A COMPARISON OF THE CLASSICAL AND INVERSE METHODS OF CALIBRATION IN REGRESSION

by

Marlin Amos Thomas

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Chairman, Dr. R. H. Myers

Dr. Boyd Harshbarger
Department of Statistics

Dr. R. G. Krutchkoff
Department of Statistics

Dr. C. Y. Kramer
Department of Statistics

Dr. L. T. McFadden
Department of Mathematics

May, 1969

Blacksburg, Virginia

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS

The author is deeply indebted to Dr. Raymond Myers for suggesting the area of research in this dissertation and for his guidance throughout the course of this effort. He is also indebted to Dr. Richard Krutchkoff for reopening this area of research in his 1967 article in Technometrics and for co-reading this manuscript. In addition he is indebted to Dr. Boyd Harshbarger, Head of the Statistics Department and Statistical Laboratory, for his guidance and encouragement.

The author expresses his thanks to Dr. R. B. Howe for developing the sequence of transformations which led to the density of d in Chapter II and to Dr. M. P. Jarnagin for his very useful suggestions in establishing the existence of the moments of d in Chapter II. Financial support for this research was provided by the U.S. Naval Weapons Laboratory, Dahlgren, Va., and the author expresses his appreciation to management there for underwriting this effort.

Finally, the author would like to express his deepest appreciation to Mrs. Lewis Whitescarver for the final typing of this manuscript and above all, to his wife, Carol, for typing the first draft and for her patient understanding during the course of this research.

CHAPTER I

INTRODUCTION

The linear calibration problem, frequently referred to as inverse regression or the discrimination problem has recently received the attention of several authors. The problem can be stated as follows: Let x represent the controlled variable and y the measured variable, and let the relation between x and y be given by

$$y = \alpha + \beta x + \varepsilon \; . \tag{1.1}$$

Consider N values of x_i, $i = 1,\ldots,N$, say known weights and the corresponding N values of y_i, $i = 1,\ldots,N$, say the readings on a scale to be calibrated. Then the model can be written

$$y_i = \alpha + \beta x_i + \varepsilon_i, \; i = 1,\ldots,N \tag{1.2}$$

where ε_i is the $i^{\underline{th}}$ reading error. If one assumes the errors to be independently and identically distributed with mean zero and variance σ^2 then a and b, the least squares estimators for α and β provide us with the least squares estimator for y for a given value of x,

$$\hat{y} = a + bx \tag{1.3}$$

where

$$a = \hat{\alpha} = \bar{y} - b\bar{x}$$

$$b = \hat{\beta} = S_{xy}/S_{xx} \; .$$

Here $\bar{x}$ and $\bar{y}$ are averages of the N values of x_i and y_i respectively and

$$S_{xy} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{yy} = \sum_{i=1}^{N} (y_i - \bar{y})^2 .$$

However, in the future, one is not interested in estimating y, the reading on the scale, for a given value of x but is rather interested in estimating x, the unknown weight of an object, for a given reading of y on the scale. Inversion of equation (1.3) provides us with

$$\hat{x}_C = (y - a)/b = \bar{x} + (y - \bar{y})/b . \tag{1.4}$$

This has recently been referred to as the Classical estimator and Ott (1966) and more recently Ott and Myers (1968) have considered it in some detail. In particular, they have determined optimal experimental designs using the criterion suggested by Box and Draper (1959), that is, the minimization of

$$\left(\int_R dx \right) J_C = \int_R E(\hat{x}_C - x)^2 \, dx \tag{1.5}$$

where R is the region of interest of x. While the $E(\hat{x}_C - x)^2$ does not exist since the moments of 1/b do not exist, by truncating the density of b, they obtained finite moments and hence a finite mean squared error; their results indicate that when the true relation is in fact first order as given by equation (1.2), then J_C can be minimized for symmetrical designs over the scaled region [-1, 1] by maximizing

$$S_{xx} = \sum_{i=1}^{N} (x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2$$

for symmetrical designs. S_{xx} is maximized by taking N/2 observations at x = -1 and N/2 observations at x = +1 when N is even and by taking (N-1)/2 observations at x = -1, (N-1)/2 observations at x = +1, and one observation at x = 0 when N is odd.

Consider rewriting equation (1.2) so that it becomes

$$x_i = \gamma + \delta\, y_i + \varepsilon_i', \; i = 1,\ldots,N \; . \qquad (1.6)$$

The least squares estimators for γ and δ are

$$\hat{\delta} = d = S_{xy}/S_{yy}$$

$$\hat{\gamma} = c = \bar{x} - d\bar{y}$$

so the estimator for x becomes

$$\hat{x}_I = c + dy = \bar{x} + d(y - \bar{y}) \; . \qquad (1.7)$$

The estimator given by (1.7) has recently been referred to as the Inverse estimator. In general, estimates given by (1.4) and (1.7) will be different, and the problem confronting the statistician is that of deciding which is the better of the two. Eisenhart (1939) rejected the Inverse estimator by the following philosophical argument:

> "It does not seem to be generally realized that the fitting should be done in terms of the deviations which actually represent "error". Thus when the research worker selects the X values in advance, and holds x to these values without error, and then

> observes the corresponding y values, the errors are in the y values, so that even if he is interested in using observed values of Y to estimate X, he should nevertheless fit $Y = a + bX$ and then use the inverse of this relation to estimate X, i.e. $X = (\hat{Y} - a)/b$, with the best available estimate of Y substituted for $\hat{Y}$."

He continued:

> "Briefly stated, when the values of x have been selected by the research worked and the corresponding y values observed, the line obtained by minimizing $\Sigma(x - \hat{Y})^2$ (he means $\Sigma(x - \hat{X})^2$) is meaningless, and (4) ((1.3) in this report)) is accordingly the only correct estimate of the postulated linear relationship between X and Y, wherefore, if it is desired to reason from Y to X this must be done by means of $X = (\hat{Y} - a)/b$, namely (4) solved for X."

Hence, in 1939 the problem appeared to be solved, and from 1939 until 1967 the literature on inverse regression dealt primarily with the Classical estimator. For example, Bennett and Franklin (1954) briefly discuss the Classical estimator and develop an approximating expression for $\mathrm{Var}(\hat{x}_C)$. Brownlee (1960) discusses inverse regression in brief under the heading, "The Use of the Regression Line in Reverse" and develops an expression for a confidence interval for x using the Classical estimator. Mood and Graybill (1963) show that if the errors are distributed independently and identically $N(0, \sigma^2)$, then the maximum likelihood estimator for x is given by the Classical estimator. Other works concerning the Classical estimator include Mandel and Linning (1957), Mandel (1958), Williams (1959), Linning and Mandel (1964), and the aforementioned research by Ott (1966) and Ott and Myers (1968).

In 1967 Krutchkoff reopened the problem of deciding between the two by comparing the average squared error of the Classical estimator with that of the Inverse by Monte Carlo sampling. While a mathematical proof was not given, his empirical results indicate that the Inverse estimator is better than the truncated Classical estimator from a mean squared error point of view for all $x \in R$.

Krutchkoff's work stimulated the research effort in two succeeding reports. McClelland (1967), using Monte Carlo procedures, showed through extensive tabulation that if one calculates an Inverse confidence interval by substituting the Inverse estimator, $\hat{x}_I$, for the Classical estimator, $\hat{x}_C$, in the Classical confidence interval, then the interval will have a higher confidence than the Classical interval of the same length. Yarbrough (1968), attempting a theoretical approach to the problem, did find an upper bound for the Inverse mean squared error, $E(\hat{x}_I - x)^2$. Then by using an approximation for $E(\hat{x}_C - x)^2$, he used extensive tabulation to show that for values of x near the mid-point of the region of interest (he used $R = [0, 1]$), the Inverse estimator tended to yield smaller mean squared errors than the Classical, but near the ends of R the Classical seemed superior. Yarbrough concludes his report with the following paragraph:

> "The whole region of Method B (Inverse estimation) has so far remained outside the realm of scholarly pursuit. It is to be hoped that this area will soon be as thoroughly explored as Method A (Classical estimation)."

It is the intent of this thesis to explore Inverse estimation and compare it with Classical estimation. To begin with, if Krutchkoff's empirical results that the Inverse estimator is better than the truncated Classical from a mean squared error point of view for all $x \in R$ be true, then the Inverse estimator should be better than the Classical using any criterion based on the mean squared error, and in particular, it should provide a lower average mean squared error than the minimum J_C for the Classical estimator obtained by Ott and Myers. However, in order to compare the minimum J_C with the minimum J_I, the minimum average mean squared error for the Inverse estimator, one must first explore the experimental design problem for the Inverse estimator. The problem of optimal experimental designs when the true model is first order as given by equation (1.2) is considered first in Chapter II. In Chapter III, the minimum J's of the two estimators are compared. The results show conclusively that the Inverse estimator is better than the truncated Classical estimator from an average mean squared error standpoint. Chapter IV is devoted to obtaining optimal experimental designs for the Inverse estimator while assuming a first order model when the true model is second order and to showing the value of protection of designing the experiment for a possible model misclassification. In Chapter V the minimum J's of the two estimators are compared in the presence of a second order model, and the results show conclusively that unless the model is nearly linear with a large ratio of the first order regression coefficient to the model standard deviation and one has nearly perfect a priori information concerning the model parameters the Inverse estimator is better than the truncated Classical from an average mean squared error standpoint. Chapter VI is devoted to a comparison of optimal designs for forward regression with those of inverse regression and concluding remarks are made in Chapter VII.

CHAPTER II

LINEAR APPROXIMATION, LINEAR MODEL - OPTIMAL DESIGNS

In order to make a fair comparison between the two estimators, it will be necessary to use the same criterion for optimizing designs for the Inverse estimator as was used by Ott and Myers for optimizing designs for the truncated Classical estimator, that is, the criterion of minimum integrated or average mean squared error. It should be emphasized that this criterion is a reasonable one for the mean squared error takes into account both the variance and bias of the estimator. Also by averaging the mean squared error over the region of interest R, one is not restricted to considering the bias and variance at a single point but can rather examine the quality of $\hat{x}$ as an estimator in the entire region R. Hence, designs are sought which minimize

$$J_{IL} = \int_R E(\hat{x}_{IL} - x)^2 \, dx \Big/ \int_R dx \tag{2.1}$$

where the subscript IL on J and $\hat{x}$ indicate the Inverse estimator for x is under consideration and that the true model is linear as given in equation (1.2). E is, of course, the usual expectation operator. Without loss of generality, the range R on x will be taken as [-1, 1], the same range used by Ott and Myers.

(2.1) Derivation of J_{IL} in Terms of the Moments of d

Distributional assumptions about the ε_i will eventually be needed, so in conjunction with the earlier assumption that the ε_i are independently and identically distributed with mean zero and variance σ^2, it will also be assumed that they are normally distributed. The first step now is to find an expression for $E(\hat{x}_{IL} - x)^2$, the $MSE(\hat{x}_{IL})$ (Mean Squared Error of $\hat{x}_{IL}$). Since

$$E(\hat{x}_{IL} - x)^2 = E(\hat{x}_{IL} - E\hat{x}_{IL})^2 + (E\hat{x}_{IL} - x)^2$$

$$= Var(\hat{x}_{IL}) + (Bias\ \hat{x}_{IL})^2 , \qquad (2.2)$$

the key terms needed will be the first two moments of $\hat{x}_{IL}$. Recalling from equation (1.7) that

$$\hat{x}_{IL} = c + dy = \bar{x} + d(y - \bar{y})$$

with $d = S_{xy}/S_{yy}$ and $c = \bar{x} - d\bar{y}$,

$$E(\hat{x}_{IL}) = \bar{x} + E[d(y - \bar{y})] . \qquad (2.3)$$

Similarly,

$$E(\hat{x}_{IL}^2) = \bar{x}^2 + 2\bar{x}E[d(y - \bar{y})] + E[d^2(y - \bar{y})^2] . \qquad (2.4)$$

It is clear that d is independent of y since y is a single future observation independent of the N observations used in forming S_{xy} and S_{yy}. It will now be shown that d is also independent of $\bar{y}$. Since d is a function of S_{xy} and S_{yy}, it need only be shown that $\bar{y}$ is independent of both S_{xy} and S_{yy}. The independence of $\bar{y}$ and S_{xy} follows from the fact that $Cov(\bar{y}, S_{xy}) = 0$. This can be shown by straightforward expectations and implies independence under the normality assumption. One can also reason the independence of $\bar{y}$ and S_{xy} from elementary forward regression theory where it is proved that $\bar{y}$ is independent of $b = S_{xy}/S_{xx}$. The independence of $\bar{y}$ and S_{yy} follows from elementary statistical inference and can be shown by invoking Theorem 4.17 in Graybill (1961). Hence, expressions (2.3) and (2.4) can be written

$$E(\hat{x}_{IL}) = \bar{x} + E(d)\ E(y - \bar{y}) \tag{2.5}$$

$$E(\hat{x}^2_{IL}) = \bar{x}^2 + 2\bar{x}E(d)\ E(y - \bar{y}) + E(d^2)\ E(y - \bar{y})^2. \tag{2.6}$$

The moments of $(y - \bar{y})$ are easily found to be

$$E(y - \bar{y}) = \beta(x - \bar{x}) \tag{2.7}$$

$$E(y - \bar{y})^2 = \sigma^2(1 + 1/N) + \beta^2(x - \bar{x})^2\ . \tag{2.8}$$

Incorporating (2.7) and (2.8) into (2.5) and (2.6), the moments of $\hat{x}_{IL}$ can be expressed as

$$E(\hat{x}_{IL}) = \bar{x} + \beta(x - \bar{x})\ E(d) \tag{2.9}$$

$$E(\hat{x}^2_{IL}) = \bar{x}^2 + 2\bar{x}\beta(x - \bar{x})\ E(d) + [\sigma^2(1 + 1/N) + \beta^2(x - \bar{x})^2]\ E(d^2)\ . \tag{2.10}$$

Also, the variance of $\hat{x}_{IL}$ can be expressed as

$$\begin{aligned} Var(\hat{x}_{IL}) = {} & [\sigma^2(1 + 1/N) + \beta^2(x - \bar{x})^2]\ E(d^2) \\ & - \beta^2(x - \bar{x})^2\ (Ed)^2\ . \end{aligned} \tag{2.11}$$

Using (2.2) and the fact that

$$(\hat{x}_{IL} - E\hat{x}_{IL})^2 = (x - \bar{x})^2 + \beta^2(x - \bar{x})^2\ (Ed)^2 - 2\beta(x - \bar{x})^2\ E(d)\ ,$$

the MSE $(\hat{x}_{IL})$ can now be expressed as

$$E(\hat{x}_{IL} - x)^2 = [\sigma^2(1 + 1/N) + \beta^2(x - \bar{x})^2]\ E(d^2)$$

$$- 2\beta(x - \bar{x})^2\ E(d) + (x - \bar{x})^2\ . \tag{2.12}$$

Since the moments of d do not involve x, the integration necessary for (2.1) can be performed before specific expressions for these moments are obtained. Integrating term by term and rearranging terms it is established that

$$J_{IL} = \{2/3 + 2\bar{x}^2 - E(d)\ [4\bar{x}^2\ \beta + (4/3)\ \beta]$$

$$+ E(d^2)\ [2\sigma^2(1 + 1/N) + (2/3)\ \beta^2 + 2\bar{x}^2\ \beta^2]\}/2\ . \tag{2.13}$$

Expressions are now needed for the first two moments of d.

(2.2) Derivation of the Moments of d

First the density of $d = S_{xy}/S_{yy}$ will be derived. It seems apparent that S_{xy} and S_{yy} are dependent since they are both functions of the same random variables, the y_i. However, this does not guarantee dependence, and to resolve this problem, consider finding the $Cov(S_{xy}, S_{yy})$. It is clear that a non-zero covariance will establish that S_{xy} and S_{yy} are dependent. Under the normality assumption, it is easily shown that

$$S_{xy} \sim N(\beta S_{xx}, \sigma^2 S_{xx}) \tag{2.14}$$

$$S_{yy} \sim \sigma^2\ \chi^2_\nu(\lambda) \tag{2.15}$$

where $\nu = N\text{-}1$ and the non-centrality parameter $\lambda = \beta^2 S_{xx}/(2\sigma^2)$. Using the fact that the expected value of a non-central chi-square variable is $\nu + 2\lambda$, it is seen that

$$E(S_{yy}) = \nu\sigma^2 + \beta^2 S_{xx} \ . \tag{2.16}$$

Expanding S_{xy} and S_{yy}, taking term by term expectations of the product $S_{xy}S_{yy}$, and using the above results in the relationship $\mathrm{Cov}(S_{xy}, S_{yy}) = E(S_{xy}S_{yy}) - E(S_{xy})\,E(S_{yy})$, it can be shown that $\mathrm{Cov}(S_{xy}, S_{yy}) = 2\sigma^2\beta S_{xx}$ and hence that S_{xy} and S_{yy} are dependent. The ratio d, however, can be decomposed into independent components by observing that

$$S_{yy} = S_{xy}^2/S_{xx} + SS_e \tag{2.17}$$

(see, for example, Wine (1964) or Graybill (1961)). The first term on the right hand side of (2.17) is the sum of squares due to regression from elementary regression theory, and the second term is the usual residual sum of squares which is well known to be a central chi-square variable with N-2 degrees of freedom. One can now write

$$d = S_{xy}/S_{yy} = u/(u^2/s + v) \tag{2.18}$$

where $u = S_{xy}$, $v = SS_e$, $s = S_{xx}$ and u and v are independent. Due to the independence of u and v, their joint density can be written as

$$f(u, v) = \frac{v^{\theta-1}\, e^{-v/(2\sigma^2) - (u - \beta s)^2/(2\sigma^2 s)}}{\sqrt{2\pi\sigma^2 s}\ \Gamma(\theta)\ (2\sigma^2)^{\theta}} \tag{2.19}$$

for $-\infty < u < \infty$ and $0 < v < \infty$ where $\theta = (N-2)/2$. Consider now the density of (w, z) where $w = u/v^{1/2}$ and $z = v^{1/2}$. The Jacobian of transformation is $2z^2$ so

$$g(w, z) = f(wz, z^2)\ 2z^2$$

$$= \frac{2z^{2\theta}\, e^{-z^2/(2\sigma^2) - (wz - \beta s)^2/(2\sigma^2 s)}}{\sqrt{2\pi\sigma^2 s}\ \Gamma(\theta)\ (2\sigma^2)^\theta} \tag{2.20}$$

for $-\infty < w < \infty$ and $0 < z < \infty$. Next consider the density of (m, n) where $m = w$ and $n = z(w^2/s + 1)$. The Jacobian of transformation is $1/(m^2/s + 1)$ and the density of (m, n) is

$$h(m, n) = g(m, n/(m^2/s + 1))[1/(m^2/s + 1)]$$

$$= \frac{2n^{2\theta}\ e^{-A}}{\sqrt{2\pi\sigma^2 s}\ \Gamma(\theta)\ (2\sigma^2)^\theta\ (m^2/s + 1)^{2\theta + 1}} \tag{2.21}$$

for $-\infty < m < \infty$ and $0 < n < \infty$ where

$$A = \left[\frac{n}{m^2/s + 1}\right]^2 /(2\sigma^2) + \left[\frac{mn}{m^2/s + 1} - \beta s\right]^2 /(2\sigma^2 s) .$$

A little reflection on the above sequence of transformations reveals that

$$\frac{m}{n} = \frac{w}{z(w^2/s + 1)} = \frac{u/v^{1/2}}{v^{1/2}(u^2/(vs) + 1)} = \frac{u}{(u^2/s + v)} = d . \tag{2.22}$$

Equations (2.21) and (2.22) can now be used to find the density of d by first deriving the density of (d, n) and then finding the marginal density of d. Consider first the density of (d, n) where $d = m/n$. The Jacobian of transformation is n and the density of (d, n) is given by

$$k(d, n) = h(dn, n)n$$

$$= \frac{2n^{N-1}e^{-A'}}{\Gamma(\theta)\ (2\sigma^2)^\theta \sqrt{2\pi\sigma^2 s}\ (d^2n^2/s + 1)^{N-1}}$$

for $-\infty < d < \infty$ and $0 < n < \infty$. Hence, the density of d is

$$f(d) = \int_0^\infty \frac{2n^{N-1}e^{-A'}}{\Gamma(\theta)\ (2\sigma^2)^\theta \sqrt{2\pi\sigma^2 s}\ (d^2n^2/s + 1)^{N-1}}\, dn \qquad (2.23)$$

for $-\infty < d < \infty$ where

$$A' = \left[\frac{n}{d^2n^2/s + 1}\right]^2 / (2\sigma^2) + \left[\frac{dn^2}{d^2n^2/s + 1} - \beta s\right]^2 / (2\sigma^2 s) .$$

The $r^{\underline{th}}$ moment of d can now be expressed as

$$E(d^r) = \int_{-\infty}^{\infty} d^r f(d)dd . \qquad (2.24)$$

Unfortunately, the integral of equation (2.24) cannot be expressed in closed form and one must resort to numerical integration. This renders equation (2.24) useless from a practical standpoint as it is not conducive to efficient numerical integration.

As an alternate approach, rather than using directly the density of d, consider using the joint density of (u, v) in (2.19) and writing

$$E(d^r) = E\{[u/(u^2/s + v)]^r\}$$

$$= \int_0^\infty \int_{-\infty}^\infty [u/(u^2/s + v)]^r \; g(u, v) \; dudv \; . \qquad (2.25)$$

Using the transformation from (u, v) to (z, w) where $z = u$ and $w = v^{1/2}$, equation (2.25) for $r = 1, 2$ becomes

$$E(d) = Ks \int_0^\infty \int_{-\infty}^\infty (zw^{N-3} \; e^{-B})/(z^2 + w^2 \; s) \; dzdw \qquad (2.26)$$

$$E(d^2) = Ks^2 \int_0^\infty \int_{-\infty}^\infty (z^2 \; w^{N-3} \; e^{-B})/(z^2 + w^2 \; s)^2 \; dzdw \qquad (2.27)$$

where

$$K = 2/\left[\Gamma(\theta) \; (2\sigma^2)^\theta \sqrt{2\pi\sigma^2 s}\right]$$

$$B = w^2/(2\sigma^2) + (z - \beta s)^2/(2\sigma^2 s) \; .$$

Equations (2.26) and (2.27) also cannot be expressed in closed form; however, they do lend themselves quite well to numerical integration.

Before proceeding, it will be necessary to establish that the above moments exist. It will be sufficient to establish the existence of $E(d^2)$ since $E(d^2) \geq (E(d))^2$. In a later section in this chapter on Error Analysis, bounds will be placed on the integral in (2.27) for $|z| \geq 1$ and $w \geq 1$ (see page 18) thus establishing existence for those range of values. So for the present discussion, it will suffice to establish existence for $|z| \leq 1$ and $w \leq 1$. Consider

$$I = \iint_G z^2 w^{N-3}/(z^2 + w^2 s)^2 \, dz dw$$

where G is a semicircle of radius R centered at the origin. It is clear that the integrand in I is greater than or equal to the integrand in (2.27) in the region G. It is now convenient to transform to polar coordinates. Letting $z = r \cos \theta$ and $w = r \sin \theta$, one obtains

$$I = \int_0^{\pi} \int_0^{R} \frac{r^{N-4} \cos^2 \theta \sin^{N-4} \theta}{(\cos^2 \theta + s \sin^2 \theta)^2} \, dr d\theta$$

$$= \frac{R^{N-3}}{N-3} \int_0^{\pi} \frac{\cos^2 \theta \sin^{N-4} \theta}{(\cos^2 \theta + s \sin^2 \theta)^2} \, d\theta \; . \qquad (2.28)$$

Setting $R = 2$ to encompass the region in question, namely, $|z| \leq 1$, $w \leq 1$, it is seen that I is finite provided $s > 0$ and $N \geq 4$. One can reason as follows. The denominator cannot vanish in the integrand in (2.28) for $s > 0$ and the numerator is bounded for $N \geq 4$. Hence, the integrand is bounded and the integral exists provided $s > 0$ and $N \geq 4$ thus establishing the existence of (2.27) in the region $|z| \leq 1$, $w \leq 1$.

Expressions (2.26) and (2.27) for $E(d)$ and $E(d^2)$ can now be substituted into (2.13) yielding an expression for J_{IL} which can be evaluated numerically. Note that J_{IL} is a function of N, $\bar{x}$, s, β and σ and that in order to evaluate J_{IL}, these quantities will have to be assigned fixed values. Note also that β, the slope of the regression line, and $\bar{x}$ can take on both positive and negative values while N, s, and σ are strictly positive. It is seen that it is highly desirable to combine or eliminate some of these variables. Fortunately, one can show that β and σ can be considered as a single quantity $\gamma = |\beta/\sigma|$. The next section is devoted to this proof.

(2.3) Proof that J_{IL} is a Function of N, $\bar{x}$, s, and $\gamma = |\beta/\sigma|$

Letting $\gamma' = \beta/\sigma$ in (2.13) yields

$$2\,J_{IL} = 2/3 + 2\bar{x}^2 - E(d)\,\sigma\,\gamma'\,[4\bar{x}^2 + 4/3]$$
$$+ E(d^2)\,\sigma^2[2(1 + 1/N) + (2/3)\,\gamma'^2 + 2\bar{x}^2\,\gamma'^2]\ . \quad (2.29)$$

Note that E(d) has a coefficient σ and $E(d^2)$ has a coefficient σ^2. It will now be proved that σ E(d) and σ^2 $E(d^2)$ are both functions of N, s, and γ' and thus $2J_{IL}$ is a function of N, $\bar{x}$, s, and γ'. In the expression for E(d) in (2.26), consider the transformation from (z, w) to (m, n) where $m = w/\sigma$ and $n = z/\sigma$. The Jacobian of the transformation is σ^2 and the range of integration remains the same so

$$E(d) = \frac{s^{1/2}}{\Gamma(\theta)\,2^{\theta-1}\,\sqrt{2\pi}\,\sigma}\int_0^{\infty}\int_{-\infty}^{\infty}\frac{nm^{N-3}\,e^{-D}}{n^2 + m^2\,s}\,dndm$$

where

$$D = m^2/2 + (n - \gamma'\,s)^2/(2s)\ .$$

Hence, σ E(d) is a function of only N, s, and γ'. Using the same transformation in (2.27) for $E(d^2)$ yields

$$E(d^2) = \frac{s^{3/2}}{\Gamma(\theta)\ 2^{\theta-1}\ \sqrt{2\pi}\ \sigma^2} \int_0^\infty \int_{-\infty}^\infty \frac{n^2 m^{N-3}\ e^{-D}}{(n^2 + m^2\ s)^2}\, dndm\ ,$$

so $\sigma^2\ E(d^2)$ is also a function of only N, s, and γ'. Thus, the result is established that $2J_{IL}$ is a function of N, $\bar{x}$, s, and γ'.

The effect of a change in sign in β and hence γ' will now be examined. Suppose the β in the integrand in (2.26) is positive (negative). Changing the sign of β to negative (positive) merely results in a mirror image of the integrand. Since the integrand is negative for $z < 0$ and positive for $z > 0$, the result in E(d) is a change in sign. But in J_{IL} in (2.13), E(d) has a coefficient β so changing the sign of β has no effect on β E(d). A change in sign in β in the integrand in (2.27) also results in a mirror image of the integrand, but the integrand is positive or zero for all z so there is no effect on $E(d^2)$ or $\beta^2\ E(d^2)$ in J_{IL} in (2.13). Hence, changing the sign of β has no effect on J_{IL} and with the above results, it can be concluded that J_{IL} is a function of N, $\bar{x}$, s, and $\gamma = |\beta/\sigma|$.

In order to minimize $2J_{IL}$ with respect to the design variables, integrals (2.26) and (2.27) will eventually have to be evaluated numerically. And since the range of integration of $w = (0, \infty)$ and the range of integration of $z = (-\infty, \infty)$, finite quantities a, b, and c will have to be found so that if one numerically integrates over the limits $w = (0, a)$ and $z = (b, c)$, the integral of the neglected region will be negligible. These finite limits will clearly depend on the magnitude of N, s, β, and σ; they are derived in the next section.

(2.4) Error Analysis

It will be convenient to divide the first two quadrants of the (z, w) plane into four regions, the region z = (b, c) and w = (0, a) over which numerical integration will be performed to obtain E(d) and E(d^2), and the three additional rectangular regions, the integral over which will comprise the error in obtaining the moments by integrating over finite limits. The quantities a, b, and c are sought such that bounds can be placed on the error

$$E = E_a + E_b + E_c$$

$$= \int_a^{\infty}\int_{-\infty}^{\infty} k_i(z, w)\ dzdw + \int_0^a\int_{-\infty}^b k_i(z, w)\ dzdw$$

$$+ \int_0^a\int_c^{\infty} k_i(z, w)\ dzdw,\ i = 1, 2$$

where k_1 represents the integrand in (2.26) and k_2 the integrand in (2.27). Each region must be considered separately for both E(d) and E(d^2). Expressions for a, b, and c and the bounds on the corresponding errors will be derived for E(d^2); for E(d) the derivation will be omitted and only the results will be shown. For E(d^2),

$$E_a^{(2)} = Ks^2 \int_a^{\infty}\int_{-\infty}^{\infty} \frac{z^2\ w^{N-3}}{(z^2 + w^2\ s)^2}\ e^{-w^2/(2\sigma^2)\ -\ (z\ -\ \beta s)^2/(2\sigma^2\ s)}\ dzdw\ ,$$

but

$$\frac{z^2 w^{N-3}}{(z^2 + w^2 s)^2} \le \frac{z^2 w^{N-3}}{a^4 s^2}$$

for all z and $w \ge a$ so that

$$E_a^{(2)} \le (K/a^4) \int_a^{\infty} w^{N-3} e^{-w^2/(2\sigma^2)} dw \int_{-\infty}^{\infty} z^2 e^{-(z - \beta s)^2/(2\sigma^2 s)} dz . \tag{2.30}$$

The superscript (2) indicates the error is for $E(d^2)$. Similarly, a superscript (1) will be placed on errors for $E(d)$. The last integral on the right hand side of (2.30) is simply $\sqrt{2\pi\sigma^2 s}$ times the second moment of a normal variable with mean βs and variance $\sigma^2 s$ so

$$E_a^{(2)} \le \left[K\sqrt{2\pi\sigma^2 s} \; (\sigma^2 s + \beta^2 s^2)/a^4 \right] \int_a^{\infty} w^{N-3} e^{-w^2/(2\sigma^2)} dw .$$

Using the transformation $v = w^2/(2\sigma^2)$, obtaining K from page 14, and simplifying the multiplying constants yields

$$E_a^{(2)} \le \frac{(\sigma^2 s + \beta^2 s^2)}{a^4} \int_{a^2/(2\sigma^2)}^{\infty} \frac{v^{(N-4)/2} e^{-v}}{\Gamma[(N-2)/2]} dv .$$

The above integral is the well known incomplete gamma function and can be written in Pearson's notation as

$$E_a^{(2)} \leq \frac{(\sigma^2 s + \beta^2 s^2)}{a^4} \left\{ 1 - I \left[\frac{a^2}{2\sigma^2 \sqrt{(N-2)/2}} , \frac{N-4}{2} \right] \right\} \qquad (2.31)$$

where

$$I(u, p) = \int_0^{u\sqrt{p+1}} \frac{v^p e^{-v}}{\Gamma(p+1)} dv .$$

The incomplete gamma function has been extensively tabulated by Pearson (1957), and Pearson's tables can be used to determine "a" interatively for any desired bound on $E_a^{(2)}$. An example will be given at the end of this section.

Next consider $E_b^{(2)}$ where

$$E_b^{(2)} = Ks^2 \int_0^a \int_{-\infty}^b \frac{z^2 w^{N-3} e^{-B}}{(z^2 + w^2 s)^2} dzdw .$$

For $z \leq b \leq -1$, $z^2/(z^2 + w^2 s)^2 \leq 1$ so that

$$E_b^{(2)} \leq Ks^2 \int_0^a w^{N-3} e^{-w^2/(2\sigma^2)} dw \int_{-\infty}^b e^{-(z - \beta s)^2/(2\sigma^2 s)} dz$$

$$\leq s^2 \; I\left[\frac{a^2}{2\sigma^2 \sqrt{(N-2)/2}}, \frac{N-4}{2}\right] \int_{-\infty}^{b} \frac{e^{-(z-\beta s)^2/(2\sigma^2 s)}}{\sqrt{2\pi\sigma^2 s}} \, dz$$

$$\leq s^2 \; \Phi \; [(b - \beta s)/\sqrt{\sigma^2 s}] \tag{2.32}$$

where

$$\Phi(z) = (1/\sqrt{2\pi}) \int_{-\infty}^{z} e^{-t^2/2} \, dt \; .$$

Here the bound on $E_b^{(2)}$ can be made as small as desired by using the cumulative normal table for proper selection of b.

An argument similar to the one above leads to a bound on $E_c^{(2)}$ for $c \geq 1$. The result is

$$E_c^{(2)} \leq s^2 \; \{1 - \Phi[(c - \beta s)/\sqrt{\sigma^2 s}]\} \; . \tag{2.33}$$

Note that the bound on $E_a^{(2)}$ is valid for all $a > 0$ while the bound on $E_b^{(2)}$ is valid only for $b \leq -1$ and the bound on $E_c^{(2)}$ is valid only for $c \geq +1$. This range of values is sufficient for small values of β. However, as β increases, the appreciable portion of the integrand shifts drastically to the right and to always maintain $b \leq -1$ results in inefficient numerical integration. It should be observed that it also shifts to the left as β increases negatively but since β appears in J_{IL} only in absolute value, negative values of β will not have to be considered. Consider now the derivation of a bound on $E_b^{(2)}$ where b is not restricted to values of $b \leq -1$. $E_b^{(2)}$ is given by

$$E_b^{(2)} = Ks^2 \int_0^a \int_{-\infty}^b \frac{z^2 w^{N-3} e^{-B}}{(z^2 + w^2 s)^2} dzdw ,$$

but $z^2 w^2/(z^2 + w^2 s)^2 \leq \max(1, 1/s^2)$ for all z and w except $z = w = 0$. Therefore,

$$E_b^{(2)} \leq Ks^2 \max(1, 1/s^2) \int_0^a w^{N-5} e^{-w^2/(2\sigma^2)} dw \int_{-\infty}^{\infty} e^{-(z - \beta s)^2/(2\sigma^2 s)} dz$$

$$\leq \frac{s^2 \max(1, 1/s^2)}{(N-4)\sigma^2} \Phi[(b - \beta s)/\sqrt{\sigma^2 s}] \qquad (2.34)$$

for all real b provided $N \geq 5$.

Before proceeding to bounds for the errors for E(d), it should be noted that bounds (2.31), (2.32), and (2.33) are those referred to on page 14 in reference to establishing the existence of $E(d^2)$. It should also be noted that while $E(d^2)$ exists for $N \geq 4$, the bound shown in (2.34) is valid only for $N \geq 5$.

