5-25-1960

A Statistical Technique for Predicting a Two Dimensional Vector with Application

Richard E. Vogel

A STATISTICAL TECHNIQUE

VOGEL

A STATISTICAL TECHNIQUE FOR PREDICTING A TWO DIMENSIONAL VECTOR WITH APPLICATION

By

Richard E. Vogel

A Thesis

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Mathematics

The University of New Mexico

1960

This thesis, directed and approved by the candidate's committee, has been accepted by the Graduate Committee of the University of New Mexico in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

E. F. Castetter
Dean

May 25, 1960
Date

Thesis committee

Royal K. Yeagler
Chairman

J. V. Lewis

J. Magne-Soelliffmidt

ACKNOWLEDGMENT

The author wishes to express his appreciation to R. K. Zeigler and R. H. Moore for their constructive criticisms and helpful suggestions, and to O. W. Stopinski for his advice on the meteorological application discussed in this report.

ABSTRACT

The problem of multiple regression analysis where the dependent and independent variables are components of a two dimensional vector is discussed, and a complete statistical development of the solution of estimators for the parameters in the model is given. The theory regarding predictions and confidence statements about such predictions is also developed. A computer code was written for the IBM 704 computer which solves the above problem and a description of the code appears in the Appendix.

The statistical model was applied to a meteorological problem in wind forecasting at the Eniwetok Proving Ground, and prediction equations were developed and evaluated.

ACKNOWLEDGMENT

The author wishes to express his appreciation to R. K. Zeigler and R. H. Moore for their constructive criticisms and helpful suggestions, and to O. W. Stopinski for his advice on the meteorological application discussed in this report.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Page

Figure

INTRODUCTION

Among the aims or goals of the scientist is the ability to predict future events on the basis of knowledge derived from past experimental work. Solutions of such problems are usually achieved by the use of regression analysis, where the analysis yields not only the prediction but a measure of the error in the prediction as well.

A simple case of regression analysis involves an independent variable linearly related to a dependent variable. Under a set of basic assumptions, the analysis is straightforward, with the prediction equation a linear function of the independent variable and the error in the prediction in the form of a confidence statement. The theory of such an analysis becomes slightly complicated when the above relationship is extended to q independent variables and even more complicated when each of the dependent and independent variables represents components of a vector in a two dimensional plane.

We shall be mainly concerned in Chapter I with the development of the theory for the latter case. Two dimensional vector variates are usually observed in the form of magnitude and direction; however, it will be shown that transforming the vector into rectangular coordinates u and v simplifies this development. After establishing the basic definitions, assumptions, and model, we will develop estimators of the parameters in the model, the prediction functions, and the error analysis associated with these predictions.

In Chapter II we shall apply the developed theory to a specific problem, namely, that of wind forecasting at the Eniwetok Proving Ground, show how the statistical model can be used in this meteorological problem, and evaluate the effectiveness of the technique.

Finally, we shall describe an IBM 704 computer code for solving the problem.

CHAPTER I

STATISTICAL DEVELOPMENT

1. Definitions

Although the mathematical notation is defined throughout this paper, there may exist some terminology not familiar to the reader. Therefore, definitions of basic terms used in this paper are given.

A continuous distribution function is a function, f(x), which has the following properties:

a. $f(x) \geq 0 \qquad -\infty < x < \infty$

b. $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

All distribution functions used in this paper are assumed to be continuous.

A joint distribution function is a function of two or more variables $f(x_1, x_2, \ldots, x_n)$ which satisfy the following conditions:

a. $f(x_1, \ldots, x_n) \geq 0 \qquad -\infty < x_1, x_2, \ldots, x_n < \infty$

b. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1, dx_2, \ldots, dx_n = 1.$

A <u>normal distribution function</u> is a function of one variable and two parameters represented by the following equation:

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty.$$

A <u>bivariate normal distribution function</u> is a function in two variables and five parameters represented by the following equation:

$$f\left(x, y;\, \mu_x \mu_y,\, \sigma_x \sigma_y, \rho\right) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho\left(x-\mu_x\right)\left(y-\mu_y\right)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} \qquad -\infty < x < \infty, \quad -\infty < y < \infty.$$

A <u>chi square distribution function</u> is a function represented by the following equation:

$$f\left(\chi^2; d\right) = \frac{1}{\left(\frac{d}{2}-1\right)!\; 2^{d/2}} \left(\chi^2\right)^{d-2/2} e^{-\chi^2/2} \qquad 0 < \chi^2 < \infty,$$

where d is the parameter of the distribution and commonly referred to as the degrees of freedom.

The <u>"F" distribution function</u> is a function represented by the following equation:

$$h(F; a, b) = \frac{\left(\frac{a+b-2}{2}\right)!}{\left(\frac{a-2}{2}\right)!\left(\frac{b-2}{2}\right)!} \left(\frac{a}{b}\right)^{a/2} \frac{F^{a-2/2}}{\left(1+\frac{aF}{b}\right)\left(\frac{a+b}{2}\right)} \qquad F > 0,$$

where a and b are the parameters of the function and commonly referred to as the degrees of freedom.

The "T^2" distribution function is a function represented by the following equation:

$$T^2 = 2h(F; 2, b)\frac{N-1}{N-2},$$

where F(2, b) is the "F" distribution with parameters 2 and b, and N the number of observations in the sample. T^2 is often referred to as Hotelling's T.

If a random sample $(x_1, x_2, \ldots, x_n)$ is drawn from a population defined by the distribution function f(x), the joint distribution function of the sample is

$$g(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) .$$

R. A. Fisher has called this joint distribution function the likelihood of the sample.

If f(x) is a distribution function, then the expectation or expected value of g(x), denoted by the letter E, is defined by the following equation:

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)\ f(x)\ dx \qquad -\infty < g(x) < \infty.$$

The moments of a distribution function are the expected values of the powers of the variable which has the given distribution. The first moment, E[x], is called the mean.

The variance is denoted by σ_x^2 and defined by $\sigma_x^2 = E\left[(x-E(x))^2\right]$. The variance is a measure of the dispersion about the mean.

The covariance of two jointly distributed variables x and y is denoted by σ_{xy} and defined by

$$\sigma_{xy} = E[(x-E(x))(y-E(y))].$$

The coefficient of correlation between x and y is denoted by ρ_{xy} and defined by

$$\rho_{xy} = \frac{E[(x-E(x))(y-E(y))]}{\sqrt{E\left[(x-E(x))^2\right]\cdot E\left[(y-E(y))^2\right]}}$$

2. Basic Assumptions and Model

Let us denote the u and v components of the dependent vector variate by the variables x_1 and x_2, respectively. The u and v components of q/2 independent vector variates are treated as separate variables and are denoted by z_g, where $g = 1, \ldots, q$. The observed data will be given and N sets of numbers $\left(z_{1\alpha}, z_{2\alpha}, \ldots, z_{q\alpha}, x_{1\alpha}, x_{2\alpha}\right)$ for $\alpha = 1, \ldots, N$ and will be denoted as a sample of size N, where the z's are free from error.

Let us make the following assumptions:[1]

1. For every observed value of $\left(z_1, \ldots, z_q\right)$, x_1 and x_2 have a bivariate normal distribution with means a linear function of the z's. These means are given by the following equations:

$$\mu_1 = \sum_{g=1}^{q} b_{1g} z_g \tag{1}$$

and

$$\mu_2 = \sum_{g=1}^{q} b_{2g} z_g \ . \tag{2}$$

2. The variances of x_1 and x_2 are independent of $z_1, \ldots, z_q$ and denoted by σ_1^2 and σ_2^2, respectively. The correlation coefficient between x_1 and x_2 is denoted by ρ.

3. The N sets of observed numbers are stochastically independent (i.e., the sample is random).

Under the above assumptions the statistical model can be written as

follows:

$$f\left(x_1, x_2\right) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x_1 - \sum_{g=1}^{q} b_{1g} z_g}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1 - \sum_{g=1}^{q} b_{1g} z_g}{\sigma_1}\right)\left(\frac{x_2 - \sum_{g=1}^{q} b_{2g} z_g}{\sigma_2}\right) + \left(\frac{x_2 - \sum_{g=1}^{q} b_{2g} z_g}{\sigma_2}\right)^2\right]\right\}. \quad (3)$$

Now if we let*

$$V = \begin{pmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1^2\left(1-\rho^2\right)} & \frac{-\rho}{\sigma_1\sigma_2\left(1-\rho^2\right)} \\ \frac{-\rho}{\sigma_1\sigma_2\left(1-\rho^2\right)} & \frac{1}{\sigma_2^2\left(1-\rho^2\right)} \end{pmatrix}, \quad (4)$$

then $f\left(x_1, x_2\right)$ becomes

$$f\left(x_1, x_2\right) = \frac{1}{2\pi}\sqrt{|V|} \quad \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}\left(x_1 - \sum_{g=1}^{q} b_{ig} z_g\right)\sigma^{ij}\left(x_j - \sum_{g=1}^{q} b_{jg} z_g\right)\right]\right\}, \quad (5)$$

where $|V|$ is the determinant of equation 4.

3. Parameter Estimation by the Theory of Maximum Likelihood

If a random sample of size N is drawn from the bivariate normal distribution, the joint distribution function of the sample becomes

---

*The matrix of V is the inverse of the variance-covariance matrix.

$$\prod_{\alpha=1}^{N} f\left(x_{1\alpha}, x_{2\alpha}\right) = \left(\frac{1}{2\pi}\right)^{N} |V|^{\frac{N}{2}} \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N} \sum_{i=1}^{2} \sum_{j=1}^{2} \left[\left(x_{1\alpha} - \sum_{g=1}^{q} b_{ig} z_{g\alpha}\right) \cdot \sigma^{ij}\left(x_{j\alpha} - \sum_{g=1}^{q} b_{jg} z_{g\alpha}\right)\right]\right\}. \tag{6}$$

On the basis of these N sets of observations, we wish to determine estimates for the parameters σ^{ij} and the b's. The method of maximum likelihood is chosen to estimate these parameters since it gives estimators which have very desirable properties.[2] This method gives estimates of the σ's and b's, based on the sampling distribution, as $\hat{\sigma}$'s and $\hat{b}$'s that maximize the likelihood function, which in this particular case is given by equation 7.

$$L = \prod_{\alpha=1}^{N} f\left(x_{1\alpha}, x_{2\alpha}\right). \tag{7}$$

It may be noted that the point at which the likelihood function is maximized is equal to the point at which the natural logarithm or the likelihood function is maximized. Hence, maximizing the natural logarithm of this likelihood function has the advantage of working with summations rather than products.

If we regard the parameters as continuous variables, the maximizing value of the natural logarithm of the likelihood function can be obtained by setting the partial derivatives of this function, taken with respect to the parameters, equal to zero and solving for the parameters in the resulting equations.

The natural logarithm of the likelihood function which we will work with is given by equation 8.

$$\ln L = -N \ln 2\pi + \frac{N}{2} \ln |V| - \frac{1}{2}\sum_{\alpha=1}^{N}\left[\sigma^{11}\left(x_{1\alpha} - \sum_{g=1}^{q} b_{1g} z_{g\alpha}\right)^2\right.$$

$$\left. + 2\sigma^{12}\left(x_{1\alpha} - \sum_{g=1}^{q} b_{1g} z_{g\alpha}\right)\left(x_{2\alpha} - \sum_{g=1}^{q} b_{2g} z_{g\alpha}\right) + \sigma^{22}\left(x_{2\alpha} - \sum_{g=1}^{q} b_{2g} z_{g\alpha}\right)^2\right]. \qquad (8)$$

The partial derivatives of equation 8 are given as follows:

$$\frac{\partial \ln L}{\partial b_{11}} = -\frac{1}{2}\sum_{\alpha=1}^{N}\left[2\sigma^{11}\left(-x_{1\alpha} z_{1\alpha} + z_{1\alpha}\sum_{g=1}^{q} b_{1g} z_{g\alpha}\right) + 2\sigma^{12}\left(-x_{2\alpha} z_{1\alpha} + \sum_{g=1}^{q} b_{2g} z_{g\alpha}\right)\right] = 0$$

$$\frac{\partial \ln L}{\partial b_{21}} = -\frac{1}{2}\sum_{\alpha=1}^{N}\left[2\sigma^{22}\left(-x_{2\alpha} z_{1\alpha} + z_{1\alpha}\sum_{g=1}^{q} b_{2g} z_{g\alpha}\right) + 2\sigma^{12}\left(-x_{1\alpha} z_{1\alpha} + z_{1\alpha}\sum_{g=1}^{q} b_{2g} z_{g\alpha}\right)\right] = 0$$

$$\frac{\partial \ln L}{\partial b_{12}} = -\frac{1}{2}\sum_{\alpha=1}^{N}\left[2\sigma^{11}\left(-x_{1\alpha} z_{1\alpha} + z_{2\alpha}\sum_{g=1}^{q} b_{1g} z_{g\alpha}\right) + 2\sigma^{12}\left(-x_{2\alpha} z_{2\alpha} + z_{2\alpha}\sum_{g=1}^{q} b_{2g} z_{g\alpha}\right)\right] = 0$$

.
.
.

The generalized form of the above equations is given by equation 9.

$$\frac{\partial \ln L}{\partial b_{kf}} = \sum_{\alpha=1}^{N}\sum_{j=1}^{2}\left[\sigma^{kj}\left(x_{j\alpha} z_{f\alpha} - \sum_{g=1}^{q} b_{jg} z_{g\alpha} z_{j\alpha}\right)\right] = 0. \qquad (9)$$

Let

$$c_{jf} = \sum_{\alpha=1}^{N} x_{j\alpha} z_{f\alpha} \qquad \text{and} \qquad a_{gf} = \sum_{\alpha=1}^{N} z_{g\alpha} z_{f\alpha},$$

then equation 9 becomes

$$\frac{\partial \ln L}{\partial b_{kf}} = \sum_{j=1}^{2}\left[\sigma^{kj}\left(c_{jf} - \sum_{g=1}^{q} b_{jg} a_{gf}\right)\right] = 0. \qquad (10)$$

Let us examine the following 2q equations:

$$\frac{\partial \ln L}{\partial b_{11}} = \sigma^{11}\left(c_{11} - \sum_{g=1}^{q} b_{1g} a_{g1}\right) + \sigma^{12}\left(c_{21} - \sum_{g=1}^{q} b_{2g} a_{g1}\right) = 0$$

$$\frac{\partial \ln L}{\partial b_{12}} = \sigma^{11}\left(c_{12} - \sum_{g=1}^{q} b_{1g} a_{g2}\right) + \sigma^{12}\left(c_{22} - \sum_{g=1}^{q} b_{2g} a_{g2}\right) = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial \ln L}{\partial b_{1q}} = \sigma^{11}\left(c_{1q} - \sum_{g=1}^{q} b_{1g} a_{gq}\right) + \sigma^{12}\left(c_{2q} - \sum_{g=1}^{q} b_{2g} a_{gq}\right) = 0$$

$$\frac{\partial \ln L}{\partial b_{21}} = \sigma^{21}\left(c_{11} - \sum_{g=1}^{q} b_{1g} a_{g1}\right) + \sigma^{22}\left(c_{21} - \sum_{g=1}^{q} b_{2g} a_{g1}\right) = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial \ln L}{\partial b_{2q}} = \sigma^{21}\left(c_{1q} - \sum_{g=1}^{q} b_{1g} a_{gq}\right) + \sigma^{22}\left(c_{2q} - \sum_{g=1}^{q} b_{2g} a_{gq}\right) = 0.$$

These equations can be summarized in the matrix form as

$$\begin{pmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{pmatrix} \begin{pmatrix} \left(c_{11} - \sum_{g=1}^{q} b_{1g} a_{g1}\right) & \cdots & \left(c_{1q} - \sum_{g=1}^{q} b_{1g} a_{gq}\right) \\ \left(c_{21} - \sum_{g=1}^{q} b_{2g} a_{g2}\right) & \cdots & \left(c_{2q} - \sum_{g=1}^{q} b_{2q} a_{gq}\right) \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} \qquad (11)$$

Since the matrix $\begin{pmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{pmatrix}$ is in itself an inverse matrix, it is non-singular

and has an inverse. If we multiply both sides of equation 11 by the inverse of the V matrix we get the following result;

$$\begin{pmatrix} \left(c_{11} - \sum_{g=1}^{q} b_{1g} a_{g1}\right) & \cdots & \left(c_{1q} - \sum_{g=1}^{q} b_{1g} a_{gq}\right) \\ \left(c_{21} - \sum_{g=1}^{q} b_{2g} a_{g1}\right) & \cdots & \left(c_{2q} - \sum_{g=1}^{q} b_{2g} a_{gq}\right) \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix}. \tag{12}$$

Equating elements of the matrices, we may write

$$c_{kf} - \sum_{g=1}^{q} b_{kg} a_{gf} = 0. \tag{13}$$

Summarizing the 2q equations in matrix notation, equation 13 can be written as

$$C - B\,A = \Phi, \tag{14}$$

where

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1q} \\ c_{21} & \cdots & c_{2q} \end{pmatrix},\ B = \begin{pmatrix} b_{11} & \cdots & b_{1q} \\ b_{21} & \cdots & b_{2q} \end{pmatrix},\ A = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{q1} & \cdots & a_{gq} \end{pmatrix},$$

and

$$\Phi = \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix}.$$

For A non-singular, the solution of the equation for B yields the estimator $\hat{B}$, where each b_{ij} is estimated by $\hat{b}_{ij}$, and is given by equation 15.

$$\hat{B}\,A\,A^{-1} = C\,A^{-1}$$
$$\hat{B} = C\,A^{-1}. \tag{15}$$

In estimating the σ^{ij} parameters from equation 8 by the method of maximum likelihood, the partial derivatives with respect to σ^{ij} are needed. The general form of these partials is given by equation 16.

$$\frac{\partial \ln L}{\partial \sigma^{ij}} = \frac{N}{2}\,\frac{\text{cofactor } \sigma^{ij}}{|V|} - \frac{\gamma}{2}\sum_{\alpha=1}^{N}\left(x_{i\alpha} - \sum_{g=1}^{q} b_{ig} z_{g\alpha}\right)\left(x_{j\alpha} - \sum_{g=1}^{q} b_{jg} z_{g\alpha}\right) = 0 \tag{16}$$

$\gamma = 2$ for $i \neq j$ and $\gamma = 1$ for $i = j$; $\quad i = 1, 2 \quad j = 1, 2,$

where

$$|V| = \begin{vmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{vmatrix} = \sigma^{11}\sigma^{22} - \sigma^{12}\sigma^{21}$$

and

$$\frac{\text{cofactor } \sigma^{ij}}{|V|} = \sigma_{ij}$$

The solutions of equation 16 will obviously involve the b's, which we have already solved for in equation 15. Substituting the $\hat{b}$'s for the b's and solving equation 16 for σ_{ij}, we get the maximum likelihood estimator $\hat{\sigma}_{ij}$.

$$\hat{\sigma}_{ij} = \frac{1}{N}\sum_{\alpha=1}^{N}\left(x_{i\alpha} - \sum_{g=1}^{q} \hat{b}_{ig} z_{g\alpha}\right)\left(x_{j\alpha} - \sum_{g=1}^{q} \hat{b}_{jg} z_{g\alpha}\right). \tag{17}$$

Although it is true that we set out to find the maximum likelihood estimator of σ^{ij} and ended with estimators for σ_{ij}, we are justified since the maximum likelihood estimators have the property of invariance[2] under this transformation.

