National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

RANDOMIZING FOR VALIDITY IN EXPERIMENTAL DESIGN

BY

CHENJIANG HU 

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

FALL 1994

ISBN 0-315-95194-X

Canada

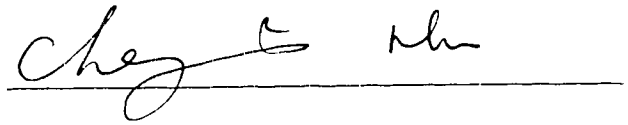
UNIVERSITY OF ALBERTA
RELEASE FORM

NAME OF AUTHOR: Chenjiang Hu

TITLE OF THESIS: **Randomizing for Validity in Experimental Design**

DEGREE FOR WHICH THESIS WAS PRESENTED: Doctor of Philosophy

YEAR THE DEGREE GRANTED: 1994



Chenjiang Hu

Permanent Address:

201-10533-83 Avenue

Edmonton, Alberta

Canada T6E 2E1

Date: July 26, 1994

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled

Randomizing for Validity in Experimental Design

submitted by **Chenjiang Hu** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy.**

Peter M. Hooper (Supervisor)

D. Wiens

N. G. N. Prasad

D. Kelker

Francis Yeh

for David R. Bellhouse (External Examiner)

Date: 21 July 94

TO MY FAMILY

ABSTRACT

The widespread use of randomization in experimental design is due in large part to its validation of the usual normal theory analysis of experimental data. This thesis investigated how randomization tends to validate the usual statistical analysis for treatment effects in a comparative experiment.

The first chapter gave a review of general background for the topic on validation of randomization. In Chapter 2, we discussed the validity measures $\mathrm{cv}(\phi)$ in Hooper (1989) and $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ in Hooper (1993) with regard to interpretation and computation. We introduced the Beta-related validity measure $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$, i.e., the coefficients of variation of the first two randomization moments of the Beta statistic. By approximating the randomization distribution of the Beta statistic by a Beta distribution, we can interpret indirectly $\mathrm{cv}(\phi)$ by $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$. A relation between $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ and $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ was derived, so the Beta-related measure can be estimated at the same cost as $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$. We also derived an approximate Laguerre series expansion for $\mathrm{cv}(\phi)$, with leading terms related to $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$.

In Chapter 3, We investigated methods of randomization for asymmetric models of the unit effects. Cox (1982) and Bellhouse (1986) suggested the truncated classical randomized design (TCRD). We proposed a sequential procedure for generating a sequence of TCRDs to assist in the selection of an appropriate truncation constant κ. Under a general linear model, Hooper (1993) proposed NORDs to achieve acceptable levels of validity and efficiency. In order to improve NORD with respect to efficiency and the selection of tuning constant we introduced the truncated NORD (TNORD). The TNORD is invariant under reparameterization of the treatment effects model and can be generated by the same sequential procedure as the TCRD. We derived some formulae for the minimization of the distance function for the TNORD. A simulation study was presented showing that the same levels of validity and efficiency

can be achieved by suitable choice of the tuning or truncation constant in these three methods.

Chapter 4 deals with optimal randomized design. A randomized design is defined to be optimal for validity if it minimizes a validity criterion. A general sufficient condition was obtained for an uniform distribution in a subset of $\mathcal{P}_n$ to be optimal for validity. Using the general sufficient condition, we obtained that some classical randomized designs, such as CRD and randomized blocks design, are optimal for validity among all distributions on a certain subset of $\mathcal{P}_n$. Under analysis of covariance models, however, the uniform distribution on a certain subset of $\mathcal{P}_n$ may not be optimal for validity. Therefore, under asymmetric models for unit effects, NORD or TNORD may have higher validity than TCRD in some circumstances, although TCRD is more easily generated.

ACKNOWLEDGMENTS

It is with feelings of warmest affection and sincerest appreciation that I express my thanks to my supervisor, Professor Peter M. Hooper, for his consistent encouragement, invaluable guidance and numerious suggestions throughout my study and research at the University of Alberta.

It is with great pleasure that I would like to thank Professors D. Wiens, N. G. N. Prasad and D. Kelker for their valuable advice and help during my study in the Department of Statistics and Applied Probability. I also wish to express my thanks to my external examiner, Professor David R. Bellhouse, for his critical reading of my thesis.

My thanks also go to the department secretaries for their kindness and helpfulness.

Last but not the least, I wish to expres my gratitude to Professor Peter M. Hooper and the Department of Statistics and Applied Probability at University of Alberta for their continued support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Review and Summary

The widespread use of randomization in experimental design is due in large part to its validation of the usual normal theory analysis of experimental data. This notion that randomization validates inference has been developed in various ways.

British statisticians defined validity in terms of whether the normal theory estimators of certain first- and second-order quantities are unbiased under a randomization model. Fisher (1925 and 1935) and Yates (1933) proposed an adequacy criterion for randomization of an experiment with a single error term in the analysis of variance. This criterion requires that the expectations of the treatment mean square and the error mean square are equal in the absence of treatment effects, the expectations being obtained by averaging over all possible outcomes of the randomization. This is called 'weak validity', in contrast with the 'strong validity' of Grundy and Healy (1950), which requires the mean square based on any set of treatment differences to have the same expectation as the error mean square. Grundy and Healy's strong validity was generalized by Nelder (1965a and 1965b) to experiments with simple block structure. Nelder used the block structure of an experiment to define strata and defined a randomization strategy to be valid if all estimators of normalized treatment contrasts

within each stratum have the same variance under randomization. Sufficient combinatorial conditions for weak and strong validity of randomization are well known for some common classes of designs with single error strata. Bailey and Rowley (1987) generalized these conditions to orthogonal designs with multiple error strata given by an association scheme on the plots and with a treatment decomposition also defined by an association scheme. Constrained randomization methods were developed to provide validity while avoiding undesirable designs; see Bailey (1986) for a review.

In North America, statisticians typically described validity by the approximate agreement between randomization tests and normal-theory tests. Fisher argued that, when testing the hypothesis of no treatment effect in an agricultural experiment, the normal-theory significance level usually approximates the corresponding randomization significance level and the latter derives its validity solely from the physical act of randomization. The tendency for agreement between normal theory and randomization tests was investigated by many authors. Eden and Yates (1933) showed close agreement between the randomization and normal theory tests in an empirical investigation of a real data set. In a simulation study, Kempthorne and Doerfler (1969) found agreement between the randomization and the normal theory tests based on the F statistic. For the randomized blocks design and the Latin square design, Welch (1937) and Pitman (1938) compared the randomization moments of a Beta statistic with the normal theory moments. Hoeffding (1952) and Robinson (1973) proved that under general conditions the randomization tests are asymptotically as powerful as the corresponding normal theory tests. Davis and Speed (1988) obtained an Edgeworth expansion of the randomization distribution of the F-ratio statistic and used this to compare the randomization and normal theory significance levels.

Hooper (1989) introduced a related notion of validity as a measure of agreement between model-assisted probabilities and randomization probabilities. Here 'model-assisted probabilities' are probabilities calculated under a model incorporating both

the randomization and normal theory assumptions and 'randomization probabilities' are conditional probabilities with all random variables held fixed except the randomization. Hooper carried out a simulation study involving several classical designs to examine how validity depends on the design and the size of the experiment. Increasing the size of the experiment generally improves the validity supplied by the randomization. In the simulation study, a coefficient of variation measure was introduced to evaluate the validity supplied by randomization of an experiment.

Hooper (1993) developed a new family of randomized designs, termed nearly orthogonal randomized designs (NORDs), intended to achieve high levels of efficiency and validity within the context of a general linear model for treatment and unit effects. A diagnostic measure of validity was proposed to aid in the selection of a tuning constant in NORDs that determines a trade-off between validity and efficiency. Cox (1982) suggested truncated completely randomized designs (TCRDs) in the context of analysis of covariance models and Bellhouse (1986) investigated the validity supplied by TCRDs. The TCRD also requires a tuning (truncation) constant.

In this thesis, I investigate how randomization tends to validate normal theory methods of analysis. First, I study methods to measure the validity supplied by randomization. I introduce a validity measure which is related to the randomization moments of a Beta statistic. Relationships among the two measures of Hooper and my measure are investigated by analytic and numerical methods. Second, I study randomization methods that achieve acceptable levels of efficiency and validity. A new method, called Truncated NORD (TNORD), is presented to improve NORDs. Results of a simulation study comparing TNORDs, NORDs and TCRDs are described. A sequential procedure is proposed to determine a value for the truncation constant that achieves a trade-off between efficiency and validity.

All of my investigations are based on a general linear model for treatment and unit effects in a comparative experiment, proposed by Hooper (1993). Before presenting

my results, I state this model and the related normal theory analysis in the next section.

1.2 Normal Theory Model and Inference

Consider a comparative experiment in which treatments are applied to experimental units and responses are observed. Although randomization can be described in several ways, see Preece, Bailey and Patterson (1978), the following treatment assignments are used in following discussions. Let the units involved in the experiment be numbered from 1 to n and imagine a set of labels also numbered from 1 to n. We choose a fixed arrangement of treatments among labels and then assign the treatment associated with label Gi to unit i, for $i = 1, \ldots, n$, where G is a random permutation. A randomized design is thus defined as a distribution of G given the arrangement of treatments among labels. For example, in a randomized blocks design labels within blocks are randomly permuted and then blocks of labels are permuted.

Let Y_i denote the response from unit $G^{-1}i$. Thus Y_i is the response from a fixed treatment and a random unit. We assume that treatment effects and unit effects are additive, so we have

$$Y_i = \theta_i + U_{G^{-1}i}, \qquad i = 1, \ldots, n, \tag{1.2.1}$$

where θ_i is the effect from the treatment associated with label i and U_i is the effect from unit $G^{-1}i$. It is convenient to express model (1.2.1) in vector form. Define column vectors $Y = (Y_1, \ldots, Y_n)^T, \theta = (\theta_1, \ldots, \theta_n)^T$ and $U = (U_1, \ldots, U_n)^T$. For each permutation g of $\{1, \ldots, n\}$, let $\tilde{g}$ be the $n \times n$ permutation matrix defined by

$$(\tilde{g}U)_i = U_{g^{-1}i}, \qquad i = 1, \ldots, n.$$

The mapping $g \to \tilde{g}$ is a group isomorphism. We shall suppress the tilde and let G denote both a random permutation and the corresponding random permutation

matrix. Let $\mathcal{P}_n$ denote both the group of permutations of $\{1,\ldots,n\}$ and the group of $n \times n$ permutation matrices. So, model (1.2.1) can be written

$$Y = \theta + GU. \tag{1.2.2}$$

We assume a linear model for the treatment effects, using a convenient canonical parameterization:

$$\theta = A_0\beta_0 + A\beta, \tag{1.2.3}$$

where β_0 is $q_0 \times 1$ vector of nuisance parameters and β is the $q \times 1$ vector of parameters of interest. The matrices $A_0 : n \times q_0$ and $A : n \times q$ are known and the columns of $[A_0 : A]$ are orthonormal. Let $\mathcal{R}(A_0)$ denote the column space of A_0 and assume that $\mathcal{R}(A_0)$ contains $\mathbf{1}_n = (1,\ldots,1)^T$. In a factorial design, for example, $A\beta$ might represent the main effects and low order interactions. If all treatment contrasts in the model are of interest then $A_0 = (1/\sqrt{n})\mathbf{1}_n$.

A linear model for the unit effects is also assumed:

$$U = B\gamma + KV \tag{1.2.4}$$

where B is a non-random $n \times r$ matrix, γ is a non-random $r \times 1$ vector, K is a random matrix taking values in $\mathcal{P}_n(B) = \{J \in \mathcal{P}_n; \mathcal{R}(JB) = \mathcal{R}(B)\}$ and V is a random $n \times 1$ vector. The columns of B are orthonormal and the column space $\mathcal{R}(B)$ contains the vector $\mathbf{1}_n$. The components of V are independent with common distribution $N(0,\sigma^2)$. The matrix B is known; γ, K, V and σ^2 are unknown. The random permutation matrix K and random vector V are *not* assumed independent and the distribution of K is unknown. Model (1.2.4) may represent the analysis of covariance models and spatial models for unit effects. Also, models for unbalanced designs and for missing data can be described by the linear model (1.2.4).

The permutation matrix K is included in the model to allow some dependence between the error terms while permitting an exact normal theory analysis under

randomization. The components of KV are not assumed independent but, as shown later, K is 'randomized out' leaving a vector GKV with independent and identically distributed random variables.

Combining equations (1.2.2), (1.2.3) and (1.2.4), we have

$$Y = A_0\beta_0 + A\beta + GB\gamma + GKV. \tag{1.2.5}$$

We shall require that the distribution of G satisfies the following conditions:

$$\text{rank}[GB : A_0 : A] = \text{rank}[GB : A_0] + \text{rank}(A), \tag{1.2.6}$$

$$G \text{ is distributed independently of } (K, V), \tag{1.2.7}$$

$$GJ \sim G \text{ for all } J \in \mathcal{P}_n(B), \tag{1.2.8}$$

where '$\sim$' means 'is distributed as'. Condition (1.2.6) ensures that all linear functions of β are estimable. From conditions (1.2.7) and (1.2.8) we obtain that GK is distributed independently of V and then GKV has the same distribution as V and GKV is distributed independently of $\mathcal{R}(GB) = \mathcal{R}(GKB)$.

The usual normal theory methods of inference on β are conditioned on $\mathcal{R}(GB)$ and can be carried out as follows. We define $X_1 = \text{GramSchmidt}\ (X_2)$ to mean 'construct X_1 by applying the Gram-Schmidt orthonormalization process to the columns of X_2, from left to right'. The number of columns of X_1 equals the rank of X_2 and $JX_1 = \text{GramSchmidt}\ (JX_2)$ for $J \in \mathcal{O}_n$, the group of $n \times n$ orthogonal matrices. Put $s = \text{rank}\ [GB : A_0 : A]$ and define the $n \times s$ matrix X and the $n \times q$ matrix C by

$$X = [GB : * : C] = \text{GramSchmidt}[GB : A_0 : A]. \tag{1.2.9}$$

The columns of C are thus an orthonormal basis for the subspace obtained by projecting $\mathcal{R}(A)$ onto the orthogonal complement of $\mathcal{R}([GB : A_0])$. Condition (1.2.6) ensures that C^TA is non-singular. Note from (1.2.9) that C is a function of G and

GK results in the same C as G for any $K \in \mathcal{P}_n(B)$. Thus C is a function of GK if $K \in \mathcal{P}_n(B)$.

The least squares estimator of β can be written

$$\hat{\beta} = (C^T A)^{-1} C^T Y$$

and the conditional distribution of $\hat{\beta}$ given $\mathcal{R}(GB)$ is multivariate normal with mean β and covariance matrix

$$\sigma^2 (A^T C C^T A)^{-1}. \tag{1.2.10}$$

Let

$$\hat{\sigma}^2 = \frac{Y^T (I_n - XX^T) Y}{n - s}.$$

Tests and confidence regions can be based on the pivotal quantity

$$Q(G, G^T Y, \beta) = \hat{\sigma}^{-1} C^T (Y - A\beta). \tag{1.2.11}$$

For example, if we are interested in testing the hypothesis $H_0 : \beta = 0$, then the F statistic can be constructed from the pivotal quantity (1.2.11), i.e.

$$F(G, G^T Y) = \frac{SS_T/q}{SS_E/(n-s)} =: \frac{\|Q(G, G^T Y, 0)\|^2}{q} = \frac{Y^T C C^T Y}{q\hat{\sigma}^2} \tag{1.2.12}$$

where $SS_T = Y^T C C^T Y = \|C^T Y\|^2$ is the sum of squares for treatments and $SS_E = Y^T (I - XX^T) Y = \|(I - XX^T) Y\|^2$ is the sum of squares for error. Here $\|M\|$ denote the Euclidean norm of the matrix M, i.e. $\|M\|^2 = \mathrm{tr}(MM^T)$. The conditional distribution of the F statistic given $\mathcal{R}(GB)$ is the $F_{q,n-s}$ distribution. The Beta statistic, defined by

$$T(G, G^T Y) = \frac{SS_T}{SS_T + SS_E}, \tag{1.2.13}$$

is a one-to-one function of the F statistic.

In model (1.2.5), let $W = U - B\gamma$ be the vector of random errors associated with experimental units. Under the null hypothesis $H_0 : \beta = 0$, we have $Q(G, G^T Y, 0) = Q(G, W, 0)$ and we can replace $G^T Y$ by W in all the above equations, for example,

$\hat{\beta} = (C^T A)^{-1} C^T GW$ and $SS_T = \|C^T GW\|^2$. Thus, the F and Beta statistics can be denoted by $F(G, W)$ and $T(G, W)$ respectively.

If all the treatment contrasts in the model are of interest, then $A_0 = (1/\sqrt{n})\mathbf{1}_n$, $s = r + q$ and

$$X = [GB : C] = \text{GramSchmidt}[GB : A] = \text{GramSchmidt}[GB : A_0 : A].$$

Moreover, we have under H_0

$$\begin{aligned} SS_T + SS_E &= \|(I - GBB^T G^T)Y\|^2 \\ &= \|(I - GBB^T G^T)GW\|^2 \\ &= \|(I - BB^T)W\|^2, \end{aligned}$$

so the denominator of the Beta statistic, under the null hypothesis, is invariant under randomization. The Beta statistic $T(G, W)$ can be written

$$T(G, W) = \frac{\|C^T GW\|^2}{\|(I - BB^T)W\|^2}. \tag{1.2.14}$$

1.3 Criteria

Based on model (1.2.5), we can evaluate randomized designs with regard to efficiency, validity and relevance.

Our criterion of validity is concerned with the dependence of probabilities and expectations on well-founded assumptions about the unit errors W. Let us consider the problem of testing the null hypothesis $H_0 : \beta = 0$ against the alternative $H_0 : \beta \neq 0$ under model (1.2.5). Suppose that H_0 is rejected if $f\{Q(G, G^T Y, 0)\} > c$, where f is a continuous real-valued function. Under the null hypothesis, the test function can be written

$$\phi(G, W) = \begin{cases} 1 & f\{Q(G, W, 0)\} > c \\ 0 & f\{Q(G, W, 0)\} \leq c \end{cases} \tag{1.3.1}$$

If $f(x) = \frac{x^T x}{q}$, we obtain the F test. We shall refer to $E\{\phi(G,W)\}$ as the model-assisted probability since it involves the distributions of both the randomization G and the random error W. This probability is uniquely determined under model (1.2.5). By conditioning on $\mathcal{R}(GB)$, we can implement a normal theory analysis on β. $E\{\phi(G,W) \mid \mathcal{R}(GB)\}$ is referred to as the normal-theory probability, which can be evaluated under normal theory and does not depend on $\mathcal{R}(GB)$ under the null hypothesis. Thus $E\{\phi(G,W)\} = E\{\phi(G,W) \mid \mathcal{R}(GB)\}$ under the null hypothesis, i.e., the model-assisted significance level is the normal theory significance level. Under the alternative hypothesis we still have $E\{\phi(G,W)\} = E\{\phi(G,W) \mid \mathcal{R}(GB)\}$ for classical randomized designs, however this equality does not hold for all randomization strategies. Our criteria for validity are based on a comparison of the randomization and model-assisted significance levels, i.e., $E\{\phi(G,W) \mid W\}$ and $E\{\phi(G,W)\}$ under the null hypothesis. We shall refer to $E\{\phi(G,W) \mid W\}$ and $E\{\phi(G,W)\}$ as the probabilities under the null hypothesis in the following context. Validity will be described in terms of the extent of agreement between the randomization probability $E\{\phi(G,W) \mid W\}$ and the model-assisted probability $E\{\phi(G,W)\}$. Validity is thus a matter of degree. Large experiments typically supply more validity than do small experiments.

The validity can be evaluated by the variance of the randomization probability $E\{\phi(G,W) \mid W\}$. If under model (1.2.5) the variance of $E\{\phi(G,W) \mid W\}$ is small, the randomization and model-assisted probabilities are then likely to be in close agreement. The model-assisted probability is thus determined primarily by the physical act of randomization, provided that the realized value of W is consistent with model (1.2.5). Hooper (1989) defined the coefficient of variation of $E\{\phi(G,W) \mid W\}$ as a measure of validity. In the next chapter we will consider two other measures of validity as well as this measure.

Efficiency is related to the precision of estimators of parameters of interest and is usually defined in terms of a scalar function of the covariance matrix (1.2.10). Efficiency depends on both the design selected and the parameterization used in the model. In the treatment-effects model we adopted a convenient parameterization such that the columns of $[A_0 : A]$ are orthonormal. Suppose that the natural parameterization is $\theta = \tilde{A}_0\delta_0 + \tilde{A}\delta$, with δ being the parameter vector of interest, and that $[A_0 : A] = \text{GramSchmidt}[\tilde{A}_0 : \tilde{A}]$. We then have $\delta = R\beta$ for a known $q \times q$ non-singular matrix R. From expression (1.2.10), the conditional covariance matrix of the least squares estimator for δ is $\sigma^2 R(A^T CC^T A)^{-1} R^T$. The matrix R depends on the selection of treatment combinations and the definition of the treatment-effects model. For the D-optimality criterion, the selection of a design with high efficiency does not depend on the parameterization of the treatment effects model; i.e., we have $\det[R(A^T CC^T A)^{-1} R^T] = \det[(A^T CC^T A)^{-1}]\det(R^T R)$ and this can be minimized as a function of G independently of R. More generally, the various efficiency criteria suggest that G should be chosen to make $(A^T CC^T A)^{-1}$ small in some sense (perhaps depending on R). By definition we have $(A^T CC^T A)^{-1} \geq I_q$, i.e. $(A^T CC^T A)^{-1} - I_q$ is non-negative definite. If G can be chosen so that the subspace $\mathcal{R}([GB : A_0])$ is orthogonal to $\mathcal{R}(A)$ then $C = A$ and $(A^T CC^T A)^{-1} = I_q$. If exact orthogonality is not possible then, for efficiency, G should be chosen to make the subspace $\mathcal{R}([GB : A_0])$ nearly orthogonal to $\mathcal{R}(A)$ in some sense. Many classical randomized designs achieve full efficiency.

Relevance concerns the interpretation of probabilities after the randomization has been carried out. At that time it may seem more appropriate to use the model-based probability $E\{\phi(G, W) \mid G\}$. However, under model (1.2.5) this probability depends on unknown nuisance parameters. The model-assisted probability can be viewed as an estimate of the unknown model-based probability provided the randomization scheme is restricted to a homogeneous family of designs. Homogeneity implies that we have

approximately the same opinion about $E\{\phi(G,W)\,|\,G\}$ for all possible designs G. The width of the non-parametric prediction intervals for an efficiency measure can be used to evaluate the relevance of a randomized design.

Chapter 2

Evaluating the Validity Supplied by Randomization

2.1 Introduction

One of the benefits attributed to experimental randomization is its tendency to validate the usual normal theory analysis. How much validity does a randomized design provide? To answer this question, we must have some quantity to measure the degree of validity supplied by a randomized design.

The present chapter deals with quantitative measures of validity. In Section 2.2, we review the two validity measures in Hooper (1989) and Hooper (1993). The two measures are discussed with regard to interpretation and computation. We propose a validity measure in Section 2.3, which is easier to interpret than the measure of Hooper (1993) and easier to compute than the measure of Hooper (1989). A relationship between the new measure and the measure in Hooper (1993) is given in Section 2.4. Theorems in Section 2.4 are proved in Section 2.5. An asymptotic analytic expression is derived for the validity measure of Hooper (1989) in Section 2.6 and 2.7. In Section 2.8, some simulation results are presented for comparison of these three validity measures and some suggestions and comments are provided.

All the discussions in this chapter are based on the model (1.2.5) with $A_0 = (1/\sqrt{n})\mathbf{1}_n$, i.e. all the treatment contrasts in the model are of interest. The assumptions about model (1.2.5) in Section 1.2 are adopted here.

2.2 Two Validity Measures in Hooper (1989 and 1993)

The model-assisted significance level $E\{\phi(G,W)\}$ is a good predictor of the randomization significance level $E\{\phi(G,W) \mid W\}$ if the variance of $E\{\phi(G,W) \mid W\}$ is small. Under model (1.2.5), the model-assisted significance level $E\{\phi(G,W)\}$ is the normal theory significance level $E\{\phi(G,W) \mid \mathcal{R}(GB)\}$. In this situation, randomization provides high validity for the normal theory analysis of the experimental data under the model (1.2.5). Given a test function (1.3.1), Hooper (1989) suggested the coefficient of variation of $E\{\phi(G,W) \mid W\}$

$$\mathrm{cv}(\phi) = \frac{\mathrm{var}^{\frac{1}{2}}\{E[\phi(G,W) \mid W]\}}{E\{\phi(G,W)\}} \tag{2.2.1}$$

as a validity measure. Let α denote the normal theory size $E\{\phi(G,W)\}$. $\mathrm{cv}(\phi)$ is the relative measure of the variability of the conditional size $E\{\phi(G,W) \mid W\}$ about its mean α. An approximate 95% prediction interval for $E\{\phi(G,W) \mid W\}$ is given by

$$\alpha - 2\alpha\,\mathrm{cv}(\phi) \le E\{\phi(G,W) \mid W\} \le \alpha + 2\alpha\,\mathrm{cv}(\phi) \tag{2.2.2}$$

Thus, if $\alpha = .05$ and $\mathrm{cv}(\phi) = .10$, then the randomization significance level of the test will usually be between .04 and .06 under model (1.2.5).

The validity measure $\mathrm{cv}(\phi)$ in (2.2.1) can be used to compare the validity of randomized designs for different unit-effects models and different numbers of units. Usually, $\mathrm{cv}(\phi)$ must be estimated by simulation methods.

Hooper (1993) proposed another validity measure to aid in the selection of a tuning constant for Nearly Orthogonal Randomized Designs (NORDs). This measure

is also motivated by the variance of $E\{\phi(G,W)\,|\,W\}$. The following result is implicit in Hoeffding (1952): if G_1 and G_2 are independent replicates of G, distributed independently of W, then

$$\text{var}\{E[\phi(G,W)\,|\,W]\} = \text{cov}\{\phi(G_1,W),\phi(G_2,W)\}.$$

Let (G_1,C_1) and (G_2,C_2) be independent replicates of (G,C). If $Q(G_1,W,0)$ and $Q(G_2,W,0)$ are nearly uncorrelated then $\text{var}\{E[\phi(G,W)\,|\,W]\}$ will be small. Since $\hat{\sigma}^2 \rightarrow_p \sigma^2$ as $n-s \rightarrow \infty$, where $\rightarrow_p$ denotes convergence in probability, we consider the covariance between $C_1^TG_1W$ and $C_2^TG_2W$. The conditional joint distribution of these two vectors, given (G_1K,G_2K), is multivariate normal with mean 0 and covariance matrix

$$\text{cov}\left\{\begin{pmatrix} C_1^TG_1W \\ C_2^TG_2W \end{pmatrix} \,|\, G_1K,G_2K\right\} = \sigma^2 \begin{pmatrix} I_q & C_1^TG_1G_1^TC_2 \\ C_2^TG_2G_1^TC_1 & I_q \end{pmatrix}.$$

If $\|C_1^TG_1G_2^TC_2\|^2 \rightarrow_p 0$ as $n \rightarrow \infty$ then $C_1^TG_1W$ and $C_2^TG_2W$ are asymptotically independent and $\text{var}\{E[\phi(G,W)\,|\,W]\} \rightarrow 0$; see Section 6 in Hooper (1993) for a proof. Note that $\|C_1^TG_1G_2^TC_2\|^2$ takes values in the interval $[0,q]$. For validity, the distribution of $\|C_1^TG_1G_2^TC_2\|^2$ should be concentrated close to 0. To aid interpretation Hooper (1993) introduced a standardized variable:

$$\text{val}(G_1,G_2) = \frac{\|C_1^TG_1G_2^TC_2\|^2 - E\|D_1^TD_2\|^2}{\{\text{var}(\|D_1^TD_2\|^2)\}^{\frac{1}{2}}} \tag{2.2.3}$$

where D_1 is uniformly distributed in the set $\mathcal{W} = \{M : n \times q, M^TM = I_q, M^TB = 0\}$ and D_2 is an independent replicate of D_1. Consequently, he defined a validity measure as the mean and standard deviation of $\text{val}(G_1,G_2)$, denoted by μ_{val} and σ_{val}, which are equivalent to the first two moments of $\|C_1^TG_1G_2^TC_2\|^2$. The following expressions were derived in Hooper (1993):

$$E\|D_1^TD_2\|^2 = \frac{q^2}{n-r}$$

and

$$\mathrm{var}(\|D_1^T D_2\|^2) = \frac{2q^2(n-r-q)^2}{(n-r-1)(n-r)^2(n-r+2)}.$$

Hooper (1993) also proved that, for all randomization strategies,

$$E\{\|C_1^T G_1 G_2^T C_2\|^k\} \geq E\{\|D_1^T D_2\|^k\} \qquad \text{for } k=2 \text{ and } k=4.$$

Thus $\mu_{\mathrm{val}} \geq 0$ and if $\mu_{\mathrm{val}} = 0$ then $\sigma_{\mathrm{val}} \geq 1$. Optimal validity for given (n,q,r) occurs when $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}}) = (0,1)$. It was shown in Hooper (1993) that $\mu_{\mathrm{val}} = 0$ for the CRD, randomized blocks design and Latin square design. As the distribution of $\|C_1^T G_1 G_2^T C_2\|^2$ becomes more spread out, both μ_{val} and σ_{val} tend to increase. If randomization is severely restricted then μ_{val} can be large and σ_{val} small. A non-randomized design has $\sigma_{\mathrm{val}} = 0$.

By (2.2.2), the validity measure $\mathrm{cv}(\phi)$ tells us in which range the randomization significance level $E\{\phi(G,W) \mid W\}$ should be. If two randomized designs with different unit-effects models and different numbers of units have the same value of $\mathrm{cv}(\phi)$, then we say that these two randomized designs provide the same degree of validity. However, this interpretation can not apply to the validity measure $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$, since the same value of $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ for two randomized designs with different unit-effects models and different numbers of units does not imply the same value of $\mathrm{cv}(\phi)$.

For NORDs and Truncated Completely Randomized Designs (TCRDs) (Cox 1982) under model (1.2.5) with given unit-effects model or given (n,q,r), both validity measures monotonically vary as the tuning constant c or the truncation constant κ increases. Then both measures can be used to aid in the selection of the tuning constant c in NORDs or the truncation constant κ in TCRDs. But the measure $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ is preferred since it is more easily estimated than is the measure $\mathrm{cv}(\phi)$.

The following U-statistics were suggested by Hooper (1993) to estimate μ_{val} and σ_{val} based on m replicates (G_i, C_i) of (G,C):

$$\hat{\mu}_{\mathrm{val}} = \frac{2}{m(m-1)} \sum_{i<j} \mathrm{val}(G_i, G_j) \tag{2.2.4}$$

and

$$\hat{\sigma}^2_{\text{val}} = \frac{1}{m(m-1)(m-2)(m-3)} \sum_{(i,j,k,l)\in\mathcal{P}(m,4)} \{\text{val}(G_i, G_j) - \text{val}(G_k, G_l)\}^2 \quad (2.2.5)$$

where $\mathcal{P}(m,4)$ is the set of permutations of four numbers from $\{1,\ldots,m\}$.

Hooper (1989) used the following method to estimate $\text{cv}(\phi)$. Let W_i and $G_{ij} (i = 1,\ldots,h; j = 1,\ldots,m)$ be independent replicates of W and G. Put $X_i = \sum_{j=1}^m \phi(G_{ij}W_i)$ and $Z_i = X_i(X_i-1)/m(m-1) - 2\,\alpha\, X_i/m + \alpha^2$. For h large, an approximate 95% confidence interval for $\text{var}\{E[\phi(G,W)\,|\,W]\}$ is given by

$$\max\{0, \bar{Z} - 2\,S_Z/\sqrt{h}\} \le \text{var}\{E[\phi(G,W)\,|\,W]\} \le \max\{0, \bar{Z} + 2\,S_Z/\sqrt{h}\}, \quad (2.2.6)$$

where $\bar{Z} = h^{-1}\sum_{i=1}^h Z_i$ and $S_Z^2 = (h-1)^{-1}\sum_{i=1}^h (Z_i - \bar{Z})^2$. A confidence interval for $\text{cv}(\phi)$ is obtained by transforming the bounds in (2.2.6). Hooper (1989) showed that when $\text{var}\{E[\phi(G,W)\,|\,W]\}$ is small it is desirable to take $\frac{h}{m}$ small for more precise estimation of $\text{cv}(\phi)$.

For classical randomized designs, we can generate a replicate of G at a little cost of computational time. However, it is much more expensive to generate a replicate of G for NORDs and TCRDs so that the above method will be not practical. Alternatively, a more efficient method is to employ the generalized U statistics to estimate $\text{cv}(\phi)$ based on m replicates G_i of G and h replicates W_i of W. Using a result in Hoeffding (1952), we have

$$\text{var}\{E[\phi(G,W)\,|\,W]\} = \text{cov}[\phi(G_1,W), \phi(G_1,W)].$$

Thus, $\text{var}\{E[\phi(G,W)\,|\,W]\}$ can be estimated by the unbiased generalized U-statistic:

$$\frac{2}{m(m-1)h} \sum_{i<j}^{m} \sum_{k=1}^{h} \phi(G_i,W_k)\phi(G_j,W_k) - \alpha^2. \quad (2.2.7)$$

We use square root of (2.2.7) divided by α to estimate $\text{cv}(\phi)$. The variance of the estimator (2.2.7) is expressed as

$$\frac{2}{m(m-1)h} \{\, 2(m-2)(h-1)\text{var}\{E[\phi(G_1,W)\phi(G_2,W)\,|\,G_1]\}$$

$$+ (h-1)\text{var}\{E[\phi(G_1,W)\phi(G_2,W) \mid G_1,G_2]\}$$
$$+ \frac{(m-2)(m-3)}{2}\text{var}\{E[\phi(G_1,W)\phi(G_2,W) \mid W]\}$$
$$+ 2(m-2)\text{var}\{E[\phi(G_1,W)\phi(G_2,W) \mid G_1,W]\}$$
$$+ \text{var}[\phi(G_1,W)\phi(G_2,W)]\ \} \qquad (2.2.8)$$

and can also be estimated by a generalized U-statistic (see Lee 1990).

The method in (2.2.6) and the generalized U-statistic method for $\text{cv}(\phi)$ require generating replicates of both G and W whereas (2.2.4) and (2.2.5) for $(\mu_{\text{val}}, \sigma_{\text{val}})$ only need replicates of G. Thus the validity measure $(\mu_{\text{val}}, \sigma_{\text{val}})$ is more easily estimated using simulation methods than is the validity measure $\text{cv}(\phi)$.

In order to seek a compromise between computation and interpretation, we will introduce a new validity measure in the next section. This new measure allows an indirect approximate interpretation and can be estimated at the same cost as the validity measure $(\mu_{\text{val}}, \sigma_{\text{val}})$.