The bounds for the errors for E(d) are listed below. These bounds are more difficult to derive than those for $E(d^2)$ since the integrand in (2.26) is negative for $z < 0$ and two of the three regions described above must be further subdivided to obtain both upper and lower bounds on the errors. The bounds are as follows:

$$-GH/a^2 \leq E_a^{(1)} \leq G(\beta s + H)/a^2, \ a > 0 \qquad (2.35)$$

$$-s\Phi[(b - \beta s)/\sqrt{\sigma^2 s}] \leq E_b^{(1)} \leq 0, \ b \leq -1 \qquad (2.36)$$

$$\min(-1, -1/s)\ C\Phi[(b - \beta s)/\sqrt{\sigma^2 s}] \leq E_b^{(1)} \leq 0,\ b \leq 0 \tag{2.37}$$

$$\min(-1, -1/s)\ C\Phi(-\beta s/\sqrt{\sigma^2 s}) \leq E_b^{(1)} \leq$$

$$\max(1, 1/s)\ C\{\Phi[(b - \beta s)/\sqrt{\sigma^2 s}] - \Phi(-\beta s/\sqrt{\sigma^2 s})\},\ b \geq 0 \tag{2.38}$$

$$0 \leq E_c^{(1)} \leq s\{1 - \Phi[(c - \beta s)/\sqrt{\sigma^2 s}]\},\ c \geq 1 \tag{2.39}$$

where

$$G = 1 - I\left[\frac{a^2}{2\sigma^2\sqrt{(N-2)/2}}, \frac{N-4}{2}\right]$$

$$H = \sqrt{\sigma^2 s/(2\pi)}\ e^{-\beta^2 s/(2\sigma^2)}$$

$$C = \frac{s\Gamma[(N-3)/2]}{\Gamma[(N-2)/2]\sqrt{2\sigma^2}}.$$

Note the multiple bounds above for $E_b^{(1)}$, each valid for the specified range of values on b. Such multiple bounds are not necessary for $E_c^{(1)}$ or $E_c^{(2)}$ since for $\beta > 0$, c will always be greater than or equal to one.

Since it was discovered that values of E(d) and $E(d^2)$ were greater than 10^{-6} for a practical range of values of N, s, β, and σ, it was decided to maintain E at less than 10^{-6} to provide accuracy of three significant digits for all numerical results presented in this thesis. This was accomplished by maintaining $E_i^{(j)}$ at less than $(30)(10^{-8})$ for $i = a, b, c$ and $j = 1, 2$. To determine "a," Pearson's tables can be used to select a value of u such that I(u, p) is close to one. Then

$$\frac{a^2}{2\sigma^2 \sqrt{(N-2)/2}} = u \tag{2.40}$$

is solved for "a" and $E_a^{(1)}$ and $E_a^{(2)}$ are determined. These bounds may be smaller than specified at the expense of a large "a" which would result in inefficient numerical integration or they could be larger than specified by yielding a small "a." In either case, the "a" for each must be adjusted by selecting a value of u such that I(u, p) is smaller or larger as the case dictates. For example, consider the case $N = 10$, $\beta = 5$, $\sigma = 1$, and $s = 10$. Since $p = (N-4)/2$, for $N = 10$, $p = 3$. From Pearson's tables, $I(u, p) = I(12.4, 3) = 1.0000000 \geq .99999995$ so $1 - I(12.4, 3) \leq (5)(10^{-8})$. Solving (2.40) for "a" yields $a = 7.04$. Using (2.31),

$$E_a^{(2)} \leq \frac{[10 + (25)(100)](5)(10^{-8})}{(7.04)^4}$$

$$= (5.10)(10^{-8})$$

which is very tolerable for $E(d^2)$. Using (2.35),

$$H = (10/(2\pi))^{1/2} e^{-125} = 0.0$$

and

$$E_a^{(1)} \leq \frac{(5)\ (10^{-8})\ (5)\ (10)}{(7.04)^2}$$

$$= (5.04)\ (10^{-8})$$

which is very tolerable for E(d). Hence, (0, a) = (0, 7.04) would be an appropriate range of integration on w for both E(d) and $E(d^2)$.

To determine b and c the following rules usually provide values for which the resulting bounds are tolerable.

(1) Set $(c - \beta s)/(\sigma^2 s)^{1/2} = 6.0$ and solve for c. (2.41)

(2) Set $(b - \beta s)/(\sigma^2 s)^{1/2} = -6.0$ and solve for b. (2.42)

Of course, if the resulting errors are not within the specified bounds, the values of b and c must be adjusted by taking a smaller or larger value than 6.0 in (2.41) and (2.42). Continuing with the previous example, use of (2.41) results in

$$c = \beta s + (6.0)\ (10)^{1/2} = 68.97$$

and use of (2.42) results in

$$b = \beta s - (6.0)\ (10)^{1/2} = 31.03\ .$$

Using (2.34),

$$E_b^{(2)} \leq (100/6)\ \phi(-6)$$

$$= (1.75)\ (10^{-8})$$

since $\Phi(-6) \leq (10.5)\ (10^{-10})$. Hence, the value of b provides a tolerable error $E_b^{(2)}$ for $E(d^2)$. Next, using (2.33) results in $E_c^{(2)} \leq (10.5)\ (10^{-8})$ which is also very acceptable. Using (2.38) yields $0 \leq E_b^{(1)} \leq (4.11)\ (10^{-9})$ and using (2.39) yields $E_c^{(1)} \leq (1.05)\ (10^{-8})$ both of which are acceptable. Hence, values of $b = 31.03$ and $c = 68.97$ for both $E(d)$ and $E(d^2)$ would result in errors which are within the specified bounds.

Reviewing the work in this chapter, an expression has been derived for J_{IL} in terms of the first two moments of the ratio $d = S_{xy}/S_{yy}$. Furthermore, exact expressions for these moments have been developed and proved to exist for $N \geq 4$ and $s > 0$. However, the expressions for these moments involve double integrals over infinite limits which cannot be expressed in closed form. In this section, finite quantities a, b, and c have been developed such that one can extract the moments by numerically integrating over these finite limits and place bounds on the resulting errors. While a general set of working rules have been developed for determining a, b, and c, it must be emphasized that the resulting errors must be checked in each case to insure they are within the specified bounds, that is, for each set of fixed quantities N, s, β, and σ. Examination of the expressions for these bounds is testimony that the first two moments of d can be extracted numerically only by considerable effort. In light of this, an approximation to the moments of d is sought which, at least for some range of values on the parameters involved, will be close enough to the exact moments to use safely and thus lighten the numerical effort. It is to be emphasized that the approximation sought will be used only for the range of values of the parameters where it is safe in terms of closeness to the exact moments. Also, the ensuing approximation for the moments of d in the linear model under consideration will lay the groundwork for an approximation to the moments of d when the true model is quadratic. It will be shown in Chapter IV that when the true model is quadratic, the method of extracting the moments by numerical integration is outside the realm of practical feasibility. It is clear that if a Taylor expansion is used in this approximation, it will involve product moments of the form $E(S_{xy}^r S_{yy}^p)$ and an efficient method is needed to obtain them. The "straightforward" method used to obtain $E(S_{xy}\ S_{yy})$ earlier in establishing the dependence of these quantities was extremely inefficient, and to obtain moments by this method for $r > 1$ and $p > 1$ would be a Herculean job. The obvious method of obtaining such product moments efficiently is by using the joint moment generating function of S_{xy} and S_{yy}. In preparation for an approximation to the moments of d, the joint moment generating function of S_{xy} and S_{yy} will be developed in the next section.

(2.5) Derivation of the Joint Moment Generating Function of S_{xy} and S_{yy}

In this section, vectors will be represented by underlined lower case letters and matrices will be represented by capital letters. Let the vector of observation y_i, $i = 1,\ldots,N$ be represented by $\underline{y}' = (y_1,\ldots,y_N)$. The distribution of $\underline{y}$ is N dimensional multivariate normal denoted by

$$\underline{y} \sim N_N\ (\underline{\mu},\ \sigma^2 I) \tag{2.43}$$

where $\underline{\mu}' = (\alpha + \beta x_1,\ldots,\alpha + \beta x_N)$, I is the identity matrix, and the subscript N means N dimensional. Clearly the density

$$f(\underline{y}) = \frac{e^{-(\underline{y}-\underline{\mu})'(\underline{y}-\underline{\mu})/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} \qquad (2.44)$$

for $-\infty < y_i < \infty$, $i = 1,\ldots,N$. Now let

$$w = S_{xy} = \sum_{i=1}^{N} (x_i - \bar{x})\, y_i = \sum_{i=1}^{N} a_i y_i = \underline{y}'\underline{a}$$

$$z = S_{yy} = \sum_{i=1}^{N} (y_i - \bar{y})^2 = \underline{y}' A\, \underline{y}$$

where $a_i = x_i - \bar{x}$ and A is the NxN symmetric idempotent matrix of rank $\nu = N-1$ with elements $1-1/N$ on the main diagonal and off diagonal elements of $-1/N$. By definition, the joint moment generating function of w and z is

$$\phi(t_1, t_2) = E(e^{t_1 w+t_2 z}) = \int \cdots \int e^{t_1 w+t_2 z} f(\underline{y})\, d\underline{y}$$

$$= [1/(2\pi\sigma^2)]^{N/2} \int \cdots \int e^{-(\underline{y}-\underline{\mu})'(\underline{y}-\underline{\mu})/(2\sigma^2) + t_1 w+t_2 z}\, d\underline{y} \qquad (2.45)$$

where the range of integration on each of the N integrals is $(-\infty, \infty)$ and $d\underline{y}$ means $dy_1 dy_2 \ldots dy_N$. Consider now the transformation $\underline{z} = \underline{y}-\underline{\mu}$. The Jacobian of transformation is unity and the exponent in (2.45) becomes

$$-\underline{z}'\underline{z}/(2\sigma^2) + t_1(\underline{z}+\underline{\mu})'\,\underline{a} + t_2(\underline{z}+\underline{\mu})'\,A(\underline{z}+\underline{\mu})$$

which upon expansion of the last term and simplification becomes

$$-\underline{z}'(I-2\sigma^2 t_2 A)\underline{z}/(2\sigma^2) + \underline{z}'(2t_2\underline{\mu}'A+t_1\underline{a}')'$$

$$+ t_2\underline{\mu}' A\underline{\mu} + t_1\underline{a}'\underline{\mu} \ . \qquad (2.46)$$

The last two terms in (2.46) are constants with respect to integration and the following identity will be useful in representing the first two terms as a quadratic form in $\underline{z}$ and an additional constant:

$$(\underline{z}'\underline{g} - \underline{z}'V\underline{z}/2)/\sigma^2 = [\underline{g}'V^{-1}\underline{g} - (\underline{z}-V^{-1}\underline{g})' V(\underline{z}-V^{-1}\underline{g})]/(2\sigma^2) \ , \qquad (2.47)$$

when V is symmetric of order N and non-singular. Identity (2.47) can be proved by simply expanding the right hand side. Letting

$$\underline{g} = (2t_2 A\underline{\mu}+t_1\underline{a})\sigma^2$$

$$V = (I-2\sigma^2 t_2 A)$$

and using (2.47) in the first two terms of (2.46) results in the following exponent:

$$\underline{g}'V^{-1}\underline{g}/(2\sigma^2) - (\underline{z}-V^{-1}\underline{g})' V(\underline{z}-V^{-1}\underline{g})/(2\sigma^2)$$

$$+ t_2\underline{\mu}' A\underline{\mu} + t_1\underline{a}'\underline{\mu} \ . \qquad (2.48)$$

Incorporating (2.48) into (2.45) yields

$$\phi(t_1, t_2) = \left(\frac{e^{\underline{g}'V^{-1}\underline{g}/(2\sigma^2) + t_1\underline{a}'\underline{\mu} + t_2\underline{\mu}'A\underline{\mu}}}{(2\pi\sigma^2)^{N/2}}\right)$$

$$\int \cdots \int e^{-(\underline{z}-V^{-1}\underline{g})' V(\underline{z}-V^{-1}\underline{g})/(2\sigma^2)} \, d\underline{z}$$

$$= |V|^{-1/2} e^{\underline{g}'V^{-1}\underline{g}/(2\sigma^2) + t_1\underline{a}'\underline{\mu} + t_2\underline{\mu}'A\underline{\mu}} \tag{2.49}$$

which is the joint moment generating function of S_{xy} and S_{yy} expressed in very general terms. It must now be simplified. Consider first

$$|V|^{-1/2} = |I-2\sigma^2 t_2 A|^{-1/2} .$$

Since A is symmetric idempotent of rank N-1, there exists an orthogonal matrix P such that PAP' is diagonal with N-1 ones and one zero on the main diagonal. (See Theorem 1.60 in Graybill.) Furthermore, $P(I-2\sigma^2 t_2 A)P'$ is diagonal with N-1 terms of the form $1-2\sigma^2 t_2$ and one 1 on the main diagonal so that

$$|I-2\sigma^2 t_2 A|^{-1/2} = (1-2\sigma^2 t_2)^{-(N-1)/2} . \tag{2.50}$$

It is easily seen that

$$\underline{\mu}'A\underline{\mu} = \sum_{i=1}^{N} [E(y_i - \bar{y})]^2 = \beta^2 S_{xx} \tag{2.51}$$

with $\underline{\mu}$ and A defined on page 27 and that

$$\underline{a}'\underline{\mu} = \beta S_{xx} . \tag{2.52}$$

It remains to simplify the first term in the exponent of (2.49), namely $\underline{g}'V^{-1}\underline{g}/(2\sigma^2)$. Clearly $V^{-1} = (I-2\sigma^2 t_2 A)^{-1}$ is of the form $(rI+qJ)^{-1}$ which is well known to equal $aI+bJ$ with $a = 1/r$ and

$b = -q/[r(Nq+r)]$ where N is the dimension of I and J and J is the matrix with each element equal to unity. In the above case, $r = 1-2\sigma^2 t_2$ and $q = 2\sigma^2 t_2/N$ so $a = 1/r = 1/[1-2\sigma^2 t_2]$ and $b = -2\sigma^2 t_2/[N(1-2\sigma^2 t_2)]$ and $(I-2\sigma^2 t_2 A)^{-1}$ is of the form $1/[1-2\sigma^2 t_2]$ times the following matrix:

$$\begin{bmatrix} 1-2\sigma^2 t_2/N & -2\sigma^2 t_2/N & -2\sigma^2 t_2/N & \cdots & -2\sigma^2 t_2/N \\ & 1-2\sigma^2 t_2/N & -2\sigma^2 t_2/N & \cdots & -2\sigma^2 t_2/N \\ & & 1-2\sigma^2 t_2/N & \cdots & -2\sigma^2 t_2/N \\ \text{sym.} & & & & 1-2\sigma^2 t_2/N \end{bmatrix} . \quad (2.53)$$

Expanding $\underline{g}'V^{-1}\underline{g}/(2\sigma^2)$, one obtains

$$(4t_2^2\underline{\mu}'AV^{-1}A\underline{\mu} + 4t_1t_2\underline{\mu}'AV^{-1}\underline{a} + t_1^2\underline{a}'V^{-1}\underline{a})\sigma^2/2$$

which upon substituting (2.53) for V^{-1}, the appropriate expressions for $\underline{\mu}$, $\underline{A}$, and $\underline{a}$ and using the fact that $\Sigma a_i = \Sigma(x_i - \bar{x}) = 0$ reduces to

$$(t_1^2 S_{xx} + 4t_1t_2S_{xx}\beta + 4t_2^2S_{xx}\beta^2)\sigma^2/[2(1-2\sigma^2 t_2)] .$$

Hence, the joint moment generating function for S_{xy} and S_{yy} becomes

$$\phi(t_1, t_2) = \exp\{(t_1^2S_{xx} + 4t_1t_2S_{xx}\beta + 4t_2^2S_{xx}\beta^2)\sigma^2/$$

$$[2(1-2\sigma^2 t_2)] + (t_1\beta S_{xx} + t_2\beta^2 S_{xx})\}$$

$$(1-2\sigma^2 t_2)^{-(N-1)/2}$$

$$= \exp\{(\sigma^2 S_{xx} t_1^2/2 + S_{xx}\beta t_1 + S_{xx}\beta^2 t_2)/$$

$$(1-2\sigma^2 t_2)\}(1-2\sigma^2 t_2)^{-(N-1)/2} . \qquad (2.54)$$

One can now investigate special cases of (2.54) as checks on whether it is, in fact, the correct moment generating function. It is observed that

$$\phi(0,\ 0) = 1$$

as it should be and that

$$\phi(t_1,\ 0) = e^{t_1\beta S_{xx} + t_1^2\sigma^2 S_{xx}/2}$$

which is the moment generating function of a normal variable with mean βS_{xx} and variance $\sigma^2 S_{xx}$ as it should be, namely, the moment generating function of S_{xy}. It is also observed that

$$\phi(0,\ t_2) = \exp\{(4t_2^2 S_{xx}\beta^2\sigma^2)/[2(1-2\sigma^2 t_2)]$$

$$+ t_2\beta^2 S_{xx}\}(1-2\sigma^2 t_2)^{-(N-1)/2} \qquad (2.55)$$

which should be the moment generating function of S_{yy}. It is shown in Graybill (1961) that if the vector $\underline{y}$ is distributed $N_N(\underline{\mu},\ \sigma^2 I)$, the moment generating function of $S_{yy} = \underline{y}'A\underline{y}$ is given by

$$\phi(t_2) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{1}{(1-2\sigma^2 t_2)^{(N-1+2i)/2}} \qquad (2.56)$$

with $\lambda = \underline{\mu}'A\underline{\mu}/(2\sigma^2)$ which in this case is $\lambda = \beta^2 S_{xx}/(2\sigma^2)$. It is not entirely clear that (2.55) and (2.56) are equal. However, substitution of λ for $\beta^2 S_{xx}/(2\sigma^2)$ in the exponent of (2.55), yields

$$\phi(0,\ t_2) = \exp[-\ \lambda + \lambda/(1-2\sigma^2 t_2)]/(1-2\sigma^2 t_2)^{(N-1)/2}$$

and expanding $\exp[\lambda/(1-2\sigma^2 t_2)]$ into a Taylor series, it is readily established that (2.55) and (2.56) are of the same form. Note also that if one partially differentiates (2.54) with respect to t_1 and t_2 and sets $t_1 = t_2 = 0$, he obtains

$$E(wz) = S_{xx}\beta(N+1)\sigma^2 + S_{xx}^2\beta^3$$

and

$$\mathrm{Cov}(w,\ z) = E(wz) - E(w)E(z) = 2\sigma^2\beta S_{xx}$$

which is precisely the same result obtained earlier by expanding wz and taking expectations term by term. Hence, one can be certain that $\phi(t_1,\ t_2)$ in (2.54) is of correct form.

The groundwork is now laid for an approximation to the moments of d to be discussed next.

(2.6) Approximation to the Moments of d

A method of approximating the first two moments of a ratio of random variables suggested by Kendall and Stuart (1963) is that of expanding the ratio in a bivariate Taylor series through the first order about the respective means and then using the expectation of the expansion for the approximation to the first moment and the expectation of the square of the expansion for the second moment. This procedure was attempted and the resulting approximate moments were compared with the exact moments obtained by numerically integrating (2.26) and (2.27).

The results were very poor, the approximate moments deviating twenty to forty percent below the exact moments. To improve the approximation, it was decided to use a second order Taylor expansion of the ratio, taking the expectation of the ratio as an approximation to the first moment and the expectation of the square of the second order expansion as an approximation to the second moment.

Taylor's theorem for two variables needs no introduction. If $f(w, z)$ and its first $n+1$ derivatives are continuous, $f(w, z)$ can be expressed as

$$f(w, z) = \sum_{i=1}^{n} (1/i!) \left[(w-a)(\partial/\partial a) + (z-b)(\partial/\partial b)\right]^i f(a, b) + R_n \tag{2.57}$$

where R_n is the remainder after n terms of the expansion, a and b are fixed, and

$$(\partial/\partial a)^r (\partial/\partial b)^s f(a, b) \text{ means } \partial^{r+s} f(w, z)/\partial w^r \partial z^s$$

evaluated at $w = a$ and $z = b$. As in the previous section, let $w = S_{xy}$, $-\infty < w < \infty$, and $z = S_{yy}$, $0 < z < \infty$. Letting $f(w, z) = w/z$, $a = E(w) = \mu_w$, $b = E(z) = \mu_z$, $n = 2$, and deleting the remainder in (2.57) provides the following expansion:

$$f(w, z) = w/z \cong \mu_w/\mu_z + (w-\mu_w)/\mu_z$$
$$-\mu_w(z-\mu_z)/\mu_z^2 - (w-u_w)(z-u_z)/\mu_z^2 + \mu_w(z-\mu_z)^2/\mu_z^3 \,. \tag{2.58}$$

Taking term by term expectations of (2.58) over both w and z results in

$$E(w/z) = E(d) \cong \mu_w/\mu_z - \mathrm{Cov}(w,\ z)/\mu_z^2 + \mu_w^2\sigma_z^2/\mu_z^3$$

$$= (\mu_w\mu_z^2 - \mu_z\ \mathrm{Cov}(w,\ z) + \mu_w\sigma_z^2)/\mu_z^3\ . \qquad (2.59)$$

Squaring (2.58) and taking term by term expectations, one obtains

$$\mu_z^6\ E(w/z)^2 \cong \mu_z^4\mu_w^2 + \mu_z^4E(w^2) + 3\mu_w^2\mu_z^2E(z^2)$$

$$+ \mu_z^2E[(w-\mu_w)^2\ (z-\mu_z)^2] + \mu_w^2(z-\mu_z)^4$$

$$- 4\mu_z^3\mu_wE(wz) - 2\mu_z^3E[(w-\mu_w)^2\ (z-\mu_z)]$$

$$+ 4\mu_w\mu_z^2E[(w-\mu_w)\ (z-\mu_z)^2] - 2\mu_w^2\mu_zE(z-\mu_z)^3$$

$$- 2\mu_w\mu_zE[(w-\mu_w)\ (z-\mu_z)^3]\ . \qquad (2.60)$$

Expanding the central moments in (2.60) and collecting terms provides

$$\mu_z^6E(w/z)^2 \cong \mu_w^2\mu_z^4 + 4\mu_z^4E(w^2) + 6\mu_w^2\mu_z^2E(z^2)$$

$$+ \mu_z^2E(w^2z^2) - 4\mu_z^3E(w^2z) + 8\mu_w\mu_z^2E(wz^2)$$

$$- 10\mu_w\mu_z^3E(wz) - 2\mu_w\mu_zE(wz^3) + \mu_w^2E(z^4)$$

$$- 4\mu_w^2\mu_zE(z^3)\ . \qquad (2.61)$$

The joint moment generating function of w and z derived in the previous section can now be used to express the moments of w and z in (2.59) and (2.61) in terms of β, σ, N, and S_{xx} by using the fact that

$$E(w^p z^q) = \partial^{p+q}\,\phi(t_1,\ t_2)/\partial t_1^p\ \partial t_2^q\Big|_{t_1 = t_2 = 0}\ .$$

Recall the expression

$$\phi(t_1 t_2) = \exp\{(\sigma^2 S_{xx} t_1^2/2 + t_1 S_{xx}\ \beta + t_2\ \beta^2 S_{xx}/$$

$$(1-2\sigma^2 t_2)\}\ (1-2\sigma^2 t_2)^{-(N-1)/2}\ . \qquad (2.62)$$

To differentiate (2.62) in its present form repeatedly for the higher order moments would be hopeless from an algebraic standpoint. Therefore, consider the following representation of (2.62). Let

$$g(t_2) = 1-2\sigma^2 t_2$$

$$k(t_1,\ t_2) = \sigma^2 S_{xx} t_1^2/2 + t_1 S_{xx}\ \beta + t_2\ \beta^2 S_{xx}$$

and write

$$\phi(t_1,\ t_2) = [g(t_2)]^{-\nu/2}\ e^{g^{-1}(t_2)\ k(t_1,\ t_2)} \qquad (2.63)$$

where $\nu = N-1$. Taking the natural logarithm of both sides of $\phi(t_1,\ t_2)$ in (2.63) yields

$$\ln\ \phi(t_1 t_2) = (-\nu/2)\ \ln\ g(t_2) + g^{-1}(t_2)\ k(t_1,\ t_2) \qquad (2.64)$$

which can be used quite efficiently to express the higher moments as functions of the lower ones. Initially, it is helpful to compute

$$\partial k(t_1,\ t_2)/\partial t_1 = \sigma^2 S_{xx} t_1 + S_{xx}\beta \qquad (2.65)$$

$$\partial^2 k(t_1,\ t_2)/\partial t_1^2 = \sigma^2 S_{xx} \qquad (2.66)$$

$$\partial k(t_1, t_2)/\partial t_2 = \beta^2 S_{xx} \tag{2.67}$$

$$\partial^2 k(t_1, t_2)/\partial t_1 \partial t_2 = 0 \tag{2.68}$$

$$g'(t_2) = - 2\sigma^2 \tag{2.69}$$

and recognize that $k(0, 0) = 0$, $g(0) = 1$, and $\phi(0, 0) = 1$. Differentiating both sides of (2.64) with respect to t_1 yields

$$\partial\phi(t_1, t_2)/\partial t_1 = \phi(t_1, t_2)g^{-1}(t_2)\ \partial k(t_1, t_2)/\partial t_1$$

which upon using (2.65) and (2.69) and setting $t_1 = t_2 = 0$ provides

$$E(w) = \beta S_{xx} . \tag{2.70}$$

This result was known at the onset, but it illustrates the method to be employed. Continuing in this fashion, one obtains

$$E(w^2) = \sigma^2 S_{xx} + (Ew)^2 \tag{2.71}$$

$$E(wz) = S_{xx}\beta[E(z) + 2\sigma^2] \tag{2.72}$$

$$E(wz^2) = S_{xx}\beta[E(z^2) + 4\sigma^2 E(z) + 8\sigma^4] \tag{2.73}$$

$$E(wz^3) = S_{xx}\beta[E(z^3) + 6\sigma^2 E(z^2) + 24\sigma^4 E(z) + 48\sigma^6] \tag{2.74}$$

$$E(w^2z) = S_{xx}\beta[E(wz) + 2\sigma^2 E(w)] + \sigma^2 S_{xx}[E(z) + 2\sigma^2] \tag{2.75}$$

$$E(w^2z^2) = S_{xx}\beta[E(wz^2) + 4\sigma^2 E(wz) + 8\sigma^4 E(w)]$$

$$+ \sigma^2 S_{xx}[E(z^2) + 4\sigma^2 E(z) + 8\sigma^4] . \tag{2.76}$$

To obtain the moments of z, the same method is employed but it is found convenient to work with the marginal moment generating function $\phi(0, t_2)$ rather than $\phi(t_1, t_2)$. The results are

$$E(z) = \beta^2 S_{xx} + \nu\sigma^2 \tag{2.77}$$

$$E(z^2) = 4\sigma^2 \beta^2 S_{xx} + 2\nu\sigma^4 + (Ez)^2 \tag{2.78}$$

$$E(z^3) = 24\sigma^4 \beta^2 S_{xx} + 8\nu\sigma^6 + E(z)[8\sigma^2 \beta^2 S_{xx} + 4\nu\sigma^4] + E(z^2)\, E(z) \tag{2.79}$$

$$E(z^4) = 192\sigma^6 \beta^2 S_{xx} + 48\nu\sigma^8 + E(z)[72\sigma^4 \beta^2 S_{xx} + 24\nu\sigma^6] + E(z^2)[12\sigma^2 \beta^2 S_{xx} + 6\nu\sigma^4] + E(z^3)\, E(z) \,. \tag{2.80}$$

The first moment of z can now be substituted into (2.78) yielding $E(z^2)$ in terms of β, σ, N, and S_{xx}. Similarly, E(z) and $E(z^2)$ can then be substituted in (2.79) yielding $E(z^3)$ in terms of β, σ, N, and S_{xx}. Continuing in this fashion, all the necessary moments can be expressed as functions of β, σ, N, and S_{xx}. Substitution of these quantities into (2.59) and (2.61) yields, after considerable algebraic manipulation, the following approximations for the first and second moments of d denoted by $E_A(d)$ and $E_A(d^2)$ respectively:

$$\mu_z^3 E_A(d) = (\nu^2\sigma^4 + \beta^2 S_{xx}^2 + 2N\sigma^2 \beta^2 S_{xx})\, \beta S_{xx} \tag{2.81}$$

$$\mu_z^6 E_A(d^2) = \beta^{10} S_{xx}^6 + (4\nu + 5)\,\sigma^2\,\beta^8 S_{xx}^5$$

$$+ (6\nu^2 + 10\nu - 8)\,\sigma^4\,\beta^6 S_{xx}^4$$

$$+ (4\nu^3 + 6\nu^2 - 10\nu + 104)\,\sigma^6\,\beta^4 S_{xx}^3$$

$$+ (\nu^4 + 2\nu^3 + 8\nu^2 - 32\nu)\,\sigma^8\,\beta^2 S_{xx}^2$$

$$+ (\nu^4 - 2\nu^3 + 8\nu^2)\,\sigma^{10}\,S_{xx}\,. \qquad (2.82)$$

To access the worth of the above approximations, integrals (2.26) and (2.27) for the exact moments of d were evaluated numerically for a wide range of values of β and σ and two values each of N and S_{xx}. The double numerical integration was performed on the IBM 360 digital computer at Virginia Polytechnic Institute Computing Center using the bivariate analog to the parabolic rule; the accuracy was held to at least three significant digits. Equations (2.81) and (2.82) were then evaluated for the same values of the parameters and the percent deviation from the exact moments noted. Some representative results for the first moment of d are shown in Table I and those for the second moment are shown in Table II. The values of β are not shown in the tables but can be obtained using the relationship $\beta = \sigma\gamma$ (only positive values of β were considered). The results indicate that the approximation for both moments are very close to the exact moments provided $\gamma \geq 5$, the deviation of $E_A(d)$ from $E_E(d)$ being zero percent to three significant digits and the deviation of $E_A(d^2)$ from $E_E(d^2)$ being well within one half percent. It is also noted that as N increases, the worth of the approximations increases. For example, for $N = 15$ the deviation is

TABLE I

Comparison of $E_A(d)$ and $E_E(d)$

N	S_{xx}	σ	$\gamma=\|\beta/\sigma\|$	$E_E(d)$	$E_A(d)$	Percent deviation from $E_E(d)$
6	6	1.0	.1	.119	.119	.00
		.5	1.0	1.21	1.20	-.83
		.2	2.5	1.84	1.84	.00
		.2	5.0	.980	.980	.00
		.5	10.0	.199	.199	.00
6	3	1.0	.1	.0597	.0598	-.17
		.5	1.0	.819	.820	.12
		.2	2.5	1.70	1.68	-1.18
		.2	5.0	.960	.960	.00
		.5	10.0	.198	.198	.00
15	14	1.0	.1	.0991	.0992	.10
		.5	1.0	1.04	1.04	.00
		.2	2.5	1.75	1.75	.00
		.2	5.0	.967	.967	.00
		.5	10.0	.198	.198	.00
15	7	1.0	.1	.0498	.0498	.00
		.5	1.0	.688	.688	.00
		.2	2.5	1.56	1.56	.00
		.2	5.0	.935	.935	.00
		.5	10.0	.197	.197	.00

TABLE II

Comparison of $E_A(d^2)$ and $E_E(d^2)$

N	S_{xx}	σ	$\gamma=\|\beta/\sigma\|$	$E_E(d^2)$	$E_A(d^2)$	Percent deviation from $E_E(d^2)$
6	6	1.0	.1	.403	.226	-43.92
		.5	1.0	1.68	1.53	-8.93
		.2	2.5	3.49	3.43	-1.72
		.2	5.0	.967	.966	-.10
		.5	10.0	.0397	.0397	.00
6	3	1.0	.1	.200	.112	-44.00
		.5	1.0	.916	.783	-14.52
		.2	2.5	3.02	2.90	-3.97
		.2	5.0	.934	.930	-.43
		.5	10.0	.0393	.0393	.00
15	14	1.0	.1	.0905	.0719	-20.55
		.5	1.0	1.12	1.10	-1.78
		.2	2.5	3.11	3.10	-.32
		.2	5.0	.937	.937	.00
		.5	10.0	.0393	.0393	.00
15	7	1.0	.1	.0435	.0340	-21.84
		.5	1.0	.519	.508	-2.16
		.2	2.5	2.46	2.45	-.41
		.2	5.0	.879	.878	-.11
		.5	10.0	.0387	.0387	.00

within one half percent for values of γ as low as 2.5. The effect of S_{xx} on the approximations is that the approximations are closer to the exact moments when S_{xx} takes on its maximum value but the loss in accuracy for smaller values is very slight.

The results of this section will drastically reduce the computational labor in minimizing J_{IL} with respect to the design variables.

(2.7) Optimal Values of the Design Variables

All the machinery is now at hand to minimize J_{IL} with respect to the design variables $\bar{x}$ and $s = S_{xx}$ and hence optimize designs using the minimum average mean squared error criterion. Recall that the form of $2J_{IL}$ is given by

$$2J_{IL} = 2/3 + 2\bar{x}^2 - E(d)[4\bar{x}^2 \beta + (4/3)\beta]$$

$$+ E(d^2)[2\sigma^2(1+1/N) + (2/3)\beta^2 + 2\bar{x}^2\beta^2] . \qquad (2.83)$$

The terms in (2.83) involving $\bar{x}$ are

$$2\bar{x}^2[1 - 2\beta E(d) + \beta^2 E(d^2)]$$

so $2J_{IL}$ can be written as

$$2J_{IL} = 2/3 - E(d)\ (4/3)\beta + E(d^2)[2\sigma^2(1+1/N) + (2/3)\beta^2]$$

$$+ 2\bar{x}^2 E(1-\beta d)^2 . \qquad (2.84)$$

It is clear upon examining (2.84) that $2J_{IL}$ is minimized when $\bar{x} = 0$ regardless of the value of s which appears in the moments of d. Hence, the first requirement that must be satisfied in order that (2.83) be minimized is that $\bar{x} = 0$. While it is true that for $N \geq 4$, there exists an infinite number of designs for which $\bar{x} = 0$, the obvious choice is to select a symmetrical design with respect to the center of the region of interest R. Symmetrical designs have the property that not only produces $\bar{x} = 0$ but that also produces any odd design moment equal to zero, i.e., any quantity of the form $\Sigma\, x_i^r/N = 0$ when r is odd. This will be important in later chapters in this thesis.