4. Distribution of the $\hat{b}$ Parameters

Let us begin by examining the case where q = 2 and write the $\hat{B}$ solution of equation 15 in terms of the x's and z's.

$$\begin{pmatrix} \hat{b}_{11} & \hat{b}_{12} \\ \hat{b}_{21} & \hat{b}_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{\sum_{\alpha=1}^{N} x_{1\alpha} z_{1\alpha} \sum_\alpha z_{2\alpha}^2 - \sum_\alpha x_{1\alpha} z_{2\alpha} \sum_\alpha z_{1\alpha} z_{2\alpha}}{D} & \dfrac{-\sum_\alpha x_{1\alpha} z_{2\alpha} \sum_\alpha z_{1\alpha} z_{2\alpha} + \sum_\alpha x_{1\alpha} z_{2\alpha} \sum_\alpha z_{1\alpha}^2}{D} \\ \dfrac{\sum_\alpha x_{2\alpha} z_{1\alpha} \sum_\alpha z_{2\alpha}^2 - \sum_\alpha x_{2\alpha} z_{2\alpha} \sum_\alpha z_{1\alpha} z_{2\alpha}}{D} & \dfrac{-\sum_\alpha x_{2\alpha} z_{1\alpha} \sum_\alpha z_{1\alpha} z_{2\alpha} + \sum_\alpha x_{2\alpha} z_{2\alpha} \sum_\alpha z_{1\alpha}^2}{D} \end{pmatrix}, \quad (18)$$

where

$$D = \left[\sum_\alpha z_{1\alpha}^2 \sum_\alpha z_{2\alpha}^2 - \left(\sum_\alpha z_{1\alpha} z_{2\alpha} \right)^2 \right].$$

Since the independent variables, z, are considered constant in this particular case, we note from equation 18 that there exists a linear relationship between $\hat{b}_{k\ell}$ and x_k ($k = 1, 2$; $\ell = 1, 2$). Because the x_k's are distributed normally, so will be the $\hat{b}_{k\ell}$'s.[3]

The expected value of $\hat{b}_{11}$ is given by the following:

$$E\left(\hat{b}_{11}\right) = E\left[\frac{\sum_\alpha x_{1\alpha} z_{1\alpha} \sum_\alpha z_{2\alpha}^2 - \sum_\alpha x_{1\alpha} z_{2\alpha} \sum_\alpha z_{1\alpha} z_{2\alpha}}{D} \right]$$

$$= E\left[\frac{\sum_\alpha \left[x_{1\alpha} \left(z_{1\alpha} \sum_\alpha z_{2\alpha}^2 - z_{2\alpha} \sum_\alpha z_{1\alpha} z_{2\alpha} \right) \right]}{D} \right]$$

(Equation continued on next page)

$$= \frac{\sum_\alpha \left[E\left(x_{1\alpha}\right)\left(z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\right]}{D}$$

$$= \frac{\sum_\alpha \left[\left(b_{11}z_{1\alpha}+b_{12}z_{2\alpha}\right)\left(z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\right]}{D}$$

$$= b_{11}\left[\frac{\sum_\alpha z_{1\alpha}^2 \sum_\alpha z_{2\alpha}^2 - \left(\sum_\alpha z_{1\alpha}z_{2\alpha}\right)^2}{D}\right]$$

$$+ \; b_{12}\left[\frac{\sum_\alpha z_{2\alpha}^2 \sum_\alpha z_{1\alpha}z_{2\alpha} - \sum_\alpha z_{2\alpha}^2 \sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\right]$$

$$E\left(\hat{b}_{11}\right) = b_{11}. \tag{19}$$

The variance of $\hat{b}_{11}$ is

$$E\left[\left(\hat{b}_{11} - b_{11}\right)^2\right] = E\left[\left(\hat{b}_{11} - E\left(\hat{b}_{11}\right)\right)^2\right]$$

$$= E\left[\left(\frac{\sum_\alpha \left[x_{1\alpha}\left(z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\right]}{D} - E\,\frac{\sum_\alpha \left[x_{1\alpha}\left(z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\right]}{D}\right)^2\right]$$

$$= E\left[\left(\sum_\alpha\left(x_{1\alpha} - E\left(x_{1\alpha}\right)\right)\frac{\left(z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)}{D}\right)^2\right]$$

(Equation continued on next page)

$$= E\left[\sum_\alpha\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)^2\left(\frac{z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\right)^2\right.$$
$$\left. + E\ [\text{cross products}]\right].$$

Examining the first cross product we get

$$E\left[\left(x_{11}-E\left(x_{11}\right)\right)\left(x_{12}-E\left(x_{12}\right)\right)\ \text{(constant as far as expected values are concerned)}\right]. \quad (20)$$

Under the basic assumptions of the model, x_{11} is independent of x_{12} so we can write equation 20 as

$$E\left(x_{11}-E\left(x_{11}\right)\right)\ E\left(x_{12}-E\left(x_{12}\right)\right)\cdot \text{constant}$$

but

$$E\left(x_{11}-E\left(x_{11}\right)\right) = E\left(x_{11}\right)-E\left(x_{11}\right) = 0.$$

Thus the expected value of the first cross product equals zero. By the same reasoning we can show that the expected value of all the cross products equals zero.

$$E\left[\sum_\alpha\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)^2\left(\frac{z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\right)^2\right]$$

$$= \sigma_{11}\sum_\alpha\left(\frac{z_{1\alpha}\sum_\alpha z_{2\alpha}^2 - z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\right)^2$$

$$= \sigma_{11}\sum_\alpha\left[\frac{z_{1\alpha}^2\left(\sum_\alpha z_{2\alpha}^2\right)^2 - 2z_{1\alpha}z_{2\alpha}\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{1\alpha}z_{2\alpha} + z_{2\alpha}^2\left(\sum_\alpha z_{1\alpha}z_{2\alpha}\right)^2}{D^2}\right]$$

(Equation continued on next page)

$$= \sigma_{11}\left[\frac{\sum_\alpha z_{2\alpha}^2}{D}\;\frac{\left(\sum_\alpha z_{1\alpha}^2 \sum_\alpha z_{2\alpha}^2 - 2\left(\sum_\alpha z_{1\alpha} z_{2\alpha}\right)^2 + \sum_\alpha \left(z_{1\alpha} z_{2\alpha}\right)^2\right)}{D}\right]$$

$$= \sigma_{11}\left[\frac{\sum_\alpha z_{2\alpha}^2}{D}\;\frac{\left(\sum_\alpha z_{1\alpha}^2 \sum_\alpha z_{2\alpha}^2 - \left(\sum_\alpha z_{1\alpha} z_{2\alpha}\right)^2\right)}{D}\right]$$

$$= \sigma_{11} \frac{\sum_\alpha z_{2\alpha}^2}{D} = \sigma_{11} a^{11}.$$

Thus $\hat{b}_{11}$ is distributed normally with mean b_{11} and variance $\sigma_{11} a^{11}$. Similarly, it can be shown that $\hat{b}_{k\ell}$ is distributed normally with mean $b_{k\ell}$ and variance $\sigma_{kk} a^{k\ell}$.

While in general the maximum likelihood estimator gives a biased estimate of the true parameter,[2] it should be noted that in this case the $\hat{b}_{k\ell}$ estimator is unbiased. To demonstrate this property for $q \geq 2$, let us show that the expected value of $\hat{B}$ is B.

Let

$$X_\alpha = \begin{pmatrix} x_{1\alpha} \\ x_{2\alpha} \end{pmatrix}, \quad Z_\alpha = \begin{pmatrix} z_{1\alpha} \\ z_{2\alpha} \\ \cdot \\ \cdot \\ \cdot \\ z_{q\alpha} \end{pmatrix}, \text{ and } Z'_\alpha = \begin{pmatrix} z_{1\alpha} & z_{2\alpha} & \ldots & z_{q\alpha} \end{pmatrix}$$

then

$$C = \sum_{\alpha=1}^{N} X_\alpha Z'_\alpha. \tag{21}$$

Taking the expected value of equation 15 we get[3]

$$
\begin{aligned}
E(\hat{B}) &= E\left(C\,A^{-1}\right) \\
&= E\left(\sum_{\alpha=1}^{N} X_{\alpha}\, Z'_{\alpha}\, A^{-1}\right) \\
&= \sum_{\alpha=1}^{N} E\left(X_{\alpha}\right) Z'_{\alpha}\, A^{-1} \\
&= \sum_{\alpha=1}^{N} B\, Z_{\alpha}\, Z'_{\alpha}\, A^{-1} \qquad (22) \\
&= B \sum_{\alpha=1}^{N} Z_{\alpha}\, Z'_{\alpha}\, A^{-1} \\
&= B\, A\, A^{-1} \\
&= B.
\end{aligned}
$$

Since the expected value of this estimator equals the true parameter, the estimator is unbiased.

Let us now examine the joint distribution of the $\hat{b}_{k\ell}$ estimators in equation 18. This joint distribution is again normal since the $\hat{b}_{k\ell}$'s are linear functions of the x_k's.[4]

First, the covariance between $\hat{b}_{11}$ and $\hat{b}_{12}$ is

$$
E\left[\left(\hat{b}_{11} - b_{11}\right)\left(\hat{b}_{12} - b_{12}\right)\right] = E\left[\left(\hat{b}_{11} - E\left(\hat{b}_{11}\right)\right)\left(\hat{b}_{12} - E\left(\hat{b}_{12}\right)\right)\right]
$$

$$
= E\left(\left[\left(\frac{\sum_{\alpha} x_{1\alpha} z_{1\alpha} \sum_{\alpha} z_{2\alpha}^{2} - \sum_{\alpha} x_{1\alpha} z_{2\alpha} \sum_{\alpha} z_{1\alpha} z_{2\alpha}}{D}\right)\right.\right.
$$

(Equation continued on next page)

$$-E\left(\frac{\sum_\alpha x_{1\alpha}z_{1\alpha}\sum_\alpha z_{2\alpha}^2-\sum_\alpha x_{1\alpha}z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\right)$$

$$\cdot\left[\left(\frac{-\sum_\alpha x_{1\alpha}z_{1\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}+\sum_\alpha x_{1\alpha}z_{2\alpha}\sum_\alpha z_{1\alpha}^2}{D}\right)\right.$$

$$\left.\left.-E\left(\frac{-\sum_\alpha x_{1\alpha}z_{1\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}+\sum_\alpha x_{1\alpha}z_{2\alpha}\sum_\alpha z_{1\alpha}^2}{D}\right)\right]\right)$$

$$= E\left(\frac{\sum_\alpha\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)z_{1\alpha}\sum_\alpha z_{2\alpha}^2-\sum_\alpha\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right]}{D}\right.$$

$$\left.\cdot\frac{\sum_\alpha\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)z_{2\alpha}\sum_\alpha z_{1\alpha}^2-\sum_\alpha\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)z_{1\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right]}{D}\right)$$

$$= E\left(\frac{\sum_\alpha\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)\left(z_{1\alpha}\sum_\alpha z_{2\alpha}^2-z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\right]}{D}\right.$$

$$\left.\cdot\frac{\sum_\alpha\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)\left(z_{2\alpha}\sum_\alpha z_{1\alpha}^2-z_{1\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\right]}{D}\right)$$

$$= E\left(\frac{\sum_\alpha\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)^2\left(z_{1\alpha}\sum_\alpha z_{2\alpha}^2-z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\left(z_{2\alpha}\sum_\alpha z_{1\alpha}^2-z_{1\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}\right)\right]}{D^2}\right)$$

$$+ E\text{ (cross products)}.$$

Using the same type of argument that was used earlier, the expected value of the cross products equals zero. Thus,

$$E\left[\left(\hat{b}_{11}-b_{11}\right)\left(\hat{b}_{12}-b_{12}\right)\right] = \frac{\sum_\alpha\left(E\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)^2\right]\left[z_{1\alpha}z_{2\alpha}\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{1\alpha}^2-z_{1\alpha}^2\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{2\alpha}z_{1\alpha}\right.\right.}{D^2}$$

(Equation continued on next page)

$$+\frac{-z_{2\alpha}^2\sum_\alpha z_{1\alpha}^2\sum_\alpha z_{1\alpha}z_{2\alpha}+z_{1\alpha}z_{2\alpha}\left(\sum_\alpha z_{1\alpha}z_{2\alpha}\right)^2\Bigg]\Bigg)}{D^2}$$

$$= \sigma_{11}\left[\frac{\sum_\alpha z_{1\alpha}z_{2\alpha}\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{1\alpha}^2-\sum_\alpha z_{1\alpha}^2\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{1\alpha}z_{2\alpha}-\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{1\alpha}^2\sum_\alpha z_{1\alpha}z_{2\alpha}}{D^2}\right.$$

$$\left.+\frac{\sum_\alpha z_{1\alpha}z_{2\alpha}\left(\sum_\alpha z_{1\alpha}z_{2\alpha}\right)^2}{D^2}\right] \tag{23}$$

$$= \sigma_{11}\left[-\frac{\sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\left(\frac{-\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{1\alpha}^2+\sum_\alpha z_{1\alpha}^2\sum_\alpha z_{2\alpha}^2+\sum_\alpha z_{2\alpha}^2\sum_\alpha z_{1\alpha}^2}{D}\right.\right.$$

$$\left.\left.-\frac{\left(\sum_\alpha z_{1\alpha}z_{2\alpha}\right)^2}{D}\right)\right]$$

$$= \sigma_{11}\left[-\frac{\sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\left(\frac{D}{D}\right)\right] = \sigma_{11}\frac{\text{cofactor } a_{12}}{D} = \sigma_{11}a^{12}.$$

Thus the covariance between $\hat{b}_{11}$ and $\hat{b}_{12}$ is given by equation 23. Similarly, it can be shown that the covariance between $\hat{b}_{1\ell}$ and $\hat{b}_{1m}$ $(\ell, m = 1, 2)$ is equal to $\sigma_{11}a^{\ell m}$, and the covariance between $\hat{b}_{2\ell}$ and $\hat{b}_{2m}$ is $\sigma_{22}a^{\ell m}$.

Next the covariance between $\hat{b}_{11}$ and $\hat{b}_{21}$ is

$$E\left[\left(\hat{b}_{11}-b_{11}\right)\left(\hat{b}_{21}-b_{21}\right)\right] = E\left[\left(\hat{b}_{11}-E\left(\hat{b}_{11}\right)\right)\left(\hat{b}_{21}-E\left(\hat{b}_{21}\right)\right)\right]$$

$$= E\left(\sum_\alpha\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)\left(\frac{z_{1\alpha}\sum_\alpha z_{2\alpha}^2}{D}-\frac{z_{2\alpha}\sum_\alpha z_{1\alpha}z_{2\alpha}}{D}\right)\right]\right.$$

(Equation continued on next page)

$$\cdot \sum_{\alpha}\left[\left(x_{2\alpha}-E\left(x_{2\alpha}\right)\right)\left(\frac{z_{1\alpha}\sum_{\alpha} z_{2\alpha}^2}{D}-\frac{z_{2\alpha}\sum_{\alpha} z_{1\alpha}z_{2\alpha}}{D}\right)\right]\Bigg)$$

$$= E\left(\sum_{\alpha}\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)\left(x_{2\alpha}-E\left(x_{2\alpha}\right)\right)\left(\frac{z_{1\alpha}^2\left(\sum_{\alpha} z_{2\alpha}^2\right)^2}{D^2}\right.\right.\right.$$

$$\left.\left.\left.+\frac{-2z_{1\alpha}z_{2\alpha}\sum_{\alpha} z_{2\alpha}^2\sum_{\alpha} z_{1\alpha}z_{2\alpha}+z_{2\alpha}^2\left(\sum_{\alpha} z_{1\alpha}z_{2\alpha}\right)^2}{D^2}\right)\right]\right)$$

+ E (cross products).

Again, the expected value of the cross products equal zero.

$$= \sum_{\alpha}\left(E\left[\left(x_{1\alpha}-E\left(x_{1\alpha}\right)\right)\left(x_{2\alpha}-E\left(x_{2\alpha}\right)\right)\right]\left[\frac{z_{1\alpha}^2\left(\sum_{\alpha} z_{2\alpha}^2\right)^2}{D^2}\right.\right.$$

$$\left.\left.+\frac{-2z_{1\alpha}z_{2\alpha}\sum_{\alpha} z_{2\alpha}^2\sum_{\alpha} z_{1\alpha}z_{2\alpha}+z_{2\alpha}^2\left(\sum_{\alpha} z_{1\alpha}z_{2\alpha}\right)^2}{D^2}\right]\right) \tag{24}$$

$$= \sigma_{12}\left(\frac{\sum_{\alpha} z_{2\alpha}^2}{D}\left[\frac{\sum_{\alpha} z_{1\alpha}^2 \sum_{\alpha} z_{2\alpha}^2-2\left(\sum_{\alpha} z_{1\alpha}z_{2\alpha}\right)^2+\left(\sum_{\alpha} z_{1\alpha}z_{2\alpha}\right)^2}{D}\right]\right)$$

$$= \sigma_{12}\left[\frac{\sum_{\alpha} z_{2\alpha}^2}{D}\left(\frac{D}{D}\right)\right] = \sigma_{12}\ \frac{\text{cofactor } a_{11}}{D} = \sigma_{12}a^{11}.$$

Equation 24 gives the equation of the covariance between $\hat{b}_{11}$ and $\hat{b}_{21}$. By following the same technique we can show the covariance between $\hat{b}_{1\ell}$ and $\hat{b}_{2m}$ is equal to $\sigma_{12}a^{\ell m}$.

Let

$$C_1 = \left(c_{11} c_{12} \ldots c_{1q}\right), \qquad C_2 = \left(c_{21} c_{22} \ldots c_{2q}\right)$$

$$\hat{\beta}_1 = \left(\hat{b}_{11} \hat{b}_{12} \ldots b_{1q}\right), \qquad \hat{\beta}_2 = \left(\hat{b}_{21} \hat{b}_{22} \ldots b_{2q}\right).$$

From equation 22

$$E\left(\hat{\beta}_1\right) = \beta_1 = \left(\hat{b}_{11} \hat{b}_{12} \ldots \hat{b}_{1q}\right)$$

and

$$E\left(\hat{\beta}_2\right) = \beta_2 = \left(\hat{b}_{21} \hat{b}_{22} \ldots \hat{b}_{2q}\right).$$

The covariance between $\hat{\beta}_\ell$ and $\hat{\beta}_m$ $(\ell, m = 1, 2)$ is

$$\begin{aligned}
E\left[\left(\hat{\beta}_\ell - \beta_\ell\right)'\left(\hat{\beta}_m - \beta_m\right)\right] &= E\left[\left(\hat{\beta}_\ell - E\left(\hat{\beta}_\ell\right)\right)'\left(\hat{\beta}_m - E\left(\hat{\beta}_m\right)\right)\right] \\
&= E\left[\left(C_\ell A^{-1} - E\left(C_\ell A^{-1}\right)\right)\left(C_m A^{-1} - E\left(C_m A^{-1}\right)\right)\right] \\
&= E\left[\left(C_\ell - E\left(C_\ell\right)\right) A^{-1}\left(C_m - E\left(C_m\right)\right) A^{-1}\right] \\
&= E\left[\sum_{\alpha=1}^{N}\left(x_{\ell\alpha} - E\left(x_{\ell\alpha}\right)\right) Z'_\alpha A^{-1} \sum_{\nu=1}^{N}\left(x_{m\nu} - E\left(x_{m\nu}\right)\right) Z_\nu A^{-1}\right] \\
&= A^{-1} \sum_{\alpha, \nu=1}^{N}\left[E\left(x_{\ell\alpha} - E\left(x_{\ell\alpha}\right)\right)\left(x_{m\nu} - E\left(x_{m\nu}\right)\right) Z'_\alpha Z_\nu A^{-1}\right] \qquad (25) \\
&= A^{-1} \sum_{\alpha, \nu=1}^{N}\left[\delta_{\alpha, \nu} \sigma_{\ell m} Z'_\alpha Z_\nu A^{-1}\right], \\
&= A^{-1} \sum_{\alpha=1}^{N}\left[\sigma_{\ell m} Z'_\alpha Z_\alpha A^{-1}\right] \\
&= A^{-1} \sigma_{\ell m} A A^{-1} = \sigma_{\ell m} A^{-1},
\end{aligned}$$

where $\delta_{\alpha,\nu}$ is the Kronicker delta.

Equation 25 can be summarized into a covariance matrix[3] 26.

$$E\left[\left(\hat{\beta}_{\ell}-\beta_{\ell}\right)'\left(\hat{\beta}_{m}-\beta_{m}\right)\right] = \begin{pmatrix} \sigma_{11}A^{-1} & \sigma_{12}A^{-1} \\ \sigma_{21}A^{-1} & \sigma_{22}A^{-1} \end{pmatrix}. \tag{26}$$

For example, if one were interested in the covariance between $\hat{b}_{13}$ and $\hat{b}_{24}$ it would be equal to $\sigma_{12}a^{34}$.

5. <u>An Unbiased Estimator for σ_{ij}</u>

As pointed out earlier, the maximum likelihood estimator may be a biased estimator. Mood[2] points out that while there is no real harm in this, some claim that unbiasedness is a good property since it forces the distribution of the estimator to be centered at the true parameter value.