2.3 A Validity Measure Related to the Beta Statistic

The F test is used for testing the hypothesis $H_0 : \beta = 0$ in normal theory analysis. In this section, we consider the F test and the test function $\phi(G,W)$ defined under the null hypothesis by (1.3.1) with $f(x) = \frac{x^T x}{q}$. For validity, we seek agreement between the randomization probability $E\{\phi(G,W) \mid W\}$ and the model-assisted probability $E\{\phi(G,W)\}$. Under model (1.2.5), the model-assisted distribution of the F statistic (1.2.12) is the F distribution. If the randomization distribution of the F statistic is close to the F distribution then the randomization probability $E\{\phi(G,W) \mid W\}$ should be close to the nominal size α, which is the model-assisted probability $E\{\phi(G,W)\}$.

It is difficult to work with the randomization distribution of the F statistic because its denominator is a function of the randomization. The denominator of the Beta statistic (1.2.13), however, is invariant under randomization, so the randomization distribution of the Beta statistic is determined by the numerator. Since the Beta statistic is equivalent to the F statistic, we consider agreement between the randomization and model-assisted distributions of the Beta statistic. In earlier papers, Welch (1937) and Pitman (1938) calculated the randomization moments of the Beta statistic and found close agreement with the corresponding normal theory moments for the randomized blocks design and the Latin square design.

It was shown in Section 1.2 that $SS_T = \|C^T GW\|^2$ and $SS_T + SS_E = \|(I - BB^T)W\|^2$ under H_0. Thus the Beta statistic (1.2.13) can be written:

$$T(G, W) = \frac{\|C^T GW\|^2}{\|(I - BB^T)W\|^2} \tag{2.3.1}$$

Since $W = KV$ and $K \in \mathcal{P}_n(B)$ under model (1.2.5), we have

$$T(G, KV) = \frac{\|C^T GKV\|^2}{\|(I - BB^T)V\|^2} \tag{2.3.2}$$

In Section 1.2, we showed that C is also a function of GK. Thus, from (2.3.2), $T(G, KV)$ is a function of GK and V. $T(G, W)$ has the Beta distribution with parameters $(\frac{q}{2}, \frac{n-r-q}{2})$ and is independent of GK. Thus the model-assisted distribution of the Beta statistic $T(G, W)$ is the Beta distribution with parameters $(\frac{q}{2}, \frac{n-r-q}{2})$, so the first two moments of model-assisted distribution of $T(G, W)$ are given by:

$$E\,[\,T(G, W)\,] = \frac{q}{n-r} \tag{2.3.3}$$

and

$$E\,[\,T^2(G, W)\,] = \frac{q(q+2)}{(n-r)(n-r+2)}. \tag{2.3.4}$$

Since the randomization moments of $T(G, W)$ are unbiased estimates of the corresponding model-assisted moments, the coefficients of variation of the first two randomization moments of $T(G, W)$ can be used to measure the extent of agreement

between the randomization and model-assisted distributions of $T(G,W)$. Therefore we define

$$\mathrm{cv}(T) = \frac{\mathrm{var}^{\frac{1}{2}}\{E[T(G,W)\,|\,W]\}}{E[T(G,W)]} \tag{2.3.5}$$

and

$$\mathrm{cv}(T^2) = \frac{\mathrm{var}^{\frac{1}{2}}\{E[T^2(G,W)\,|\,W]\}}{E[T^2(G,W)]}. \tag{2.3.6}$$

Based on the Beta-related validity measure (2.3.5) and (2.3.6), approximate 95% prediction intervals for the first two randomization moments $E[T(G,W)\,|\,W]$ and $E[T^2(G,W)\,|\,W]$ are given by

$$E(T) - 2E(T)\,\mathrm{cv}(T) \le E[T(G,W)\,|\,W] \le E(T) + 2E(T)\,\mathrm{cv}(T) \tag{2.3.7}$$

and

$$E(T^2) - 2E(T^2)\,\mathrm{cv}(T^2) \le E[T^2(G,W)\,|\,W] \le E(T^2) + 2E(T^2)\,\mathrm{cv}(T^2) \tag{2.3.8}$$

respectively.

The randomization distribution of $T(G,W)$ is discrete on the interval $[0,1]$ and cannot be explicitly determined for most randomization strategies. In order to compare the randomization distribution of $T(G,W)$ with the model-assisted one, we may approximate it by the Beta density curve:

$$p(t) = \text{const.} \times t^{m_1-1}(1-t)^{m_2-1}. \tag{2.3.9}$$

Here m_1 and m_2 are chosen so that the first two moments of this curve agree with the true moments of $T(G,W)$. We may find a confidence region for (m_1,m_2) through the validity measure $[\mathrm{cv}(T),\mathrm{cv}(T^2)]$ by converting the prediction intervals (2.3.7) and (2.3.8).

The Beta-related measure $[\mathrm{cv}(T),\mathrm{cv}(T^2)]$ can be estimated by simulation. The generalized U-statistic method used to estimate $\mathrm{cv}(\phi)$ in Section 2.2 can be adopted for the Beta-related measure $[\mathrm{cv}(T),\mathrm{cv}(T^2)]$.

However, that method requires replicates of both W and G. In the next section, we will derive a relation between the measures $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ and $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$. From this relation, we can estimate $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ without generating the replicates of W.

2.4 A Relationship Between measures $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ and $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$

The validity measure $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ introduced by Hooper (1993) is equivalent to the first two moments of the random variable $\|C_1^T G_1 G_2^T C_2\|^2$. The following theorem expresses the Beta-related validity measure (2.3.5) and (2.3.6) in terms of the first two moments of random variable $\|C_1^T G_1 G_2^T C_2\|^2$ and the first moment of $\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2$.

Theorem 2.4.1 *Under model* (1.2.5) *and the assumptions given in Section* 1.2, *we have*

$$\{\mathrm{cv}(T)\}^2 = \frac{2(n-r)}{q^2(n-r+2)} \left\{ E\|C_1^T G_1 G_2^T C_2\|^2 - \frac{q^2}{n-r} \right\} \tag{2.4.1}$$

and

$$\begin{aligned} \{\mathrm{cv}(T^2)\}^2 \;=\; & \frac{8(n-r)(n-r+2)}{q^2(q+2)^2(n-r+4)(n-r+6)} \left\{ (q+2)^2 [E\|C_1^T G_1 G_2^T C_2\|^2 - \frac{q^2}{n-r}] \right. \\ & + E\|C_1^T G_1 G_2^T C_2\|^4 + 2E\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2 \\ & \left. - \frac{q^2(q+2)^2}{(n-r)(n-r+2)} \right\} \end{aligned} \tag{2.4.2}$$

where (G_1, C_1) *and* (G_2, C_2) *are independent replicates of* (G, C).

The proof of Theorem 2.4.1 will be given in the next section since it requires some results on calculating moments of the distribution of two correlated quadratic forms of independent normal variables.

From Theorem 2.4.1, it follows that $\mathrm{cv}(T)$ only depends on μ_{val} and $\mathrm{cv}(T^2)$ is related to $E\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2$ as well as $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$. We can estimate the validity measure $(\mathrm{cv}(T), \mathrm{cv}(T^2))$ through estimating $E\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2$ and

$(\mu_{\text{val}}, \sigma_{\text{val}})$ using simulation methods. Thus the cost of computation for $[\text{cv}(T), \text{cv}(T^2)]$ is almost the same as that for $(\mu_{\text{val}}, \sigma_{\text{val}})$.

Optimal validity occurs when $[\text{cv}(T), \text{cv}(T^2)] = (0, 0)$. A randomized design with $[\text{cv}(T), \text{cv}(T^2)]$ close to $(0, 0)$ has a high degree of validity. From (2.4.1) and (2.4.2), we have that $\text{cv}(T) \to 0$ and $\text{cv}(T^2) \to 0$ if $E\|C_1^T G_1 G_2^T C_2\|^2 \to 0$ as $(n - r) \to \infty$.

Let $\lambda_1 \geq \cdots \geq \lambda_q$ be the ordered eigenvalues of $C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1$. We have $\lambda_1 \leq 1$, $\lambda_q \geq 0$,

$$\|C_1^T G_1 G_2^T C_2\|^2 = \sum_{i=1}^{q} \lambda_i$$

and

$$\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2 = \sum_{i=1}^{q} \lambda_i^2.$$

It follows that

$$E\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2 \leq E\|C_1^T G_1 G_2^T C_2\|^2$$

and

$$E\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2 \leq E\|C_1^T G_1 G_2^T C_2\|^4.$$

Thus, the convergence rate of $\text{cv}(T^2)$ is the same as that of $E\|C_1^T G_1 G_2^T C_2\|^2$ as $(n - r) \to \infty$. If $q = 1$, then $\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2 = \|C_1^T G_1 G_2^T C_2\|^4$, so the Beta-related measure $[\text{cv}(T), \text{cv}(T^2)]$ is equivalent to the measure $(\mu_{\text{val}}, \sigma_{\text{val}})$.

For the classical randomized designs, explicit expressions can be derived for the first two moments of $\|C_1^T G_1 G_2^T C_2\|^2$ and $E\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2$. Hooper (1993) showed that $E\|C_1^T G_1 G_2^T C_2\|^2 = \frac{q^2}{n-r}$ for the classical randomization strategies, such as CRD, randomized blocks design and Latin square design. Thus, $E[T(G, W) \mid W] = E[T(G, W)]$, that is, the first randomization moment of $T(G, W)$ does not depend on the uncontrolled random errors W, for classical randomized designs. In the following theorem, we give the explicit expressions for $E\|C_1^T G_1 G_2^T C_2\|^4$ and $E\|C_1^T G_1 G_2^T C_2 C_2^T G_2 G_1^T C_1\|^2$ for the CRD. Explicit expressions for randomized blocks

design and Latin square design can also be derived but more tedious calculations are required.

Under model (1.2.5), $B = (1/\sqrt{n})\mathbf{1}_n, r = 1$ and $C = A$ for the CRD. We then have that $C_1 = C_2, G_1 G_2^T \sim G$ and $C_1^T G_1 G_2^T C_2 \sim C^T G C$.

Theorem 2.4.2 *For the CRD, we have*

$$E\|C^T G C\|^4 = \frac{q}{n(n-1)(n-2)(n-3)}\{n^2(n+1)q^2a^2 - 2n(n-1)q(q+2)a + (q+2)^2 + n(n-3)(q^2+2)\} \tag{2.4.3}$$

and

$$E\|C^T G C C^T G^T C\|^2 = \frac{q}{n(n-1)(n-2)(n-3)}\{n^2(n+1)q^2a^2 - 2n(n-1)q(q+2)a + (q+2)^2 + n(n-3)(2q+1)\}, \tag{2.4.4}$$

where $A = (a_{ij})$ *and* $a = q^{-2}\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{q} a_{ij}^2 a_{ik}^2$.

The proof of Theorem 2.4.2 will be given in the next section. The equation (2.4.3) is implicit in Hooper (1993). (2.4.4) is derived by the same method as in Hooper (1993).

By (2.4.3) and (2.4.4), (2.4.2) is simplified for CRD:

$$[\mathrm{cv}(T^2)]^2 = \frac{24n(n+1)^2}{(q+2)^2(n-2)(n-3)(n+3)(n+5)}\left\{qa - \frac{(n-1)(q+2)}{n(n+1)}\right\}^2. \tag{2.4.5}$$

From (2.4.5), an appropriate choice of the matrix A can make $\mathrm{cv}(T^2)$ equal to 0.

2.5 Proof of Theorems

Before giving the proof of Theorem 2.4.1, I state some results on calculating the moments of the distribution of two correlated quadratic forms of independent normal

variables. The notation used before the proof of Theorem 2.4.1 in this section differs from that used elsewhere in this thesis. Let $A \otimes B$ be the Kronecker product of matrices A and B. Denote by E_{ij} the $n \times n$ matrix taking value 1 at the (i,j)th entry and 0 otherwise. Define $H_{ij} = E_{ij} + E_{ji}$.

Lemma 2.5.1 *Let $V \sim N(0, I_n)$. We have*

$$E[VV^T \otimes VV^T] = I_n \otimes I_n + \frac{1}{2}\sum_{ij} H_{ij} \otimes H_{ij} \tag{2.5.1}$$

and

$$\begin{aligned}
&E[VV^T \otimes VV^T \otimes VV^T \otimes VV^T] = I_n \otimes I_n \otimes I_n \otimes I_n \\
&+ \frac{1}{2}\sum_{ij}\{H_{ij} \otimes H_{ij} \otimes I_n \otimes I_n + H_{ij} \otimes I_n \otimes H_{ij} \otimes I_n + H_{ij} \otimes I_n \otimes I_n \otimes H_{ij} \\
&+ I_n \otimes H_{ij} \otimes H_{ij} \otimes I_n + I_n \otimes H_{ij} \otimes I_n \otimes H_{ij} + I_n \otimes I_n \otimes H_{ij} \otimes H_{ij}\} \\
&+ \sum_{ijk}\{H_{ij} \otimes H_{ik} \otimes H_{jk} \otimes I_n + H_{ij} \otimes H_{ik} \otimes I_n \otimes H_{jk} + H_{ij} \otimes I_n \otimes H_{ik} \otimes H_{jk} \\
&+ I_n \otimes H_{ij} \otimes H_{ik} \otimes H_{jk}\} + \frac{1}{4}\sum_{ijkl}\{H_{ij} \otimes H_{kl} \otimes H_{ij} \otimes H_{kl} \\
&+ H_{ij} \otimes H_{kl} \otimes H_{kl} \otimes H_{ij} + H_{ij} \otimes H_{ij} \otimes H_{kl} \otimes H_{kl}\} \\
&+ \sum_{ijkl}\{H_{ij} \otimes H_{kl} \otimes H_{ik} \otimes H_{jl} + H_{ij} \otimes H_{ik} \otimes H_{kl} \otimes H_{jl} \\
&+ H_{ij} \otimes H_{ik} \otimes H_{jl} \otimes H_{kl}\}
\end{aligned} \tag{2.5.2}$$

where $\sum_{ij}$ denotes the summation over the indices i and j, $\sum_{ijk}$ and $\sum_{ijkl}$ are defined in a similar manner.

The expression (2.5.1) was derived by Magnus and Neudecker (1979). Using the same method, I derived expression (2.5.2) by a straightforward calculation. From Lemma 2.5.1, the following results are immediately obtained.

Lemma 2.5.2 *Let $V \sim N(0, I_n)$ and let A, B, C and D be symmetric matrices of order n. Then*

$$E\{V^TAV \cdot V^TBV\} = \mathrm{tr}A\,\mathrm{tr}B + 2\mathrm{tr}AB$$

and

$$E\{V^TAV \cdot V^TBV \cdot V^TCV \cdot V^TDV\} = \mathrm{tr}A\,\mathrm{tr}B\,\mathrm{tr}C\,\mathrm{tr}D + 2\{\mathrm{tr}AB\,\mathrm{tr}C\,\mathrm{tr}D$$
$$+ \mathrm{tr}AC\,\mathrm{tr}B\,\mathrm{tr}D + \mathrm{tr}AD\,\mathrm{tr}B\,\mathrm{tr}C + \mathrm{tr}BC\,\mathrm{tr}A\,\mathrm{tr}D + \mathrm{tr}BD\,\mathrm{tr}A\,\mathrm{tr}C$$
$$+ \mathrm{tr}CD\,\mathrm{tr}A\,\mathrm{tr}B\} + 8\{\mathrm{tr}A\,\mathrm{tr}BCD + \mathrm{tr}B\,\mathrm{tr}ACD + \mathrm{tr}C\,\mathrm{tr}ABD + \mathrm{tr}D\,\mathrm{tr}ABC\}$$
$$+ 4\{\mathrm{tr}AB\,\mathrm{tr}CD + \mathrm{tr}AC\,\mathrm{tr}BD + \mathrm{tr}AD\,\mathrm{tr}BC\}$$
$$+ 16\{\mathrm{tr}ABCD + \mathrm{tr}ACBD + \mathrm{tr}ABDC\}.$$

Proof: We observe that

$$E\{V^TAV \cdot V^TBV\} = \mathrm{tr}\{[A \otimes B]\, E[VV^T \otimes VV^T]\}$$

and

$$E\{V^TAV \cdot V^TBV \cdot V^TCV \cdot V^TDV\} = \mathrm{tr}\{[A \otimes B \otimes C \otimes D]\, E[VV^T \otimes VV^T \otimes VV^T \otimes VV^T]\}.$$

the proof is completed by applying Lemma 2.5.1 and simplifying. ∎

The results in Lemma 2.5.2 can also be found in Rao and Kleffe (1988).

Lemma 2.5.3 *Under the assumptions of Lemma* 2.5.2 *, we have*

(*i*) $\mathrm{cov}(V^TAV, V^TBV) = 2\mathrm{tr}AB,$

(*ii*) $\mathrm{cov}\{(V^TAV)^2, (V^TBV)^2\} = 8\{\mathrm{tr}A\,\mathrm{tr}B\,\mathrm{tr}AB + 2\mathrm{tr}A\,\mathrm{tr}AB^2$
$+ 2\mathrm{tr}A^2B\,\mathrm{tr}B + (\mathrm{tr}AB)^2 + 4\mathrm{tr}A^2B^2 + 2\mathrm{tr}(AB)^2\},$

(*iii*) *if further A and B are idempotent with rank q, then*

$$\mathrm{cov}\{(V^TAV)^2, (V^TBV)^2\} = 8\{(q+2)^2\mathrm{tr}AB + (\mathrm{tr}AB)^2 + 2\mathrm{tr}(AB)^2\}$$

On the covariance of two correlated Beta random variables, we have the following result.

Lemma 2.5.4 *Let $V \sim N(0, \sigma^2 I_n)$. Let A and B be symmetric matrices of order n and let C be a symmetric idempotent matrix of order n. If $CA = A$ and $CB = B$, then $\left(\frac{V^TAV}{V^TCV}, \frac{V^TBV}{V^TCV}\right)$ is independent of V^TCV.*

Proof: Consider inference on σ^2 based on CV. Since $CV \sim N(0, \sigma^2 C)$, V^TCV is complete sufficient for σ^2. From $CA = A$ and $CB = B$, we obtain that V^TAV and V^TBV are functions of CV. Moreover, $\left(\frac{V^TAV}{V^TCV}, \frac{V^TBV}{V^TCV}\right)$ is distributed free of σ^2 and so is an ancillary statistic. By Basu's theorem (Lehmann, 1983), $\left(\frac{V^TAV}{V^TCV}, \frac{V^TBV}{V^TCV}\right)$ is independent of V^TCV. ∎

Lemma 2.5.5 *Let x, y and z be random variables. If (x, y) is independent of z, then we have that*

$$\operatorname{cov}(xz, yz) = E(z^2)\operatorname{cov}(x, y) + \operatorname{var}(z)\,E(x)\,E(y). \tag{2.5.3}$$

Proof:

$$\begin{aligned}
&\operatorname{cov}(xz, yz) = E\{[xz - E(xz)][yz - E(yz)]\} \\
= \;& E\{[xz - zE(x) + zE(x) - E(x)E(z)][yz - zE(y) + zE(y) - E(y)E(z)]\} \\
= \;& E\{z^2[x - E(x)][y - E(y)] + E(x)E(y)[z - E(z)]^2\} \\
= \;& E(z^2)\operatorname{cov}(x, y) + \operatorname{var}(z)\,E(x)\,E(y).
\end{aligned}$$

∎

Lemma 2.5.6 *Under the assumptions of Lemma 2.5.4, we have that*

$$\begin{aligned}
\operatorname{cov}\left\{\frac{(V^TAV)^i}{(V^TCV)^i}, \frac{(V^TBV)^i}{(V^TCV)^i}\right\} \;=\; & \frac{\operatorname{cov}[(V^TAV)^i, (V^TBV)^i]}{E\,[(V^TCV)^{2i}]} \\
& - \operatorname{var}[(V^TCV)^i]\,\frac{E[(V^TAV)^i]\,E[(V^TBV)^i]}{[E(V^TCV)^i]^2 E[(V^TCV)^{2i}]}, \\
& i = 1, 2, \ldots.
\end{aligned} \tag{2.5.4}$$

Proof: By Lemma 2.5.4, $\left[\frac{(V^TAV)^i}{(V^TCV)^i}, \frac{(V^TBV)^i}{(V^TCV)^i}\right]$ is independent of $(V^TCV)^i$, for $i = 1, 2, \ldots$. Applying Lemma 2.5.5, we have

$$\begin{aligned}\operatorname{cov}[(V^TAV)^i, (V^TBV)^i] &= \operatorname{cov}\left\{\frac{(V^TAV)^i}{(V^TCV)^i}(V^TCV)^i, \frac{(V^TBV)^i}{(V^TCV)^i}(V^TCV)^i\right\} \\ &= E[(V^TCV)^{2i}]\operatorname{cov}\left\{\frac{(V^TAV)^i}{(V^TCV)^i}, \frac{(V^TBV)^i}{(V^TCV)^i}\right\} \\ &\quad + \operatorname{var}[(V^TCV)^i]\, E\left[\frac{(V^TAV)^i}{(V^TCV)^i}\right] E\left[\frac{(V^TBV)^i}{(V^TCV)^i}\right], \\ &\qquad i = 1, 2, \ldots. \end{aligned} \tag{2.5.5}$$

Noting that $E\left[\frac{(V^TAV)^i}{(V^TCV)^i}\right] = \frac{E[(V^TAV)^i]}{E[(V^TCV)^i]}$ and $E\left[\frac{(V^TBV)^i}{(V^TCV)^i}\right] = \frac{E[(V^TBV)^i]}{E[(V^TCV)^i]}$, we obtain (2.5.4) by solving for $\operatorname{cov}\left[\frac{(V^TAV)^i}{(V^TCV)^i}, \frac{(V^TBV)^i}{(V^TCV)^i}\right]$ in (2.5.5). ∎

Proof of Theorem 2.4.1 :

By a result in Hoeffding (1952), we have

$$\operatorname{var}\{E[T^i(G, W) \mid W]\} = \operatorname{cov}\{T^i(G_1, W), T^i(G_2, W)\}, \qquad i = 1, 2, \ldots$$

where G_1 and G_2 are independent replicates of G. Note that

$$T(G, KV) = \frac{\|C^TGKV\|^2}{\|(I - BB^T)V\|^2}$$

and C depends on GK. Thus $T(G, KV)$ is also a function of GK and V. By conditioning and noting that $E[T^i(G_1, W) \mid G_1K]$ $(i = 1, 2, \ldots)$ are constant, we obtain

$$\begin{aligned}&\operatorname{cov}\{T^i(G_1, W), T^i(G_2, W)\} = E\{\operatorname{cov}[T^i(G_1, KV), T^i(G_2, KV) \mid G_1K, G_2K]\} \\ &= E\left\{\operatorname{cov}\left[\frac{\|C_1^TG_1KV\|^{2i}}{\|(I - BB^T)V\|^{2i}}, \frac{\|C_2^TG_2KV\|^{2i}}{\|(I - BB^T)V\|^{2i}} \mid G_1K, G_2K\right]\right\}, \quad i = 1, 2, \ldots.\end{aligned} \tag{2.5.6}$$

Furthermore, it follows from Lemma 2.5.4 that

$$\left(\frac{\|C_1^TG_1KV\|^{2i}}{\|(I - BB^T)V\|^{2i}}, \frac{\|C_2^TG_2KV\|^{2i}}{\|(I - BB^T)V\|^{2i}}\right)$$

is independent of $\|(I-BB^T)V\|^{2i}$ for given (G_1K, G_2K). By Lemma 2.5.6, we have that

$$\begin{aligned}
&\operatorname{cov}\left\{\frac{\|C_1^TG_1KV\|^{2i}}{\|(I-BB^T)V\|^{2i}}, \frac{\|C_2^TG_2KV\|^{2i}}{\|(I-BB^T)V\|^{2i}} \mid G_1K, G_2K\right\} \\
&= \frac{\operatorname{cov}[\|C_2^TG_2KV\|^{2i}, \|C_2^TG_2KV\|^{2i} \mid G_1K, G_2K]}{E[\|(I-BB^T)V\|^{4i}]} \\
&- \operatorname{var}[\|(I-BB^T)V\|^{2i}] \frac{E[\|C_2^TG_2KV\|^{2i} \mid G_1K]\, E[\|C_2^TG_2KV\|^{2i} \mid G_2K]}{[E(\|(I-BB^T)V\|^{2i})]^2\, E(\|(I-BB^T)V\|^{4i})}, \\
&\qquad\qquad i = 1, 2, \ldots. \qquad (2.5.7)
\end{aligned}$$

Since $V^T(I-BB^T)V$ is distributed with $\chi^2(n-r)$, then we have that

$$E\|(I-BB^T)V\|^{2i} = (n-r)(n-r+2)\cdots[n-r+2(i-1)], \qquad i = 1, 2, \ldots.$$

It follows that

$$\operatorname{var}[\|(I-BB^T)V\|^2] = 2(n-r) \qquad (2.5.8)$$

and

$$\operatorname{var}[\|(I-BB^T)V\|^4] = 8(n-r)(n-r+2)(n-r+3). \qquad (2.5.9)$$

Also $\|C^TGKV\|^2$ is distributed with $\chi^2(q)$ for given GK, so we have

$$E\|C^TGKV\|^{2i} = q(q+2)\cdots[q+2(i-1)], \qquad i = 1, 2, \ldots. \qquad (2.5.10)$$

Applying Lemma 2.5.3, we can obtain the conditional covariances in (2.5.7) for $i = 1, 2$:

$$\begin{aligned}
&\operatorname{cov}[\|C_1^TG_1KV\|^2, \|C_2^TG_2KV\|^2 \mid G_1K, G_2K] \\
&= 2\operatorname{tr}(K^TG_1^TC_1C_1^TG_1KK^TG_2^TC_2C_2^TG_2K) \\
&= 2\|C_1^TG_1G_2^TC_2\|^2 \qquad (2.5.11)
\end{aligned}$$

and

$$\operatorname{cov}[\|C_1^TG_1KV\|^2, \|C_2^TG_2KV\|^2 \mid G_1K, G_2K]$$

$$
\begin{aligned}
&= 8\{(q+2)^2\mathrm{tr}(K^TG_1^TC_1C_1^TG_1KK^TG_2^TC_2C_2^TG_2K) \\
&+ [\mathrm{tr}(K^TG_1^TC_1C_1^TG_1KK^TG_2^TC_2C_2^TG_2K)]^2 \\
&+ 2\mathrm{tr}(K^TG_1^TC_1C_1^TG_1KK^TG_2^TC_2C_2^TG_2K)^2\} \\
&= 8\{(q+2)^2\|C_1^TG_1G_2^TC_2\|^2 + \|C_1^TG_1G_2^TC_2\|^4 + 2\|C_1^TG_1G_2^TC_2C_2^TG_2G_1^TC_1\|^2\}.
\end{aligned}
\tag{2.5.12}
$$

Substituting (2.5.8), (2.5.9), (2.5.10), (2.5.11) and (2.5.12) back into (2.5.7) and using (2.5.6), we obtain that

$$
\begin{aligned}
&\mathrm{cov}[T(G_1,W),T(G_2,W)] \\
&= E\left\{\mathrm{cov}\left[\frac{\|C_1^TG_1KV\|^2}{\|(I-BB^T)V\|^2},\frac{\|C_2^TG_2KV\|^2}{\|(I-BB^T)V\|^2}\,|\,G_1K,G_2K\right]\right\} \\
&= E\left\{\frac{2\|C_1^TG_1G_2^TC_2\|^2}{(n-r)(n-r+2)} - 2(n-r)\frac{q\cdot q}{(n-r)^2(n-r)(n-r+2)}\right\} \\
&= \frac{2}{(n-r)(n-r+2)}\left\{E\|C_1^TG_1G_2^TC_2\|^2 - \frac{q^2}{n-r}\right\}
\end{aligned}
\tag{2.5.13}
$$

and

$$
\begin{aligned}
&\mathrm{cov}[T^2(G_1,W),T^2(G_2,W)] \\
&= E\left\{\mathrm{cov}\left[\frac{\|C_1^TG_1KV\|^4}{\|(I-BB^T)V\|^4},\frac{\|C_2^TG_2KV\|^4}{\|(I-BB^T)V\|^4}\,|\,G_1K,G_2K\right]\right\} \\
&= E\left\{\frac{8}{(n-r)(n-r+2)(n-r+4)(n-r+6)}\left[(q+2)^2\|C_1^TG_1G_2^TC_2\|^2\right.\right. \\
&+ \left.\|C_1^TG_1G_2^TC_2\|^4 + 2\|C_1^TG_1G_2^TC_2C_2^TG_2G_1^TC_1\|^2\right] \\
&- \left.\frac{(n-r)(n-r+2)(n-r+3)q^2(q+2)^2}{(n-r)^2(n-r+2)^2(n-r)(n-r+2)(n-r+4)(n-r+6)}\right\} \\
&= \frac{8}{(n-r)(n-r+2)(n-r+4)(n-r+6)}\left\{(q+2)^2\left[E\|C_1^TG_1G_2^TC_2\|^2 - \frac{q^2}{n-r}\right]\right. \\
&+ \left. E\|C_1^TG_1G_2^TC_2\|^4 + 2E\|C_1^TG_1G_2^TC_2C_2^TG_2G_1^TC_1\|^2 - \frac{q^2(q+2)^2}{(n-r)(n-r+2)}\right\}
\end{aligned}
\tag{2.5.14}
$$

Therefore, combining (2.5.13) and (2.5.14) with (2.3.3) and (2.3.4) yields at once (2.4.1) and (2.4.2). This completes the proof.

Proof of Theorem 2.4.2:

Since $B = (1/\sqrt{n})\mathbf{1}_n, r = 1$ and $C = A$ for CRD, we then have $C_1 = C_2$ and $G_1 G_1^T \sim G$. Thus $C_1^T G_1 G_2^T C_2 \sim C^T G C$ and G is uniformly distributed on the group $\mathcal{P}_n$. Part of the following derivation can be found in Hooper (1993).

We wish to calculate

$$E(\|C^T G C\|^4) = \sum\sum\sum\sum E\{(a^T G b)^2 (c^T G d)^2 \tag{2.5.15}$$

and

$$E(\|C^T G C C^T G^T C\|^2) = \sum\sum\sum\sum E\{a^T G c\, b^T G c\, a^T G d\, b^T G d\}, \tag{2.5.16}$$

where the summations are over columns a, b, c and d of C. Let $\mathcal{S}_4$ denote the set of $\mathcal{P}_n$-orbits on $\{1, \ldots, n\}^4$, i.e. each $S \in \mathcal{S}_4$ has the form $S = \mathcal{P}_n(i,j,k,l) = \{(gi, gj, gk, gl) : g \in \mathcal{P}_n\}$. Let $a = (a_1, \ldots, a_n)^T, b = (b_1, \ldots, b_n)^T, c = (c_1, \ldots, c_n)^T$ and $d = (d_1, \ldots, d_n)^T$. For every $g \in \mathcal{P}_n$, a simple calculation leads to

$$(a^T g b)^2 (c^T g d)^2 = \sum_i \sum_j \sum_k \sum_l a_i a_j c_k c_l b_{g^{-1}i} b_{g^{-1}j} d_{g^{-1}k} d_{g^{-1}l}$$

and

$$a^T g c\, b^T g c\, a^T g d\, b^T g d = \sum_i \sum_j \sum_k \sum_l a_i b_j a_k b_l c_{g^{-1}i} c_{g^{-1}j} d_{g^{-1}k} d_{g^{-1}l}.$$

Let $\#S$ denote number of elements in S. Noting that G has the uniform distribution on the group $\mathcal{P}_n$, we then have

$$\begin{aligned}
E\{(a^T G b)^2 (c^T G d)^2\} &= \frac{1}{\#\mathcal{P}_n} \sum_{g \in \mathcal{P}_n} \left\{ \sum_i \sum_j \sum_k \sum_l a_i a_j c_k c_l b_{g^{-1}i} b_{g^{-1}j} d_{g^{-1}k} d_{g^{-1}l} \right\} \\
&= \frac{1}{\#\mathcal{P}_n} \sum_{g \in \mathcal{P}_n} \left\{ \sum_i \sum_j \sum_k \sum_l a_i a_j c_k c_l b_{gi} b_{gj} d_{gk} d_{gl} \right\} \\
&= \sum_i \sum_j \sum_k \sum_l \left\{ \frac{1}{\#\mathcal{P}_n} \sum_{g \in \mathcal{P}_n} a_i a_j c_k c_l b_{gi} b_{gj} d_{gk} d_{gl} \right\} \\
&= \sum_{S \in \mathcal{S}_4} \sum_{(i,j,k,l) \in S} a_i a_j c_k c_l \{ \frac{1}{\#\mathcal{P}_n} \sum_{g \in \mathcal{P}_n} b_{gi} b_{gj} d_{gk} d_{gl} \}
\end{aligned}$$

$$= \sum_{S\in\mathcal{S}_4}\left\{\sum_{(i,j,k,l)\in S} a_ia_jc_kc_l\right\}\left\{\frac{1}{\#\mathcal{P}_n}\frac{\#\mathcal{P}_n}{\#S}\sum_{(i,j,k,l)\in S} b_ib_jd_kd_l\right\}$$

$$= \sum_{S\in\mathcal{S}_4}\frac{1}{\#S}\left\{\sum_{(i,j,k,l)\in S} a_ia_jc_kc_l\right\}\left\{\sum_{(i,j,k,l)\in S} b_ib_jd_kd_l\right\}. \quad (2.5.17)$$

Similarly, we can obtain

$$E\{a^TGcb^TGca^TGdb^TGd\} = \sum_{S\in\mathcal{S}_4}\frac{1}{\#S}\left\{\sum_{(i,j,k,l)\in S} c_ic_jd_kd_l\right\}\left\{\sum_{(i,j,k,l)\in S} a_ib_ja_kb_l\right\}. \quad (2.5.18)$$

For $S \in \mathcal{S}_4$ and two columns a and b of C, define

$$u(S) = \sum_{(i,j,k,l)\in S} a_ia_jb_kb_l$$

and

$$w(S) = \sum_{(i,j,k,l)\in S} a_ib_ja_kb_l.$$

There are 15 $\mathcal{P}_n$-orbits on $\{1,\ldots,n\}^4$:

$S_1 = \mathcal{P}_n(1,1,1,1), \#S_1 = n, u(S_1) = w(S_1) = \sum_{i=1}^n a_i^2b_i^2$;

$S_{21} = \mathcal{P}_n(2,1,1,1), S_{22} = \mathcal{P}_n(1,2,1,1), S_{23} = \mathcal{P}_n(1,1,2,1)$ and $S_{24} = \mathcal{P}_n(1,1,1,2)$, $\#S_{2k} = n(n-1), k = 1,\ldots,4, u(S_{2k}) = w(S_{2k}) = -\sum_{i=1}^n a_i^2b_i^2, k = 1,\ldots,4$;

$S_{31} = \mathcal{P}_n(1,1,2,2), S_{32} = \mathcal{P}_n(1,2,1,2)$ and $S_{33} = \mathcal{P}_n(1,2,2,1)$, $\#S_{3k} = n(n-1)$, $k = 1,2,3$, if $a = b, u(S_{3k}) = w(S_{3k}) = 1 - \sum_{i=1}^n a_i^4, k = 1,2,3$, if $a \neq b, u(S_{31}) = w(S_{32}) = 1 - \sum_{i=1}^n a_i^2b_i^2, u(S_{32}) = w(S_{31}) = u(S_{33}) = w(S_{33}) = -\sum_{i=1}^n a_i^2b_i^2$;

$S_{41} = \mathcal{P}_n(1,1,2,3), S_{42} = \mathcal{P}_n(1,2,1,3), S_{43} = \mathcal{P}_n(1,2,3,1), S_{44} = \mathcal{P}_n(2,1,1,3)$, $S_{45} = \mathcal{P}_n(2,1,3,1)$ and $S_{46} = \mathcal{P}_n(2,3,1,1), \#S_{4k} = n(n-1)(n-2), k = 1,\ldots,6$, if $a = b, u(S_{4k}) = w(S_{4k}) = 2\sum_{i=1}^n a_i^4 - 1, k = 1,\ldots,6$, if $a \neq b, u(S_{41}) = w(S_{42}) = u(S_{46}) = w(S_{45}) = 2\sum_{i=1}^n a_i^2b_i^2 - 1$, the remaining $u(S_{4k}) = w(S_{4k}) = 2\sum_{i=1}^n a_i^2b_i^2$;

$S_5 = \mathcal{P}_n(1,2,3,4), \#S_5 = n(n-1)(n-2)(n-3)$, if $a = b, u(S_5) = w(S_5) = 1 - 6\sum_{i=1}^n a_i^2b_i^2$, if $a \neq b, u(S_5) = w(S_5) = 3(1 - 2\sum_{i=1}^n a_i^4)$.

