ASUU JOURNAL OF SCIENCE
A Journal of Research and Development
Vol 5 Nos 1 & 2, December, 2018; pp.112-135

Relative Efficiency of OLS and GLS Estimators in Split-Plot Design for Second-Order Response Surface Model Under Equivalence Conditions

[1]Saleh, I. M. and [2]Akeyede, I.

Abstract

In many industrial experiments, applications of completely randomised designs (CRDs) are often not practicable due to the presence of one or more hard-to-change factors which require restrictions on randomisation. In situations like these, split-plot experiments are practicable. The split-plot experiments have two separate randomisations which make the error structure different from that in CRDs, which affects response modelling, choice of design and the method of estimation. In this study, the statistical techniques for the analysis of split-plot experiments were examined. Response surface based modelling approach using generalised least square (GLS) which requires the use of restricted maximum likelihood (REML) estimation method for the model parameters was considered. The idea is to make comparison between GLS-estimation by REML and the ordinary least squares (OLS) that require the use of maximum likelihood estimation method. Furthermore, the method of OLS-GLS estimation was examined for it advantage property of yielding best linear unbiased estimates. In the study, the methods of estimations were compared with the aim of selecting the best under equivalent conditions using the simulated and secondary data. The data were analysed using R statistical package. The results showed that OLS estimator was not consistent for the model under study while the GLS is consistent but highly technical. However, the OLS-GLS is found to be less technical and more consistent.

***Keywords**:*
Response Surface Model, Split-Plot Design, OLS, GLS, Equivalence Condition.

[1,] 1Saleh, I. M. and [2,] Akeyede, I.
1,2Department of Mathematics, Federal University Lafia, Lafia, Nigeria
Email: 1salehibro22@gmail.com, 2imamakeyede@gmail.com.

Introduction

In conducting experiments, factors are often segregated with regards to their adoptability to change at every experimental run. Noticing the cost, time consumption and size that is involve in the changing processes, easy-to-change factors are often considered as less costly, less time consuming and not large enough in sizes. They are therefore allowed to change freely at every experimental run. However the hard-to-change factors are assumed the contrary, they are therefore restricted and allow changing less frequently. This results to cost saving, time management and efficiency (Montgomery, 2008 and Kirk, 2009).

In most situations involving industrial experiments, factors are broadly categorised into two namely, factors with costly or hard-to-change levels and those factors that levels may be relatively easy-to-change, both popularly known as whole plot factor(s) and split-plot factors respectively. For Example, A unit of Local Small Scale Block Industry uses a vibrating compacting machine for hollow blocks production seeks to investigate the causes of changes in their product quality and made available the following 2 factors: Production Machine (M), brand of cement (C) at two levels each.

Table 1: Split-plot Design for Small scale block industry

Runs	Production Unit			
	M_2	M_2	M_1	M_1
1	C_2	C_1	C_2	C_1
2	C_1	C_2	C_1	C_2

statistical softwareand make a report about the validity of the claims reported by Vinning *et.al* (2005) and on equivalent designs.This study therefore aimed at investigating the performance of OLS and GLS estimators in estimating second order response surface modelling for the analysis of Split-Plot Designs. Furthermore, the method of OLS-GLS estimation was examined for it advantage property of yielding best linear unbiased estimates. the methods of estimations were compared with the aim of selecting the best under equivalent conditions.

Response Surface Methodology

Box and Wilson 1951 introduced statistical methods to optimise processes using a series of design experiments which are basically referred to as Response Surface Methodology (RSM).This method brought about an active research area for industrial Statisticians. The earlier development of RSM was basically in the context of chemical industrial experimentation in Chemical industries. Since that time, its gained wider application in different areas of human endeavours ranging from Agricultural, Food science, Environmental Studies, Biological sciences and lots more.

Response Surface methodology can be seen as combination of mathematical and statistical approach applicable for the modelling and analysis of industrial problems and optimizing processes under study. The goal of the Response Surface Methodology is to detect within little number of experiments the settings among variety of input variables at which the process response is at maximum or minimum In industrial experiments, regression models are formulated to provide a parametric equation

suitable for response prediction at any level within the experimental model may be fit, repeating the procedure continually, you will eventually reach the vicinity of the optimum. The model (1) is usually indicated in response surface modelling by lack of fit. Then additional experiments are carried out to obtain more precise estimates of the optimum. Continually, when the search is relatively close to the optimum, a model that includes curvature is applied to approximate the response. The commonly used is model (2), which is considered adequate.