Let us first investigate the biasedness of the $\hat{\sigma}_{11}$ estimator given by equation 17.

$$\hat{\sigma}_{11} = \sum_{\alpha=1}^{N}\left(x_{1\alpha}-\sum_{g=1}^{q} b_{1g}z_{g\alpha}\right)^2 = \sum_{\alpha=1}^{N}\left[\left(x_{1\alpha}-\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}\right)+\left(\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}-\sum_{g=1}^{q}b_{1g}z_{g\alpha}\right)\right]^2$$

$$= \sum_{\alpha=1}^{N}\left[\left(x_{1\alpha}-\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}\right)^2+2\left(x_{1\alpha}-\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}\right)\sum_{g=1}^{q}\left(\hat{b}_{1g}-b_{1g}\right)z_{g\alpha} + \left(\sum_{g=1}^{q}\left(\hat{b}_{1g}-b_{1g}\right)z_{g\alpha}\right)^2\right] \tag{27}$$

From equation 15 we can demonstrate that

$$c_{kf}-\sum_{g=1}^{q}\hat{b}_{kg}a_{gf} = 0.$$

The cross product terms of equation 27 can be written in the above form differing only by a multiplication factor

$$2\sum_{h=1}^{q}\left(x_{1\alpha}-\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}\right)z_{h\alpha}\left(\hat{b}_{1h}-b_{1h}\right).$$

Hence, the cross product terms equal zero. Rewriting equation 27 we get

$$\sum_{\alpha=1}^{N}\left(x_{1\alpha}-\sum_{g=1}^{q}b_{1g}z_{g\alpha}\right)^2 = \sum_{\alpha=1}^{N}\left(x_{1\alpha}-\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}\right)^2+\sum_{\alpha=1}^{N}\left[\sum_{g=1}^{q}\left(\hat{b}_{1g}-b_{1g}\right)z_{g\alpha}\right]^2$$

or

$$\sum_{\alpha=1}^{N}\left(x_{1\alpha}-\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}\right)^2 = \sum_{\alpha=1}^{N}\left(x_{1\alpha}-\sum_{g=1}^{q}b_{1g}z_{g\alpha}\right)^2-\sum_{\alpha=1}^{N}\left[\sum_{g=1}^{q}\left(\hat{b}_{1g}-b_{1g}\right)z_{g\alpha}\right]^2. \qquad (28)$$

Let us now divide both sides of equation 28 by N and take the expected value.

$$E\left[\frac{1}{N}\sum_{\alpha=1}^{N}\left(x_{1\alpha}-\sum_{g=1}^{q}\hat{b}_{1g}z_{g\alpha}\right)^2\right] = E\left[\frac{1}{N}\sum_{\alpha=1}^{N}\left(x_{1\alpha}-\sum_{g=1}^{q}b_{1g}z_{g\alpha}\right)^2\right]$$

$$-E\left(\frac{1}{N}\sum_{\alpha=1}^{N}\left[\sum_{g=1}^{q}\left(\hat{b}_{1g}-b_{1g}\right)z_{g\alpha}\right]^2\right)$$

$$=\frac{1}{N}\left(\sum_{\alpha=1}^{N}\sigma_{11}\right)-\frac{1}{N}E\left(\sum_{g=1}^{q}\sum_{h=1}^{q}\left[\left(\hat{b}_{1g}-b_{1g}\right)\left(\hat{b}_{1h}-b_{1h}\right)\sum_{\alpha=1}^{N}z_{g\alpha}z_{h\alpha}\right]\right).$$

From equation 23 we may write

$$E\left(\hat{\sigma}_{11}\right) = \sigma_{11}-\frac{1}{N}\left[\sigma_{11}\frac{\text{cofactor } a_{11}}{D}a_{11}+\sigma_{11}\frac{\text{cofactor } a_{12}}{D}a_{12}+\ldots\right.$$

(Equation continued on next page)

$$+\sigma_{11}\frac{\text{cofactor } a_{1q}}{D}a_{1q}+\sigma_{11}\frac{\text{cofactor } a_{21}}{D}a_{21}+\ldots+\sigma_{11}\frac{\text{cofactor } a_{qq}}{D}a_{qq}\Bigg]$$

$$=\sigma_{11}-\frac{1}{N}\Bigg[\sigma_{11}\left(\frac{\left(\text{cofactor } a_{11}\right)a_{11}+\left(\text{cofactor } a_{12}\right)a_{12}+\ldots\left(\text{cofactor } a_{1q}\right)a_{1q}}{D}\right)$$

$$+\sigma_{11}\left(\frac{\left(\text{cofactor } a_{21}\right)a_{21}+\ldots\left(\text{cofactor } a_{q1}\right)a_{21}}{D}\right)+\ldots$$

$$+\sigma_{11}\left(\frac{\left(\text{cofactor } a_{q1}\right)a_{q1}+\ldots\left(\text{cofactor } a_{qq}\right)a_{qq}}{D}\right)\Bigg]$$

$$=\sigma_{11}-\frac{1}{N}\left[\sigma_{11}(1+1+\ldots+1)\right]$$

$$E\left(\hat{\sigma}_{11}\right)=\frac{N\sigma_{11}}{N}-\frac{q\sigma_{11}}{N}=\frac{N-q}{N}\sigma_{11}.$$

If we multiply $E\left(\hat{\sigma}_{11}\right)$ by $\frac{N}{N-q}$, this gives us an unbiased estimator of σ_{11}, and the expected value of this estimator equals σ_{11}, thus the estimator is unbiased. Using this same type of argument, we can show the general form of the unbiased $\hat{\sigma}_{ij}$ estimator to be

$$\hat{\sigma}_{ij}=\frac{1}{N-q}\sum_{\alpha=1}^{N}\left(x_{i\alpha}-\sum_{g=1}^{N}\hat{b}_{ig}z_{g\alpha}\right)\left(x_{j\alpha}-\sum_{g=1}^{q}\hat{b}_{jg}z_{g\alpha}\right). \tag{29}$$

6. Predictions and the Confidence Ellipse

The regression equations $x_1=\sum_{g=1}^{q}b_{1g}z_g$ and $x_2=\sum_{g=1}^{q}b_{2g}z_g$ have been estimated by $x_1=\sum_{g=1}^{q}\hat{b}_{1g}z_g$ and $x_2=\sum_{g=1}^{q}\hat{b}_{2g}z_g$ from a sample of N sets of observations. We are now interested in predicting an x_1 and x_2 given a particular set of z's. These predictions are quite straightforward from the above estimated regression equations; however, such predictions are useful only if they actually reproduce the true value of the dependent variables or are within

certain limits of error specified by the experimenter. If this particular set of z's is known to have been drawn from the same population as the z's used in determining the regression equations, then it is possible to set up a confidence statement regarding the possible error associated with this joint prediction.

Let the estimated predicted value for the ith (i = 1, 2) dependent variable equal x_{ip}. Then

$$x_{ip} = \sum_{g=1}^{q} \hat{b}_{ig} z_g \qquad \text{and} \qquad E\left(x_{ip}\right) = \sum_{g=1}^{q} b_{ig} z_g .$$

Since x_{ip} is equal to a linear combination of the $\hat{b}$'s which are normally distributed, then x_{ip} is normally distributed.
The variates

$$u_i = x_i - \sum_{g=1}^{q} \hat{b}_{ig} z_g, \qquad i = 1, 2 \tag{30}$$

are also normally distributed since they are linear functions of normally distributed variables. Again, by the same type of reasoning the joint distribution of $u_1 u_2$ is normal.

Now

$$\begin{aligned} E\left(u_i\right) &= E\left[x_i - \sum_{g=1}^{q} \hat{b}_{ig} z_g\right] \\ &= E\left(x_i\right) - E\left[\sum_{g=1}^{q} \hat{b}_{ig} z_g\right] \\ &= \sum_{g=1}^{q} b_{ig} z_g - \sum_{g=1}^{q} b_{ig} z_g = 0. \end{aligned}$$

The variance, denoted by σ_{ip}^2, of u_i is

$$\sigma_{ip}^2 = E\left[\left(u_i - E\left(u_i\right)\right)^2\right] = E\left[u_i^2 - 2u_i E\left(u_i\right) + \left(E\left(u_i\right)\right)^2\right] = E\left(u_i^2\right)$$

$$= E\left[\left(x_i - \sum_{g=1}^{q} \hat{b}_{ig} z_g\right)^2\right] = E\left(\left[\left(x_i - \sum_{g=1}^{q} b_{ig} z_g\right) + \left(\sum_{g=1}^{q} b_{ig} z_g - \sum_{g=1}^{q} \hat{b}_{ig} z_g\right)\right]^2\right)$$

$$= E\left(\left[\left(x_i - \sum_{g=1}^{q} b_{ig} z_g\right) + \left(\sum_{g=1}^{q} \left(b_{ig} - \hat{b}_{ig}\right) z_g\right)\right]^2\right) \tag{31}$$

$$= E\left[\left(x_i - \sum_{g=1}^{q} b_{ig} z_g\right)^2\right] + 2E\left[\left(x_i - \sum_{g=1}^{q} b_{ig} z_g\right) \sum_{g=1}^{q} \left(b_{ig} - \hat{b}_{ig}\right) z_g\right]$$

$$+ E\left[\sum_{g=1}^{q} \sum_{h=1}^{q} \left(b_{ig} - \hat{b}_{ig}\right)\left(b_{ih} - \hat{b}_{ih}\right) z_g z_h\right]$$

Now

$$E\left[\left(x_i - \sum_{g=1}^{q} b_{ig} z_g\right) \sum_{g=1}^{q} \left(b_{ig} - \hat{b}_{ig}\right) z_g\right] = E\left[\sum_{g=1}^{q} \left(b_{ig} - \hat{b}_{ig}\right)\left(x_i - \sum_{h=1}^{q} b_{ih} z_h\right) z_g\right]. \tag{32}$$

Since x_i is independent of the $\hat{b}$'s, we may write

$$E\left[\sum_{g=1}^{q} \left(b_{ig} - \hat{b}_{ig}\right)\left(x_i - \sum_{h=1}^{q} b_{ih} z_h\right) z_g\right] = \sum_{g=1}^{q} E\left(b_{ig} - \hat{b}_{ig}\right) E\left(x_i - \sum_{h=1}^{q} b_{ih} z_h\right) z_g = 0.$$

Hence, equation 31 becomes

$$\sigma_{ip}^2 = E\left[\left(x_i - \sum_{g=1}^{q} b_{ig} z_g\right)^2\right] + E\left[\sum_{g=1}^{q} \sum_{h=1}^{q} \left(b_{ig} - \hat{b}_{ig}\right)\left(b_{ih} - \hat{b}_{ih}\right) z_g z_h\right]$$

$$\sigma_{ip}^2 = \sigma_{ii} + \sum_{g=1}^{q} \sum_{h=1}^{q} \sigma_{ii} a^{gh} z_g z_h.$$

σ^2_{ip} is approximated by s^2_{ip}, where

$$s^2_{ip} = \hat{\sigma}_{ii} + \sum_{g=1}^{q} \sum_{h=1}^{q} \hat{\sigma}_{ii} a^{gh} z_g z_h \tag{33}$$

and $\hat{\sigma}_{ii}$ is the unbiased estimator.

Now

$$\begin{aligned}
E(u_1u_2) &= E\left[\left(x_1 - \sum_{g=1}^{q} \hat{b}_{1g} z_g\right)\left(x_2 - \sum_{h=1}^{q} \hat{b}_{2h} z_h\right)\right] \\
&= E\left(\left[\left(x_1 - \sum_{g=1}^{q} b_{1g} z_g\right) + \left(\sum_{g=1}^{q} (b_{1g} - \hat{b}_{1g}) z_g\right)\right]\left[\left(x_2 - \sum_{h=1}^{q} b_{2h} z_h\right) + \left(\sum_{h=1}^{q} (b_{2h} - \hat{b}_{2h}) z_h\right)\right]\right) \\
&= E\left[\left(x_1 - \sum_{g=1}^{q} b_{1g} z_g\right)\left(x_2 - \sum_{h=1}^{q} b_{2h} z_h\right)\right] + E\left(x_1 - \sum_{g=1}^{q} b_{1g} z_g\right) E\left(\sum_{h=1}^{q} (b_{2h} - \hat{b}_{2h}) z_h\right) \qquad (34) \\
&\quad + E\left(x_2 - \sum_{h=1}^{q} b_{2h} z_h\right) E\left(\sum_{g=1}^{q} (b_{1g} - \hat{b}_{1g}) z_g\right) + E\left[\sum_{g=1}^{q} \sum_{h=1}^{q} (b_{1g} - \hat{b}_{1g})(b_{2h} - \hat{b}_{2h}) z_g z_h\right] \\
&= \sigma_{12} + \sigma_{12} \sum_{g=1}^{q} \sum_{h=1}^{q} a^{gh} z_g z_h .
\end{aligned}$$

The correlation between u_1 and u_2 is denoted by ρ_p and is

$$\rho_p = \frac{E(u_1u_2)}{\sqrt{E(u_1^2)E(u_2^2)}} = \frac{\sigma_{12}\left(1 + \sum_{g=1}^{q} \sum_{h=1}^{q} a^{gh} z_g z_h\right)}{\sqrt{\sigma_{11}\sigma_{12}\left(1 + \sum_{g=1}^{q} \sum_{h=1}^{q} a^{gh} z_g z_h\right)^2}} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{12}}} .$$

ρ_p is approximated by r_p, where

$$r_p = \frac{\hat{\sigma}_{12}}{\sqrt{\hat{\sigma}_{11}\hat{\sigma}_{12}}} \tag{35}$$

The joint distribution function of $u_1 u_2$ is given by equation 36.

$$g(u_1, u_2) = \frac{1}{2\pi\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}} \exp\left\{-\frac{1}{2(1-\rho_p^2)}\left[\left(\frac{u_1 - E(u_1)}{\sigma_{1p}}\right)^2 - 2\rho_p \frac{(u_1 - E(u_1))(u_2 - E(u_2))}{\sigma_{1p}\sigma_{2p}} + \left(\frac{u_2 - E(u_2)}{\sigma_{2p}}\right)^2\right]\right\}. \tag{36}$$

Since u_1 is a function of x_1, and u_2 a function of x_2 we can write equation 36 as

$$g(x_1, x_2) = \frac{1}{2\pi\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}} \exp\left\{-\frac{1}{2(1-p_p^2)}\left[\left(\frac{x_1 - \sum_{g=1}^{q}\hat{b}_{1g}z_g}{\sigma_{1p}}\right)^2 - 2\rho_p \frac{\left(x_1 - \sum_{g=1}^{q}\hat{b}_{1g}z_g\right)\left(x_2 - \sum_{g=1}^{q}\hat{b}_{2g}z_g\right)}{\sigma_{1p}\sigma_{2p}} + \left(\frac{x_2 - \sum_{g=1}^{q}\hat{b}_{2g}z_g}{\sigma_{2p}}\right)^2\right]\right\}. \tag{37}$$

Under a linear orthogonal transformation it is possible to find two normally distributed and stochastically independent linear functions of the variables[1] x_1 and x_2.

Let

$$x_1 - \sum_{g=1}^{q}\hat{b}_{1g}z_g = y_1 \cos\theta - y_2 \sin\theta$$

and

$$x_2 - \sum_{g=1}^{q} \hat{b}_{2g} z_g = y_1 \sin\theta + y_2 \cos\theta,$$

where y_1 and y_2 are the transformed variables. The joint distribution function of $g(x_1, x_2)$ becomes

$$h\left[x_1(y_1, y_2), x_2(y_1, y_2)\right] \quad |J|,$$

where J is the Jacobian of the transformation.

Now

$$|J| = \begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{vmatrix} = \cos^2\theta + \sin^2\theta = 1;$$

hence, equation 37 becomes

$$h(y_1 y_2) = \frac{1}{2\pi\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_2^2}} e^{-\frac{1}{2}G(y_1, y_2)}, \tag{38}$$

where

$$G(y_1, y_2) = \frac{1}{1-\rho_2^2}\left[y_1^2\left(\frac{\cos^2\theta}{\sigma_{1p}^2} + \frac{\sin^2\theta}{\sigma_{2p}^2} - \rho_p \frac{\sin 2\theta}{\sigma_{1p}\sigma_{2p}}\right)\right.$$
$$\left. + y_2^2\left(\frac{\sin^2\theta}{\sigma_{1p}^2} + \frac{\cos^2\theta}{\sigma_{2p}^2} + \rho_p \frac{\sin 2\theta}{\sigma_{1p}\sigma_{2p}}\right) - y_1 y_2\left(\frac{\sin 2\theta}{\sigma_{1p}^2} - \frac{\sin 2\theta}{\sigma_{2p}^2} + \frac{2\rho_p \cos 2\theta}{\sigma_{1p}\sigma_{2p}}\right)\right].$$

Since the correlation coefficient of (y_1, y_2) depends on θ, let us solve θ for which the correlation coefficient is zero. This is done by equating the coefficient of y_1, y_2 to zero in equation 38.

$$\left(\frac{1}{\sigma_{1p}^2}-\frac{1}{\sigma_{2p}^2}\right)\sin 2\theta + \frac{2\rho_p}{\sigma_{1p}\sigma_{2p}}\cos 2\theta = 0$$

$$\tan 2\theta = \frac{2\rho_p\sigma_{1p}\sigma_{2p}}{\sigma_{1p}^2-\sigma_{2p}^2}\,. \tag{39}$$

With this value of θ, y_1 and y_2 become stochastically independent, so that equation 38 becomes

$$f(y_1,y_2) = \frac{1}{2\pi\sigma_{y_1}\sigma_{y_2}}\exp\left[-\frac{1}{2}\left(\frac{y_1^2}{\sigma_{y_1}^2}+\frac{y_2^2}{\sigma_{y_2}^2}\right)\right], \tag{40}$$

where $\sigma_{y_1}^2$ and $\sigma_{y_2}^2$ are the variances of y_1, y_2, respectively.

If we set y_1 and y_2 equal to zero in equation 38 and equation 40 we get

$$\sigma_{y_1}\sigma_{y_2} = \sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}\,. \tag{41}$$

Similarly, setting y_1 and y_2 equal to 1 and inserting the relationship of equation 41 into equation 40 we get

$$\frac{1}{\sigma_{y_1}^2}+\frac{1}{\sigma_{y_2}^2} = \frac{1}{1-\rho_p^2}\left(\frac{1}{\sigma_{1p}^2}+\frac{1}{\sigma_{2p}^2}\right)\,. \tag{42}$$

From equations 41 and 42 we obtain the following two equations in two unknowns:

$$\left(\sigma_{y_1}+\sigma_{y_2}\right)^2 = \sigma_{1p}^2+\sigma_{2p}^2+2\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}$$

$$\left(\sigma_{y_1}-\sigma_{y_2}\right)^2 = \sigma_{1p}^2+\sigma_{2p}^2-2\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}\,,$$

from which σ_{y_1} and σ_{y_2} are determined. Hence,

$$\sigma_{y_1} = \frac{\sqrt{\sigma_{1p}^2+\sigma_{2p}^2+2\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}} + \sqrt{\sigma_{1p}^2+\sigma_{2p}^2-2\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}}}{2}$$

and is approximated by

$$s_{y_1} = \frac{\sqrt{s_{1p}^2+s_{2p}^2+2s_{1p}s_{2p}\sqrt{1-r_p^2}} + \sqrt{s_{1p}^2+s_{2p}^2-2s_{1p}s_{2p}\sqrt{1-r_p^2}}}{2},$$

and

$$\sigma_{y_2} = \frac{\sqrt{\sigma_{1p}^2+\sigma_{2p}^2+2\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}} - \sqrt{\sigma_{1p}^2+\sigma_{2p}^2-2\sigma_{1p}\sigma_{2p}\sqrt{1-\rho_p^2}}}{2}$$

and is approximated by

$$s_{y_2} = \frac{\sqrt{s_{1p}^2+s_{2p}^2+2s_{1p}s_{2p}\sqrt{1-r_p^2}} - \sqrt{s_{1p}^2+s_{2p}^2-2s_{1p}s_{2p}\sqrt{1-r_p^2}}}{2}.$$

It can be shown that the quadratic form of the exponential of a bivariate normal distribution is distributed as chi-square with two degrees of freedom.[2] The 95% probability that x_1, x_2 lie inside some region is therefore

$$P\left(\frac{1}{1-\rho_p^2}\left[\left(\frac{x_1-\sum_{g=1}^{q}\hat{b}_{1g}z_g}{\sigma_{1p}}\right)^2 - \frac{2\rho_p\left(x_1-\sum_{g=1}^{q}\hat{b}_{1g}z_g\right)\left(x_2-\sum_{g=1}^{q}\hat{b}_{2g}z_g\right)}{\sigma_{1p}\sigma_{2p}} + \left(\frac{x_2-\sum_{g=1}^{q}\hat{b}_{2g}z_g}{\sigma_{2p}}\right)^2\right] \le \chi^2_{.95}\right) = .95 \qquad (43)$$

Equation 43 indicates this region is an ellipse and its boundary is given by

$$\frac{1}{1-\rho_p^2}\left[\left(\frac{x_1-\sum_{g=1}^{q}\hat{b}_{1g}z_g}{\sigma_{1p}}\right)^2-\frac{2\rho_p\left(x_1-\sum_{g=1}^{q}\hat{b}_{1g}z_g\right)\left(x_2-\sum_{g=1}^{q}\hat{b}_{2g}z_g\right)}{\sigma_{1p}\sigma_{2p}}\right.$$

$$\left.+\left(\frac{x_2-\sum_{g=1}^{q}\hat{b}_{2g}z_g}{\sigma_{2p}}\right)^2\right]=5.991, \qquad (44)$$

since $\chi^2_{.95}=5.991$ for two degrees of freedom.

The application of equation 44 assumes that the parameters σ_{1p}, σ_{2p}, and ρ_p are known. Since this is unlikely, equations 43 and 44 must be modified. If we replace these population parameters by their estimators s_{1p}, s_{2p}, and r_p in the quadratic form of the exponential, this quadratic form is then distributed as Hotelling's T^2 and

$$T^2=\frac{1}{1-r_p^2}\left[\left(\frac{x_1-\sum_{g=1}^{q}\hat{b}_{1g}z_g}{s_{1p}}\right)^2-\frac{2r_p\left(x_1-\sum_{g=1}^{q}\hat{b}_{1g}z_g\right)\left(x_2-\sum_{g=1}^{q}\hat{b}_{2g}z_g\right)}{s_{1p}s_{2p}}\right.$$

$$\left.+\left(\frac{x_2-\sum_{g=1}^{q}\hat{b}_{2g}z_g}{s_{2p}}\right)^2\right]. \qquad (45)$$

The variable $\frac{N-2}{N-1}\,\frac{T^2}{2}$ has an "F" distribution with 2 and N-2 degrees of freedom. Thus, using the "F" distribution, one can compute the appropriate confidence regions.