Substituting the above $u(S)$'s and $w(S)$'s into (2.5.17) and (2.5.18), we obtain (2.4.3) and (2.4.4). (2.4.5) comes out immediately from the substitution of (2.5.17) and (2.5.18) into (2.4.2) and after simplification. This completes the proof.

2.6 An Analytical Approximation to cv(ϕ)

The validity measures cv(ϕ) and ($\mu_{\text{val}}, \sigma_{\text{val}}$) were reviewed and discussed with regard to interpretation and computation in Section 2.2. In this section, we will develop an analytical approximation to cv(ϕ), where ϕ is the F-test, to study the relationship between the two measures under the null hypothesis.

In the following development, we approximate a joint distribution of two correlated F-ratios of normal vectors by joint distribution of two correlated quadratic forms of normal vectors. In fact, this means that we approximate cv(ϕ) by cv(ϕ_χ), where ϕ_χ is the χ^2-test used when the standard error σ of the error term V in model (1.2.5) is assumed known. But an empirical study in Hooper (1989) shows that cv(ϕ_χ) decreases very slowly as the experiment size n increases. This means that the approximation of cv(ϕ_χ) to cv(ϕ) is very poor for small and even moderate size n. I also conducted an empirical study comparing the joint distribution of two correlated F-ratios of normal vectors with the joint distribution of two correlated quadratic forms of normal vectors. The result shows that a large value of n is required to obtain a good approximation. Consequently, I do not recommend using this analytical approximation in practice. However, it does provide some insight on the relationship between cv(ϕ) and ($\mu_{\text{val}}, \sigma_{\text{val}}$).

The F statistic (1.2.12) can be written:

$$F(G,W) = \frac{(n-r-q)}{q}\frac{\|C^TGW\|^2}{\|(I-XX^T)GW\|^2} = \frac{(n-r-q)}{q}\frac{\|C^TGKV\|^2}{\|(I-XX^T)GKV\|^2}.$$

The F test is then defined by

$$\phi_F(G,W) = \begin{cases} 1 & F(G,W) > c \\ 0 & F(G,W) \le c \end{cases} \tag{2.6.1}$$

where c is a constant such that the normal theory significance level of (2.6.1) is α, i.e. $E[\phi_F(G,W)] = \alpha$. Hence, the coefficient of variation validity measure is

$$\text{cv}(\phi_F) = \alpha^{-1}\text{var}^{\frac{1}{2}}\{E[\phi_F(G,W)\,|\,W]\}. \tag{2.6.2}$$

By conditioning, we have

$$\begin{aligned}\text{var}\{E[\phi_F(G,W)|W]\} &= \text{cov}\{\phi_F(G_1,W),\phi_F(G_2,W)\} \\ &= E\{\text{cov}[\phi_F(G_1,W),\phi_F(G_2,W)\,|\,G_1K,G_2K]\} \\ &= E\{P[F(G_1,W)>c,F(G_2,W)>c\,|\,G_1K,G_2K]\}-\alpha^2.\end{aligned}$$

Here, the second equality comes from the fact that $E[\phi_F(G_1,W)|\,G_1K]$ and $E[\phi_F(G_2,W)|\,G_2K]$ are constant and then

$$\text{cov}\{E[\phi_F(G_1,W)|\,G_1K],E[\phi_F(G_2,W)|\,G_2K]\}=0.$$

Since $\frac{\|(I-XX^T)GKV\|^2}{n-r-q}$ converges in probability to σ^2 under some mild conditions (see Hooper 1989), we approximate

$$P[F(G_1,W)>c,F(G_2,W)>c\,|\,G_1K,G_2K] \tag{2.6.3}$$

by

$$P[\|C_1^TG_1KV\|^2>cq\sigma^2,\|C_2^TG_2KV\|^2>cq\sigma^2\,|\,G_1K,G_2K] \tag{2.6.4}$$

when n is sufficiently large (an empirical result shows this approximation is poor for small and even moderate n).

Quantity (2.6.4) involves the joint distribution of two correlated quadratic forms of normal vectors. A series expansion is derived for this probability in the next section.

For $\alpha>-1$, the Laguerre polynomial of degree k is defined by

$$\begin{aligned}L_k^\alpha(x) &= \frac{x^{-\alpha}e^x}{k!}\frac{d^k}{dx^k}(x^{\alpha+k}e^{-x}) \\ &= \sum_{j=0}^{k}\frac{\Gamma(\alpha+k+1)}{\Gamma(\alpha+j+1)}\frac{(-x)^j}{j!(k-j)!};\end{aligned}$$

see Szegö (1975) for more details. Let

$$a_k=\frac{\Gamma(k)}{\Gamma(k+\frac{q}{2})}(\frac{cq}{2})^{\frac{q}{2}}e^{-\frac{cq}{2}}L_{k-1}^{\frac{q}{2}}(\frac{cq}{2}). \tag{2.6.5}$$

Moreover, we define

$$\mu'_k = \sum_{m=1}^{k} \sum_{i_1 j_1 + \cdots + i_m j_m = k} \left(\frac{\mathrm{tr}M^{i_1}}{2i_1}\right)^{j_1} \cdots \left(\frac{\mathrm{tr}M^{i_m}}{2i_m}\right)^{j_m} \frac{k!}{j_1! j_2! \cdots j_m!} \tag{2.6.6}$$

with $i_1 < \cdots < i_m$ for $k \geq 1$, where $M = C_2^T G_2 G_1^T C_1 C_1^T G_1 G_2^T C_2$. We have from results in the next section that

$$P[\|C_1^T G_1 K V\|^2 > cq\sigma^2, \|C_2^T G_2 K V\|^2 > cq\sigma^2 \mid G_1 K, G_2 K] = \alpha^2 + \sum_{k=1}^{\infty} \frac{\mu'_k}{k!} a_k^2 .$$

Hence, we approximate $\mathrm{var}\{E[\phi_F(G, W)|W]\}$ by

$$\sum_{k=1}^{\infty} \frac{E(\mu'_k)}{k!} a_k^2 \tag{2.6.7}$$

For the case $q = 1$, (2.6.7) has the following simple form:

$$\sum_{k=1}^{\infty} \frac{\Gamma(k + \frac{1}{2})}{k!\Gamma(\frac{1}{2})} a_k^2 E\|C_1^T G_1 G_2^T C_2\|^{2k} . \tag{2.6.8}$$

From (2.6.5), we see that a_k depends solely on the critical value of the normal theory significance level and the number of treatments. In (2.6.6), μ'_k is a combinatorial function of the random matrix M, so the expectation $E(\mu'_k)$ is determined solely by the randomization strategy. The first two terms in (2.6.7) involve $\mathrm{tr}M, (\mathrm{tr}M)^2$, and $\mathrm{tr}(M^2)$. The measure $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ is equivalent to $[E(\mathrm{tr}M), E(\mathrm{tr}M)^2]$. Usually, the convergence of series (2.6.7) is determined by both of the term $\frac{a_k^2}{k!}$ and the expectation $E(\mu'_k)$. If the randomized design is such that $E(\mathrm{tr}M)$ is small, the series (2.6.7) converges quickly and then only the first few terms need to be calculated to approximate $\mathrm{var}\{E[\phi_F(G, W)|W]\}$.

Since $E\|C_1^T G_1 G_2^T C_2\|^{2k} \leq 1$ in the case $q = 1$, then the series (2.6.8) is controlled by

$$\sum_{k=1}^{\infty} \frac{\Gamma(k + \frac{1}{2})}{k!\Gamma(\frac{1}{2})} a_k^2. \tag{2.6.9}$$

Table 2.1 lists the numerical values of first 15 terms and their partial sums in the series (2.6.9) for $c = 2.706, 3.841$ and 6.635, which are the $90, 95$ and 99 percentiles of χ_1^2 respectively. From Table 2.1, we find that the series (2.6.8) converges very quickly.

Table 2.1: The numerical values for the series (2.6.9)

Order	$c = 2.706$		$c = 3.841$		$c = 6.635$	
k	term	partial sum	term	partial sum	term	partial sum
1	.0287705	.0287705	.0131262	.0131262	.0013871	.0013871
2	.0001036	.0288741	.0003868	.0135130	.0007637	.0021508
3	.0004484	.0293225	.0006831	.0141961	.0000517	.0022025
4	.0003640	.0296865	.0001659	.0143621	.0000143	.0022168
5	.0001259	.0298124	.0000089	.0143710	.0000201	.0022369
6	.0000245	.0298369	.0000022	.0143731	.0000047	.0022416
7	.0000016	.0298386	.0000064	.0143796	.0000001	.0022416
8	.0000002	.0298387	.0000045	.0143841	.0000003	.0022419
9	.0000010	.0298397	.0000018	.0143860	.0000004	.0022424
10	.0000010	.0298407	.0000005	.0143864	.0000002	.0022426
11	.0000006	.0298413	.0000001	.0143865	.0000000	.0022426
12	.0000003	.0298415	.0000000	.0143865	.0000000	.0022426
13	.0000001	.0298416	.0000000	.0143865	.0000000	.0022426
14	.0000000	.0298417	.0000000	.0143865	.0000000	.0022426
15	.0000000	.0298417	.0000000	.0143865	.0000000	.0022426

2.7 Joint Probability of Two Correlated Quadratic Forms

The notation in this section is self-contained though I try to avoid inconsistency with the remainder of the thesis.

Kibble (1941) expressed the joint density function of two correlated chi-square variates as a series of the products of Laguerre polynomials. Krishnamoorthy and Parthasarathy (1951) generalized Kibble's result to the multivariate correlated chi-square distribution. Jensen (1970) derived the joint density function for the generalized multivariate correlated chi-square distribution. The Laguerre series expansion for the joint density function of the correlated quadratic forms was derived by Khatri, Krishnaiah and Sen (1977). The results of Kibble (1941), Krishnamoorthy and Parthasarathy (1951) and Jensen (1970) are the special cases of Khatri, Krishnaiah and Sen (1977).

Without loss of generality, we assume $\sigma^2 = 1$ in this section so that $V \sim N(0, I_n)$. Define $Y_1 = V^T C_1 C_1^T V$ and $Y_2 = V^T C_2 C_2^T V$ where C_1 and C_2 are two fixed $n \times q$ matrices with orthonormal columns. Applying the result of Khatri, Krishnaiah and Sen (1977) we can find the joint density function for Y_1 and Y_2. However much simplification is required to obtain a clear expression. It is easier to derive the joint density function for Y_1 and Y_2 by adapting the methods used in Kibble (1941), Krishnamoorthy and Parthasarathy (1951) and Jensen (1970). Furthermore, we find the Laguerre series expansion for the joint distribution function of Y_1 and Y_2. Following is the detailed procedure of derivation.

The following Lemma is well known; see Rao (1973).

Lemma 2.7.1 *Assume that* $U = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N(0, \Sigma)$, *where* $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. *Let* $W_1 = {U_1}^T U_1$ *and* $W_2 = {U_2}^T U_2$. *Then the joint characteristic function of* W_1 *and* W_2

is given by

$$E\{\exp(it_1W_1 + it_2W_2)\} = |I - 2iT\Sigma|^{-\frac{1}{2}} \tag{2.7.1}$$

where $T = \begin{pmatrix} t_1 I & 0 \\ 0 & t_2 I \end{pmatrix}$.

Applying Lemma 2.7.1, we have that the joint c.f. of Y_1 and Y_2 is given by

$$\varphi(t_1, t_2) = (1-2it_1)^{-\frac{q}{2}}(1-2it_2)^{-\frac{q}{2}}|I_q - \frac{2it_1 2it_2}{(1-2it_1)(1-2it_2)}C_2^T C_1 C_1^T C_2|^{-\frac{1}{2}} \tag{2.7.2}$$

In order to expand (2.7.2) as a power series of the variable $\frac{2it_1 2it_2}{(1-2it_1)(1-2it_2)}$, we need the following lemma, which can be found in Jensen (1970).

Lemma 2.7.2 *Assume that M is a symmetric matrix with the largest eigenvalue less than 1. Let $\psi(t) = |I - tM|^{-\frac{1}{2}}$, where t can be a real or complex variable. Then we have*

$$\log\psi(t) = \sum_{j=1}^{\infty} \mathrm{tr}M^j \frac{t^j}{2j}$$

and

$$\psi(t) = \sum_{k=0}^{\infty} \mu'_k \frac{t^k}{k!},$$

where $\mu'_0 = 1$ and

$$\mu'_k = \sum_{m=1}^{k} \sum_{i_1 j_1 + \cdots + i_m j_m = k} \left(\frac{\mathrm{tr}M^{i_1}}{2i_1}\right)^{j_1} \cdots \left(\frac{\mathrm{tr}M^{i_m}}{2i_m}\right)^{j_m} \frac{k!}{j_1! j_2! \cdots j_m!} \tag{2.7.3}$$

with $i_1 < \cdots < i_m$ for $k \geq 1$.

By Lemma 2.7.2, we can obtain the power series expansion for the characteristic function of the two correlated quadratic forms.

Theorem 2.7.3 *Assume that $V \sim N(0, I_n)$. Define $Y_1 = V^T C_1 C_1^T V$ and $Y_2 = V^T C_2 C_2^T V$ where C_1 and C_2 are two fixed $n \times q$ matrices with orthonormal columns. Then the joint c.f. of Y_1 and Y_2 is given by*

$$\varphi(t_1, t_2) = \sum_{k=0}^{\infty} \frac{\mu'_k}{k!} \frac{(2it_1)^k (2it_2)^k}{(1-2it_1)^{k+\frac{q}{2}}(1-2it_2)^{k+\frac{q}{2}}} \tag{2.7.4}$$

where μ'_k is given in (2.7.3) with $M = C_2^T C_1 C_1^T C_2$.

Lemma 2.7.4 *Let W be distributed with χ^2_q and $L^{\alpha}_n(x)$ be the Laguerre polynomial of degree n. Then*

$$\frac{(-1)^k k!\Gamma(\frac{q}{2})}{\Gamma(k+\frac{q}{2})} E\{L_k^{\frac{q}{2}-1}(\frac{W}{2})\exp(itW)\} = \frac{(2it)^k}{(1-2it)^{k+\frac{q}{2}}} \tag{2.7.5}$$

Proof: From the definition of Laguerre polynomial, we know that

$$\begin{aligned} L^{\alpha}_n(x) &= \frac{x^{-\alpha}e^x}{n!}\frac{d^n}{dx^n}(x^{\alpha+n}e^{-x}) \\ &= \sum_{k=0}^{n}\frac{\Gamma(\alpha+n+1)}{\Gamma(\alpha+k+1)}\frac{(-x)^k}{k!(n-k)!}, \qquad n=0,1,2,\cdots \end{aligned}$$

Thus,

$$\begin{aligned} E\{L_k^{\frac{q}{2}-1}(\frac{W}{2})\exp(itW)\} &= \int_0^{\infty} L_k^{\frac{q}{2}-1}(\frac{w}{2})\exp(itw)\frac{1}{2^{\frac{q}{2}}\Gamma(\frac{q}{2})}w^{\frac{q}{2}-1}e^{-\frac{w}{2}}\,dw \\ &= \frac{1}{\Gamma(\frac{q}{2})}\int_0^{\infty} L_k^{\frac{q}{2}-1}(y)e^{(2it-1)y}y^{\frac{q}{2}-1}\,dy \\ &= \frac{\Gamma(k+\frac{q}{2})}{\Gamma(\frac{q}{2})}\sum_{j=0}^{k}\frac{(-1)^j}{\Gamma(j+\frac{q}{2})j!(k-j)!}\int_0^{\infty} y^{j+\frac{q}{2}-1}e^{-(1-2it)y}\,dy \\ &= \frac{\Gamma(k+\frac{q}{2})}{\Gamma(\frac{q}{2})}\sum_{j=0}^{k}\frac{(-1)^j}{j!(k-j)!}(1-2it)^{-(j+\frac{q}{2})} \\ &= \frac{\Gamma(k+\frac{q}{2})}{\Gamma(\frac{q}{2})k!(1-2it)^{k+\frac{q}{2}}}\sum_{j=0}^{k}\binom{k}{j}(-1)^j(1-2it)^{k-j} \\ &= \frac{\Gamma(k+\frac{q}{2})}{\Gamma(\frac{q}{2})k!}\frac{(-1)^k(2it)^k}{(1-2it)^{k+\frac{q}{2}}} \end{aligned}$$

■

Combining Theorem 2.7.3 and Lemma 2.7.4 and using the inversion theorem, we can obtain the Laguerre series expansion for the joint density function of Y_1 and Y_2 in the following theorem. The proof for the absolute convergence in the theorem is similar to that in Krishnamoorthy and Parthasarathy (1951).

Theorem 2.7.5 *Using the same assumptions in Theorem* 2.7.3 , *we have that the joint density function of* Y_1 *and* Y_2 *is given by*

$$f(y_1,y_2)=f(y_1)f(y_2)\sum_{k=0}^{\infty}\frac{\mu_k'}{k!}\left[\frac{\Gamma(\frac{q}{2})k!}{\Gamma(k+\frac{q}{2})}\right]^2 L_k^{\frac{q}{2}-1}(\frac{y_1}{2})L_k^{\frac{q}{2}-1}(\frac{y_2}{2}),\quad y_1\geq 0,y_2\geq 0. \tag{2.7.6}$$

If the largest eigenvalue of $M=C_2^TC_1C_1^TC_2$ *is less than 1, then the series in* (2.7.6) *is absolutely convergent for* $y_1\geq 0,y_2\geq 0$.

Proof: (2.7.6) comes immediately from Theorem 2.7.3 and Lemma 2.7.4. It is trivial to show the absolute convergence of (2.7.6) for at least one of y_1 and y_2 equal to zero. Following is the proof of the absolute convergence of (2.7.6) for $y_1>0$ and $y_2>0$. Let ρ be the largest eigenvalue of $M=C_2^TC_1C_1^TC_2$. Then $\mathrm{tr}M^i\leq q\rho^i$, for all integer i. Noting a combinatorial identity:

$$\frac{\Gamma(p+k)}{\Gamma(p)}=\sum_{m=1}^{k}\sum_{i_1j_1+\cdots+i_mj_m=k}\frac{p^{j_1+j_2+\cdots+j_m}}{i_1^{j_1}\cdots i_m^{j_m}}\frac{k!}{j_1!\cdots j_m!}$$

we have

$$\begin{aligned}
\mu_k' &= \sum_{m=1}^{k}\sum_{i_1j_1+\cdots+i_mj_m=k}\left(\frac{\mathrm{tr}M^{i_1}}{2i_1}\right)^{j_1}\cdots\left(\frac{\mathrm{tr}M^{i_m}}{2i_m}\right)^{j_m}\frac{k!}{j_1!j_2!\cdots j_m!}\\
&\leq \sum_{m=1}^{k}\sum_{i_1j_1+\cdots+i_mj_m=k}\left(\frac{q\rho^{i_1}}{2i_1}\right)^{j_1}\cdots\left(\frac{q\rho^{i_m}}{2i_m}\right)^{j_m}\frac{k!}{j_1!j_2!\cdots j_m!}\\
&= \rho^k\sum_{m=1}^{k}\sum_{i_1j_1+\cdots+i_mj_m=k}\frac{(\frac{q}{2})^{j_1+j_2+\cdots+j_m}}{i_1^{j_1}\cdots i_m^{j_m}}\frac{k!}{j_1!\cdots j_m!}\\
&= \rho^k\frac{\Gamma(\frac{q}{2}+k)}{\Gamma(\frac{q}{2})}
\end{aligned}$$

Hille (1926) deduced that

$$L_k^{\alpha}(x)=\frac{1}{\sqrt{\pi}}e^{\frac{x}{2}}x^{-\frac{1}{2}(\alpha+\frac{1}{2})}k^{\frac{1}{2}(\alpha-\frac{1}{2})}\cos[2\sqrt{kx}-\pi(\frac{1}{4}+\frac{\alpha}{2})]+O(k^{\frac{1}{2}(\alpha-\frac{1}{2})}),\ k\to\infty. \tag{2.7.7}$$

So, it follows that

$$L_k^{\frac{q}{2}-1}(\frac{y}{2}) = \frac{2^{\frac{1}{4}(q-1)}}{\sqrt{\pi}} e^{\frac{y}{4}} y^{-\frac{1}{4}(q-1)} k^{\frac{1}{4}(q-3)} \cos[\sqrt{2ky} - \frac{\pi}{4}(q-1)] + O(k^{\frac{1}{4}(q-4)}), \ k \to \infty. \tag{2.7.8}$$

It is well known that

$$\frac{\Gamma(k+\frac{q}{2})}{k!} \approx k^{\frac{q}{2}-1} \tag{2.7.9}$$

Combining (2.7.8) with (2.7.9), we conclude that

$$\left[\frac{k!}{\Gamma(k+\frac{q}{2})}\right]^2 \left| L_k^{\frac{q}{2}-1}(\frac{y_1}{2}) L_k^{\frac{q}{2}-1}(\frac{y_2}{2}) \right| \leq \tau_1(y_1, y_2, q) k^{-\frac{1}{2}(q-1)}, \quad k \to \infty,$$

where $\tau_1(y_1, y_2, q)$ is a constant depending on y_1, y_2 and q. Hence, we obtain

$$\frac{\mu_k'}{k!} \left[\frac{k!}{\Gamma(k+\frac{q}{2})}\right]^2 \left| L_k^{\frac{q}{2}-1}(\frac{y_1}{2}) L_k^{\frac{q}{2}-1}(\frac{y_2}{2}) \right| \leq \frac{\tau_2(y_1, y_2, q)}{\Gamma(\frac{q}{2})} \rho^k k^{-\frac{1}{2}}, \quad k \to \infty,$$

where $\tau_2(y_1, y_2, q)$ is another constant depending on y_1, y_2 and q. Therefore, if $\rho < 1$, the series in (2.7.6) is absolutely convergent for $y_1 > 0$ and $y_2 > 0$. ■

Lemma 2.7.6 *Let $f(y)$ be the density function of χ^2 with q degrees of freedom. Then for any positive constant c, we have*

$$\int_c^\infty L_k^{\frac{q}{2}-1}(\frac{y}{2}) f(y)\, dy = -\frac{e^{-\frac{c}{2}}(\frac{c}{2})^{\frac{q}{2}}}{\Gamma(\frac{q}{2})k} L_{k-1}^{\frac{q}{2}}(\frac{c}{2}) \tag{2.7.10}$$

Proof: Transforming $x = \frac{y}{2}$, we rewrite the left hand side of (2.7.10) as follows:

$$\begin{aligned}
\int_c^\infty L_k^{\frac{q}{2}-1}(\frac{y}{2}) f(y)\, dy &= \int_{\frac{c}{2}}^\infty L_k^{\frac{q}{2}-1}(x) f(2x) 2\, dx \\
&= \int_{\frac{c}{2}}^\infty \left[\sum_{j=0}^k \frac{\Gamma(k+\frac{q}{2})}{\Gamma(j+\frac{q}{2})} \frac{(-x)^j}{j!(k-j)!}\right] \frac{x^{\frac{q}{2}-1} e^{-x}}{\Gamma(\frac{q}{2})}\, dx \\
&= \frac{\Gamma(k+\frac{q}{2})}{\Gamma(\frac{q}{2})} \sum_{j=0}^k \frac{(-1)^j}{j!(k-j)!\Gamma(j+\frac{q}{2})} \int_{\frac{c}{2}}^\infty x^{j+\frac{q}{2}-1} e^{-x}\, dx \\
&= \frac{\Gamma(k+\frac{q}{2})}{k!\Gamma(\frac{q}{2})} \sum_{j=0}^k (-1)^j \binom{k}{j} Q(j+\frac{q}{2}, \frac{c}{2}),
\end{aligned} \tag{2.7.11}$$

where $Q(\alpha,t) = \frac{1}{\Gamma(\alpha)}\int_t^\infty x^{\alpha-1}e^{-x}\,dx, t \le 0$, is the incomplete Gamma function. Using the recursion formula $Q(\alpha,t) = \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)} + Q(\alpha-1,t)$, for $\alpha > 0$ and $t \ge 0$, we have

$$Q(j+\frac{q}{2},t) = t^{\frac{q}{2}}e^{-t}\sum_{i=0}^{j-1}\frac{t^i}{\Gamma(i+\frac{q}{2}+1)} + Q(\frac{q}{2},t) \qquad (2.7.12)$$

Letting $t = \frac{c}{2}$, substituting (2.7.12) into (2.7.11) and using the combinatorial identity

$$\sum_{j=i+1}^{k}(-1)^j\binom{k}{j} = (-1)^{i+1}\binom{k-1}{i} \qquad \text{for } k \ge 1 \text{ and } i \ge 0,$$

we obtain

$$\begin{aligned}
\int_c^\infty L_k^{\frac{q}{2}-1}(\tfrac{y}{2})f(y)\,dy &= \frac{\Gamma(k+\frac{q}{2})}{k!\Gamma(\frac{q}{2})}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\left[t^{\frac{q}{2}}e^{-t}\sum_{i=0}^{j-1}\frac{t^i}{\Gamma(i+\frac{q}{2}+1)} + Q(\frac{q}{2},t)\right] \\
&= \frac{t^{\frac{q}{2}}e^{-t}\Gamma(k+\frac{q}{2})}{k!\Gamma(\frac{q}{2})}\sum_{j=1}^{k}(-1)^j\binom{k}{j}\sum_{i=0}^{j-1}\frac{t^i}{\Gamma(i+\frac{q}{2}+1)} \\
&\quad + \frac{\Gamma(k+\frac{q}{2})}{k!\Gamma(\frac{q}{2})}Q(\frac{q}{2},t)\sum_{j=0}^{k}(-1)^j\binom{k}{j} \\
&= \frac{t^{\frac{q}{2}}e^{-t}\Gamma(k+\frac{q}{2})}{k!\Gamma(\frac{q}{2})}\sum_{i=0}^{k-1}\frac{t^i}{\Gamma(i+\frac{q}{2}+1)}\sum_{j=i+1}^{k}(-1)^j\binom{k}{j} \\
&= \frac{t^{\frac{q}{2}}e^{-t}\Gamma(k+\frac{q}{2})}{k!\Gamma(\frac{q}{2})}\sum_{i=0}^{k-1}\frac{(-1)^{i+1}t^i}{\Gamma(i+\frac{q}{2}+1)}\binom{k-1}{i} \\
&= -\frac{t^{\frac{q}{2}}e^{-t}}{k!\Gamma(\frac{q}{2})}\sum_{i=0}^{k-1}\frac{\Gamma(k+\frac{q}{2})}{\Gamma(i+\frac{q}{2}+1)}\frac{(-t)^i}{i!(k-1-i)!} \\
&= -\frac{t^{\frac{q}{2}}e^{-t}}{k!\Gamma(\frac{q}{2})}L_{k-1}^{\frac{q}{2}}(t)
\end{aligned}$$

■

Let c be a constant such that $\int_c^\infty f(y)\,dy = \alpha$. By Theorem 2.7.5 and Lemma 2.7.6, we have

$$\begin{aligned}
P(Y_1 > c, Y_2 > c) &= \int_c^\infty \int_c^\infty f(y_1)f(y_2)\,dy_1 dy_2 \\
&\quad + \int_c^\infty \int_c^\infty \sum_{k=1}^\infty \frac{\mu_k'}{k!}\left[\frac{\Gamma(\frac{q}{2})k!}{\Gamma(k+\frac{q}{2})}\right]^2 L_k^{\frac{q}{2}-1}(\frac{y_1}{2})L_k^{\frac{q}{2}-1}(\frac{y_2}{2})f(y_1)f(y_2)\,dy_1 dy_2 \\
&= \alpha^2 + \sum_{k=1}^\infty \frac{\mu_k'}{k!}\left[\frac{k!\Gamma(\frac{q}{2})}{\Gamma(k+\frac{q}{2})}\int_c^\infty L_k^{\frac{q}{2}-1}(\frac{y_1}{2})f(y_1)\,dy_1\right]^2 \\
&= \alpha^2 + \sum_{k=1}^\infty \frac{\mu_k'}{k!}\left[\frac{\Gamma(k)}{\Gamma(k+\frac{q}{2})}(\frac{c}{2})^{\frac{q}{2}}e^{-\frac{c}{2}}L_{k-1}^{\frac{q}{2}}(\frac{c}{2})\right]^2
\end{aligned}$$

Therefore, we obtain the following power series expansion for the joint probability of two correlated quadratic forms.

Theorem 2.7.7 *Let c be the $(1-\alpha)$ quantile of the χ_q^2 distribution. We have*

$$P(Y_1 > c, Y_2 > c) = \alpha^2 + \sum_{k=1}^\infty \frac{\mu_k'}{k!}\,a_k^2, \tag{2.7.13}$$

where μ_k' is defined in (2.7.3) *with $M = C_2^T C_1 C_1^T C_2$ and*

$$a_k = \frac{\Gamma(k)}{\Gamma(k+\frac{q}{2})}(\frac{c}{2})^{\frac{q}{2}}e^{-\frac{c}{2}}L_{k-1}^{\frac{q}{2}}(\frac{c}{2}).$$

Furthermore, the series in (2.7.13) *is absolutely convergent for $c \geq 0$.*

Proof: (2.7.13) comes immediately from Theorem 2.7.5 and Lemma 2.7.6. It is trivial to show the absolute convergence of (2.7.6) for c equal to zero. Following is the proof of the absolute convergence of (2.7.13) for $c > 0$. With the same argument in the proof of Theorem 2.7.5, we can show that

$$\frac{\mu_k'}{k!}\,a_k^2 \leq \frac{\tau_3(c,q)}{\Gamma(\frac{q}{2})}\rho^k k^{-\frac{3}{2}}, \quad k \to \infty,$$

where $\tau_3(c,q)$ is a constant depending on c and q. Therefore, the series in (2.7.13) is absolutely convergent for $c > 0$. ∎

For the case $q = 1$, μ'_k has the simple form

$$\mu'_k = \frac{\Gamma(k + \frac{1}{2})}{\Gamma(\frac{1}{2})} \rho^k,$$

where $\rho = \|C_1^T C_2\|^2 = (C_1^T C_2)^2$.

2.8 The Simulation Study

Three validity measures were discussed in the previous sections. Hooper (1989) used $\text{cv}(\phi)$ to investigate validity in small experiments. $(\mu_{\text{val}}, \sigma_{\text{val}})$ was used in Hooper (1993) to assist in the selection of the tuning constant λ for NORD. Here, we provide some simulation results to show the use of the Beta-related measure to evaluate the validity of the two-sided t-test in completely randomized designs and randomized block designs. In the empirical study the Beta-related measure is compared with $\text{cv}(\phi)$.

2.8.1 Validity of the t-test in completely randomized designs

We first consider the completely randomized design in experiments with zero, one, or two covariates. The analysis of covariance model is adopted for the unit effects if there are one or two covariates associated with units. Only two treatments are compared in the experiments. There are $n = 2k$ units available and the general additive linear model (1.2.5) is assumed for analysis. The two-sided t-test is used to test the null hypothesis of no difference between the two treatment effects. We want to investigate how the validity of the t-test depends on the experiment size and how much the addition of covariates reduces validity. The data sets used in Bellhouse (1986) are employed for values of covariates for $n = 10$ to 20. Several truncated data sets are employed for $n = 6$ and 8.

Table 2.2: Approximate 95% confidence intervals for cv(ϕ) for the t-test in a Completely Randomized Design for the data sets from Bellhouse (1986)

Size	Lvl	No	One covariate		Two covariates	
n	α	covariate	Data 1	Data 2	Data 1	Data 2
6	.10	.553, .629	.633, .701	.593, .666	.720, .788	.614, .683
	.05	.994, 1.004	1.059, 1.170	1.017, 1.100	1.143, 1.283	1.006, 1.076
	.01	2.722, 3.437	2.320, 3.037	2.572, 3.291	2.436, 3.598	2.644, 3.369
8	.10	.248, .289	.277, .314	.295, .337	.328, .376	.306, .344
	.05	.513, .575	.476, .546	.504, .567	.547, .621	.487, .549
	.01	1.315, 1.452	1.315, 1.410	1.428, 1.658	1.452, 1.686	1.347, 1.525
10	.10	.144, .180	.146, .176	.148, .175	.150, .172	.153, .177
	.05	.247, .296	.267, .307	.250, .294	.262, .303	.242, .277
	.01	.750, .839	.730, .831	.722, .820	.697, .794	.608, .691
12	.10	.072, .100	.077, .106	.079, .115	.086, .112	.084, .107
	.05	.145, .185	.143, .181	.162, .204	.163, .206	.163, .197
	.01	.448, .505	.443, .506	.449, .505	.494, .555	.498, .558
14	.10	.038, .058	.035, .054	.040, .052	.036, .061	.047, .060
	.05	.089, .125	.085, .112	.092, .118	.085, .123	.100, .126
	.01	.335, .386	.300, .351	.302, .350	.301, .343	.339, .389
16	.10	.016, .032	.014, .030	.022, .037	.030, .049	.024, .038
	.05	.048, .070	.054, .078	.060, .088	.068, .100	.064, .089
	.01	.205, .247	.215, .262	.227, .275	.236, .284	.233, .280
18	.10	.014, .030	.013, .036	.016, .030	.011, .026	.013, .029
	.05	.049, .075	.048, .076	.049, .068	.040, .058	.048, .071
	.01	.191, .236	.194, .239	.178, .221	.159, .197	.178, .222
20	.10	.002, .024	.009, .025	.000, .021	.000, .019	.002, .024
	.05	.021, .042	.020, .040	.022, .054	.038, .066	.041, .064
	.01	.117, .159	.121, .160	.146, .186	.170, .210	.182, .227

Table 2.3: Values of $\text{cv}(T^2)$ for the CRD with no covariate

No. of Units n	$\text{cv}(T^2)$	$E(T) = \frac{1}{n-1}$	$E(T^2) = \frac{3}{n^2-1}$
4	.29096	.3333	.20
6	.15474	.20	.0857
8	.10578	.1428	.0476
10	.07907	.1111	.0303
12	.06223	.0909	.0210
14	.05073	.0769	.0154
16	.04242	.0666	.0118
18	.03618	.0588	.0093
20	.03134	.0526	.0075
30	.01787	.0345	.0033
50	.00865	.0204	.0012

Table 2.4: Values of $[\text{cv}(T), \text{cv}(T^2)]$ for the CRD with one or two covariates

No. of	One covariate		Two covariates	
Units n	Data 1	Data 2	Data 1	Data 2
6	.02242, .15043	.05010, .15512	.07745, .17401	.10343, .16431
8	.03475, .09228	.01727, .10731	.05699, .12854	.02748, .08709
10	.01409, .07451	.01918, .07855	.01960, .06836	.02496, .06594
12	.00653, .05754	.01395, .06203	.00855, .06133	.00956, .06082
14	.00366, .04711	.01155, .05226	.00000, .04138	.01260, .05202
16	.00000, .03863	.00560, .04067	.01163, .04426	.00595, .03784
18	.00000, .03334	.00000, .03402	.00455, .03412	.00813, .03484
20	.00684, .03259	.00000, .02724	.00610, .03189	.00000, .02933

Table 2.2 displays approximate 95% confidence intervals of $\mathrm{cv}(\phi)$ for the t-test in the CRD. The method in Hooper (1989) is used here to obtain an approximate 95% confidence interval of $\mathrm{cv}(\phi)$ based on 500 replicates of W and 5000 replicates of G. $\mathrm{cv}(\phi)$ decreases rapidly as the experiment size increases. The addition of one or two covariates in the experiment does not yield a substantial increase in $\mathrm{cv}(\phi)$.

