x_4 + (0)x_6 + (0)x_7 + (0)x_9 &= 234 \\
(0)x_1 + (0.30910)x_4 + (0)x_6 + (0)x_7 + (1)x_9 &= 234 \\
(0.05253)x_1 + (0)x_4 + (0)x_6 + (0)x_7 + (0)x_9 &= 182.
\end{align*}
\]

The problem is: the input coefficients are liable to vary during the operation of the program, what will be the variability of the outcomes? Now we will use the results of Chapter IV to answer that question. We assume that the operating input coefficients are normally distributed about the corresponding values used in planning the program, as mean values. From Tables III and IV we can get unbiased estimates of the variances of the input coefficients.
pertaining to land and capital. To get the variance of the labor input coefficients for any crop pertaining to any particular month, we multiply the corresponding variance from the yearly input coefficients by the square of the corresponding percentage of yearly labor requirement.¹ For example, to get the variance of the labor input coefficient for flax for the month of July we take the variance of yearly labor input coefficients for flax which is (.035087) and multiply it with $(30.5/100)^2$, getting .003263. We will get the same result if we separately write the series of input coefficients for the month of July and then compute the variances. We will see in the following pages that we need only three variances which are obtained as suggested above.

1) Variance of the labor input coefficient for July for flax which is (.003262968)

2) Variance of the labor input coefficient for May for corn which is found to be (.000227125)

3) Variance of the labor input coefficient for July for corn which is found to be (.000053646).

¹We are using the well-known property that the variance of $ax$ is equal to $a^2$ times the variance of $x$. 
Table 11
Monthly labor input coefficients
(hours per bushel)

<table>
<thead>
<tr>
<th>Year</th>
<th>July Corn</th>
<th>July Flax</th>
<th>May Corn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1928</td>
<td>.03210</td>
<td>.19878</td>
<td>.06599</td>
</tr>
<tr>
<td>1929</td>
<td>.03121</td>
<td>.18553</td>
<td>.06417</td>
</tr>
<tr>
<td>1930</td>
<td>.02809</td>
<td>.13916</td>
<td>.05775</td>
</tr>
<tr>
<td>1931</td>
<td>.03074</td>
<td>.22265</td>
<td>.07965</td>
</tr>
<tr>
<td>1932</td>
<td>.04012</td>
<td>.21004</td>
<td>.08249</td>
</tr>
<tr>
<td>1933</td>
<td>.02203</td>
<td>.28915</td>
<td>.04530</td>
</tr>
<tr>
<td>1934</td>
<td>.02956</td>
<td>.31807</td>
<td>.06077</td>
</tr>
<tr>
<td>1935</td>
<td>.02554</td>
<td>.23193</td>
<td>.05250</td>
</tr>
<tr>
<td>1936</td>
<td>.04494</td>
<td>.24199</td>
<td>.09240</td>
</tr>
<tr>
<td>1937</td>
<td>.02412</td>
<td>.16995</td>
<td>.05022</td>
</tr>
<tr>
<td>1938</td>
<td>.02120</td>
<td>.15045</td>
<td>.04359</td>
</tr>
<tr>
<td>1939</td>
<td>.02007</td>
<td>.27831</td>
<td>.04126</td>
</tr>
<tr>
<td>1940</td>
<td>.02007</td>
<td>.18554</td>
<td>.04126</td>
</tr>
<tr>
<td>1941</td>
<td>.02042</td>
<td>.20241</td>
<td>.04199</td>
</tr>
<tr>
<td>1942</td>
<td>.01783</td>
<td>.20241</td>
<td>.03666</td>
</tr>
<tr>
<td>1943</td>
<td>.01904</td>
<td>.18554</td>
<td>.03915</td>
</tr>
<tr>
<td>1944</td>
<td>.02554</td>
<td>.30086</td>
<td>.05251</td>
</tr>
<tr>
<td>1945</td>
<td>.02412</td>
<td>.12368</td>
<td>.05022</td>
</tr>
<tr>
<td>1946</td>
<td>.01872</td>
<td>.11133</td>
<td>.03848</td>
</tr>
<tr>
<td>1947</td>
<td>.02956</td>
<td>.15147</td>
<td>.06077</td>
</tr>
<tr>
<td>1948</td>
<td>.01863</td>
<td>.13020</td>
<td>.03830</td>
</tr>
<tr>
<td>1949</td>
<td>.02390</td>
<td>.19531</td>
<td>.04913</td>
</tr>
<tr>
<td>1950</td>
<td>.02215</td>
<td>.23191</td>
<td>.04555</td>
</tr>
<tr>
<td>1951</td>
<td>.02375</td>
<td>.28915</td>
<td>.04833</td>
</tr>
<tr>
<td>1952</td>
<td>.01672</td>
<td>.20241</td>
<td>.03437</td>
</tr>
<tr>
<td>Means</td>
<td>.02555</td>
<td>.20593</td>
<td>.05253</td>
</tr>
</tbody>
</table>
7.2 Examination of the input coefficients series.

We may examine the relevant series of the input coefficients for the period 1928-1952 to see how far they are random and how far the assumption of normality is justified. Table 11 shows these time series. First of all, a close examination of the time series shows that the series are pretty much random. Next, we form frequency distributions of these series and notice they can be considered as approximately normal.