Using the linear orthogonal transformation, we have shown that the

exponential of the bivariate normal (equation 37) can be replaced by

$$\left(\frac{y_1^2}{\sigma_{y_1}^2} + \frac{y_2^2}{\sigma_{y_2}^2}\right) \text{and estimated by} \left(\frac{y_1^2}{s_{y_1}^2} + \frac{y_2^2}{s_{y_2}^2}\right),$$

where y_1^2, y_2^2, $s_{y_1}^2$, $s_{y_2}^2$ have been defined earlier. Because T has the property of invariance under homogeneous linear transformations of the variates,[5] the boundary of the 95% confidence ellipse becomes

$$\left(\frac{y_1^2}{s_{y_1}^2} + \frac{y_2^2}{s_{y_2}^2}\right) = T^2 = 2h\left(F_{.95};\ 2,\ N-2\right) \frac{N-1}{N-2}\ . \tag{46}$$

Hotelling[5] also points out that for large N, s_{y_1} and s_{y_2} are sufficiently accurate estimates of σ_{y_1} and σ_{y_2} to permit the use of the χ^2 distribution.

While this linear orthogonal transformation was not really necessary to construct the confidence ellipse, it certainly facilitates the actual graphing of this function.

Figure 1 illustrates a prediction and its confidence ellipse.

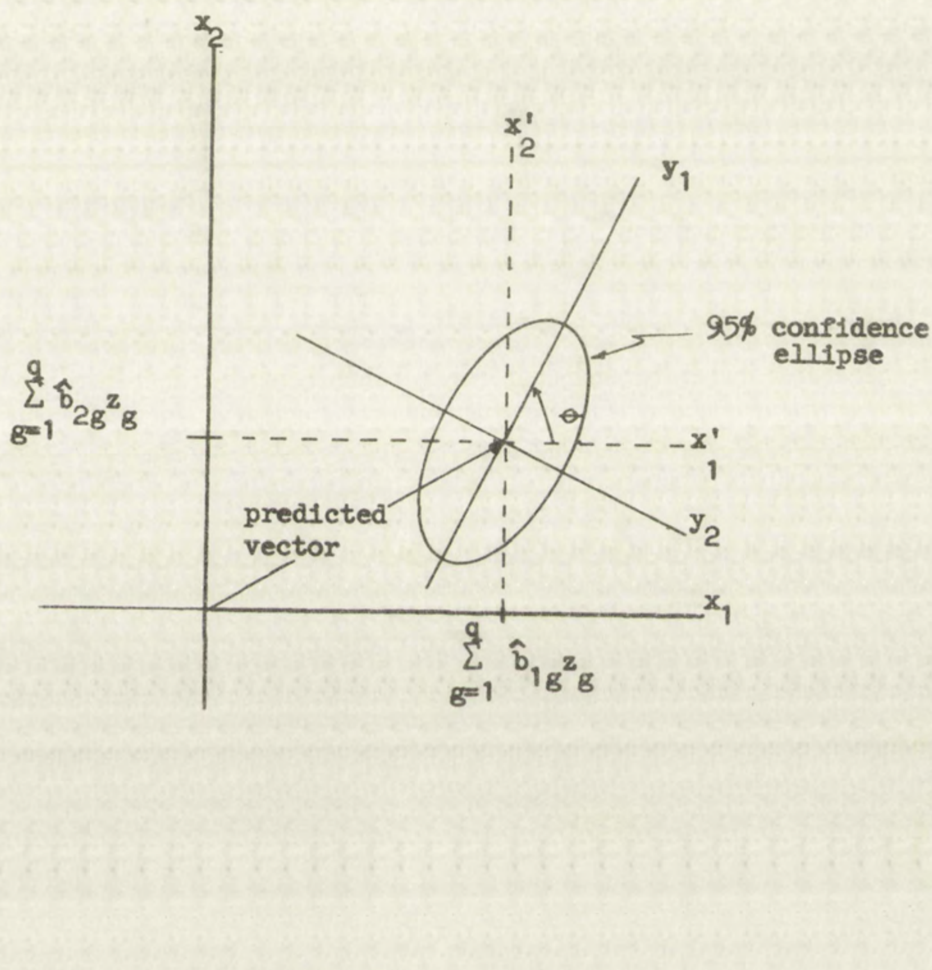

Fig. 1.—Example of prediction vector and its 95% confidence ellipse.

CHAPTER II

A METEOROLOGICAL APPLICATION OF THE THEORY

1. Introduction

The problem of weather forecasting can be characterized into three classes:[6] (1) synoptic-empirical, (2) physical-mathematical, and (3) statistical. The synoptic-empirical approach is strictly a subjective type of forecasting based on the knowledge and experience of the forecaster. Although this method has been somewhat successful, it does leave much to be desired. The physical-mathematical approach is based on formulating the physical laws that govern the motions of the atmosphere. This approach is still in its infancy but has already accomplished revolutionary strides in meteorology.[7] The statistical method, with which we will be mainly concerned, argues that, under a set of past parameters, the atmosphere behaved in a certain way, and this behavior should then be reproducible in the future when preceded by a similar set of parameters. It should be pointed out that the statistical approach enables one to associate a probability statement with each prediction.

The statistical approach developed in Chapter I might be applied to a variety of problems of prediction; however, the specific problem with which we will be concerned is that of making a 12 hour prediction of the wind hodograph at the Bikini Atoll of Eniwetok Proving Grounds, located in the Marshall Islands.

2. The Data Problem

The collection of data was a serious problem which in many ways hampered the solution of the prediction problem.

First, the weather reporting stations in the proving ground and surrounding areas are widely scattered. The closest network we were able to obtain included the following stations: Bikini, Eniwetok, Kusaie, Kwajalein, and Majuro. The distances between these stations range from approximately 130 to 600 miles.

Second, the observations taken at these stations were not always complete, and often, because of mechanical difficulties, data were not recorded. Although most of the observations were taken from land, those at Bikini were taken from aboard ship, which may or may not have introduced errors.

Finally, the observing equipment and techniques were not precise. Rapp,[8] in an attempt to evaluate the reliability of wind data, found that under laboratory conditions the average probable error of the east-west component was 1.24 knots and of the north-south component, 1.62 knots. However, in considering data available in the field, there are other sources of error which are introduced in the coding and transmission steps. Palmer et al.[9] attempted to assess errors in the winds under field conditions and found an average probable error of approximately 4 knots in the wind vector.

A set of data taken during the "Redwing" operation (1956) was used as the basic population. Because of the variability in the wind patterns from season to season, we chose to work with a two month period, namely, June and July. It was also recognized that there undoubtedly exists a between-years within-season variability, but since no data were available from other years from all the stations, we were forced to use this small population.

The winds were normally observed with GMD-1A equipment at each station and reported in angles and speed in 5000 foot increments above

2000 feet and more frequent intervals at the lower altitude. Because of equipment failure, balloon burst, etc., all the wind observations did not terminate at the same altitude; thus, we did not have the same number of observations at each level.

Each wind observation was resolved into its u and v components, and an average u and v component for a 10,000 foot layer was computed at the 10,000, 20,000, 30,000, and 40,000 foot levels. For example, the average u component at the 20,000 foot level was computed as follows:

$$u_{\overline{20}} = \frac{u_{15}+u_{20}+u_{25}}{3},$$

where u_g equals the u component at g-thousand feet. If any of the u_g components was missing, the denominator was reduced accordingly. Using these average layer wind components, u and v, had the following advantages: (1) The wind component represented a layer instead of a particular level, (2) the effect of small scale turbulence at a particular level was reduced in the averaging process, and (3) the effect of small errors in the observing system was reduced.

Summarizing, we collected a set of data consisting of u and v components for four levels, at each of five stations, observed every 6 hours during the two month period with the exception of missing data. Hypothesizing that each u and v component at each level at each station during this two month period was distributed normally, the first four moments were computed and tests of normality were performed.

The χ^2 goodness of fit test[10] is a comparison of an observed sample distribution with a theoretical frequency distribution, which in our case is the normal distribution. This test requires the construction of a frequency distribution of the sample observations, which was performed according to

Mann and Wald,[11] who suggest that class intervals be chosen so that the expected number of observations is the same for all classes. If $\sum_{i=1}^{P} \frac{\left(f_i - F_i\right)^2}{F_i}$, where P is the number of classes, f_i the observed frequency for the ith class, and F_i the theoretical frequency for the ith class, is greater than $\chi^2_{.99}$ for P-3 degrees of freedom, the hypothesis that the sample distribution is normal is rejected.

An alternate test for normality is the testing of the third and fourth moments. Skewness, a measure of symmetry, has a value 0, and kurtosis, a measure of peakedness, has a value 3 for the normal distribution. For convenience we rescaled our kurtosis values to a 0 base. The standard deviation of skewness and kurtosis was also calculated so that the hypothesis that skewness or kurtosis of the sample distribution equals 0 could be tested for normality.[12]

Computations were done on the IBM 704 computer and the results appear in tables 1 through 5. With the exception of a few cases, there is no evidence to reject the hypothesis of normality of the data at these five stations. Wind data were available from the island of Wake and are presented in table 6 to illustrate the extent of non-normality of the wind that can exist at a station.

A sample of the average u and v components for each of the four levels at Bikini was taken and plotted. (See figures 2 through 5.) A 95% confidence ellipse was also plotted, based on the sample moments of the assumed bivariate normal distribution. These figures illustrate the correlation which exists between the u and v components and a sample of the rotated elliptical type of distribution of the observed winds found in that particular area of the globe.

TABLE 1

BIKINI WIND STATISTICS FOR JUNE AND JULY 1956

Level	Component	N	Average	Standard Deviation	Skewness	S. D. of Skewness	Kurtosis	S. D. of Kurtosis	$\sum_{i=1}^{10} \frac{(f_i - F_i)^2}{F_i}$	$\chi^2_{.99}(7)$
10,000	$u_{\overline{10}}$	177	-10.78	4.98	.066	.183	-.414	.363	18.31	18.48
	$v_{\overline{10}}$	177	.19	4.02	.293	.183	.603	.363	8.43	18.48
20,000	$u_{\overline{20}}$	177	- 2.77	6.13	-.015	.183	-.669	.363	7.80	18.48
	$v_{\overline{20}}$	177	.61	6.02	-.129	.183	.369	.363	7.92	18.48
30,000	$u_{\overline{30}}$	177	5.98	8.54	.336	.183	.520	.363	8.82	18.48
	$v_{\overline{30}}$	177	2.58	8.32	-.220	.183	-.321	.363	7.30	18.48
40,000	$u_{\overline{40}}$	176	14.11	14.08	.036	.183	-.005	.364	10.36	18.48
	$v_{\overline{40}}$	176	3.78	10.52	-.022	.183	-.686	.364	5.63	18.48

TABLE 2

ENIWETOK WIND STATISTICS FOR JUNE AND JULY 1956

Level	Component	N	Average	Standard Deviation	Skewness	S. D. of Skewness	Kurtosis	S. D. of Kurtosis	$\sum_{i=1}^{10} \frac{(f_i - F_i)^2}{F_i}$	$\chi^2_{.99}(7)$
10,000	$u_{\overline{10}}$	197	-11.15	5.75	-.195	.173	.125	.345	8.63	18.48
	$v_{\overline{10}}$	197	1.52	4.32	.506[a]	.173	2.00[a]	.345	9.41	18.48
20,000	$u_{\overline{20}}$	197	- 4.47	6.15	-.083	.173	-.649	.345	9.85	18.48
	$v_{\overline{20}}$	197	.04	5.95	-.041	.173	-.508	.345	3.59	18.48
30,000	$u_{\overline{30}}$	198	2.77	10.44	-.068	.173	-.192	.344	9.27	18.48
	$v_{\overline{30}}$	198	1.31	8.76	-.110	.173	-.796	.344	17.96	18.48
40,000	$u_{\overline{40}}$	197	9.93	16.26	-.285	.173	.002	.345	9.14	18.48
	$v_{\overline{40}}$	197	3.17	11.47	.127	.173	-.741	.345	9.92	18.48

[a]The hypothesis that the sample moments are the same as the moments of the normal distribution is rejected under a "t" test.

TABLE 3

KUSAIE WIND STATISTICS FOR JUNE AND JULY 1956

Level	Component	N	Average	Standard Deviation	Skewness	S. D. of Skewness	Kurtosis	S. D. of Kurtosis	$\sum_{i=1}^{10} \frac{\left(f_i - F_i\right)^2}{F_i}$	$\chi^2_{.99}(7)$
10,000	$u_{\overline{10}}$	199	-11.23	6.37	-.205	.172	.492	.343	10.10	18.48
	$v_{\overline{10}}$	199	3.84	3.80	-.117	.172	.743	.343	21.56[b]	18.48
20,000	$u_{\overline{20}}$	200	-10.62	6.02	-.201	.172	.737	.342	6.50	18.48
	$v_{\overline{20}}$	200	4.27	5.02	-.099	.172	.173	.342	6.10	18.48
30,000	$u_{\overline{30}}$	199	- 6.64	7.99	.460[a]	.172	.681	.343	16.63	18.48
	$v_{\overline{30}}$	199	3.22	4.58	-.551[a]	.172	1.32[a]	.343	11.90	18.48
40,000	$u_{\overline{40}}$	200	.46	12.38	.082	.172	-.643	.342	6.90	18.48
	$v_{\overline{40}}$	200	1.03	6.62	.330	.172	-.089	.342	8.57	18.48

[a]The hypothesis that the sample moments are the same as the moments of the normal distribution is rejected under a "t" test.

[b]The hypothesis that the sample distribution is normal is rejected under the χ^2 test at the 99% confidence level.

TABLE 4

KWAJALEIN WIND STATISTICS FOR JUNE AND JULY 1956

Level	Component	N	Average	Standard Deviation	Skewness	S. D. of Skewness	Kurtosis	S. D. of Kurtosis	$\sum_{i=1}^{10} \frac{(f_i - F_i)^2}{F_i}$	$\chi^2_{.99}(7)$
10,000	$u_{\overline{10}}$	190	-12.35	6.07	-.105	.176	-.013	.351	12.63	18.48
	$v_{\overline{10}}$	190	.70	3.96	-.324	.176	.924[a]	.351	5.47	18.48
20,000	$u_{\overline{20}}$	171	- 6.36	5.26	-.126	.186	.293	.369	12.92	18.48
	$v_{\overline{20}}$	171	.88	5.79	.190	.186	.588	.369	14.09	18.48
30,000	$u_{\overline{30}}$	166	- 1.32	7.14	.144	.188	-.553	.375	7.13	18.48
	$v_{\overline{30}}$	166	2.67	6.72	-.441	.188	.343	.375	13.28	18.48
40,000	$u_{\overline{40}}$	161	6.60	12.42	-.195	.191	.076	.380	4.40	18.48
	$v_{\overline{40}}$	161	3.47	8.43	-.194	.191	-.052	.380	6.02	18.48

[a]The hypothesis that the sample moments are the same as the moments of the normal distribution is rejected under a "t" test.

TABLE 5

MAJURO WIND STATISTICS FOR JUNE AND JULY 1956

Level	Component	N	Average	Standard Deviation	Skewness	S. D. of Skewness	Kurtosis	S. D. of Kurtosis	$\sum_{i=1}^{10} \frac{(f_i - F_i)^2}{F_i}$	$\chi^2_{.99}(7)$
10,000	$u_{\overline{10}}$	204	-12.71	6.23	-.009	.170	-.208	.339	6.39	18.48
	$v_{\overline{10}}$	204	.86	4.11	1.08[a]	.170	3.23[a]	.339	9.40	18.48
20,000	$u_{\overline{20}}$	195	- 8.92	7.17	.186	.174	-.124	.346	5.77	18.48
	$v_{\overline{20}}$	195	1.15	5.61	.448	.174	-.178	.346	18.17	18.48
30,000	$u_{\overline{30}}$	193	- 4.23	8.74	.007	.175	-.696	.348	14.41	18.48
	$v_{\overline{30}}$	193	2.91	5.77	-.181	.175	-.873	.348	13.68	18.48
40,000	$u_{\overline{40}}$	191	5.47	12.73	.151	.176	-.154	.350	2.56	18.48
	$v_{\overline{40}}$	191	2.87	7.42	-.0003	.176	1.156[a]	.350	11.16	18.48

[a]The hypothesis that the sample moments are the same as the moments of the normal distribution is rejected under a "t" test.

TABLE 6

WAKE WIND STATISTICS FOR JUNE AND JULY 1956

Level	Component	N	Average	Standard Deviation	Skewness	S. D. of Skewness	Kurtosis	S. D. of Kurtosis	$\sum_{i=1}^{10} \frac{(f_i - F_i)^2}{F_i}$	$\chi^2_{.99}(7)$
10,000	$u_{\overline{10}}$	204	-8.47	5.19	-.444[a]	.170	-.215	.339	17.57	18.48
	$v_{\overline{10}}$	204	.90	5.64	1.024[a]	.170	2.580[a]	.339	19.32[b]	18.48
20,000	$u_{\overline{20}}$	199	-3.40	7.66	-.561[a]	.172	.027	.343	18.17	18.48
	$v_{\overline{20}}$	199	-1.90	8.21	.949[a]	.172	2.689[a]	.343	18.84[b]	18.48
30,000	$u_{\overline{30}}$	198	1.81	13.86	.209	.173	-.680	.344	29.78[b]	18.48
	$v_{\overline{30}}$	198	-4.29	13.24	.508[a]	.173	-.252	.344	21.24[b]	18.48
40,000	$u_{\overline{40}}$	191	7.17	23.98	.229	.176	-1.196[a]	.350	35.54[b]	18.48
	$v_{\overline{40}}$	191	-4.98	19.67	.287	.176	-.395	.350	13.37	18.48

[a]The hypothesis that the sample moments are the same as the moments of the normal distribution is rejected under a "t" test.

[b]The hypothesis that the sample distribution is normal is rejected under the χ^2 test at the 99% confidence level.

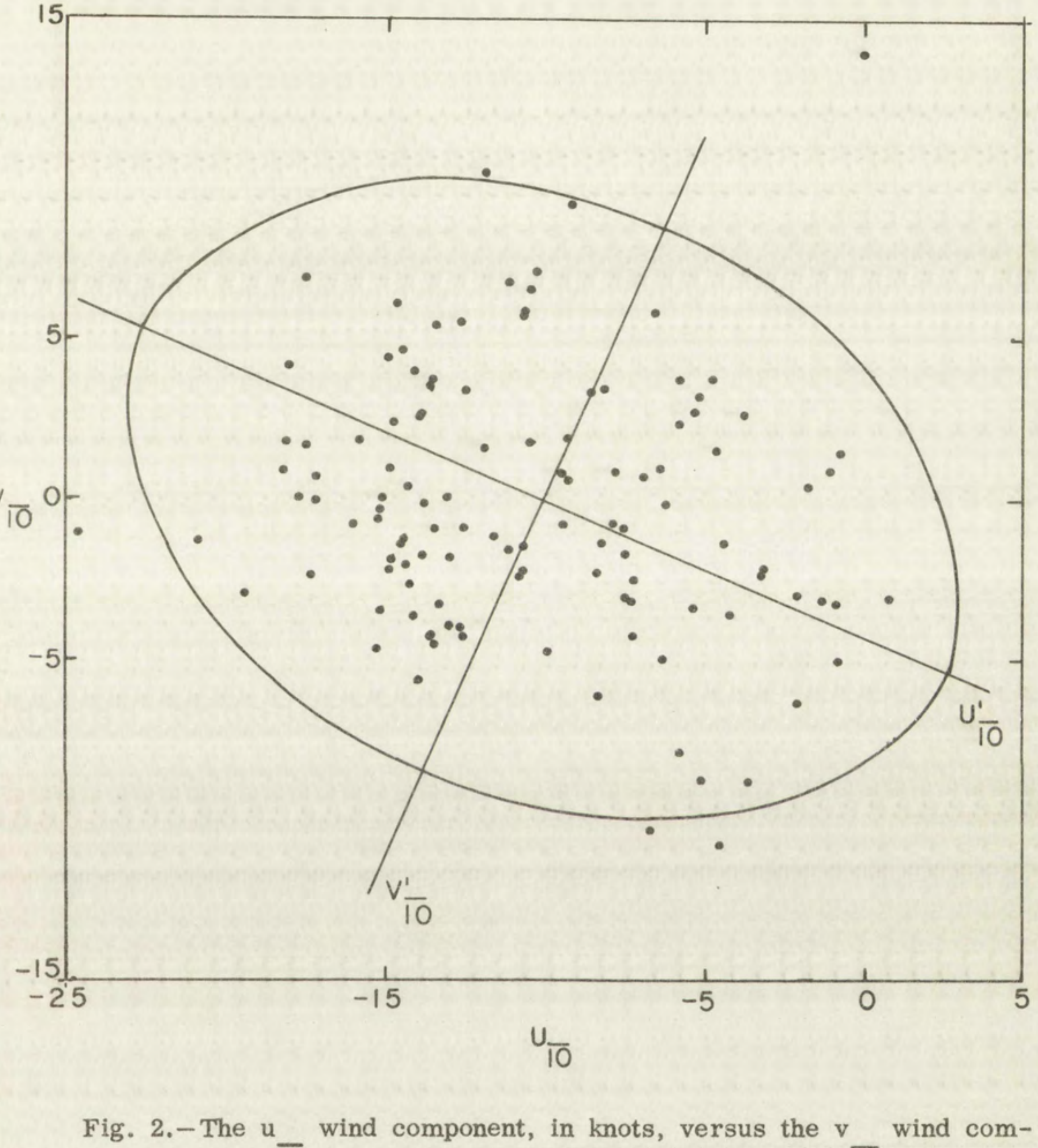

Fig. 2.—The $u_{\overline{10}}$ wind component, in knots, versus the $v_{\overline{10}}$ wind component, in knots, at Bikini during June and July 1956. Based on the assumption that the $u_{\overline{10}}$ and $v_{\overline{10}}$ have a bivariate normal distribution, the 95% confidence ellipse is also given.