In the following, we use the Beta-related validity measure $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$. For CRDs with no covariate the results in Section 2.4 yield $\mathrm{cv}(T) = 0$ and, since $a = \frac{1}{2k}$,

$$\mathrm{cv}(T^2) = \frac{4\sqrt{2(n-2)}}{\sqrt{3n(n-3)(n+3)(n+5)}}.$$

By (2.3.3) and (2.3.4), we obtain $E[T(G,W)] = (n-1)^{-1}$ and $E[T^2(G,W)] = 3(n^2-1)^{-1}$. The values of $\mathrm{cv}(T^2)$, $E[T(G,W)]$ and $E[T^2(G,W)]$ are tabulated for some values of $n = 4$ to 50 in Table 2.3.

For the CRDs with one or two covariates, we use the generalized U-statistics to estimate the Beta-related validity measure $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$. In order to compare these estimates here with those confidence intervals of $\mathrm{cv}(\phi)$, we also use here 500 replicates of W and 5000 replicates of G. The method is similar to that for $\mathrm{cv}(\phi)$ in Section 2.2. The estimated values of $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ are listed in Table 2.4.

From Table 2.4 and Table 2.3, we also found that $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ decreases rapidly as the experiment size increases and the addition of one or two covariates in the experiment does not yield much increase in $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$.

2.8.2 Validity of the t-test in randomized blocks designs

In this subsection, we consider the randomized blocks design in experiments for comparison of two treatments. The two-sided t-test is used to test the null hypothesis of no difference between the two treatment effects. Table 2.5 lists the values of validity measure $\mathrm{cv}(\phi)$ for the t-test in randomized blocks designs. We see that $\mathrm{cv}(\phi)$ increases as the number of blocks increases with n fixed. The same can be seen in Table 2.6

Table 2.5: Approximate 95% confidence intervals for $\mathrm{cv}(\phi)$ for the t-test in randomized blocks designs

No. of Units n	No. of Blocks	Level α		
		.10	.05	.01
8	0	.248, .289	.513, .575	1.315, 1.452
	2	.4187, .4743	.7682, .8576	1.7528, 2.0734
	4	.6479, .7322	1.2197, 1.2654	2.7878, 3.7381
12	0	.072, .100	.145, .185	.448, .505
	2	.1244, .1517	.2029, .2483	.6334, .7078
	3	.1483, .1857	.2766, .3258	.8918, .9891
	6	.3067, .3670	.5152, .5833	1.4123, 1.6086
16	0	.016, .032	.048, .070	.205, .247
	2	.0315, .0495	.0782, .1157	.2762, .3303
	4	.0685, .0964	.1455, .1893	.4461, .5050
	8	.1384, .1778	.2701, .3174	.8080, .8984
20	0	.002, .024	.021, .042	.117, .159
	2	.0115, .0256	.0358, .0468	.1509, .1914
	5	.0264, .0418	.0724, .0955	.2945, .3413
	10	.0747, .1062	.1547, .2044	.4845, .5428

when we use $\mathrm{cv}(T^2)$ to evaluate the validity. $\mathrm{cv}(T^2)$ in Table 2.6 is estimated by the generalized U-statistics based on 500 replicates of W and 5000 replicates of G. Note that $\mathrm{cv}(T) = 0$ in randomized blocks designs.

Table 2.6: Values of cv(T^2) for randomized blocks designs

No. of	No. of Blocks							
Units n	0	2	3	4	5	6	8	10
8	.10578	.14945		.15743				
12	.06223	.07518	.09468			.11053		
16	.04242	.04542		.06816			.08313	
20	.03134	.03763			.04990			.06568

2.8.3 A single valued validity measure and its empirical relationship to cv(ϕ)

In the next chapter, we will discuss several methods of randomization with respect to validity and efficiency and will empirically investigate whether these randomized designs make the same trade-off between validity and efficiency. It is helpful to use a single valued validity measure in this investigation. Under the general linear model (1.2.5), Hooper (1993) described validity in terms of whether the distribution of the random variable $\|C_1^T G_1 G_2^T C_2\|^2$ is concentrated close to 0 and defined validity measure as the mean and the standard deviation of val(G_1, G_2); see Section 2.2 for details. The second moment of val(G_1, G_2) can also be used to measure the closeness of $\|C_1^T G_1 G_2^T C_2\|^2$ to 0. We define

$$\mu 2_{\text{val}} = \{E\text{val}^2(G_1, G_2)\}^{\frac{1}{2}} = \sqrt{\mu_{\text{val}}^2 + \sigma_{\text{val}}^2}. \tag{2.8.1}$$

$\mu 2_{\text{val}}$ can be estimated by

$$\widehat{\mu 2}_{\text{val}} = \sqrt{\hat{\mu}_{\text{val}}^2 + \hat{\sigma}_{\text{val}}^2}, \tag{2.8.2}$$

where $(\hat{\mu}_{\text{val}}, \hat{\sigma}_{\text{val}})$ is defined by (2.2.4) and (2.2.5) in Section 2.2. An unbiased estimate of $\mu 2_{\text{val}}^2$ is the U-statistic:

$$\frac{2}{m(m-1)} \sum_{i<j} \text{val}^2(G_i, G_j). \tag{2.8.3}$$

It can be shown that $\widehat{\mu 2}^2_{\text{val}}$ and (2.8.3) are asympotically equivalent as m tends to ∞. The standard error of $\widehat{\mu 2}_{\text{val}}$ can thus be approximately determined by that of (2.8.3).

The measure cv(ϕ) provides a better interpretation than the other validity measures but is more expensive to estimate than $(\mu_{\text{val}}, \sigma_{\text{val}})$ for TCRD and NORD, since the cost of generating a G in these designs is very high. In the following, we consider the empirical study in Section 3 of Chapter 3. For each data set, we generate a sequence of TCRDs with a decreasing sequence of κ. For each TCRD in the sequence, we calculate $\widehat{\mu 2}_{\text{val}}$ by (2.8.2) and estimate three cv(ϕ) for $\alpha = .1, .05$ and $.01$ by the generalized U-statistics in Section 2.2. Thus, we obtain a sequence of $\widehat{\mu 2}_{\text{val}}$ and three sequences of estimated cv(ϕ) for $\alpha = .1, .05$ and $.01$ respectively. For each data set, we present superimposed plots of estimated cv(ϕ) against $\widehat{\mu 2}_{\text{val}}$ for $\alpha = .1(\cdot), .05(\triangle)$ and $.01(+)$. Figures 2.1 - 2.6 give the superimposed plots for 24 data sets from Bellhouse (1986). All the Figures show a clear linear relationship between the estimated cv(ϕ) and $\widehat{\mu 2}_{\text{val}}$. Note that the scale on the y-axis decreases as the number n of units increases, which causes the plots for larger n to appear more random.

We also drew the plots of the estimated cv(ϕ) against $\hat{\mu}_{\text{val}}$ and $\hat{\sigma}_{\text{val}}$ and the linear relationship was again found. The same result was obtained for the NORD. This suggests that $\mu_{\text{val}}, \sigma_{\text{val}}$, or $\mu 2_{\text{val}}$ can be used as a surrogate for cv(ϕ) in determining an appropriate value for the truncation constant κ.

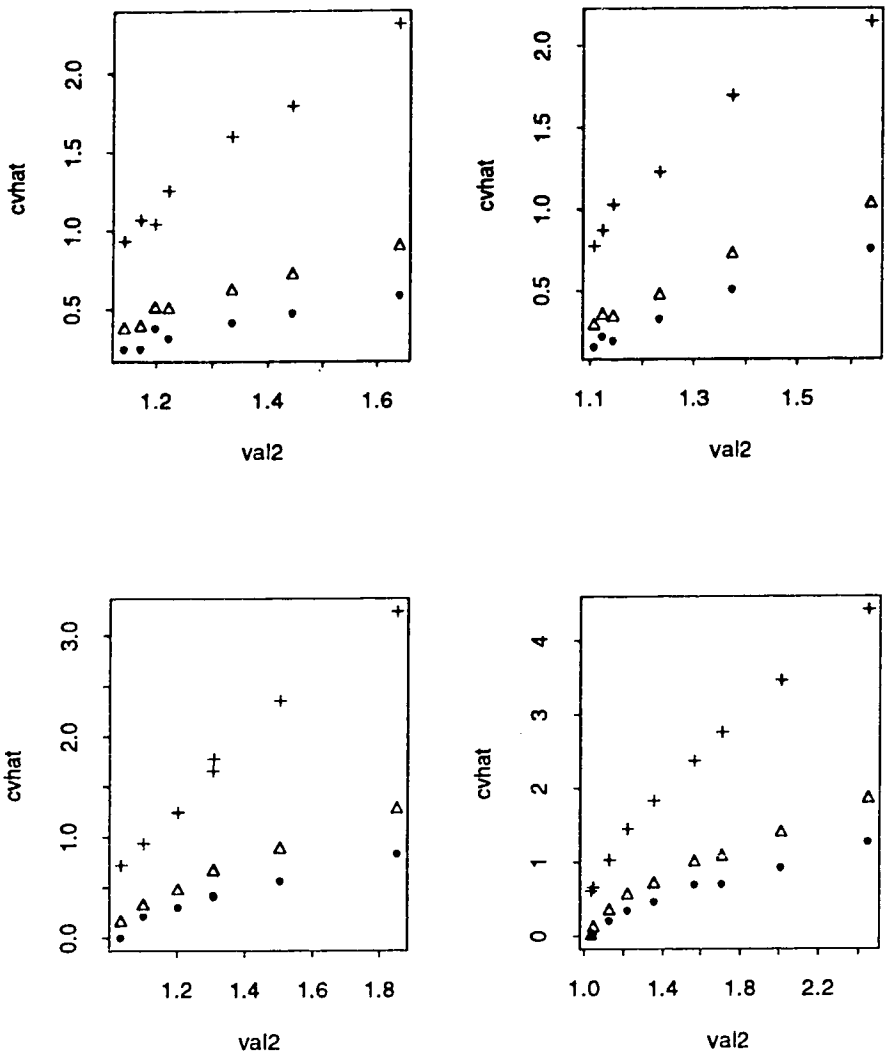


Figure 2.1: Plots of estimated cv(ϕ) vs. $\widehat{\mu 2}_{\text{val}}$ for $\alpha = .1(\cdot), .05(\Delta)$ and $.01(+)$ in analysis of covariance model with $n = 10$, one covariate for the upper two plots and two covariates for the lower two plots

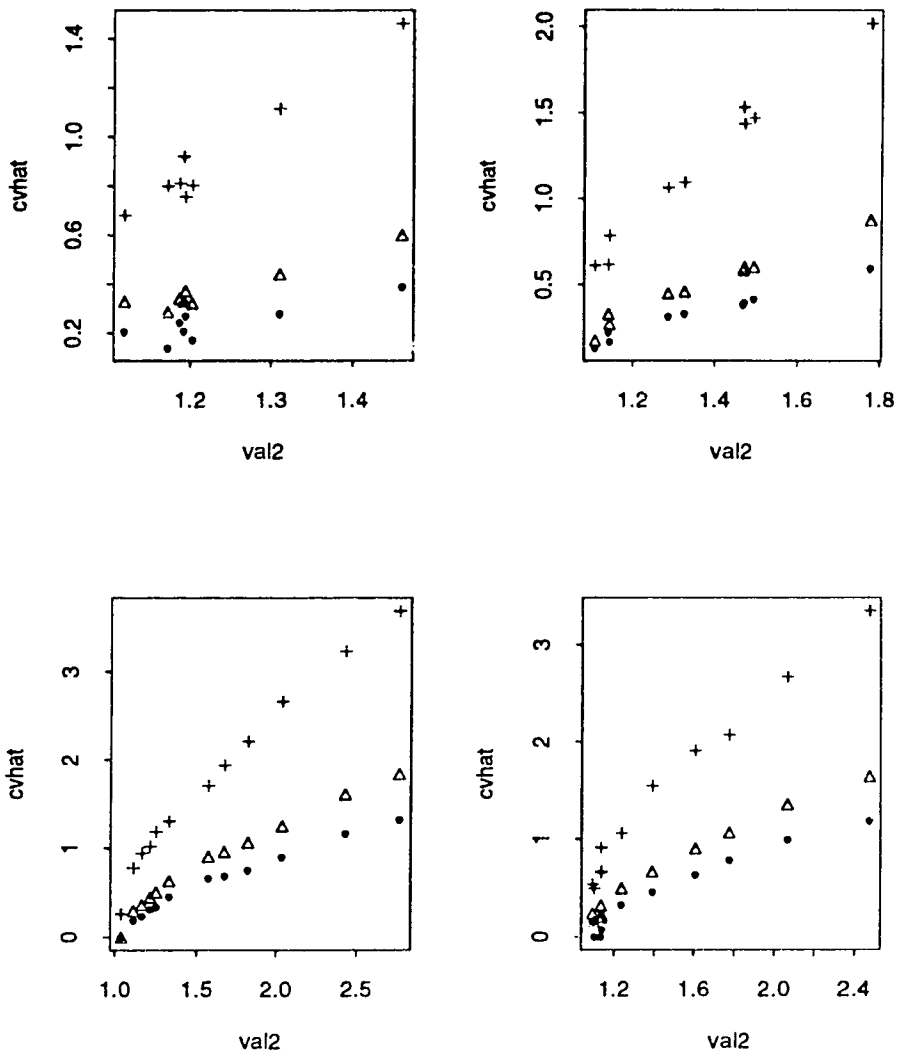


Figure 2.2: Plots of estimated cv(ϕ) vs. $\widehat{\mu 2}_{\text{val}}$ for $\alpha = .1(\cdot), .05(\triangle)$ and $.01(+)$ in analysis of covariance model with $n = 12$, one covariate for the upper two plots and two covariates for the lower two plots

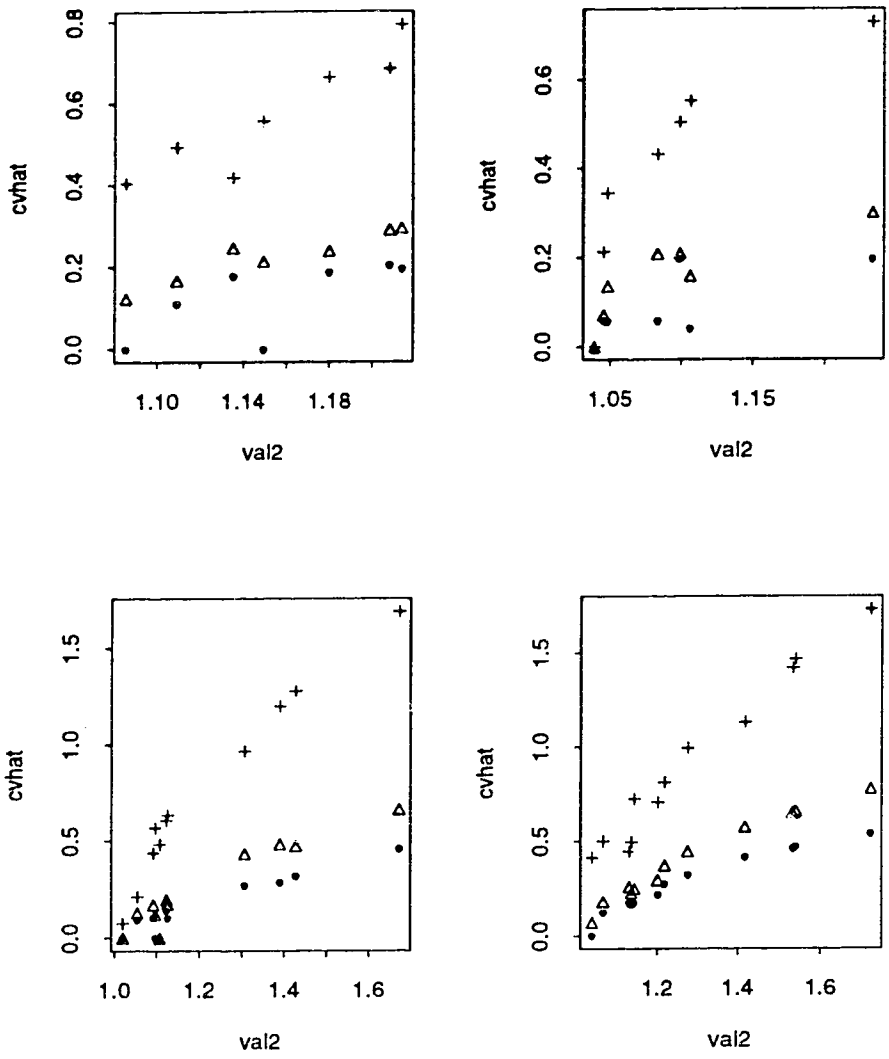


Figure 2.3: Plots of estimated cv(ϕ) vs. $\widehat{\mu 2}_{\text{val}}$ for $\alpha = .1(\cdot), .05(\triangle)$ and $.01(+)$ in analysis of covariance model with $n = 14$, one covariate for the upper two plots and two covariates for the lower two plots

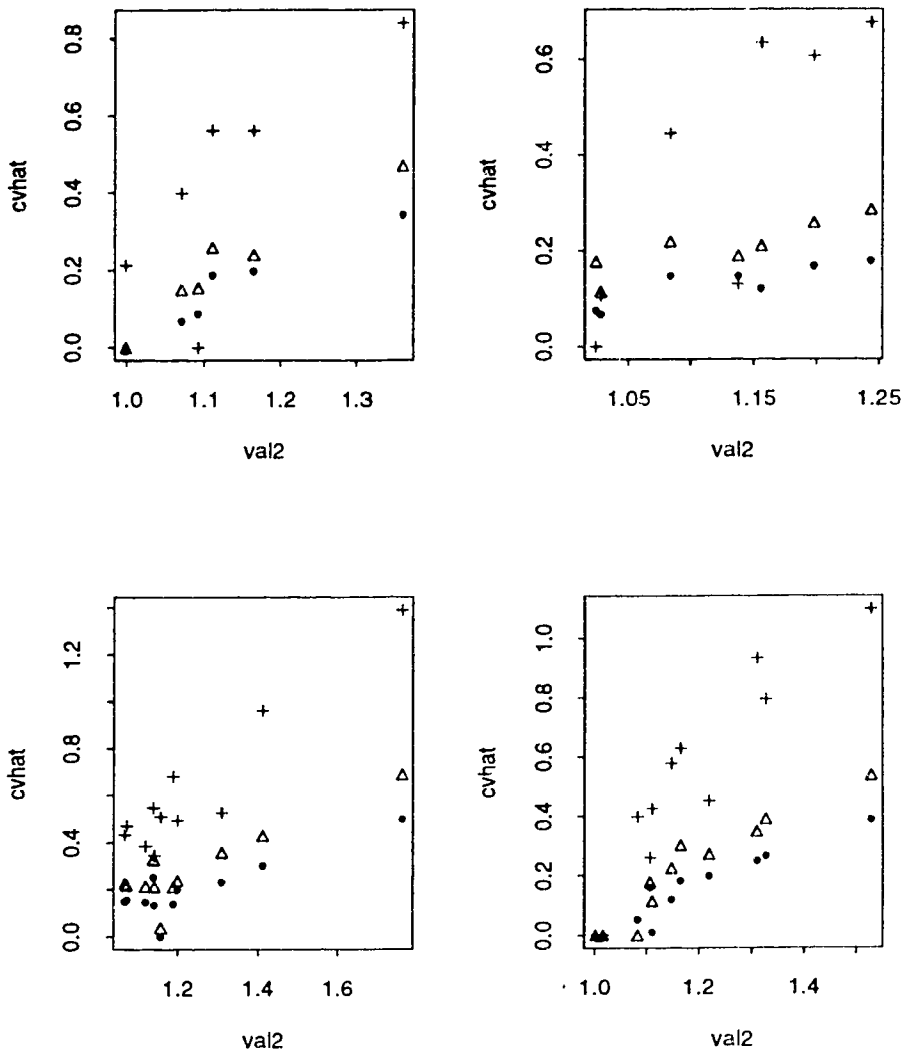


Figure 2.4: Plots of estimated cv(ϕ) vs. $\widehat{\mu 2}_{\text{val}}$ for $\alpha = .1(\cdot), .05(\triangle)$ and $.01(+)$ in analysis of covariance model with $n = 16$, one covariate for the upper two plots and two covariates for the lower two plots

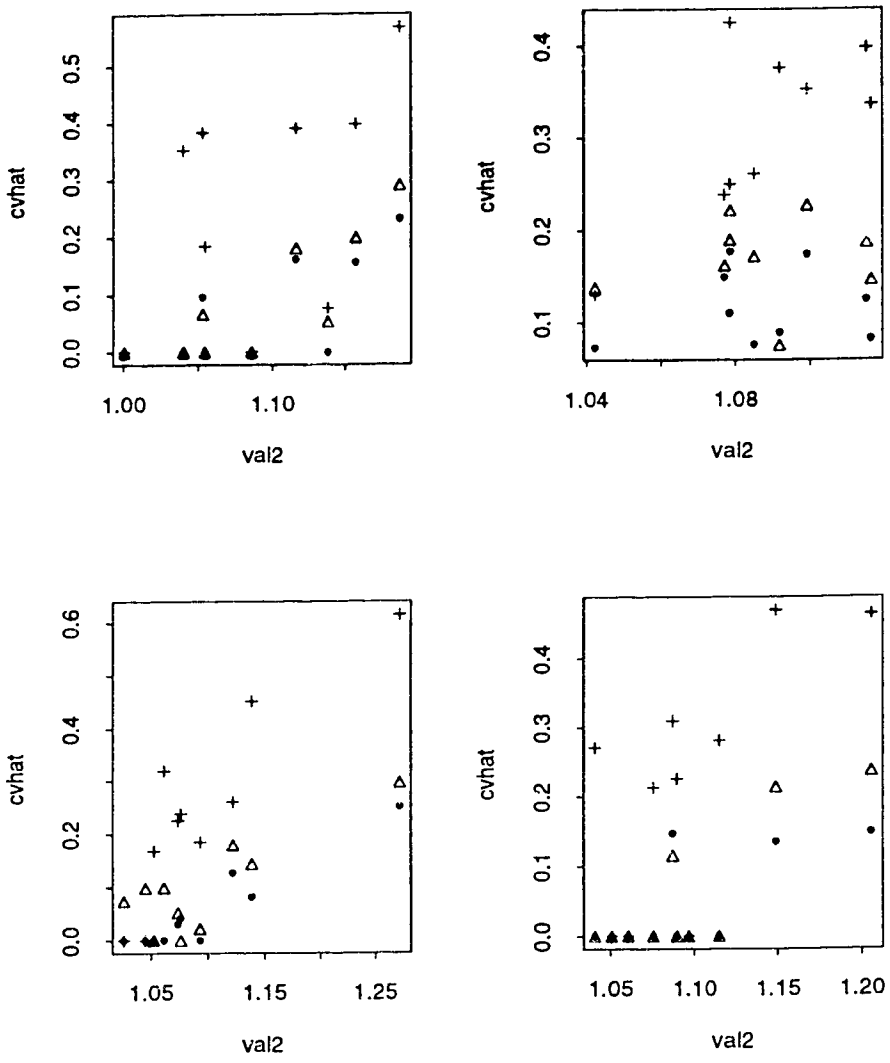


Figure 2.5: Plots of estimated cv(ϕ) vs. $\widehat{\mu 2}_{\mathbf{val}}$ for $\alpha = .1(\cdot), .05(\triangle)$ and $.01(+)$ in analysis of covariance model with $n = 18$, one covariate for the upper two plots and two covariates for the lower two plots

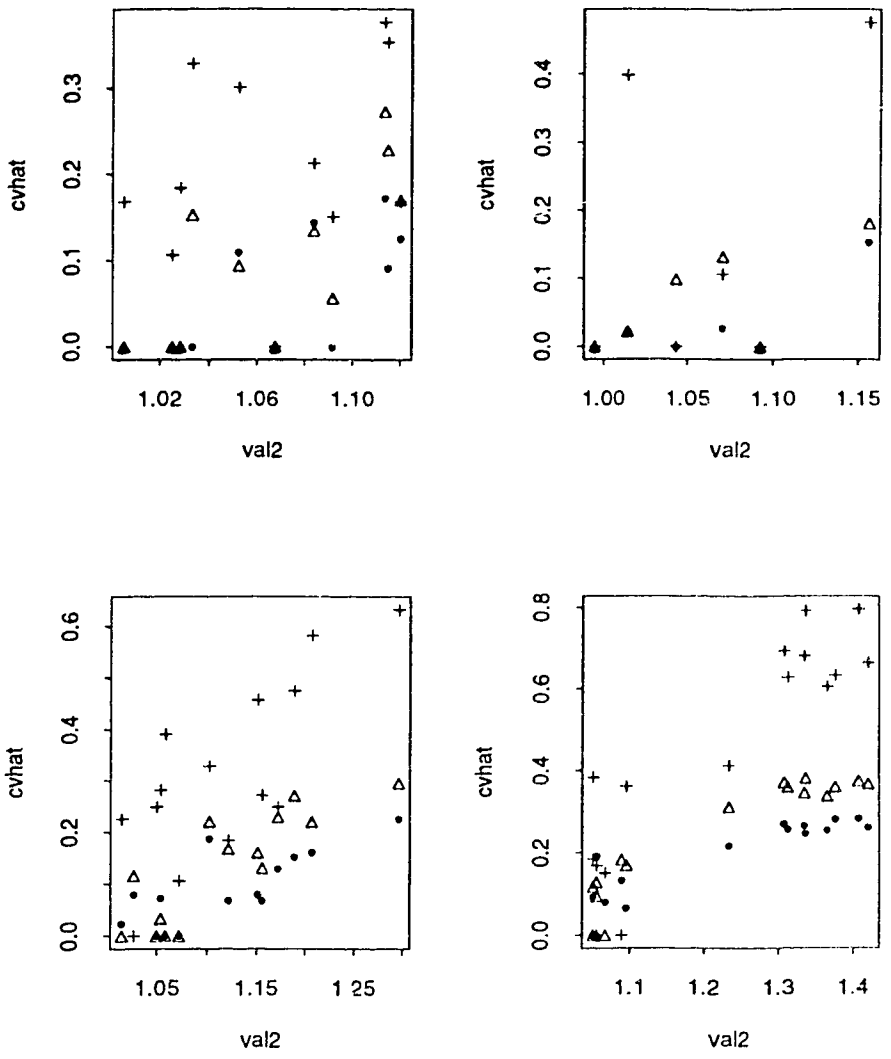


Figure 2.6: Plots of estimated cv(ϕ) vs. $\widehat{\mu 2}_{\text{val}}$ for $\alpha = .1(\cdot), .05(\triangle)$ and $.01(+)$ in analysis of covariance model with $n = 20$, one covariate for the upper two plots and two covariates for the lower two plots

Chapter 3

Methods of Randomization

3.1 Introduction

Randomized designs can be evaluated with regard to efficiency, validity and relevance. These criteria were discussed in Section 1.3. Under simple models for unit effects, classical randomization strategies usually achieve full efficiency and relevance as well as a high degree of validity for the usual analysis of variance. For example, the completely randomized design (CRD) is used when the unit effects are homogeneous; the randomized blocks design or Latin square design is preferred if the unit effects are blocked in one direction or in two directions (of course, we assume that the two block factors are additive in Latin square design). Hoeffding (1952) and Robinson (1972) showed the asymptotic validity for the classical randomized designs. Under asymmetric models of the unit effects, such as analysis of covariance models and spatial models, the CRD provides validity, however efficiency can be substantially reduced due to non-orthogonality. Thus it is desirable to develop methods of randomization with high efficiency and high validity under asymmetric models for the unit effects.

Cox (1982) and Bellhouse (1986) considered the efficient and valid use of covariate information in design of experiments and recommended truncating classical methods of randomization to ensure high levels of efficiency while maintaining validity. We

refer to the truncated classical methods of randomization as Truncated Classical Randomized Designs (TCRDs). Under a general linear model for additive treatment and unit effects, Hooper (1993) proposed the Nearly Orthogonal Randomized Designs (NORDs) to achieve acceptable levels of efficiency and validity.

This chapter is organized as follows. The TCRD is reviewed in Section 3.2. This design requires the selection of a truncation constant to determine a trade-off between efficiency and validity. In section 3.3, a sequential procedure is introduced to assist in this selection and is illustrated by analysis of covariance models. We review the NORD in section 3.4. A truncated NORD is proposed in Section 3.5 to improve NORD with respect to efficiency and selection of tuning constant. The truncated NORD is invariant under reparameterization of treatment-effects model and can be generated by the sequential procedure. In Section 3.6, we derive some formulae used in the computation of the truncated NORD. In Section 3.7, an empirical study is presented for comparison of truncated NORDs with TCRDs and NORDs with regard to statistical and computational properties.

Our discussions are all based on model (1.2.5) and the assumptions in Section 1.2 are adopted in this chapter.

3.2 Truncated Classical Randomized Designs

In comparative experiments with covariates (concomitant variables), the treatment effects are estimated by analysis of covariance. If classical methods of randomization are used, there may be some inflation of variance. In order to limit the inflation of variance, Cox (1982) suggested restricting the randomization to ensure a measure of efficiency is small.

To illustrate his idea, Cox (1982) considered an experiment with $n = 2k$ experimental units to compare two treatment effects τ_1 and τ_2, each treatment being assigned to k units. Let a $p \times 1$ vector z of covariates be available on each unit.

Assume that linear methods of analysis based on the method of least squares are used. If $\hat{\tau}_1 - \hat{\tau}_2$ is the adjusted difference between treatments after allowing for linear regression on z, we have under the usual second-order assumptions about error that

$$\text{var}(\hat{\tau}_1 - \hat{\tau}_2) = \sigma^2 \left\{ \frac{2}{k} + (\bar{z}_1 - \bar{z}_2)^T S_z^{-1} (\bar{z}_1 - \bar{z}_2) \right\}. \tag{3.2.1}$$

Here σ^2 is the residual variance per observation, $\bar{z}_1$ and $\bar{z}_2$ are the vector means of the concomitant variables for τ_1 and τ_2 in the design chosen and S_z is the matrix of sums of squares and products of z within treatments, eliminating block differences where appropriate. More generally, if there are t treatments, each assigned to k experimental units,

$$\text{ave var}(\hat{\tau}_i - \hat{\tau}_j) = \sigma^2 \left\{ \frac{2}{k} + \frac{2}{k(t-1)} \text{tr}(B_z S_z^{-1}) \right\}, \tag{3.2.2}$$

where B_z is the matrix of sums of squares and products between treatments, i.e., of k times the sum of squares and products of deviations of means.

The inflation of variance is $\frac{k}{2}(\bar{z}_1 - \bar{z}_2)^T S_z^{-1} (\bar{z}_1 - \bar{z}_2)$ in (3.2.1). Under classical randomization,

$$W_p = \frac{k}{2} d_w (\bar{z}_1 - \bar{z}_2)^T S_z^{-1} (\bar{z}_1 - \bar{z}_2) \tag{3.2.3}$$

has asymptotically a chi-squared distribution with p degrees of freedom, where d_w is the degrees of freedom within treatments, i.e., the d.f. for S_z ($d_w = t(k-1)$ when there is no blocking).

Cox (1982) discussed the circumstances under which the chance of appreciable inflation might be kept at a small number. He suggested that randomization be repeated until the constraint $\frac{W_p}{d_w} \leq b$ is achieved for some suitable b, e.g. $b \leq \frac{1}{20}$, so that negligible inflation of variance occurs.

Bellhouse (1986) examined the validity aspect of TCRDs by an empirical study. He selected 72 sets of data with one or two covariates in the experiment for comparing two treatments, where the experiment size $tk = 2k$ was taken to be $10, 12, 14, 16, 18$ and 20. The truncated randomization with $b = \frac{1}{20}$ was considered in that simulation.

He investigated the fit of the distribution of W_p to its asymptotically chi-squared distribution under classical randomization and compared the randomization significance level with its nominal level. He concluded that approximate balance of covariates and approximate validity of normal theory inference can be achieved through truncated randomization and that larger sample sizes are needed as the number of covariates increases.

The idea of Cox (1982) can be applied to general linear models for treatment and unit effects. Under model (1.2.5), the normal theory inference can be implemented by conditioning on $\mathcal{R}(GB)$. The least squares estimator of β has covariance matrix $\sigma^2(A^TCC^TA)^{-1}$. In Section 3.1, we concluded that $(A^TCC^TA)^{-1} \geq I_q$. For efficiency, we should choose designs G to make the covariance matrix small in some sense. Let $\text{eff}(G) = h\{(A^TCC^TA)^{-1}\}$ be a measure of efficiency for a particular design G, where $h(M)$ is a real-valued function of symmetric matrix M such as the trace, determinant, or largest eigenvalue. A design G is fully efficient if $A^TC = I_n$. For optimal efficiency one could adopt the uniform distribution of G on the subset $\{G \in \mathcal{P}_n : \text{eff}(G) = \min_{J \in \mathcal{P}_n} \text{eff}(J)\}$. This subset may be too small to provide much validity. For validity, we recommend uniform randomization over a subset $\{G \in \mathcal{P}_n : \text{eff}(G) \leq \kappa\}$, where κ is a constant. We will continue to call such a randomization strategy a truncated classical randomized design (TCRD).

The algorithm for generating a TCRD is as follows. Generate a design G from the uniform distribution on $\mathcal{P}_n$. Choose G if $\text{eff}(G) \leq \kappa$, otherwise repeat the preceding step.

Increasing the truncation constant κ reduces efficiency but typically improves validity, since the size of subset $\{G : \text{eff}(G) \leq \kappa\}$ increases with κ. We would like to determine the truncation constant κ to achieve acceptable levels of efficiency and validity for a given model. The measures of validity in Chapter 1 and the measures of efficiency given above can be used to select the truncation constant κ.

With the aid of the measures of validity and efficiency, we propose a sequential procedure in the next section to select an appropriate κ which defines a TCRD with a reasonable trade-off between efficiency and validity.