In order to ensure computational simplicity, the response surface model for a CRDin (1) can be expressed in matrix form as

$$y = X\beta + \epsilon \qquad (3)$$

$where\ y$ is $n \times 1$ vector of responses, $y^T = [y_1, y_2, \dots, y_n]$, X is a $n \times p$ model matrix containing vectors of the explanatory variables x, for p=2, then, $X^T = [1_n, x_1, x_2, x_1^2, x_2^2, x_1x_2]$ where 1_n is a n - dimensional column vector of 1's, n is the total number of experimental runs, x_1^2 *is the squared term* of the explanatory variable x_1, x_2^2 is the squared term of the explanatory variable x_2, β is $p \times 1$ vector of the unknown model parameters, for p=2, then, $\beta^T = [\beta_o, \beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12}]$, ϵ is a $n \times 1$ vector of responses measurement errors $\epsilon^T = [\epsilon_1, \epsilon_2, \dots, \epsilon_n]$, it assumed to be i.i.d such that $\epsilon \sim N(0, \sigma^2 I_n)$, I_n is an

Consequently, the model (6), is appropriate for balanced designs. In the absence of balance or there exist more complicated structure in the design such as response surface experiment, or there is missing data in the design then a more general approach based model is needed for the analysis (Kempthorne, 1952; Jones and Nachtsheim, 2009). Based on this, the response surface methodology approach was used in this study for the split-plot analysis.

Model Estimation Procedure in Split-Plot Experiment

A mixed model is used in this research work. For every split-plot design of experiments that is run with w_awhole plots, the general matrix model of response surface split-plot type with similar proposed by (Goos et al., 2004) can be written as

$$Y = X\beta + Z\delta + \varepsilon \qquad (7)$$

where y is the vector of N observed responses, $N = \sum n$, such that, $y^T = [y_1, y_2, \ldots, y_n]$, X is the model matrix containing both the whole plot factors, subplots factors settings, their interaction and any squared terms, Z is a $N \times w_a$ block form matrix of incidence, thus,

$$Z = \begin{pmatrix} 1_k & 0 & \cdots & 0 \\ 0 & 1_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1_k \end{pmatrix}$$

where the Z matrix element (i,j) is 1 if ith experimental run is in the jth whole plot and element (i,j)is 0 otherwise, k is the total number of whole plots(whole plot size), δ is a w_m×1 vector of whole plot (random intercept) errors, it is assumed to be i.i.d. normally distributed with mean zero and variance $\sigma_\delta{}^2 I$, i.e $\delta \sim N(0, \sigma_\delta{}^2 I_{w_a})$, The estimates of the variance components can be obtain using estimation techniques called Restricted Maximum Likelihood (REML). Our interest is to use REML estimation method to estimate both the factor effects and the variance components associated with the model (7) in R statistical software through 'lme4' package.

Restricted Maximum Likelihood (REML) Estimation

The method of Maximum Likelihood (ML) estimation is often subjected to criticism due to its inability to account for the loss of degrees of freedom needed to estimate the unknown regression coefficient β. Another point of concern about the ML method of estimation is that most of the properties its depend on are asymptotic, which works well for only the large sample size cases of parameter estimation but not the small sample cases.In fact, one of the common identified problems for ML estimations is under estimation of parameters variance for small sample size. The ordinary least squares estimation method are ML based approach of estimation.

The method of ordinary least squares (OLS) estimation approach may not be appropriate for model estimation of split-plot design since it totally ignores the dependent structure of the design by assuming that $\Sigma=\sigma^2$ In. The restricted maximum likelihood (REML) estimation is an estimation techniques that uses transform version of likelihood of the responses y based on the residuals from the fitted model to estimate the variance components in (8).

The REML estimators of the variance components produce the same estimates as the unbiased positive ANOVA based estimators. These provide better approach for split-plot model estimation where $\theta = (\sigma\delta^2, \sigma^2)$,, $\tilde{\beta}^G$ is (8), lr (θ) is therestricted log-likelihood function of a multivariate normal probability density function of the responses y.

The estimation of the variance components becomes necessary for their role in estimating the unknown model parameters $\tilde{\beta}G$ of β as in (11) and (12), and also for the used in drawing inferences about the model parameters. Most importantly, (Searle, 1992) showed that REML is a modern estimation approach and proven to work efficiently for unbalanced data, in the case of a balanced data, REML estimates are the same with the standard ANOVA estimates if the ANOVA estimates are positive.

When the variance components θ are estimated, the vector of the fixed affects estimates of β can be obtained using (12) as $\hat{\beta}_G$, called the best linear unbiased estimate (BLUE), as well as the covariance matrix of the model coefficients using (13).(Gilmour and Goos, 2009) reported that,the method may be inadequate when used in non-orthogonal split-plot designs with few whole plots, can produce misleading outcome. He further suggested the use of Markov Chain Monte Carlo method as an alternative method for cases like that.