Setting $\bar{x} = 0$ in (2.83) reduces $2J_{IL}$ to

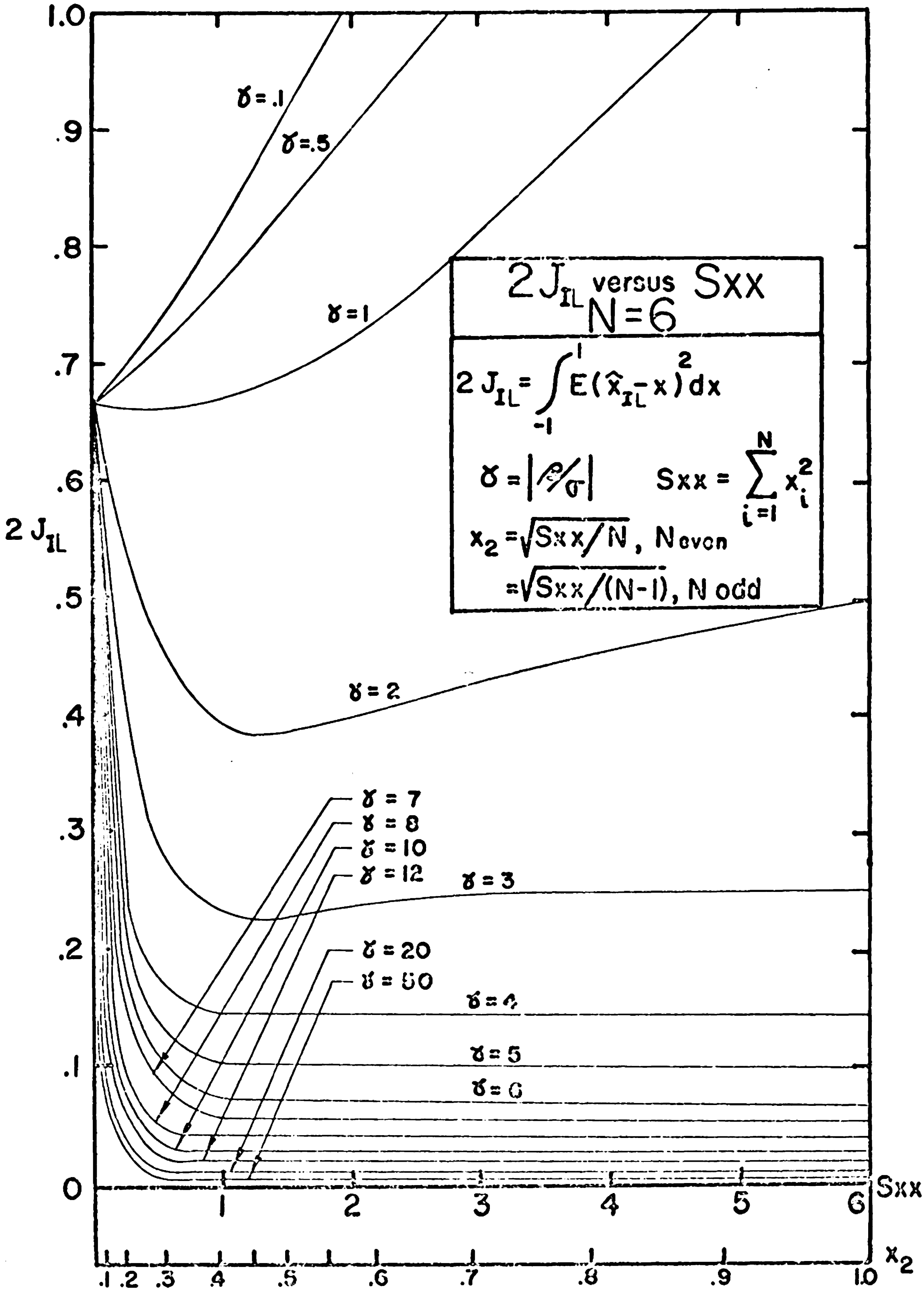
$$2J_{IL} = 2/3 - E(d)\ (4/3)\ \beta + E(d^2)[2\sigma^2(1+1/N) + (2/3)\ \beta^2]\ . \quad (2.85)$$

It is now necessary to minimize (2.85) with respect to the design variable s which appears in the moments of d. Consider first differentiating (2.85) with respect to s, setting the result equal to zero, and solving for s by a numerical procedure such as the Newton-Raphson procedure. Each iteration would require four double numerical integrations which would be extremely costly, and one could not be certain that the resulting solution, if any, actually minimized $2J_{IL}$. Also, if this procedure did produce a value of s which minimized $2J_{IL}$, it would yield no insight into the effect of moving s away from its point of minimum. In light of this, it was felt that the most efficient method of accomplishing minimization would be to evaluate (2.85) for selected values of N, s, and $\gamma = |\beta/\sigma|$ and display the results graphically, using the approximation for the moments of d for large values of γ and numerical integration for small values of γ. It was decided to use five values of N, N = 6, 10, 15, 25, 50 and fourteen values of γ, γ = .1, .5, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 20, 50 and for each (γ, N) combination, evaluate $2J_{IL}$ for as many values of s as necessary to obtain a "good" plot. The number of points for each curve ranged between fourteen and twenty. As in the previous numerical integration used to compare $E_A(d^r)$ with $E_E(d^r)$, accuracy was held to at least three significant digits. While it was felt that the approximation for the moments of d could be used safely for $\gamma \geq 5$ in optimizing designs, the approximation was restricted to values of $\gamma \geq 10$ because the resulting minimum value of $2J_{IL}$ would eventually be compared with the minimum 2J of the Classical estimator and in the event of small differences, the half percent (or less) deviation from the exact moments could be important.

The results are displayed in Figures 1-5 where $2J_{IL}$ is plotted as a function of S_{xx} for fixed values of γ and N and show that unlike the Classical estimator, the optimal design using the Inverse estimator does not always occur where s is maximum. The value of s which provides a minimum J_{IL} is very much dependent on $\gamma = |\beta/\sigma|$ and N. For large values of γ, the minimum J_{IL} does occur where s is maximum, but as γ decreases, the minimum occurs at decreasingly smaller values of s.

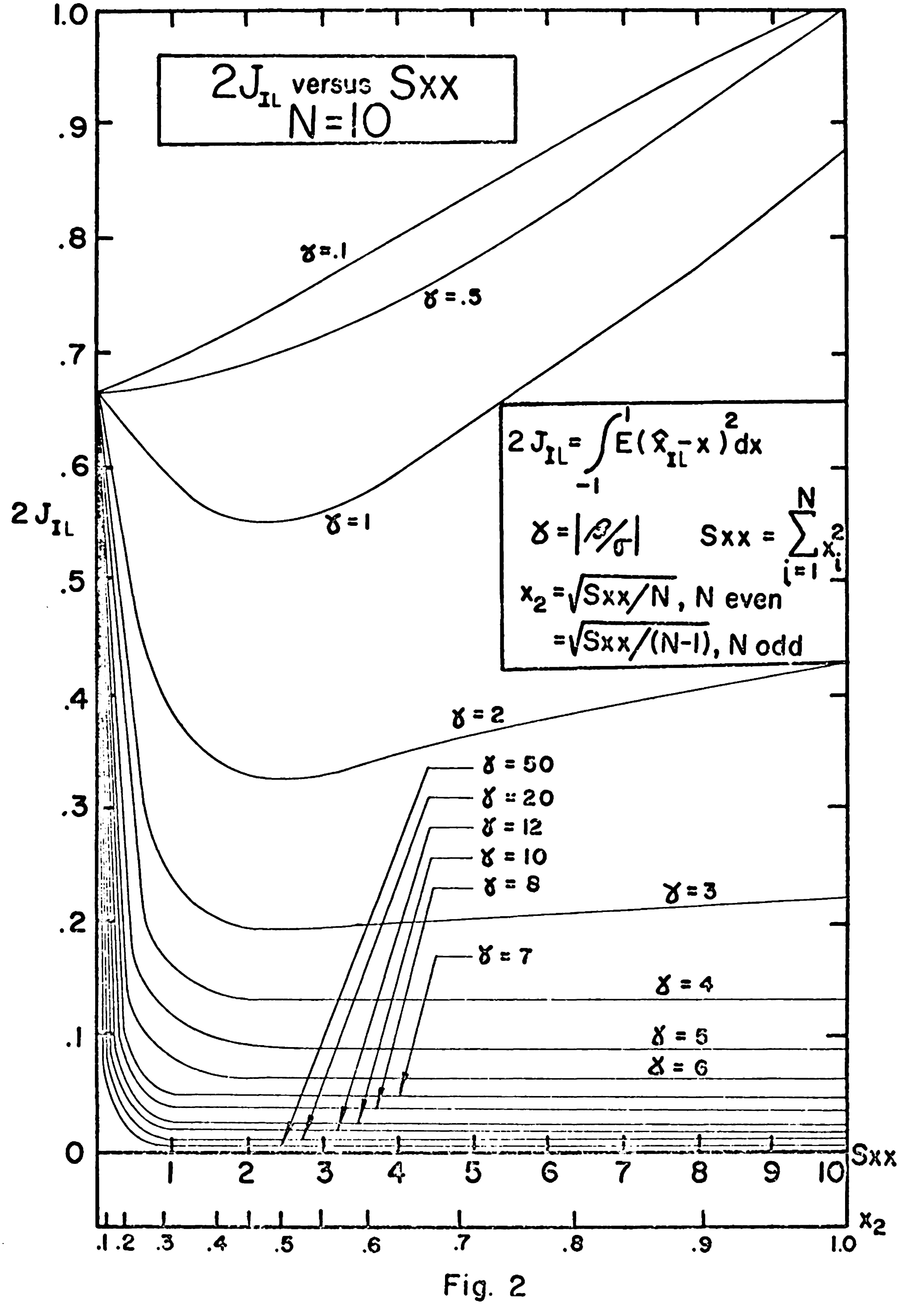
If experimental plans are restricted to "two-point" designs when N is even with N/2 observations at $x_1 = -x_2$ and N/2 observations at x_2 and to "three-point" designs when N is odd with (N-1)/2 observations at $x_1 = -x_2$, one observation at $x = 0$, and (N-1)/2 observations at x_2 the value of x_2, x_2(min), corresponding to the value of s at which J_{IL} is minimized, s(min), is obtained by solving s = s(min) for x_2, namely, solving $\Sigma x_i^2 = s(\min)$ for x_2. The results are simply

$$x_2(\min) = \sqrt{s(\min)/N}, \text{ N even} \tag{2.86}$$



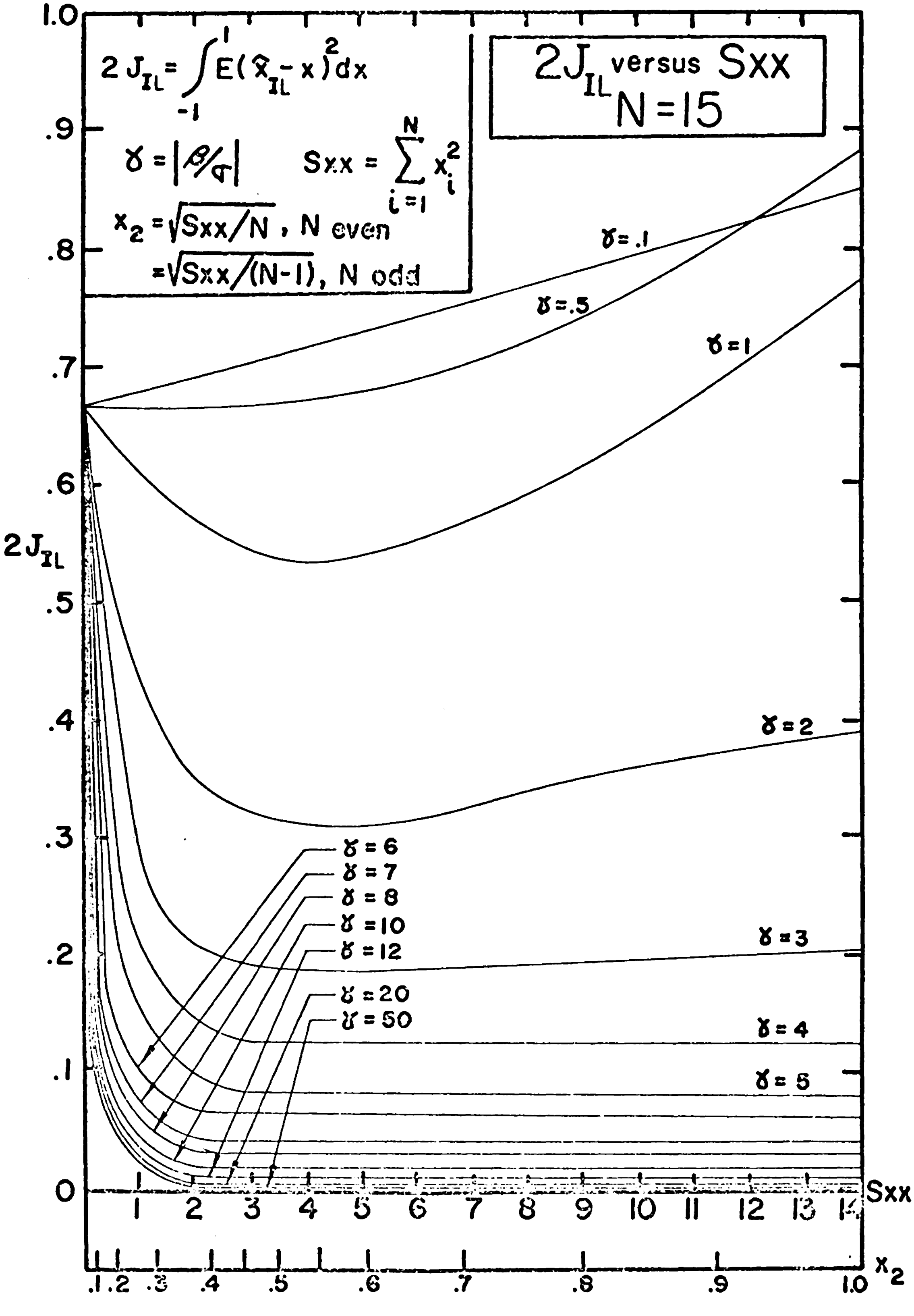
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
2 J_IL
γ = .1
γ = .5
γ = 1
γ = 2
γ = 3
γ = 4
γ = 5
γ = 6
γ = 7
γ = 8
γ = 10
γ = 12
γ = 20
γ = 50
2 J_IL versus Sxx
N = 6
2 J_IL = ∫_{-1}^{1} E(x̂_IL − x)^2 dx
γ = |β/σ|
Sxx = Σ_{i=1}^{N} x_i^2
x_2 = √(Sxx/N), N even
= √(Sxx/(N−1)), N odd
1 2 3 4 5 6
Sxx
.1 .2 .3 .4 .5 .6 .7 .8 .9 1.0
x_2

Fig. 1



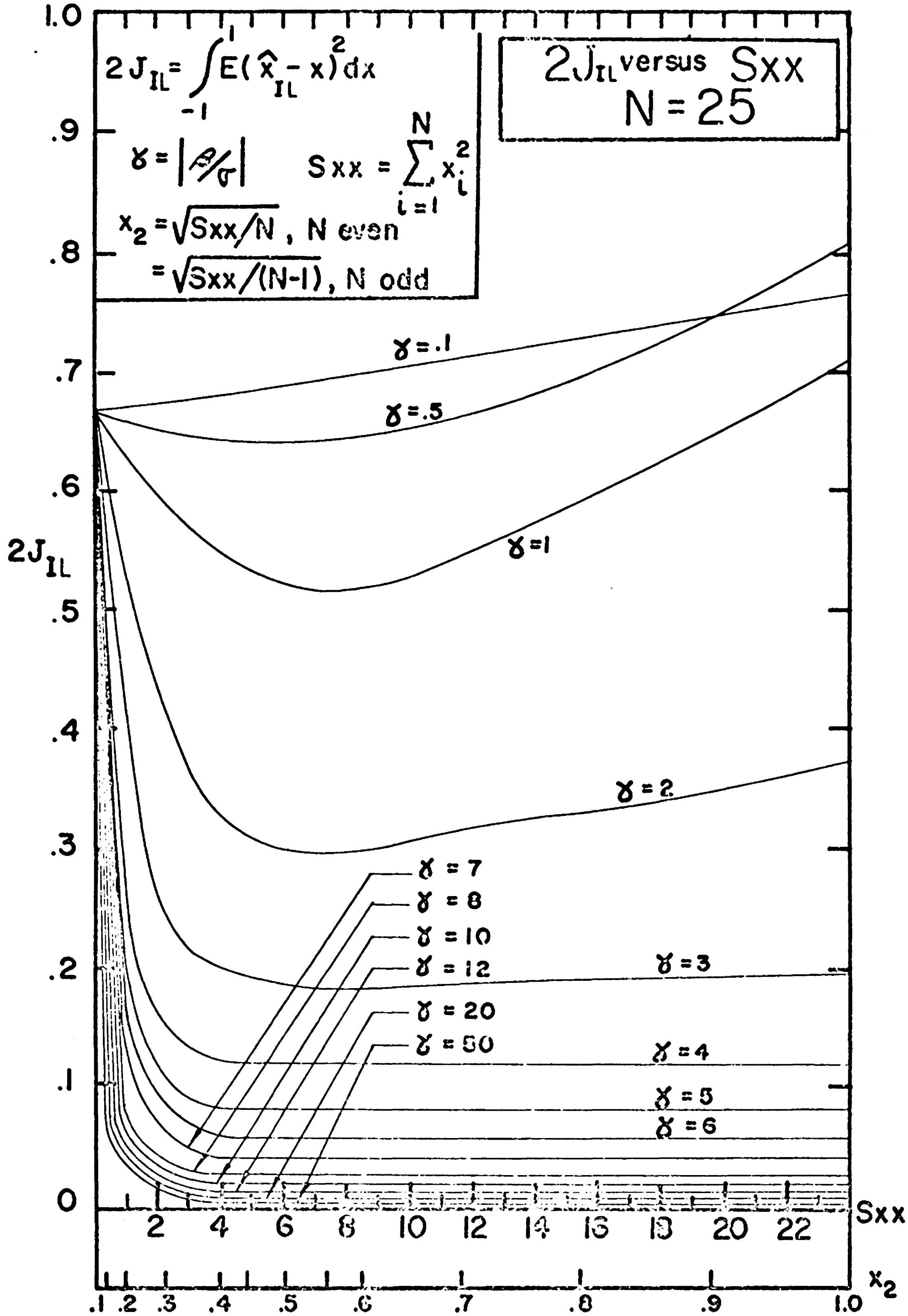
$2J_{IL}$ versus Sxx N=10
$2J_{IL} = \int_{-1}^{1} E(\hat{x}_{IL} - x)^2 dx$
$\gamma = |\beta/\sigma|$
$Sxx = \sum_{i=1}^{N} x_i^2$
$x_2 = \sqrt{Sxx/N}$, N even
$= \sqrt{Sxx/(N-1)}$, N odd
$\gamma = .1$
$\gamma = .5$
$\gamma = 1$
$\gamma = 2$
$\gamma = 3$
$\gamma = 4$
$\gamma = 5$
$\gamma = 6$
$\gamma = 7$
$\gamma = 8$
$\gamma = 10$
$\gamma = 12$
$\gamma = 20$
$\gamma = 50$
$2J_{IL}$
Sxx
x_2

Fig. 2



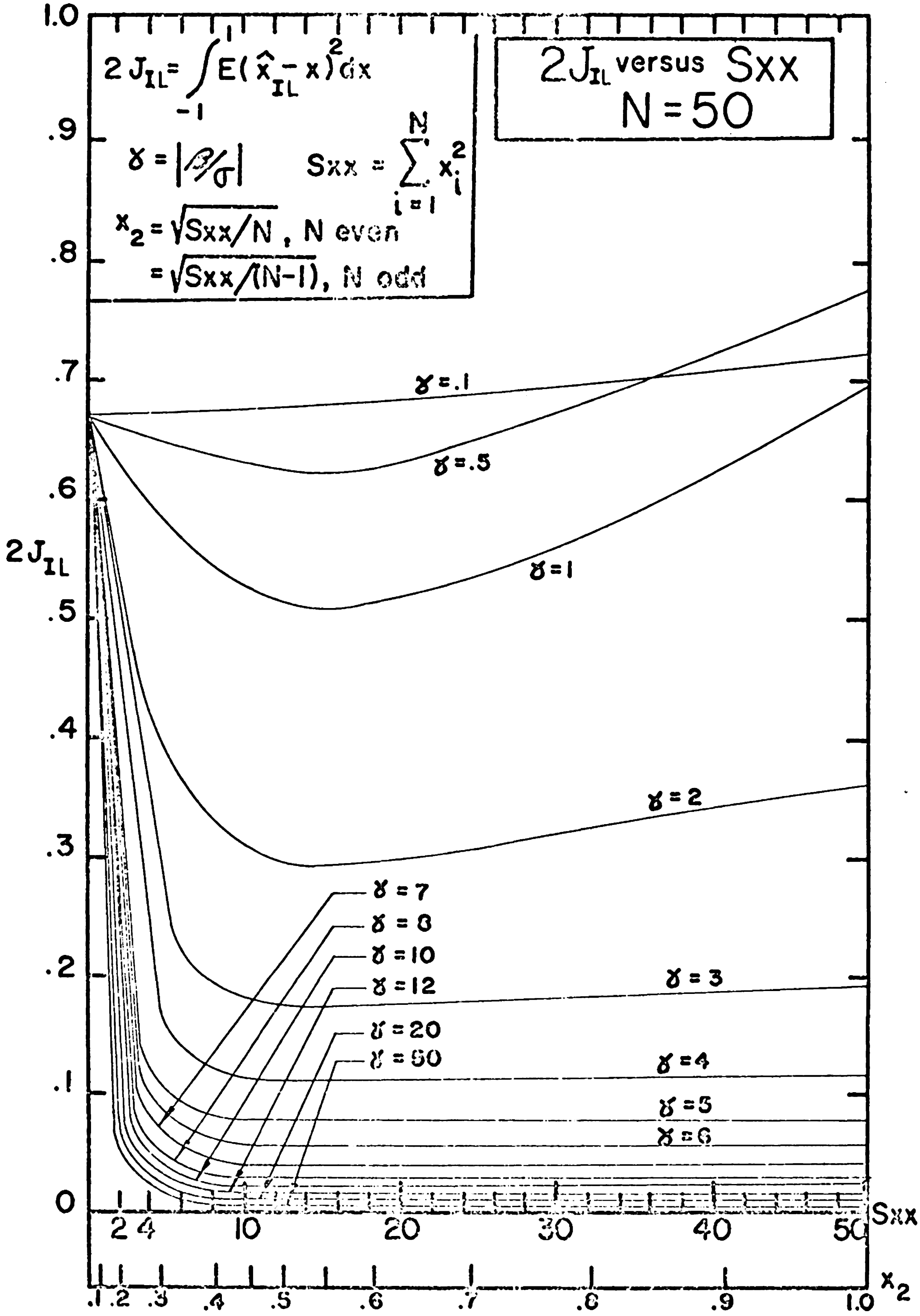
$2 J_{IL} = \int_{-1}^{1} E(\hat{x}_{IL} - x)^2 dx$
$\gamma = |\beta/\sigma|$
$Sxx = \sum_{i=1}^{N} x_i^2$
$x_2 = \sqrt{Sxx/N}$, N even
$= \sqrt{Sxx/(N-1)}$, N odd
$2J_{IL}$ versus Sxx
N=15
$2J_{IL}$
Sxx
x_2
γ=.1
γ=.5
γ=1
γ=2
γ=3
γ=4
γ=5
γ=6
γ=7
γ=8
γ=10
γ=12
γ=20
γ=50

Fig. 3



$2J_{IL}$ versus Sxx
N = 25
$2J_{IL} = \int_{-1}^{1} E(\hat{x}_{IL} - x)^2 dx$
$\gamma = |\beta/\sigma|$
$Sxx = \sum_{i=1}^{N} x_i^2$
$x_2 = \sqrt{Sxx/N}$, N even
$= \sqrt{Sxx/(N-1)}$, N odd
$2J_{IL}$
γ=.1
γ=.5
γ=1
γ=2
γ=3
γ=4
γ=5
γ=6
γ=7
γ=8
γ=10
γ=12
γ=20
γ=50
Sxx
x_2

Fig. 4



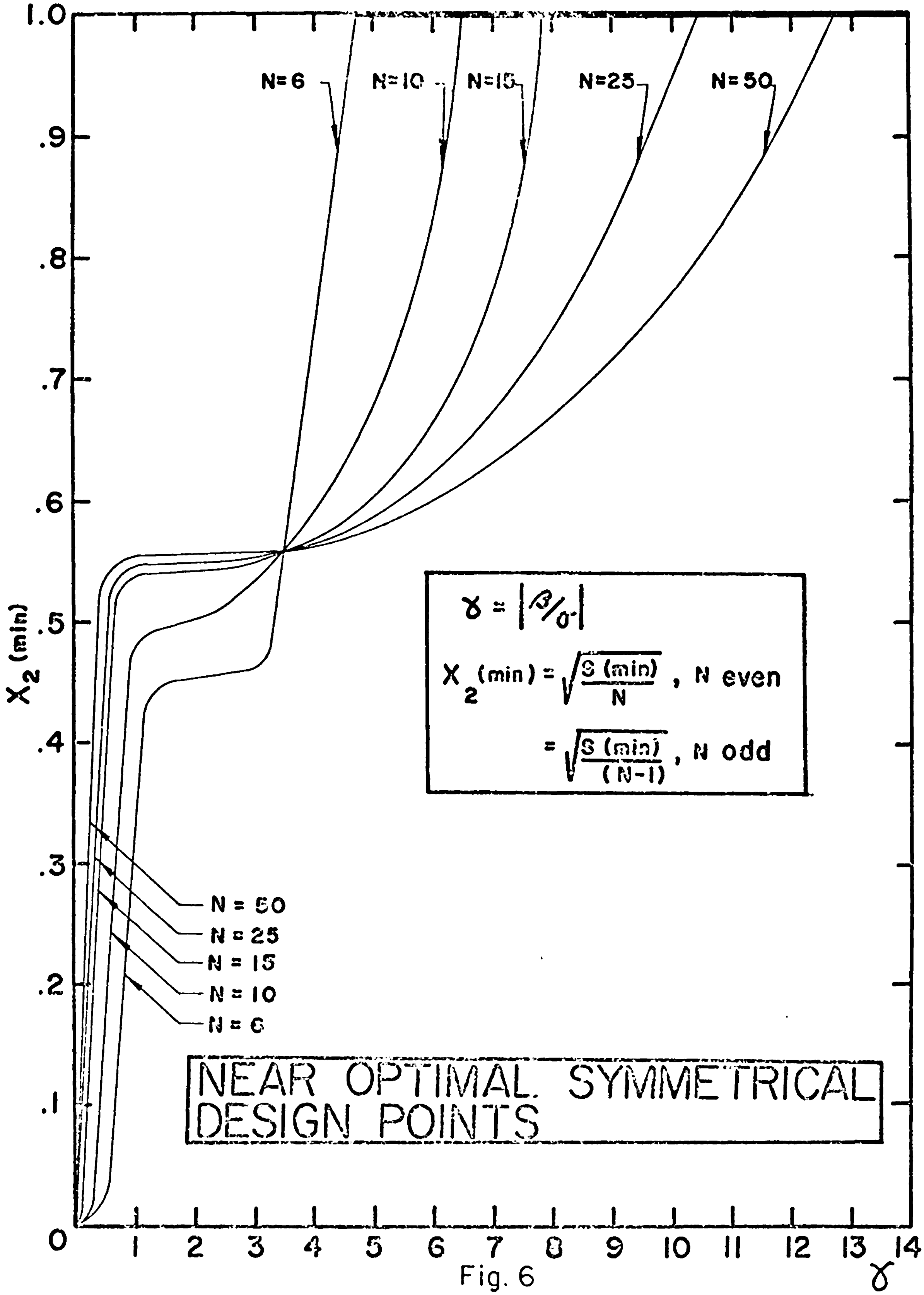
2J_IL versus Sxx
N = 50
2 J_IL = ∫_{-1}^{1} E(x̂_IL − x)² dx
γ = |β/σ|
Sxx = Σ_{i=1}^{N} x_i²
x_2 = √(Sxx/N), N even
= √(Sxx/(N−1)), N odd
2 J_IL
γ = .1
γ = .5
γ = 1
γ = 2
γ = 3
γ = 4
γ = 5
γ = 6
γ = 7
γ = 8
γ = 10
γ = 12
γ = 20
γ = 50
Sxx
x_2

Fig. 5

$$x_2(\min) = \sqrt{s(\min)/(N-1)},\ N \text{ odd}. \tag{2.87}$$

The supplementary x_2 scale on Figures 1-5 enables one to easily determine $x_2(\min)$ for each (γ, N) combination. Values of $x_2(\min)$ thus obtained were used to construct Figure 6, a graph of "Near Optimal Designs" showing values of x_2 for "two-point" (N even) and "three-point" (N odd) designs which provide a minimum J_{IL}. The adjective "near" is used since the values of $x_2(\min)$ are not mathematical minima in the true sense of the word but were obtained from the calculations used in constructing Figures 1-5 which were based on the accuracy of three significant digits.

It should be noted that the optimal designs derived above depend on knowledge of the unknown quantity $\gamma = |\beta/\sigma|$. This implies that optimal designs cannot be attained in a practical situation. Hence, some rule is necessary for constructing designs for the Inverse estimator assuming no knowledge of γ. This discussion will be deferred until the next chapter where the minimum J's of the Inverse and Classical estimators are compared.



N=6
N=10
N=15
N=25
N=50
$\gamma = |\beta/\sigma|$
$X_2(\min) = \sqrt{\frac{S(\min)}{N}}$, N even
$= \sqrt{\frac{S(\min)}{(N-1)}}$, N odd
N = 50
N = 25
N = 15
N = 10
N = 6
NEAR OPTIMAL SYMMETRICAL DESIGN POINTS
X_2 (min)
γ

Fig. 6

CHAPTER III

COMPARISON OF AVERAGE MSE'S - TRUE MODEL LINEAR

In this chapter, an expression for the average or integrated mean squared error will be developed for the truncated Classical estimator. Then, using the design which minimizes this criterion, the resulting average MSE will be compared with the minimum average MSE for the Inverse estimator. Also, since in some cases the optimal design for the Inverse estimator depends on unknown parameters and cannot be realistically used without a priori information on these parameters, a further comparison will be made between estimators for designs which are always attainable.

First, an expression is needed for

$$J_{CL} = \int_R E(\hat{x}_{CL} - x)^2 \, dx \Big/ \int_R dx \tag{3.1}$$

where the subscript CL on J and $\hat{x}$ indicate that the Classical estimator is under consideration and that the true model is linear as given by equation (1.2). The range R on x will be taken as [-1, 1]. Recall from equation (1.4) that

$$\hat{x}_{CL} = \bar{x} + (y - \bar{y})/b \tag{3.2}$$

where b is the least squares estimator for β in (1.2), i.e., $b = S_{xy}/S_{xx}$. This is identical to the Inverse estimator, $\hat{x}_{IL}$, if we replace d with 1/b. Furthermore, since b is independent of both y and $\bar{y}$, the derivation of J_{CL} is identical with the derivation of J_{IL} with 1/b replacing d. Hence, one can immediately write

$$J_{CL} = \{(2/3) + 2\bar{x}^2 - E(1/b)\ [4\bar{x}^2\ \beta + (4/3)\ \beta]$$

$$+ E(1/b^2)\ [2\sigma^2(1+1/N) + (2/3)\ \beta^2 + 2\bar{x}^2\ \beta^2]\}/2. \qquad (3.3)$$

It is meaningless to consider the above, however, because Ott has shown that the moments of $1/b$ are infinite. Since it is unrealistic to consider values of b which are more than four standard deviations from β, Ott and Myers truncated the density of b at $\delta_1 = \beta - 4\sigma_b$ and $\delta_2 = \beta + 4\sigma_b$ with $\delta_1 > 0$ for $\beta > 0$ and thereby obtained finite results for $E(1/b^r)$. Rewriting (3.3) in terms of the finite moments, one obtains

$$J_{CL} = \{(2/3) + 2\bar{x}^2 - E_T(1/b)\ [4\bar{x}^2\ \beta + (4/3)\ \beta]$$

$$+ E_T(1/b^2)\ [2\sigma^2(1+1/N) + (2/3)\ \beta^2 + 2\bar{x}^2\ \beta^2]\}/2 \qquad (3.4)$$

where the subscript T indicates that the expectations are over the truncated density of b, that is,

$$E_T(1/b^r) = (1/k) \int_{\delta_1}^{\delta_2} (1/b^r)\ f(b)\ db$$

$$= 1/[(2\pi\sigma_b^2)^{1/2}\ k] \int_{\delta_1}^{\delta_2} (1/b^r)\ e^{-(b-\beta)^2/(2\sigma_b^2)}\ db \qquad (3.5)$$

under the normality assumption on the y_i. The quantity σ_b^2 above is well known to equal σ^2/S_{xx} and k is the truncation constant which is given by

$$k = \int_{\delta_1}^{\delta_2} f(b)\ db$$

$$= P[\delta_1 = \beta - 4\sigma_b < b < \delta_2 = \beta + 4\sigma_b]$$

$$= 2\Phi(4) - 1 = .9999366576\ .$$

The same truncation points can be used for $\beta < 0$ with the provision that $\delta_2 < 0$.

It is observed that the moments $E_T(1/b^r)$ are functions of σ, β, and $s = S_{xx}$ so that J_{CL} is a function of σ, β, $\bar{x}$, s, and N. As in the Inverse case, it is a simple matter to show that β and σ in (3.4) can be considered as the single quantity $\gamma = |\beta/\sigma|$ so that J_{CL} is a function of γ, $\bar{x}$, s, and N. Letting $\beta = \sigma\gamma'$ in (3.4), it is observed that σ appears only as a coefficient of $E_T(1/b)$ and σ^2 appears only as a coefficient of $E_T(1/b^2)$. Using the transformation $w = b/\sigma$ in (3.5) for r = 1 and letting $\beta = \sigma\gamma'$ results in

$$\sigma\ E_T(1/b) = \frac{s^{1/2}}{k\sqrt{2\pi}} \int_{\delta_1'}^{\delta_2'} \frac{1}{w}\ e^{-(s/2)\ (w-\gamma')^2}\ dw \tag{3.6}$$

for $\beta > 0$ where $\delta_1' = \gamma' - 4/s^{1/2}$ and $\delta_2' = \gamma' + 4/s^{1/2}$ so $\sigma\ E_T(1/b)$ is a function of only s and γ'. Using the same transformation in (3.5) for r = 2 results in

$$\sigma^2\ E_T(1/b^2) = \frac{s^{1/2}}{k\sqrt{2\pi}} \int_{\delta_1'}^{\delta_2'} \frac{1}{w^2}\ e^{-(s/2)\ (w-\gamma')^2}\ dw \tag{3.7}$$

and $\sigma^2 E_T(1/b^2)$ is a function of only s and γ'. The effect of a change in the sign of β is merely a change in sign of $E_T(1/b)$ and in identical results for $E_T(1/b^2)$. Hence, $\sigma E_T(1/b)$ and $\sigma^2(E_T(1/b^2)$ are unchanged so J_{CL} is a function of $\bar{x}$, s, N, and γ, the same as J_{IL}.