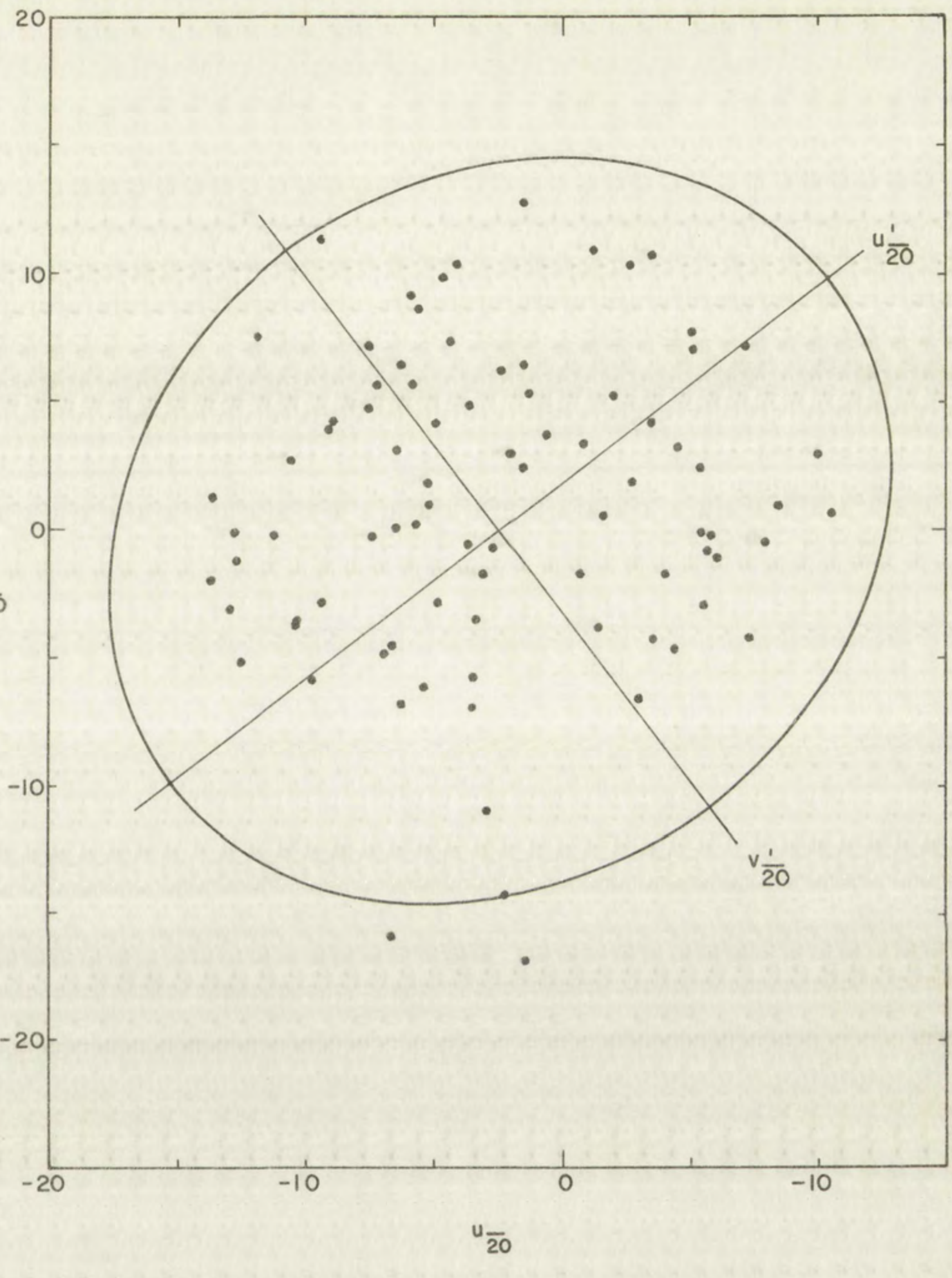

Fig. 3.–The $u_{\overline{20}}$ wind component, in knots, versus the $v_{\overline{20}}$ wind component, in knots, at Bikini during June and July 1956. Based on the assumption that the $u_{\overline{20}}$ and $v_{\overline{20}}$ have a bivariate normal distribution, the 95% confidence ellipse is also given.

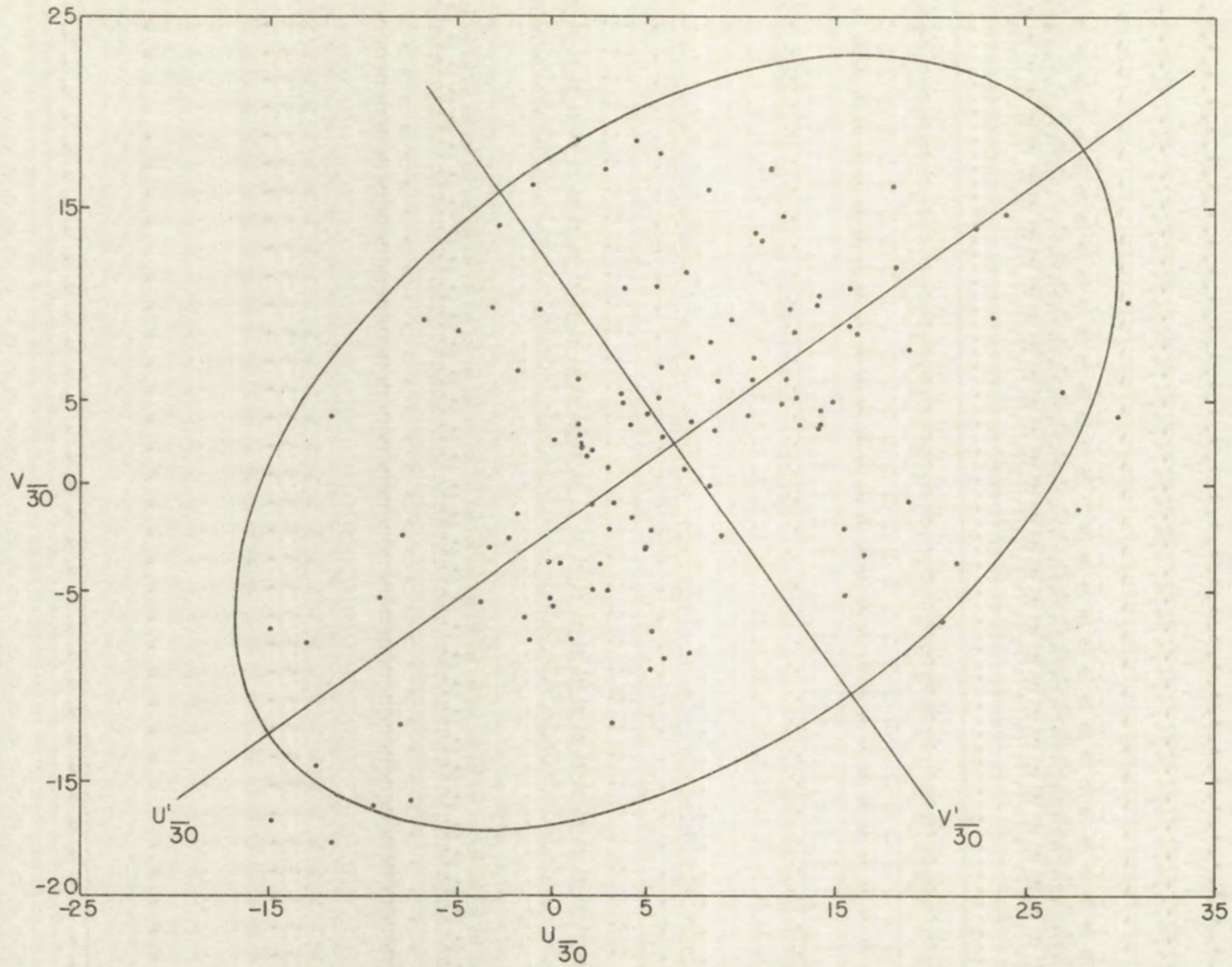

Fig. 4.—The $u_{\overline{30}}$ wind component, in knots, versus the $v_{\overline{30}}$ wind component, in knots, at Bikini during June and July 1956. Based on the assumption that the $u_{\overline{30}}$ and $v_{\overline{30}}$ have a bivariate normal distribution, the 95% confidence ellipse is also given.

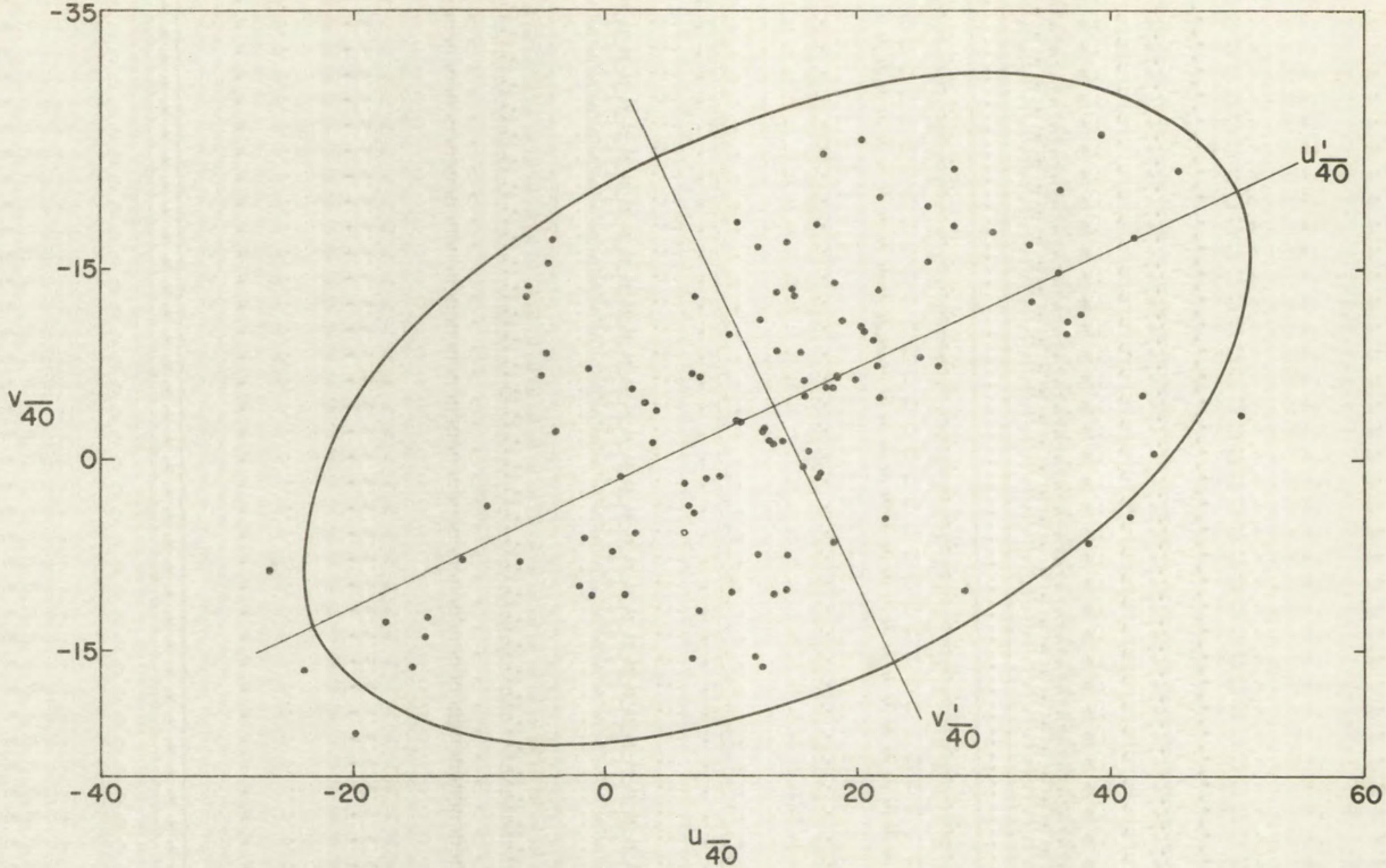

Fig. 5.—The $u_{\overline{40}}$ wind component, in knots, versus the $v_{\overline{40}}$ wind component, in knots, at Bikini during June and July 1956. Based on the assumption that the $u_{\overline{40}}$ and $v_{\overline{40}}$ have a bivariate normal distribution, the 95% confidence ellipse is also given.

3. Development of the Prediction Equations

Let us first denote the average u and v wind components at Bikini as the dependent variables for each level. The average u and v components at Bikini, Eniwetok, Kusaie, Kwajalein, and Majuro at the same level observed 12 hours prior to the dependent observation at Bikini will be the independent variables. N sets of data at each level were then drawn from the basic populations, always remembering that every variable must have been observed in order to fulfill the requirement of a set of data.

Assuming that (1) the average u and v components at each level are distributed as a bivariate normal with means

$$\mu_u = \sum_{g=1}^{q} b_{1g} z_g \text{ and } \mu_v = \sum_{g=1}^{q} b_{2g} z_g,$$

(2) the variances of the u and v components at Bikini are independent of the independent variables, and (3) the N sets of data are independent, we can apply the statistical model of Chapter I. Although it is virtually impossible to exhibit the validity of all of the above assumptions from the data, we can demonstrate through correlation analysis that there does exist a relationship between the dependent and independent variables. Even though these relationships may not be fully understood from the standpoint of theoretical meteorology, such relationships have been used by the subjective forecaster with success and certainly enough to justify our use of them. Figure 6 is a plot of the dependent variable Bikini $u_{\overline{10}}$ versus the independent variable Bikini $u_{\overline{10}}$ 12 hours prior. Figure 7 is a plot of the dependent variable Bikini $v_{\overline{10}}$ versus the independent variable Kwajalein $v_{\overline{10}}$ 12 hours prior. These figures help to illustrate some of the relationships or correlations that were found.

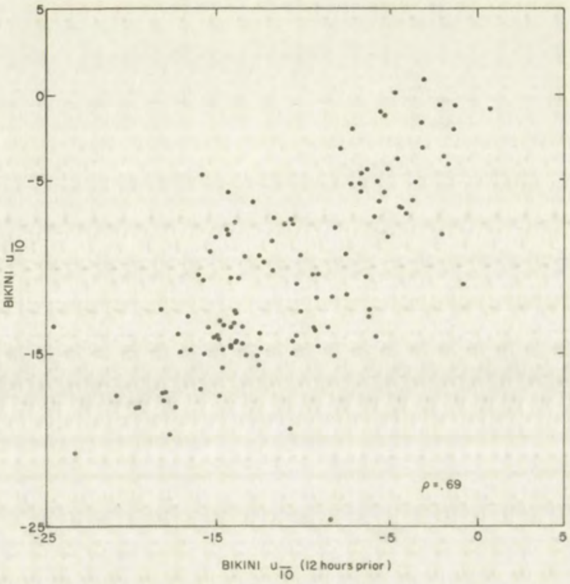

Fig. 6.—The $u_{\overline{10}}$ wind component, in knots, at Bikini versus the $u_{\overline{10}}$ wind component, in knots, 12 hours prior at Bikini. The correlation coefficient between these two variables is .69.

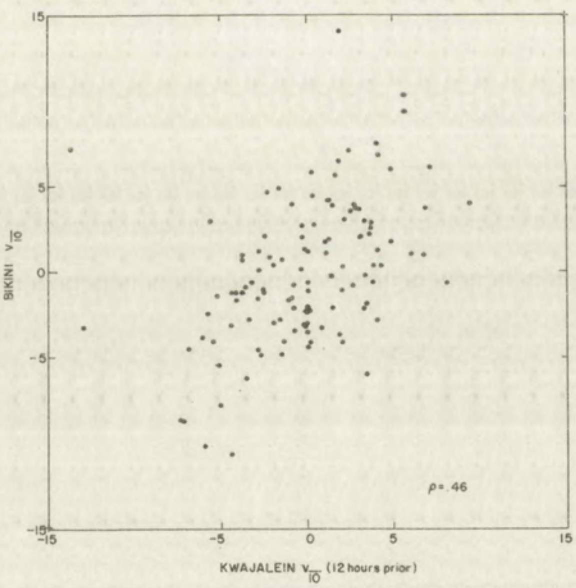

Fig. 7.—The $v_{\overline{10}}$ wind component, in knots, at Bikini versus the $v_{\overline{10}}$ wind component, in knots, 12 hours prior at Kwajalein. The correlation coefficient between these two variables is .46.

Although the theoretical development of our statistical model assures us a solution, if one exists, the actual solution for a particularly large problem is almost impossible without the aid of a high speed computer. Thus, a computer code was written specifically for this problem. A description of this code appears in the Appendix.

The first problem attempted was the prediction of the average wind for the 10,000 foot level. One hundred nineteen sets of data were drawn from the present populations. The mean and standard deviation of $u_{\overline{10}}$ and $v_{\overline{10}}$ components at Bikini were computed and are given as follows:

$$\bar{u}_{\overline{10}} = -10.56 \text{ knots} \qquad \text{S.D. of } u_{\overline{10}} = 5.19 \text{ knots}$$

$$\bar{v}_{\overline{10}} = -\ .04 \text{ knots} \qquad \text{S.D. of } v_{\overline{10}} = 4.14 \text{ knots.}$$

These moments are felt to be in agreement with the computed moments of the parent populations. If one knew no meteorology and predicted the $u_{\overline{10}}$ component to fall in the range of $-10.56 \pm 1.96\ (5.19)$ knots every time, we would expect him to be right about 95 percent of the time. Thus, if we are to develop a better forecasting technique, we must improve our estimates and reduce the error interval on the estimate significantly.

A slightly modified 12 hour persistence type forecast using regression analysis was tried. The $u_{\overline{10}}$ and $v_{\overline{10}}$ components at Bikini were the dependent variables and the $u_{\overline{10}}$ and $v_{\overline{10}}$ components 12 hours prior at Bikini were the independent variables. The analysis indicated that we had significantly reduced the variability* between the observed and calculated value for the

---

*The "F" test at the 95% level was used in the comparison of two variances.

$u_{\overline{10}}$ dependent variable, as compared to the standard deviation of the $u_{\overline{10}}$ dependent variable. Although the $v_{\overline{10}}$ component S. D. (FIT) was reduced, it was not reduced significantly. This measure of variability used throughout the paper will be referred to as the S. D. (FIT) and is defined as follows:

$$S.D.(FIT) = \frac{\sum_{i=1}^{N}\left(\chi_{obs.i} - \chi_{cal.i}\right)^2}{N-q}$$

where $\chi_{obs.i}$ is the ith observed dependent variable, $\chi_{cal.i}$ is the ith calculated dependent variable, N is the number of sets of data, and q is the number of independent variables.

In an attempt to further reduce the variability, more independent variables, namely the $u_{\overline{10}}$ and $v_{\overline{10}}$ components at Eniwetok, Kusaie, Kwajalein, and Majuro 12 hours prior were added. While the S. D. (FIT) for each dependent variable was reduced, it was not reduced significantly over the 12 hour persistence type forecast; however, the new S. D. (FIT) for the $v_{\overline{10}}$ component was significantly reduced as compared to the standard deviation of the $v_{\overline{10}}$ component at Bikini.

Computations of the multiple correlation coefficient, ρ, the S. D. (FIT), the regression or prediction equation, and the standard deviation of each coefficient in the regression equation at each dependent component for the above problems appear in tables 7 and 8. In order to facilitate reading these and later tables, let us define

x_1 = the predicted average u component at Bikini

x_2 = the predicted average v component at Bikini

z_1 = 1

z_2 = the average u component at Bikini 12 hours prior

z_3 = the average v component at Bikini 12 hours prior

z_4 = the average u component at Eniwetok 12 hours prior

z_5 = the average v component at Eniwetok 12 hours prior

z_6 = the average u component at Kusaie 12 hours prior

z_7 = the average v component at Kusaie 12 hours prior

z_8 = the average u component at Kwajalein 12 hours prior

z_9 = the average v component at Kwajalein 12 hours prior

z_{10} = the average u component at Majuro 12 hours prior

z_{11} = the average v component at Majuro 12 hours prior

b_{ij} = the coefficient of the j independent variable for the ith prediction equation.