3.3 A Sequential Procedure for TCRD

The validity and efficiency of a TCRD usually must be estimated by a sample of G using simulation. For a given model, the truncation constant κ is chosen so that the corresponding TCRD achieves acceptable levels of efficiency and validity. Under the general linear model (1.2.5), we introduce a sequential procedure for generating TCRDs to assist in the selection of κ.

This sequential procedure starts with generating m G's from the CRD and calculating the order statistics of an efficiency measure based on the sample. Use the $(h+1)$th order statistic of the efficiency measure (say, $\frac{h}{m} = \frac{1}{2}$ or $\frac{2}{3}$) as the truncation constant κ of the next TCRD. In the second step, we want to generate another m G's from the TCRD to estimate its efficiency and validity. Since there are h G's generated from the CRD whose efficiency measures are less than the chosen truncation constant, we can keep the most efficient h G's. Thus we only need to generate an additional $(m-h)$ G's from the TCRD. Given the value of κ determined in the previous step, the combined m G's approximate a random sample from the TCRD with this κ and can be used to obtain the $(h+1)$th order statistic of the efficiency measure for the truncation constant κ of the next TCRD. We repeat the above procedure until some stopping rule is satisfied.

Let m be the number of G's generated for each TCRD. For convenience, we let m be even and let $1 < h < m$ be an integer. The sequential procedure is described as follows:

Step 1. Choose κ_0 large enough so that $\mathcal{P}_n = \{G : \text{eff}(G) \leq \kappa_0\}$. Generate m independent matrices G distributed uniformly on $\mathcal{P}_n$ and compute the corresponding

efficiency measure value for each G. Let $e_{0(1)} \leq e_{0(2)} \leq \cdots \leq e_{0(m)}$ denote the ordered efficiencies and $G_{01}, G_{02}, \ldots, G_{0m}$ be the corresponding permutation matrices.

Step 2. Let κ_1 be the $(h+1)$th order statistic of the efficiency measure in step 1, i.e., let $\kappa_1 = e_{0(h+1)}$. Keep $G_{01}, G_{02}, \ldots, G_{0h}$, the first h most efficient designs in step 1. Generate $m - h$ additional permutation matrices, each distributed uniformly on $\{G : \text{eff}(G) \leq \kappa_1\}$, and compute the corresponding efficiency measure value for each design. For the combined set of m G's, let $e_{1(1)} \leq e_{1(2)} \leq \cdots \leq e_{1(m)}$ denote the ordered efficiencies and let $G_{11}, G_{12}, \ldots, G_{1m}$ be the corresponding permutation matrices.

Step 3. Repeat step 2 until $|\kappa_i - \kappa_{i-1}| \leq \varepsilon$ for a given small constant ε.

This sequential procedure produces a sequence of TCRDs starting with the CRD. For each TCRD in the sequence, we can compute the values of an efficiency measure and a validity measure. Thus we obtain a sequence of efficiencies and a sequence of validities and obtain a plot of validity versus efficiency. Examining the plot, we can select the TCRD with an appropriate κ to achieve acceptable levels of efficiency and validity.

The following result is given in Reiss (1989). Let $X_1, \ldots, X_m$ be m random vectors from a continuous distribution $F(\cdot)$ and let ψ be a real-valued function. The ψ-order statistics $X_{(1)} \leq_\psi \cdots \leq_\psi X_{(m)}$ are defined according to the ascending order of the ψ function values of $X_1, \ldots, X_m$, i.e., $\psi(X_{(1)}) \leq \cdots \leq \psi(X_{(m)})$. Given $\psi(X_{(h+1)}) = c$, $X_{(1)} \leq_\psi \cdots \leq_\psi X_{(h)}$ are the ψ-order statistics of h random vectors from the distribution $F(\cdot)$ truncated outside of $\{x : \psi(x) < c\}$. In the sequential procedure, the distribution of the random permutation matrix G is discrete. If n is not small, then the distribution of $\text{eff}(G)$ approximates a continuous distribution. Thus, we keep in Step 2 the most efficient h permutation matrices from Step 1. Given $e_{0(h+1)}$, the h permutation matrices in the original order of their outcome are approximately each distributed uniformly on $\{G : \text{eff}(G) \leq e_{0(h+1)}\}$. Hence, given the value of κ_i

determined in Step 1, the m permutation matrices in Step 2 approximate a random sample on $\{G : \text{eff}(G) \leq \kappa_1\}$.

Generally, how large $\frac{h}{m}$ is chosen depends on many aspects of the model. If the experimental size n is small, the set $\mathcal{P}_n$ contains a few elements. Thus $\frac{h}{m}$ should be chosen to be relatively large so that enough values of κ are obtained to produce a clear curve in the plot of the estimated validity against the estimated efficiency. When n is large, however, a small value of $\frac{h}{m}$ may be chosen to save computational time. In our empirical study on analysis of covariance models, the size n is from 10 to 20 and an appropriate value of $\frac{h}{m}$ is taken to be 2/3.

For small values of κ, the validity measure of the TCRD may become too large for the design to be of interest. The stopping rule in Step 3 can be modified to take validity into account. Let val_0 be the value of the validity measure of the CRD and val_i be that of the TCRDs in the sequence. An alternative stopping rule can be that if

$$\frac{\text{val}_i - \text{val}_0}{\text{val}_0} \geq c,$$

then we stop iteration, where c is a large positive constant. We can also combine this alternative stopping rule with that in Step 3 for use in practice.

The following proposition asserts the convergence in probability of the sequence $\{\kappa_i\}$ to the optimum efficiency.

Proposition 3.3.1 *Let π_0 be the minimum value of the efficiency measure, i.e.,*

$$\pi_0 = \min_{G \in \mathcal{P}_n} \text{eff}(G).$$

Then $\kappa_i \to \pi_0$ in probability as $i \to \infty$.

The proof of Proposition 3.3.1 is based on the following Lemma about sequential sampling from a truncated distribution:

Lemma 3.3.2 *Let $X_{11}, X_{12}, \ldots, X_{1m}$ be a random sample from a distribution $F(\cdot)$ and $X_{1(h+1)}$ be the $(h+1)$th order statistic of the sample, where m and h are fixed integers with $1 \leq h < m$. Given $X_{1(h+1)}$, we take another sample $X_{21}, X_{22}, \ldots, X_{2m}$ from the distribution $F(\cdot)$ truncated on the right by $X_{1(h+1)}$ and let $X_{2(h+1)}$ be the $(h+1)$th order statistic of the sample. According to the above sequential sampling procedure, we obtain a random process $\{X_{k(h+1)}\}$. Let $x_0 = \inf\{x : F(x) > 0\}$. Then we have $X_{k(h+1)} \to x_0$ in probability as $k \to \infty$.*

Proof: The random process $\{X_{k(h+1)}\}$ can be defined another way. Suppose we have a double sequence $\{Y_{ij}\}(i = 1, 2, \ldots, j = 1, 2, \ldots)$ which are independently distributed with cdf $F(\cdot)$. Let $X_{11}, X_{12}, \ldots, X_{1m}$ be the first random variables from the first row sequence $\{Y_{1j}\}(j = 1, 2, \ldots)$. Thus $X_{1(h+1)} = Y_{1(h+1)}$ where $Y_{1(h+1)}$ is the $(h+1)$th order statistic of $\{Y_{11}, Y_{12}, \ldots, Y_{1m}\}$. Let $X_{21}, X_{22}, \ldots, X_{2m}$ be the first m random variables from the second row sequence $\{Y_{2j}\}(j = 1, 2, \ldots)$ such that $Y_{2j} \leq X_{1(h+1)}$, and take the corresponding hth order statistic $X_{2(h+1)}$. Continuing the procedure, we obtain the random process $\{X_{k(h+1)}\}$. Let $Y_{k(m)}$ be the largest order statistic of the first m random variables of the kth row sequence $\{Y_{kj}\}(j = 1, 2, \ldots)$. It is easy to see $X_{k(h+1)} \leq Y_{k(m)}$ and then

$$X_{k(h+1)} = \min\{X_{1(h+1)}, X_{2(h+1)}, \ldots, X_{k(h+1)}\} \leq \min\{Y_{1(m)}, Y_{2(m)}, \ldots, Y_{k(m)}\}.$$

For any x such that $F(x) > 0$, we have that as $k \to \infty$,

$$\begin{aligned} P\left\{\min\{Y_{1(m)}, Y_{2(m)}, \ldots, Y_{k(m)}\} > x\right\} &= P\{Y_{1(m)} > x, Y_{2(m)} > x, \ldots, Y_{k(m)} > x\} \\ &= P^k\{Y_{1(m)} > x\} \\ &= [1 - F^m(x)]^k \to 0. \end{aligned}$$

Since $\min\{Y_{1(m)}, Y_{2(m)}, \ldots, Y_{k(m)}\} \geq x_0$, it follows that for all $\epsilon > 0$,

$$\begin{aligned} & P\left\{|\min\{Y_{1(m)}, Y_{2(m)}, \ldots, Y_{k(m)}\} - x_0| > \epsilon\right\} \\ = \; & P\left\{\min\{Y_{1(m)}, Y_{2(m)}, \ldots, Y_{k(m)}\} > x_0 + \epsilon\right\} \\ = \; & [1 - F^m(x_0 + \epsilon)]^k \to 0, \end{aligned}$$

as $k \to \infty$. Hence, $X_{k(h+1)} \leq \min\{Y_{1(m)}, Y_{2(m)}, \ldots, Y_{k(m)}\} \to x_0$ in probability as $k \to \infty$. ■

The mean time for generating a G from a TCRD with κ is proportional to $\frac{N}{N_\kappa}$, where N and N_κ are the number of permutations in $\mathcal{P}_n$ and $\{G \in \mathcal{P}_n : \text{eff}(G) \leq \kappa\}$ respectively. In the sequential procedure, κ_i is a random sequence. If m is large, we can expect N_{κ_i} to be around $\rho^{i-1}N$, where $\rho = \frac{k}{m}$. Therefore, the mean time for generating a G from the TCRD in each iteration of the sequential procedure behaves like a power function of the number of iterations. Although it is very cheap to generate a G from the CRD, the cost in the later iterations may be very high if we need a large number of iterations to select an appropriate truncation constant κ. The number of iterations is not too large in our empirical study, so the sequential procedure is computationally cheap.

To illustrate how to select κ for TCRD by the sequential procedure, we consider analysis of covariance models and employ data sets in Bellhouse (1986) for values of covariates. Table 2.2 in Section 2.8 lists the approximate 95% confidence intervals of $\text{cv}(\phi)$ for the CRDs and shows that the CRDs provide acceptable levels of validity for the analysis of covariance models with one or two covariates when the number of units are larger than 10. However, the CRDs have low efficiency under the analysis of covariance models. Hence, we need to use a TCRD with appropriate κ to improve efficiency without reducing much validity.

In order to compare with the result of Cox (1982), we use $\frac{W_p}{d_w}$ defined in (3.2.3) as the efficiency measure $\text{eff}(G)$ and use $\mu 2_{\text{val}} = \sqrt{\mu_{\text{val}}^2 + \sigma_{\text{val}}^2}$ to measure the validity of a TCRD. It can be shown that $\frac{W_p}{d_w}$ is equivalent to $\text{tr}(A^T C C^T A)^{-1} - q$ under model (1.2.5). Here, only two treatments are compared in experiments, so the criterion $\text{tr}(A^T C C^T A)^{-1}$ is equal to $\det\{(A^T C C^T A)^{-1}\}$. For each data set, we generate a sequence of TCRDs with a decreasing sequence of κ_i by the sequential procedure.

For each TCRD, we generate $m = 60$ replicates of G to estimate the validity and efficiency. The 95 percentile of the distribution of eff(G) is used to measure the efficiency of a TCRD and is estimated by eff57, the 57th order statistic of eff(G) based on 60 replicates of G. $\mu 2_{\text{val}}$ is estimated by $\widehat{\mu 2}_{\text{val}} = \sqrt{\hat{\mu}^2_{\text{val}} + \hat{\sigma}^2_{\text{val}}}$, where $(\hat{\mu}_{\text{val}}, \hat{\sigma}_{\text{val}})$ is defined by (2.2.4) and (2.2.5) in Section 2.2, or see Hooper (1993). We choose $\frac{k}{m}$ to be $\frac{2}{3}$ in the sequential procedure.

Tables 3.1 - 3.6 contain the values of κ, $\widehat{\mu 2}_{\text{val}}$ and eff57 obtained by the sequential procedure for each of 24 data sets from Bellhouse (1986). The corresponding plots of $\widehat{\mu 2}_{\text{val}}$ against eff57 are displayed on Figures 3.1 - 3.6. We should note that the points in each plot are correlated since at each iteration in the sequential procedure we keep $\frac{2}{3}$ of G's at previous iteration to estimate the efficiency and validity. This yields a smoother plot than without keeping any G at each iteration.

Table 3.1: The estimated $\mu 2_{\text{val}}$ and eff57 in analysis of covariance model with $n = 10$

One Covariate						Two Covariates					
Data 1			Data 2			Data 1			Data 2		
κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57
∞	1.062	.3745	∞	1.079	.6133	∞	1.043	.9981	∞	1.023	1.964
.0984*	1.122	.0795	.0939*	1.192	.0834	.2878*	1.155	.2598	.3126*	1.16	.2864
.0263	1.207	.0218	.0539*	1.173	.0488	.1766*	1.266	.1620	.1411	1.288	.1369
.0143	1.277	.0142	.0196	1.321	.0182	.1188	1.308	.1090	.0995	1.218	.0958
.0083	1.413	.0059	.0086	1.570	.0086	.0556	1.408	.0521	.0534	1.438	.0494
.0024	1.599	.0024				.0240	1.687	.0235	.0334	1.510	.0334
						.0158	1.956	.0157	.0284	1.774	.0283
						.0068	2.244	.0068	.0194	1.983	.0189
									.0084	2.240	.0076

For those plots with $n = 10, 12$ and 14, the relationship between $\widehat{\mu 2}_{\text{val}}$ and eff57 can be well fitted by a curve. We can select the asterisked κ as a truncation constant which determines a TCRD with substantial improvement in efficiency and little reduction in validity. Smaller values of κ yield substantial improvement in efficiency but also substantial reduction in validity. Plots for larger values of n show a more random

Table 3.2: The estimated $\mu 2_{val}$ and eff57 in analysis of covariance model with $n = 12$

One Covariate						Two Covariates					
Data 1			Data 2			Data 1			Data 2		
κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57
∞	1.071	.6063	∞	1.032	.6971	∞	0.998	.6244	∞	1.046	.6503
.1237*	1.077	.1006	.0788*	1.101	.0605	.2665	1.060	.2512	.2212	1.159	.1949
.0438	1.174	.0438	.0450	1.104	.0421	.1615*	1.100	.1577	.1344*	1.202	.1304
.0116	1.193	.0116	.0233	1.144	.0160	.1089	1.165	.1028	.0974	1.225	.0940
.0047	1.293	.0047	.0102	1.196	.0102	.0809	1.268	.0779	.0823	1.342	.0814
.0024	1.346	.0024	.0040	1.277	.0040	.0603	1.418	.0564	.0576	1.375	.0557
.0009	1.488	.0009	.0014	1.451	.0014	.0508	1.675	.0508	.0511	1.574	.0503
						.0397	1.781	.0390	.0336	1.805	.0324
						.0312	1.931	.0304	.0282	2.141	.0281
									.0156	2.370	.0156

Table 3.3: The estimated $\mu 2_{val}$ and eff57 in analysis of covariance model with $n = 14$

One Covariate						Two Covariates					
Data 1			Data 2			Data 1			Data 2		
κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57
∞	1.055	.4592	∞	1.068	.2564	∞	1.016	.8010	∞	1.164	.6662
.0846	1.040	.0682	.0714	1.107	.0614	.2238*	1.056	.1943	.2466	1.086	.2269
.0318*	1.065	.0288	.0229*	1.145	.0185	.0990	1.244	.0901	.1263	1.113	.1122
.0130	1.098	.0115	.0078	1.114	.0073	.0453	1.165	.0436	.0752*	1.086	.0679
.0065	1.120	.0057	.0029	1.216	.0025	.0343	1.176	.0319	.0503	1.306	.0451
.0018	1.223	.0016	.0010	1.247	.0009	.0228	1.171	.0216	.0302	1.220	.0271
.0007	1.302	.0007	.0004	1.405	.0004	.0161	1.278	.0149	.0175	1.240	.0172
						.0108	1.214	.0105	.0098	1.327	.0089
						.0079	1.363	.0077	.0069	1.435	.0068
						.0058	1.424	.0054	.0051	1.522	.0048
						.0038	1.707	.0033	.0036	1.831	.0034

Table 3.4: The estimated $\mu 2_{\text{val}}$ and eff57 in analysis of covariance model with $n = 16$

One Covariate						Two Covariates					
Data 1			Data 2			Data 1			Data 2		
κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57
∞	1.076	.2222	∞	1.036	.3561	∞	1.055	.6419	∞	1.078	.5678
.0690	1.163	.0454	.1001	1.067	.0689	.1936	1.035	.1853	.1781	1.117	.1606
.0208	1.300	.0179	.0273	1.056	.0225	.1100	1.028	.0942	.1038	1.081	.0984
.0077	1.290	.0068	.0146	1.199	.0123	.0544	1.079	.0517	.0766	1.046	.0697
.0050	1.222	.0045	.0084	1.0909	.0067	.0389	1.076	.0372	.0367	1.112	.0328
.0024	1.357	.0022	.0039	1.110	.0039	.0278	1.121	.0266	.0234	1.063	.0220
.0009	1.497	.0006	.0019	1.189	.0019	.0211	1.171	.0199	.0152	1.090	.0147
			.0011	1.1265	.0011	.0139	1.181	.0125	.0096	1.039	.0089
			.0006	1.1568	.0006	.0088	1.401	.0085	.0065	1.167	.0063
			.0002	1.2140	.0002	.0059	1.555	.0057	.0044	1.215	.0040
									.0032	1.240	.0032
									.0021	1.372	.0021

Table 3.5: The estimated $\mu 2_{\text{val}}$ and eff57 in analysis of covariance model with $n = 18$

One Covariate						Two Covariates					
Data 1			Data 2			Data 1			Data 2		
κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57	κ	$\widehat{\mu 2}_{\text{val}}$	eff57
∞	0.977	.3392	∞	1.091	.2809	∞	1.041	.5709	∞	1.022	.3888
.0646	0.972	.0533	.0501	1.171	.0390	.1753	1.084	.1526	.1588	1.063	.1450
.0198	0.950	.0198	.0137	1.104	.0121	.0908	1.118	.0867	.0724	1.054	.0704
.0094	1.053	.0094	.0078	1.141	.0066	.0484	1.086	.0427	.0496	1.086	.0473
.0056	1.124	.0056	.0036	1.144	.0036	.0284	1.113	.0278	.0319	1.108	.0280
.0029	1.131	.0029	.0021	1.066	.0021	.0166	1.148	.0151	.0169	1.066	.0161
.0011	1.166	.0010	.0013	1.081	.0010	.0106	1.167	.0104	.0107	1.062	.0105
			.0006	1.083	.0006	.0074	1.127	.0073	.0067	1.090	.0064
			.0001	1.1684	.0001	.0050	1.157	.0048	.0052	1.144	.0045
						.0035	1.255	.0034	.0036	1.130	.0033
						.0026	1.288	.0025	.0021	1.090	.0020
						.0018	1.362	.0017	.0018	1.218	.0017

Table 3.6: The estimated $\mu 2_{val}$ and eff57 in analysis of covariance model with $n = 20$

One Covariate						Two Covariates					
Data 1			Data 2			Data 1			Data 2		
κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57	κ	$\widehat{\mu 2}_{val}$	eff57
∞	1.037	.1704	∞	1.121	.2788	∞	1.039	.4592	∞	1.033	.4836
.0581	1.074	.0453	.0767	1.040	.0576	.1289	1.047	.1159	.1471	1.095	.1381
.0223	1.071	.0196	.0234	0.986	.0181	.0733	1.063	.0685	.0903	1.057	.0818
.0076	1.102	.0076	.0101	1.032	.0101	.0394	1.036	.0371	.0541	1.064	.0521
.0027	1.046	.0027	.0045	1.015	.0045	.0257	1.002	.0246	.0289	1.034	.0278
.0012	1.078	.0012	.0011	1.035	.0011	.0154	0.986	.0148	.0219	1.045	.0202
						.0099	1.026	.0093	.0153	1.132	.0148
						.0067	1.106	.0065	.0116	1.172	.0105
						.0051	1.099	.0048	.0079	1.142	.0077
						.0034	1.119	.0029	.0056	1.215	.0054
						.0022	1.128	.0022	.0034	1.228	.0032
						.0014	1.034	.0013	.0025	1.292	.0022
						.0010	1.104	.0010	.0016	1.399	.0015
						.0006	1.228	.0006			

pattern. Here $\widehat{\mu 2}_{val}$ is nearly constant over a wide range of values of κ and the variation in $\widehat{\mu 2}_{val}$ is due in large part to random variation in the estimation. Note that the scale on the validity axis decreases as the number n of units increases. This is why plots for larger n appear more random. In this situation, we can select a small value of κ so that the corresponding TCRD has very high degree of efficiency and validity.

Tables 3.1 - 3.6 and Figures 3.1 - 3.6 suggest that, for $n = 10, 12$ and $n = 14$ with two covariates, the truncation constant κ might be chosen larger than the value 0.05 suggested by Cox (1982), in order to provide greater validity. For $n = 14$ with one covariate, $n = 16, 18$ and 20, it appears that κ can be chosen to be less than 0.05 without substantially reducing validity.

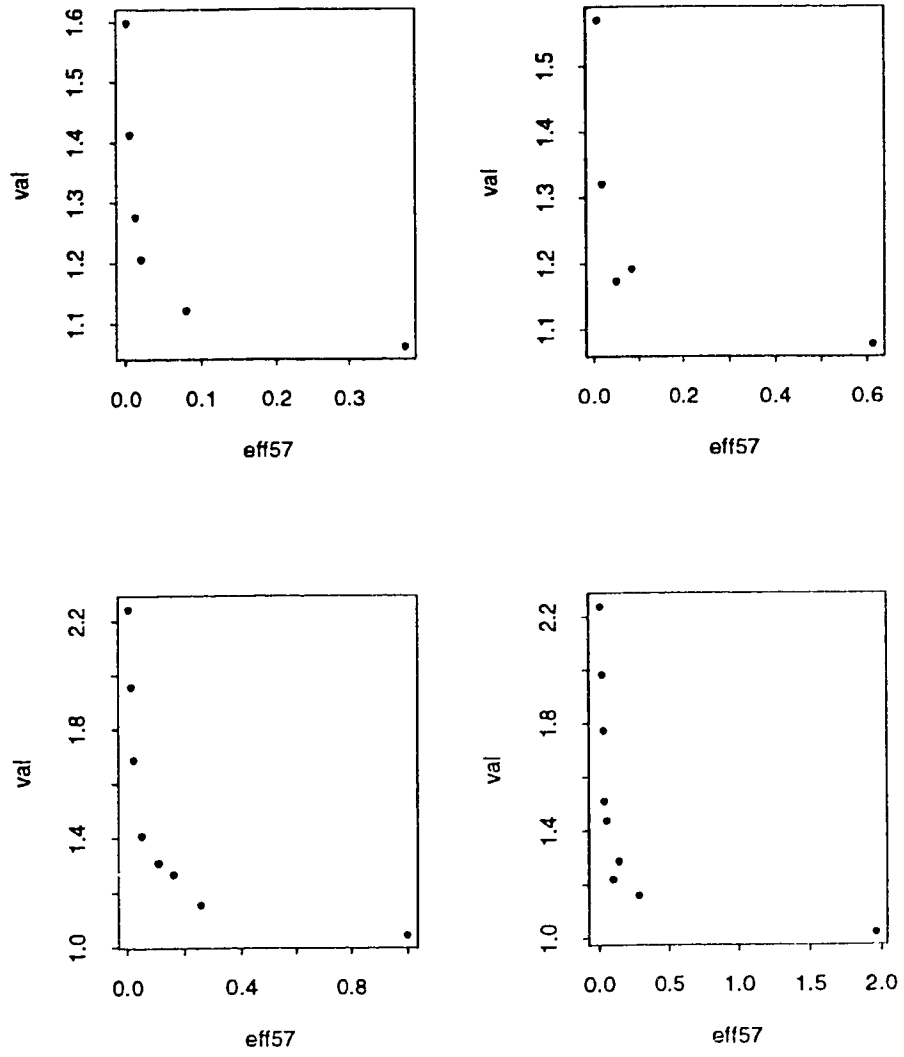


Figure 3.1: Plots of validity vs. efficiency for TCRD in analysis of covariance model with $n = 10$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

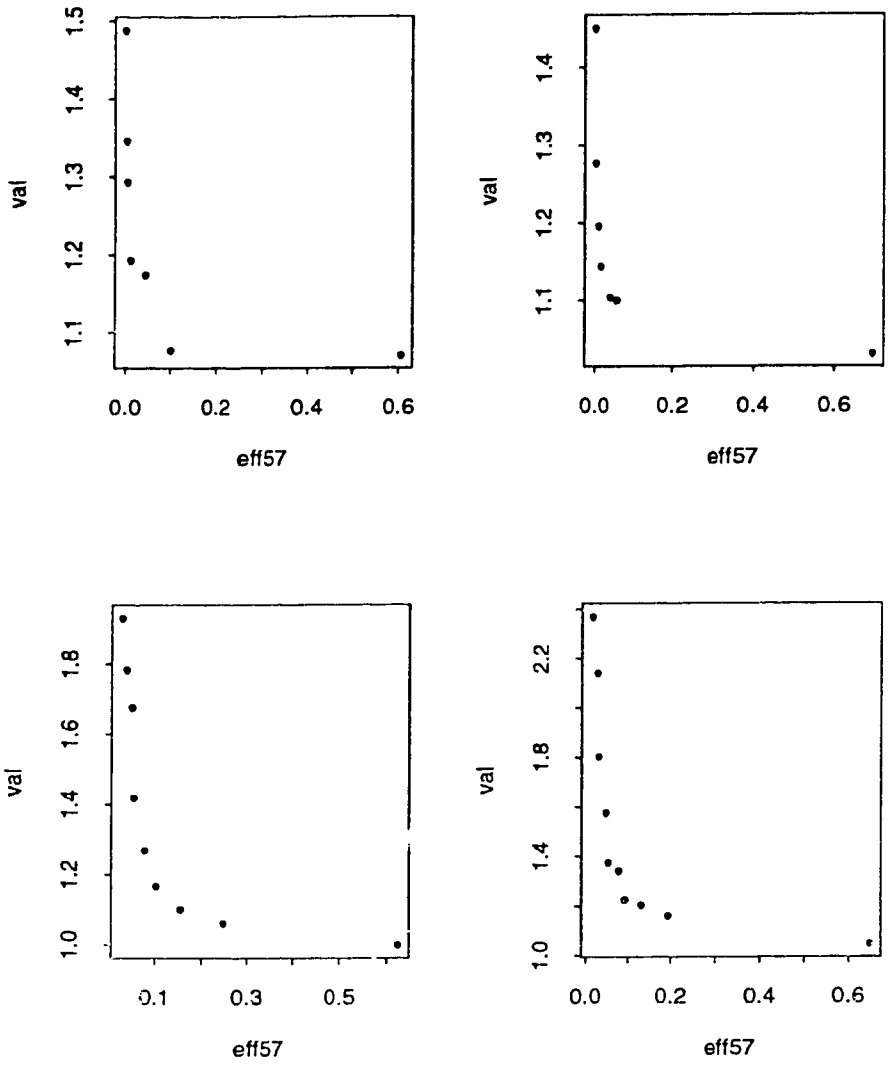


Figure 3.2: Plots of validity vs efficiency for TCRD in analysis of covariance model with $n = 12$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

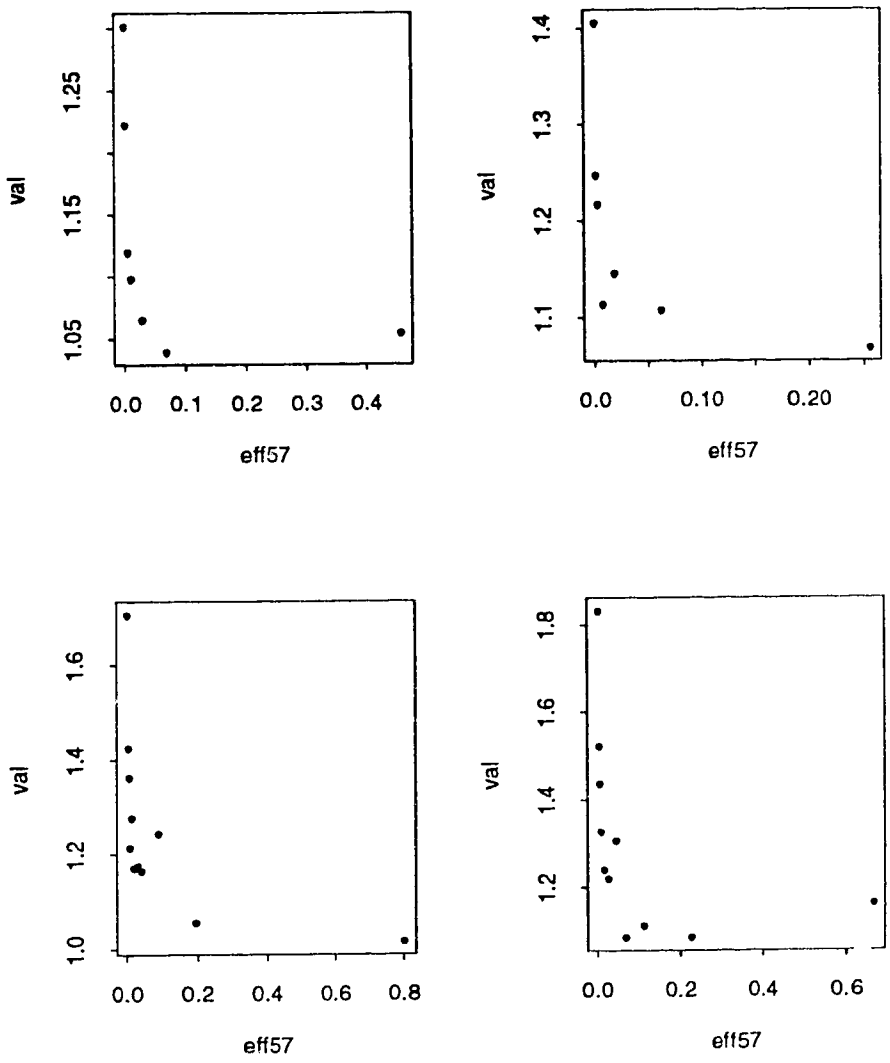


Figure 3.3: Plots of validity vs. efficiency for TCRD in analysis of covariance model with $n = 14$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

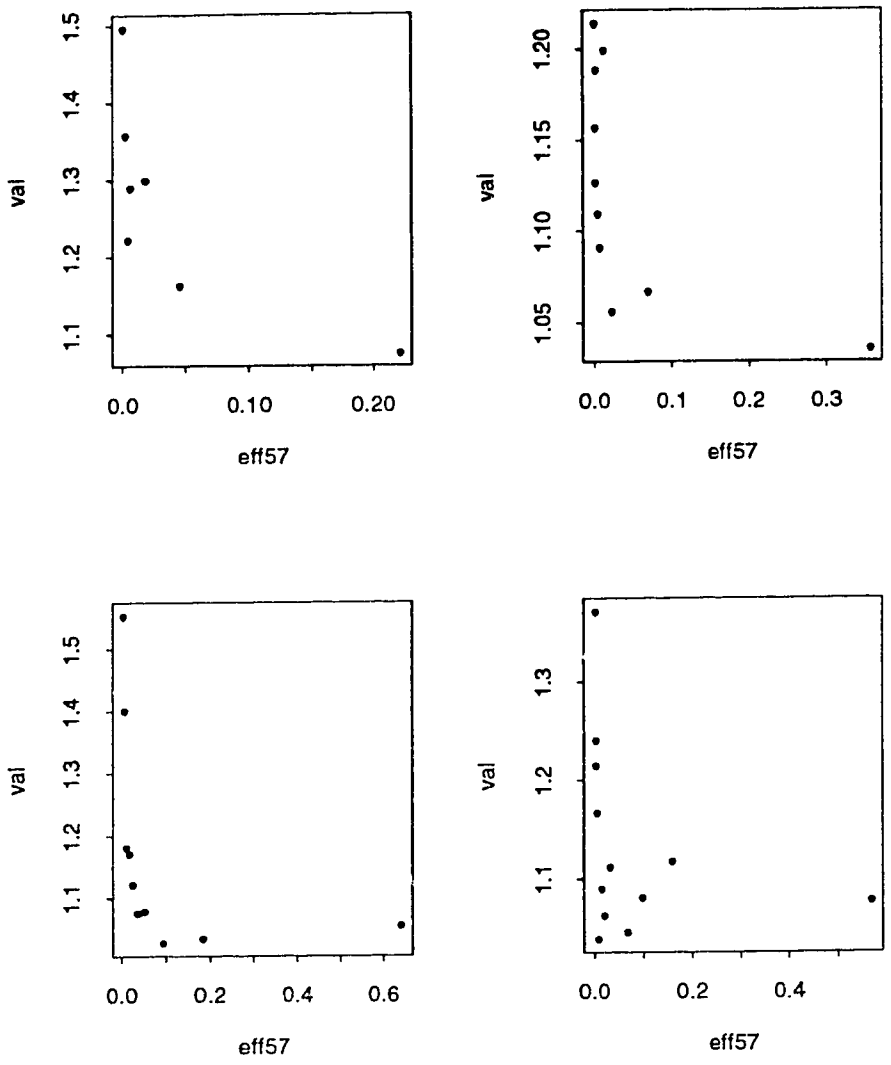


Figure 3.4: Plots of validity vs. efficiency for TCRD in analysis of covariance model with $n = 16$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