Since J_{CL} is identical to J_{IL} if one replaces the moments of d with the moments $E_T(1/b^r)$, $2J_{CL}$ can be written as in (2.84) as

$$2J_{CL} = (2/3) - E_T(1/b)\ (4/3)\ \beta + E_T(1/b^2)\ [2\sigma^2(1+1/N)$$
$$+ (2/3)\ \beta^2] + 2\bar{x}^2\ E_T(1-\beta/b)^2\ . \qquad (3.8)$$

As in the case of the Inverse estimator, it is clear upon examining (3.8) that $2J_{CL}$ is minimized when $\bar{x} = 0$ regardless of the value of s which appears in $E_T(1/b^r)$, $r = 1, 2$. It will be convenient, as before, to restrict the class of designs for which $\bar{x} = 0$ to symmetrical designs. Setting $\bar{x} = 0$ in (3.8), one obtains

$$2J_{CL} = (2/3) - E_T(1/b)\ (2/3)\ \beta$$
$$+ E_T(1/b^2)\ [2\sigma^2(1+1/N) + (2/3)\ \beta^2]\ . \qquad (3.9)$$

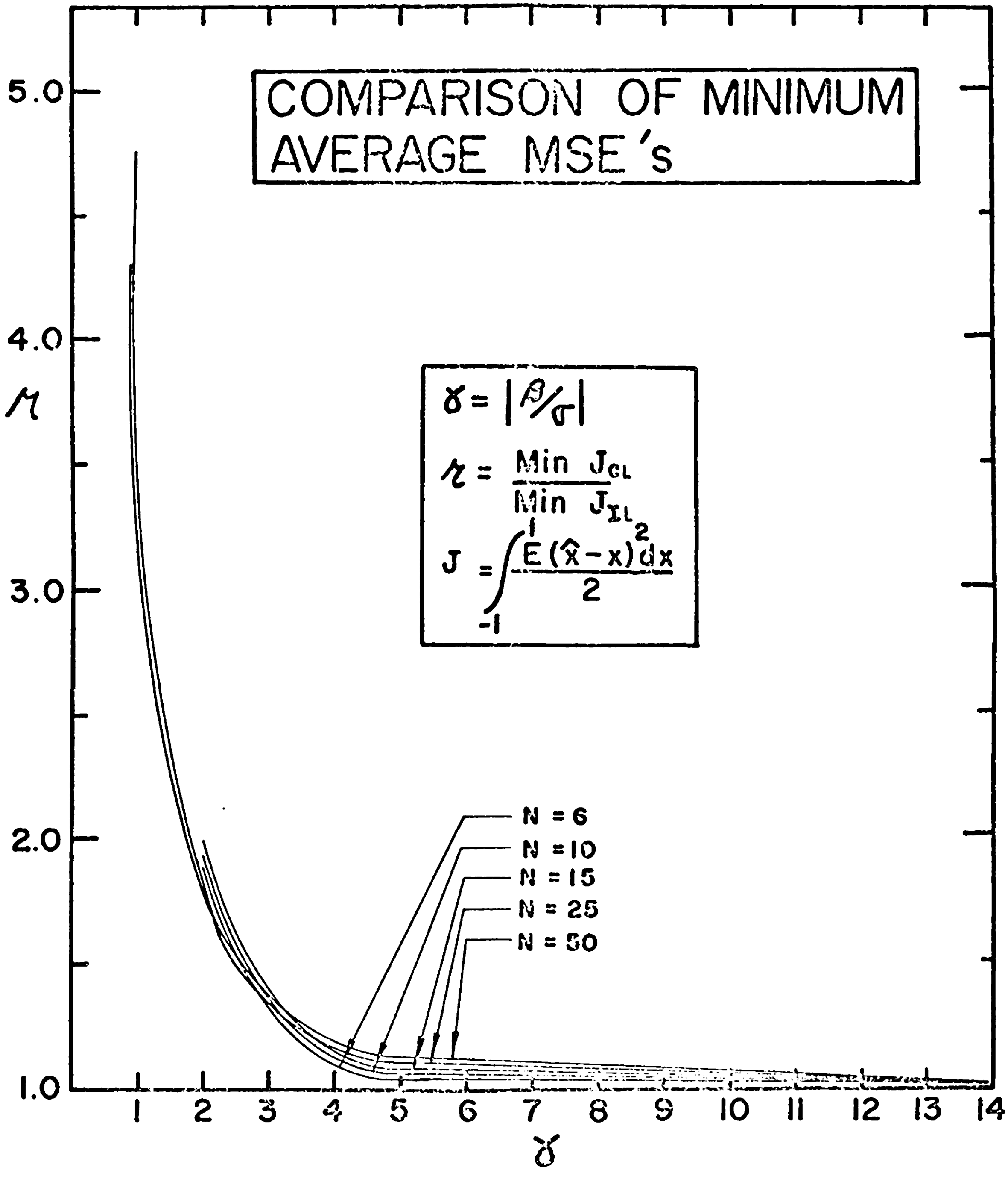
Ott and Myers, using an approximation to $E_T(1/b^r)$, $r = 1, 2$, have shown that $2J_{CL}$ in (3.9) can be minimized by always taking s maximum. For symmetrical designs (N even), s is maximized by taking N/2 observations at $x_1 = -1$ and N/2 observations at $x_2 = +1$ and for symmetrical designs (N odd), s is maximized by taking (N-1)/2 observations at $x_1 = -1$, one observation at $x = 0$, and (N-1)/2 observations at $x_2 = +1$. Hence, the minimum value of $2J_{CL}$ can easily be obtained for any specified values of γ and N by merely computing $2J_{CL}$ in (3.9) with

$s = N$ for N even or $s = N - 1$ for N odd. The moments, $E_T(1/b^r)$, can be evaluated by using the approximation developed by Ott and Myers or for better precision, by numerical integration.

Since the minimum of both J_{CL} and J_{IL} are functions of γ and N, they can easily be compared for selected values of these quantities. However, for any comparison, one must take the truncation into consideration. The requirement that $\delta_1 > 0$ in terms of the transformed parameter γ requires that $\delta_1' > 0$ which implies that $s > 16/\gamma^2$. If, for example $\gamma = 1$, one would necessarily need $s > 16$ or equivalently $N \geq 18$ or the design would be meaningless in terms of the average MSE criterion since J_{CL} would not exist. With this restriction in mind, the comparison of the minimum J's is confined to values of $\gamma \geq 1$, smaller values of γ requiring unrealistically large values of N. The minimum J_{CL} was computed for $\gamma = 1(1)\ 8, 10, 12, 20, 50$ and $N = 6, 10, 15, 25, 50$ where the (γ, N) combinations were confined due to the truncation restriction. To obtain accuracy comparable with that for the Inverse estimator, the moments $E_T(1/b^r)$ were evaluated by numerical integration maintaining at least three significant digit accuracy. Corresponding values of the minimum J_{IL} were then taken from the calculations from which Figures 1-5 were constructed. The measure of comparison was

$$r = \text{Min } J_{CL}/\text{Min } J_{IL} . \tag{3.10}$$

Figure 7 shows plots of r as a function of γ for the selected values of N. As γ increases, r converges to 1.000. In all cases except $N = 50$, the value of r at $\gamma = 50$ is 1.000. For $N = 50$, r is equal



COMPARISON OF MINIMUM AVERAGE MSE's
$\gamma = |\beta/\sigma|$
$r = \dfrac{\text{Min } J_{CL}}{\text{Min } J_{IL}}$
$J = \int_{-1}^{1} \dfrac{E(\hat{x}-x)^2 dx}{2}$
N = 6
N = 10
N = 15
N = 25
N = 50
5.0
4.0
3.0
2.0
1.0
r
1
2
3
4
5
6
7
8
9
10
11
12
13
14
γ

Fig. 7

to 1.001 at $\gamma = 50$. These results show conclusively that the Inverse estimator is better than the truncated Classical estimator from an average mean squared error point of view for the range of values of the parameters studied which exceeds the practical range. They also show that the Inverse estimator becomes increasingly superior as γ decreases. This was expected in view of Krutchkoff's empirical results. However, the extremely large ratios of the average squared errors exhibited by Krutchkoff by Monte Carlo sampling for small values of γ and small values of N can never be realized using the truncation in this chapter since for small γ and small N, the mean squared error and hence the average mean squared error for this truncated Classical estimator simply does not exist. The largest value of r detected in this investigation was $r = 4.750$ at $\gamma = 1$ for $N = 25$. Of course, one could truncate at less than four standard deviations, say p standard deviations with $p < 4$, and the truncated Classical would exist for $s > p^2/\gamma^2$. This would enable one to consider smaller (γ, N) combinations, but in allowing p to become too small, one loses the practical aspects of truncation, i.e., the truncation assumption would become invalid. Krutchkoff truncated the density of b at .001 in absolute value for all values of β and σ_b.

When using the truncated Classical estimator, the optimal symmetrical design can always be used by merely taking N/2 observations at both $x_1 = -1$ and $x_2 = +1$ for N even and for N odd by taking $(N-1)/2$ observations at both $x_1 = -1$ and $x_2 = +1$ and one observation at $x = 0$. However, when one uses the Inverse estimator, the optimal design is a function of N and the unknown ratio γ so that the optimal design is not always useable. If one has a very good a priori estimate of γ, Figure 6 can be used to design optimal experiments. However, in most practical situations only a rough a priori estimate of γ is available at best, and it seems appropriate to provide some rule for designing experiments for the Inverse estimator assuming no knowledge of γ. Referring to the x_2 scale on the abscissa of Figures 1-5, if one were to use a symmetrical "two-point" (N even) or "three-point" (N odd) design with x_2 in the vicinity of .45 to .55 (depending on N), the resulting J_{IL} would depart very little from Min J_{IL}. It is well to point out, however, that without knowledge of γ, this procedure could lead to a design which would result in an average MSE greater than the minimum average MSE for the Classical which can always be attained if the true model is linear as shown in (1.2). While it would result in a design very near optimal for small γ, the increase in J_{IL} over Min J_{IL} for large γ and small N could be as much as 13.5%. Reference to Figure 7 reveals that for large values of γ, say $\gamma \geq 10$, r is very near 1.000. For example, for $N = 6$ and $\gamma = 10$, $r = 1.016$. Hence, the increase in J_{IL} for large γ would be sufficient to increase J_{IL} above Min J_{CL} which can always be attained. In fact, any value for x_2 other than +1 will yield a J_{IL} which is greater than Min J_{CL} for large enough γ. In view of this, without knowledge of γ the only "safe" procedure is to use the "end-point" design with $x_2 = +1$ if one is intent on obtaining a value of J_{IL} which is always less than or equal to Min J_{CL} for all values of γ. Figure 8 is a plot of r' as a function of γ for the selected values of N where r' is defined as

$$r' = \text{Min } J_{CL}/J_{IL}(x_2 = 1) \ . \tag{3.11}$$

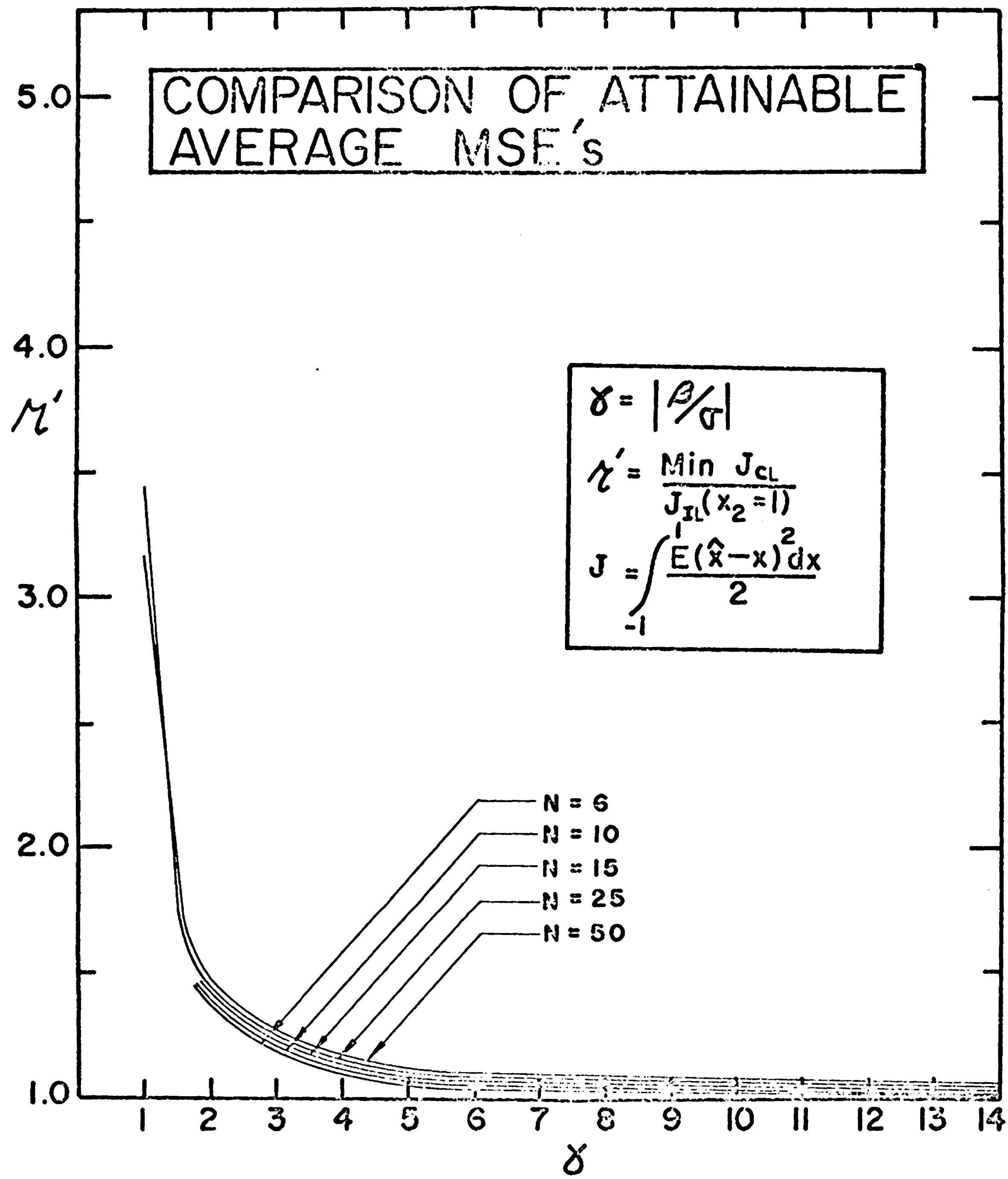

COMPARISON OF ATTAINABLE AVERAGE MSE's
$\gamma = |\beta/\sigma|$
$\eta' = \frac{\text{Min } J_{CL}}{J_{IL}(x_2=1)}$
$J = \int_{-1}^{1} \frac{E(\hat{x}-x)^2 dx}{2}$
N = 6
N = 10
N = 15
N = 25
N = 50
η'
γ
5.0
4.0
3.0
2.0
1.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14

Fig. 8

This figure reflects the decrease in the average mean squared error which can always be attained by using the Inverse estimator. A decrease cannot be guaranteed for all values of γ if any other design is used. However, the Inverse estimator is still fairly robust with respect to design for large γ so that if one uses a design other than the "end-point," say a "two-point" design with $x_2 \cong .5$ or a design with a uniform spread of points, he stands to decrease the average MSE by a greater percentage if γ is small than he stands to increase it if γ is large.

The results of the investigation reported in this chapter are simply that the minimum average MSE which can be attained using the Inverse estimator is smaller than the corresponding minimum average MSE which can be attained using the truncated Classical estimator, the degree of superiority increasing with decreasing γ. Furthermore, without a priori knowledge of γ, the "end-point" design for the Inverse estimator, while not optimal, still provides an average MSE which is smaller than the minimum average MSE for the truncated Classical estimator but the degree of superiority is not so pronounced for small values of γ. These results are substantiated by the original Monte Carlo study conducted by Krutchkoff (1967).

The above results hinge on the key assumption that the true model is linear as given in equation (1.2). While the experimenter may suspect that the relationship between x and y is approximately linear in some region R on x, more often than not, he is not certain that the relationship between x and y is exactly linear in this region. If the relationship is of higher order, say quadratic, then this chapter as well as Chapter II shed no light on the effect such a model misclassification has on the average MSE's or how the resulting average MSE's compare. These topics will be explored in Chapters IV and V.

CHAPTER IV

LINEAR APPROXIMATION, QUADRATIC MODEL - OPTIMAL DESIGNS

In this chapter, the assumption concerning the structure of the true model is that it is no longer linear over the range R on x as given by equation (1.2) but is rather quadratic of the form

$$y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \beta_2(x_i^2 - \overline{x^2}) + \varepsilon_i, \; i = 1,\ldots,N \qquad (4.1)$$

where $\overline{x^2} = \Sigma\, x_i^2/N$. The Inverse estimator will still be used to estimate x as if the true model were linear. This situation could arise in one of several ways. First, one may assume that the true model is linear when it is quadratic, that is, the experimenter misclassifies the model. Also, the experimenter may know that the true model is not linear but proceeds with a linear approximation for simplicity and/or ease in computation. In either case, the basic problem is the same, namely, if the experimenter designs for a linear model when the true model is quadratic, how far will the resulting J (the average MSE) deviate from the minimum J that could be attained with the correct model assumption and conversely? To answer these questions, one must first determine optimal experimental designs for the Inverse estimator when the true model is quadratic as given by equation (4.1). The first step is, as before, to find an expression for the average mean squared error.

(4.1) Derivation of J_{IQ} in Terms of the Moments of d

The same distributional assumptions about the ε_i will be used here as were used in Chapter II, namely, that the ε_i are distributed independently and identically normal with mean zero and variance σ^2. The only change is in the structure of the true model. Recall that the Inverse estimator for x is given by

$$\hat{x}_{IL} = \bar{x} + d(y-\bar{y})$$

so that

$$E_Q(\hat{x}_{IL}) = \bar{x} + E_Q[d(y-\bar{y})] \tag{4.2}$$

$$E_Q(\hat{x}^2_{IL}) = \bar{x}^2 + 2\bar{x}E_Q[d(y-\bar{y})] + E_Q[d^2(y-\bar{y})^2] \tag{4.3}$$

where the subscript Q indicates that the true model is quadratic. One can now show that $\bar{y}$ and S_{yy} are independent and that $\mathrm{Cov}(S_{xy}, \bar{y}) = 0$ as in Chapter II, implying that d and $\bar{y}$ are independent even in the presence of quadratic effect. Hence, (4.2) and (4.3) can be written

$$E_Q(\hat{x}_{IL}) = \bar{x} + E_Q(d)\ E_Q(y-\bar{y}) \tag{4.4}$$

$$E_Q(\hat{x}^2_{IL}) = \bar{x}^2 + 2\bar{x}E_Q(d)\ E_Q(y-\bar{y}) + E_Q(d^2)\ E_Q(y-\bar{y})^2\ . \tag{4.5}$$

The moments of $y-\bar{y}$ are found to be

$$E_Q(y-\bar{y}) = E[\beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2}) + (\varepsilon-\bar{\varepsilon})]$$

$$= \beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2}) \tag{4.6}$$

$$E_Q(y-\bar{y})^2 = \mathrm{Var}_Q(y-\bar{y}) + [E_Q(y-\bar{y})]^2$$

$$= \sigma^2(1+1/N) + [\beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2})]^2 \tag{4.7}$$

where $\bar{\varepsilon} = \Sigma\ \varepsilon_i/N$. One can now write the moments of $\hat{x}_{IL}$ as

$$E_Q(\hat{x}_{IL}) = \bar{x} + E_Q(d)[\beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2})] \tag{4.8}$$

$$E_Q(\hat{x}^2_{IL}) = \bar{x}^2 + 2\bar{x}E_Q(d)[\beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2})]$$
$$+ E_Q(d^2)[\sigma^2(1+1/N) + \{\beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2})\}^2] \tag{4.9}$$

and the mean squared error as

$$E_Q(\hat{x}_{IL}-x)^2 = (x-\bar{x})^2 - 2E_Q(d)\ (x-\bar{x})[\beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2})]$$
$$+ E_Q(d^2)[\sigma^2(1+1/N) + \{\beta_1(x-\bar{x}) + \beta_2(x^2 - \overline{x^2})\}^2] \ . \tag{4.10}$$

Expanding (4.10) and integrating term by term results in

$$2J_{IQ} = \int_{-1}^{1} E_Q(\hat{x}_{IL}-x)^2\ dx$$
$$= 2/3 + 2\bar{x}^2 - E_Q(d)[\beta_1(4/3 + 4\bar{x}^2) + \beta_2(4\bar{x}\ \overline{x^2} - 4\bar{x}/3)]$$
$$+ E_Q(d^2)[2\sigma^2(1+1/N) + \beta_1^2(2/3 + 2\bar{x}^2)$$
$$+ \beta_1\beta_2(4\bar{x}\ \overline{x^2} - 4\bar{x}/3) + \beta_2^2(2/5 - 4\overline{x^2}/3 + 2\overline{x^2}^2)] \tag{4.11}$$

where the subscript IQ on J indicates that the Inverse estimator is under consideration and that the true model is quadratic. For symmetrical designs, (4.11) reduces to

$$2J_{IQ} = 2/3 - (4/3)\beta_1\ E_Q(d) + E_Q(d^2)[2\sigma^2(1+1/N)$$
$$+ (2/3)\beta_1^2 + \beta_2^2(2/5 - 4\overline{x^2}/3 + 2\overline{x^2}^2)] \ . \tag{4.12}$$

Only symmetrical designs will be under consideration since it was shown in Chapter II that if the true model is linear, the first requirement that must be satisfied in order that J_{IL} be minimized is that $\bar{x} = 0$. This infinite class of designs for which $\bar{x} = 0$ was then further restricted to the class of symmetrical designs for reasons which will soon become evident. Most of the theory which follows will be developed first for the general case and then reduced to the symmetrical case. To proceed further requires expressions for the first two moments of d in the presence of a quadratic effect.

(4.2) The Moments of d in the Presence of a Quadratic Effect

It will be well to recall how the moments of d were obtained in the linear case. First, S_{yy} was decomposed into independent components as

$$S_{yy} = S_{xy}^2/S_{xx} + SS_e \tag{4.13}$$

where $SS_e = \Sigma (y_i - \hat{y}_i)^2 = \Sigma [y_i - \bar{y} - b(x_i - \bar{x})]^2$, the residual or error sum of squares from forward regression where $b = \hat{\beta} = S_{xy}/S_{xx}$. The ratio d was then written as

$$d = S_{xy}/S_{yy} = u/[u^2/s + v] \tag{4.14}$$

with $u = S_{xy}$, $s = S_{xx}$, $v = SS_e$ and the moments of d were obtained by numerically integrating

$$E(d^r) = \int_0^\infty \int_{-\infty}^\infty \{u/[u^2/s + v]\}^r \, g(u, v) \, du dv \,. \tag{4.15}$$

The joint density of (u, v) was easily constructed since u and v are independent with

$$u \sim N(\beta S_{xx}, \sigma^2 S_{xx})$$

$$v \sim \chi^2_{N-2}(0) \ .$$

For the case at hand where the true model is quadratic, the partition in (4.13) still holds because it is purely algebraic and does not depend on the true form of the model. The change due to the model structure lies in the densities of u and v. Considering u first, it is clear that u is still a function of independent normal variables and hence normal but with mean

$$E(u) = E(S_{xy}) = \sum(x_i - \bar{x})\ E(y_i)$$

$$= \beta_1 S_{xx} + \beta_2 \sum(x_i - \bar{x})\ (x_i^2 - \overline{x^2}) \qquad (4.16)$$

and variance the same as in the linear case. For symmetrical designs, the coefficient of β_2 becomes

$$\sum x_i(x_i^2 - \overline{x^2}) = \sum x_i^3 - \overline{x^2} \sum x_i = 0 \ .$$

Hence, if designs are restricted to be symmetrical, the density of u does not change in the presence of a quadratic effect. One is not so fortunate with the density of v. This is because SS_e now contains an independent "lack of fit" contribution which in itself is a non-central chi-square variable. The non-centrality parameter λ is found by evaluating

$$\lambda = \sum[E\{y_i - \bar{y} - b(x_i - \bar{x})\}]^2/(2\sigma^2) \quad . \tag{4.17}$$

Evaluating (4.17) involves the expectation of $b = S_{xy}/S_{xx}$ which upon using (4.16) is found to be

$$E(b) = \beta_1 + C\beta_2$$

where

$$C = \sum(x_i - \bar{x})\ (x_i^2 - \overline{x^2})/S_{xx}$$

which for symmetrical designs has been shown to be zero. Hence, b is unbiased for β_1 even in the presence of a quadratic effect if a symmetrical design is used. The remaining expectations in (4.17) are easily accomplished and λ is found to be

$$\lambda = \beta_2^2\sum(x_i^2 - \overline{x^2})^2/(2\sigma^2) \tag{4.18}$$

for symmetrical designs. The quantity $\Sigma(x_i^2 - \overline{x^2})^2$ in the non-centrality parameter will play a key role in the latter portion of this chapter and will henceforth be designated as S_Q.

Using the transformation from (u, v) to (z, w) in (4.15) where $z = u$ and $w = v^{1/2}$, the expressions for the first two moments of d become

$$E_Q(d) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i K(i)s}{i!} \int_0^{\infty}\int_{-\infty}^{\infty} \frac{zw^{2i+N-3}e^{-B}}{z^2 + sw^2}\, dzdw \tag{4.19}$$

$$E_Q(d^2) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i K(i)s^2}{i!} \int_0^{\infty}\int_{-\infty}^{\infty} \frac{z^2w^{2i+N-3}e^{-B}}{(z^2 + sw^2)^2}\, dzdw \tag{4.20}$$

where λ is given by (4.18), B is the same as in the linear case, namely,

$$B = w^2/(2\sigma^2) + (z - \beta_1 s)^2/(2\sigma^2 s)$$

and

$$K(i) = 2/[(2\sigma^2)^{(1/2)(2i+N-2)}\Gamma[(1/2)(2i+N-2)]\ (2\pi\sigma^2 s)^{1/2}]\ .$$

It is observed that (4.19) and (4.20) are functions of β_1, β_2, σ, N, S_Q, and $s = N\overline{x^2}$ and that J_{IQ} in (4.12) is a function of the same quantities. It will now be shown that J_{IQ} can be considered as a function of γ_1, γ_2, N, S_Q, and $\overline{x^2}$, where $\gamma_1 = |\beta_1/\sigma|$ and $\gamma_2 = |\beta_2/\sigma|$. First, letting $\beta_1 = \sigma\gamma_1'$ and $\beta_2 = \sigma\gamma_2'$ in (4.12) results in

$$2J_{IQ} = 2/3 - (4/3)\gamma_1'\sigma\ E_Q(d) + \sigma^2 E_Q(d^2)[2(1+1/N)$$

$$+ (2/3)\ (\gamma_1')^2 + (\gamma_2')^2\ (2/5 - 4\overline{x^2}/3 + \overline{2x^2}^{\,2}\ )] \qquad (4.21)$$

where one notes $E_Q(d)$ has a coefficient σ and $E_Q(d^2)$ has a coefficient σ^2. Using the transformation (z, w) to (m, n) in both (4.19) and (4.20) with $m = w/\sigma$ and $n = z/\sigma$ provides a Jacobian of σ^2 and the moments of d become

$$E_Q(d) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i s^{1/2}}{i!\ 2^{(1/2)(2i+N-4)}\Gamma[(1/2)(2i+N-2)]\ (2\pi)^{1/2}\ \sigma}$$

$$\int_0^{\infty}\int_{-\infty}^{\infty} nm^{2i+N-3}\ e^{-D}/(n^2 + m^2 s)\ dndm \qquad (4.22)$$

$$E_Q(d^2) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i s^{3/2}}{i!\; 2^{(1/2)(2i+N-4)}\Gamma[(1/2)(2i+N-2)]\; (2\pi)^{1/2}\sigma^2}$$

$$\int_0^{\infty}\int_{-\infty}^{\infty} n^2 m^{2i+N-3}\; e^{-D}/(n^2 + m^2 s)^2 \; dndm \qquad (4.23)$$

where D is the same as for the linear case, namely,

$$D = m^2/2 + (n - \gamma_1' s)^2/(2s)$$

and λ becomes $(\gamma_2')^2\; S_Q/2$. Hence, both $\sigma\; E_Q(d)$ and $\sigma^2\; E_Q(d^2)$ are functions of only γ_1', γ_2', N, S_Q, and $\overline{x^2}$. Also, β_2 appears in J_{IQ} only as a squared quantity so a change in sign in β_2 and hence γ_2' has no effect on J_{IQ}. A change in sign in β_1 and hence γ_1' has the same effect on $E_Q(d)$ as it had on E(d), namely, a change in sign so that the sign of $\beta_1\; E_Q(d)$ remains unchanged. A change in sign in β_1 has no effect on $E_Q(d^2)$ so $\beta_1^2\; E_Q(d^2)$ remains unchanged which establishes that J_{IQ} is a function of N, S_Q, $\overline{x^2}$, $\gamma_1 = |\beta_1/\sigma|$, and $\gamma_2 = |\beta_2/\sigma|$.

At this point, it becomes very clear that the moments of d are strongly influenced by the non-centrality parameter λ which depends on the degrees of freedom for lack of fit and hence on the number of distinct points in the design. It is only natural then for one to begin investigating the effect of design on J_{IQ} with the "two-point" design for N even.

(4.3) N Even - Symmetrical "Two-Point" Designs

Consider the "two-point" symmetrical design with N/2 observations at $x_1 = -x_2$ and N/2 observations at x_2. Since there are two parameters to estimate with N/2 observations at each of two distinct points, the breakdown in degrees of freedom is two degrees of freedom for regression, N/2 - 1 degrees of freedom for pure error at each of two points yielding a total of N-2 degrees of freedom for pure error. Hence, there are N-2-(N-2) = 0 degrees of freedom for lack of fit which means that the non-centrality parameter λ should be zero in this case. This is easily verified since $x_i^2 = x_2^2$ for all i and $\overline{x^2} = \Sigma x_i^2/N = Nx_2^2/N = x_2^2$ so that S_Q and hence λ is zero. This means that the density of v and hence the moments of d are the same as in the linear case. In fact, since the moments of d are identical to those for the linear case, one can write

$$2J_{IQ} = 2/3 - (4/3)\ \beta_1\ E(d) + E(d^2)\ [2\sigma^2(1+1/N)$$

$$+ (2/3)\ \beta_1^2] + E(d^2)\ \beta_2^2\ (2/5 - 4\overline{x^2}/3 + 2\overline{x^2}^2)$$

$$= 2J_{IL} + E(d^2)\ \beta_2^2\ (2/5 - 4\overline{x^2}/3 + 2\overline{x^2}^2) \tag{4.24}$$

and use the values of $2J_{IL}$ and $E(d^2)$ from the linear case in computing $2J_{IQ}$. It is interesting to note that the coefficient of $E(d^2)\ \beta_2^2$ in (4.24), which will be designated as

$$f(\overline{x^2}) = 2/5 - 4\overline{x^2}/3 + 2\overline{x^2}^2\ , \tag{4.25}$$

occurs in the average bias squared in <u>forward</u> regression when a linear approximation is used for a quadratic model. This will be shown in Chapter VI when optimal designs for forward and inverse regression are compared.

The term $E(d^2)\ \beta_2^2\ f(\overline{x^2})$ in (4.24) vanishes if $\beta_2 = 0$ and J_{IQ} degenerates to J_{IL}. However, if $\beta_2 > 0$, this term is strictly positive and can only increase J_{IQ}. It appears then that if one were to minimize $f(\overline{x^2})$, J_{IQ} should be minimized for at least some range of values on β_2. Rewriting (4.25) as

$$f(\overline{x^2}) = 2[(\overline{x^2} - 1/3)^2 + 4/45] \tag{4.26}$$

reveals that $f(\overline{x^2})$ is minimized when $\overline{x^2} = 1/3$. For the case at hand, $\overline{x^2} = x_2^2$ so $\overline{x^2} = 1/3$ implies that $x_2^2 = 1/3$ or that $x_2 = .58$. Hence, for some range of values on β_2, J_{IQ} should be minimized for a "two-point" design by taking N/2 observations each at $x_1 = -.58$ and $x_2 = .58$. This is borne out in Figures 9-12 for N = 10 which shows plots of $2J_{IQ}$ as a function of x_2 for values of $\gamma_1 = 1, 5, 8, 12$ and $\gamma_2 = 0, 1, 3, 5, 8, 12$. The quantities γ_1 and γ_2 are the standardized regression coefficients referred to earlier, namely, $\gamma_1 = |\beta_1/\sigma|$ and $\gamma_2 = |\beta_2/\sigma|$. In each case where $0 < \gamma_2 \leq \gamma_1$, J_{IQ} attains its minimum where x_2 is approximately .58. For values of $\gamma_2 > \gamma_1$, a relative minimum occurs at about $x_2 = .58$, but the absolute minimum occurs where x_2 is much closer to the design center. However, it is difficult to imagine an experimenter misclassifying a model in which $\gamma_2 > \gamma_1$ so that for any practical situation, it could be agreed that the optimal "two-point" design occurs where $x_2 = .58$.

It is well to point out that the optimal designs above are <u>optimal "two-point" designs</u> and not overall optimal designs. There is no evidence at this point that any "two-point" design is optimal from the standpoint of minimizing J_{IQ}. Some light can be shed on this by examining the "three-point" symmetric design for N odd.