The average wind prediction problems at 20,000 feet were next tried. Eighty-five sets of data were drawn from the parent populations. The means and standard deviations of the $u_{\overline{20}}$ and $v_{\overline{20}}$ components at Bikini of these 85 sets are given as follows:

$$\bar{u}_{\overline{20}} = -2.51 \text{ knots} \qquad \text{S.D. of } u_{\overline{20}} = 6.23 \text{ knots}$$

$$\bar{v}_{\overline{20}} = .91 \text{ knot} \qquad \text{S.D. of } v_{\overline{20}} = 6.03 \text{ knots.}$$

These are consistent with the computed moments of the parent population.

The modified 12 hour persistence type forecast using regression analysis was again attempted. The S.D. (FIT) for each dependent variable was significantly reduced as compared to the standard deviation of $u_{\overline{20}}$ and $v_{\overline{20}}$ components at Bikini. By adding more independent variables, namely, the $u_{\overline{20}}$ and $v_{\overline{20}}$ at Eniwetok, Kusaie, Kwajalein, and Majuro 12 hours prior, we were able to significantly reduce the S.D. (FIT) of the $v_{\overline{20}}$ component as

TABLE 8

10,000 FOOT, FORECAST EQUATIONS USING ALL "12 HOUR PRIOR" STATIONS

$u_{\overline{10}}$ Component

S. D. (FIT) = 3.47 knots $\rho = .77$

$x_1 = \sum_{j=1}^{11} b_{1j} z_j$ N = 119

$b_{11} = -1.7425$	S. D. $= 9.1100 \times 10^{-1}$
$b_{12} = 3.9364 \times 10^{-1}$	S. D. $= 1.0920 \times 10^{-1}$
$b_{13} = -5.1531 \times 10^{-2}$	S. D. $= 1.1326 \times 10^{-1}$
$b_{14} = 1.1846 \times 10^{-1}$	S. D. $= 9.9867 \times 10^{-2}$
$b_{15} = -2.6500 \times 10^{-1}$	S. D. $= 9.3090 \times 10^{-2}$
$b_{16} = 7.1310 \times 10^{-2}$	S. D. $= 8.6728 \times 10^{-2}$
$b_{17} = -1.1645 \times 10^{-2}$	S. D. $= 1.2002 \times 10^{-1}$
$b_{18} = 1.9527 \times 10^{-2}$	S. D. $= 7.3778 \times 10^{-2}$
$b_{19} = 3.0418 \times 10^{-1}$	S. D. $= 1.0283 \times 10^{-1}$
$b_{110} = 1.2336 \times 10^{-1}$	S. D. $= 8.5063 \times 10^{-2}$
$b_{111} = -2.0233 \times 10^{-1}$	S. D. $= 9.2820 \times 10^{-2}$

TABLE 8 - Continued

v_{10} Component

S. D. (FIT) = 3.47 knots	ρ = .60
$x_2 = \sum_{j=1}^{11} b_{2j} z_j$	N = 119
$b_{11} = -9.4209 \times 10^{-1}$	S. D. = 9.1101×10^{-1}
$b_{12} = 4.1120 \times 10^{-2}$	S. D. = 1.0920×10^{-1}
$b_{13} = 4.1517 \times 10^{-1}$	S. D. = 1.1326×10^{-1}
$b_{14} = -1.0019 \times 10^{-1}$	S. D. = 9.9868×10^{-2}
$b_{15} = -1.0843 \times 10^{-1}$	S. D. = 9.3091×10^{-2}
$b_{16} = 1.3072 \times 10^{-1}$	S. D. = 8.6729×10^{-2}
$b_{17} = -8.7602 \times 10^{-2}$	S. D. = 1.2003×10^{-1}
$b_{18} = -4.4151 \times 10^{-3}$	S. D. = 7.3778×10^{-2}
$b_{19} = 3.2567 \times 10^{-1}$	S. D. = 1.0283×10^{-1}
$b_{110} = -1.6592 \times 10^{-1}$	S. D. = 8.5063×10^{-2}
$b_{111} = -3.9273 \times 10^{-2}$	S. D. = 9.2821×10^{-2}

compared to the S.D. (FIT) of the $v_{\overline{20}}$ component for the 12 hour persistence type forecast. Under the same comparison, the $u_{\overline{20}}$ S.D. (FIT) was reduced, but not significantly.

The computations for the two above problems appear in tables 9 and 10.

The next problem attempted was the average wind forecast at the 30,000 foot level. A sample of 115 sets of data was drawn from the parent populations, and the means and standard deviations of the $u_{\overline{30}}$ and $v_{\overline{30}}$ components at Bikini are given as follows:

$\bar{u}_{\overline{30}} = 6.30$ knots   S.D. of $u_{\overline{30}} = 9.50$ knots

$\bar{v}_{\overline{30}} = 2.92$ knots   S.D. of $v_{\overline{30}} = 8.35$ knots.

These appear consistent with the computed moments of the parent populations at Bikini.

As in the previous problems, the modified 12 hour persistence type forecast was done, and again the S.D. (FIT) for the two dependent variables at Bikini indicated a significant reduction over their $u_{\overline{30}}$ and $v_{\overline{30}}$ component standard deviations. The further addition of $u_{\overline{30}}$ and $v_{\overline{30}}$ components at Eniwetok, Kusaie, Kwajalein, and Majuro 12 hours prior reduced the S.D. (FIT) of the $u_{\overline{30}}$ component significantly as compared to the S.D. (FIT) of the $u_{\overline{30}}$ for the 12 hour persistence type forecast. Although the $v_{\overline{30}}$ S.D. (FIT) was reduced, it was not reduced significantly. Tables 11 and 12 give the computations for the two above problems.

Finally, the 40,000 foot average wind forecast was tried. One hundred thirteen sets of data were drawn from the parent populations and the means

TABLE 9

20,000 FOOT, 12 HOUR PERSISTENCE TYPE
FORECAST EQUATIONS

u_{20} Component

S. D. (FIT) = 4.89 knots

$\rho = .63$

N = 85

$$x_1 = \sum_{j=1}^{3} b_{1j} z_j$$

$b_{11} = -5.2172 \times 10^{-1}$	S. D. $= 6.1042 \times 10^{-1}$
$b_{12} = 6.1844 \times 10^{-1}$	S. D. $= 8.4689 \times 10^{-2}$
$b_{13} = -8.2577 \times 10^{-3}$	S. D. $= 7.8540 \times 10^{-2}$

v_{20} Component

S. D. (FIT) = 4.42 knots

$\rho = .69$

N = 85

$$x_2 = \sum_{j=1}^{3} b_{2j} z_j$$

$b_{21} = 7.1233 \times 10^{-2}$	S. D. $= 5.5136 \times 10^{-1}$
$b_{22} = -2.3468 \times 10^{-2}$	S. D. $= 7.6496 \times 10^{-2}$
$b_{23} = 6.0687 \times 10^{-1}$	S. D. $= 7.0942 \times 10^{-2}$

TABLE 10

20,000 FOOT, FORECAST EQUATIONS USING ALL "12 HOUR PRIOR" STATIONS

$u_{\overline{20}}$ Component

S.D. (FIT) = 4.31 knots $\qquad \rho = .76$

$$x_1 = \sum_{j=1}^{11} b_{1j} z_j \qquad N = 85$$

$b_{11} = -3.2446$	S.D. = 1.3945
$b_{12} = 3.6676 \times 10^{-1}$	S.D. = 1.4909×10^{-1}
$b_{13} = 6.6509 \times 10^{-2}$	S.D. = 1.2285×10^{-1}
$b_{14} = 2.4797 \times 10^{-1}$	S.D. = 1.2282×10^{-1}
$b_{15} = -2.5272 \times 10^{-1}$	S.D. = 1.0169×10^{-1}
$b_{16} = -3.1026 \times 10^{-1}$	S.D. = 1.0904×10^{-1}
$b_{17} = 3.1559 \times 10^{-1}$	S.D. = 1.1566×10^{-1}
$b_{18} = 2.8995 \times 10^{-1}$	S.D. = 1.2004×10^{-1}
$b_{19} = 8.5643 \times 10^{-2}$	S.D. = 1.1808×10^{-1}
$b_{110} = -1.9732 \times 10^{-2}$	S.D. = 1.0149×10^{-1}
$b_{111} = 8.3422 \times 10^{-2}$	S.D. = 1.0774×10^{-1}

TABLE 10 - Continued

$v_{\overline{20}}$ Component

S.D.(FIT) = 3.23 knots $\rho = .86$

N = 85

$$x_2 = \sum_{j=1}^{11} b_{2j} z_j$$

$b_{21} = -2.1811$	S.D. = 1.0439
$b_{22} = 2.7534 \times 10^{-1}$	S.D. = 1.1161×10^{-1}
$b_{23} = 2.0585 \times 10^{-1}$	S.D. = 9.1967×10^{-2}
$b_{24} = 6.7171 \times 10^{-2}$	S.D. = 9.1949×10^{-2}
$b_{25} = -9.9632 \times 10^{-2}$	S.D. = 7.6126×10^{-2}
$b_{26} = 1.3927 \times 10^{-1}$	S.D. = 8.1633×10^{-2}
$b_{27} = -1.0731 \times 10^{-3}$	S.D. = 8.6587×10^{-2}
$b_{28} = -3.0740 \times 10^{-1}$	S.D. = 8.9861×10^{-2}
$b_{29} = 3.0347 \times 10^{-1}$	S.D. = 8.8397×10^{-2}
$b_{210} = -3.6770 \times 10^{-1}$	S.D. = 7.5974×10^{-2}
$b_{211} = 1.0668 \times 10^{-2}$	S.D. = 8.0660×10^{-2}

TABLE 11

30,000 FOOT, 12 HOUR PERSISTENCE TYPE FORECAST EQUATIONS

u_{30} Component

S. D. (FIT) = 6.23 knots $\qquad \rho = .76$

$$x_1 = \sum_{j=1}^{3} b_{1j} z_j \qquad N = 115$$

$b_{11} = 1.7420 \qquad$ S. D. $= 6.8876 \times 10^{-1}$

$b_{12} = 7.7808 \times 10^{-1} \qquad$ S. D. $= 7.0576 \times 10^{-2}$

$b_{13} = 1.6181 \times 10^{-2} \qquad$ S. D. $= 7.6448 \times 10^{-2}$

v_{30} Component

S. D. (FIT) = 5.38 knots $\qquad \rho = .77$

$$x_2 = \sum_{j=1}^{3} b_{2j} z_j \qquad N = 115$$

$b_{21} = 1.4746 \times 10^{-1} \qquad$ S. D. $= 5.9547 \times 10^{-1}$

$b_{22} = 1.2043 \times 10^{-1} \qquad$ S. D. $= 6.1017 \times 10^{-2}$

$b_{23} = 6.8876 \times 10^{-1} \qquad$ S. D. $= 6.6093 \times 10^{-2}$

TABLE 12

30,000 FOOT, FORECAST EQUATIONS USING ALL "12 HOUR PRIOR" STATIONS

u_{30} Component

S.D. (FIT) = 4.64 knots $\rho = .88$

$$x_1 = \sum_{j=1}^{11} b_{1j} z_j$$

N = 115

b_{11}	$= 1.5192$	$S.D. = 8.5471 \times 10^{-1}$
b_{12}	$= 5.3537 \times 10^{-1}$	$S.D. = 8.3553 \times 10^{-2}$
b_{13}	$= 1.8632 \times 10^{-1}$	$S.D. = 9.7656 \times 10^{-2}$
b_{14}	$= 2.5579 \times 10^{-1}$	$S.D. = 7.1538 \times 10^{-2}$
b_{15}	$= -2.7715 \times 10^{-1}$	$S.D. = 8.3009 \times 10^{-2}$
b_{16}	$= 9.9599 \times 10^{-2}$	$S.D. = 8.1281 \times 10^{-2}$
b_{17}	$= 1.2013 \times 10^{-1}$	$S.D. = 1.2009 \times 10^{-1}$
b_{18}	$= 1.3203 \times 10^{-1}$	$S.D. = 9.3407 \times 10^{-2}$
b_{19}	$= 1.1195 \times 10^{-2}$	$S.D. = 1.0607 \times 10^{-1}$
b_{110}	$= 6.6047 \times 10^{-2}$	$S.D. = 7.5467 \times 10^{-2}$
b_{111}	$= 1.4876 \times 10^{-1}$	$S.D. = 9.3594 \times 10^{-2}$

TABLE 12 - Continued

v_{30} Component

S. D. (FIT) = 4.73 knots

ρ = .84

N = 115

$$x_2 = \sum_{j=1}^{11} b_{2j} z_j$$

$b_{21} = -1.4794$	S. D. $= 8.7166 \times 10^{-1}$
$b_{22} = 9.3749 \times 10^{-2}$	S. D. $= 8.5210 \times 10^{-2}$
$b_{23} = 5.6219 \times 10^{-1}$	S. D. $= 9.9593 \times 10^{-2}$
$b_{24} = 4.7428 \times 10^{-2}$	S. D. $= 7.2956 \times 10^{-2}$
$b_{25} = 1.7503 \times 10^{-2}$	S. D. $= 8.4656 \times 10^{-2}$
$b_{26} = -3.6054 \times 10^{-2}$	S. D. $= 8.2893 \times 10^{-2}$
$b_{27} = 5.3219 \times 10^{-4}$	S. D. $= 1.2247 \times 10^{-1}$
$b_{28} = 2.6310 \times 10^{-2}$	S. D. $= 9.5260 \times 10^{-2}$
$b_{29} = 3.4078 \times 10^{-2}$	S. D. $= 1.0817 \times 10^{-1}$
$b_{210} = -1.1495 \times 10^{-1}$	S. D. $= 7.6964 \times 10^{-2}$
$b_{211} = 3.7193 \times 10^{-1}$	S. D. $= 9.5451 \times 10^{-2}$

and standard deviations of the $u_{\overline{40}}$ and $v_{\overline{40}}$ components at Bikini are given as follows:

$$\bar{u}_{\overline{40}} = 13.49 \text{ knots} \qquad \text{S.D. of } u_{\overline{40}} = 15.51 \text{ knots}$$

$$\bar{v}_{\overline{40}} = 3.61 \text{ knots} \qquad \text{S.D. of } v_{\overline{40}} = 10.98 \text{ knots.}$$

These appear consistent with the computed moments of the parent populations at Bikini.

The modified 12 hour persistence type forecast was again applied, and a significant reduction in the S.D. (FIT) for each dependent component was found as compared to its standard deviation. In a further attempt to reduce the S.D. (FIT) of the 12 hour persistence type forecast, more independent variables were added, specifically, the $u_{\overline{40}}$ and $v_{\overline{40}}$ components at Eniwetok, Kusaie, Kwajalein, and Majuro 12 hours prior. The analysis indicated that, although the S.D. (FIT) in both components was reduced, they were not reduced significantly. The results appear in tables 13 and 14.

4. Evaluation of the Prediction Equations

Many experimenters feel that testing the prediction equations on a set of independent data is necessary for evaluating the equations. Actually, the evaluation is more of a test on whether the sample of data from which the equations were derived is representative of the entire population. At any rate, a randomly chosen subset of the "20,000 foot" sets of data was deliberately removed for this test before the regression analysis was performed. This explains why the 20,000 foot level problem had the smallest number of sets of data.

Using the regression equations of table 10, predictions and their

TABLE 13

40,000 FOOT, 12 HOUR PERSISTENCE TYPE FORECAST EQUATIONS

u_{40} Component

S. D. (FIT) = 9.90 knots

$\rho = .77$

N = 113

$$x_1 = \sum_{j=1}^{3} b_{1j} z_j$$

$b_{11} = 2.3988$	S. D. = 1.2813
$b_{12} = 7.8249 \times 10^{-1}$	S. D. = 7.1417×10^{-2}
$b_{13} = 8.4587 \times 10^{-2}$	S. D. = 9.4264×10^{-2}

v_{40} Component

S. D. (FIT) = 7.39 knots

$\rho = .75$

N = 113

$$x_2 = \sum_{j=1}^{3} b_{2j} z_j$$

$b_{21} = -9.0276 \times 10^{-1}$	S. D. = 9.5641×10^{-1}
$b_{22} = 1.6718 \times 10^{-1}$	S. D. = 5.3307×10^{-2}
$b_{23} = 6.0398 \times 10^{-1}$	S. D. = 7.0360×10^{-2}

TABLE 14

40,000 FOOT, FORECAST EQUATIONS USING ALL "12 HOUR PRIOR" STATIONS

$u_{\overline{40}}$ Component

S. D. (FIT) = 9.42 knots $\qquad \rho = .81$

$$x_1 = \sum_{j=1}^{11} b_{1j} z_j \qquad N = 113$$

$b_{11} = 1.9660$	S. D. = 1.5310
$b_{12} = 4.6287 \times 10^{-1}$	S. D. = 1.3344×10^{-1}
$b_{13} = -2.1688 \times 10^{-2}$	S. D. = 1.6843×10^{-1}
$b_{14} = 2.8590 \times 10^{-1}$	S. D. = 1.1139×10^{-1}
$b_{15} = 6.0727 \times 10^{-2}$	S. D. = 1.2915×10^{-1}
$b_{16} = 5.1113 \times 10^{-2}$	S. D. = 1.1971×10^{-1}
$b_{17} = 2.2241 \times 10^{-1}$	S. D. = 1.7486×10^{-1}
$b_{18} = 7.7209 \times 10^{-2}$	S. D. = 1.2524×10^{-1}
$b_{19} = -1.6910 \times 10^{-1}$	S. D. = 1.8465×10^{-1}
$b_{110} = 1.2711 \times 10^{-1}$	S. D. = 1.1345×10^{-1}
$b_{111} = 4.1771 \times 10^{-1}$	S. D. = 1.5964×10^{-1}

TABLE 14 - Continued

$v_{\overline{40}}$ Component

S.D. (FIT) = 6.80 knots $\rho = .81$

$x_2 = \sum_{j=1}^{11} b_{2j} z_j$ N = 113

$b_{21} = -2.5314$	S.D. = 1.1061
$b_{22} = 3.8018 \times 10^{-2}$	S.D. = 9.6406×10^{-2}
$b_{23} = 3.0230 \times 10^{-1}$	S.D. = 1.2169×10^{-1}
$b_{24} = 1.5665 \times 10^{-1}$	S.D. = 8.0473×10^{-2}
$b_{25} = 1.1212 \times 10^{-1}$	S.D. = 9.3304×10^{-2}
$b_{26} = -2.2187 \times 10^{-2}$	S.D. = 8.6486×10^{-2}
$b_{27} = 1.4182 \times 10^{-1}$	S.D. = 1.2633×10^{-1}
$b_{28} = 5.1534 \times 10^{-2}$	S.D. = 9.0477×10^{-2}
$b_{29} = 1.8103 \times 10^{-1}$	S.D. = 1.3341×10^{-1}
$b_{210} = 5.9850 \times 10^{-2}$	S.D. = 8.1964×10^{-2}
$b_{211} = 3.9973 \times 10^{-1}$	S.D. = 1.1534×10^{-1}

confidence ellipses based on the ten sets of independent data were made. The predicted vector, denoted by P, the 95% confidence ellipse for the prediction, and the observed vector, denoted by O, were plotted for the ten sets of data and appear in figures 8, 9, and 10. The computations for figure 8 appear as an example in the Appendix. Since only one of the ten observed vectors fell outside its predicted confidence ellipse, it appears that the test on independent data is consistent with the developed statistical theory.

5. Conclusions

Although the derived prediction equations were able to significantly reduce the error in the predicted wind as compared to the error in predicting the average wind every time, the real value of this type of forecasting should be compared with other existing techniques. Unfortunately, this type of comparison is dependent upon the requirements of the forecast and this cannot be generalized. For example, one forecasting technique may be very good most of the time but poor in predicting extreme conditions, whereas another technique might work in just the opposite manner. Which one is the better?

Nevertheless, it is felt that, because of the simplicity in making a prediction from the derived regression equations, the technique should be employed as a first approximation and modified according to the forecaster's subjective views. The combination of the two techniques could prove most successful.