<table>
<thead>
<tr>
<th>July-Labor-Corn Input coefficient range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>.015 - .020</td>
<td>5</td>
</tr>
<tr>
<td>.020 - .025</td>
<td>10</td>
</tr>
<tr>
<td>.025 - .030</td>
<td>5</td>
</tr>
<tr>
<td>.030 - .035</td>
<td>2</td>
</tr>
<tr>
<td>.035 - .040</td>
<td>1+1 (including .0402)</td>
</tr>
<tr>
<td>.040 - .045</td>
<td>1</td>
</tr>
</tbody>
</table>
### July-Labor-Flax

<table>
<thead>
<tr>
<th>Input coefficient range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10 - .15</td>
<td>4</td>
</tr>
<tr>
<td>.15 - .20</td>
<td>8</td>
</tr>
<tr>
<td>.20 - .25</td>
<td>8</td>
</tr>
<tr>
<td>.25 - .30</td>
<td>3</td>
</tr>
<tr>
<td>.30 - .35</td>
<td>2</td>
</tr>
</tbody>
</table>

### May-Labor-Corn

<table>
<thead>
<tr>
<th>Input coefficient range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>.03 - .04</td>
<td>5</td>
</tr>
<tr>
<td>.04 - .05</td>
<td>8</td>
</tr>
<tr>
<td>.05 - .06</td>
<td>5</td>
</tr>
<tr>
<td>.06 - .07</td>
<td>4</td>
</tr>
<tr>
<td>.07 - .08</td>
<td>1</td>
</tr>
<tr>
<td>.08 - .09</td>
<td>1</td>
</tr>
<tr>
<td>.09 - .10</td>
<td>1</td>
</tr>
</tbody>
</table>

These cumulative distributions are not presented in an attempt to justify the assumption of normality. They give a rough idea of the distributions. Roughly speaking, these distributions give higher frequency closer to the means and lesser frequency if we move away from the means. Also we notice that the frequency groups adjacent to the modal group or groups give same frequency. Therefore, we may
retain the assumption of normality to exhibit the working of theory developed in Chapter IV. Alternatively, we could have used the mean and the variance computed from these cumulative frequency distributions, to make the program and then to make statistical predictions. It may be noticed that the procedure adopted by us, that is, to use ordinary arithmetic averages to plan the program and ordinary sample variances to make statistical predictions, is not necessarily the best in all cases. Nor is it used with implication that this is the only workable procedure. In fact, in another problem, a more detailed examination and expert advice may determine the expected values of the input coefficients and possible estimates of variances. In this case, the analysis of Chapter IV will give more accurate limits. However, since our object is to exhibit, as an example, the working of the theory of statistical predictions about the outcomes, we took a simple course.

7.3 Preliminary analysis for predictions.

Now we apply the results of Section 4.4 of Chapter IV to determine the confidence limits of the corn yield ($x_1$) and the flax yield ($x_4$).

We notice from 7.1, equation (7.1.1) that $x_1 = \frac{\bar{x}_1}{\beta}$ and

$$x_2 = \frac{\bar{x}_4}{\beta}$$  \hspace{1cm} (7.3.1)
where

$\theta_1 =
\begin{bmatrix}
148 & .09249 & 1 & 0 & 0 \\
1800 & .96956 & 0 & 1 & 0 \\
234 & .21186 & 0 & 0 & 0 \\
234 & .30910 & 0 & 0 & 1 \\
182 & 0 & 0 & 0 & 0
\end{bmatrix}
$

$= 182 \cdot .21186 = 38.55352$ \hspace{1cm} (7.3.2)

$\theta_4 =
\begin{bmatrix}
.02274 & 148 & 1 & 0 & 0 \\
.31772 & 1800 & 0 & 1 & 0 \\
.02555 & 234 & 0 & 0 & 0 \\
.0 & 234 & 0 & 0 & 1 \\
.05253 & 182 & 0 & 0 & 0
\end{bmatrix}
$

$= 182 \cdot .02555 - 234 \cdot .05253 = 7.64192$ \hspace{1cm} (7.3.3)

$\beta =
\begin{bmatrix}
.02274 & .09249 & 1 & 0 & 0 \\
.31772 & .96956 & 0 & 1 & 0 \\
.02555 & .21186 & 0 & 0 & 0 \\
.0 & .30910 & 0 & 0 & 1 \\
.05253 & 0 & 0 & 0 & 0
\end{bmatrix}
$

$= .05253 \cdot .21186 = .0111290059.$ \hspace{1cm} (7.3.4)
We notice also that these quantities $\alpha_1$ and $\alpha_4$ and $\beta$ correspond to the formula (4.1.6) and (4.1.8). Now, using formulae (4.1.7) and (4.1.9) we get the variances $\sigma^2$ for numerators and $\sigma_B^2$ for the denominator.