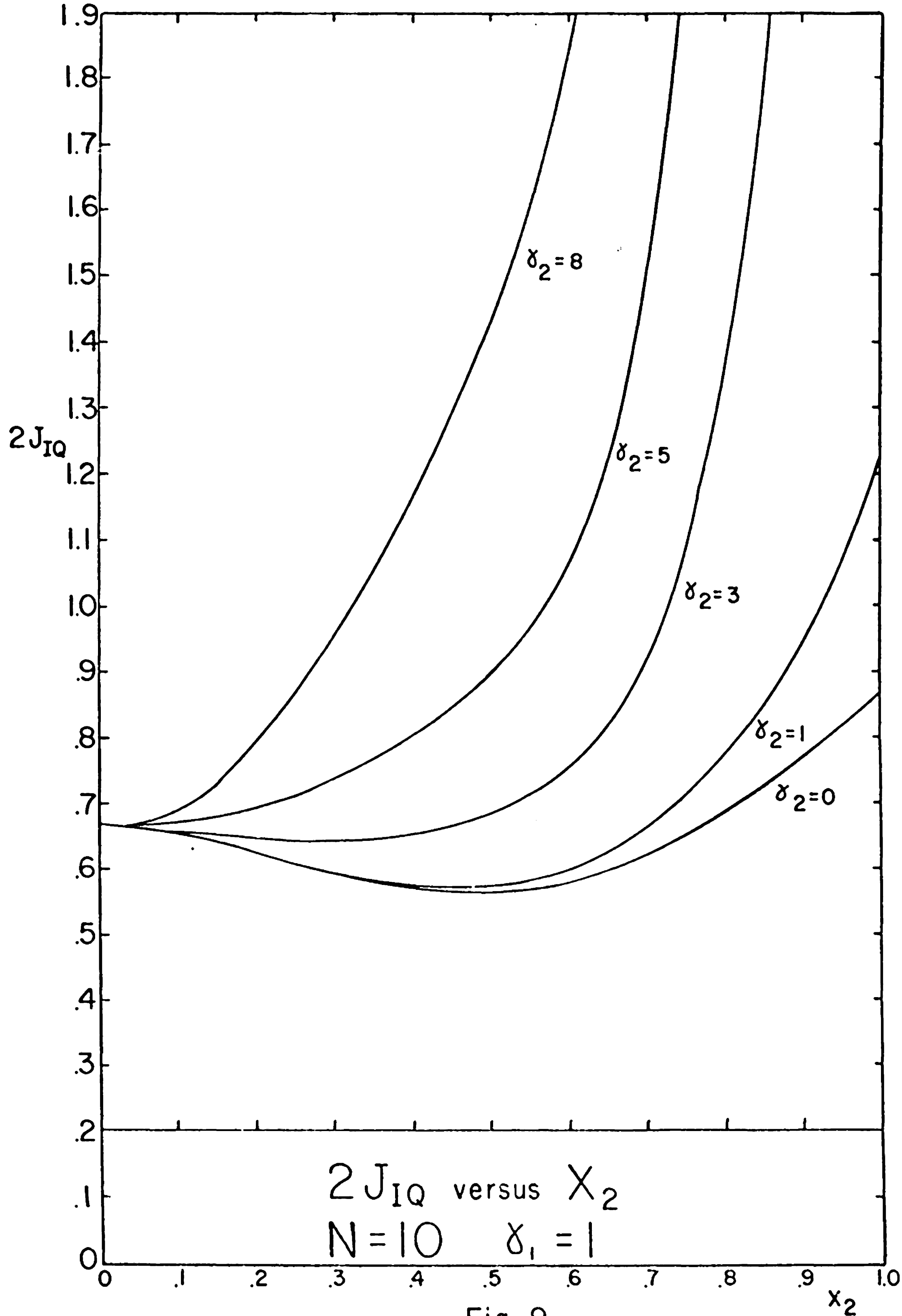

1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
$2J_{IQ}$
$\gamma_2=8$
$\gamma_2=5$
$\gamma_2=3$
$\gamma_2=1$
$\gamma_2=0$
$2J_{IQ}$ versus X_2
$N=10 \quad \gamma_1=1$
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
x_2

Fig. 9

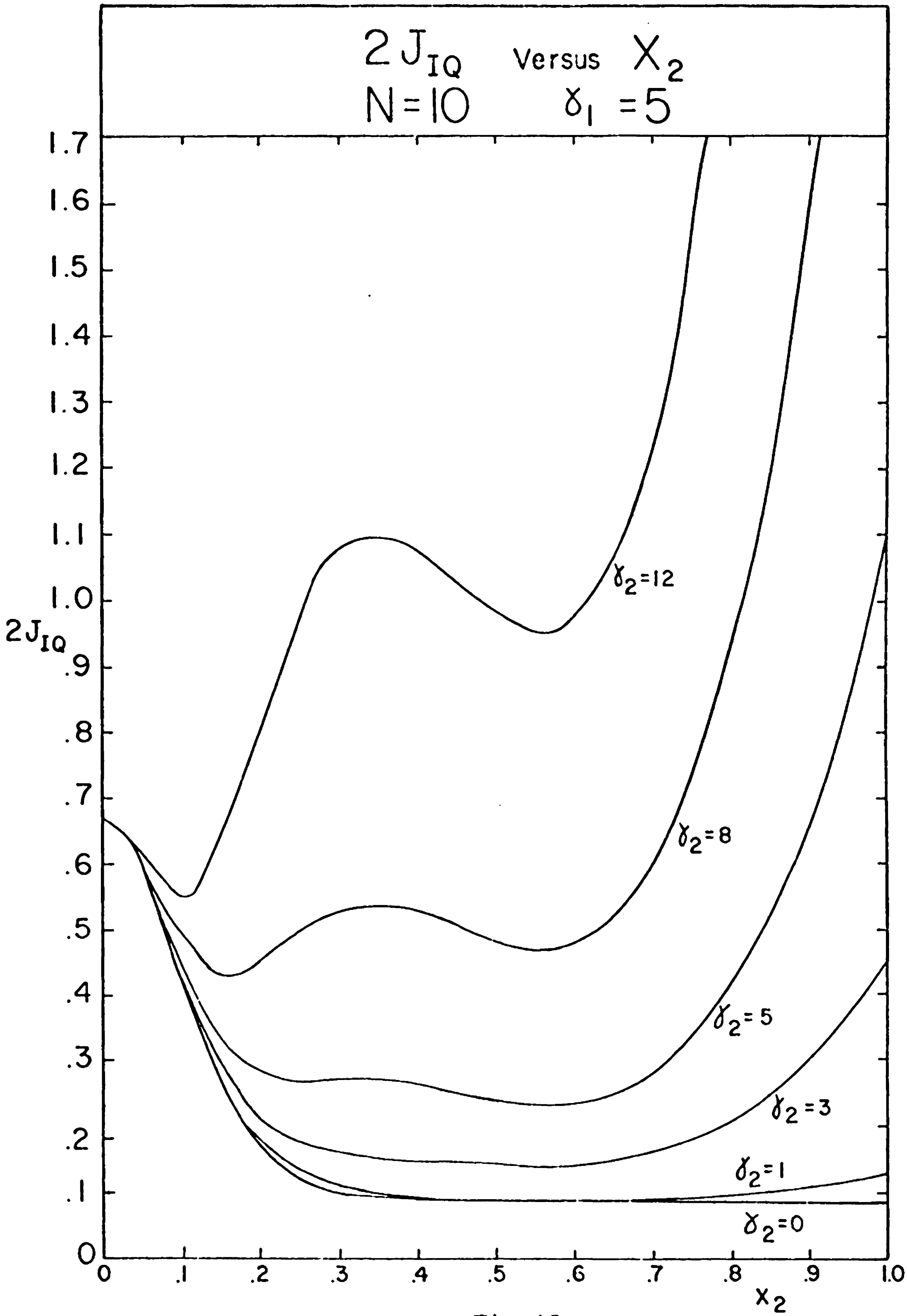

2 J_{IQ} versus X_2
N=10 γ_1 =5
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
2J_{IQ}
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
x_2
γ_2=12
γ_2=8
γ_2= 5
γ_2 = 3
γ_2=1
γ_2=0

Fig. 10

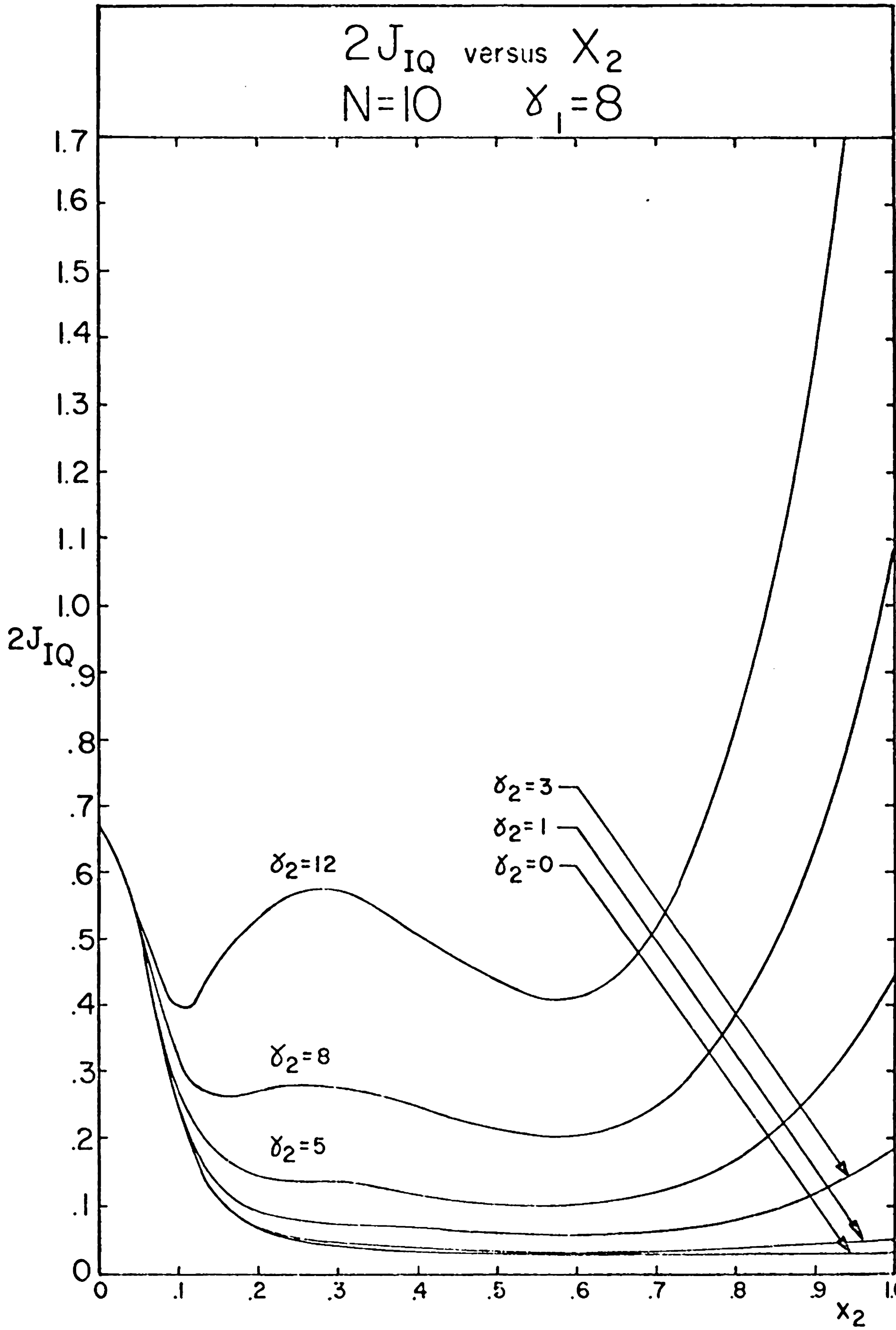
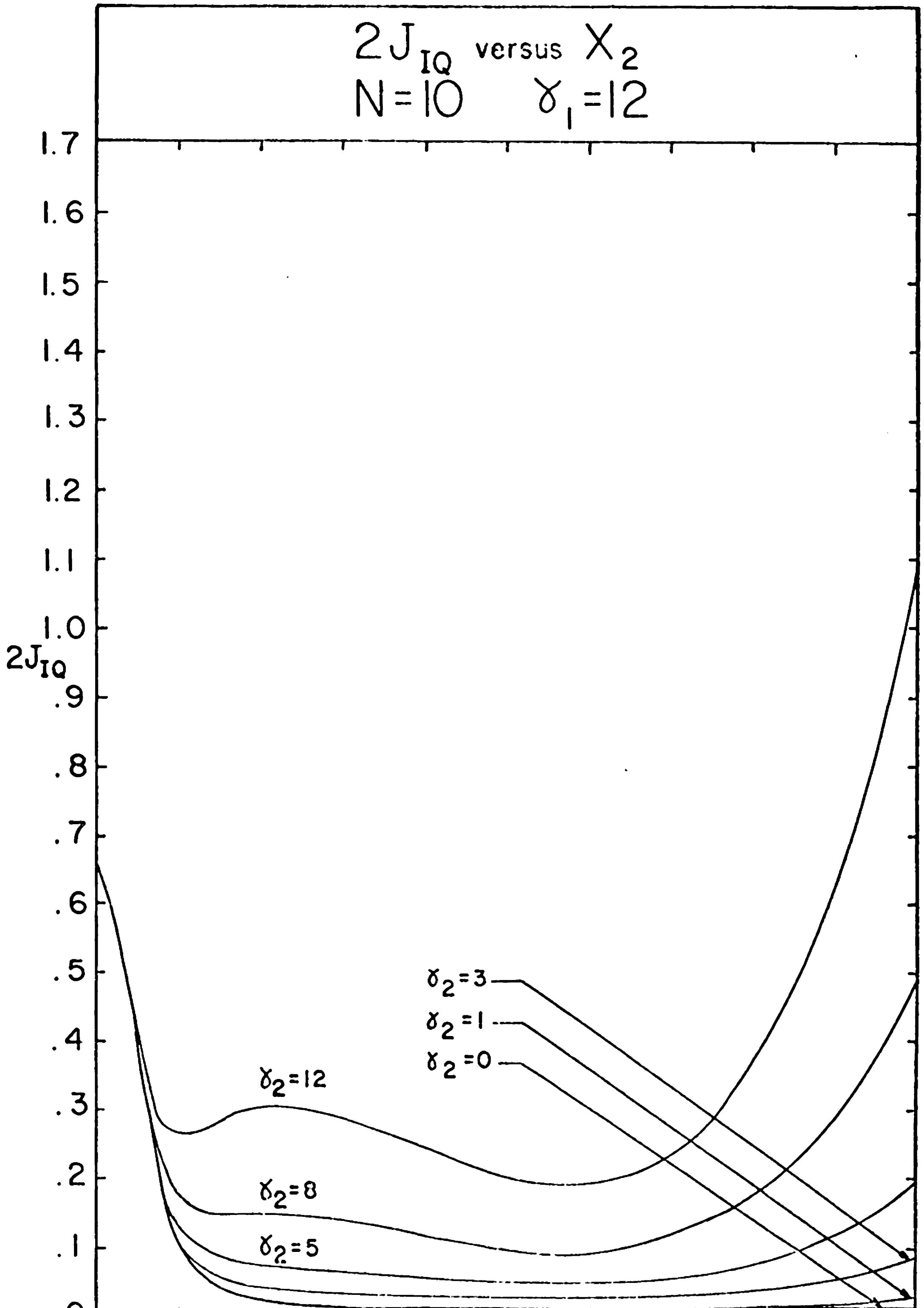

2 J_{IQ} versus X_2
N=10 γ_1=8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
2J_{IQ}
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
γ_2=12
γ_2=8
γ_2=5
γ_2=3
γ_2=1
γ_2=0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
x_2

Fig. 11



2J_IQ versus X_2
N=10 γ_1=12
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
2J_IQ
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
γ_2=3
γ_2=1
γ_2=0
γ_2=12
γ_2=8
γ_2=5
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
x_2

Fig. 12

(4.4) N Odd - Symmetrical "Three-Point" Designs

Only one "three-point" design will be discussed here, namely, the design with a single center point and $(N-1)/2$ observations at each $x_1 = -x_2$ and x_2. Referring to λ in (4.18), it is seen that it is no longer zero but now becomes

$$2\sigma^2\lambda = \beta_2^2 S_Q = \beta_2^2\left[\sum x_i^4 - N\overline{x^2}^2\right]$$

$$= \beta_2^2[(N-1)\ x_2^4 - (N-1)^2\ x_2^4/N]$$

$$= \beta_2^2\ x_2^4(N-1)/N\ . \tag{4.27}$$

This was expected in view of the fact that SS_e for a "three-point" design contains a single degree of freedom contribution for lack of fit. This shows that v is now a non-central chi-square variable so the expressions for the moments of d are those given in equations (4.19) and (4.20) with λ given by (4.27). Examination of these equations reveals that to obtain the moments of d by numerical integration is outside the realm of practical feasibility. To begin with, since the finite quantities for the range of integration derived in Chapter II when v was a central chi-square variable depended on N, they now depend on $N + 2i$ so that each value of i would require different limits of integration. Also, from a pure economics standpoint, in the linear case when v was a central chi-square variable, the machine time required to compute a single value of $2J_{IL}$ for fixed values of s, N, and γ was appreciable. Here one is summing over double integrals and to obtain three place accuracy, a manyfold increase in machine time could be expected.

To gain some insight into the problem, the moments of d were approximated by Monte Carlo procedures. N values of values of y_i in equation (4.1) were selected by fixing the design, β_0, β_1, β_2, and σ^2 and selecting the values of ε_i randomly from a normal density with mean zero and variance σ^2. The N values of y_i and the N design values of x_i were used to obtain a single value of S_{xy} and S_{yy}. The entire process was repeated n times obtaining n values of S_{xy} and S_{yy}, and the average values of S_{xy}/S_{yy} and $(S_{xy}/S_{yy})^2$ were then used to approximate $E_Q(d)$ and $E_Q(d^2)$ respectively. The moments obtained in this fashion were remarkably accurate as indicated by their standard errors, that is, the sample standard deviations of the average values of $(S_{xy}/S_{yy})^r$, $r = 1, 2$. Using $n = 1000$ for the case $N = 9$ revealed that $Av(S_{xy}/S_{yy})^r \pm 3$ standard errors provided nearly two place accuracy for $E_Q(d^2)$ and nearly three place accuracy for $E_Q(d)$ if $\overline{x^2} \geq .2$.

The entire Monte Carlo results for the case $\gamma_1 = 5$, $\gamma_2 = 0, 1, 3, 5, 8$ and $N = 9$ (one center point and four observations each at $x_1 = -x_2$ and x_2) are displayed in Figure 13 where $2J_{IQ}$ is represented as a function of x_2. This figure reveals some interesting results. First, the predominant role played by $f(\overline{x^2})$ in (4.25) is quite clear. It was shown earlier that $f(\overline{x^2})$ is minimized when $\overline{x^2} = 1/3$. For the "three-point" design under discussion, $\overline{x^2} = x_2^2(N-1)/N$ so $\overline{x^2} = 1/3$ implies that

$$x_2 = \sqrt{N/[3(N-1)]} \; . \tag{4.28}$$

Using $N = 9$ in (4.28), one obtains $x_2 = .61$, and Figure 13 shows that in each case except $\gamma_2 = 0$, $2J_{IQ}$ is minimized when x_2 is approximately .61. Also, if one compares Figure 13 with Figure 10, the "two-point"

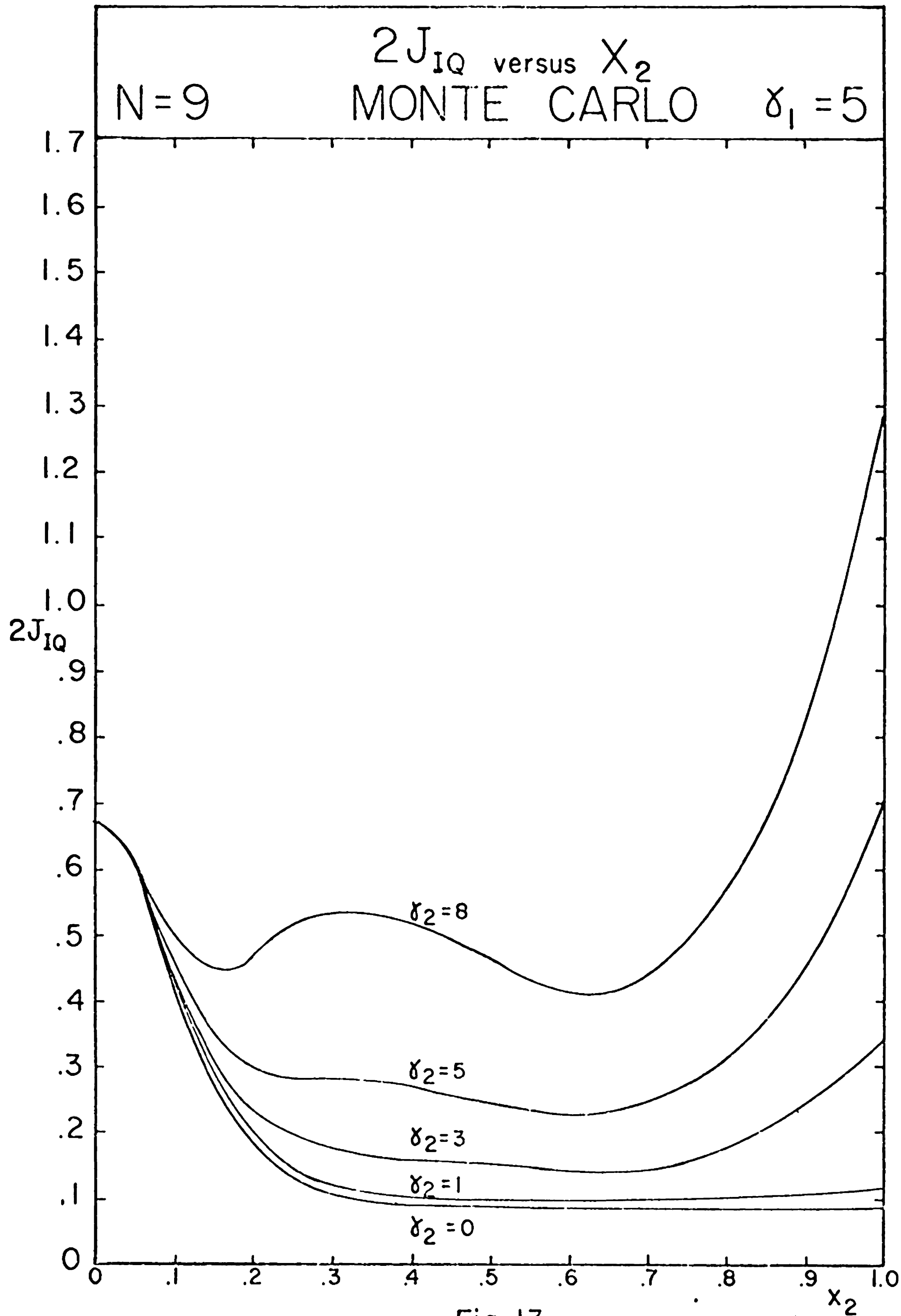

2J_IQ versus X_2
N=9
MONTE CARLO
γ_1 =5
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
2J_IQ
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
x_2
γ_2=8
γ_2=5
γ_2=3
γ_2=1
γ_2 =0

Fig. 13

design for $N = 10$, he sees that the inclusion of the single center point had the effect of drastically reducing $2J_{IQ}$ at $x_2 = 1.0$. More importantly, it is noted that the Min $2J_{IQ}$ is lower for all $\gamma_2 > 0$ for $N = 9$ than it is for $N = 10$. This means that one is getting better precision with one less observation which implies that the "two-point" design is clearly inefficient and hence not over-all optimal.

In this section and the one previous, two and three point designs have been examined and found to be optimized when the pseudo bias term $f(\overline{x^2})$ is minimized which occurs when $\overline{x^2} = 1/3$. For the "two-point" design, this occurs when $x_2 = .58$, but for the "three-point" design with a single center point, the optimal value of x_2 depends on N and ranges from .58 for infinite N to .667 for $N = 4$. Again, there is no evidence that either the two or three point design is over-all optimal from the standpoint of minimizing J_{IQ}, but it was found that the "three-point" design provided a lower Min J_{IQ} than the "two-point" design for a larger value of N. In the quest for an over-all optimal design, it is clear that one needs a rapid, efficient means of evaluating J_{IQ} for many values of the model parameters and the design variables $\overline{x^2}$ and S_Q. While the Monte Carlo procedure used in this section is more efficient than numerically integrating (4.19) and (4.20), a more rapid means of evaluating these quantities is sought. The approximation developed in Chapter II was rapid, efficient, and accurate provided γ_1 was not too small, and the next section will be devoted to revising it for the present quadratic case.

(4.5) Approximation to the Moments of d - Quadratic Model

Expressions (2.59) and (2.61) hold in general as approximations for $E(d) = E(w/z)$ and $E(d^2) = E(w/z)^2$ respectively. However, the joint moment generating function of $(w,z) = (S_{xy}, S_{yy})$ does change as well as the moments of z and the product moments of w and z. Recall that even in the presence of a quadratic effect, the density of $w = S_{xy}$ does not change if designs are restricted to the symmetrical class of designs. In order to express (2.59) and (2.61) in terms of the model parameters and the design variables, it will be necessary to obtain the moments involved when the true model is quadratic as given by equation (4.1). To obtain these moments, the joint moment generating function of (w,z) will have to first be derived under the quadratic model assumption.

The general expression for the joint moment generating function of $w = S_{xy}$ and $z = S_{yy}$ developed in Chapter II is given in equation (2.49) as

$$\phi(t_1, t_2) = |V|^{-1/2}\, e^{\underline{g}'\, V^{-1}\, \underline{g}/(2\sigma^2) + t_1\underline{a}'\underline{\mu} + t_2\underline{\mu}' A\underline{\mu}} \tag{4.29}$$

with

$$\underline{g} = (2t_2 A\underline{\mu} + t_1\underline{a})\sigma^2$$

$$V = (I - 2\sigma^2 t_2 A)$$

$$\underline{a}' = (x_1 - \bar{x}, \ldots, x_N - \bar{x})$$

and where A is symmetric indempotent matrix of dimension NxN of rank $\nu = N-1$ with elements $1-1/N$ on the main diagonal and off diagonal elements of $-1/N$ and where $\underline{\mu} = E(\underline{y})$. In simplifying (4.29) for the linear case, the mean vector $\underline{\mu}'$ was equal to $(\alpha + \beta x_1, \ldots, \alpha + \beta x_N)$. If the true model is quadratic, the only change in (4.29) is that

$$\underline{\mu}' = (\beta_0 + \beta_1(x_1 - \bar{x}) + \beta_2(x_1^2 - \overline{x^2}), \ldots, \beta_0 + \beta_1(x_N - \bar{x}) + \beta_2(x_N^2 - \overline{x^2})) \quad (4.30)$$

In simplifying (4.29) with the present value of $\underline{\mu}$ given by (4.30), some results can be salvaged from Chapter II, namely, V^{-1} and $|V|^{-1/2}$ remain the same since they do not involve $\underline{\mu}$. The quantity $\underline{\mu}'A\underline{\mu}$ is easily simplified as

$$\begin{aligned} \underline{\mu}'A\underline{\mu} &= \sum_{i=1}^{N} [E(y_i - \bar{y})]^2 \\ &= \sum_{i=1}^{N} [\beta_1(x_i - \bar{x}) + \beta_2(x_i^2 - \overline{x^2})]^2 \\ &= \beta_1^2 S_{xx} + \beta_2^2 S_Q \end{aligned} \quad (4.31)$$

since the cross product term vanishes for symmetrical designs. It is further found that

$$\underline{a}'\underline{\mu} = \beta_1 S_{xx} , \quad (4.32)$$

which is the same as in the linear case. Using (4.30) for $\underline{\mu}$ in $\underline{g}$ and using the form of V^{-1} derived in Chapter II results in

$$\begin{aligned} \underline{g}' V^{-1} \underline{g}/(2\sigma^2) = \{&4t_2^2(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 4t_1 t_2 \beta_1 S_{xx} \\ &+ t_1^2 S_{xx}\} \sigma^2/[2(1 - 2\sigma^2 t_2)] . \end{aligned} \quad (4.33)$$

Finally, incorporating (4.31), (4.32), (4.33), and $|V|^{-1/2}$ from Chapter II into (4.29) yields

$$\phi(t_1 t_2) = \exp\{[4t_2^2(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 4t_1 t_2 \beta_1 S_{xx}$$

$$+ t_1^2 S_{xx}]\ \sigma^2/[2(1 - 2\sigma^2 t_2)] + t_1 \beta_1 S_{xx}$$

$$+ t_2(\beta_1^2 S_{xx} + \beta_2^2 S_Q)\}\ (1 - 2\sigma^2 t)^{-\nu/2}$$

$$= \exp\{[t_1^2 S_{xx}\ \sigma^2/2 + t_1 \beta_1 S_{xx} + t_2(\beta_1^2 S_{xx} + \beta_2^2 S_Q)]/$$

$$(1 - 2\sigma^2 t_2)\}\ (1 - 2\sigma^2 t_2)^{-\nu/2} \tag{4.34}$$

where $\nu = N-1$ and $S_Q = \Sigma(x_i^2 - \overline{x^2})^2$. This is the joint moment generating function of S_{xy} and S_{yy} when the true model is quadratic and where all odd design moments are taken as zero. Performing the usual checks on (4.34), it can be verified that $\phi(0, 0) = 1$ and that $\phi(t_1, 0)$ and $\phi(0, t_2)$ are of correct form so that one can rest assured that (4.34) is of correct form.

Equation (4.34) must now be used to extract the moments needed to express the approximations for $E_Q(d)$ and $E_Q(d^2)$ given in (2.59) and (2.61) as functions of the model parameters and the design variables. The procedure used was identical to that discussed in Chapter II; the results are listed below:

$$E_Q(w) = \beta_1 S_{xx} \tag{4.35}$$

$$E_Q(w^2) = \sigma^2 S_{xx} + [E_Q(w)]^2 \tag{4.36}$$

$$E_Q(z) = \beta_1^2 S_{xx} + \beta_2^2 S_Q + \nu\sigma^2 \tag{4.37}$$

$$E_Q(z^2) = 4\sigma^2(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 2\nu\sigma^4 + [E_Q(z)]^2 \qquad (4.38)$$

$$E_Q(z^3) = 24\sigma^4(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 8\nu\sigma^6$$

$$+ E_Q(z)[8\sigma^2(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 4\nu\sigma^4]$$

$$+ E_Q(z^2)\ E_Q(z) \qquad (4.39)$$

$$E_Q(z^4) = 192\sigma^6(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 48\nu\sigma^8$$

$$+ E_Q(z)[72\sigma^4(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 24\nu\sigma^6]$$

$$+ E_Q(z^2)[12\sigma^2(\beta_1^2 S_{xx} + \beta_2^2 S_Q) + 6\nu\sigma^4]$$

$$+ E_Q(z^3)\ E_Q(z) \qquad (4.40)$$

$$E_Q(wz) = S_{xx}\beta_1[E_Q(z) + 2\sigma^2] \qquad (4.41)$$

$$E_Q(wz^2) = S_{xx}\beta_1[E_Q(z^2) + 4\sigma^2\ E_Q(z) + 8\sigma^4] \qquad (4.42)$$

$$E_Q(wz^3) = S_{xx}\beta_1[E_Q(z^3) + 6\sigma^2\ E_Q(z^2) + 24\sigma^4\ E_Q(z) + 48\sigma^6] \qquad (4.43)$$

$$E_Q(w^2z) = S_{xx}\beta_1[2\sigma^2\ E_Q(w) + E_Q(wz)]$$

$$+ \sigma^2 S_{xx}[2\sigma^2 + E_Q(z)] \qquad (4.44)$$

$$E_Q(w^2z^2) = S_{xx}\beta_1[E_Q(wz^2) + 4\sigma^2\ E_Q(wz) + 8\sigma^4\ E_Q(w)]$$

$$+ \sigma^2 S_{xx}[E_Q(z^2) + 4\sigma^2\ E_Q(z) + 8\sigma^4]\ . \qquad (4.45)$$

An attempt to substitute the above quantities into (2.59) and (2.61) and obtain simplified expressions for the approximate moments was not made. Instead, a computer program was written to compute $2J_{IQ}$ in (4.12) for specified values of the model parameters and design variables using (2.59) and (2.61) for $E_Q(d)$ and $E_Q(d^2)$ respectively by computing expressions (4.35) to (4.45) in the order given above for the specified values and then substitution of these numerical quantities into (2.59) and (2.61). This procedure saved considerable algebraic manipulation.

The accuracy of the approximations was checked against the Monte Carlo values of the moments used in the construction of Figure 13 and against additional Monte Carlo values for "three-point" designs for $N = 10$ and 15. As in the linear case, the goodness of the approximations appears to depend primarily on the magnitude of γ_1 and to some extent on the magnitude of $\overline{x^2}$. The results indicate that provided $\overline{x^2}$ is not less than about .20, the approximate moments are quite accurate for γ_1 as low as 5, being within about one percent of the Monte Carlo moments. For $\gamma_1 = 2.5$ and $N = 10$, the deviations of the approximate moments from the Monte Carlo moments are within about three percent. However, as N increases, the correspondence between the two improves. For example, for $\gamma_1 = 2.5$ and $N = 15$, the deviations of the approximate moments from the Monte Carlo moments are within about one percent. Some of the numerical results in the remainder of this chapter and in Chapter V include values of $\gamma_1 = 2.5$ with the understanding that as N decreases, the worth of the approximations for this small value of γ_1 decreases.

These approximations provide one with the necessary tools to evaluate $2J_{IQ}$ rapidly and efficiently for a large number of values of the model parameters and the design variables in the search for overall optimality.

(4.6) Optimal Values of the Design Variables

In Chapter II where the true model was assumed linear, J_{IL} was a function of the model parameters and two design variables, $\bar{x}$ and $s = S_{xx} = N\overline{x^2}$. However, J_{IL} could be factored such that the portion containing $\bar{x}$ was strictly positive for non-zero $\bar{x}$ and a non-zero $\bar{x}$ could thus only increase J_{IL}. Hence, the first requirement necessary to minimize J_{IL} was to choose designs for which $\bar{x} = 0$. (The class of designs for which $\bar{x} = 0$ was then restricted to the class of symmetrical designs since they provide not only $\bar{x} = 0$ but also all odd design moments equal to zero and, as has been observed, simplified the present quadratic case considerably.) After $\bar{x}$ was set equal to zero, only one design variable remained, i.e., S_{xx} and the value of this variable which minimized J_{IL} was found by plotting J_{IL} as a function of S_{xx} for a wide range of values of the model parameters.