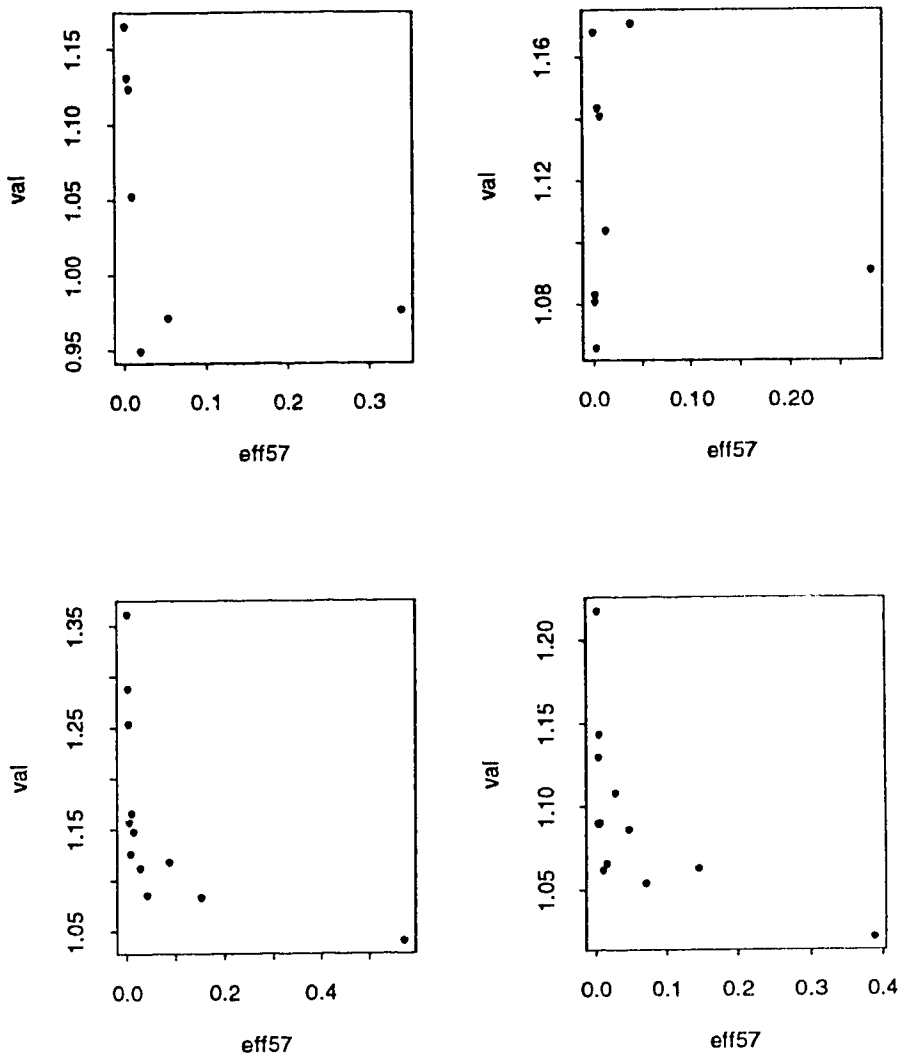


Figure 3.5: Plots of validity vs. efficiency for TCRD in analysis of covariance model with $n = 18$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

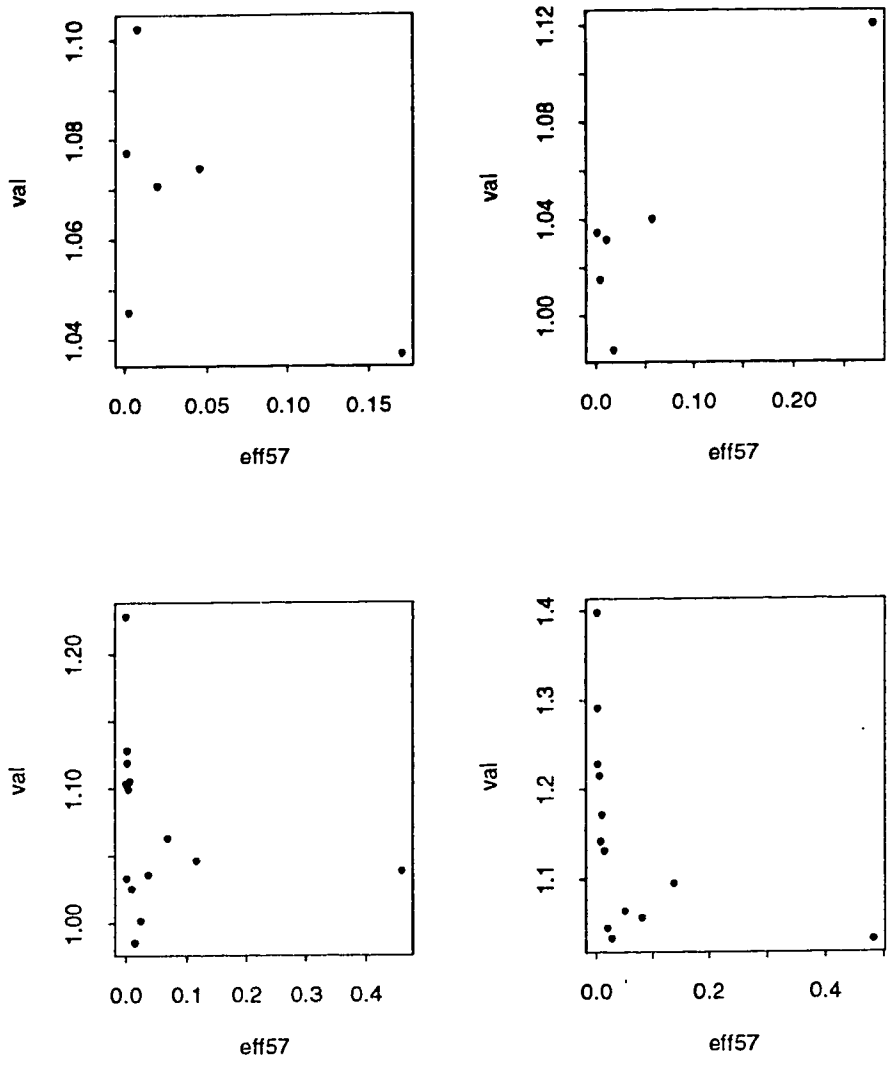


Figure 3.6: Plots of validity vs. efficiency for TCRD in analysis of covariance model with $n = 20$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

3.4 Nearly Orthogonal Randomized Designs

The NORD is motivated by the criteria of validity and efficiency. The two criteria are partially in conflict with each other. Restricting G to a class of designs that are optimal in some sense may leave too few designs to provide much validity. In seeking a compromise between efficiency and validity, Hooper (1993) considered distributions on the $n \times n$ orthogonal matrix group $\mathcal{O}_n$ as an extension of the notion of randomization, and found an optimal distribution on $\mathcal{O}_n$ that achieves full efficiency and full validity. The NORD is defined as a distribution on $\mathcal{P}_n$ that approximates the optimal one on $\mathcal{O}_n$.

An optimal distribution on $\mathcal{O}_n$ can be defined as follows. Let H_1 be a non-random orthogonal matrix such that $A^T H_1 B = 0$. Let H_2 be distributed uniformly on the subgroup $\mathcal{O}_n(B) = \{J \in \mathcal{O}_n : \mathcal{R}(JB) = \mathcal{R}(B)\}$ and independently of $W = KV$. If we replace G by $H = H_1 H_2$ in definition (1.2.9) then we obtain $C = A$. Thus H is fully efficient. Hooper (1993) also proved that the conditional distribution of $Q(H, W, 0)$ given W is the same for all vectors W with $(I_n - BB^T)W \neq 0$. Hence H provides full validity.

Let H be a random orthogonal matrix with distribution defined in the above way. Hooper (1993) defined NORD as a random permutation matrix G which minimizes a measure of distance between G and H. The measure of distance involves only $\mathcal{R}(G^T A)$ and $\mathcal{R}(H^T A)$. It is shown in Hooper (1993) that $H^T A$ has the same distribution as D where $[B : D] = \text{GramSchmidt}[B : Z]$ with $Z \sim N_{n \times q}(0, I_n \otimes I_q)$. Thus we can generate a NORD by generating D and then finding a G such that $\mathcal{R}(G^T A)$ is closest to $\mathcal{R}(D)$.

The NORD Algorithm is described as follows:

(a) Generate an $n \times q$ matrix Z of independent, standard normal random variables.

(b) Construct D via $[B : D] = \text{GramSchmidt}\,[B : Z]$. Choose $G \in \mathcal{P}_n$ to minimize

$$\text{dis}(G) = q - \|A^T G D\|^2 + \lambda \|A^T G B\|^2. \tag{3.4.1}$$

(c) If condition (1.2.6) fails then repeat the preceding steps.

Distance function (3.4.1) was defined on an *ad hoc* basis taking into consideration efficiency, validity and tractability. The first term $q - \|A^TGD\|^2$ addresses both efficiency and validity and the second term $\lambda\|A^TGB\|^2$ permits greater emphasis on efficiency. λ is a tuning constant to determine a trade-off between efficiency and validity. Increasing λ forces $\mathcal{R}(GB)$ to be more nearly orthogonal to $\mathcal{R}(A)$, which improves efficiency and reduces validity. Hooper (1993) used a simulated annealing algorithm to carry out the minimization in step (b).

3.5 Truncated Nearly Orthogonal Randomized Designs

The main idea of the NORD is to define a distribution on $\mathcal{P}_n$ to approximate the optimal one on $\mathcal{O}_n$. This is carried out by minimizing the distance function (3.4.1) in which a constant λ needs to be selected appropriately to achieve acceptable levels of validity and efficiency. However, the tuning constant λ in NORD does not directly relate to any efficiency measure so that we cannot use a sequential procedure to assist in the selection of λ as we did with the TCRD. If we instead minimize a distance between $\mathcal{R}(G^TA)$ and $\mathcal{R}(H^TA)$ subject to an efficiency measure being small, the sequential procedure for TCRD can be adopted here.

In this section, our discussion is based on the general linear model (1.2.5). In the treatment-effects model (1.2.3), we used a convenient canonical parameterization such that the columns of $[A_0 : A]$ are orthonormal. It is desirable to adopt a randomization strategy that is invariant under reparameterization of the treatment-effects model. Following Hooper (1993), we choose $\text{dis}(G) = q - \|A^TGD\|^2$ as the distance function. Minimizing $\text{dis}(G)$ makes the random subspace $\mathcal{R}(G^TA)$ close to $\mathcal{R}(D) \sim \mathcal{R}(H^TA)$. Recall that $\mathcal{R}(A)$ is the subspace of $\mathcal{R}^n$ corresponding to the treatment-effects of

interest after removal of nuisance parameters. Thus $\mathcal{R}(A)$ does not depend on the parameterization of the treatment-effects model. Since dis(G) is a distance between the column spaces $\mathcal{R}(G^TA)$ and $\mathcal{R}(D)$, the minimization of dis(G) is invariant under reparameterization of the treatment-effects model. In Section 1.3, we discussed the efficiency of a randomized design and showed that the D-optimality criterion is invariant under reparameterization, i.e., the minimization of $\det[(A^TCC^TA)^{-1}]$ does not depend on the parameterization of the model for treatment effects. Thus we use $\text{eff}(G) = \det[(A^TCC^TA)^{-1}]$ as the efficiency measure in the following modified NORD. Because we truncate the efficiency measure and then minimize the distance function, we call the new randomization strategy a truncated NORD (TNORD).

The TNORD algorithm is described as follows.

(a) Generate an $n \times q$ matrix Z of independent standard normal random variables.

(b) Construct D via $[B : D] = \text{GramSchmidt}\,[B : Z]$. Choose $G \in \mathcal{P}_n$ to minimize $\text{dis}(G) = q - \|A^TGD\|^2$ subject to the constraint $\text{eff}(G) \leq \kappa$, where $\text{eff}(G) = \det[(A^TCC^TA)^{-1}]$.

The constrained minimization problem in (b) can be approximated by minimizing the unrestricted distance function:

$$\text{dis}_u(G) = q - \|A^TGD\|^2 + \nu\,[\text{eff}(G) - \kappa]^+, \tag{3.5.1}$$

where ν is large positive constant and

$$[\text{eff}(G) - \kappa]^+ = \begin{cases} \text{eff}(G) - \kappa & \text{if } \text{eff}(G) > \kappa \\ 0 & \text{if } \text{eff}(G) \leq \kappa\,. \end{cases}$$

When $\nu \longrightarrow \infty$, the solution to minimizing the unrestricted distance (3.5.1) converges to that of the constrained minimization problem. The unrestricted minimization can be carried out with the simulated annealing algorithm in Hooper (1993). The simulated annealing algorithm to minimize (3.5.1) and some formulae used in this algorithm are described in the next section.

Like the truncation constant in TCRD, the selection of κ makes a trade-off between validity and efficiency. A larger κ determines a TNORD with higher validity and lower efficiency. As with TCRD, a sequential procedure can be used to select an appropriate truncation constant κ so that the TNORD achieves acceptable levels of efficiency and validity.

3.6 The Algorithm to Minimize $\mathrm{dis}_u(G)$ and Some Formulae

The solution G to (3.5.1) is approximated by a product of transpositions. We adopt the simulated annealing algorithm in Hooper (1993) to minimize (3.5.1). This algorithm is based on an algorithm given by Press *et al.* (1986).

Algorithm to Minimize $\mathrm{dis}_u(G)$

Generate an initial random permutation G

Generate independent random transpositions $H_1, \ldots, H_{10}$

Set the initial temperature $t = \max\{\mathrm{dis}_u(H_iG) - \mathrm{dis}_u(G), i = 1, \ldots, 10\}$

Do 100 times (at most 100 temperature steps)

  Set count = 0

  Do $100n$ times (at most $100n$ transpositions per step)

    Generate a random transposition H

    Set $\Delta = \mathrm{dis}_u(HG) - \mathrm{dis}_u(G)$

    If $\Delta < 0$ then

      Replace G by HG

      Increment count by 1 (count number of successes)

    Else

      Generate a uniform $(0, 1)$ random variable U

If $U < \exp(-\Delta/t)$ then replace G by HG

End if

If count $\geq 10n$ then exit do loop (at most $10n$ successes per step)

Repeat

If count $= 0$ then exit do loop (stop when there are no successes)

Replace t by $0.9t$ Repeat

Return G as solution

In the simulated annealing algorithm, we need to compute the difference of the loss (3.5.1) by a transposition. Before deriving a formula for this difference, we introduce some general results about transpositions.

In the following, we will let G and H denote both permutations of $\{1, 2, \ldots, n\}$ and $n \times n$ permutation matrices. Let H be a transposition such that $HGi = Gi$ if $i \neq k$ or l, $HGk = Gl$, and $HGl = Gk$. Let X be an $n \times n$ matrix and denote its (i, j) element by x_{ij}. Write $A^T = (a_1, \cdots, a_n)$, i.e., a_j is the jth row vector of the coefficient matrix A in the treatment-effects model. We then have the following results.

Lemma 3.6.1

$$A^T HGXG^T H^T A = A^T GXG^T A + fg^T + gf^T,$$

where $g = a_{Gk} - a_{Gl}$ *and* $f = \sum_{i \neq k \text{ or } l}(x_{il} - x_{ik})a_{Gi} + \frac{1}{2}(x_{ll} - x_{kk})(a_{Gk} + a_{Gl})$.

Proof: Noting $A^T HG = (a_{HG1}, \cdots, a_{HGn})$, we obtain

$$\begin{aligned}
& A^T HGXG^T H^T A \\
= & \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} a_{HGi} a_{HGj}^T \\
= & \sum_{i \neq k \text{ or } l}\sum_{j \neq k \text{ or } l} x_{ij} a_{Gi} a_{Gj}^T + \sum_{j \neq k \text{ or } l} x_{kj} a_{Gl} a_{Gj}^T + \sum_{j \neq k \text{ or } l} x_{lj} a_{Gk} a_{Gj}^T \\
& + x_{kk} a_{Gl} a_{Gl}^T + x_{kl} a_{Gl} a_{Gk}^T + x_{lk} a_{Gk} a_{Gl}^T + x_{ll} a_{Gk} a_{Gk}^T \\
& + \sum_{i \neq k \text{ or } l} x_{ik} a_{Gi} a_{Gl}^T + \sum_{i \neq k \text{ or } l} x_{il} a_{Gi} a_{Gk}^T
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}a_{Gi}a_{Gj}^T + \sum_{j\neq k \text{ or } l} x_{kj}(a_{Gl}-a_{Gk})a_{Gj}^T + \sum_{j\neq k \text{ or } l} x_{lj}(a_{Gk}-a_{Gl})a_{Gj}^T \\
&\quad + x_{kk}(a_{Gl}a_{Gl}^T - a_{Gk}a_{Gk}^T) + x_{ll}(a_{Gk}a_{Gk}^T - a_{Gl}a_{Gl}^T) \\
&\quad + \sum_{i\neq k \text{ or } l} x_{ik}a_{Gi}(a_{Gl}-a_{Gk})^T + \sum_{i\neq k \text{ or } l} x_{il}a_{Gi}(a_{Gk}-a_{Gl})^T \\
&= A^TGXG^TA + \sum_{j\neq k \text{ or } l} (x_{kj}-x_{lj})(a_{Gl}-a_{Gk})a_{Gj}^T \\
&\quad + (x_{kk}-x_{ll})(a_{Gl}a_{Gl}^T - a_{Gk}a_{Gk}^T) + \sum_{i\neq k \text{ or } l} (x_{ik}-x_{il})a_{Gi}(a_{Gl}-a_{Gk})^T \\
&= A^TGXG^TA + \sum_{j\neq k \text{ or } l} (x_{kj}-x_{lj})(a_{Gl}-a_{Gk})a_{Gj}^T \\
&\quad + \frac{1}{2}(x_{kk}-x_{ll})(a_{Gl}-a_{Gk})(a_{Gk}+a_{Gl})^T + \frac{1}{2}(x_{kk}-x_{ll})(a_{Gk}+a_{Gl})(a_{Gl}-a_{Gk})^T \\
&\quad + \sum_{i\neq k \text{ or } l} (x_{ik}-x_{il})a_{Gi}(a_{Gl}-a_{Gk})^T \\
&= A^TGXG^TA + (a_{Gl}-a_{Gk})\left\{\sum_{j\neq k \text{ or } l} (x_{kj}-x_{lj})a_{Gj}^T + \frac{1}{2}(x_{kk}-x_{ll})(a_{Gk}+a_{Gl})^T\right\} \\
&\quad + \left\{\sum_{i\neq k \text{ or } l} (x_{ik}-x_{il})a_{Gi} + \frac{1}{2}(x_{kk}-x_{ll})(a_{Gk}+a_{Gl})\right\}(a_{Gl}-a_{Gk})^T \\
&= A^TGXG^TA + (a_{Gk}-a_{Gl})\left\{\sum_{j\neq k \text{ or } l} (x_{lj}-x_{kj})a_{Gj}^T + \frac{1}{2}(x_{ll}-x_{kk})(a_{Gk}+a_{Gl})^T\right\} \\
&\quad + \left\{\sum_{i\neq k \text{ or } l} (x_{il}-x_{ik})a_{Gi} + \frac{1}{2}(x_{ll}-x_{kk})(a_{Gk}+a_{Gl})\right\}(a_{Gk}-a_{Gl})^T.
\end{aligned}
$$

Letting $g = a_{Gk} - a_{Gl}$ and $f = \sum_{i\neq k \text{ or } l}(x_{il}-x_{ik})a_{Gi} + \frac{1}{2}(x_{ll}-x_{kk})(a_{Gk}+a_{Gl})$, we then have

$$A^THGXG^TH^TA = A^TGXG^TA + fg^T + gf^T.$$

This completes the proof. ∎

Lemma 3.6.2 *Suppose that A^TGXG^TA and $A^THGXG^TH^TA$ are nonsingular. Let $M = A^TGXG^TA$ and $\Delta = (1+f^TM^{-1}g)^2 - f^TM^{-1}fg^TM^{-1}g$. We then have*

$$(A^THGXG^TH^TA)^{-1} = (M+fg^T+gf^T)^{-1}$$

$$= M^{-1} - \frac{1}{\Delta} M^{-1}\{(1 + f^T M^{-1} g)(fg^T + gf^T) - g^T M^{-1} g f f^T - f^T M^{-1} f g g^T\} M^{-1}$$

Proof: By Lemma 3.6.1, we have

$$A^T H G X G^T H^T A = M + fg^T + gf^T.$$

Applying twice the Sherman-Morrison formula for the inverse of the matrix, we have after a simplification

$$\begin{aligned}
& (M + fg^T + gf^T)^{-1} \\
= \; & (M + fg^T)^{-1} - \frac{(M + fg^T)^{-1} g f^T (M + fg^T)^{-1}}{1 + f^T (M + fg^T)^{-1} g} \\
= \; & M^{-1} - \frac{M^{-1} f g^T M^{-1}}{1 + g^T M^{-1} f} \\
& - \frac{M^{-1}(I - \frac{f g^T M^{-1}}{1 + g^T M^{-1} f}) g f^T (I - \frac{M^{-1} f g^T}{1 + g^T M^{-1} f}) M^{-1}}{1 + f^T (M^{-1} - \frac{M^{-1} f g^T M^{-1}}{1 + g^T M^{-1} f}) g} \\
= \; & M^{-1} - \frac{M^{-1} f g^T M^{-1}}{1 + g^T M^{-1} f} \\
& - \frac{M^{-1}\{g f^T - \frac{f g^T M^{-1} g f^T}{1 + g^T M^{-1} f} - \frac{g f^T M^{-1} f g^T}{1 + g^T M^{-1} f} + \frac{f g^T M^{-1} g f^T M^{-1} f g^T}{(1 + g^T M^{-1} f)^2}\} M^{-1}}{1 + f^T M^{-1} g - \frac{f^T M^{-1} f g^T M^{-1} g}{1 + g^T M^{-1} f}} \\
= \; & M^{-1} - \frac{M^{-1} f g^T M^{-1}}{1 + g^T M^{-1} f} - \Delta^{-1} (1 + g^T M^{-1} f)^{-1} M^{-1} \\
& \{(1 + g^T M^{-1} f)^2 g f^T - g^T M^{-1} g (1 + g^T M^{-1} f) f f^T \\
& - f^T M^{-1} f (1 + g^T M^{-1} f) g g^T + g^T M^{-1} g f^T M^{-1} f f g^T\} M^{-1} \\
= \; & M^{-1} - \frac{1}{\Delta} M^{-1}\{(1 + f^T M^{-1} g)(fg^T + gf^T) \\
& - g^T M^{-1} g f f^T - f^T M^{-1} f g g^T\} M^{-1}
\end{aligned}$$

This completes the proof. ∎

Lemma 3.6.3 *Suppose that A^TGXG^TA is nonsingular. Let $M = A^TGXG^TA$ and $\Delta = (1 + f^TM^{-1}g)^2 - f^TM^{-1}fg^TM^{-1}g$. We then have*

$$\det(M + fg^T + gf^T) = \Delta \det(M)$$

and $M + fg^T + gf^T$ is nonsingular if and only if $\Delta \neq 0$.

Proof: It is well known that

$$\begin{aligned} \det\begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} &= \det(\Gamma_{11})\det(\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}) \\ &= \det(\Gamma_{22})\det(\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}), \end{aligned}$$

where Γ_{11} and Γ_{22} are nonsingular square matrices (may be of different orders); see page 23 of Rao (1973). Letting $\Gamma_{11} = M$, $\Gamma_{12} = u$, $\Gamma_{21} = -v^T$ and $\Gamma_{22} = 1$, we have

$$\begin{aligned} \det\begin{pmatrix} M & u \\ -v^T & 1 \end{pmatrix} &= \det(M)(1 + v^TM^{-1}u) \\ &= \det(M + uv^T). \end{aligned}$$

Hence, for any nonsingular $p \times p$ matrix M and two vectors u and v of dimension p, we have

$$\det(M + uv^T) = \det(M)(1 + v^TM^{-1}u). \tag{3.6.1}$$

Applying twice (3.6.1), we obtain

$$\begin{aligned} \det(M + fg^T + gf^T) &= \det(M + fg^T)\,[1 + f^T(1 + g^TM^{-1}f)g] \\ &= \det(M)\,(1 + g^TM^{-1}f)\left[1 + f^TM^{-1}g - \frac{f^TM^{-1}f\,g^TM^{-1}g}{1 + g^TM^{-1}f}\right] \\ &= \Delta \det(M). \end{aligned}$$

This completes the proof. ∎

We are now in the place to derive some formulae for the difference of the distance function (3.5.1) by a transposition in the simulated annealing algorithm. We assume $A_0 = \frac{1}{\sqrt{n}}\mathbf{1}_n = \frac{1}{\sqrt{n}}(1, 1, \cdots, 1)^T$, so $A^T CC^T A = I_q - A^T GBB^T G^T A$. Thus we have

$$\mathrm{dis}_u(G) = q - \|A^T GD\|^2 + \nu\,[\det(I_q - A^T GBB^T G^T A)^{-1} - \kappa]^+, \tag{3.6.2}$$

where

$$[\det(I_q - A^T GBB^T G^T A)^{-1} - \kappa]^+$$
$$= \begin{cases} \det(I_q - A^T GBB^T G^T A)^{-1} - \kappa, & \text{if } \det(I_q - A^T GBB^T G^T A)^{-1} > \kappa \\ 0, & \text{if } \det(I_q - A^T GBB^T G^T A)^{-1} \le \kappa\,. \end{cases}$$

Let

$$\mathrm{dis}_{u1}(G) = q - \|A^T GD\|^2 = q - \mathrm{tr}(A^T GDD^T G^T A)$$

and

$$\mathrm{dis}_{u2}(G) = \det(I_q - A^T GBB^T G^T A)^{-1}.$$

By Lemma 3.6.1, it follows that

$$\mathrm{dis}_{u1}(HG) - \mathrm{dis}_{u1}(G) = -\,\mathrm{tr}(f_x g^T + g f_x^T) = -\,2g^T f_x,$$

where $x_{ij} = (X)_{ij} = (DD^T)_{ij}$, $f_x = \sum_{i \neq k \text{ or } l}(x_{il} - x_{ik})a_{Gi} + \frac{1}{2}(x_{ll} - x_{kk})(a_{Gk} + a_{Gl})$ and $g = a_{Gk} - a_{Gl}$. Noting that $\mathrm{dis}_{u2}(G) = [\det(I_q - A^T GBB^T G^T A)]^{-1}$ and applying Lemma 3.6.3, we obtain

$$\mathrm{dis}_{u2}(HG) - \mathrm{dis}_{u2}(G) = (\Delta_y^{-1} - 1)\,\mathrm{dis}_{u2}(G),$$

where $\Delta_y = (1 + f_y^T M_y^{-1} g)^2 - f_y^T M_y^{-1} f_y g^T M_y^{-1} g)$ with

$$M_y = I_q - A^T GBB^T G^T A = A^T G(I_n - Y)G^T A,$$

$y_{ij} = (Y)_{ij} = (BB^T)_{ij}$ and $f_y = \sum_{i \neq k \text{ or } l}(y_{ik} - y_{il})a_{Gi} + \frac{1}{2}(y_{kk} - y_{ll})(a_{Gk} + a_{Gl})$. Here M_y^{-1} can be revised by the formula in Lemma 3.6.2 at each iteration.

Hence, if $\mathrm{dis}_{u2}(G) > \kappa$ and $\Delta_y^{-1}\,\mathrm{dis}_{u2}(G) > \kappa$,

$$\mathrm{dis}_u(HG) - \mathrm{dis}_u(G) = -2g^T f_x + \nu\,(\Delta_y^{-1} - 1)\,\mathrm{dis}_{u2}(G);$$

if $\mathrm{dis}_{u2}(G) > \kappa$ and $\Delta_y^{-1}\,\mathrm{dis}_{u2}(G) \le \kappa$,

$$\mathrm{dis}_u(HG) - \mathrm{dis}_u(G) = -2g^T f_x - \nu\,\{\mathrm{dis}_{u2}(G) - \kappa\};$$

if $\mathrm{dis}_{u2}(G) \le \kappa$ and $\Delta_y^{-1}\,\mathrm{dis}_{u2}(G) > \kappa$,

$$\mathrm{dis}_u(HG) - \mathrm{dis}_u(G) = -2g^T f_x + \nu\,\{\Delta_y^{-1}\,\mathrm{dis}_{u2}(G) - \kappa\};$$

if $\mathrm{dis}_{u2}(G) \le \kappa$ and $\Delta_y^{-1}\,\mathrm{dis}_{u2}(G) \le \kappa$,

$$\mathrm{dis}_u(HG) - \mathrm{dis}_u(G) = -2g^T f_x.$$

The time for generating a G from TNORD depends on the above simulated annealing algorithm to minimize distance function (3.6.2). The numbers of multiplications required to compute g, f_x and f_y are all of order n. The numbers of multiplications required to compute $g^T f_x$ and Δ_y are of order q and q^2 respectively, which are independent of n. Thus, the number of multiplications required to compute the change in the distance (3.6.2) by a transposition is of order n. Since the number of iterations is also of order n, the algorithm is of the same order n^2 as that for NORD.

3.7 Comparison of TNORDs with TCRDs and NORDs

The TNORD algorithm typically defines a weighted randomized design like the NORD algorithm, while TCRDs are defined by the uniform distribution on a subset of $\mathcal{P}_n$. The NORD requires the selection of a tuning constant λ to determine a trade-off between validity and efficiency, as well the TCRD and the TNORD require the selection

of a truncation constant κ. Increasing λ in NORD reduces validity and improves efficiency; increasing κ in TCRD and TNORD improves validity but reduces efficiency. One may question whether the same levels of validity and efficiency can be achieved by suitable choice of the tuning or truncation constant in these three methods. To investigate this question we conducted an empirical study under analysis of covariance models.

Instead of taking the natural parameterization used by Cox (1982) and Bellhouse (1986), we adopt the more general notation of Section 1.2. We write

$$Y = A_0\beta_0 + A\beta + GB\gamma + GKV, \tag{3.7.1}$$

where β_0 is $q_0 \times 1$ vector of nuisance parameters, β is the $q \times 1$ vector of parameters of interest and γ is the vector of regression coefficients of responses on covariates. The columns of $[A_0 : A]$ are orthonormal and $\mathcal{R}(A_0)$ contains $\mathbf{1}_n = (1, \ldots, 1)^T$. Let x_i be a vector of values of the ith covariate, $i = 1, \ldots, r-1$. B is obtained by applying the Gram-Schmidt orthonormalization procedure to the matrix $[\mathbf{1}_n : x_1 : \cdots : x_{r-1}]$, i.e., $B = \text{GramSchmidt}\,[\mathbf{1}_n : x_1 : \cdots : x_{r-1}]$. In this empirical study we are interested in comparison of two treatments, so $q_0 = 1$, $q = 1$ and $A_0 = \mathbf{1}_n$.

Values for covariates are taken from data sets in Bellhouse (1986). To see how λ and κ tune the efficiency and validity for NORD, TCRD and TNORD, for each data set we generate a sequence of TCRDs and a sequence of TNORDs with decreasing values of κ by the sequential procedure described in Section 3.3. We choose ρ to be $\frac{2}{3}$ in the sequential procedure. A sequence of NORDs with increasing values of λ is also generated. From each design, we generate $m = 60$ replicates of G to estimate the validity and efficiency.

Since the D-optimality criterion is invariant under reparameterization, we use $\text{eff}(G) = \det\{(A^TCC^TA)^{-1}\} - 1$ to measure the efficiency of the least squares estimator of β for a particular design G. However, the D-optimality criterion is equivalent

to the A-optimality criterion, $\mathrm{tr}\{(A^TCC^TA)^{-1}\}$, in this empirical study for comparison of two treatments. The 95 percentile of the distribution of $\mathrm{eff}(G)$ is used to measure the efficiency of a randomized design and is estimated by eff57, the 57th order statistic of $\mathrm{eff}(G)$ based on 60 replicates of G. To simplify our comparison, we use a real-valued validity measure $\mu 2_{\mathrm{val}} = \sqrt{\mu_{\mathrm{val}}^2 + \sigma_{\mathrm{val}}^2}$ to evaluate the validity of a randomized design. $\mu 2_{\mathrm{val}}$ is estimated by $\widehat{\mu 2}_{\mathrm{val}} = \sqrt{\hat{\mu}_{\mathrm{val}}^2 + \hat{\sigma}_{\mathrm{val}}^2}$, where $(\hat{\mu}_{\mathrm{val}}, \hat{\sigma}_{\mathrm{val}})$ is defined by (2.2.4) and (2.2.5) in Section 2.2, or see Hooper (1993).

We examine how the estimated validity measure $\widehat{\mu 2}_{\mathrm{val}}$ varies with the estimated efficiency measure eff57 in TCRD, NORD and TNORD as the tuning or truncation constant increases. For each data set, we present superimposed plots of $\widehat{\mu 2}_{\mathrm{val}}$ against eff57 for NORD, TCRD and TNORD. Figures 3.7 to 3.12 give scatterplots for 24 data sets from Bellhouse (1986). In all figures, the scatterplots for NORD, TCRD and TNORD are denoted by three symbols: $\cdot$, $\triangle$ and $+$ respectively. From these figures, we can see that the scatterplots gradually change from a smooth type to a random pattern type when the number n of units increases. However, we should notice that the difference in appearance for these scatterplots also depends on the scale used for the y-axis (validity). Based on this empirical study, it appears that, for comparing two treatments with one or two covariates, the three methods achieve the same trade-off between validity and efficiency.

We also point out that there is larger variation in the plots for NORDs than in those for TCRDs and TNORDs. This is expected because we generate 60 independent G's for each NORD so that the points in the plot are independent. However we use 40 G's from the previous design and generate additional 20 G's for each TCRD or TNORD, so neighbouring points have positive correlation and the plots appear smoother.