That is

$$\sigma_{(1)}^2 = (.0032629706)(182)^2$$
$$= 106.108263804$$ (7.3.5)

$$\sigma_{(4)}^2 = (.0002271250)(234)^2$$
$$+ (.0000536460)(182)^2$$
$$= 1.42134266808$$ (7.3.6)

also from formula (4.1.10) we get the covariances of the numerators and the denominators.

$$\sigma_{B1} = (.05253)(182)(.0032629706)$$
$$= (.0311954999)$$ (7.3.7)

$$\sigma_{B4} = (.21186)(234)(.0002271250)$$
$$= (.01125977619)$$ (7.3.8)

First of all, we notice that the assumption $\beta > 3 \sigma_B$ for the application of results of sections (4.3) and (4.4) is not satisfied since

$$3 \cdot \sigma_B = .013143$$

and

$$\beta = .011129.$$
However, we notice that $\beta < 2.54 \sigma_B$. Now we go back and check the reason which necessitated this assumption. We check Geary's article\(^1\) and see that in deriving the distribution of a quotient $x/y$ we are concerned that $y$ may assume negative values. We say (since $y$ is a normal variate) if $E(y) > 3 \sigma_y$ then it is very unlikely that $y$ will assume negative values. But even if $\beta > 3 \sigma_B$, there is some small probability, (.00135) that the variable in the denominator may become negative. On the other hand, if $E(y) > 2.54 \sigma_y$, this probability is (.0055). Therefore, if we could tolerate a probability of .00135, we can also ignore a probability of .0055 and say that the denominator is not likely to become negative. Hence, at least for our purpose, to exhibit the working of probability approach, we may relax the assumption of $\beta > 3 \sigma_B$ and proceed with our analysis.

7.4 Confidence limits for the activity levels.

Now we apply the equation (4.4.4) of Chapter IV to get the confidence limits. Suppose we want 5 per cent confidence limits. Then the probability that absolute value of a standardized normal variate $z$ is greater than 1.96 is .05; therefore, we take $\gamma = 1.96$ and compute

Therefore, the roots of the following equation gives the 95 per cent confidence limits for corn yield.

\[(0.000501)x^2 - (0.6185542733)x + (1071.549201) = 0.\]  

The roots are 2088 and 10282, approximately that is, the

\[P[2088 < x < 10282] = .95.\]  

A comparison of these limits with the expected value 3464.5 bushels of corn in our schedule shows that the distribution
of corn yield is skew on the right. Since the lower limit is the one which any planning unit is mostly concerned with, 2088 is a reasonable lower limit.

We should also notice that the equation (4.4.4) is derived from the formula (4.4.1), that is

\[
P \left[ \left| \frac{E(D) - E(N)}{\sqrt{\sigma_D^2 - 2Z\sigma_{ND} + \sigma_N^2}} \right| \leq \gamma \right] = a
\]

with equal probabilities on both ends of the normal distribution, but since the distribution of \( x_i \), that is, the corn yield is apparently skew to the right we would have gotten better limits if we could use, say, a probability of 0.01 on left-hand tail and a probability of 0.04 on the right-hand tail. This can be done with the help of Section 4.6, using the cumulative distribution.

To get the confidence limits for flax yield, we get the roots of the equation

\[
(.0000501)Z^2 - (.0835827403)Z + (3.774664166) = 0 . \tag{7.4.6}
\]

The roots are 14.65 and 16.25, whereas the expected value was 686.7. This shows that

1) The flax yields involve much more variation. This is also apparent if we look at the data of average yields for the period 1928-1952, in Table 1.
2) The distribution of flax yield \((x_4)\) is also skew to the right though to a lesser degree than that of corn.

Before the discussion on the confidence limits of activity levels is closed, a very important fact needs emphasis. It should be noticed that the variables \(x_1\) and \(x_4\) are not independently distributed, so it is very unlikely that the extreme limits for both activities will be raised simultaneously. Even if they were independent, the probability of a simultaneous occurrence of the lowest extremes will be the product of the individual probabilities which, therefore, will be very small. A rigorous analysis of the probability of joint occurrences needs a study of the joint distribution of the activity levels which is very tedious. However, an indirect inference can be made by considering the limits of the profit function since it is a linear function of those activity levels.

7.5 **Confidence limits of the profit function.**

The expected profit 

\[
y = (1.56)(346.5)+(3.81)(686.7376) = 8021
\]

(7.5.1)

which is equal to 

\[
\frac{dy}{\beta}
\]
where
\[ dy = (1.56)(38.55852) + (3.81)(7.64192) = 89.267 \quad (7.5.2) \]
(assuming (4.1.10)) and \( \beta = .011129 \) \quad (7.5.3)
as in Section 7.3.

Assuming prices to be constant, we have from (4.1.11),
\[ \sigma_N^g(y) = (1.56)^2(108.082638) + (3.87)^2(14.2134266) \]
\[ = 469.3534299 \quad (7.5.4) \]
and
\[ \sigma_B^g = .000019198 \quad (7.5.5) \]
as in Section 7.3.

Also
\[ \sigma_{BN}(y) = (1.56)(.311954999) + (3.81)(.0112497764) \]
\[ = .09156427928 \quad \text{(using equation } (4.1.12)) \quad (7.5.6) \]

Therefore,
\[ (\beta^2 - \gamma^2 \sigma_B^2) = .0000501 \text{ as before} \]
\[ 2[\beta dy - \gamma^2 \sigma_{BN}(y)] \]
\[ = 2(.011129)(89.267) - (1.96)^2(.09156473) \]
\[ = 1.2833947666. \quad (7.5.7) \]
and
\[ (\delta y^2 - \gamma^2 \sigma_N^2(y)) \]
Therefore, the roots of the following equation gives the 95 per cent confidence limits

\[(0.000501)^2 - (1.2833947686)z + 6165.529102 = 0\]

and the roots are 6297.95 and 18866.65 or

6298 and 18866.