In the present case where a linear approximation is being used for a quadratic model and designs are restricted to be symmetrical, two design variables are also present, $S_{xx} = N\overline{x^2}$ and S_Q, but unlike the previous case, neither one can be eliminated at the onset. Recalling the form of J_{IQ} in (4.12), $\overline{x^2}$ is present in the pseudo bias term $f(\overline{x^2})$ and in the expressions for the moments of d. The variable S_Q appears only in the non-centrality parameter in the expressions for the moments of d. To minimize J_{IQ}, one of these variables must be held fixed while varying the other and conversely. Since $\overline{x^2} = 1/3$ minimizes $f(\overline{x^2})$, it is only natural to begin by holding $\overline{x^2}$ fixed at 1/3 and varying S_Q for fixed values of the model parameters N, γ_1, and γ_2. For each combination of the values of N, γ_1, and γ_2 considered, S_Q was varied from its maximum value (which depends on N) to zero in increments of .05. Values of the model parameters considered were N = 5(1)20(2)30(5)50, γ_1 = 2.5, 5, 8, 12, and γ_2 = 1, 3, 5, 8, 12. The results show that for $\bar{x}^2 = 1/3$, the value of S_Q which minimizes J_{IQ} is nearly independent of the parameters γ_1 and γ_2 and depends primarily on N, the number of observations. Representative results for three values of N, N = 9, 15, 24 are shown in Tables III, IV, and V. For each combination of values of γ_1 and γ_2, the tables show the range of values of the design variable S_Q which minimizes $2J_{IQ}$, designated by Opt S_Q, the resulting minimum value of $2J_{IQ}$, designated by Min $2J_{IQ}$, the value of $2J_{IQ}$ resulting from setting S_Q = N/10 for all values of γ_1 and γ_2, designated by $2J_{IQ}$ (N/10), and the percent increase in $2J_{IQ}$ (N/10) above Min $2J_{IQ}$. For example, consider Table III for N = 9 and enter the row for $\gamma_1 = 8$ and $\gamma_2 = 3$. All values of S_Q ranging between 1.05 and 1.60 inclusive result in a minimum value of $2J_{IQ}$ of .0576. However, a value of S_Q = N/10 = .9 provides a $2J_{IQ}$ of .0578, only .35 percent greater than the minimum value. For all values of N = 5(1)20(2)30(5)50, the results are basically the same, that is, a value of S_Q = N/10 provides a value of $2J_{IQ}$ which is very close to Min $2J_{IQ}$, usually well within one percent, and in over half of the 500 cases considered, the value of $2J_{IQ}$ evaluated at S_Q = N/10 is identical to the value of Min $2J_{IQ}$ to three significant digits. Slightly larger deviations occur for small values of N, γ_1, and γ_2, the largest being 3.75% at $\gamma_1 = 2.5$, $\gamma_2 = 3$ for N = 5. For N = 6, the largest deviations is 2.65% at $\gamma_1 = 2.5$, $\gamma_2 = 3$. As N increases, the percentages for low γ_1 and γ_2 continue to decrease as

TABLE III

OPTIMAL VALUES OF S_Q

$\overline{x^2} = 1/3$

N = 9

γ_1	γ_2	Opt S_Q	Min $2J_{IQ}$	$2J_{IQ}(.9)$	% > Min $2J_{IQ}$
2.5	1.0	3.50	.256	.262	2.34
	3.0	[1.10,1.40]	.331	.334	.91
	5.0	1.00	.420	.421	.24
	8.0	[.90,1.00]	.518	.518	.00
	12.0	[.80, .90]	.584	.584	.00
5.0	1.0	[3.35,3.50]	.0897	.0910	1.45
	3.0	[0.95,1.70]	.130	.131	.77
	5.0	[0.80,1.15]	.196	.196	.00
	8.0	[0.80, .95]	.304	.304	.00
	12.0	[0.80, .90]	.420	.420	.00
8.0	1.0	[2.60,3.50]	.0384	.0386	.52
	3.0	[1.05,1.60]	.0576	.0578	.35
	5.0	[0.95,1.00]	.0927	.0928	.11
	8.0	[0.75,1.05]	.164	.164	.00
	12.0	[0.80, .90]	.265	.265	.00
12.0	1.0	[0.65,3.50]	.0177	.0177	.00
	3.0	[1.00,1.70]	.0269	.0270	.37
	5.0	[0.95,1.05]	.0446	.0447	.22
	8.0	[0.85, .95]	.0840	.0840	.00
	12.0	[0.75, .95]	.151	.151	.00

TABLE IV

OPTIMAL VALUES OF S_Q

$\overline{x^2} = 1/3$

N = 15

γ_1	γ_2	Opt S_Q	Min $2J_{IQ}$	$2J_{IQ}(1.5)$	% > Min $2J_{IQ}$
2.5	1.0	[3.90,6.10]	.245	.248	1.22
	3.0	[1.80,1.90]	.322	.323	.31
	5.0	[1.40,1.70]	.415	.415	.00
	8.0	[1.40,1.50]	.514	.514	.00
	12.0	[1.40,1.50]	.582	.582	.00
5.0	1.0	[4.60,6.80]	.0847	.0852	.59
	3.0	[1.70,1.95]	.125	.126	.80
	5.0	[1.30,1.80]	.192	.192	.00
	8.0	[1.35,1.50]	.301	.301	.00
	12.0	[1.35,1.45]	.418	.419	.24
8.0	1.0	[2.80,9.40]	.0360	.0361	.28
	3.0	[1.75,2.00]	.0553	.0554	.18
	5.0	[1.40,1.65]	.0907	.0907	.00
	8.0	[1.20,1.65]	.162	.162	.00
	12.0	[1.25,1.50]	.264	.264	.00
12.0	1.0	[2.20,10.4]	.0165	.0166	.61
	3.0	[1.10,2.65]	.0258	.0258	.00
	5.0	[1.20,1.85]	.0436	.0436	.00
	8.0	[1.35,1.50]	.0830	.0830	.00
	12.0	[1.20,1.50]	.150	.150	.00

TABLE V

OPTIMAL VALUES OF S_Q

$\overline{x^2} = 1/3$

N = 24

γ_1	γ_2	Opt S_Q	Min $2J_{IQ}$	$2J_{IQ}(2.4)$	% > Min $2J_{IQ}$
2.5	1.0	[4.30,7.30]	.239	.240	.42
	3.0	[2.10,3.10]	.318	.318	.00
	5.0	[2.20,2.50]	.411	.411	.00
	8.0	[2.10,2.40]	.512	.512	.00
	12.0	[2.10,2.30]	.581	.582	.17
5.0	1.0	[5.00,8.40]	.0819	.0821	.24
	3.0	[1.80,3.60]	.123	.123	.00
	5.0	[1.90,2.80]	.190	.190	.00
	8.0	[1.95,2.50]	.300	.300	.00
	12.0	[2.10,2.25]	.417	.418	.24
8.0	1.0	[1.00,13.0]	.0347	.0347	.00
	3.0	[1.95,3.40]	.0541	.0541	.00
	5.0	[2.10,2.55]	.0895	.0895	.00
	8.0	[1.85,2.60]	.161	.161	.00
	12.0	[2.05,2.35]	.263	.264	.38
12.0	1.0	[0.00,19.0]	.0159	.0159	.00
	3.0	[1.15,4.25]	.0252	.0252	.00
	5.0	[1.75,2.90]	.0430	.0430	.00
	8.0	[2.00,2.40]	.0825	.0825	.00
	12.0	[1.80,2.60]	.150	.150	.00

shown in Tables III, IV, and V, that is, for $N = 9$, the largest deviation is 2.34% at $\gamma_1 = 2.5$, $\gamma_2 = 1$ but for $N = 15$ the largest deviation is 1.22% and for $N = 24$, all percentages are well below one. In the 500 cases considered, only twenty-one deviations are greater than one percent and they all occur for low values of N, γ_1, and γ_2. Considering that approximations are being used for the moments of d, one could not get much closer than one percent to Min $2J_{IQ}$ and have meaningful results since the results would fall outside the range of accuracy of the approximations. Also, concerning the larger deviations which occur for small values of γ_1, eleven of the twenty-one deviations greater than one percent occur for small N and $\gamma_1 = 2.5$, and as aforementioned, the approximations for the moments of d at $\gamma_1 = 2.5$ become less accurate as N decreases. Hence, for small N and $\gamma_1 = 2.5$, one could not get much closer than three percent to Min $2J_{IQ}$ and have meaningful results.

The preceding results were obtained by fixing the value of $\overline{x^2} = 1/3$. Referring to the two and three point designs discussed earlier in the chapter, the pseudo bias term $f(\overline{x^2})$ appears to be predominant since minimizing it with $\overline{x^2} = 1/3$ minimized J_{IQ} for these designs. However, since $\overline{x^2}$ also appears in the moments of d, it is not entirely clear that minimizing $f(\overline{x^2})$ will minimize J_{IQ} in general. To insure that this is the case, $2J_{IQ}$ was evaluated over a range of values of S_Q while fixing $\overline{x^2}$ at values other than 1/3. The results are shown in Table VI for the case $N = 10$; $\gamma_1 = 2.5, 5, 8, 12$; $\gamma_2 = 1, 3, 5, 8, 12$. The entries in the table are the minimum values that $2J_{IQ}$ can attain by varying S_Q while holding $\overline{x^2}$ fixed at values of .20, .25, .30, .33, .36, and .50. It is readily seen that the values of Min $2J_{IQ}$ in

TABLE VI

VALUES OF MIN $2J_{IQ}$

$\overline{x^2}$ = .20, .25, .30, .33, .36, .50

N = 10

γ_1	γ_2	$\overline{x^2}$=.20	$\overline{x^2}$=.25	$\overline{x^2}$=.30	$\overline{x^2}$=.33	$\overline{x^2}$=.36	$\overline{x^2}$=.50
2.5	1.0	.257	.255	.254	.253	.253	.263
	3.0	.346	.336	.330	.329	.329	.346
	5.0	.442	.429	.421	.419	.419	.445
	8.0	.537	.526	.518	.517	.517	.540
	12.0	.597	.590	.585	.584	.584	.599
5.0	1.0	.0915	.0898	.0888	.0884	.0882	.0901
	3.0	.139	.133	.130	.129	.129	.140
	5.0	.214	.203	.196	.195	.195	.220
	8.0	.331	.315	.305	.303	.304	.340
	12.0	.447	.431	.422	.420	.420	.456
8.0	1.0	.0396	.0386	.0380	.0378	.0376	.0380
	3.0	.0625	.0593	.0575	.0570	.0574	.0628
	5.0	.104	.0969	.0931	.0922	.0925	.107
	8.0	.184	.172	.166	.163	.164	.192
	12.0	.293	.276	.267	.265	.266	.305
12.0	1.0	.0184	.0178	.0175	.0174	.0173	.0174
	3.0	.0294	.0278	.0269	.0266	.0266	.0294
	5.0	.0504	.0468	.0448	.0444	.0445	.0523
	8.0	.0963	.0888	.0846	.0838	.0842	.102
	12.0	.172	.159	.152	.151	.152	.182

the column $\overline{x^2} = .33$ are lower than the corresponding values of Min $2J_{IQ}$ for values of $\overline{x^2} < .33$. They are also lower than the values of Min $2J_{IQ}$ for $\overline{x^2} = .50$. Comparing the values of Min $2J_{IQ}$ in the columns $\overline{x^2} = .33$ and $\overline{x^2} = .36$, it is seen that for three values of γ_1 and γ_2, the value of Min $2J_{IQ}$ is slightly smaller for $\overline{x^2} = .36$ than for $\overline{x^2} = .33$. These values all occur at $\gamma_1 \geq 5$ and $\gamma_2 = 1$, that is, $\gamma_1 = 5$, $\gamma_2 = 1$; $\gamma_1 = 8$, $\gamma_2 = 1$; $\gamma_1 = 12$, $\gamma_2 = 1$. For all other values of γ_1 and γ_2, the values of Min $2J_{IQ}$ are lower for $\overline{x^2} = .33$. This is understandable if one considers that for $\gamma_1 = 1$, the model is very close to linear, and it has been shown in Chapter II that optimal designs for the linear model depend on both γ_1 and N and do not occur at precisely $\overline{x^2} = 1/3$ or equivalently $S_{xx} = N/3$. In fact, the figures used to construct Figure 2 for $N = 10$ reveal that for $\gamma_1 = 5$, 8, and 12, the optimal designs occur at values of $S_{xx} > N/3$ while for $\gamma_1 = 2.5$, the optimal design occurs when S_{xx} is approximately equal to $N/3$. Hence, it is understandable that a value of $\overline{x^2} > 1/3$ yields a lower $2J_{IQ}$ than $\overline{x^2} = 1/3$ for small γ_2 and $\gamma_1 \geq 5$.

The results of this section are simply that if γ_2 has any magnitude whatever, say ≥ 1, designs which are very near optimal can be obtained for all $\gamma_1 \geq 2.5$ by requiring that $\overline{x^2} = 1/3$ and $S_Q = N/10$. The restriction on γ_1 does not involve the work in this section but is required because the moments of d are approximated and the goodness of the approximation is primarily dependent on a value of $\gamma_1 \geq 2.5$. Hence, values of γ_1 less than 2.5 were not considered.

(4.7) Optimal "Multi-Point" Designs

Designs which satisfy

$$\overline{x^2} = 1/3 \tag{4.46}$$

$$S_Q = \sum_{i=1}^{N} (x_i^2 - \overline{x^2})^2 = N/10 \tag{4.47}$$

will be termed "near optimal", and specific designs which satisfy these conditions will be developed in the present section. It will first be shown that except for the case where N is a multiple of 19, there does not exist a near optimal "three-point" design and under no conditions does there exist a near optimal "two-point" design. Consider a two or three point design with N_1 observations at x_1, N_0 observations at $x_0 = 0$ ($N_0 \geq 0$), and N_2 observations at x_2 with $N_1 = N_2$ and $x_1 = -x_2$ for symmetry and, of course, $N_1 + N_0 + N_2 = N$. To satisfy (4.46) requires that $\overline{x^2} = 1/3$, but for the present case,

$$\overline{x^2} = \sum x_i^2/N = 2N_2x_2^2/N \ , \tag{4.48}$$

and setting (4.48) equal to 1/3 results in

$$x_2^2 = N/(6N_2) \ . \tag{4.49}$$

For the present case,

$$S_Q = \sum x_i^4 - N\overline{x^2}^2 = (2N_2 - 4N_2^2/N)\ x_2^4 \ ,$$

and substituting (4.49) for x_2^2 results in

$$S_Q = (N/18)\ (N/N_2 - 2) \ . \tag{4.50}$$

For a "two-point" symmetrical design, $N_0 = 0$ and $N_2 = N/2$. Substitution of $N_2 = N/2$ in (4.50) yields $S_Q = 0$ and for near optimality, S_Q must equal N/10 so a near optimal "two-point" design does not exist. The fact that $S_Q = 0$ for a "two-point" design was, of course, known at the onset, for it was this term which forced the non-centrality parameter λ to zero in the earlier discussion of "two-point" design. For a "three-point" symmetrical design, setting (4.50) equal to N/10 and simplifying results in

$$N_2 = N/3.8 \; . \tag{4.51}$$

It is clear that unless N is a multiple of 19, N_2 will not be integral and a near optimal design will not exist. Consider now the case where N is a multiple of 19 and let N = c19 with c integral. Using (4.51), $N_2 = 5c$ and $N_0 = N - 2N_2 = 19c - 10c = 9c$. Also, using (4.49), $x_2^2 = N/(6N_2) = 19c/(30c) = 19/30$ and $x_2 = .796$. Hence, when N = c19, conditions for near optimality can be obtained by taking 5c observations at $x_1 = -\,.796$, 9c observations at $x_0 = 0$, and 5c observations at $x_2 = +\,.796$. In general, it is seen that to achieve a near optimal design, more than three points will have to be used.

Consider now a four or five point design with N_i observations at x_i, i = 0, 1, 2, 3, 4 taking $x_0 = 0$ and with $N_1 = N_4$, $N_2 = N_3$, $x_1 = -\,x_4$, and $x_2 = -\,x_3$ for symmetry and $2N_3 + 2N_4 + N_0 = N$. By taking $N_0 = 0$, one has a "four-point" symmetrical design, and $N_0 > 0$ produces a "five-point" symmetrical design. Under these conditions,

$$\overline{x^2} = 2(N_3x_3^2 + N_4x_4^2)/N$$

and letting $x_3 = kx_4$, $k \geq 0$, for simplicity results in

$$\overline{x^2} = 2(k^2N_3 + N_4)\, x_4^2/N \,. \tag{4.52}$$

Setting (4.52) equal to 1/3 to satisfy condition (4.46) results in

$$x_4^2 = N/[6(k^2N_3 + N_4)] \,. \tag{4.53}$$

Expressing S_Q in terms of the N_i, x_4, N, and k and using (4.46), one obtains

$$\begin{aligned} S_Q &= 2(x_3^4N_3 + x_4^4N_4) - N/9 \\ &= 2x_4^4(k^4N_3 + N_4) - N/9 \,. \end{aligned}$$

Using (4.53) for x_4^2, one obtains

$$S_Q = (N/18)[N(k^4N_3 + N_4)/(k^2N_3 + N_4)^2 - 2] \,, \tag{4.54}$$

and setting (4.54) equal to N/10 to satisfy (4.47) yields

$$(k^4N_3 + N_4)/(k^2N_3 + N_4)^2 = 3.8/N \,. \tag{4.55}$$

The above is simply a quadratic equation in k^2 and after letting $H = 3.8/N$ can be shown to produce the solution

$$k^2 = \frac{HN_3N_4 \pm \sqrt{H(N_3N_4^2 + N_3^2N_4) - N_3N_4}}{N_3 - HN_3^2} \,. \tag{4.56}$$

All positive integral values of N_3 and N_4 such that $2N_3 + 2N_4 \leq N$ which yield real positive solutions for k^2 such that

$$k^2N_3 + N_4 \geq N/6 \tag{4.57}$$

and

$$kx_4 = x_3 \leq 1 \tag{4.58}$$

provide designs which satisfy (4.46) and (4.47) and are hence, near optimal. Condition (4.57) is necessary to insure that the resulting value of x_4 is less than or equal to one. The conditions on N_3 and N_4 which result in satisfactory solutions for k^2 are, in general, difficult to ascertain. However, some insight can be gained by examining the discriminant in (4.56). For real solutions of k^2, $H(N_3N_4^2 + N_3^2N_4) - N_3N_4$ must be greater than or equal to zero which implies that

$$2(N_3 + N_4)/N \geq .526 \ . \tag{4.59}$$

This means that the ratio of non-center points to the total number of observations must be greater than or equal to .526, that is, more than half the observations must be non-center points in order to attain near optimality. As for the remaining conditions, the easiest approach is to merely list all possible candidates (N_3, N_4) such that (4.59) holds and compute k^2, retaining those values of (N_3, N_4) which yield satisfactory solutions and discarding those which do not. This is not as awesome as it appears since the number of candidates is quite small unless N is excessively large. For example, for N = 10, the candidates (N_3, N_4) are

(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (3,1), (3,2), (4,1) .

The first can be immediately discarded since $2(1+1)/10 = .4 < .526$ and

(4.59) does not hold. Also one need only consider one permutation of (N_3, N_4) since (N_3, N_4) and (N_4, N_3) will result in the same design. This narrows the class of candidates to

$$(1,2),\ (1,3),\ (1,4),\ (2,2),\ (2,3)\ . \tag{4.60}$$

These candidates were substituted into (4.56), and six acceptable values of k resulted which gave rise to six near optimal designs. These are shown in Table VII. Note that the first two designs are "four-point" designs while the remaining four are "five-point" designs. The reason there are more optimal designs than candidates is that each candidate admits two solutions for k, both of which may be acceptable. Each of these six designs satisfy conditions (4.46) and (4.47) are thus equivalent from the standpoint of minimizing J_{IQ}. However, some of these may be easier to attain experimentally than others. Also the "four-point" designs provide two degrees of freedom for lack of fit while the "five-point" designs provide three degrees of freedom for lack of fit.

To show the contrast between the near optimal designs developed in this section and the "two-point" designs developed earlier in the chapter, the first design listed in Table VII was used to construct plots of $2J_{IQ}$ as a function of x_4 by letting $x_3 = kx_4 = .4259x_4$ for $\gamma_1 = 5, 8,$ and 12 and $\gamma_2 = 0, 3, 5, 8,$ and 12. These plots are shown in Figures 14, 15, and 16. In comparing these three figures with the corresponding "two-point" designs in Figures 10, 11, and 12, note that the Min $2J_{IQ}$ is considerably lower for the "four-point" design for all $\gamma_2 > 0$ and that these minima for all $\gamma_2 > 0$ occur where x_4 is approximately .98 and hence where x_3 is approximately .42.

TABLE VII

NEAR OPTIMAL DESIGNS FOR N=10

N_0	N_3	x_3	N_4	x_4
0	4	.4186	1	.9827
0	3	.2741	2	.8489
4	2	.6392	1	.9217
4	1	.5115	2	.8382
2	3	.4931	1	.9680
2	2	.3409	2	.8468

x_i = value of x at $i^{\underline{th}}$ design point, $i = 0,\ldots,4$

$x_0 = 0$

N_i = number of observations at x_i

$N_1 = N_4$ $\qquad x_1 = -x_4$

$N_2 = N_3$ $\qquad x_2 = -x_3$

N_0 = number of center points

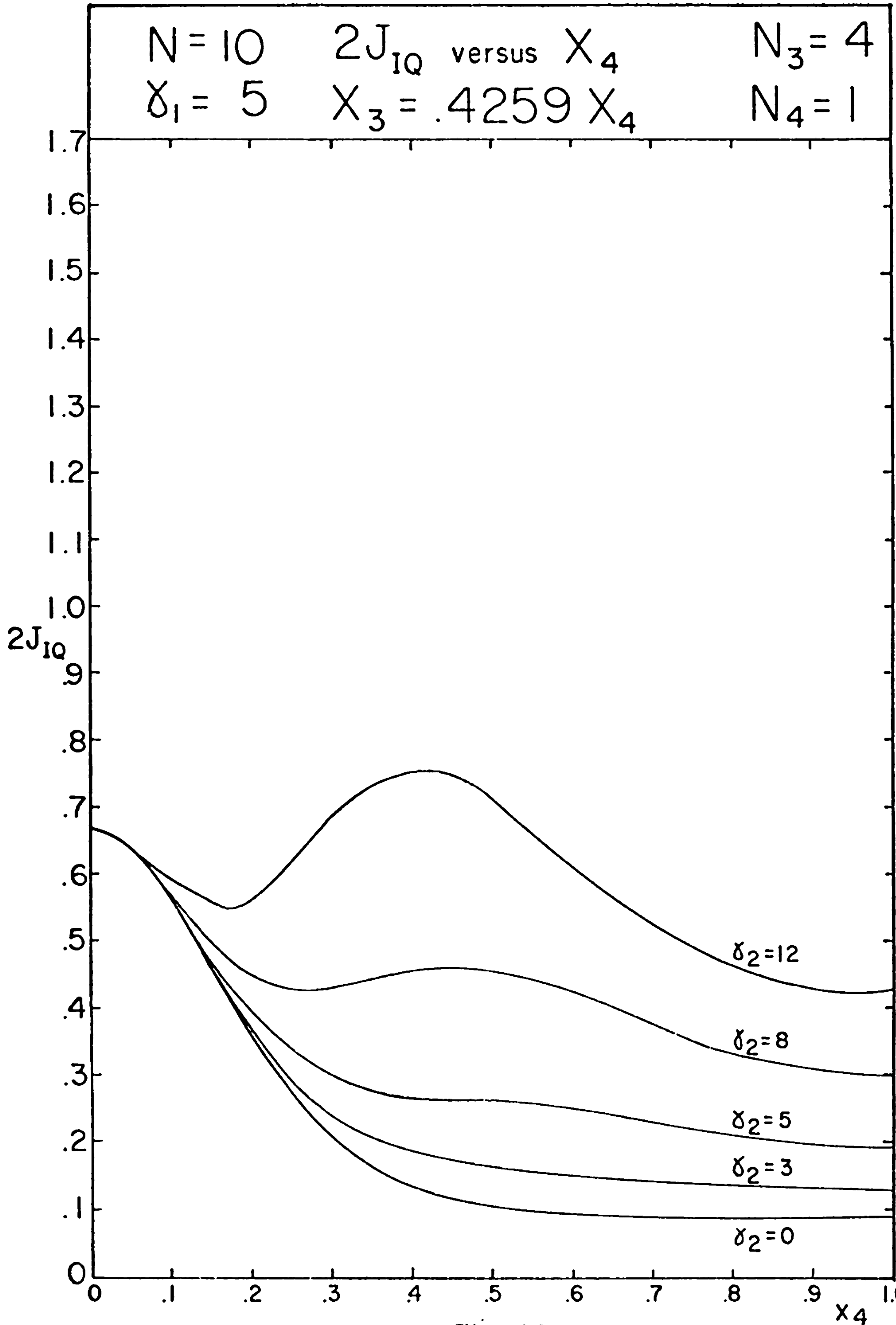

N = 10
2J_IQ versus X_4
N_3 = 4
γ_1 = 5
X_3 = .4259 X_4
N_4 = 1
2J_IQ
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
γ_2 = 12
γ_2 = 8
γ_2 = 5
γ_2 = 3
γ_2 = 0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
x_4

Fig. 14

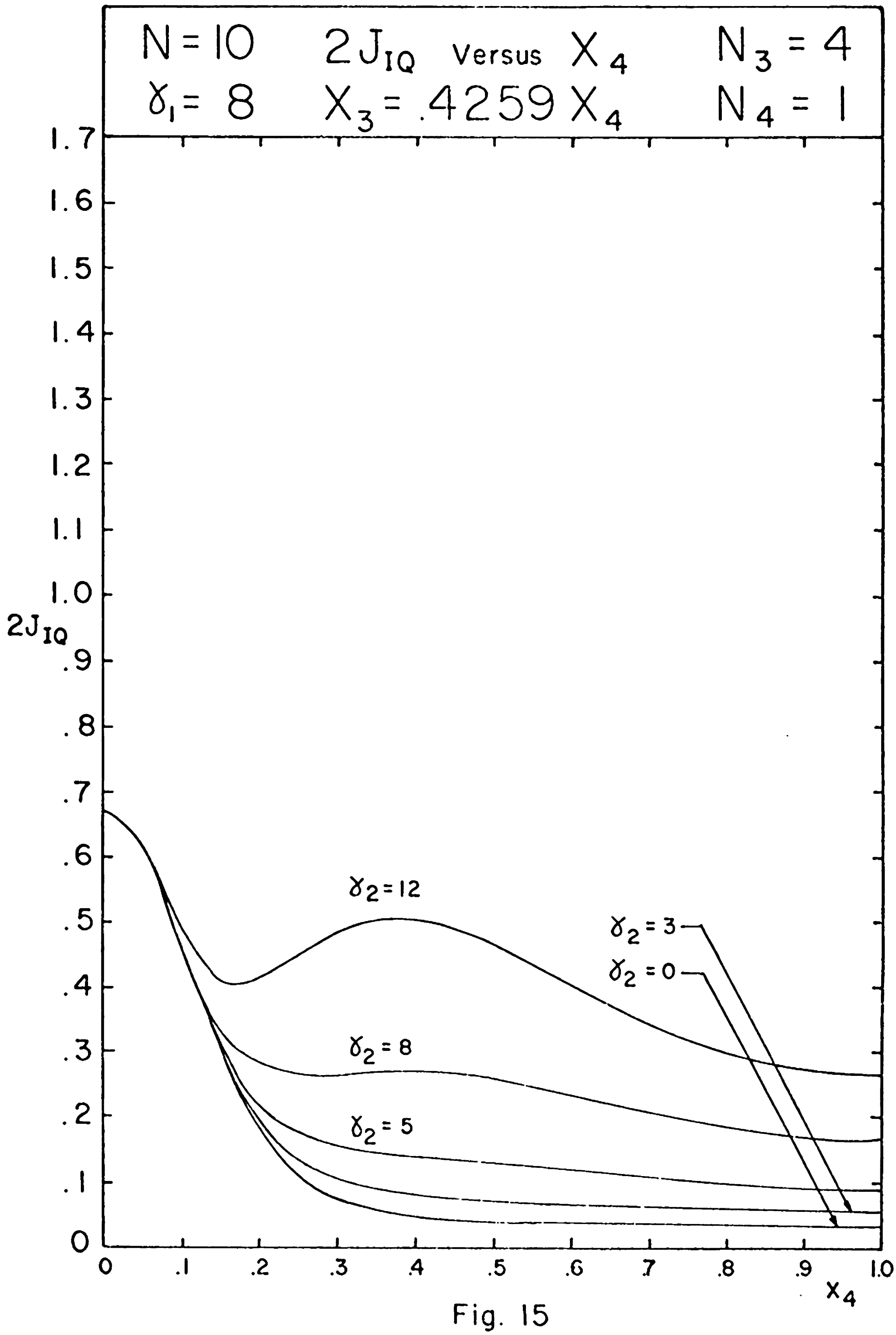

N = 10
2J_IQ versus X_4
N_3 = 4
γ_1 = 8
X_3 = .4259 X_4
N_4 = 1
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
2J_IQ
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
γ_2 = 12
γ_2 = 8
γ_2 = 5
γ_2 = 3
γ_2 = 0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
X_4

Fig. 15

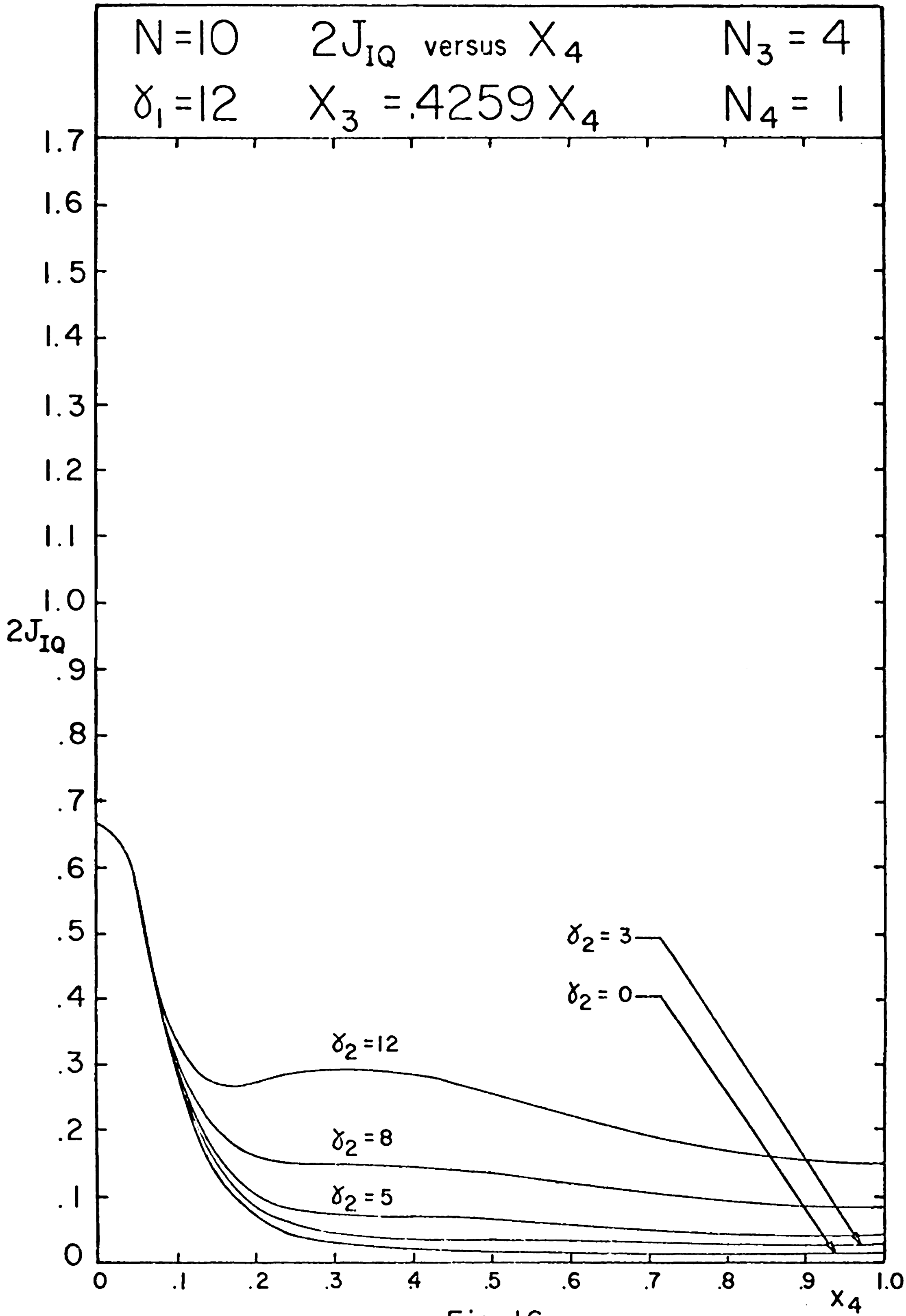

N = 10
2J_IQ versus X_4
N_3 = 4
γ_1 = 12
X_3 = .4259 X_4
N_4 = 1
2J_IQ
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
γ_2 = 3
γ_2 = 0
γ_2 = 12
γ_2 = 8
γ_2 = 5
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
X_4

Fig. 16

While the method of obtaining near optimal designs has been given and is straightforward for any value of N, it does require some effort in evaluating (4.56) repeatedly for the various candidates (N_3, N_4). Since a computer program has been written for this purpose, it was considered appropriate to list some near optimal designs for each value of N for the potential user. For N = 5(1)20(2)30(5)50, Tables VIII, IX, X, and XI show two near optimal "four-point" designs and two near optimal "five-point" designs for N even and four near optimal "five-point" designs for N odd since "four-point" near optimal designs do not exist for N odd due to lack of symmetry. For low values of N even, two "four-point" and two "five-point" designs do not exist, and for low values of N odd, four "five-point" designs do not exist and the listing is exhaustive. For larger values of N, the listing is far from exhaustive, and should the potential user have difficulty in attaining the listed design points experimentally, the method outlined in this section can be used to generate additional designs.