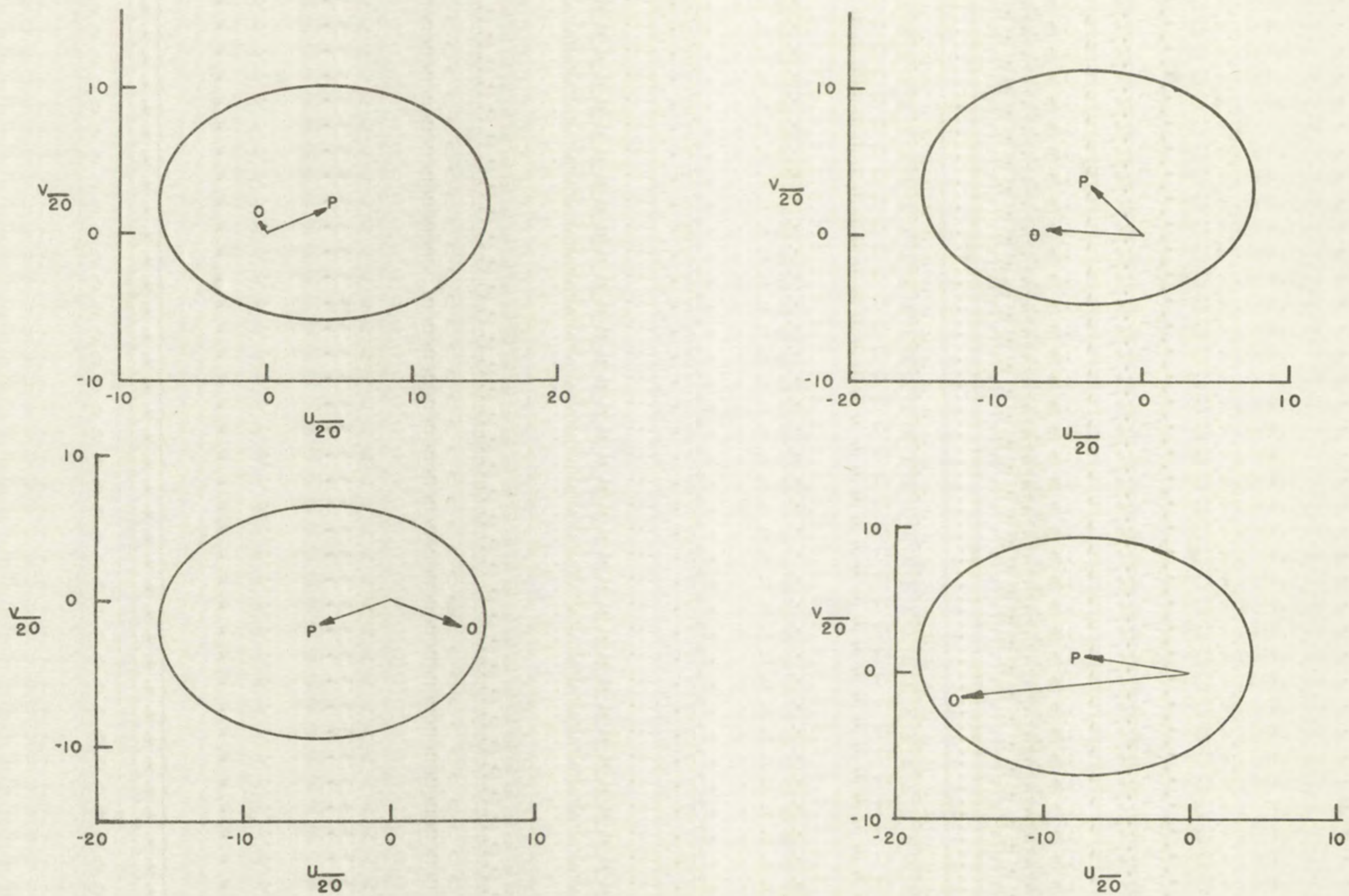

Fig. 8.—Four comparisons between a predicted vector, P, and an observed vector, O, in knots, at 20,000 feet at Bikini. The 95% confidence ellipse associated with each prediction is also given.

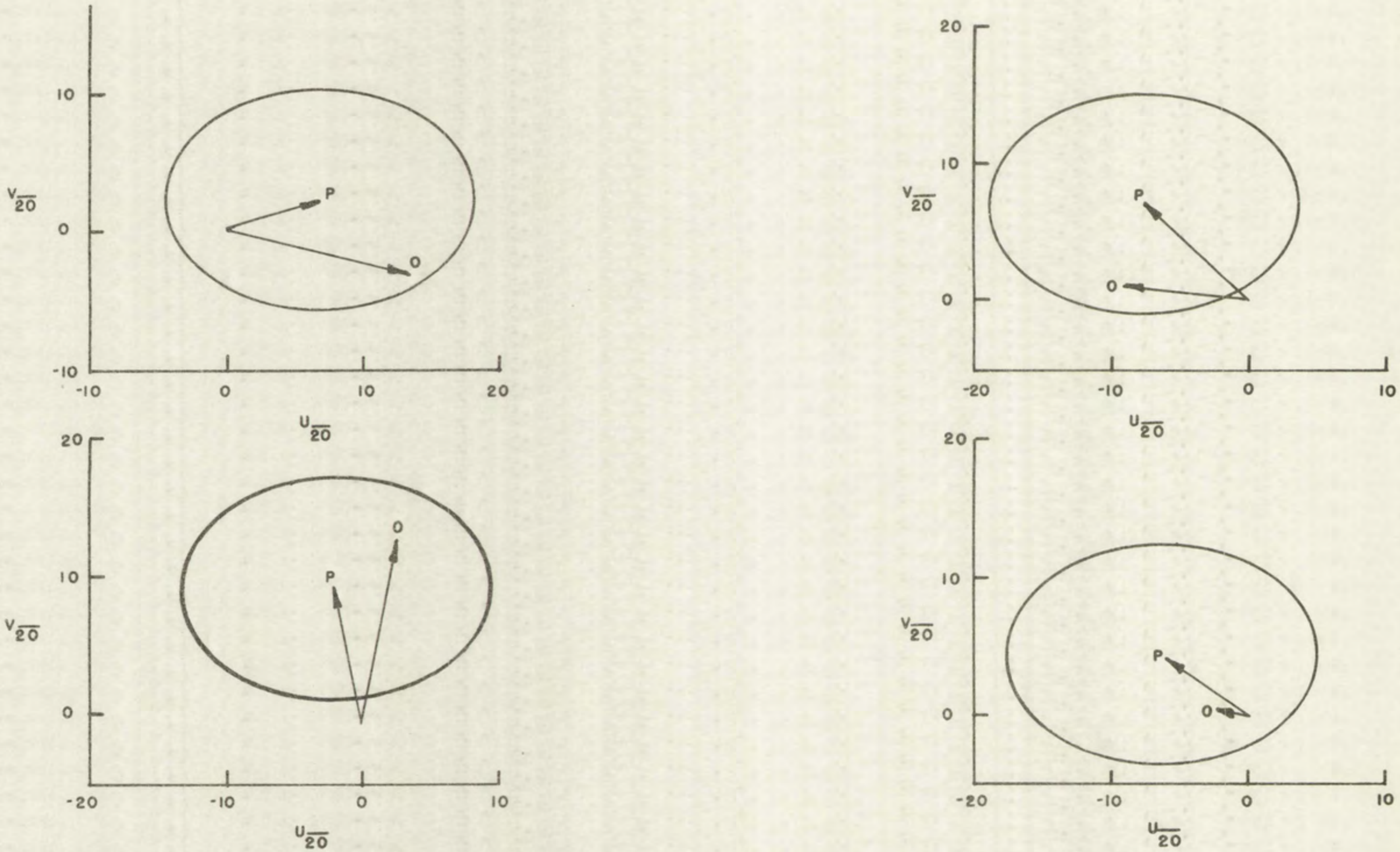

Fig. 9.—Four comparisons between a predicted vector, P, and an observed vector, O, in knots, at 20,000 feet at Bikini. The 95% confidence ellipse associated with each prediction is also given.

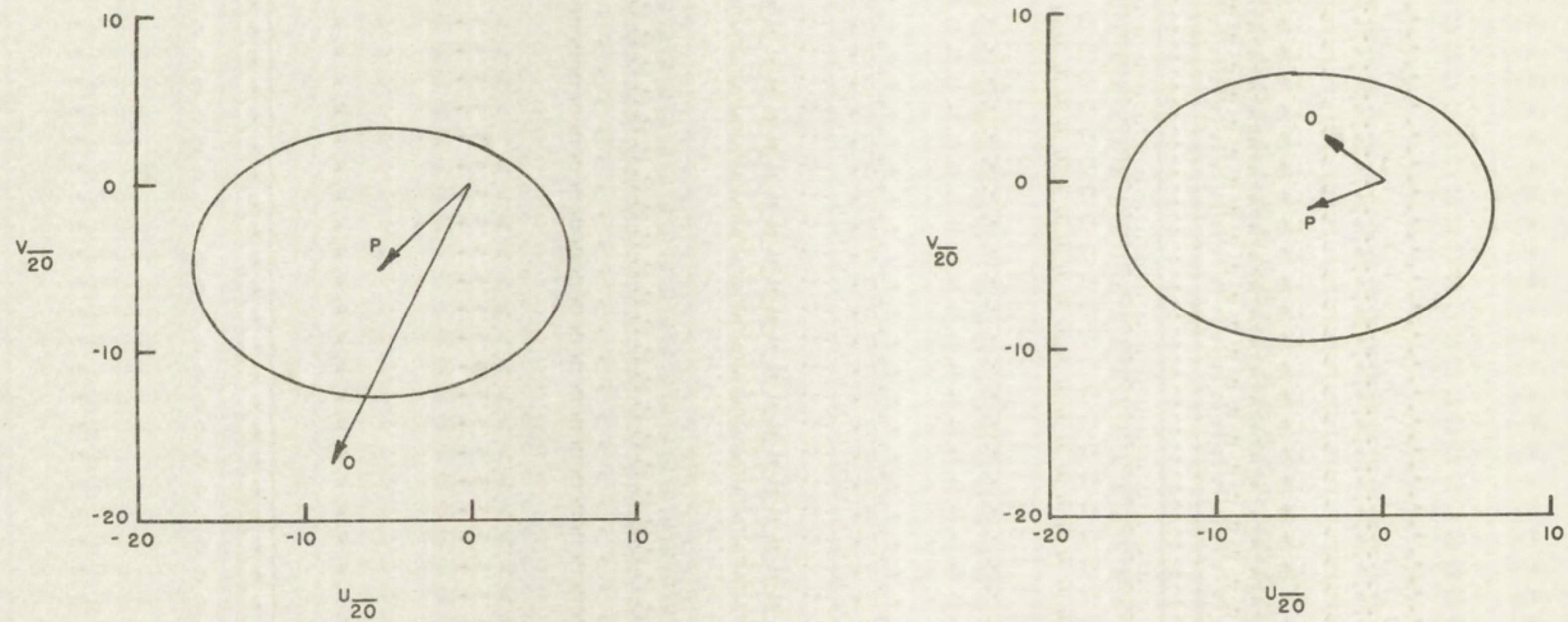

Fig. 10.—Two comparisons between a predicted vector, P, and an observed vector, O, in knots, at 20,000 feet at Bikini. The 95% confidence ellipse associated with each prediction is also given.

APPENDIX

CORRELATION AND MULTIPLE REGRESSION ANALYSIS CODE FOR THE VECTOR PREDICTION PROBLEM

Purpose: The code performs a correlation and multiple regression analysis for the vector prediction problem. Given N sets of data, two dependent variables, and q independent variables, we compute the following:

1. Mean and standard deviation for each independent and dependent variable.
2. Simple correlation coefficients between the independent variables.
3. Simple correlation coefficients between the dependent and independent variables.
4. Estimates of the parameters, $\sigma^2_{x_1}$, $\sigma^2_{x_2}$, $\sigma_{x_1 x_2}$, b_{1i}, b_{2i} ($i = 1, \ldots, q$), of the bivariate normal distribution given below:

$$f\left(x_1, x_2\right) = \frac{1}{2\pi\sigma_{x_1}\sigma_{x_2}\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2\left(1-\rho^2\right)}\left[\left(\frac{x_1 - \sum_{i=1}^{q} b_{1i} z_i}{\sigma_{x_1}}\right)^2 - \frac{2\rho\left(x_1 - \sum_{i=1}^{q} b_{1i} z_i\right)\left(x_2 - \sum_{i=1}^{q} b_{2i} z_i\right)}{\sigma_{x_1}\sigma_{x_2}} + \left(\frac{x_2 - \sum_{i=1}^{q} b_{2i} z_i}{\sigma_{x_2}}\right)^2\right]\right\}$$

where

$$\rho = \frac{\sigma_{x_1 x_2}}{\sigma_{x_1} \sigma_{x_2}}.$$

5. Estimates of the errors on the b coefficients.
6. The least squares matrix and inverse.
7. For each dependent variable:
   a. Standard deviation of the fit,
   b. Multiple correlation coefficient,
   c. Each observed dependent variable, its computed value from the fitted regression equation, and their difference.
8. From a set of independent variables a prediction of each dependent variable from its fitted regression equation. Errors on each prediction and the correlation coefficient between the two predictions are given, and from this information an equation of the rotated confidence ellipse is computed.

Restrictions: This code is written in Fortran II for a "32K" IBM 704 computer; however, it may be used on a smaller core 704 by recompiling the source deck with smaller dimension statements. The present code will handle up to 25 independent variables and 200 sets of data.

Input: Identification, control, and prediction information is loaded by cards, while the data is loaded on a tape "off line" and the tape is used as the input medium.

Output: The computed information is written on a tape which is listed "off line." A sample listing of the output follows.

```
PREDICTION PROBLEM FOR HØT AIR      DICK VOGEL
              STATISTICS OF THE INPUT DATA

MEANS AND STANDARD DEVIATIONS OF THE DEPENDENT VARIABLE
MEAN X1=-2.5069556E 00   S.D. X1= 6.2325963E 00          MEAN X2= 9.0857763E-01   S.D. X2= 6.0250787E 00

MEANS AND STANDARD DEVIATIONS OF THE INDEPENDENT VARIABLES
MEAN Z 1= 1.0000000E 00      S.D. Z 1= 0.
MEAN Z 2=-3.1932870E 00      S.D. Z 2= 6.3796901E 00
MEAN Z 3= 1.2562818E 00      S.D. Z 3= 6.8791367E 00
MEAN Z 4=-4.6562348E 00      S.D. Z 4= 6.6459018E 00
MEAN Z 5= 7.3313655E-01      S.D. Z 5= 6.3241128E 00
MEAN Z 6=-9.4201918E 00      S.D. Z 6= 5.1207366E 00
MEAN Z 7= 5.1273931E 00      S.D. Z 7= 4.5060142E 00
MEAN Z 8=-6.1635457E 00      S.D. Z 8= 5.5589718E 00
MEAN Z 9= 1.2129570E 00      S.D. Z 9= 5.8189769E 00
MEAN Z10=-8.5206701E 00      S.D. Z10= 6.9288037E 00
MEAN Z11= 1.6707526E 00      S.D. Z11= 5.5490654E 00
```

SIMPLE CORRELATION COEFFICIENTS BETWEEN INDEPENDENT VARIABLES

1.000	0.157	0.767	0.042	0.316	-0.032	0.546	0.112	0.503	0.019
0.157	1.000	0.347	0.670	0.235	0.354	-0.280	0.627	-0.373	0.422
0.767	0.347	1.000	0.200	0.335	0.095	0.257	0.157	0.256	0.071
0.042	0.670	0.200	1.000	0.150	0.313	-0.257	0.402	-0.254	0.289
0.316	0.235	0.335	0.150	1.000	0.139	0.044	0.293	0.215	0.356
-0.032	0.354	0.095	0.313	0.139	1.000	-0.177	0.336	-0.193	0.305
0.546	-0.280	0.257	-0.257	0.044	-0.177	1.000	-0.171	0.594	-0.162
0.112	0.627	0.157	0.402	0.293	0.336	-0.171	1.000	-0.291	0.551
0.503	-0.373	0.256	-0.254	0.215	-0.193	0.594	-0.291	1.000	-0.128
0.019	0.422	0.071	0.289	0.356	0.305	-0.162	0.551	-0.128	1.000

SIMPLE CORRELATION COEFFICIENTS DEPENDENT VS INDEPENDENT

DEPENDENT VARIABLE X1

0.632	0.090	0.542	-0.113	0.020	0.159	0.487	0.112	0.294	0.041

DEPENDENT VARIABLE X2

0.084	0.689	0.263	0.399	0.261	0.297	-0.442	0.654	-0.548	0.393

```
PREDICTION PROBLEM FOR HOT AIR       DICK VOGEL

ESTIMATES OF THE PARAMETERS OF THE BIVARIATE NORMAL

SIGMA PARAMETERS

SIGMA SQUARED (1,1)= 1.8579298E 01        SIGMA SQUARED (2,2)= 1.0412527E 01
                SIGMA SQUARED (1,2)= 1.9072783E 00


B PARAMETERS

B1 1=-3.2445779E 00   S.D.= 1.3944671E 00        B2 1=-2.1811235E 00   S.D.= 1.0439308E 00
B1 2= 3.6676491E-01   S.D.= 1.4908762E-01        B2 2= 2.7533653E-01   S.D.= 1.1161049E-01
B1 3= 6.6509491E-02   S.D.= 1.2284671E-01        B2 3= 2.0585150E-01   S.D.= 9.1965929E-02
B1 4= 2.4796962E-01   S.D.= 1.2282403E-01        B2 4= 6.7171388E-02   S.D.= 9.1948951E-02
B1 5=-2.5272081E-01   S.D.= 1.0168860E-01        B2 5=-9.9632280E-02   S.D.= 7.6126467E-02
B1 6=-3.1026377E-01   S.D.= 1.0904373E-01        B2 6= 1.3927288E-01   S.D.= 8.1632692E-02
B1 7= 3.1558573E-01   S.D.= 1.1566211E-01        B2 7=-1.0731285E-03   S.D.= 8.6587372E-02
B1 8= 2.8995170E-01   S.D.= 1.2003521E-01        B2 8=-3.0739874E-01   S.D.= 8.9861175E-02
B1 9= 8.5643107E-02   S.D.= 1.1807981E-01        B2 9= 3.0346793E-01   S.D.= 8.8397311E-02
B110=-1.9732119E-02   S.D.= 1.0148570E-01        B210=-3.6769741E-01   S.D.= 7.5974579E-02
B111= 8.3422368E-02   S.D.= 1.0774453E-01        B211= 1.0667875E-02   S.D.= 8.0660084E-02
```

PREDICTION PROBLEM FOR HOT AIR   DICK VOGEL

A MATRIX

8.500E 01	-2.714E 02	1.068E 02	-3.958E 02	6.232E 01	-8.007E 02
4.358E 02	-5.239E 02	1.031E 02	-7.243E 02	1.420E 02	
-2.714E 02	4.286E 03	2.383E 02	3.995E 03	-5.760E 01	3.424E 03
-1.470E 03	3.299E 03	2.088E 01	4.180E 03	-3.958E 02	
1.068E 02	2.383E 02	4.109E 03	8.369E 02	2.528E 03	-3.104E 02
1.469E 03	-1.558E 03	2.237E 03	-2.403E 03	1.532E 03	
-3.958E 02	3.995E 03	8.369E 02	5.553E 03	4.169E 02	4.686E 03
-1.790E 03	3.238E 03	3.050E 01	4.363E 03	-4.418E 02	
6.232E 01	-5.760E 01	2.528E 03	4.169E 02	3.405E 03	-1.803E 02
1.070E 03	-1.144E 03	1.317E 03	-1.467E 03	9.552E 02	
-8.007E 02	3.424E 03	-3.104E 02	4.686E 03	-1.803E 02	9.746E 03
-3.836E 03	5.040E 03	-2.371E 02	7.463E 03	-4.880E 02	
4.358E 02	-1.470E 03	1.469E 03	-1.790E 03	1.070E 03	-3.836E 03
3.940E 03	-3.058E 03	1.269E 03	-4.218E 03	1.368E 03	
-5.239E 02	3.299E 03	-1.558E 03	3.238E 03	-1.144E 03	5.040E 03
-3.058E 03	5.825E 03	-1.099E 03	6.385E 03	-1.296E 03	
1.031E 02	2.088E 01	2.237E 03	3.050E 01	1.317E 03	-2.371E 02
1.269E 03	-1.099E 03	2.969E 03	-1.864E 03	1.666E 03	
-7.243E 02	4.180E 03	-2.403E 03	4.363E 03	-1.467E 03	7.463E 03
-4.218E 03	6.385E 03	-1.864E 03	1.020E 04	-1.622E 03	
1.420E 02	-3.958E 02	1.532E 03	-4.418E 02	9.552E 02	-4.880E 02
1.368E 03	-1.296E 03	1.666E 03	-1.622E 03	2.824E 03	