We compare these with the expected profit, $8021.

(i) It is noticed that the distribution of the profit function seems skew to the right, therefore, confidence limits with equal probability on both ends are not efficient. However, since the lower limit is the one we are most concerned with, the results seem very satisfactory.

(ii) Also, we notice that if the lower limits of both corn and flax were realized simultaneously, the profit would have been

\[(1.56)(2088) + (3.81)(46.5) = 3434\]

which seems very unlikely. In fact, in the following section, we will see that the probability that the profit is less than $5000 is practically zero. Before closing this discussion on confidence limits, it is necessary to emphasize that the whole analysis used above is approximate and therefore the limits are approximate, too.
7.6 **Statistical predictions based on the cumulative distributions.**

Now we will exhibit the application of Section 4.6 of Chapter IV.

(1) First of all, we investigate the probability that corn yield may be less than 2000 bushels. Then using the notations of Section 4.6, we have

\[ V = 2000 \quad (7.6.1) \]

\[ h = \frac{\beta}{\sigma_B} = \frac{0.011129}{0.004381} = 2.54 \quad (7.6.2) \]

and

\[ k_1 = \frac{k_1}{\sigma_1} = \frac{(2000)\beta}{\left[ \sigma_1^2 - 2\sigma_{B1} - (2000) + (2000)^2 \sigma_B^2 \right]^{1/2}} \]

\[ = \frac{38.55852 - (2000)(0.011129)}{[108.082638 - 2(0.0311954999)(2000) + (2000)^2(0.000019198)]^{1/2}} \]

\[ = 2.01 \quad (7.6.3) \]

\[ \rho = \frac{\sigma_{B1} - (2000)\sigma_B^2}{\left[ \sigma_{D1}^2 - 2(2000)\sigma_{B1} + (2000)^2 \sigma_B^2 \right]^{1/2}} \]

\[ = \frac{0.0311954999 - (2000)(0.000019198)}{(0.004381)([108.082638 - 2(0.0311954999)(2000) + (2000)^2(0.000019198)]^{1/2}} \]

\[ = 0.212 \quad (7.6.4) \]
Therefore, the probability required, that is

\[ P\left[ x_1 \geq 2000 \right] = \left[ \int_{-h}^{h} \int_{-k_1}^{k_1} N(0,0,1,1,1,1) \, dx \, dy \right] + \int_{-h}^{h} \int_{-k_1}^{k_1} N(0,0,1,1,1,1) \, dx \, dy \]

where \( N(0,0,1,1,1,1,1,1) \), is standardized normal bivariate distribution with \( \rho = 0.212 \).

To evaluate these integrals, we refer to Pearson's tables for statisticians and biometricians\(^1\) and, using the necessary interpolation, we get the required probability

\[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N(\rho = 0.212) \, dx \, dy = 0.000239 \]

and

\[ \int_{-h}^{h} \int_{-k}^{k} N(\rho = 0.212) = 1 - \int_{-h}^{h} N(0,1) \, dx \, dy - \int_{-k}^{k} N(0,1) \, dy \]

\[ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N(\rho) \, dx \, dy \]

\[ = 1 - (0.0055) - (0.0222) + 0.0000239 \]

\[ = 0.9723239 \]

\(^1\)Pearson, Karl. Tables for Statisticians and Biometricians, Part II, Tables VIII and IX, Cambridge Univ. Press.
therefore

\[ P(x_1 \geq 2000) = 0.9723479 \] \hspace{1cm} (7.6.5)

therefore, the probability that

\[ (x_1 \leq 2000) = 0.0277 \] \hspace{1cm} (7.6.6)

which approximately checks with the lower limit obtained in preceding section.

Similarly we investigate the probability that corn yield may be greater than (10,000).

As before

\[ h = 2.54 \]

\[ k_1 = \frac{a_1 - (10,000)h}{\left[ \sigma_1^2 - 2\sigma_{B1}(10,000) + (10,000)^2 \sigma_{B1}^2 \right]^{1/2}} \]

\[ = \frac{(38.558852) - (10,000)(0.01129)}{[108.082638] - 2(0.0311954999)(10,000) + (10,000)^2(0.00009193)]^{1/2}} \]

\[ = -1.94 \] \hspace{1cm} (7.6.7)

\[ \rho = \frac{\sigma_{B1} - (10,000)\sigma_B}{\sigma_B(\sigma_1^2 - 2\sigma_{B1}(10,000) + (10,000)^2 \sigma_{B1}^2)^{1/2}} \]
\[
\frac{(.0311954999)-(10,000)(.000019198)}{(.004387)[(108.082638)-2(.311954999)(10,000) + (10,000)(.000019198)]}
\]

\[= -.97 \quad (7.6.8)\]