To attain near optimality, it has been shown that more than three distinct design points are required, the smallest number being four, and thus designs in this section have been developed for both four and five design points, the development being identical by letting $N_0 = 0$ for the "four-point" design and $N_0 > 0$ for the "five-point" design. These designs provide two and three degrees of freedom respectively for lack of fit as aforementioned. Should the experimenter feel that more degrees of freedom are needed for lack of fit, designs with more than five points can be developed which satisfy conditions (4.46) and (4.47) by simply extending the work in this section. For example, for

TABLE VIII

NEAR OPTIMAL DESIGNS

N = 5, 6, 7, 8, 9, 11, 12

N	N_0	N_3	x_3	N_4	x_4
5	1	1	.3409	1	.8468
6	0	2	.3313	1	.8835
	2	1	.4917	1	.8707
7	1	2	.4134	1	.9083
	3	1	.6423	1	.8684
8	0	2	.1308	2	.8060
	0	3	.3883	1	.9386
	2	2	.4895	1	.9242
	2	1	.1862	2	.8058
9	1	2	.2525	2	.8284
	1	3	.4419	1	.9561
	3	2	.5634	1	.9301
	3	1	.3672	2	.8262
10	See Table VII, page 99.				
11	1	3	.3337	2	.8658
	3	2	.4185	2	.8611
	3	3	.5430	1	.9741
	5	1	.6681	2	.8328
12	0	3	.1308	3	.8060
	0	4	.3313	2	.8835
	2	2	.1607	3	.8059
	4	3	.5924	1	.9732

TABLE IX

NEAR OPTIMAL DESIGNS

N = 13, 14, 15, 16, 17

N	N_0	N_3	x_3	N_4	x_4
13	1	4	.3734	2	.8969
	3	2	.2698	3	.8208
	5	2	.5647	2	.8744
	5	1	.3946	3	.8187
14	0	4	.2439	3	.8358
	0	5	.3651	2	.9129
	2	3	.2838	3	.8350
	4	3	.4888	2	.8990
15	1	4	.2918	3	.8484
	3	4	.4519	2	.9174
	5	2	.4287	3	.8431
	7	1	.7071	3	.8165
16	0	4	.1308	4	.8060
	0	6	.3883	2	.9386
	2	4	.3353	3	.8596
	4	4	.4895	2	.9242
17	1	4	.1993	4	.8177
	3	4	.3760	3	.8695
	5	4	.5265	2	.9286
	7	4	.6852	1	.9774

TABLE X

NEAR OPTIMAL DESIGNS

N = 18, 19, 20, 22, 24, 26

N	N_0	N_3	x_3	N_4	x_4
18	0	5	.2247	4	.8288
	0	7	.4053	2	.9617
	2	5	.3659	3	.8814
	4	4	.4148	3	.8779
19	1	7	.4285	2	.9699
	3	4	.2987	4	.8381
	5	4	.4523	3	.8847
	7	4	.6007	2	.9282
20	0	5	.1308	5	.8060
	0	8	.4186	2	.9827
	2	7	.4512	2	.9768
	4	5	.4289	3	.8970
22	0	6	.2113	5	.8245
	0	7	.3071	4	.8670
	2	5	.2323	5	.8242
	4	5	.3689	4	.8641
24	0	6	.1308	6	.8060
	0	7	.2570	5	.8411
	2	5	.1435	6	.8059
	4	5	.3074	5	.8399
26	0	7	.2014	6	.8215
	0	10	.4002	3	.9543
	2	7	.3100	5	.8556
	4	6	.3371	5	.8546

TABLE XI

NEAR OPTIMAL DESIGNS

N = 28, 30, 35, 40, 45, 50

N	N_0	N_3	x_3	N_4	x_4
28	0	7	.1308	7	.8060
	0	9	.3125	5	.8704
	2	8	.3330	5	.8694
	4	5	.1552	7	.8059
30	0	10	.3313	5	.8835
	0	9	.2741	6	.8489
	2	8	.2918	6	.8484
	6	5	.2473	7	.8189
35	1	10	.2813	7	.8487
	3	9	.2975	7	.8482
	5	10	.4134	5	.9083
	7	7	.3409	7	.8468
40	0	10	.1308	10	.8060
	0	13	.3183	7	.8742
	6	7	.1568	10	.8059
	10	10	.4895	5	.9242
45	1	16	.3841	6	.9255
	5	10	.2525	10	.8284
	9	8	.2841	10	.8279
	13	6	.3319	10	.8270
50	0	13	.1719	12	.8139
	0	18	.3690	7	.9167
	2	12	.1790	12	.8139
	10	10	.3409	10	.8468

six or seven design points with N_i observations at x_i, $i = 0, 1, \ldots, 6$ and $N_{i+1} = N_{6-i}$ as well as $x_{i+1} = -x_{6-i}$ for $i = 0, 1, 2$, by letting $x_4 = k_4 x_6$ and $x_5 = k_5 x_6$, one obtains the following equation analogous to (4.55):

$$\frac{N_4 k_4^4 + N_5 k_5^4 + N_6}{(N_4 k_4^2 + N_5 k_5^2 + N_6)^2} = \frac{3.8}{N} . \tag{4.61}$$

This is a quadratic equation in both k_4^2 and k_5^2 and can be solved for k_4^2, say, by assigning values to k_5^2, N_4, N_5, and N_6, retaining those values of k_5^2, N_4, N_5, and N_6 which yield satisfactory solutions for k_4^2 and discarding those which do not. This is clearly no job for a desk calculator, but with a high speed digital computer, it poses no problem.

(4.8) Cost of Protecting Against a Possible Quadratic Effect

In Chapter II it was shown that if the true model is linear, the optimal experimental design for the Inverse estimator is dependent on both γ_1 and N. However, it was also shown that the "end-point" design with N/2 observations at both $x_1 = -1$ and $x_2 = +1$ for N even and (N-1)/2 observations at both $x_1 = -1$ and $x_2 = +1$ and one observation at $x = 0$ for N odd, while not optimal, provides an average MSE which is smaller than the corresponding average MSE for the Classical estimator for all γ_1 and N. Hence, if the true model is linear and one does not have a very good a priori estimate of γ_1 (which is likely to be the case), one can use the "end-point" design, and be assured of obtaining a smaller average MSE than can be attained using the Classical. In this chapter it has been shown that if the true model is quadratic, designs which are very close to optimal for the Inverse estimator can be constructed by requiring that $\overline{x^2} = 1/3$ and $S_Q = N/10$. Since the existence of a quadratic effect is difficult to ascertain, the obvious questions to be answered are the following: If one protects against a possible quadratic effect in the selection of the design, what increase in the average MSE will result if the true model is linear? Also, if one does not protect against a possible quadratic effect in the selection of the design and a quadratic effect is present, what increase in the average MSE will result? These questions can be answered by examining both J_{IL} and J_{IQ} when designing for a linear model and when designing for a possible quadratic effect for selected values of N, γ_1, and γ_2.

Values of $2J_{IL}$ and $2J_{IQ}$ for both a linear design ("end-point") and a quadratic design ($\overline{x^2} = 1/3$ and $S_Q = N/10$) are shown in Table XII for twelve combinations of values of γ_1 and γ_2 for N = 10 and N = 24. Entries under $2J_{IL}^{(L)}$ are values of twice the average MSE obtained by designing for a linear model (as indicated by the superscript (L)) when the true model is linear. Likewise, entries under $2J_{IQ}^{(L)}$ are values of twice the average MSE obtained by designing for a linear model when the true model is quadratic. Entries under $2J_{IL}^{(Q)}$ are values of twice the average MSE obtained by designing for a quadratic model (as indicated by the superscript (Q)) when the true model is linear. And entries under $2J_{IQ}^{(Q)}$ are values of twice the average MSE obtained by designing for a quadratic model when the true model is quadratic. The "percent deviation from $2J_{IQ}^{(Q)}$" is the percent increase that will result if one designs for a linear model and the true model is really quadratic, that is, the cost of misclassifying a quadratic

TABLE XII

COST OF PROTECTION AND MISCLASSIFICATION

N	Y_1	Y_2	$2J_{IL}^{(L)}$	$2J_{IQ}^{(L)}$	% deviation from $2J_{IQ}^{(Q)}$	$2J_{IL}^{(Q)}$	% deviation from $2J_{IL}^{(L)}$	$2J_{IQ}^{(Q)}$
10	5	1	.0865	.127	41.9	.0856	-1.04	.0895
	5	3	.0865	.451	250.	.0856	-1.04	.129
	5	5	.0865	1.10	464.	.0856	-1.04	.195
	8	1	.0347	.0511	34.5	.0355	2.30	.0380
	8	3	.0347	.182	218.	.0355	2.30	.0572
	8	5	.0347	.443	380.	.0355	2.30	.0923
	8	8	.0347	1.08	563.	.0355	2.30	.163
	12	1	.0156	.0229	31.6	.0163	4.49	.0174
	12	3	.0156	.0817	206.	.0163	4.49	.0267
	12	5	.0156	.199	348.	.0163	4.49	.0444
	12	8	.0156	.485	479.	.0163	4.49	.0838
	12	12	.0156	1.07	609.	.0163	4.49	.151
24	5	1	.0796	.120	46.2	.0768	-3.52	.0821
	5	3	.0796	.439	257.	.0768	-3.52	.123
	5	5	.0796	1.08	468.	.0768	-3.52	.190
	8	1	.0322	.0485	39.8	.0322	.00	.0347
	8	3	.0322	.178	229.	.0322	.00	.0541
	8	5	.0322	.438	389.	.0322	.00	.0895
	8	8	.0322	1.07	565.	.0322	.00	.161
	12	1	.0145	.0218	37.1	.0147	1.38	.0159
	12	3	.0145	.0804	219.	.0147	1.38	.0252
	12	5	.0145	.198	360.	.0147	1.38	.0430
	12	8	.0145	.483	485.	.0147	1.38	.0825
	12	12	.0145	1.07	613.	.0147	1.38	.150

model as linear. Similarly, the "percent deviation from $2J_{IL}^{(L)}$" is the percent increase that will result if one designs for a quadratic model and the true model is linear, that is, the cost of protecting against a possible quadratic effect. For example, consider the case $N = 10$, $\gamma_1 = 8$, and $\gamma_2 = 5$. If one designs for a linear model and the true model is linear with $\gamma_1 = 8$, twice the average MSE is .0347 while if the true model is quadratic with $\gamma_1 = 8$ and $\gamma_2 = 5$, twice the average MSE is .443. If one designs for a quadratic model and the true model is linear with $\gamma_1 = 8$, twice the average MSE is .0355 while if the true model is quadratic with $\gamma_1 = 8$ and $\gamma_2 = 5$, twice the average MSE is .0923. The percentages involved are calculated as follows: If one designs for a linear model and it is quadratic, twice the average MSE is .443 but had the design been for a quadratic model, twice the average MSE would be .0923 so misclassification costs an increase of 380% of .0923. However, if one designs for a quadratic model and it is linear, twice the average MSE is .0355 but had the design been for a linear model, twice the average MSE would be .0347 so the cost of protection is only 2.3% above .0347. This example is quite typical of the results in general as exhibited in Table XII, namely, the cost of protecting against a possible quadratic effect is very small in comparison with the cost of misclassifying a quadratic model as linear.

The cost of misclassifying depends on the magnitude of γ_2. If γ_2 is small, of course, the cost is small, but even for a γ_2 as small as one, the cost of misclassifying is at least 30% and as γ_2 increases, the cost increases rapidly into the hundreds of percent. On the other hand, the cost of protection depends on the magnitude of γ_1. For small values of γ_1, the cost is actually negative, that is, one obtains a decrease in the average MSE if the model is linear. This appears to be unrealistic, but is is easily explained. Recall that for linear designs, the "end-point" design was used which is not optimal, the optimal design for small γ_1 being closer to the design center. By protecting against a quadratic effect by requiring that $\overline{x^2} = 1/3$, one is actually closer to the optimal design than with the "end-point" design and hence the negative cost. As γ_1 increases, the cost of protection also increases but not nearly as rapidly or as high as the cost of misclassification. Table XII shows a maximum cost of 4.49% but for small values of N and large values of γ_1, the cost will be greater than this quantity. The maximum costs detected in this investigation at $\gamma_1 = 50$ were 9.30% for N = 6, 5.61% for N = 10, 3.89% for N = 15, 2.25% for N = 25, and 1.34% for N = 50.

The results of this section are quite clear. Unless γ_1 is excessively large, the cost of protecting against a possible quadratic effect ranges from a small negative percent to about five percent. However, the cost of misclassifying a quadratic model as linear can easily be several hundred percent for moderate values of γ_2. These results lead to the recommendation that unless one is very certain that his model is linear, the designs developed in this section should be used.

In Chapter II, it was shown that the "end-point" design for the Inverse estimator provides an average MSE which is less than the corresponding average MSE for the truncated Classical estimator for all γ_1 and N <u>if the true model is linear</u>. In the next chapter, the two estimators will be compared under the assumption that <u>the true model is quadratic</u>.

CHAPTER V

COMPARISON OF AVERAGE MSE'S - TRUE MODEL QUADRATIC

In order to compare the average MSE of the Inverse estimator with that of the truncated Classical when the true model is quadratic as given by (4.1), it will first be necessary to derive the average MSE for the truncated Classical estimator under the quadratic assumption. Using the same reasoning employed to develop an expression for J_{CL} in Chapter III, it is easily seen that the average MSE for the truncated Classical estimator is the same as for the Inverse estimator in (4.12) with the moments of d replaced by the moments of 1/b. Hence, one can immediately write

$$2J_{CQ} = \int_{-1}^{1} E_Q(\hat{x}_{CL} - x)^2 \, dx$$

$$= 2/3 - E(1/b)\ (4/3)\ \beta_1 + E(1/b^2)\ [2\sigma^2(1+1/N)$$

$$+ (2/3)\ \beta_1^2 + \beta_2^2(2/5 - 4\overline{x^2}/3 + 2\overline{x^2}^2)] \qquad (5.1)$$

for symmetrical designs where the subscript CQ on J indicates that the Classical estimator is under consideration and that the true model is quadratic. As in Chapter III, it is meaningless to work with expression (5.1) since the moments of 1/b do not exist. However, using the reasoning of Ott and Myers, one can replace the moments of 1/b with the moments, $E_T(1/b^r)$, r = 1, 2, and thus obtain a meaningful expression for $2J_{CQ}$. It is interesting to note that in the Inverse case, the moments of d change with the inclusion of the quadratic term, but in the Classical case under consideration, the moments of 1/b remain the same.

This is because for symmetrical designs, the density of b is the same for both models. Hence, the expression for the moments of 1/b remain the same as in Chapter III, i.e., the expression given in (3.5). Incorporating (3.5) for $r = 1, 2$ into (5.1) it is seen that (5.1) contains the single design variable $\overline{x^2}$. One must now minimize (5.1) with respect to $\overline{x^2}$.

In their minimization of (5.1), Ott and Myers used the following approximations for the moments of 1/b over the truncated density of b:

$$E_T(1/b) \cong 1/\beta_1 + \sigma^2/(N\overline{x^2}\beta_1^3) \qquad (5.2)$$

$$E_T(1/b^2) \cong 1/\beta_1^2 + 3\sigma^2/(N\overline{x^2}\beta_1^4) + 3\sigma^4/(N^2\overline{x^2}^2\beta_1^6) \ . \qquad (5.3)$$

These approximations can be obtained by expanding 1/b into a second order Taylor series and using term by term expectations in the expansion for the first moment and term by term expectations in the square of the expansion for the second moment. The details are given in Ott (1966). Substituting (5.2) and (5.3) into (5.1) and letting $u = \overline{x^2}$, $\gamma_1 = |\beta_1/\sigma|$, and $\gamma_2 = |\beta_2/\sigma|$, one obtains the following approximation for the average MSE for the truncated Classical estimator:

$$\begin{aligned} 2J_{CQ} \cong\ & 2/(3Nu\gamma_1^2) + 2(1+1/N)/\gamma_1^2 \\ & + \gamma_2^2(2/5 - 4u/3 + 2u^2)/\gamma_1^2 + 6(1+1/N)/(Nu\gamma_1^4) \\ & + 3\gamma_2^2(2/(5u) - 4/3 + 2u)/(N\gamma_1^4) \\ & + 6(1+1/N)/(N^2u^2\gamma_1^6) + 2/(N^2u^2\gamma_1^4) \\ & + 3\gamma_2^2(2/(5u^2) - 4/(3u) + 2)/(N^2\gamma_1^6) \ . \end{aligned} \qquad (5.4)$$

It is noted that after substituting $\gamma_1 = |\beta_1/\sigma|$ into (5.4), β_1, β_2, and σ vanish so that, similar to the situation of J_{IQ}, β_1, β_2, and σ can be considered with the two parameters γ_1 and γ_2. Equation (5.4) corresponds to equation (18) in Ott and Myers. However, they represented the quadratic model in terms of Legendre polynomials by writing

$$y_i = c_0 + c_1P_1(x_i) + c_2P_2(x_i) + \varepsilon_i$$

where

$$P_1(x_i) = x_i$$

$$P_2(x_i) = (3x_i^2 - 1)/2$$

so that in order to note that their equation (18) is identical to (5.4) above, it is necessary to relate the regression coefficients of the Legendre polynomial representation to the regression coefficients in (4.1). By letting $c_1 = \beta_1$, $c_2 = 2\beta_2/3$, and $\beta_1^2 = \gamma_1^2\sigma^2$ in their (18), one sees that (18) and (5.4) are identical. To minimize (5.4) with respect to the design variable $\overline{x^2} = u$, it is necessary to differentiate (5.4) with respect to u and solve the resultant set equal to zero. After differentiation and simplification, one obtains

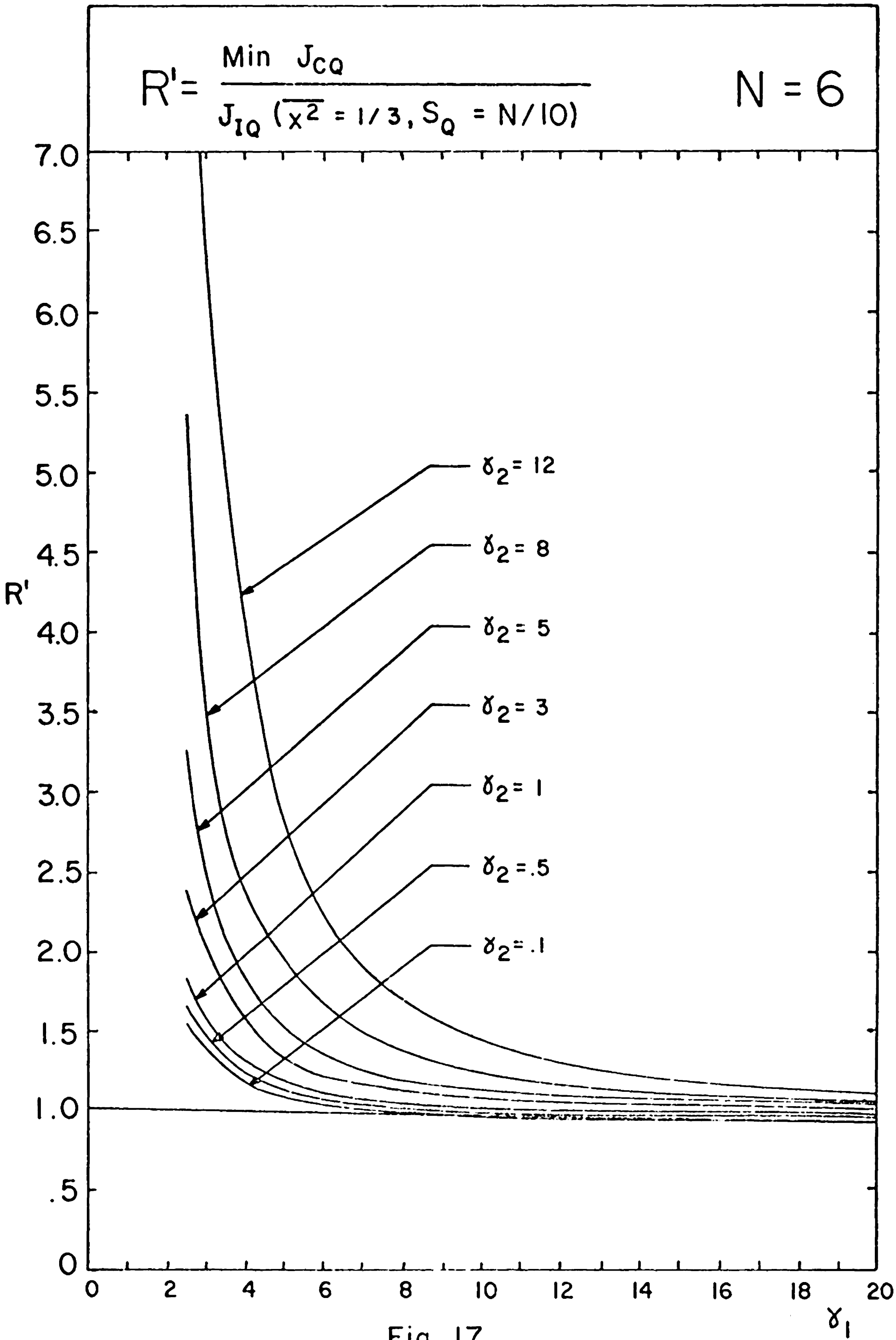
$$\begin{aligned}
&(4\gamma_2^2/\gamma_1^2)u^4 + [6\gamma_2^2/(N\gamma_1^4) - 4\gamma_2^2/(3\gamma_1^2)]u^3 \\
&+ [4\gamma_2^2/(N^2\gamma_1^6) - 6\gamma_2^2/(5N\gamma_1^4) - 6(1+1/N)/(N\gamma_1^4) \\
&- 2/(3N\gamma_1^2)]u - 12(1+1/N)/(N^2\gamma_1^6) - 4/(N^2\gamma_1^4) \\
&- 12\gamma_2^2/(5N^2\gamma_1^6) = 0 \,. \qquad (5.5)
\end{aligned}$$

This equation is of exactly the same form as (19) in Ott and Myers if one makes the necessary correspondence between regression coefficients and between β_1, σ, and γ_1 and divides each term by $4\gamma_2^2/\gamma_1^2$. Ott and Myers have shown that for a given set of parameters, (19) and hence (5.5) above either has a unique positive root less than or equal to one which minimizes (5.4) or $2J_{CQ}$ in (5.4) is a decreasing function of u in the interval [0, 1]. Therefore, to obtain optimal designs for the truncated Classical estimator, one must first solve (5.5) for the unique positive value of $u = \overline{x^2}$ and take $u_0 = u$ if $0 < u \leq u' \leq 1$ and for $u > u'$, take $u_0 = u'$ where u' is largest value less than or equal to one which will insure that all design points are within the region of interest [-1, 1]. The resulting value, u_0, can then be used to construct optimal designs as follows. For a "two-point" design for N even, take N/2 observations at each $x_1 = -x_2$ and x_2 where $x_2 = u_0^{1/2}$. Note that if $u > u' = 1$ here, a value of $u_0 = u' = 1$ will insure that x_2 lies in the region of interest. For a "three-point" design with a single center point for N odd, take (N-1)/2 observations at each $x_1 = -x_2$ and x_2 where $x_2 = [Nu_0/(N-1)]^{1/2}$ and one observation at $x = 0$. The above value of x_2 is reasoned by merely expressing $\overline{x^2}$ as $(N-1)x_2^2/N$ and solving $(N-1)x_2^2/N = u_0$ for x_2. It is noted that if N is odd, a value of $u_0 > u' = (N-1)/N$ results in a value of $x_2 > 1$. Hence, in this case, for values of $u > (N-1)/N$, one must take $u_0 = (N-1)/N$ to insure that the resulting value of x_2 lies within the region of interest. While the concern of this chapter is not obtaining specific optimal designs for the truncated Classical estimator (Ott and Myers have done this) but to determine the minimum value that $2J_{CQ}$ can attain, it is noticed that if one uses more than two distinct design points for N even or more than three distinct design points for N odd and the solution to (5.5) is large, u_0 must be set at decreasingly smaller values to insure that each design point lies in [-1, 1]. Hence, using more than two design points for N even or more than three for N odd could result in a value of u_0 smaller than optimal and hence a value of $2J_{CQ}$ larger than could be attained by using the "two-point" (N even) or "three-point" (N odd) design. In the comparison which follows, the minimum value of $2J_{CQ}$, Min $2J_{CQ}$, will be obtained by substituting u_0 in (5.4) where u_0 is taken as one if the root u is greater than one and N is even and where u_0 is taken as (N-1)/N if the root u is greater than (N-1)/N and N is odd. This is the smallest possible value that $2J_{CQ}$ can attain for N even and N odd respectively.

In Chapter IV, it was shown that designs very close to optimal can be obtained for the Inverse estimator <u>without prior knowledge of the parameters γ_1 and γ_2</u>. However, Ott (1966) has stated that optimal designs for the truncated Classical estimator under the quadratic assumption <u>depend on partial prior knowledge of γ_1 and γ_2</u>. To compare the estimators under the quadratic model assumption, the value of J_{IQ} for the Inverse estimator obtained by using the near optimal designs developed in Chapter IV which <u>do not depend on γ_1 and γ_2</u> will be compared with the smallest possible value of J_{CQ} that can be attained for the truncated Classical estimator by <u>assuming full prior knowledge of γ_1 and γ_2</u>. The measure of comparison will be R', where

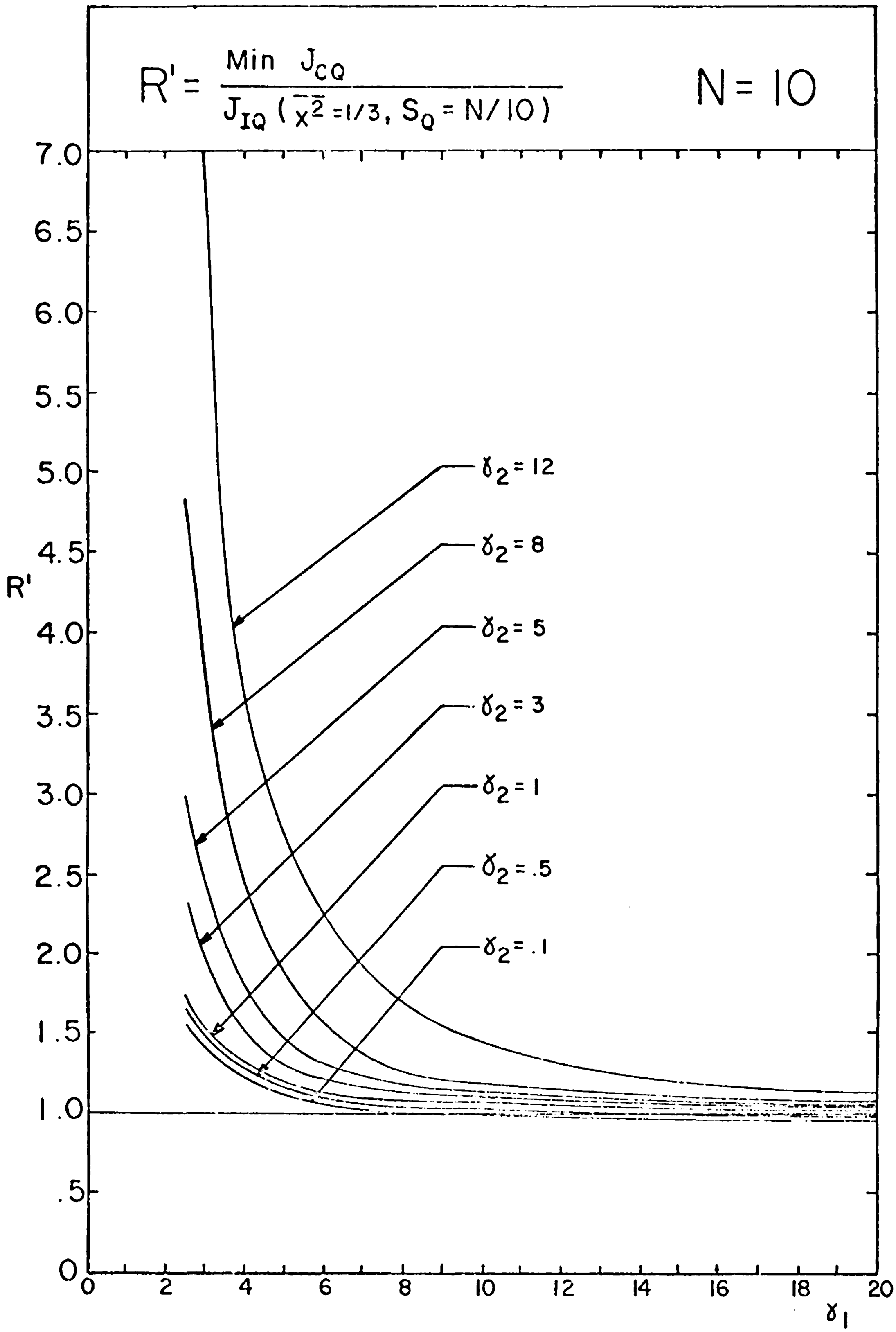
$$R' = \frac{\text{Min } J_{CQ}}{J_{IQ}(\overline{x^2} = 1/3,\ S_Q = N/10)} . \qquad (5.6)$$

The quantity R' was computed for $N = 6, 10, 15, 25, 50$; $\gamma_1 = 2.5, 5, 8, 12, 16, 20$; and $\gamma_2 = .1, .5, 1, 3, 5, 8, 12$, and the results are set forth in Figures 17-21 where R' is plotted as a function of γ_1 for fixed values of N and γ_2. For each value of the parameters N, γ_1, and γ_2, Min $2J_{CQ}$ was obtained by first solving (5.5) by the Newton-Raphson procedure on the IBM 360 digital computer for the optimal value of u and then substitution of this value into (5.4). Values of $2J_{IQ}(\overline{x^2} = 1/3, S_Q = N/10)$ were taken from the previous computations in Chapter IV. The results of Figures 17-21 deserve some comments. Without any prior knowledge of the model parameters, the Inverse estimator is markedly superior, the ratio R' being greater than one, for all N considered, $\gamma_1 \leq 7$, and $\gamma_2 \geq .1$, the superiority becoming more pronounced as γ_1 decreases. For values of $\gamma_1 > 7$, small N and small γ_2, the ratio R' is slightly less than one, but as N increases, R' increases with γ_1 until for $N = 50$, R' is equal to 1.00 for $\gamma_1 = 20$ for values of γ_2 as small as .1. The reason R' is less than one for small γ_2 and large γ_1 is that the model is essentially linear and the truncated Classical estimator, assumed to have full prior knowledge of the model parameters, selects a design which is either "end-point" or very nearly "end-point," the optimal design for the linear model, while the Inverse estimator, assuming no prior knowledge, maintains the design for the quadratic model by requiring $\overline{x^2} = 1/3$ and $S_Q = N/10$. This situation was anticipated in Chapter III where it was forwarned that <u>if the true model is linear</u>, any design other than the "end-point" design for the Inverse estimator could yield as average MSE greater than the minimum average MSE for the truncated Classical estimator for large enough γ_1. However, it must be realized in this comparison that in a practical



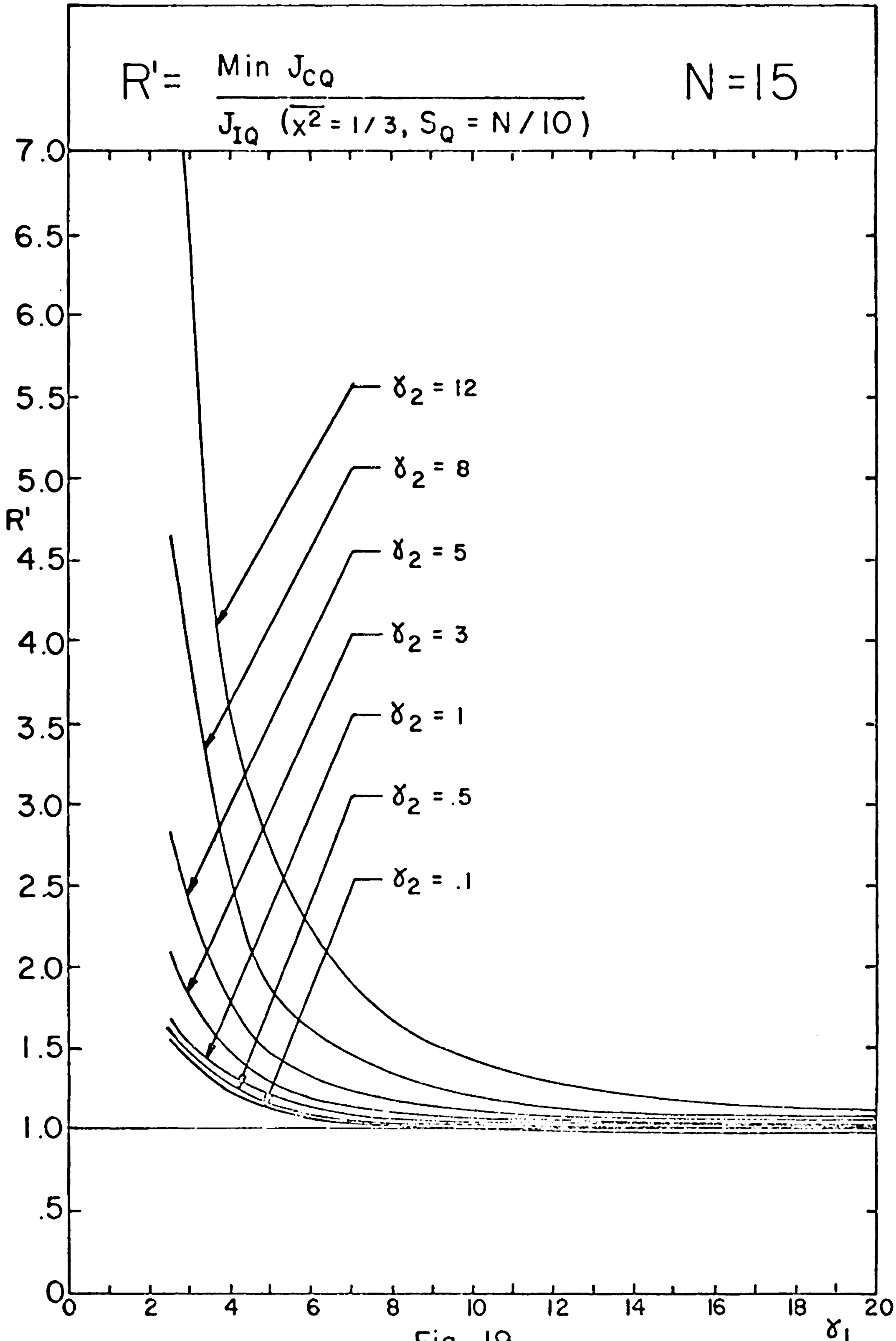
$R' = \frac{\text{Min } J_{CQ}}{J_{IQ}(\overline{x^2} = 1/3,\ S_Q = N/10)}$
N = 6
7.0
6.5
6.0
5.5
5.0
4.5
R'
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0
0
2
4
6
8
10
12
14
16
18
20
γ_1
$\gamma_2 = 12$
$\gamma_2 = 8$
$\gamma_2 = 5$
$\gamma_2 = 3$
$\gamma_2 = 1$
$\gamma_2 = .5$
$\gamma_2 = .1$

Fig. 17



$R' = \frac{\text{Min } J_{CQ}}{J_{IQ}(\overline{x^2} = 1/3,\ S_Q = N/10)}$
N = 10
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0
R'
$\gamma_2 = 12$
$\gamma_2 = 8$
$\gamma_2 = 5$
$\gamma_2 = 3$
$\gamma_2 = 1$
$\gamma_2 = .5$
$\gamma_2 = .1$
0 2 4 6 8 10 12 14 16 18 20
γ_1

Fig. 18



R'= Min J_CQ / J_IQ (x̄² = 1/3, S_Q = N/10)
N = 15
γ2 = 12
γ2 = 8
γ2 = 5
γ2 = 3
γ2 = 1
γ2 = .5
γ2 = .1
R'
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0
0
2
4
6
8
10
12
14
16
18
20
γ1