In the following, we consider $\frac{n}{2}$ treatments in comparative experiments with one or two covariates. Model (3.7.1) is assumed here with $q_0 = 1, q = \frac{n}{2} - 1$, and $A_0 = \mathbf{1}_n$. We

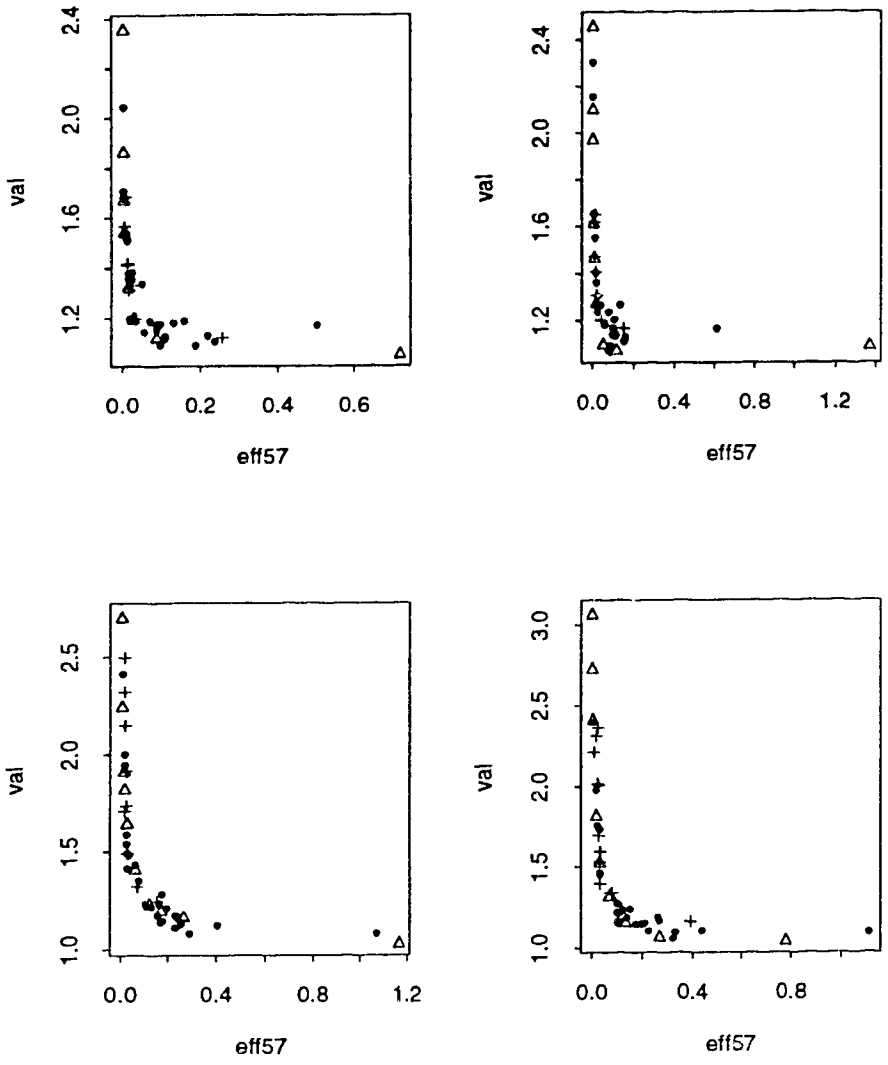


Figure 3.7: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 10$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

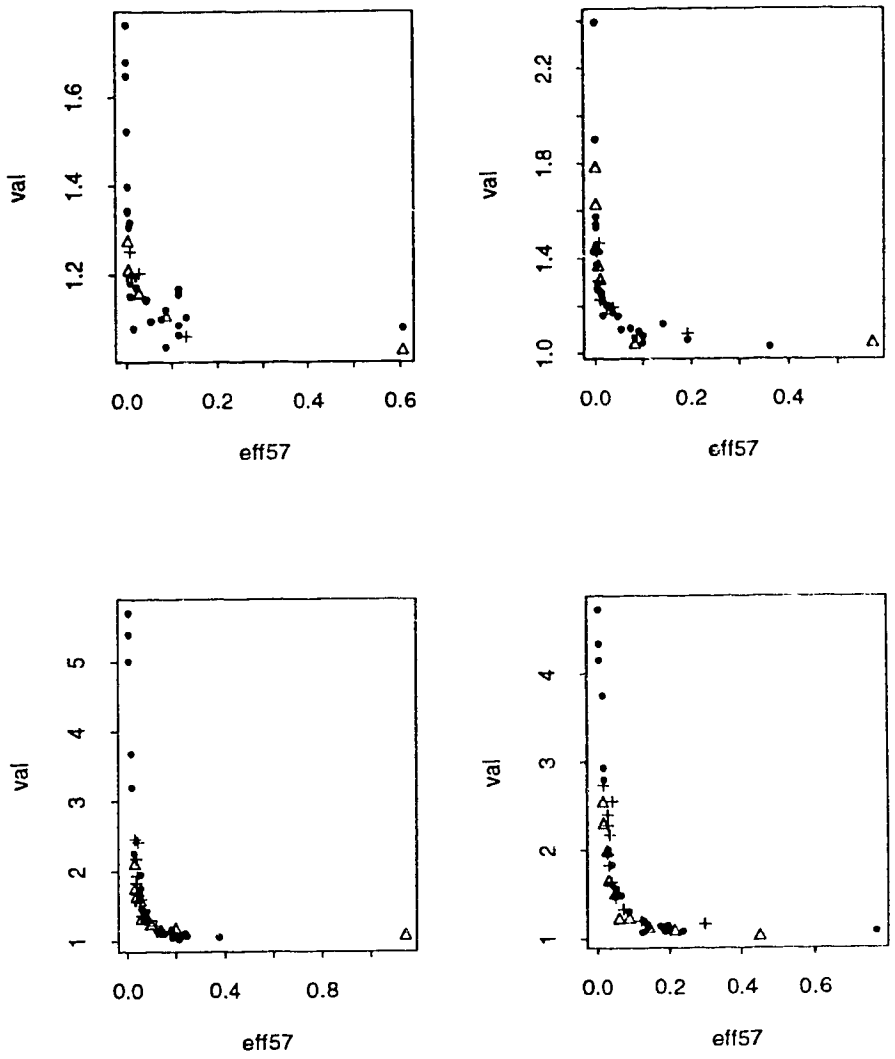


Figure 3.8: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 12$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

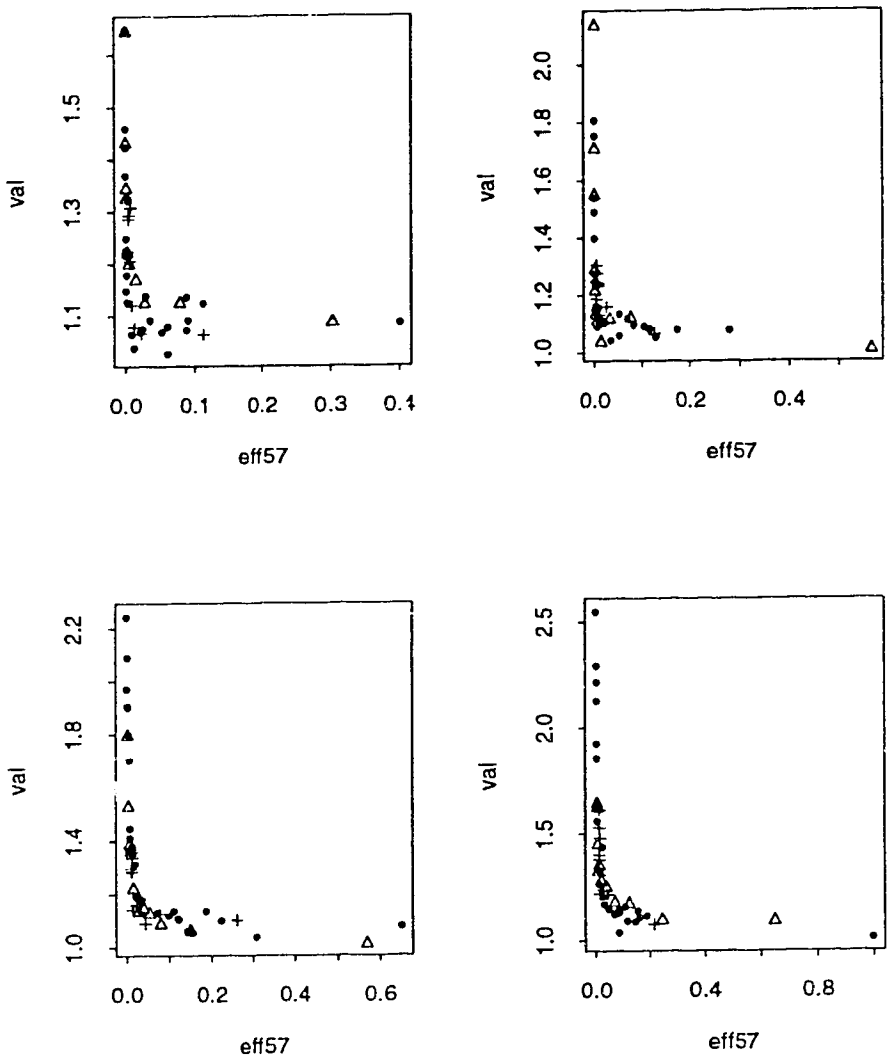


Figure 3.9: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 14$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

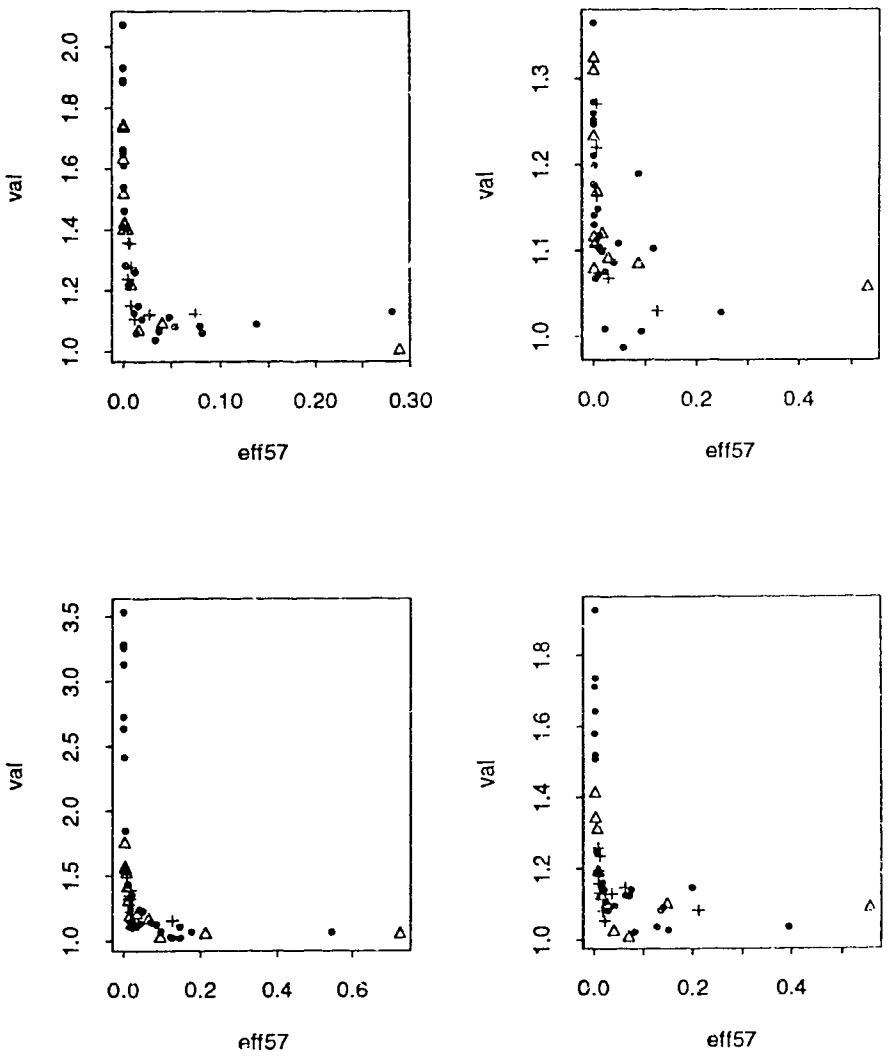


Figure 3.10: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 16$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

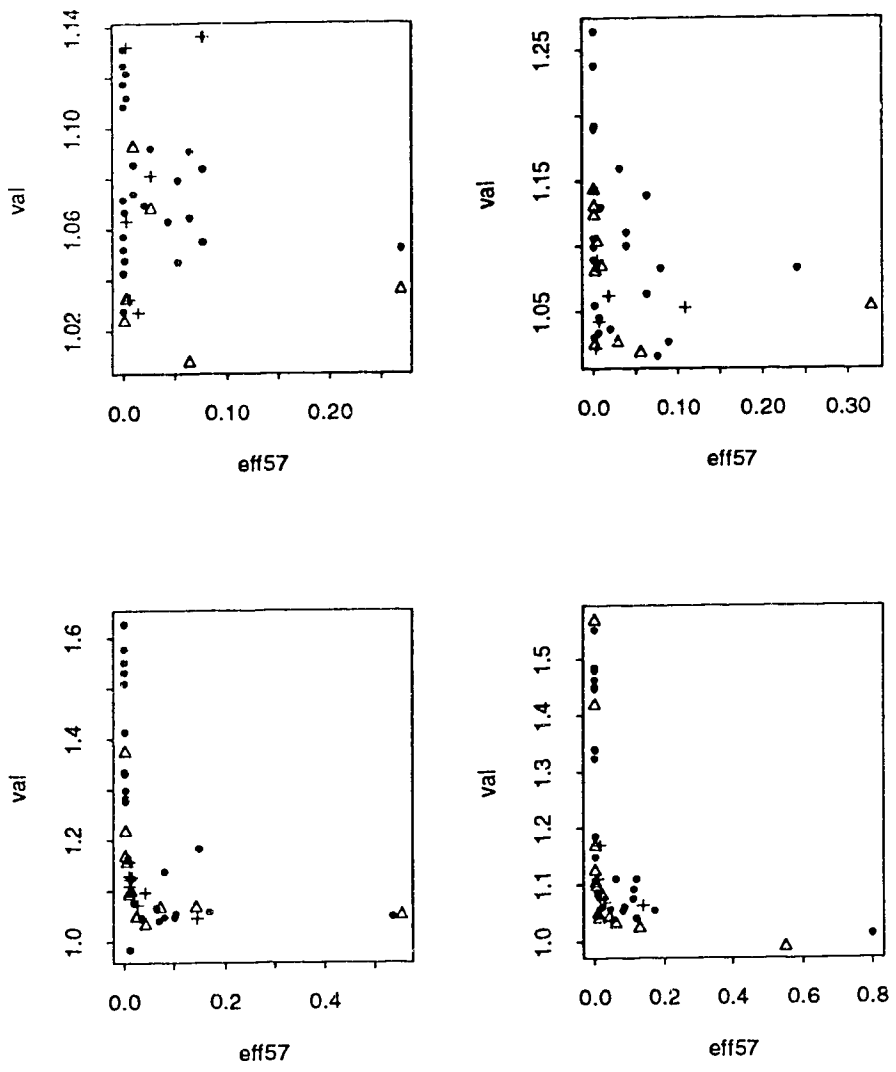


Figure 3.11: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 18$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

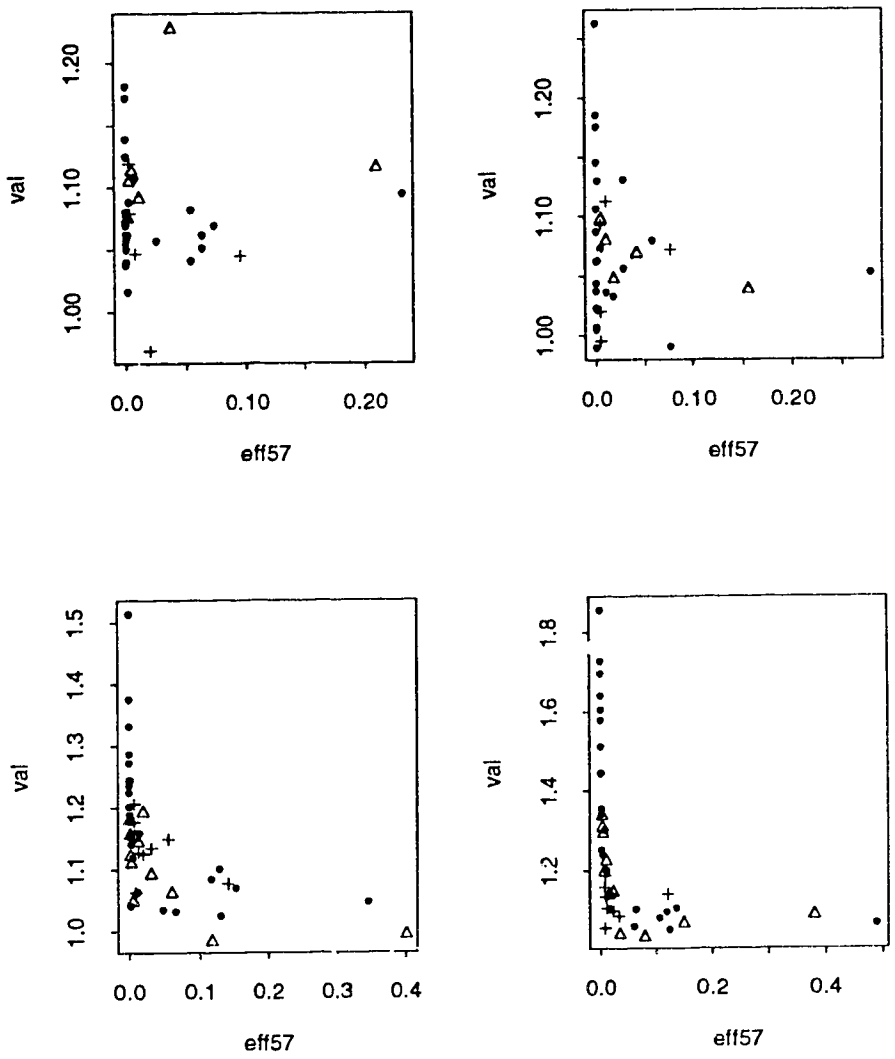


Figure 3.12: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 20$, 2 treatments, one covariate for the first two plots and two covariates for the last two plots

also use the data sets in Bellhouse (1986) for values of covariates. The superimposed plots of $\widehat{\mu 2}_{\text{val}}$ against eff57 for NORD, TCRD and TNORD are displayed in Figures 3.13 to 3.18 for those 24 data sets.

All the plots in Figures 3.13 to 3.18 show a clear curvature. The curves for TCRD and TNORD almost overlap and are below the curve for NORD. Hence, the TCRD and the TNORD provide slightly higher validity than the NORD for analysis of covariance models with more than two treatments. This is to be expected for two reasons. First, we use the 95 percentile of eff(G) to measure efficiency. The TCRD and TNORD both place an upper bound on eff(G), but the NORD does not. The NORD might perform better with respect to a different functional of eff(G), such as the median. Second, the term $\lambda\|A^TGB\|^2$ in the distance function for the NORD is not directly related to the efficiency measure $\det\{(A^TCC^TA)^{-1}\}$ when $q > 1$.

The algorithm for TCRD is very simple and the time required to generate a G from TCRD is much less than that for TNORD if κ is not too small. The time for generating a G from TNORD depends on the annealing algorithm to carry out the minimization of distance function. The algorithm is of the same order n^2 as that for NORD. In Section 3.3, we concluded that the mean time for generating a G from the TCRD in the sequential procedure is a power function of the number of iterations.

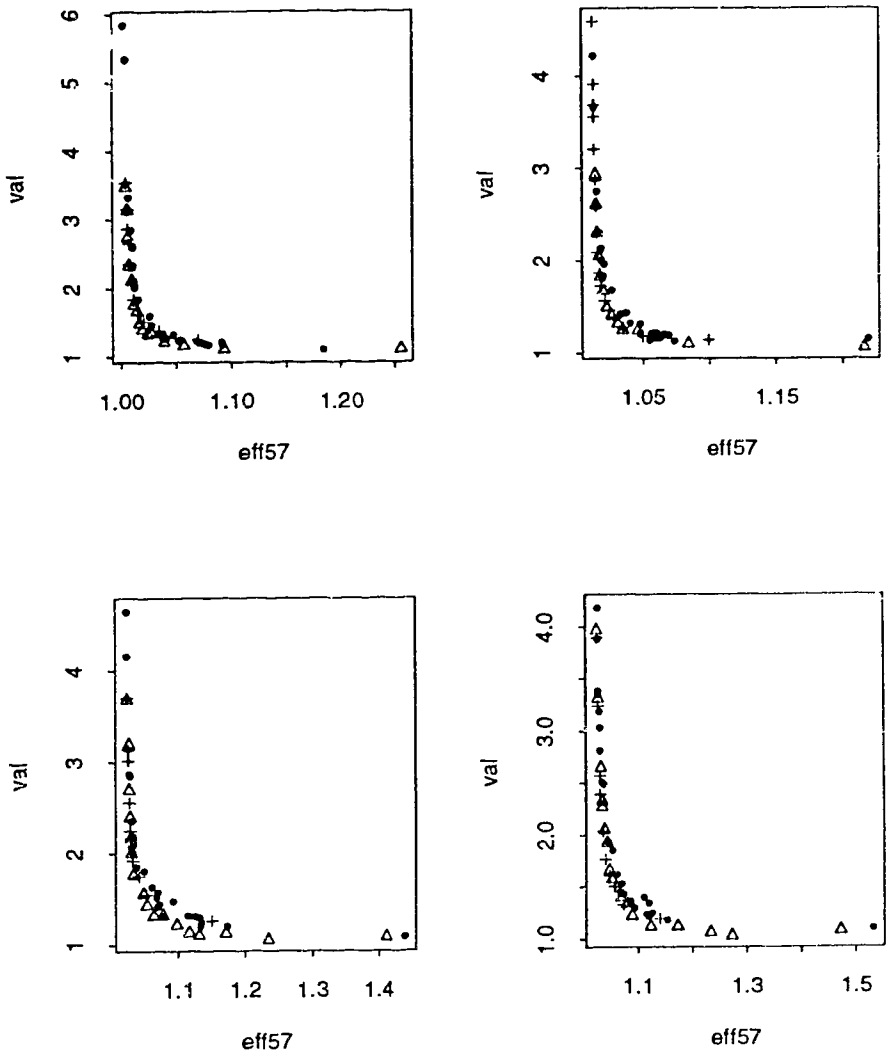


Figure 3.13: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 10$, 5 treatments, one covariate for the first two plots and two covariates for the last two plots

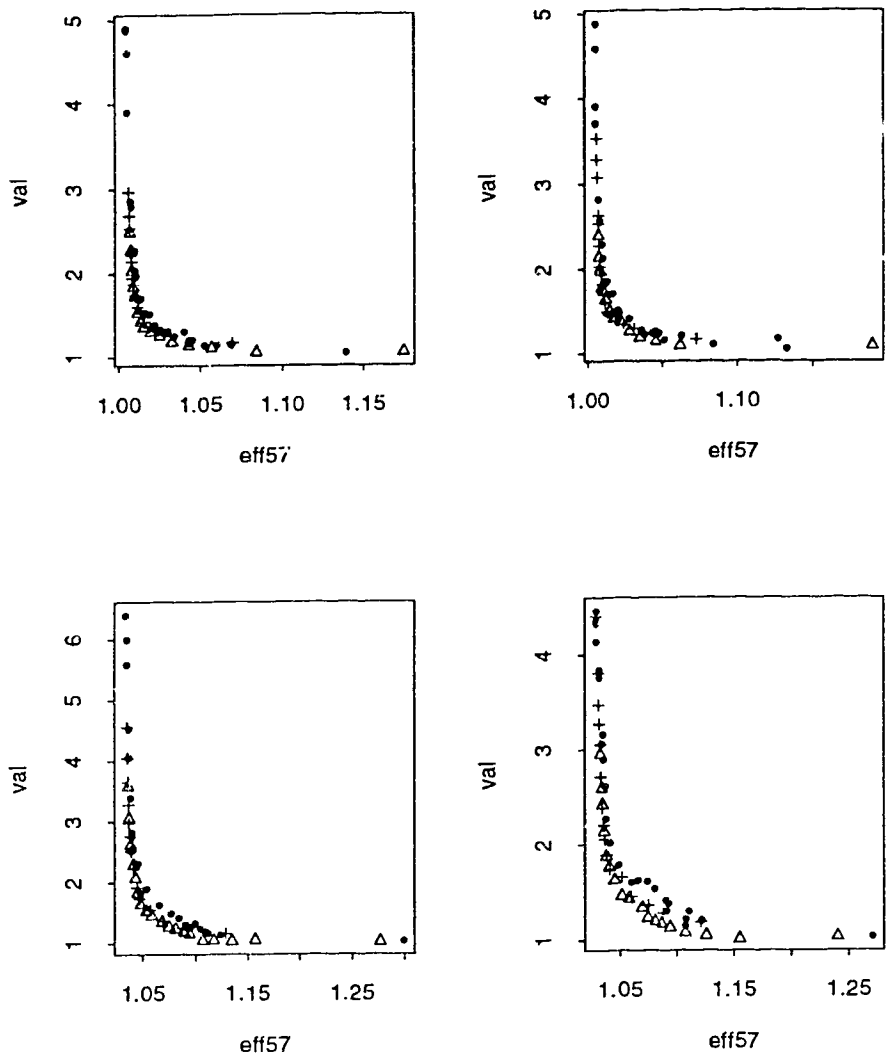


Figure 3.14: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 12$, 6 treatments, one covariate for the first two plots and two covariates for the last two plots

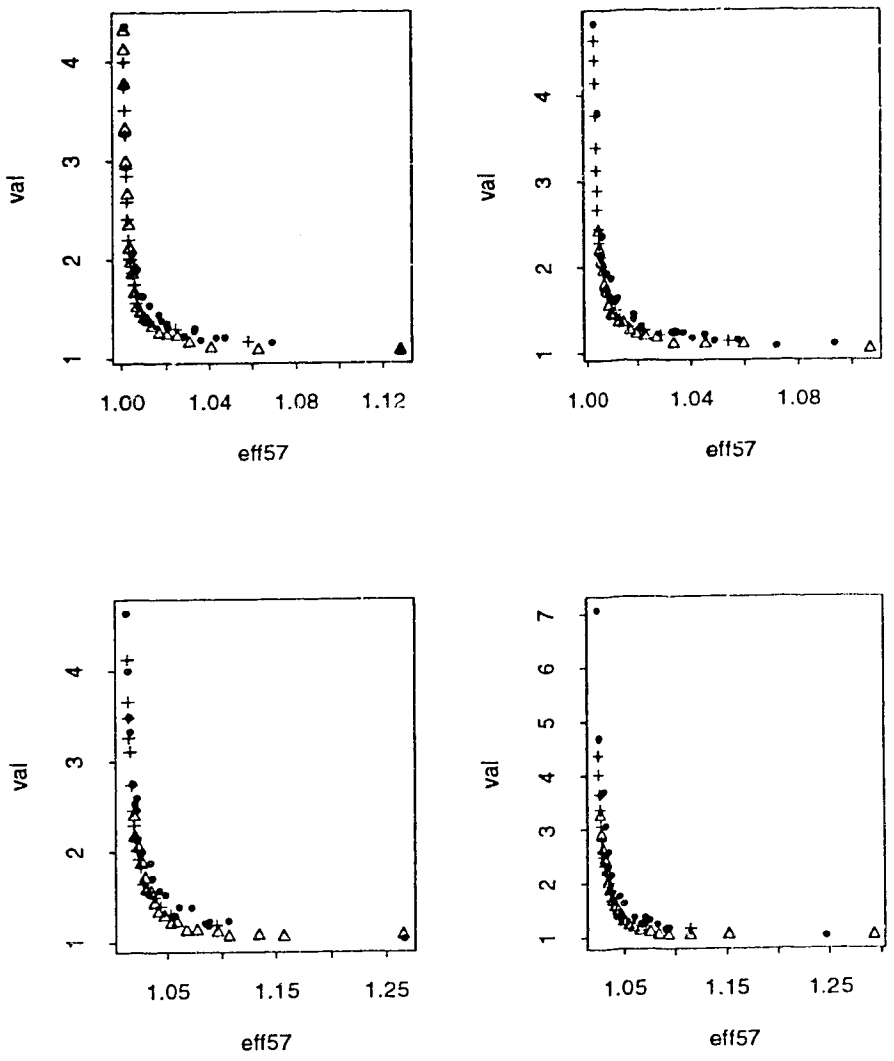


Figure 3.15: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 14$, 7 treatments, one covariate for the first two plots and two covariates for the last two plots

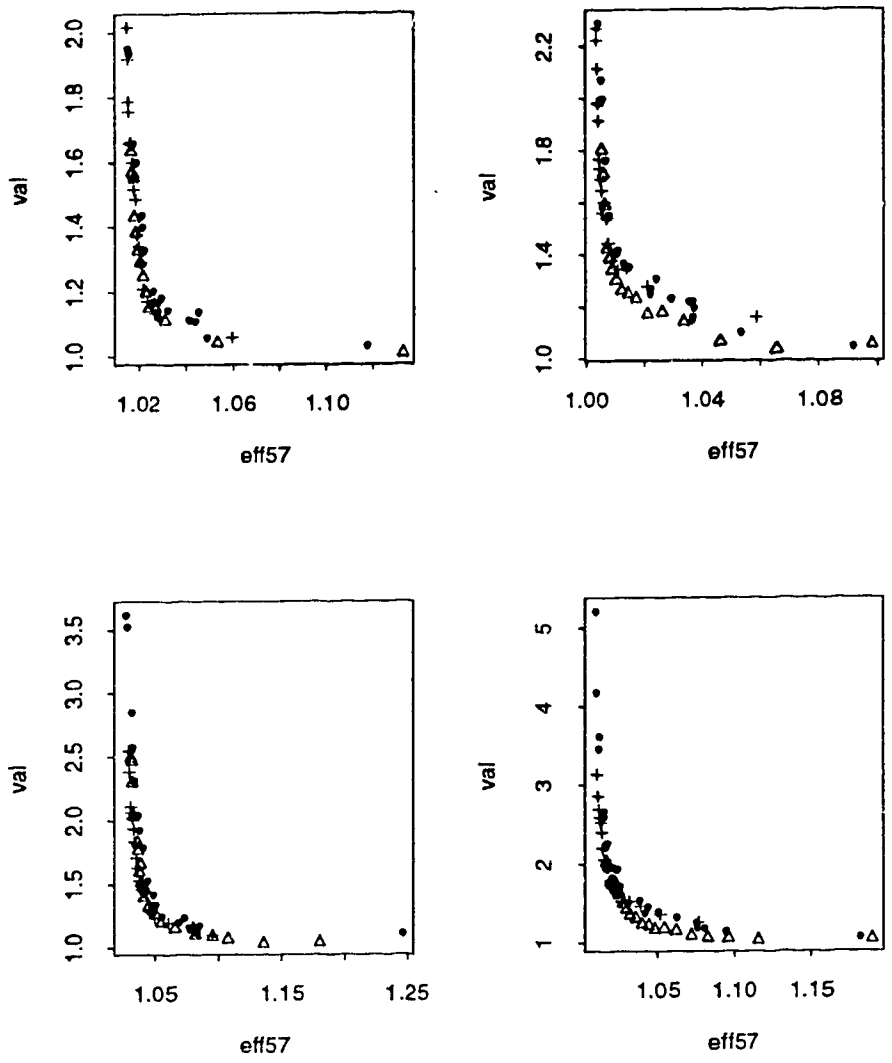


Figure 3.16: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 16$, 8 treatments, one covariate for the first two plots and two covariates for the last two plots

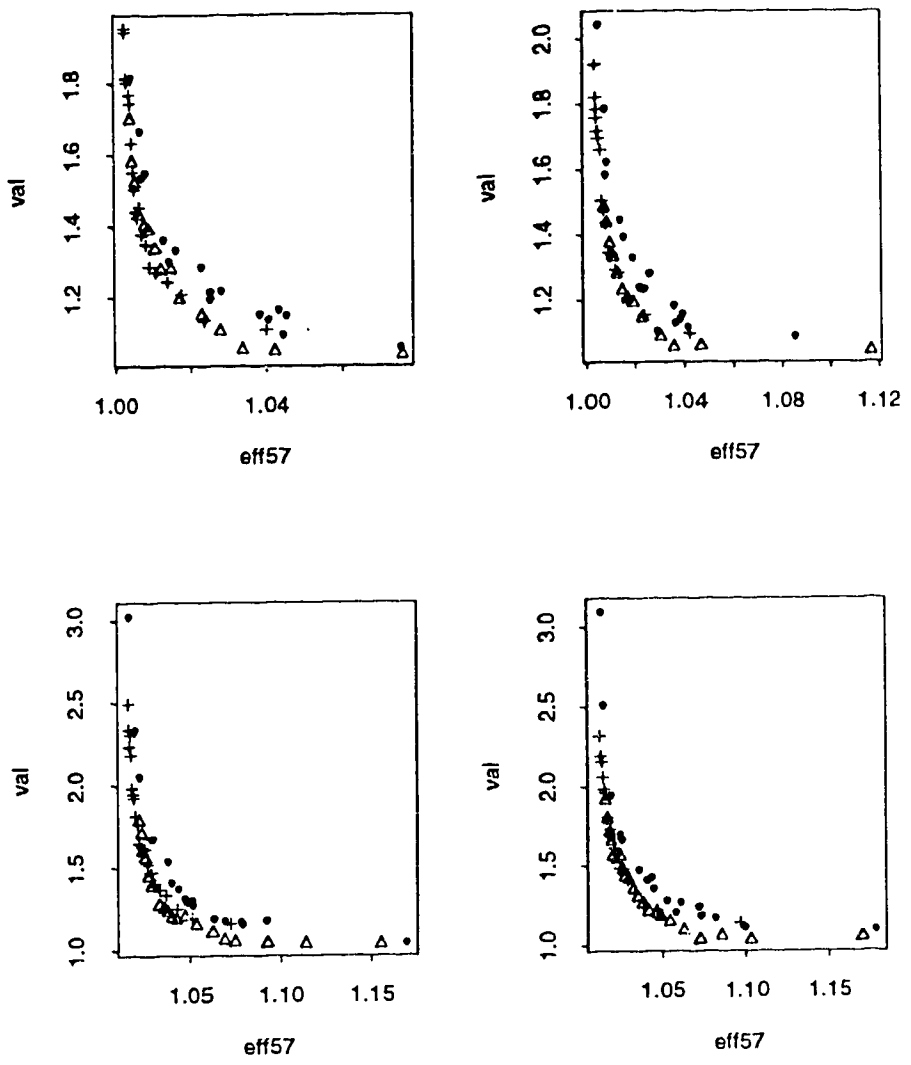


Figure 3.17: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 18$, 9 treatments, one covariate for the first two plots and two covariates for the last two plots

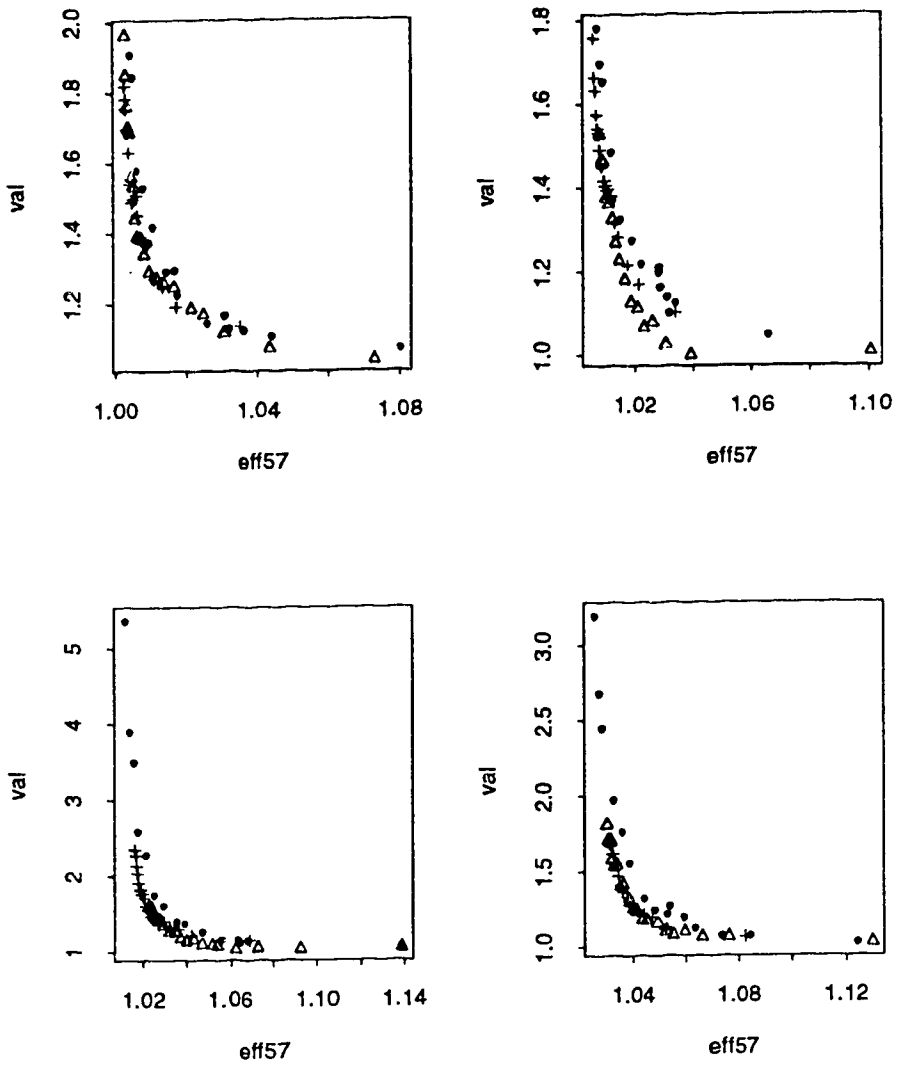


Figure 3.18: Plots of validity vs. efficiency for NORD (·), TCRD (△) and TNORD (+) in analysis of covariance model with $n = 20$, 10 treatments, one covariate for the first two plots and two covariates for the last two plots

Chapter 4

Optimal Randomized Design

4.1 Introduction

In Chapter 3 we discussed several randomized designs intended to achieve acceptable levels of efficiency and validity. The TCRD defines the uniform distribution on the subset with efficiency bounded by a constant. A TNORD is defined to minimize a distance function on that subset. The simulation results in the last chapter revealed that the TCRD and TNORD had nearly the same levels of efficiency and validity, and that both sometimes had a slightly higher level of validity than the NORD with the same level of efficiency. Is it true that the TCRD always has higher level of validity than the NORD? If not, in what situations will the TCRD be better for validity than the NORD? This chapter will give some theoretical results and several examples to clarify this issue.