Therefore, consulting the tables as before, we have

\[
\int_{-2.54}^{2.54} \int_{-1.94}^{1.94} N(\rho = -0.97) \, dx \, dy = \int_{-2.54}^{2.54} N(0,1) \, dx
\]

\[
= 0.0055 - 0.005465
\]

\[= 0.000035\]

And

\[
\int_{-2.54}^{2.54} \int_{1.94}^{1.94} N(\rho = -0.97) \, dx \, dy = \int_{-1.94}^{1.94} N(0,1) \, dy
\]

\[= \int_{2.54}^{2.54} (\rho = 0.97) \, dx \, dy\]
and, therefore,

\[ P \{x_i \geq 10,000\} = .020735 + .000035 \]

\[ = .02077. \quad (7.6.9) \]

The probability is about 2 per cent which again checks approximately the corresponding result of the preceding section.

(ii) Flax yield.

Now we try to derive the probability statements for the yield of flax. Suppose we investigate the probability that the yield of flax will not be less than 50 bushel. Then

\[ v = 50 \]

\[ h = \frac{b}{\sigma_B} = \frac{.011129}{.004381} = 2.54 \]

\[ k = \frac{\sigma^2 - (50)^2}{(\sigma^2 - 2\sigma_B)(50) = (50)^2 \sigma_B^2} \]

\[ = \frac{7.64192 - (50)(.011129)}{[(14.2134266 - 2(50)(.0112597764) + (50)^2(.000019198)]^{1/2}} \quad (7.6.10) \]
\[ \rho = \frac{\sigma_{\text{pl}} - (\bar{Y}) \sigma^2}{\sigma^2 B \left[ \sigma^2 B - 2(\bar{Y}) \sigma_{\text{pl}}^2 + (\bar{Y})^2 \sigma^2 B \right]^{1/2}} \]

\[ = \frac{(0.0112497764) - (\bar{Y})(0.000019198)}{(0.004381)[(1.42134266) - 2(\bar{Y})(0.0112497764) + (\bar{Y})^2(0.000019198)]} \]

\[ = .65 \quad \text{(7.6.11)} \]

and from tables we have

\[ \int_{2.54}^{1.95} \int_{1.95}^{\infty} N(\rho=.65) \, dx \, dy = .002516 \]

and

\[ \int_{-2.54}^{-1.95} \int_{-1.95}^{\infty} N(\rho=.65) \, dx \, dy \int_{2.54}^{\infty} N(0,1) \, dx \]

\[ - \int_{1.95}^{\infty} N(0,1) \, dy + \int_{1.95}^{\infty} \int_{2.54}^{1.95} N(\rho=.65) \, dx \, dy \]

\[ = 1 - .0055 - .0256 + .002516 \]

\[ = .771416 \]
Therefore

\[ P \left[ X_4 \geq 50 \right] = 0.971416 + 0.002516 \]
\[ = 0.973 \]  \hspace{1cm} (7.6.12)

and hence

\[ P \left[ X_4 < 50 \right] = 0.027 \]  \hspace{1cm} (7.6.13)

which approximately checks with lower the limit obtained in the preceding section.

To check the upper limit, let us investigate the probability that flax production may be higher than 1600 bushels.

\[ v = 1600 \]

\[ h = \frac{\beta}{\sigma_B} = 2.54 \text{ as before} \]

\[ k_4 = \frac{3 - (1600)\beta}{\left[ \sigma^2 - 2\sigma_B (1600) + (1600)^2 \sigma_B^2 \right]^{1/2}} \]

\[ = \frac{7.64192 - (1600)(.011129)}{[(14.2134266) - 2(1600)(.0112597764) + (1600)^2(.000019198)]^{1/2}} \]

\[ = -1.94 \]  \hspace{1cm} (7.6.14)

\[ \rho = \frac{\sigma_{Bh} - (1600) \sigma_B^2}{\sigma_B \left[ \sigma^2 - 2(1600) \sigma_B + (1600)^2 \sigma_B^2 \right]^{1/2}} \]
and consulting the tables we have

\[
\int_{-1.94}^{2.54} \int_{-1.94}^{2.54} N(\rho = -0.85) \, dx \, dy = \int_{-2.54}^{2.54} \int_{1.94}^{2.54} N(\rho = -0.85) \, dx \, dy
\]

\[
= \frac{(0.0112597764) - (1600)(0.000019198)}{0.004381[(14.2134\, 266) - 2(1600)(0.0112597764) + (1600)^2(0.000019198)]^{1/2}}
\]

\[
= -0.85
\]

(7.6.15)

and

\[
= \int_{1.94}^{2.54} \int_{1.94}^{2.54} N(\rho = -0.85) \, dx \, dy
\]

\[
= 0.055 - 0.04450
\]

= 0.00105

and

\[
= \int_{2.54}^{\infty} \int_{-1.94}^{2.54} N(\rho = -0.85) \, dx \, dy
\]

\[
= \int_{1.94}^{\infty} N(0,1) \, dy - \int_{1.94}^{2.54} \int_{1.94}^{2.54} N(\rho = 0.85) \, dx \, dy
\]

\[
= 0.0262 - 0.04450
\]

= 0.02175
and hence,

\[ P \left( X_4 > 1600 \right) = 0.00105 + 0.02175 \]
\[ = 0.0228 . \quad (7.6.16) \]

This checks approximately with the corresponding upper limit obtained in the preceding section.