General Equivalent Conditions

Based on the conditions from(Vining et al.,2005) given for the proof of equivalent condition, we can deduce the following summary that the condition holds generally for a design that is A balanced design Orthogonal but not necessarily the same design in the subplot model matrix with exception of the pure quadratics and the subplot model matrix sum to a constant. Such that the relation in (14)is satisfied. Equivalence can also be achieved when the subplot pure quadratic

$$XF = \sigma^2 X + \sigma_\delta{}^2 X(X^T X)^{-1} X^T Z Z^T X$$

$$X(X^T X)^{-1} X^T Z Z^T X = [ZZ^T]X \qquad (16)$$

When (16) hold, OLS-GLS equivalent can be achieved without requiring knowing the variance components or the matrix (F. Parker et al. 2007)pointed that to construct equivalent estimation designs, there is need for suitable choices for the p×p matrix F such that it satisfies the equivalent condition in (14). Thus, we can be expressed from (15) as

$$F = \sigma^2 I_n + \sigma_\delta{}^2 U$$

where

$$U = (X^T X)^{-1} X^T Z Z^T X \qquad (17)$$

Then, condition for equivalence can be defined in terms of $U, thus$

$$XU = ZZ^T X \qquad (18)$$

The column of model matrix within each whole plot is sums by the right hand side of equation (18), the column sums must equals the model matrix multiply by U to achieve equivalence condition.

Equivalence Condition in Matrix Form of a Second-Order Response Surface Model

Looking at the equivalent condition in (18), depends only on the model matrix and the incidence block matrix, we can write the second-order response surface model matrix X as

$$X = \begin{pmatrix} 1_k & W_1 & S_{b_1} & S_{q_1} \\ 1_k & W_2 & S_{b_2} & S_{q_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1_k & W_a & S_{b_a} & S_{q_a} \end{pmatrix} \qquad (19)$$

where n is the length of vectors 1's represented by I.

we then have that

$$XU = \begin{pmatrix} n1 & nW_1 & 1a^T{}_o & 11^T S_{q_1} \\ \vdots & \vdots & \vdots & \vdots \\ n1 & nW_a & 1a^T{}_o & 11^T S_{q_a} \end{pmatrix}$$

also,

$$ZZ^T X = \begin{pmatrix} n1 & nW_1 & 1a^T{}_o & S_{q_1} Q \\ \vdots & \vdots & \vdots & \vdots \\ n1 & nW_a & 1a^T{}_o & S_{q_a} Q \end{pmatrix}.$$

We can finally state that the equivalent condition

$$XU = ZZ^T X \qquad (21)$$

holds for $i = 1,2,\dots,a$, if and only if

Data Analysis and Comparison $11^T S_{q_i} = S_{q_i} Q$ (22)

The estimation methods are compared for various values of fixed effect estimates and the variances obtained from results in R-statistical package. The design used are partly created on published data (Kowalski and Potcner, 2003; Jones and Nachtsheim, 2009). Furthermore, datawasreframed by changing some design points from the original design which requires augmentation. As a result of that, the designs are formed for the sake to carry out non-equivalent and equivalent designs estimation. The data were analysed using the two estimators OLS and GLS when the equivalence conditions hold and when violated. The summary of the results are presented in tabular form to ease reference, and comments were made beneath each compared pairs of the table (See tables 2, 3, 4 and 5).

the correct methods. At 5% level of significance, it can be seen that GLS estimate reported that the estimate effect of parameters is significantly the same with the fixed value based on the results. Also the results of GLS and OLS are inconsistent in the reframe and real data analyses.

Table 3: Analysis Summary for (Kowalski and Potcner, 2003) Baking industry Design