4.2 Optimal Design for Validity

In this section we assume that the treatment effects matrix A in the general linear model (1.2.5) is orthogonal to the block effects matrix B. Let $\mathcal{G}$ be a subset of $\{G \in \mathcal{P}_n : \|A^T G B\|^2 = 0\}$. In the following discussion, we only consider randomized designs defined on $\mathcal{G}$. A randomized design on $\mathcal{G}$ is defined by its distribution on $\mathcal{G}$. We choose the expectation of the random variable $\|C_1^T G_1 G_2^T C_2\|^2$ as our criterion for

validity, where G_1 and G_2 are independent replicates of G. We have $C_1 = C_2 = A$ here. We will say that a randomized design is optimal for validity if it minimizes

$$E\|A^T G_1 G_2^T A\|^2, \tag{4.2.1}$$

among all randomized designs on $\mathcal{G}$.

In order to minimize (4.2.1), we only need to consider distributions of column spaces $\mathcal{R}(G^T A)$. Let $A_1, \cdots, A_k$ be k matrices whose column spaces list all $\mathcal{R}(G^T A)$ for $G \in \mathcal{G}$. For a randomized design G, put $w_i = P\{\mathcal{R}(G^T A) = \mathcal{R}(A_i)\}$. We then have

$$E\|A^T G_1 G_2^T A\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \|A_i^T A_j\|^2. \tag{4.2.2}$$

It follows from (4.2.2) that the criterion (4.2.1) only depends on the weights w_i on the column spaces $\mathcal{R}(A_1), \ldots, \mathcal{R}(A_k)$, i.e., randomized designs on $\mathcal{G}$ with the same weights w_i must have the same validity. Our goal is to find the weights $w^T = (w_1, w_2, \ldots, w_k)$ that minimize (4.2.2) subject to $\sum_{j=1}^{k} w_j = 1$ and $\min\{w_j\} \geq 0$. The randomized design on $\mathcal{G}$ with those weights is optimal for validity.

We have the following result about the coefficient matrix in (4.2.2).

Lemma 4.2.1 *Let $M = (d_{ij})$, where $d_{ij} = \|A_i^T A_j\|^2$ for $1 \leq i, j \leq k$,(d_{ij} is the square of a similarity measure between column spaces $\mathcal{R}(A_i)$ and $\mathcal{R}(A_j)$). We then have that M is non-negative definite.*

Proof: Note that

$$d_{ij} = \|A_i^T A_j\|^2 = \mathrm{tr}(A_i^T A_j A_j^T A_i) = \mathrm{tr}(A_i A_i^T A_j A_j^T) = \{Vec(A_i A_i^T)\}^T Vec(A_j A_j^T).$$

Letting $\Gamma = [Vec(A_1 A_1^T) : \cdots : Vec(A_k A_k^T)]$, we have $M = \Gamma^T \Gamma$, so M is non-negative definite. ∎

By Lemma 4.2.1, quantity (4.2.2) has a minimum and its minimal point is defined by the following equations:

$$Mw = \lambda \mathbf{1}_k \text{ for some } \lambda \text{ and } w^T \mathbf{1}_k = 1, \tag{4.2.3}$$

provided that the solution satisfies $\min\{w_j\} \geq 0$. Hence, we immediately have the following basic theorem.

Theorem 4.2.2 *The randomized design on $\mathcal{G}$ with the equal weights $w = (\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$ is optimal for validity among all randomized designs on $\mathcal{G}$ under the criterion* (4.2.1) *if $\mathbf{1}_k$ is an eigenvector of M.*

Applying the basic theorem, we obtain two results.

Theorem 4.2.3 *If $\mathcal{G}$ is a subgroup of $\mathcal{P}_n$ then the uniform distribution on $\mathcal{G}$ is optimal for validity among all distributions on $\mathcal{G}$ under the criterion* (4.2.1).

Proof: Since $\mathcal{G}$ is a group, for any pair (i, l) there exists a $G \in \mathcal{G}$ such that $\mathcal{R}(G^T A_i) = \mathcal{R}(A_l)$. Moreover, $\{\mathcal{R}(G^T A_1), \mathcal{R}(G^T A_2), \ldots, \mathcal{R}(G^T A_k)\}$ is a permutation of $\{\mathcal{R}(A_1), \mathcal{R}(A_2), \ldots, \mathcal{R}(A_k)\}$ for any $G \in \mathcal{G}$. Hence, we have

$$\sum_{j=1}^{k} d_{ij} = \sum_{j=1}^{k} \|A_i^T A_j\|^2 = \sum_{j=1}^{k} \|(G^T A_i)^T G^T A_j\|^2 = \sum_{j=1}^{k} \|A_l^T A_j\|^2 = \sum_{j=1}^{k} d_{lj},$$

for any pair (i, l). This means that $\mathbf{1}_k$ is an eigenvector of M. By Theorem 4.2.2, a distribution on $\mathcal{G}$ with the equal weights $w = (\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$ is optimal for validity under the criterion (4.2.1).

Since $\mathcal{G}$ is a group, the subsets $\{G : \mathcal{R}(G^T A) = \mathcal{R}(A_i)\}$ are cosets of the subgroup $\{G : \mathcal{R}(G^T A) = \mathcal{R}(A)\}$ and hence are disjoint and have the same number of elements. Therefore, the uniform distribution on $\mathcal{G}$ has the equal weights and is then optimal for validity. ∎

Theorem 4.2.4 *Suppose $\mathcal{G}_0$ is a subset of $\mathcal{G}$ where $\mathcal{G}_0$ is a group and for each $G \in \mathcal{G}$ there exists $G_0 \in \mathcal{G}_0$ such that $\mathcal{R}(G^T A) = \mathcal{R}(G_0^T A)$. The uniform distribution in $\mathcal{G}_0$ is then optimal for validity among all distributions on $\mathcal{G}$ under criterion* (4.2.1).

Proof: Let G be a random permutation matrix taking values in $\mathcal{G}$. By assumption there exists a random G_0 taking values in $\mathcal{G}_0$ satisfying $\mathcal{R}(G^T A) = \mathcal{R}(G_0^T A)$. Now the distribution of G and the distribution of G_0 have the same validity under criterion (4.2.1). The result follows from Theorem 4.2.3. ∎

For an application of Theorem 4.2.4, consider a block design where, for each treatment, the number of replicates within a block is the same for all blocks. Let $\mathcal{G}_0 = \mathcal{P}_n(B)$, where B represents the block structure, and let $\mathcal{G}$ be a larger set of permutations preserving orthogonality between treatments and blocks; i.e., permutations such that, for each treatment, the number of replicates within a block is the same for all blocks. Note that any such treatment assignment can be represented by a $G_0 \in \mathcal{G}_0$, so the condition of Theorem 4.2.4 is satisfied. The larger set $\mathcal{G}$ may include permutations of units receiving the same treatment, i.e., permutations that do not affect treatment assignments. Theorem 4.2.4 shows that the usual method for randomizing a block design (randomly permute units within blocks and randomly permute blocks) is optimal for validity among all randomization strategies that preserve treatment-block orthogonality.

If the conditions of Theorem 4.2.4 hold and in addition we have $\mathcal{G} \supseteq \mathcal{G}\mathcal{G}_0 \equiv \{GG_0 : G \in \mathcal{G}, G_0 \in \mathcal{G}_0\}$ then the uniform distribution on $\mathcal{G}$ has the same validity as the uniform distribution on $\mathcal{G}_0$. This can be seen by partitioning $\mathcal{G}$ into left cosets of $\mathcal{G}_0$. All cosets contain the same number of elements and all permutations in the same coset determine the same treatment assignment.

If the conditions above fail then the uniform distribution on the subset $\mathcal{G}$ may not be optimal for validity. Some examples follow.

Example 4.2.5 In some circumstances, it does happen that $\mathbf{1}_k$ is not an eigenvector of M. Let

$$B = \text{GramSchmidt}\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -2 \\ 1 & -2 \end{pmatrix}, \qquad A := a = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 2 \\ -2 \end{pmatrix}.$$

Consider $\mathcal{G} = \{G \in \mathcal{P}_6 : a^T GB = 0\}$. There are 30 different column spaces $\mathcal{R}(G^T a)$ for $G \in \mathcal{G}$ and the linear system (4.2.3) results in unequal weights. The 30 column spaces for $\mathcal{R}(G^T a)$ and their associated permutations are listed in Table 4.1. In this thesis, we always refer to G as a permutation matrix. In Section 1.2, we pointed out the one-to-one correspondence between the group of permutations of $\{1, \ldots, n\}$ and the group of $n \times n$ permutation matrices. In Table 4.1, we use a permutation (i.e., an arrangement) of the index set $\{1, \ldots, 6\}$ to denote its corresponding permutation matrix G. The uniform distribution on $\mathcal{G}$ is not optimal for validity under the criterion (4.2.1).

□

Example 4.2.6 In analysis of covariance models, the uniform distribution on $\mathcal{G} = \{G \in \mathcal{P}_n : \|A^T GB\|^2 = 0\}$ may not be optimal for validity among all distributions on $\mathcal{G}$ under the criterion (4.2.1). Let

$$B = \text{GramSchmidt}\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \\ 1 & 0 \\ 1 & 3 \\ 1 & 7 \\ 1 & 2 \end{pmatrix}, \qquad A := a = \frac{1}{2\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix},$$

Table 4.1: 30 column spaces for $\mathcal{R}(G^T a)$ and their associated permutations

$2\sqrt{3}G^T a$	$G(1,2,3,4,5,6)^T$	Optimal Weights
$(1,-1,1,-1,2,-2)^T$	$(1,2,3,4,5,6)^T$	4/144
$(1,-1,1,-1,-2,2)^T$	$(1,2,3,4,6,5)^T$	4/144
$(1,-1,-1,1,2,-2)^T$	$(1,2,4,3,5,6)^T$	4/144
$(1,-1,-1,1,-2,2)^T$	$(1,2,4,3,6,5)^T$	4/144
$(1,1,-1,-1,2,-2)^T$	$(1,3,2,4,5,6)^T$	4/144
$(1,1,-1,-1,-2,2)^T$	$(1,3,2,4,6,5)^T$	4/144
$(1,-1,2,-2,1,-1)^T$	$(1,2,5,6,3,4)^T$	5/144
$(1,-1,-2,2,1,-1)^T$	$(1,2,6,5,3,4)^T$	5/144
$(1,-1,2,-2,-1,1)^T$	$(1,2,5,6,4,3)^T$	5/144
$(1,-1,-2,2,-1,1)^T$	$(1,2,6,5,4,3)^T$	5/144
$(1,2,-1,-2,1,-1)^T$	$(1,5,2,6,3,4)^T$	5/144
$(1,-2,-1,2,1,-1)^T$	$(1,6,2,5,3,4)^T$	5/144
$(1,2,-1,-2,-1,1)^T$	$(1,5,2,6,4,3)^T$	5/144
$(1,-2,-1,2,-1,1)^T$	$(1,6,2,5,4,3)^T$	5/144
$(1,2,-2,-1,1,-1)^T$	$(1,5,6,2,3,4)^T$	5/144
$(1,-2,2,-1,1,-1)^T$	$(1,6,5,2,3,4)^T$	5/144
$(1,2,-2,-1,-1,1)^T$	$(1,5,6,2,4,3)^T$	5/144
$(1,-2,2,-1,-1,1)^T$	$(1,6,5,2,4,3)^T$	5/144
$(2,1,-1,-2,1,-1)^T$	$(5,1,2,6,3,4)^T$	5/144
$(-2,1,-1,2,1,-1)^T$	$(6,1,2,5,3,4)^T$	5/144
$(2,1,-1,-2,-1,1)^T$	$(5,1,2,6,4,3)^T$	5/144
$(-2,1,-1,2,-1,1)^T$	$(6,1,2,5,4,3)^T$	5/144
$(2,1,-2,-1,1,-1)^T$	$(5,1,6,2,3,4)^T$	5/144
$(-2,1,2,-1,1,-1)^T$	$(6,1,5,2,3,4)^T$	5/144
$(2,1,-2,-1,-1,1)^T$	$(5,1,6,2,4,3)^T$	5/144
$(-2,1,2,-1,-1,1)^T$	$(6,1,5,2,4,3)^T$	5/144
$(2,-2,1,-1,1,-1)^T$	$(5,6,1,2,3,4)^T$	5/144
$(-2,2,1,-1,1,-1)^T$	$(6,5,1,2,3,4)^T$	5/144
$(2,-2,1,-1,-1,1)^T$	$(5,6,1,2,4,3)^T$	5/144
$(-2,2,1,-1,-1,1)^T$	$(6,5,1,2,4,3)^T$	5/144

and $\mathcal{G} = \{G : a^T GB = 0\}$. Define

$$[a : a_1 : a_2 : a_3] = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}.$$

It is easy to show that $\mathcal{G} = \{G : a^T GB = 0\} = \bigcup_{i=0}^{3} \mathcal{G}_i$, where $\mathcal{G}_0 = \{G : G^T a = \pm a\}$ and $\mathcal{G}_i = \{G : G^T a = \pm a_i\}, i = 1, 2, 3$. The coefficient matrix in linear system (4.2.3) is

$$M = \begin{pmatrix} 1 & 0 & 0 & .25 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ .25 & 0 & 0 & 1 \end{pmatrix},$$

and the optimal weight is $w^T = (\frac{4}{18}, \frac{5}{18}, \frac{5}{18}, \frac{4}{18})$. The TCRD with $\kappa = 0$ (the uniform distribution on $\mathcal{G}$) is not optimal for validity among all distributions on $\mathcal{G}$. In fact, by (4.2.2) the design with the optimal weight and the TCRD with $\kappa = 0$ have validity of .27778 and .28125 respectively.

□

Remarks 4.2.7 If A is no longer assumed to be orthogonal to B and $\mathcal{G}$ is an arbitrary subset of $\mathcal{P}_n$, we may still consider the problem of finding an optimal distribution on $\mathcal{G}$ under the criterion $E\|C_1^T G_1 G_2^T C_2\|^2$. A randomized design may be defined to be optimal for validity if it minimizes

$$E\|C_1^T G_1 G_2^T C_2\|^2, \tag{4.2.4}$$

among all randomized designs on $\mathcal{G}$. We can also consider distributions of column spaces $\mathcal{R}(G^T C)$ and let $A_1, \cdots, A_k$ be k matrices whose column spaces list all $\mathcal{R}(G^T C)$

for $G \in \mathcal{G}$. For a randomized design G, put $w_i = P\{\mathcal{R}(G^TC) = \mathcal{R}(A_i)\}$. We again have

$$E\|C_1^T G_1 G_2^T C_2\|^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j \|A_i^T A_j\|^2. \tag{4.2.5}$$

By the same argument as above we can obtain a similar sufficient and necessary condition for the uniform distribution on $\mathcal{G}$ to be optimal for validity. If $\mathcal{G}$ is a group, then the uniform distribution on $\mathcal{G}$ has optimal validity among all distributions on $\mathcal{G}$. If $\mathcal{G} = \{G \in \mathcal{P}_n : \text{eff}(G) \leq \kappa\}$ then it is not true in general that the uniform distribution on $\mathcal{G}$ is optimal for validity under the criterion (4.2.4). Optimal validity is difficult to investigate analytically when C depends on G.

□

Remark 4.2.8 More generally, we may define that a randomized design is optimal for validity if it minimizes

$$E\|C_1^T G_1 G_2^T C_2\|^{2i}, \tag{4.2.6}$$

where i is a positive integer. We can make the same discussions and obtain the same results as in Remarks 4.2.7.

□

4.3 Some Relationships Among NORDs, TCRDs and TNORDs in Special Cases

In general, the NORD with $\lambda = 0$ is equal to the TNORD with $\kappa = \infty$. Usually, the TCRD with $\kappa = 0$ is not equal to the TNORD with $\kappa = 0$. The first example gives a situation where the NORD with $\lambda = 0$ is equal to the TCRD with $\kappa = 0$ and the TNORD with $\kappa = 0$.

Example 4.3.1 Consider an experiment to compare two treatments in k blocks with each treatment applied once in each block. Representing the data set by the additive linear model (1.2.5), we have

$$Y = a_0\beta_0 + a\beta + GB\gamma + GKV,$$

where

$$a_0 = \frac{1}{\sqrt{2k}}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}, \quad a = \frac{1}{\sqrt{2k}}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \\ 1 \\ -1 \end{pmatrix}, \quad \text{and} \quad B = \text{GramSchmidt}\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

In this experiment, we have $n = 2k, q_0 = 1, q = 1$, and $r = k$. Let $[B : d] = \text{GramSchmidt}[B : z]$, where z is a $N(0, I_{2k})$ random vector. Thus d is uniformly distributed with $d^T d = 1$ over the column space $\mathcal{R}(B)^\perp$ and is of the form

$$d = \begin{pmatrix} d_1 \\ -d_1 \\ d_2 \\ -d_2 \\ \vdots \\ d_k \\ -d_k \end{pmatrix}.$$

The random variables $\text{sgn}(d_i), i = 1, \ldots, k$, are independent and uniformly distributed on $\{-1, 1\}$.

Consider the NORD generated by minimizing the distance function:

$$\begin{aligned} \text{dis}(G) &= 1 - (a^T G d)^2 \\ &= 1 - (a_{G1}d_1 - a_{G2}d_1 + \cdots + a_{G(n-1)}d_k - a_{Gn}d_k)^2. \end{aligned} \tag{4.3.1}$$

For given $d_1, \ldots, d_k$, the distance function (4.3.1) is minimized when $a_{G(2i-1)}$ and $a_{G(2i)}$ have different signs for $i = 1, \ldots, k$ and $a_{G(2i-1)}d_i - a_{G(2i)}d_i, i = 1, \ldots, k$, have

the same sign. Thus, this G must be in $\mathcal{G} = \{G \in \mathcal{P}_n : a^T GB = 0\}$. Moreover, the minimization of (4.3.1) only depends on the signs of $d_1, \ldots, d_k$. The NORD defined by minimizing (4.3.1), the TNORD defined by minimizing (4.3.1) over $\mathcal{G}$, and the TCRD with $\kappa = 0$ (the uniform distribution on $\mathcal{G}$) all determine the same distribution of the column space $\mathcal{R}(G^T a)$; i.e., all determine the same distribution of treatment assignments.

□

In the NORD and TNORD algorithms, the matrix D is uniformly distributed in the space $\mathcal{R}(B)^{\perp}$. The NORD and TNORD algorithms are trying to select design G such that $\mathcal{R}(G^T A)$ is close to $\mathcal{R}(D)$. In Example 1, the TNORD with $\kappa = 0$ is equivalent to the TCRD with $\kappa = 0$, i.e., the TNORD algorithm with $\kappa = 0$ defines the same distribution on $\mathcal{P}_n$ as the TCRD with $\kappa = 0$. In the next examples, we consider the TNORD and the TCRD both with $\kappa = 0$. In general, the TNORD algorithm defines a different distribution on $\{G : \|A^T GB\|^2 = 0\}$ than does the TCRD. The question is in what situations are the two equivalent, i.e., when does the TNORD algorithm result in the uniform distribution on $\{G : \|A^T GB\|^2 = 0\}$. The following are several examples to investigate this question. Example 4.3.2 describes a situation where the two methods lead to the same result. Example 4.3.3 describes a situation where the two methods produce different results and the TNORD has higher validity than the TCRD.

Example 4.3.2 Let

$$a = \sqrt{\frac{1}{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \\ 0 \end{pmatrix}, \quad B = [\frac{1}{2}\mathbf{1}_4 : b] = \begin{pmatrix} \frac{1}{2} & -\frac{3}{2\sqrt{5}} \\ \frac{1}{2} & -\frac{1}{2\sqrt{5}} \\ \frac{1}{2} & \frac{1}{2\sqrt{5}} \\ \frac{1}{2} & \frac{3}{2\sqrt{5}} \end{pmatrix} = \text{GramSchmidt} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

It is easy to check that $[\frac{1}{2}\mathbf{1}_4 : b : a : c]$ is an orthonormal basis of $\mathcal{R}^4$, where

$$c = \sqrt{\frac{6}{5}} \begin{pmatrix} -\frac{1}{3} \\ \frac{1}{6} \\ \frac{2}{3} \\ -\frac{1}{2} \end{pmatrix}.$$

Let vector d be uniformly distributed on $\{d : a^T B = 0, d^T d = 1\}$. Then

$$d = d_1 a + d_2 c,$$

with $d_1^2 + d_2^2 = 1$.

Denoting a permutation by G, we have

$$\mathcal{P}_4(B) = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \right\}.$$

Let

$$\mathcal{P}_4(a) = \mathcal{G}_1 = \{G : G^T a = \pm a\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \\ 4 \end{pmatrix} \right\}.$$

We find

$$\mathcal{G} = \{G : a^T G B = 0\} = \{G_1, G_2, G_3, G_4\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} \right\}.$$

To generate TNORD with $\kappa = 0$, we need to minimize over $\mathcal{G}$

$$\mathrm{dis}(G) = 1 - (d_1 a^T G a + d_2 a^T G c)^2.$$

For G_1 and G_2, $1 - \mathrm{dis}(G) = d_1^2$; for G_3 and G_3, $1 - \mathrm{dis}(G) = (-\frac{2}{3} d_1 + \frac{\sqrt{5}}{3} d_2)^2$. By a simple symmetry argument or direct calculation we obtain

$$P\{d_1^2 \leq (-\frac{2}{3} d_1 + \frac{\sqrt{5}}{3} d_2)^2\} = P\{d_1^2 \geq (-\frac{2}{3} d_1 + \frac{\sqrt{5}}{3} d_2)^2\} = \frac{1}{2}.$$

Hence, the TNORD with $\kappa = 0$ is defined by the uniform distribution on $\mathcal{G}$, i.e., the TNORD with $\kappa = 0$ is equivalent to the TCRD with $\kappa = 0$.

□

Example 4.3.3 (continued Example 4.2.6). Let the columns of the matrix

$$[B : a : c_1 : c_2 : c_3 : c_4 : c_5] = \text{GramSchmidt}[B : a : a_1 : a_2 : a_3 : a_4 : a_5]$$

constitute an orthonormalized basis of $\mathcal{R}^8$, where

$$[a_4 : a_5] = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ -1 & -1 \\ 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

The TNORD with $\kappa = 0$ is defined by minimizing the following distance function

$$\text{dis}(G) = 1 - (d_1 a^T G a + d_2 a^T G c_1 + d_3 a^T G c_2 + d_4 a^T G c_3 + d_5 a^T G c_4 + d_6 a^T G c_5)^2 \quad (4.3.2)$$

on $\mathcal{G}$, where $(d_1, d_2, \ldots, d_6)^T$ is uniformly distributed with constraint $\sum_{i=1}^{6} d_i^2 = 1$. In order to estimate the weights in four subsets $\mathcal{G}_i (i = 0, 1, 2, 3)$ of $\mathcal{G}$ for the TNORD, we generated 1000 random vectors each uniformly distributed with constraint $\sum_{i=1}^{6} d_i^2 = 1$. The estimated weight for the TNORD is $w^T = (.239, .270, .263, .228)$, which is closer to the optimal weight $(.2222, .2778, .2778, .2222)$ than the equal weight (i.e., the TCRD with $\kappa = 0$). By (4.2.2) the TNORD has an estimated validity of .27842 and has slightly higher validity than the TCRD.

□

Chapter 5

Conclusion and Future Research

5.1 Summary and Conclusion

There are several purposes for randomization in experimental design. An important purpose is its validation of the usual normal theory analysis. In this thesis, we have investigated how randomization validates the usual statistical analysis for treatment effects.

In Chapter 2, we discussed the validity measures $\mathrm{cv}(\phi)$ in Hooper (1989) and $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ in Hooper (1993) with regard to interpretation and computation. The measure $\mathrm{cv}(\phi)$ describes the variation in the conditional significance level $E\{\phi(G,W) \mid W\}$ about its mean $E\{\phi(G,W)\}$ and can be used to compare randomized designs of different sizes. To estimate $\mathrm{cv}(\phi)$ we need to generate replicates of both G and W. The measure $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ was proposed in Hooper (1993) to assist in the selection of a tuning constant for NORD. It is appropriate for comparison of different randomized designs under a given model, for example, for given (n,q,r) under the general linear model (1.2.5). In this thesis, we suggested a single valued measure $\mu 2_{\mathrm{val}}$ based on $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ and used it to assist in the selection of a truncation constant for TCRD and TNORD. Compared with $\mathrm{cv}(\phi)$, $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ is easier to estimate since only replicates of G are required. In order to seek a compromise between computation and

interpretation, we introduced the Beta-related validity measure $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$, i.e., the coefficients of variation of the first two randomization moments of the Beta statistic. This measure was motivated by comparing the first two randomization moments of the Beta statistic with corresponding normal-theory moments to evaluate agreement between the randomization probability $E\{\phi(G,W) \mid W\}$ and the model-assisted probability $E\{\phi(G,W)\}$. The Beta-related measure allows an indirect approximate interpretation. We derived a relation between $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ and $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$, so the Beta-related measure can be estimated at the same cost as $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$. In the last part of Chapter 2, we presented a simulation study showing that the addition of one or two covariates reduces little validity using $\mathrm{cv}(\phi)$ and $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$. Another simulation study showed that $\mathrm{cv}(\phi)$ varies monotonically with μ_{val}, σ_{val} and $\mu 2_{\mathrm{val}}$, which suggests that $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ or the single valued measure $\mu 2_{\mathrm{val}}$ can be used as a surrogate for $\mathrm{cv}(\phi)$ in determining an appropriate value for the tuning constant λ in NORD and the truncation constant κ in TCRD and TNORD.

In Chapter 3, we investigated methods of randomization for asymmetric models of the unit effects. Cox (1982) and Bellhouse (1986) considered the efficient and valid use of covariate information in design of experiments and recommended truncating classical methods of randomization to ensure high levels of efficiency while maintaining validity. This idea can be applied to general linear models for treatment and unit effects. We called such a randomization strategy a truncated classical randomized design (TCRD). A sequential procedure was proposed for generating a sequence of TCRDs to assist in the selection of an appropriate truncation constant κ. A simulation study was presented to illustrate the use of this sequential procedure under analysis of covariance models. Under the general linear model (1.2.5), Hooper (1993) proposed NORDs to achieve acceptable levels of validity and efficiency. In order to improve NORD with respect to efficiency and the selection of tuning constant we introduced the truncated NORD (TNORD). The TNORD is invariant under reparameterization of the treatment effects model and can be generated by the same

sequential procedure as the TCRD. We derived some formulae for the minimization of the distance function for the TNORD. A simulation study was presented showing that the same levels of validity and efficiency can be achieved by suitable choice of the tuning or truncation constant in these three methods.

In Chapter 4, we introduced the concept of optimal randomized design. A randomized design is defined to be optimal for validity if it minimizes a validity criterion. A general sufficient condition was obtained for an uniform distribution in a subset of $\mathcal{P}_n$ to be optimal for validity. Using the general sufficient condition, we obtained that some classical randomized designs, such as CRD and randomized blocks design, are optimal for validity among all distributions on a certain subset of $\mathcal{P}_n$. Under analysis of covariance models, however, the uniform distribution on a certain subset of $\mathcal{P}_n$ may not be optimal for validity. Therefore, under asymmetric models for unit effects, NORD or TNORD may have higher validity than TCRD in some circumstances, although TCRD is more easily generated.

5.2 Future Research

The Beta-related validity measure $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$ was proposed in Chapter 2 and was used to determine $\mathrm{cv}(\phi)$. A question of interest is whether $\mathrm{cv}(\phi)$ can be approximated by $[\mathrm{cv}(T), \mathrm{cv}(T^2)]$, i.e., what is the relationship between the two measures, at least for the F test. This is a problem for future research.

We made a first attempt by developing an analytical approximation to $\mathrm{cv}(\phi)$ in Chapter 2. By approximating $\mathrm{cv}(\phi)$ by $\mathrm{cv}(\phi_\chi)$ where ϕ_χ is the χ^2-test, we derived a Laguerre series expansion for $\mathrm{cv}(\phi)$, with leading terms related to $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$. Randomization does not provide the same level of validity for the χ^2-test as the F test, however. It is more useful to develop a series expansion for $\mathrm{cv}(\phi)$ without approximating $\mathrm{cv}(\phi)$ by $\mathrm{cv}(\phi_\chi)$ such that the series has some relation to $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$.

A sequential procedure was proposed to assist in the selection of an appropriate truncation constant for TCRD. The theoretical aspects of this procedure have not been fully developed. Some theory concerning order statistics for discrete distribution is needed for this development.

In Chapter 4, we developed some theory for optimal randomized designs under the validity criterion $E\|C_1^T G_1 G_2^T G_2\|^2$ with G restricted to a particular subset of $\mathcal{P}_n$. It is helpful to study the theory of optimal randomized designs under other validity criteria in more general subsets of $\mathcal{P}_n$. We also showed in this thesis that the CRD and randomized blocks design are optimal for validity among all randomized designs on a certain subset of $\mathcal{P}_n$. One may ask how to characterize the optimality of the Latin square design for validity in the above sense. This is another problem for future research.

Bibliography

[1] Bailey, R. A. (1986) Randomization, constrained. In *Encyclopedia of Statistical Science* (eds S. Kotz and N. L. Johnson), vol. 7, pp. 524-530. New York: Wiley.

[2] Bailey, R. A. and Rowley, C. A. (1987) Valid Randomization. *Proc. R. Soc. Lond.* A, **410**, 105-124.

[3] Bellhouse, D. R. (1986) Randomization in the analysis of covariance. *Biometrika*, **73**, 207-211.

[4] Box, G. E. P. and Watson, G. S. (1962) Robustness to non-normality of regression tests. *Biometrika*, **49**, 93-106.

[5] Cox, D. R. (1982) Randomization and concomitant variables in the design of experiments. In *Statistics and Probability: Essays in Honor of C. R. Rao* (eds G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh), pp. 197-202. Amsterdam: North-Holland.

[6] Davis, A. W. and Speed, T. P. (1988) An Edgeworth expansion for the distribution of the F-ratio under a randomization model for the randomized block design. In *Statistical Decision Theory and Related Topics IV* (eds S. S. Gupta and J. O. Berger), vol. 2, pp. 119-130. New York: Springer-Verlag.

[7] Eden, T. and Yates, F. (1933) On the validity of Fisher's z test when applied to an actual sample of non-normal data. *Journal of Agricultural Science*, **23**, 8-17.

[8] Fisher, R. A. (1925) *Statistical Methods for Research Workers.* Edinburgh: Oliver & Boyd.

[9] Fisher, R. A. (1935) *The Design of Experiments.* Edinburgh: Oliver & Boyd.

[10] Grundy, P. M. and Healy, M. J. R. (1950) Restricted randomization and quasi-Latin squares. *J. R. Statist. Soc.* B, **12**, 286-291.

[11] Hille, E. (1926) On Laguerre's series: first note. *Proc. Nat. Acad. Sci.*, **12**, 261-265.

[12] Hoeffding, W. (1952) The large sample power of tests based on permutations of the observations. *Ann. Math. Statist.*, **23**, 169-192.

[13] Hooper, P. M. (1989) Experimental randomization and validity of normal-theory inference. *J. Am. Statist. Ass.*, **84**, 576-586.

[14] Hooper, P. M. (1993) Nearly orthogonal randomized designs. *J. R. Statist. Soc.* B, **55**, 221-236.

[15] Kempthorne, O. and Doerfler, T. E. (1969) The behavior of some significance tests under experimental randomization. *Biometrika*, **56**, 231-247.

[16] Khatri, C. G., Krishnaiah, P. R. and Sen, P. K. (1977) A note on the joint distribution of correlated quadratic forms. *Journal of Statistical Planning and Inference*, **1**, 299-307.

[17] Kibble, W. F. (1941) A two-variate Gamma-type distribution. *Sankhyā*, **5**, 137-150.

[18] Krishnamoorthy, A. S. and Parthasarathy, M. (1951) A multivariate gamma distribution. *Ann. Math. Statist.*, **22**, 549-557. Erratum: (1960) *Ann. Math. Siatist.*, **31**, 229.

[19] Jensen, D. R. (1970) The joint distribution of traces of Wishart matrices and some applications. *Ann. Math. Statist.*, **41**, 133-145.

[20] Lee, A. J. (1990) *U-Statistics: Theory and Practice*. New York: Marcel Dekker.

[21] Lehmann, E. L. (1983) *Theory of Point Estimation*. New York: Wiley.

[22] Magnus, J. R. and Neudecker, H. (1979) The commutation matrix: some properties and applications. *Ann. Statist.*, **7**, 381-394.

[23] Nelder, J. A. (1965a) The analysis of randomized experiments with orthogonal block structure. I: Block structure and the null analysis of variance. *Proc. Roy. Soc. Lond.* A, **283**, 147-162.

[24] Nelder, J. A. (1965b) The analysis of randomized experiments with orthogonal block structure. II: Treatment structure and the general analysis of variance. *Proc. Roy. Soc. Lond.* A, **283**, 163-178.

[25] Pitman, E. J. G. (1938) Significance tests which may be applied to samples from any population, III: The analysis of variance test. *Biometrika*, **29**, 322-335.

[26] Preece, D. A., Bailey, R. A. and Patterson, H. D. (1978) A randomization problem in forming designs with superimposed treatments. *The Australian Journal of Statistics*, **20**, 111-125.

[27] Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986) *Numerical Recipes: the Art of Scientific Computing*. Cambridge University Press.

[28] Rao, C. R. (1973) *Linear Statistical Inference and its Applications* (2th ed.). New York: Wiley.

[29] Rao, C. R. and Kleffe, J. (1988) *Estimation of Variance Components and Applications*. New York: North Holland.

[30] Reiss, R. D. (1989) *Approximate Distributions of Order Statistics.* New York: Springer-Verlag.

[31] Robinson, J. (1973) The large sample power of permutation tests for randomization models. *Ann. Statist.*, **1**, 291-296.

[32] Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

[33] Szegö, G (1975) *Orthogonal Polynomials* (4th ed.). Am. Math. Soc., Providence, Rhode Island.

[34] Welch, B. L. (1937) On the z-test in randomized blocks and Latin squares. *Biometrika*, **29**, 21-52.

[35] Wilkinson, G. N., Eckert, S. R., Hancock, T. W. and Mayo, O (1983) Nearest neighbour (NN) analysis of field experiments (with discussion). *J. R. Statist. Soc.* B, **45**, 151-211.

[36] Yates, F. (1933) The information of Latin squares for use in field experiments. *Empire J. Exp. Agric.*, **1**, 235-244.