(iii) Profit function.

In the same manner, we deal with profit function \( y \), and find the probability that profit will not be less than $6500. We again apply Section 4.10 of Chapter IV. In this case,

\[ v = 6500 \]

\[ h = \frac{\beta}{\sigma_B} = \frac{0.011129}{0.004381} = 2.54 \]

\[
\frac{\partial y - (6500) \beta}{\sqrt{\frac{\sigma_N(y)}{\sigma_{BN}(y)} - 2(6500)\sigma_{BN}(y) + (6500)^2 \sigma_B^2}}^{1/2}
\]

and substituting the values from Section 5(c), we get

\[
k = \frac{89.267 - (6500)(0.011129)}{[469.353(1)3 - 2(6500)(0.0915647279) + (6500)^2(0.000019198)]^{1/2}} \]
\[ = 2.11 \quad (7.6.17) \]
and
\[
\rho = \frac{\sigma_{BN}(y) - (6500) \sigma_B^a}{\sigma_B \left[ \sigma_{N}(y) - 2(6500) \sigma_{BN}(y) + (6500)^a \sigma_B^a \right]^{1/a}}
\]
\[
= \frac{(915647279)-(6500)(.000019198)}{[(.004381)(469.353443)-2(.0915647279) + (6500)^a(.000019198)]^{1/a}}
\]
\[= .17 \quad (7.6.18)\]

and consulting the tables, we have
\[
\int_{-\infty}^{\infty} \int_{-2.11}^{2.54} N(p=-.17) \, dx \, dy = .0000251
\]
and
\[
\int_{-\infty}^{\infty} \int_{-2.11}^{2.54} N(p=-.17) \, dx \, dy = 1 - \int_{-\infty}^{\infty} N(0,1) \, dx
\]
\[
= 1 - .0055 - .0222 + .0000251
\]
\[= .9723502\]

and therefore,
\[
P \left[ y \geq 6500 \right] = .9723502 + .0000251
\]
therefore

\[ P \left[ y < 6500 \right] = 0.0276247. \]  

Similarly, if we find

\[ P \left[ y < 6200 \right] \]

it will be close to 0.025 and this checks with the lower limit of the preceding section.

Now for upper limits, we try $10,000$. In this case

\[
k = \frac{(89,267) - (10,000)(.011129)}{[\frac{1}{4}(69.35344299) - 2(10,000)(+.0915647279) + (10,000)^2(.000019198)^{1/2}]
\]

\[= -.932 \]  

and

\[
p = \frac{(915647279)-(10,000)(.000019198)}{(.004381)\left[\frac{1}{4}(69.35344299) - 2(10,000)(.0915647279) + (10,000)^2(.000019198)\right]^{1/2} \]

\[ = -.97 \]  

and consulting the tables, we get

\[
\int_{2.54}^{\infty} \int_{-9.32}^{\infty} N(p=-.97) \, dx \, dy = \int_{2.54}^{\infty} N(0,1) \, dx
\]
and
\[- \int_{-2.54}^{\infty} \int_{9.32}^{\infty} N(p = .97) dx \, dy = \int_{-2.54}^{\infty} N(0,1) dy \]
\[- \int_{2.54}^{\infty} \int_{9.32}^{\infty} N(p = .97) dx \, dy \]
\[= .005 - (.005) = 0 \]
and
\[- \int_{-2.54}^{\infty} \int_{9.32}^{\infty} N(p = -.97) dx \, dy \]
\[= .1762 - .005590 = .17 \]
Therefore, the probability that \( y > 10,000 \) is about 17 per cent.

Similarly, if we find the probability that profit \( y > 12,000 \), the chance is 8 per cent. Again the chance that profit be \( > 13,000 \) is 6 per cent. That clearly shows that the distribution is skew to the right and to get the point of .025 probability, we will have to investigate much higher profits.

Suppose we try \( 18,000 \).

\[ v = 18,000 \]
\[ h = 2.54 \]
\[ k = \frac{(89.267)-(18,000)(.011129)}{[(469.35443)-2(18,000)(.0915647479) + (18,000)^2(.000019198)]^{1/2}} \]

\[ = -1.9 \quad (7.6.24) \]

\[ \rho = \frac{(.0915647279)-(18,000)(.000019198)}{(.004381)(469.35443)-2(18,000)(.0915647279) + (18,000)^2(.000019198)]^{1/2}} \]

\[ = -.995 \quad (7.6.25) \]

Now, using the tables, we get

\[ \int_{-1.9}^{2.54} \int_{2.54}^{\infty} N(\rho=-.995) \, dx \, dy = \int_{2.54}^{\infty} N(0,1) \, dx \]

\[ = .0055 - .0055 = 0 \]

\[ \int_{1.9}^{2.54} \int_{1.9}^{\infty} N(\rho=.995) \, dx \, dy = \int_{1.9}^{\infty} N(0,1) \, dy \]

\[ = \]
Therefore,

$$\Pr \{ y > 18,000 \} = 0.0232 \quad (7.6.26)$$

which is slightly less than 0.025. Hence, our results obtained from cumulative distributions check with the limits obtained by quadratic formula of Section 4.8. This fact is very reassuring since the two sections are based on different approaches. The confidence limits obtained by equation (4.4.3) are based on Geary's approximation whereas the cumulative distribution inferences are based on Fieller's approach.#

### 7.7 A critical note.

We have noticed in the previous section, that, whereas the lower limits of our results seem fairly reasonable, the upper limits are very far removed from the expected values. For example, the lower limit for corn yield is 2088 bushels.