A INVERSE MATRIX

1.047E-01	-4.423E-03	8.213E-04	2.073E-03	4.129E-04	5.233E-03
-3.427E-03	3.094E-03	2.772E-04	1.220E-03	-1.623E-03	
-4.423E-03	1.196E-03	-9.470E-05	-6.819E-04	3.804E-05	-3.900E-05
8.369E-05	-3.860E-04	-1.567E-04	-2.454E-04	4.898E-05	
8.213E-04	-9.470E-05	8.123E-04	-1.777E-04	-3.478E-04	-1.025E-05
-4.120E-05	8.819E-05	-2.727E-04	1.874E-04	-7.836E-05	
2.073E-03	-6.819E-04	-1.777E-04	8.120E-04	-7.790E-06	-9.188E-05
-6.503E-05	1.084E-04	1.233E-04	3.777E-05	4.058E-05	
4.129E-04	3.804E-05	-3.478E-04	-7.790E-06	5.566E-04	6.027E-06
-6.667E-05	4.589E-05	1.852E-05	-4.162E-05	3.356E-06	
5.233E-03	-3.900E-05	-1.025E-05	-9.188E-05	6.027E-06	6.400E-04
-1.434E-05	9.029E-05	-9.239E-05	-1.461E-04	-1.499E-04	
-3.427E-03	8.369E-05	-4.120E-05	-6.503E-05	-6.667E-05	-1.434E-05
7.200E-04	2.879E-06	-7.647E-05	1.069E-05	-7.990E-05	
3.094E-03	-3.860E-04	8.819E-05	1.084E-04	4.589E-05	9.029E-05
2.879E-06	7.755E-04	-3.408E-05	-1.939E-04	2.261E-05	
2.772E-04	-1.567E-04	-2.727E-04	1.233E-04	1.852E-05	-9.239E-05
-7.647E-05	-3.408E-05	7.505E-04	1.258E-04	-2.399E-04	
1.220E-03	-2.454E-04	1.874E-04	3.777E-05	-4.162E-05	-1.461E-04
1.069E-05	-1.939E-04	1.258E-04	5.543E-04	-5.264E-05	
-1.623E-03	4.898E-05	-7.836E-05	4.058E-05	3.356E-06	-1.499E-04
-7.990E-05	2.261E-05	-2.399E-04	-5.264E-05	6.248E-04	

PREDICTION PROBLEM FOR HOT AIR   DICK VOGEL

DEPENDENT VARIABLE X1

STANDARD DEVIATION OF THE FIT = 4.3103709E 00   MULTIPLE CORRELATION COEFFICIENT = 7.6069000E-01

	X(1) OBSERVED	X(1) CALCULATED	DELTA X
1	9.2547998E-01	-4.1050373E 00	5.0305173E 00
2	5.8074000E 00	9.9698787E-01	4.8104122E 00
3	1.0141000E 01	2.7732969E 00	7.3677030E 00
4	8.5855999E 00	3.9809619E 00	4.6046380E 00
5	6.2370999E 00	1.4253970E 00	4.8117030E 00
6	1.0607000E 01	1.2586005E 00	9.3483995E 00
7	1.6995000E 00	-2.9484905E 00	4.6479905E 00
8	4.4998000E 00	3.4650174E 00	1.0347825E 00
9	3.1218000E 00	5.4936032E 00	-2.3718032E 00
10	1.5058000E 00	4.0653169E 00	-2.5595169E 00
11	-5.3945999E-01	-2.8454676E-01	-2.5491323E-01
12	-3.5383000E 00	-6.4366564E 00	2.8983564E 00
13	-1.1168000E 01	-1.0618408E 01	-5.4959166E-01
14	-1.3635000E 01	-1.3254598E 01	-3.8040197E-01
15	-1.2803000E 01	-8.0878524E 00	-4.7151476E 00
16	-1.2687000E 01	-6.1499234E 00	-6.5370765E 00
17	-6.4094000E 00	-1.0273318E 01	3.8639179E 00
18	-7.3023000E 00	-9.2603285E 00	1.9580286E 00
19	-3.2481000E 00	-7.3969763E 00	4.1488763E 00
20	-6.1847000E 00	-5.6038513E 00	-5.8084863E-01
21	-1.2473000E 01	-3.5440008E 00	-8.9289992E 00
22	-6.5175000E 00	-4.7012035E 00	-1.8162965E 00
23	-2.9453000E 00	-5.4366809E 00	2.4913809E 00
24	-1.4504000E 00	-2.9984419E 00	1.5480419E 00
25	5.9087999E 00	-1.0374374E 00	6.9462373E 00
26	5.5430000E 00	2.0959199E 00	3.4470801E 00
27	5.5473000E 00	4.5936731E 00	9.5362693E-01
28	3.6105000E 00	6.8875349E 00	-3.2770349E 00
29	2.1308000E 00	6.5201189E 00	-4.3893189E 00
30	3.5979000E 00	1.3296330E 00	2.2682670E 00
31	7.2905999E 00	3.9789514E 00	3.3116486E 00
32	2.8822000E 00	5.3263776E-02	2.8289362E 00
33	-4.6397000E 00	2.1416565E 00	-6.7813565E 00
34	-1.0337000E 01	-7.3693627E-02	-1.0263306E 01
35	3.6907000E 00	3.2782019E 00	4.1249806E-01
36	5.6258000E 00	3.6092403E 00	2.0165597E 00
37	4.1770000E 00	2.5799902E 00	1.5970098E 00
38	4.6522000E 00	3.8757389E 00	7.7646112E-01
39	8.0804000E 00	3.7989625E 00	4.2814375E 00
40	7.3173000E-01	4.2507674E 00	-3.5190374E 00
41	-4.2381000E 00	-1.3961973E 00	-2.8419027E 00
42	-1.4418000E 00	3.3319109E 00	-4.7737109E 00
43	1.3423000E 00	2.1834792E 00	-8.4117922E-01
44	2.7475000E 00	1.4827012E 00	1.2647987E 00
45	5.2523000E 00	2.1150462E 00	3.1372538E 00
46	-6.1931000E 00	4.2805242E 00	-1.0473624E 01
47	-5.4426000E 00	-5.5592492E 00	1.1664927E-01
48	-8.8155999E 00	-6.3106339E 00	-2.5049660E 00
49	-1.1780000E 01	-8.4407865E 00	-3.3392135E 00
50	-9.2952999E 00	-1.1708595E 01	2.4132950E 00
51	-1.2908000E 01	-6.9957215E 00	-5.9122785E 00
52	-1.0401000E 01	-7.4871274E 00	-2.9138725E 00
53	-3.4902000E 00	-8.1162837E 00	4.6260837E 00
54	-1.4346000E 00	-8.4190496E 00	6.9844496E 00

55	-5.6871000E 00	-5.6453788E 00	-4.1721165E-02
56	-4.5209000E 00	-4.1536232E 00	-3.6727679E-01
57	-7.4227999E 00	-4.6677387E 00	-2.7550613E 00
58	-4.8552999E 00	-4.3429472E 00	-5.1235276E-01
59	-1.8776000E 00	-1.0572230E 00	-8.2037697E-01
60	-2.5399000E 00	-2.5531733E 00	1.3273299E-02
61	-7.4363000E 00	-1.3965175E 00	-6.0397825E 00
62	-5.7673000E 00	-7.4820930E 00	1.7147931E 00
63	-6.3600000E 00	-7.3942248E 00	1.0342248E 00
64	-2.2403000E 00	-3.1499943E 00	9.0969428E-01
65	-1.3638000E 01	-2.4298388E 00	-1.1208161E 01
66	7.4214000E 00	1.7575997E 00	5.6638003E 00
67	-3.9734000E 00	-5.6682473E 00	1.6948473E 00
68	5.1902000E 00	2.2038170E 00	2.9863830E 00
69	-3.3355000E 00	-6.3675324E 00	3.0320324E 00
70	-3.3935000E 00	-2.5412114E 00	-8.5228857E-01
71	-9.6593000E 00	-6.2861578E 00	-3.3731421E 00
72	-2.8176000E 00	-3.8289172E 00	1.0113173E 00
73	-2.0627000E 00	-4.6103756E 00	2.5476757E 00
74	-1.3493000E 00	-3.0972072E 00	1.7479072E 00
75	-6.5801000E 00	-3.7720700E 00	-2.8080300E 00
76	-5.3029000E 00	-4.2915907E 00	-1.0113093E 00
77	-6.8685000E 00	-5.1023499E 00	-1.7661501E 00
78	-5.0653999E 00	-5.2582464E 00	1.9284648E-01
79	-5.1066999E 00	-3.8073839E 00	-1.2993160E 00
80	-7.1300000E 00	-5.1275889E 00	-2.0024111E 00
81	-8.9749999E 00	-5.1841120E 00	-3.7908880E 00
82	-1.2146000E 00	-5.6583854E 00	4.4437854E 00
83	-9.3032999E 00	-7.0362504E 00	-2.2670495E 00
84	-5.5906000E 00	-4.9924380E 00	-5.9816194E-01
85	-1.0252000E 01	-9.3512478E 00	-9.0075219E-01

PREDICTION PROBLEM FOR HOT AIR DICK VOGEL

DEPENDENT VARIABLE X2

STANDARD DEVIATION OF THE FIT = 3.2268448E 00 MULTIPLE CORRELATION COEFFICIENT = 8.6447258E-01

	X(2) OBSERVED	X(2) CALCULATED	DELTA X
1	3.3296000E 00	3.3955411E 00	-6.5941125E-02
2	-7.9804999E-01	3.8099424E 00	-4.6079924E 00
3	3.0135000E 00	2.4623448E 00	5.5115518E-01
4	9.2612000E-01	1.0072597E 00	-8.1139728E-02
5	-1.0998000E 00	-8.1449643E-01	-2.8530355E-01
6	6.6821000E-01	-4.5658756E-01	1.1247976E 00
7	5.3239999E-01	8.1182549E-01	-2.7942550E-01
8	-4.6641000E 00	-2.3731408E 00	-2.2909591E 00
9	-6.5972000E 00	-4.4438077E 00	-2.1533923E 00
10	-3.7808000E 00	-9.9087168E 00	6.1279168E 00
11	3.7671000E 00	3.4895639E 00	2.7753603E-01
12	-5.7979000E-01	-3.2513842E-01	-2.5465158E-01
13	-1.8811000E-01	-7.9277117E-01	6.0466118E-01
14	1.2115000E 00	2.5304835E 00	-1.3189835E 00
15	-7.6115999E-02	4.3437855E-01	-5.1049455E-01
16	-1.1667000E 00	1.2650222E-01	-1.2932022E 00
17	3.2652000E-02	9.4840114E-01	-9.1574914E-01
18	-2.3268000E-01	-8.2414285E-01	5.9146286E-01
19	-3.5107000E 00	-4.6202046E-01	-3.0486795E 00
20	-6.8640000E 00	-1.5508639E 00	-5.3131361E 00
21	-5.1751000E 00	-5.7752305E 00	6.0013056E-01
22	-4.5321000E 00	-4.9224183E 00	3.9031833E-01
23	-1.6814000E 00	-3.1452402E 00	1.4638403E 00
24	2.4846000E 00	-1.5001764E 00	3.9847764E 00
25	-1.7365000E-01	-1.0965986E 00	9.2294856E-01
26	2.4735000E 00	1.9308990E 00	5.4260102E-01
27	-5.9003000E-02	7.6004021E-01	-8.1904321E-01
28	4.2158999E 00	2.4931803E 00	1.7227197E 00
29	5.2420000E 00	1.8084352E 00	3.4335648E 00
30	1.0760000E 01	7.2457202E 00	3.5142798E 00
31	7.1740000E 00	5.0015777E 00	2.1724223E 00
32	1.8139000E 00	2.7234151E 00	-9.0951508E-01
33	-2.8694000E 00	2.6156044E 00	-5.4850044E 00
34	-3.7622000E 00	-1.3265330E-01	-3.6295467E 00
35	-4.2126999E 00	-2.7066939E 00	-1.5060061E 00
36	-2.8862000E 00	-1.5965388E 00	-1.2896612E 00
37	-1.6717000E 00	-2.1845156E 00	5.1281558E-01
38	3.2821000E 00	-1.6645934E 00	4.9466934E 00
39	-4.5602999E-01	8.4594605E-01	-1.3019761E 00
40	-1.6874000E 00	2.9454141E 00	-4.6328140E 00
41	7.3877000E 00	6.2493912E 00	1.1383088E 00
42	1.2815000E 01	9.4722123E 00	3.3427876E 00
43	1.0896000E 01	1.4175527E 01	-3.2795268E 00
44	1.0330000E 01	1.1249490E 01	-9.1948974E-01
45	7.0545000E 00	8.4421571E 00	-1.3876572E 00
46	6.4082000E 00	8.6271737E 00	-2.2189737E 00
47	8.6183000E 00	6.9306500E 00	1.6876500E 00
48	4.1815000E 00	4.9588846E 00	-7.7738464E-01
49	7.5468000E 00	8.3321680E 00	-7.8536803E-01
50	1.1338000E 01	1.2185662E 01	-8.4766197E-01
51	-3.1782000E 00	3.8645162E 00	-7.0427162E 00
52	2.7407000E 00	5.6829798E 00	-2.9422798E 00
53	4.9005000E 00	6.6334720E 00	-1.7329720E 00
54	6.4848000E 00	4.2804595E 00	2.2043405E 00

55	5.6662000E 00	5.0860754E 00	5.8012456E-01
56	9.8825999E 00	4.1375743E 00	5.7450256E 00
57	7.1352000E 00	5.2447761E 00	1.8904239E 00
58	4.1196000E 00	2.8269861E 00	1.2926139E 00
59	2.9495000E 00	5.0264142E 00	-2.0769142E 00
60	-7.1554000E-01	2.2017525E 00	-2.9172924E 00
61	4.7510999E 00	2.2978681E 00	2.4532319E 00
62	9.1789999E 00	1.0706500E 00	8.1083500E 00
63	3.0796000E 00	1.2393320E 00	1.8402680E 00
64	6.2566000E 00	2.2743589E 00	3.9822411E 00
65	-2.0521000E 00	4.3402095E-01	-2.4861209E 00
66	-4.1901000E 00	-4.3136817E 00	1.2358177E-01
67	1.0373000E 01	8.7676157E 00	1.6053842E 00
68	7.7163000E 00	6.2363074E 00	1.4799926E 00
69	-5.8191000E 00	1.2132467E-02	-5.8312324E 00
70	-6.9579000E 00	-5.5987499E 00	-1.3591501E 00
71	-5.8796000E 00	-5.5379120E 00	-3.4168798E-01
72	-1.1030000E 01	-7.6476254E 00	-3.3823746E 00
73	-1.4365000E 01	-1.0541766E 01	-3.8232343E 00
74	-1.6955000E 01	-1.0855498E 01	-6.0995018E 00
75	-1.5961000E 01	-1.0257183E 01	-5.7038167E 00
76	-6.1208000E 00	-1.0583787E 01	4.4629869E 00
77	-4.8477000E 00	-8.4421418E 00	3.5944418E 00
78	1.0000000E 00	-4.4248050E 00	5.4248050E 00
79	1.7621000E 00	-5.2203884E-01	2.2841388E 00
80	5.6451000E 00	3.7801278E 00	1.8649722E 00
81	3.8714000E 00	1.7194430E 00	2.1519570E 00
82	5.3195000E 00	-1.4593883E 00	6.7788883E 00
83	-2.9121000E 00	-3.1355634E 00	2.2346342E-01
84	2.1069000E-01	6.3325663E-01	-4.2256663E-01
85	-3.6084000E 00	-4.2343010E 00	6.2590104E-01

```
 PREDICTION PROBLEM FOR HOT AIR       DICK VOGEL
PREDICTIONS  Y1=B(1,1)+B(1,2)*Z2+...     Y2=B(2,1)+B(2,2)*Z2+...      AND THE CONFIDENCE ELLIPSE


INPUT INDEPENDENT VARIABLES

Z 2= 1.0607000E 01
Z 3= 6.6821000E-01
Z 4= 7.3620999E 00
Z 5= 4.8777000E-01
Z 6=-8.4512800E 00
Z 7= 3.4809800E-01
Z 8=-3.3222900E 00
Z 9=-3.1367100E 00
Z10=-4.6277500E 00
Z11=-2.1111200E 00



        PREDICTIONS

Y1= 3.8076874E 00   SD= 4.5440514E 00        Y2= 1.8939076E 00   SD= 3.4017834E 00

        CORRELATION BETWEEN PREDICTED VALUES = 1.3712649E-01


        95 PERCENT CONFIDENCE ELLIPSE

X**2/2.1119034E+01 +Y**2/1.1101500E+01 =5.991   ANGLE OF ROTATION 12.52 DEGREES
```

```
PREDICTION PROBLEM FOR HOT AIR      DICK VOGEL
PREDICTIONS  Y1=B(1,1)+B(1,2)*Z2+...      Y2=B(2,1)+B(2,2)*Z2+...      AND THE CONFIDENCE ELLIPSE


INPUT INDEPENDENT VARIABLES

Z 2=-2.0627000E 00
Z 3=-5.8378000E-01
Z 4=-2.8088000E 00
Z 5= 4.0710000E 00
Z 6=-9.3618399E 00
Z 7= 5.7263200E 00
Z 8=-9.0331600E 00
Z 9= 2.0558500E 00
Z10=-1.3604300E 01
Z11=-4.1313800E 00


        PREDICTIONS

Y1=-3.5727881E 00   SD= 4.4600933E 00        Y2= 3.5853577E 00   SD= 3.3389303E 00

        CORRELATION BETWEEN PREDICTED VALUES = 1.3712649E-01

        95 PERCENT CONFIDENCE ELLIPSE

X**2/2.0345833E+01 +Y**2/1.0695056E+01 =5.991   ANGLE OF ROTATION 12.52 DEGREES
```

```
 PREDICTION PROBLEM FOR HOT AIR      DICK VOGEL
PREDICTIONS  Y1=B(1,1)+B(1,2)*Z2+...     Y2=B(2,1)+B(2,2)*Z2+...     AND THE CONFIDENCE ELLIPSE


INPUT INDEPENDENT VARIABLES


Z 2=-1.3595000E 00
Z 3=-4.4119000E 00
Z 4=-2.9346000E 00
Z 5= 5.6188000E 00
Z 6=-1.4812600E 01
Z 7= 2.1416600E 00
Z 8=-8.7942300E 00
Z 9=-6.2320499E 00
Z10=-1.1477600E 01
Z11=-1.1529000E 01


        PREDICTIONS

Y1=-4.7315500E 00   SD= 4.6245941E 00        Y2=-1.3764647E 00   SD= 3.4620797E 00

        CORRELATION BETWEEN PREDICTED VALUES = 1.3712649E-01

        95 PERCENT CONFIDENCE ELLIPSE

X**2/2.1874333E+01 +Y**2/11.1498534E+01=5.991   ANGLE OF ROTATION 12.52 DEGREES
```

```
 PREDICTION PROBLEM FOR HOT AIR      DICK VOGEL
PREDICTIONS  Y1=B(1,1)+B(1,2)*Z2+...      Y2=B(2,1)+B(2,2)*Z2+...      AND THE CONFIDENCE ELLIPSE


INPUT INDEPENDENT VARIABLES

Z 2=-3.5383000E 00
Z 3=-5.7979000E-01
Z 4=-1.2544000E 01
Z 5= 5.0874000E 00
Z 6=-1.0067400E 01
Z 7= 4.8909200E 00
Z 8=-8.9802400E 00
Z 9=-9.4380399E 00
Z10=-1.9927100E 01
Z11= 3.2491900E 00


        PREDICTIONS

Y1=-7.0579122E 00   SD= 4.8898370E 00        Y2= 1.2266482E 00   SD= 3.6606466E 00

        CORRELATION BETWEEN PREDICTED VALUES = 1.3712649E-01


        95 PERCENT CONFIDENCE ELLIPSE

X**2/2.4455489E+01 +Y**2/11.2855352E+01=5.991   ANGLE OF ROTATION 12.52 DEGREES
```

LIST OF REFERENCES

1. A. Hald, Statistical Theory with Engineering Application, John Wiley and Sons, Inc., New York, 1952.

2. A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill Book Company, New York, 1950.

3. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., New York, 1958.

4. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

5. H. Hotelling, "Generalization of Students' Ratio," Ann. Math. Statist. 2 360-378 (1931).

6. W. A. Baum, and P. D. Thompson, "Long Range Weather Forecasting in the Soviet Union," Bull. Am. Meteorol. Soc. 40, 394-409 (1959).

7. "Numerical Weather Predictions," Report NAVAER 50-1P-541, Chief of Naval Operations, Washington, D. C., June 1956.

8. R. Robert Rapp, "The Effect of Variability and Instrumental Error of Measurements in the Free Atmosphere," Meteorological Papers 2, No. 1, New York University College of Engineering, New York (June 1952).

9. C. E. Palmer, F. R. Miller, and O. W. Stopinski, WT-40, Institute of Geophysics, University of California, Los Angeles, Scientific Director's Report of Atomic Weapon Tests at Eniwetok, 1951, Annex 4.5, "The Precipitation and Formation Movement of Clouds in the Central Pacific" (September 1951).

10. W. G. Cochran, "The χ^2 Test of Goodness of Fit," Ann. Math. Statist. 23, 315-345 (1952).

11. H. B. Mann and A. Wald, "On the Choice of the Number of Class Intervals in the Application of the Chi Square Test," Ann. Math. Statist. 13, 306-317 (1942).

12. R. A. Fisher, Statistical Methods for Research Workers, Hafner Publishing Company, New York, 1954.