	REML-GLS				OLS			
Fixed Effect	Estimate	StdError	t value	p-value	Estimate	Std. Error	t value	p-value
$\hat{\beta}_o$	1059.165	19.410	54.567	0.04777	1056.453	17.436	60.591	8.76e-11*
$\hat{\beta}_1$	40.276	20.682	1.947	0.0814	40.851	20.337	2.009	0.0445*
$\hat{\beta}_2$	-16.038	24.476	-0.655	0.03954*	-14.846	24.097	-10.616	0.0173*
$\hat{\beta}_3$	-19.253	50.554	-0.381	0.071582	-17.129	50.126	-10.342	0.7426
$\hat{\beta}_4$	3.591	22.525	0.159	0.07804	3.030	22.697	10.133	0.00897*
$\hat{\beta}_{11}$	-1.854	39.143	-0.047	0.06360	-5.048	8.912	-10.130	0.00900*
$\hat{\beta}_{22}$	-29.686	33.697	-0.881	0.05110	-26.038	32.370	-10.804	0.00448*
$\hat{\beta}_{33}$	-75.248	37.852	-1.988	0.05749	-78.052	7.937	-12.057	0.03872*
$\hat{\beta}_{44}$	3.294	33.661	0.098	0.07912	7.122	32.310	10.220	0.01318*
$\hat{\beta}_{12}$	88.865	70.443	1.262	0.07784	83.158	70.298	11.183	0.00275*
$\hat{\beta}_{13}$	192.625	46.890	4.108	0.06198	191.352	47.487	14.030	0.00500*
$\hat{\beta}_{14}$	35.212	17.784	1.980	0.16271	34.399	17.088	12.013	0.00840*
$\hat{\beta}_{23}$	-121.112	46.101	-2.627	0.07408	-122.084	46.119	-12.647	0.00331*
$\hat{\beta}_{24}$	-28.621	26.987	-1.061	0.32619	-27.011	26.815	-11.007	0.003473*
$\hat{\beta}_{34}$	9.310	44.959	0.207	0.84222	10.594	44.663	10.237	0.008193*

** indicates parameters that are significant at 5% level.It was observed that, in the non-equivalent design*

(table 2 and 3), the OLS method represents incorrect method of analysing split-plot design and the REML-GLS method represents the correct methods. At 5% level of significance, it can be seen that GLS estimate reported that the estimate effect of parameters is significantly the same with the fixed value based on the results. Also the results of GLS and OLS are inconsistent in the reframe and real data analyses. More so, the variance estimates of the GLS are given by $\hat{\sigma}^2+\hat{\sigma}_\delta^2$, where $\hat{\sigma}^2=6.742$ and $\hat{\sigma}_\delta^2=11.142$ as obtained from the analysis result, it implies that the total variance is 17.884, while OLS estimates for the variance is $\hat{\sigma}^2=4.036^2=16.289$. This may lead to loss of precision in inference when considering analysis of variance for the fitted model.

*indicates parameters that are significant at 5% level. The design in Table 5 was orthogonal design and also reported with zero variance at the whole plot using REML-GLS estimation in R. We run the analysis in R which confirmed the results that the design is orthogonal, thus, $\hat{\sigma}^2=6.495$, $\hat{\sigma}_\delta^2=0$ and for the OLS $\hat{\sigma}^2=2.549^2=6.497$. The parameters of the main effect and the interactions and their p-values, from the table 4 and 5, showed that GLS and OLS estimators are consistent in estimating equivalent designs.

Conclusion

The method of equivalent estimation enjoy some properties such as model estimation simplification since best linear unbiased estimates can be obtain by both OLS and GLS approach, parameter estimates are independent of variance components, designs are constructed independent of model assumptions, the practical challenges of statistical software that are capable of handling REML estimation is also alleviated among others. In this work we demonstrates some of these methods studied based on the available literature as well as applying some of the methods to generates some designs which analysis are carried out in R-statistical software via Lme4 package and the results were presented. . It is therefore recommended that OLS-GLS equivalence estimator for Order of Response Surface Model on Split-Plot Design due to its fewer technicalities and more consistency.

analyses (Vol. 152): CRC press.

Kirk, R. E. (2009). Experimental design. *Sage handbook of quantitative methods in psychology*, 23-45.

Kowalski, S. M., Cornell, J. A. & Vining, G. G. (2002). Split-plot designs and estimation methods for mixture experiments with process variables. *Technometrics,* 44(1), 72-79.

Kowalski, S. M. & Potcner, K. J. (2003). How to recognize a split-plot experiment. *Quality progress,* 36(11), 60-61.

Kulahci, M. & Tyssedal, J. (2016). Split-plot designs for multistage experimentation. *Journal of Applied Statistics*,34(2), 1-18.

Letsinger, J. D., Myers, R. H. & Lentner, M. (1996). Response surface methods for bi-randomization structures. *Journal of Quality Technology,* 28(4), 381-397.

McAllister, P. R. & Holland, B. S. (1979). On OLS estimation in singular linear models with general positive definite dispersion matrix. *Communications in Statistics-Theory and Methods,* 8(3), 299-302.

McElroy, F. W. (1967). A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased.*Journal of the American Statistical Association,* 62(320), 1302-1304.

Montgomery, D. C. (2012). *Design and Analysis of Experiments*: Wiley. 8th Edition.

Montgomery, D. C. (2008). *Design and analysis of experiments*:

Vining, G. G., Kowalski, S. M. & Montgomery, D. C. (2005). Response surface designs within a split-plot structure. *Journal of Quality Technology,* 37(2), 115.