---


Therefore, if this yield is realized on 78.78 acres of land used on corn production, the average yield will be 26.5 bushels per acre which is very reasonable. If we have a look on our data for corn yield for the period 1928-1950 in Table 1, we notice that there is one year, 1936, when the corn yield was 25 bushels per acre. On the contrary, if the upper limit of total yield, that is, 10,282 bushels, is realized, the average yield will be about 130 bushels per acre which seems fantastically high. Similarly the upper limits of the flax yield and the profit function are very high and therefore nor reliable. As we have said before, the lower limits are more important in planning. A farmer for example, is more concerned about the possibility of bankruptcy than about getting too rich. Therefore, if he is told that the program of production which is recommended to him and which is expected to yield a profit of $8,000, will bring him at least $6,200, he will have confidence in the recommendation. Thus, the lower limits serve the important purpose of assuring the results of our planning with the flexibility necessary under "risk" phenomena.

It may be instructive, however, to have a look into the cause of this over-estimation on the upper side. In Section 7.2 we wrote down the frequency distribution of the input coefficients which enter into our prediction and
and in that section we said that these could be approximated by normal distributions fairly closely. Let us investigate that fact a little further. When we have a look at the distribution and the corresponding histograms, we notice that, whereas they could be approximated as normal having equal frequency on either side of the modal group, they are truncated on the left. That is on that evidence the input coefficients have practically no chance to come close to zero. To be more precise, we can suppose, for example, that they have no chance to assume values less than the respective lower limits of the lowest frequency class. Whereas, in applying our results to get the limits of the yields, we assume a full normal distributions of the input coefficients thus allowing the input coefficients to come close to zero. Since the input coefficients are inversely proportional to the yields, the upper limits of the yields and of the profit, shown by our analysis, though perfectly valid, theoretically, may be far above the actual realization. This explains why, in practice, these upper limits will be far from prevailing anything which may actually happen.

One way to remedy this shortcoming will be to use truncated normal distributions or non-normal distributions, and get the results by developing the theory using the general procedures outlined in Chapter IV. However, this will require a great deal of work.
The technique of the simplex method of linear programming was employed to get an optimum production plan for a family farm in Ellsworth Township in Hancock County, Iowa. The input coefficients used in planning were the averages of the corresponding input coefficients computed from the input and output data for the period 1928-1952. Corn, oats, soybeans, flax, and wheat were the five crops considered in making the plan.

It was assumed that the family farm has one hundred and forty-eight acres of land and a capital of $1800. The monthly available labor was also specified. The problem was to draw up an optimum production plan which may be expected to maximize the gross profit. The prices per bushel of the outputs were assumed to be constant and the same as the 1952 average prices in Iowa. The following plan was derived: The family farm should plan to produce 3464 bushels of corn and 686 bushels of flax. The detailed schedule of inputs required is given in Tables 10a and 10c.

With the average prices of 1952, this program was expected to yield a gross profit of $8021.

Then to exhibit the practical application of the probability approach, it was assumed that the input coefficients relevant to the program, vary like normal variates with the
values used in planning as expected (mean) values. The variances of these parameters were estimated again from the series of input coefficients computed for the period 1928-1952. Using these estimates in the analysis of Section 4.4, we obtained the following 95 per cent, approximate confidence limits.

- Corn output in bushels: 2088 and 10282
- Flax output in bushels: 46 and 1625
- Gross output in dollars: 6298 and 18866

These results check very closely with the corresponding probability statements based on the cumulative distributions Section 4.6. It was noticed that (1) flax output is subject to greater variability than that of corn (2) the upper limits, seem to depart too much from the expected values compared with the corresponding lower limits; this observation indicates that the derived distributions are skew to the right.

In the end of Chapter VII, a possible reason was given for this discrepancy. A more realistic assumption about the distributions of the input coefficients would have been to assume a normal distribution truncated on the left. However, since the theoretical analysis in that case is difficult, we assume simply normal distributions. The extension of the distributions towards zero affects the
upper limits. It is feared that in all practical cases, where normality of the input coefficients is assumed, the upper limits may not be very realistic.

In planning production under risk phenomena, however, we are mostly concerned with the lower limits. Therefore, the reliable lower limits derived from this analysis will be of considerable importance to the enterprise.
IX. BIBLIOGRAPHY


Dwyer, Paul S. Linear Computations, New York, John Wiley and Sons, Inc., 1951.


Georgescu-Roegen, N. Leontief's System in the Light of Recent Results, Rev. of Econ. and Stat., 32:214-222, 1950.


Hicks, J. R. Value and Capital, Oxford, Oxford Univ. Press, 1941.


Pearson, Karl. On the Constants of Index Distributions, etc., Biometrika, 7:531-541, 1910.