Fig. 19

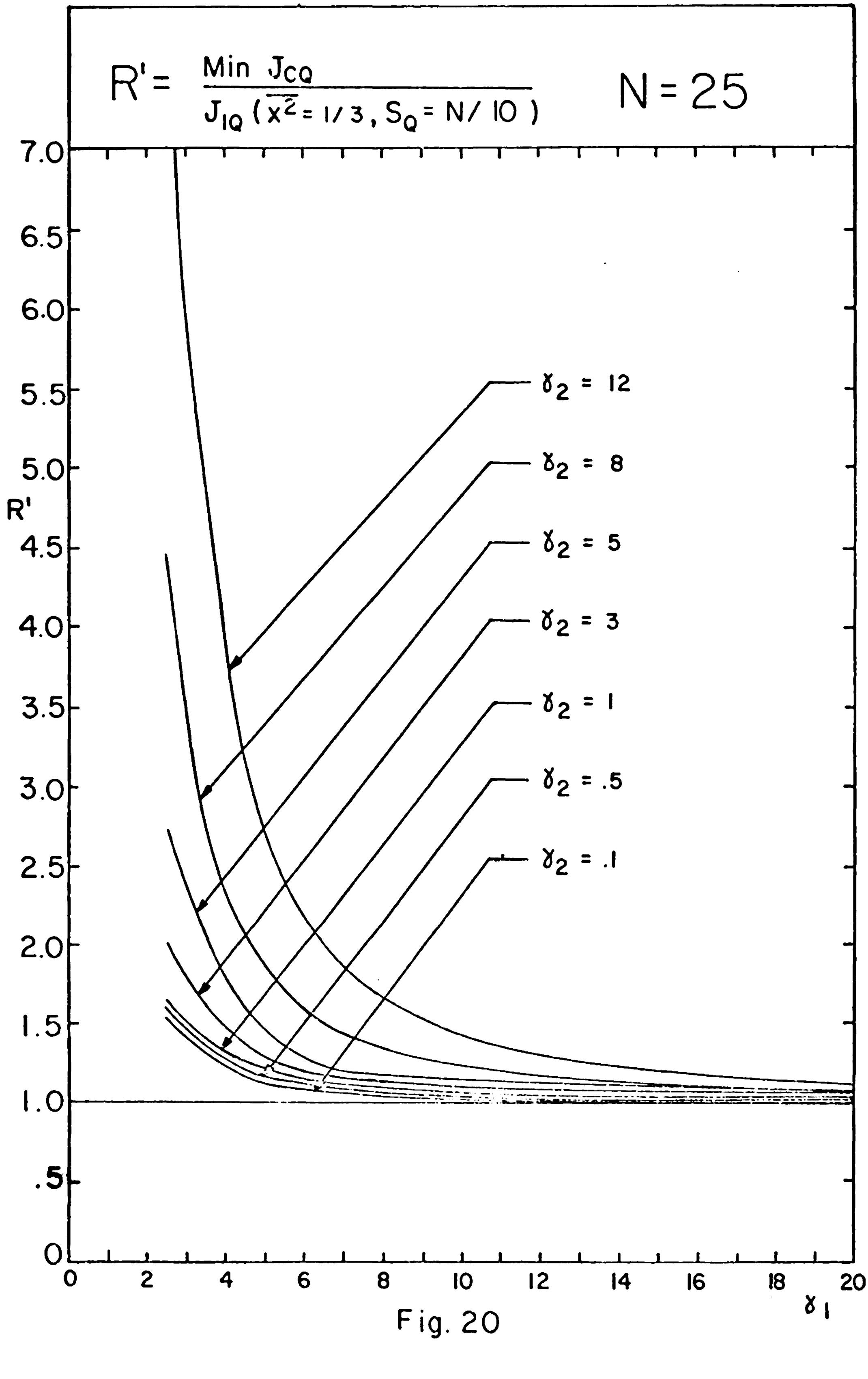

R' = Min J_CQ / J_1Q (x̄² = 1/3, S_Q = N/10)
N = 25
7.0
6.5
6.0
5.5
5.0
R'
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0
γ2 = 12
γ2 = 8
γ2 = 5
γ2 = 3
γ2 = 1
γ2 = .5
γ2 = .1
0
2
4
6
8
10
12
14
16
18
20
γ1

Fig. 20

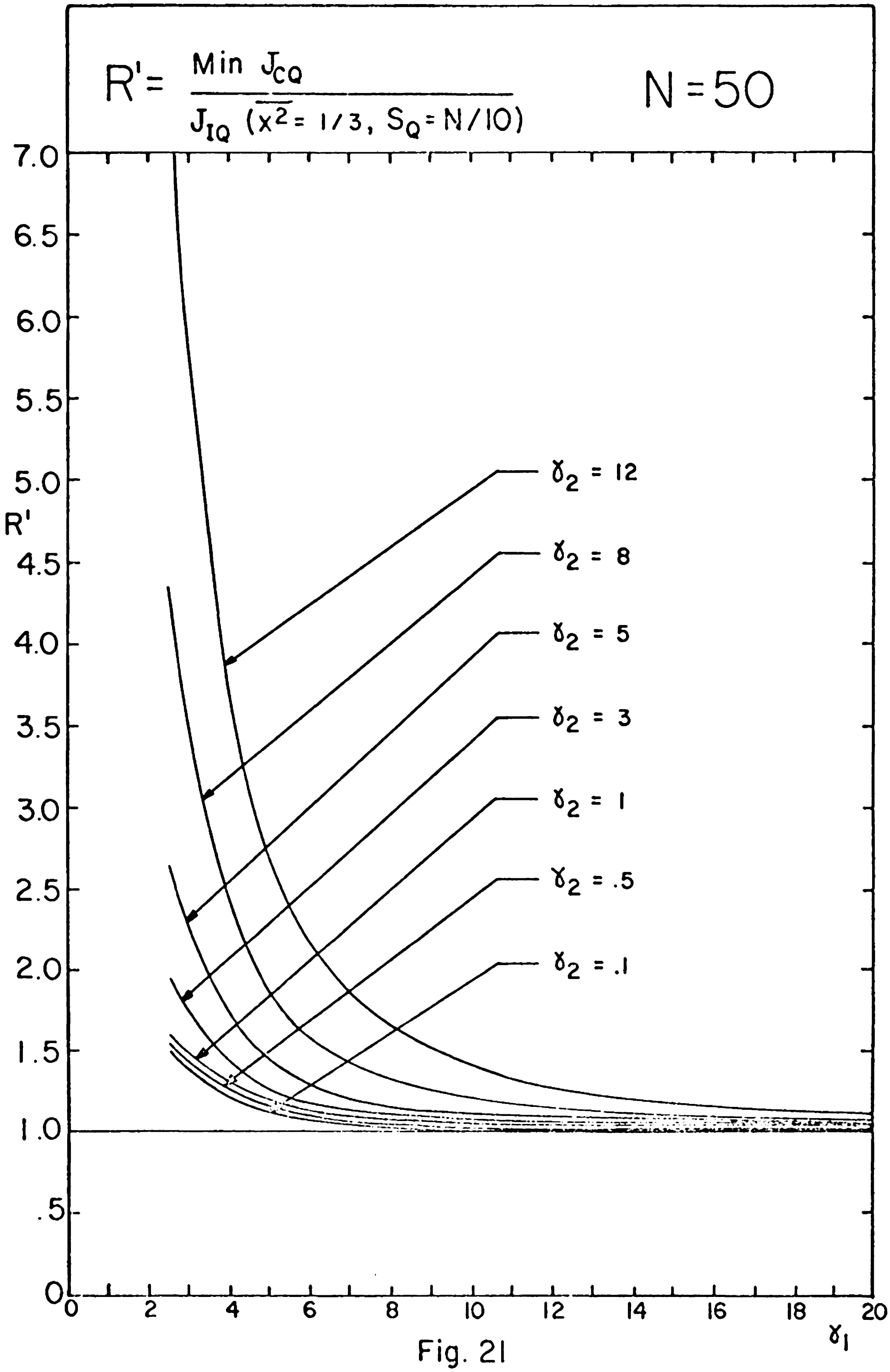

R' = Min J_CQ / J_IQ (x̄² = 1/3, S_Q = N/10)
N = 50
7.0
6.5
6.0
5.5
5.0
R'
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0
0
2
4
6
8
10
12
14
16
18
20
γ₁
γ₂ = 12
γ₂ = 8
γ₂ = 5
γ₂ = 3
γ₂ = 1
γ₂ = .5
γ₂ = .1

Fig. 21

situation, one would certainly not have full prior knowledge of the model parameters to select designs for the truncated Classical estimator so that Min J_{CQ} could never be attained. This means that in practice, the value of R' will be larger than shown in Figures 17-21, quite possibly large enough to be greater than one for all N, γ_1, and γ_2. Also, if one does have some prior information which indicates that γ_1 is large and γ_2 is small and which enables him to select designs close to optimal for the truncated Classical estimator, then this information could be used for the Inverse estimator as well to select a design for which $\overline{x^2}$ is larger than 1/3, $\overline{x^2}$ being the dominant design variable in this case since the value of S_Q has nearly no effect on J_{IQ} for small γ_2.

While the Inverse estimator is markedly superior except for large γ_1 and small γ_2, to declare that the Inverse estimator is superior to the truncated Classical estimator for all values of the quadratic model parameters would perhaps be too strong although this author believes this to be the case. The reason such a statement is not possible is that while designs for the Inverse estimator which are very close to optimal can be obtained by simply taking $\overline{x^2} = 1/3$ and $S_Q = N/10$, designs for the truncated Classical estimator depend on prior information and to some extent upon the skill of the experimenter in making prior point estimates of the parameters. These nebulous quantities are impossible to reflect in a graph and all one can do is, as has been done, examine the very best that could possibly be achieved with the truncated Classical estimator with perfect prior information realizing that the value of J_{CQ} achieved in practice would be somewhat greater than Min J_{CQ}. If it can be agreed than one cannot get closer to Min J_{CQ} than 7% when γ_1 is large and γ_2 is small, then the Inverse estimator is superior for all γ_1 and N examined since in no case was R' less than .93. On the other hand, if one believes he can get closer to Min J_{CQ} than 7% for large γ_1 and small γ_2, it cannot be stated that the truncated Classical is best for these values of the γ_1 because the same prior information which enabled one to get within 7% of Min J_{CQ} could be used to select a value of $\overline{x^2}$ greater than 1/3 and hence decrease J_{IQ}. These statements plus the ease with which near optimality for the Inverse estimator can be attained without prior information and the relatively low cost of protecting against a quadratic effect point toward superiority of the Inverse estimator for all quadratic model parameters but as aforementioned a declaration of this kind will not be made. The strongest statement that can be made is that unless the model is very nearly linear with a large γ_1, the Inverse estimator is markedly superior for the quadratic model assumption, but for the case where the model is nearly linear with a large γ_1 the better of the two estimators is too close to ascertain.

As in the linear comparisons, the above results are substantiated by Krutchkoff's Monte Carlo study. His comparisons of the average squared errors under the quadratic model assumption were made with the scaled region of interest R = [0, 1] using a "two-point" symmetrical design for N = 6 with $x_1 = .15$ and $x_2 = .85$ for both estimators. Examining the cases $\beta_1 = .5$, $\sigma = .1$, ten values of β_2 from -.5 to 10.0, and x = 0(.2)1.2, 2, 5, 10, his results show that with the exception of the point x = 0, the Inverse estimator has, in general, a smaller average squared error than the truncated Classical.

It is interesting to note in Krutchkoff's results the effect of using $R = [0, 1]$ in examining the effect of a quadratic term. When using $R = [0, 1]$, the effect of a positive quadratic term tends to decrease the average squared errors for both estimators while the effect of a negative quadratic term tends to increase them. As pointed out by Krutchkoff, the reason for this is that the positive quadratic term, in effect, increases the slope of the approximating curve while the negative quadratic term tends to decrease it. As has been demonstrated, using the concept of an average MSE, this is not the case with $R = [-1, 1]$, for in this case the quadratic term increases the average MSE independent of the sign of β_2. Some insight can be gained on the effect of using $R = [0, 1]$ from an average MSE standpoint by integrating $E_Q(\hat{x}_{IL} - x)^2$ in (4.10) and $E(\hat{x}_{IL} - x)^2$ in (2.12) over $x = [0, 1]$. For symmetrical designs ($\bar{x} = .5$), one obtains

$$\int_0^1 E(\hat{x}_{IL} - x)^2\, dx = 1/12 - E(d)\ \beta_1/6 + E(d^2)[\sigma^2(1+1/N) + \beta_1^2/12] \qquad (5.7)$$

$$\int_0^1 E_Q(\hat{x}_{IL} - x)^2\, dx = 1/12 - E_Q(d)\ (\beta_1 + \beta_2)/6 + E_Q(d^2)[\sigma^2(1+1/N) + (\beta_1 + \beta_2)^2/12 + \beta_2^2(7/60 - 2\overline{x^2}/3 + \overline{x^2}^2)] \ . \qquad (5.8)$$

Equation (5.7) above is the average MSE for a linear approximation to a linear model using $R = [0, 1]$ while (5.8) is the average MSE for a linear approximation to a quadratic model, both for the Inverse estimator. It is observed that if β_1 and β_2 are of the same sign, say positive, then increasing β_2 in (5.8) has nearly the same effect as increasing the slope in (5.7) and will, hence, decrease the average MSE. It is also observed that if β_1 and β_2 are of opposite signs, say $\beta_1 > 0$ and $\beta_2 < 0$, then increasing β_2 negatively in (5.8) has nearly the same effect as decreasing the slope in (5.7) and will, hence, increase the average MSE. These results also hold for the truncated Classical estimator by replacing the moments of d in (5.7) and (5.8) with $E_T(1/b^r)$, $r = 1, 2$. Hence, had the scaled region of interest $R = [0, 1]$ been used in this thesis rather than $R = [-1, 1]$, one would obtain results for the average MSE's comparable with Krutchkoff's results for the average squared errors, that is, a decrease in average MSE for a positive β_2 and an increase in average MSE for a negative β_2 for both estimators. It should be noted, however, that the coefficient of β_2^2 in (5.8) is, apart from the leading constant term, identical to the coefficient of β_2^2 in the expression for $2J_{IQ}$ in (4.12) using $R = [-1, 1]$. As in (4.12), this term attains its minimum when $\overline{x^2} = 1/3$. It is clear that as β_2 increases either <u>positively or negatively</u>, this term will become important so that as in minimizing $2J_{IQ}$ in (4.12) using $R = [-1, 1]$, it appears that (5.8) would be minimized when $\overline{x^2} = 1/3$. One notes that Krutchkoff's design for examining the effect of a quadratic term was one for which $\overline{x^2} = .3725$ and was, thus, close to optimal. Hence, it is felt that had $R = [0, 1]$ been used in this thesis to examine the effect of a quadratic term, there would have been little change in the resulting optimal designs. Also, the agreement between Krutchkoff's results (using $R = [0, 1]$) and the results of this chapter (using $R = [-1, 1]$) concerning the comparison of estimators under the quadratic model assumption indicates that had $R = [0, 1]$ been used, the results of this chapter would be basically the same.

CHAPTER VI

COMPARISON OF DESIGN RESULTS FOR INVERSE REGRESSION WITH THOSE OF FORWARD REGRESSION

The preceding five chapters have been devoted to deriving optimal experimental designs for the Inverse estimator and comparing the resulting average MSE's with those of the truncated Classical estimator. Before summarizing the results of these chapters, it was considered appropriate to briefly discuss optimal designs for forward regression with a single independent variable and attempt to draw some parallels between the forward and inverse methods. Suppose the true model is linear as given by (1.2) and suppose one wishes to estimate y for a given value of the independent variable x. It is well known from elementary regression theory that the least squares procedure provides one with

$$\hat{y} = a + bx = \bar{y} + b(x-\bar{x}) \tag{6.1}$$

where $b = \hat{\beta} = S_{xy}/S_{xx}$. It is also well known that this least squares estimator is unbiased if the true model is linear so in this case the mean squared error for $\hat{y}$ is simply the variance of $\hat{y}$. It is easily shown that

$$\begin{aligned}\mathrm{Var}(\hat{y}) &= \sigma^2/N + \sigma^2(x-\bar{x})^2/S_{xx} \\ &= (\sigma^2/N)\ [1 + (x-\bar{x})^2/\overline{x^2}]\ .\end{aligned} \tag{6.2}$$

Therefore,

$$2J_{FL} = \int_{-1}^{1} E(\hat{y}-y)^2 \, dx$$

$$= (\sigma^2/N) \int_{-1}^{1} [1 + (x-\bar{x})^2/\overline{x^2}] \, dx$$

$$= (2\sigma^2/N) \; [1 + (1 + 3\bar{x}^2)/(3\overline{x^2})] \tag{6.3}$$

where the subscript FL on J indicates that forward regression is under consideration and that the true model is linear. It is quite clear that in order to minimize J_{FL}, one must minimize $\bar{x}$ and maximize $\overline{x^2}$. This means that one must restrict oneself to designs for which $\bar{x} = 0$ and for which $\overline{x^2} = 1$, the maximum value of $\overline{x^2}$ when the x_i are confined to the region of interest [-1, 1], in order to attain optimality. This essentially means that the optimal design for the forward case when the true model is linear is the two or three point "end-point" design depending on whether N is even or odd. This was exactly the same conclusion reached for inverse regression using the truncated Classical estimator, and while the "end-point" design for the Inverse estimator is not optimal, it was shown that it is the only design which could be used without prior information which would insure an average MSE which is always less than or equal to the truncated Classical average MSE for all values of $\gamma = |\beta/\sigma|$.

Consider now the case where the true model is quadratic as given by (4.1) but where one is using the linear approximation given by (6.1) to estimate y due to a model misclassification. Since optimal designs for the linear model are among the infinite class of designs for which

$\bar{x} = 0$, as in the inverse case, designs for the quadratic model will be further confined to the class of symmetrical designs which provide not only $\bar{x} = 0$ but also all odd design moments equal to zero. It is clear that the average or integrated variance will remain the same as in the linear case, namely, the expression given by (6.3) but with $\bar{x} = 0$ due to the confinement to symmetrical designs. This is because $\mathrm{Var}(\hat{y})$ is dependent on the variance of the experimental errors and not on the structure of the true model. Hence, it remains to find an expression for B, the average or integrated bias squared. By definition,

$$2B = \int_{-1}^{1} [E_Q(\hat{y}) - y]^2 \, dx \tag{6.4}$$

where the subscript Q on the expectation of $\hat{y}$ indicates that the true model is quadratic. This expectation involves the expectation of both $a = \bar{y}$ and b under the quadratic model assumption. It has already been shown in Chapter IV that for symmetrical designs, b is unbiased for β_1 even in the presence of a quadratic effect. The expectations of $a = \bar{y}$ under the quadratic model assumption is simply β_0 so that both a and b are unbiased for their respective regression coefficients in the presence of a quadratic effect if designs are restricted to be symmetrical. Therefore,

$$\begin{aligned}
\mathrm{Bias}(\hat{y}) &= E_Q(\hat{y}) - y \\
&= \beta_0 + \beta_1 x - \beta_0 - \beta_1 x - \beta_2(x^2 - \overline{x^2}) \\
&= - \beta_2(x^2 - \overline{x^2})
\end{aligned} \tag{6.5}$$

and

$$2B = \beta_2^2 \int_{-1}^{1} (x^2 - \overline{x^2})^2 \, dx$$

$$= \beta_2^2 [2/5 - 4\overline{x^2}/3 + 2\overline{x^2}^2] \tag{6.6}$$

by elementary integration. Combining (6.3) with $\bar{x} = 0$ and (6.6) and dividing by two, one obtains

$$J_{FQ} = (\sigma^2/N)\ [1+1/(3\overline{x^2})] + \beta_2^2[1/5 - 2\overline{x^2}/3 + \overline{x^2}^2] \tag{6.7}$$

where the subscript FQ on J indicates that forward regression is under consideration and that the true model is quadratic. It is noticed that apart from a factor of 2, the coefficient of β_2^2 in (6.7) is the same term that appears in J_{IQ} in Chapter IV and was designated as $f(\overline{x^2})$ in equation (4.25). One cannot minimize J_{FQ} without prior knowledge of β_2 and σ^2 but computation of J_{FQ} for a wide variety of designs and model parameters reveals that unless the variance contribution in J_{FQ} overwhelms the bias contribution, designs very close to optimal can be obtained by simply minimizing the coefficient of β_2^2 in (6.7). In Chapter IV, this quantity was shown to be minimized when $\overline{x^2} = 1/3$. Hence, unless the variance contribution is very dominant, designs for which $\overline{x^2} = 1/3$ yield a value of J_{FQ} very close to Min J_{FQ}. It is noted that in order for the variance contribution to dominate J_{FQ}, σ^2 must be very large in comparison with β_2^2 which means that $\gamma_2 = |\beta_2/\sigma|$ must be very small. Therefore, unless γ_2 is very small in the forward case, designs very close to optimal can be obtained by minimizing $f(\overline{x^2})$. Designs of this type are commonly referred to as "all-bias" designs since they are obtained by ignoring the variance contribution. This result then compares with the results obtained for the Inverse estimator in Chapter IV where it was shown that unless γ_2 is very small, designs very close to optimal can be obtained by minimizing $f(\overline{x^2})$ by taking $\overline{x^2} = 1/3$ and the second design variable $S_Q = N/10$.

For one who wishes to take the approach of utilizing prior information to construct designs, David and Arens (1959) have considered optimal "two-point" designs by expressing the quadratic model in terms of Legendre polynomials as discussed in Chapter IV. For symmetrical "two-point" designs, after setting $d2J_{FQ}/dx_2 = 0$ to minimize J_{FQ} with respect to x_2 and replacing x_2^2 by u, they arrived at an equation of the form

$$3u^3 - u^2 - a' = 0 \tag{6.8}$$

where $a' = 2\sigma^2/(9Nc_2^2)$ and where c_2 is the regression coefficient of the second Legendre polynomial. Solutions to (6.8) are functions of a' or equivalently of $f = |\sigma'/c_2|$ where $\sigma'^2 = 2\sigma^2/N$, and to aid the experimenter in utilizing prior information to obtain near optimal designs, David and Arens constructed a graph which shows optimal values of x_2 as a function of f. Ott and Myers, using the same Legendre polynomial model representation, showed that for the truncated Classical estimator in the inverse case, one could construct designs using David and Arens' graph for a given set of parameters and obtain a value of $2J_{CQ}$ very close to Min $2J_{CQ}$. Testing many cases, they found there was never more than five percent difference between the value of $2J_{CQ}$ obtained by using optimal designs for the forward case and Min $2J_{CQ}$ obtained by using optimal designs for the inverse case. They concluded that for all practical purposes, David and Arens' graph of optimal designs for forward regression could be used to construct "two-point" symmetrical designs for the truncated Classical estimator in the inverse case.

The results of this chapter serve to indicate that whether the true model is linear or quadratic or whether one uses the truncated Classical estimator and depends on prior information to construct designs or uses the Inverse estimator and does not depend on prior information to construct designs, there are definite analogies between the construction of optimal designs for the forward and inverse methods of regression.

CHAPTER VII

CONCLUDING REMARKS

While it has been shown that if the true model is linear, the Inverse estimator is fairly robust with respect to design for large values of γ and that the "end-point" design for the Inverse estimator provides an average MSE which is less than or equal to the minimum average MSE that can be obtained with the truncated Classical estimator for all values of the model parameters, the most significant aspect of this investigation, in this author's opinion, centers around model misclassification. Ott and Myers have shown that under the quadratic model assumption, the average MSE for the truncated Classical estimator is somewhat insensitive to slight design changes about the optimal designs, but even so, use of the Classical estimator still requires partial prior knowledge of the parameters involved. In fact, Ott in concluding his 1966 dissertation, states "An unfortunate aspect of the work presented to this point concerning misclassification of the model using the EMS (J in this manuscript) criterion, is that the experimenter must, in order to attain the "optimum" design, assume at least partial knowledge of the parameters involved." To attain near optimality using the Inverse estimator under the quadratic model assumption, one does not need this prior information which in many circumstances is a difficult commodity to acquire. While it has not been definitely established that the near optimal designs derived for the Inverse estimator are superior to designs for the truncated Classical estimator for all values of the model parameters, it has been established that they are markedly superior, even if designs for the Classical estimator are derived under perfect prior information, except for large γ_1

and small γ_2, and for this range of values of the parameters, it certainly cannot be concluded that the Classical designs are superior unless they are derived under perfect or near perfect prior information. Hence, if one uses the near optimal designs derived for the Inverse estimator, he is certain of obtaining an average MSE which is considerably smaller than the corresponding average MSE for the truncated Classical estimator except for large γ_1 and small γ_2, and for these values of the parameters, he will not obtain an average MSE which is much larger, if any at all depending on the amount of prior information available, than one would obtain using the truncated Classical estimator. Also, if prior information is available, which indicates that γ_1 is large and γ_2 is small and enables one to construct designs which are close to optimal for the Classical estimator, this prior information can also be used to improve designs for the Inverse estimator.

The above conclusions point toward the recommendation that the Inverse estimator be used for inverse estimation. If one feels confident that the true model is linear, the "end-point" design can be used with assurance of obtaining an average MSE which is at least as small as the minimum average MSE for the truncated Classical estimator. However, due to the robustness of the Inverse estimator with respect to design for large γ_1 and the fact that for small γ_1, the optimal design occurs when $\overline{x^2}$ is less than maximum, the use of a design for which $\overline{x^2} = 1/3$, a design with a uniform spread of points, or in fact any design for which $\overline{x^2}$ is greater than about .20 will provide an average MSE which is at most 13.5% above minimum for large γ_1 (evaluated at

$\overline{x^2} = .20$, $\gamma_1 = 50$, $N = 6$). For larger values of N and/or smaller values of γ_1, the above percentage will decrease. If one feels the possibility of a quadratic effect, the designs developed in Chapter IV can be used, the cost of such protection being relatively low compared with the cost of misclassification (versus the "end-point" design). In fact, if γ_1 is small, one obtains an improvement over the "end-point" design even if the model is linear since, as aforementioned, for small γ_1, the optimal design for the linear model is closer to $\overline{x^2} = 1/3$ than it is to $\overline{x^2} = 1$. In either the linear or the quadratic case, should prior information be available, an improvement in the above designs can be obtained. Should the true model be linear and prior information be available, improvement can be made by using Figures 1-5 or Figure 6. Should the true model be quadratic and prior information indicate that γ_1 is large and γ_2 is small, improvement can be made by constructing designs for which $\overline{x^2}$ lies between 1/3 and 1 (moving toward 1 for decreasing values of γ_2) and $S_Q = \Sigma(x_i^2 - \overline{x^2})^2$ is disregarded since this design variable has nearly no effect on the average MSE for small γ_2. It is well to re-emphasize that in either the linear or the quadratic case, one does not need prior information with the Inverse estimator to obtain designs which are very close to optimal.

The one unfortunate aspect of this investigation is the complicated nature of the exact moments of d under the quadratic model assumption. This forced the use of an approximation for the moments which prohibited the investigation of optimal designs for small values of γ_1. However, while this author will not take the risk of extrapolating results, everything in this investigation points toward increased superiority of the Inverse estimator as γ_1 decreases. This was shown in Figures 7 and 8 for the linear case and in Figures 17-21 for the quadratic case. Also, from a practical standpoint, small values of γ_1, say $\gamma_1 = 1$, imply either a large error variance or small slope of the regression line or both and these conditions are not conducive to accurate calibration. Hence, it is not felt that too much has been lost by not being able to consider designs in the quadratic case for small values of γ_1.

One final point deserves comment. In a recent Note in Technometrics, Williams (1969) commented on Krutchkoff's 1967 article. In brief, Williams states that since the MSE for the Classical estimator is infinite and the MSE for the Inverse estimator is finite, the Inverse estimator is better than the Classical from a mean squared error point of view. He further states that this isn't very satisfying and then proceeds to attack the mean squared error criterion which appears to fail in this case. He states,

> "In fact, since the Classical, the unbiased, or indeed any estimator that could be derived in a theoretically justifiable manner all have infinite variances, the fact that Krutchkoff's estimator (He is referring to the Inverse estimator) has a finite variance seems to be of little account. Thus, any estimator which is constant will have finite variance, and so from the MSD (MSE in this manuscript) point of view will be preferable to the classical estimator. The establishment of the conclusion that the "inverse" regression gives an estimator with a smaller MSD, although true, does not therefore appear to be particularly illuminating."

Then he concludes,

> "The upshot of this discussion appears to be

> that the minimum variance or minimum MSD criterion is not a suitable one in problems of this kind. Since sufficient statistics exist for all the parameters requiring estimation, the data can be summarized in terms of a few of these. Confidence limits can be readily derived by familiar methods, as given in the references listed by Krutchkoff (See, for instance, Williams (1959) p. 95). These confidence limits, being based on sufficient statistics, should provide what is required in estimation for calibration purposes."

It is clear that this Note was written prior to the publication of the previously cited work by Ott and Myers. One does not discard a criterion, especially the criterion of minimum MSE, merely because it appears to fail, for to do so, one prohibits himself from investigating the effect of design and model misclassification on the estimator. This has been shown to be extremely important as it is in most estimation problems which are subject to design. Williams fails to even mention design. The only reasonable approach is to truncate the density involved to obtain a finite MSE. This is what is actually done in practice and was the approach taken by Ott and Myers in optimizing designs for the Classical estimator. Also, concerning the statement on confidence limits, it was mentioned earlier that McClelland (1967) has shown that if one calculates an Inverse confidence interval by substituting the Inverse estimator for the Classical in the Classical confidence interval, then the interval will have a higher confidence than the Classical interval of the same length. Finally, the investigation of Ott (1966) and Ott and Myers (1968) concerning the derivation of optimal designs for the truncated Classical estimator, the empirical investigation of Krutchkoff (1967), and the investigation set forth in this dissertation concerning the derivation of optimal designs for the

Inverse estimator and comparison of these with those for the truncated Classical estimator are testimonies that one cannot dispose of the calibration problem in a few short sentences. This kind of reasoning can only be likened onto that of Eisenhart, whose 1939 article suppressed the use of the better estimator of the two for nearly thirty years.

BIBLIOGRAPHY

1. Ambramowitz, M. and Stegun, I. A., 1964. Handbook of Mathematical Functions, Applied Mathematics Series 55, U.S. Department of Commerce, National Bureau of Standards.

2. Bennett, C. A. and Franklin, N. L., 1954. Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons, New York, p. 231.

3. Box, G. E. P. and Draper, N. R., 1959. A Basis for the Selection of a Response Surface Design. Journal of the Amer. Statist. Assn. 54, 622.

4. Brownlee, K. A., 1966. Statistical Theory and Methodology, John Wiley and Sons, Inc., New York, p. 346.

5. Courant, R., 1959. Differential and Integral Calculus, Vol. II, Interscience Publishers, Inc., New York, p. 257.

6. David, H. A. and Arens, B. E., 1959. Optimal Spacing in Regression Analysis. Ann. Math. Statist., 30, 1072.

7. Draper, N. R., 1966. Applied Regression Analysis, John Wiley and Sons, Inc., New York.

8. Eisenhart, C., 1939. The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research. Ann. Math. Statist. 10, 162.

9. Good, I. J., 1963. On the Independence of Quadratic Expressions. J.R. Statist. Soc. B, 25, 377.

10. Graybill, F. A., 1961. An Introduction to Linear Statistical Models, Volume I, McGraw-Hill Book Company, Inc., New York, pp. 74, 138.

11. Hemmerle, W. J., 1967. Statistical Computations on a Digital Computer, Blaisdell Publishing Company, p. 204.

12. Kendall, M. G. and Stuart, A., 1963. The Advanced Theory of Statistics, Hafner Publishing Company, New York, p. 232.

13. Krutchkoff, R. C., 1967. Classical and Inverse Regression Methods of Calibration. Technometrics 9, 425.

14. Linning, F. J. and Mandel, J., 1964. Which Measure of Precision? Analytic Chemistry 36, 25A.

15. Mandel, J., 1958. A Note on Confidence Intervals in Regression Problems. Ann. Math. Statist. 29, 903.

16. Mandel, J. and Linning, F. J., 1957. Study of Accuracy in Chemical Analysis Using Linear Calibration Curves. Analytic Chemistry 29, 743.

17. McClelland, W. T., 1967. Linear Calibration: A Comparison of an Inverse Regression Method of Confidence Intervals with the Classical Regression Method of Confidence Intervals, M.S. Dissertation, V.P.I.

18. Mood, A. M. and Graybill, F. A., 1963. Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., New York, p. 338.

19. Ott, R. L., 1966. Optimal Designs for Inverse Estimation in Regression Analysis. Ph.D. Dissertation, V.P.I.

20. ________ and Myers, R. H., 1968. Optimal Experimental Designs for Estimating the Independent Variable in Regression. Technometrics 10, 811.

21. Pearson, K., 1957. Tables of the Incomplete Γ-Function, Cambridge University Press.

22. Scarborough, J. B., 1962. Numerical Mathematical Analysis, The Johns Hopkins Press, Baltimore, pp. 133, 199.

23. Widder, D. V., 1961. Advanced Calculus, Prentice-Hall, Inc., Englewood Cliffs, p. 44.

24. Williams, E. J., 1959. Regression Analysis, John Wiley and Sons, New York, p. 10.

25. ____________, 1969. A Note on Regression Methods in Calibration. Technometrics 11, 189.

26. Wine, R. L., 1964. Statistics for Scientists and Engineers, Prentice-Hall, Inc., Englewood Cliffs, p. 511.

27. Yarbrough, G. G., 1968. A Comparison of Expected Mean Square Error for the Classical and Inverse Methods of Linear Calibration, M.S. Dissertation, V.P.I.

The vita has been removed from the scanned document

A COMPARISON OF THE CLASSICAL AND INVERSE
METHODS OF CALIBRATION IN REGRESSION

by

Marlin Amos Thomas

Abstract

The linear calibration problem, frequently referred to as inverse regression or the discrimination problem can be stated briefly as the problem of estimating the independent variable x in a regression situation for a measured value of the dependent variable y. The literature on this problem deals primarily with the Classical method where the Classical estimator is obtained by expressing the linear model as

$$y_i = \alpha + \beta x_i + \varepsilon_i ,$$

obtaining the least squares estimator for y for a given value of x and inverting the relationship. A second estimator for calibration, the Inverse estimator, is obtained by expressing the linear model as

$$x_i = \gamma + \delta y_i + \varepsilon_i'$$

and using the resulting least squares estimator to estimate x. The experimental design problem for the Inverse estimator is explored first in this dissertation using the criterion of minimizing the average or integrated mean squared error, and the resulting optimal and near optimal designs are then compared with those for the Classical estimator which were recently derived by Ott and Myers.

Optimal designs are developed for a linear approximation when the true model is linear and when it is quadratic. In both cases, the

optimal designs depend on unknown model parameters and are not realistically useable. However, designs are shown to exist which are near optimal and do not depend on the unknown model parameters. For the linear approximation to the quadratic model, these near optimal designs depend on N, the number of observations used to estimate the model parameters, and specific designs are developed and set forth in tables for N = 5(1)20(2)30(5)50.

The cost of misclassifying a quadratic model as linear is discussed from a design point of view as well as the cost of protecting against a possible quadratic effect. The costs are expressed in terms of the percent deviation from the average mean squared error that would be obtained if the model were classified correctly.

The derived designs for the Inverse estimator are compared with the recently derived designs for the Classical estimator using as a measure of comparison the ratio of minimum average mean squared errors obtained by using the optimal design for both estimators. Further comparisons are also made between optimal designs for the Classical estimator and the derived near optimal designs for the Inverse estimator using the ratio of the corresponding average mean squared errors as a measure of comparison.

Parallels are drawn between forward regression (estimating the dependent variable for a given value of the independent variable) and inverse regression using both the Classical and Inverse methods.