Tintner, Gerhard. The Distribution of Solutions of Linear Equations Whose Coefficients are Subject to Error. Unpublished, Department of Econ. and Soc., Iowa State College.


X. ACKNOWLEDGEMENTS

The author gratefully acknowledges the help and guidance of the committee in charge of his graduate work, Prof. Gerhard Tintner, Prof. Earl O. Heady, Prof. R. J. Jessen, Prof. H. P. Thielman, and Prof. B. Vinograde. He also is grateful to his other teachers and colleagues at Iowa State College for their encouragement and help. The author is much indebted to Prof. Tintner for suggesting the problem and for the active help he has given him during the course of this work. He is also obliged to Prof. Heady for setting up the empirical example and aiding in its analysis. In appreciating their help, the author should not fail to acknowledge that the responsibility of any error or shortcoming in the dissertation is entirely his own.

Dr. T. A. Bancroft, Dr. D. L. Holl, and Dr. W. G. Murray, by providing research and teaching assistantships, have made it possible for the author to finance his studies. He humbly acknowledges that help. He is also indebted to Diwan Anand Kuman, vice-chancellor, Panjab University, India, and Principal Niranjani Singh for granting him study leave with a special allowance.

Finally, the author wishes to express respect to his parents and former teachers whose association remains a constant source of inspiration to him.
XI. APPENDIX
<table>
<thead>
<tr>
<th></th>
<th>P9</th>
<th>P10</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>.05253</td>
<td>.11681</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.56</td>
<td>1.84</td>
<td>2.79</td>
<td>3.81</td>
<td>2.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.29922</td>
<td>.02274</td>
<td>.00266</td>
<td>.05862</td>
<td>.09249</td>
<td>.05081</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3.13672</td>
<td>.31772</td>
<td>.01615</td>
<td>.70612</td>
<td>.96956</td>
<td>1.00356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3.68541</td>
<td>.02555</td>
<td>.01786</td>
<td>.05485</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.23520</td>
<td>.05253</td>
<td>.27079</td>
<td>.11681</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.326</td>
<td>1.03</td>
<td>3.81</td>
<td>1.066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.326</td>
<td>-1.56</td>
<td>.19</td>
<td>-2.79</td>
<td></td>
<td></td>
<td>-1.074</td>
<td></td>
</tr>
<tr>
<td>-5.10440</td>
<td>.38792</td>
<td>.04538</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.97781</td>
<td>.04303</td>
<td>.01598</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4.40543</td>
<td>.00427</td>
<td>.01537</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.23520</td>
<td>.00722</td>
<td>.27079</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.59624</td>
<td>.00530</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.9</td>
<td>1.08</td>
<td>1.15</td>
<td>1.15</td>
<td>3.81</td>
<td>4.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.9</td>
<td>.48</td>
<td>.31</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td>8.56100</td>
<td>.44973</td>
<td>.01173</td>
<td>1</td>
<td>.00004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>---</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>.80137</td>
<td>.03724</td>
<td>.01177</td>
<td>1</td>
<td>.05157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>.67998</td>
<td>.00918</td>
<td>.01177</td>
<td>1</td>
<td>.21523</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>5.42601</td>
<td>.03918</td>
<td>.29955</td>
<td>1</td>
<td>.93185</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.67718</td>
<td>.01211</td>
<td>.00889</td>
<td>1</td>
<td>.21699</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.21</td>
<td>1.1054</td>
<td>1.14</td>
<td>2.79</td>
<td>3.81</td>
<td>3.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.21</td>
<td>.455</td>
<td>.30</td>
<td></td>
<td></td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-24.75135</td>
<td>- .57661</td>
<td>1</td>
<td></td>
<td></td>
<td>-10.54412</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3.55961</td>
<td>- .05948</td>
<td></td>
<td></td>
<td></td>
<td>-.82154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74.07189</td>
<td>1</td>
<td>1.28213</td>
<td></td>
<td></td>
<td>23.44553</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.51387</td>
<td>.31978</td>
<td>.02442</td>
<td></td>
<td></td>
<td>1.90044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.78017</td>
<td>.01177</td>
<td></td>
<td></td>
<td></td>
<td>.50091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.9</td>
<td>1.56</td>
<td>1.72</td>
<td>2.79</td>
<td>3.81</td>
<td>14.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.9</td>
<td>.88</td>
<td></td>
<td></td>
<td></td>
<td>12.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-22.056</td>
<td>-.00514</td>
<td>.00514</td>
<td></td>
<td></td>
<td>.09396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3.82241</td>
<td>-.06560</td>
<td>.01061</td>
<td></td>
<td></td>
<td>.70967</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.03611</td>
<td>1</td>
<td>.00923</td>
<td>2.22237</td>
<td>1</td>
<td>46.88100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.28513</td>
<td>.35510</td>
<td>.00923</td>
<td>1</td>
<td></td>
<td>1.30362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.70953</td>
<td>.02607</td>
<td>.00285</td>
<td></td>
<td></td>
<td>.47082</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.98</td>
<td>1.56</td>
<td>1.35</td>
<td>3.43</td>
<td>3.81</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.98</td>
<td>.51</td>
<td>.64</td>
<td></td>
<td></td>
<td>77.86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>