Integra: Journal of Integrated Mathematics and Computer Science
Journal Homepage: https://integrajimcs.com/index.php/integra/index
integra
Journal of Integrated Mathematics and Computer Science

Research Paper

Survey of Estimability Criteria, Connected Design and Testing Testable Hypotheses in Unbalanced Design

Faiz AM Elfaki[1], Edwin Russel[2]*, Widiarti[2], Mustofa Usman[2], Jamal I. Daoud[3]

[1] *Dept. of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Doha, 2713, Qatar*
[2] *Dept. of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Bandar Lampung, 35145, Indonesia*
[3] *Kulliyyah of Engineering, Department of Science and Mathematics, IIUM, Kuala Lumpur, 50728, Malaysia*
***Corresponding author**: winrush08@gmail.com

Keywords
Estimability, Linear Parameter Function, Testable Hypothesis, Unbalanced Data

Abstract
In the linear model $Y = X\beta + \epsilon$ with X having a full column rank, all β parameters can be estimated and the estimates are unique. However, in cases where X does not have a full column rank, not all β parameters can be estimated. In this paper, the problem to be discussed is how to determine parameters or parameter functions that are estimable and testable. Applications to the case of unbalanced data will be presented.

Received: 2 September 2025, Accepted: 16 November 2025
https://doi.org/10.26554/integra.20252343

1. INTRODUCTION

The idea of a parametric estimable function was first proposed by Bose [1], who stated that such a function requires that the function have an unbiased linear estimator. Since then, this concept has attracted much attention from researchers in statistical theory and linear modeling, and has become an interesting topic of study, especially when discussing cases of missing and unbalanced data, nonstandard models, and models for unbalanced data with non-full rank column design matrix X [2]. The importance of the idea of estimability lies in the fact that there is a best unbiased linear estimator (BLUE) of a linear combination of parameters if the linear combination is estimable [2, 3, 4, 5, 6]. Generally, estimability requirements are very difficult to check [2, 7, 8, 9]. For studies and developments in the concept of estimability, we can see many writings [2, 3, 10, 11, 12]. Searle [11] defined estimability in terms of the generalized inverse design matrix, while Milliken [3] defined estimability using the trace matrix. Baksalary and Kala [13] developed the concept which have been developed by Milliken by generalizing the properties of Milliken's design matrix. Alalouf and Styan [7] presented several characteristics of estimability by examining properties of their design matrices Usman [9] presented estimability criteria for missing data cases, particularly in split plot designs, while Elswick et al. [14] presented a technique for determining estimable parametric functions by using the concept of elementary row operations in their matrix design. Equation (1) is the linear model with the mean vector $X\beta$ overparameterized [10, 12] or with the design of the matrix X not having a full column rank, then two vectors β can produce the same vector $X\beta$.

$$Y = X\beta + \epsilon \tag{1}$$

In many non-standard models, for example in the case of unbalanced data in hypothesis testing, the concept of estimability becomes a very important issue. Equation (2) below is the hypothesis testing: :

$$H_o : H\beta = a \quad \text{with the alternative} \quad H_a : H\beta \neq a \tag{2}$$

A requirement for a hypothesis to be testable is that $H\beta$ must be an estimable function [2, 3, 11, 15, 16], meaning that each component of $H\beta$ is an estimable function of parameters. If $H\beta$ is an estimable function of parameters, then the hypothesis is said to be a testable hypothesis [3, 15]. By testable hypothesis, we mean that a hypothesis is testable if it can be expressed in the form of an estimable function [16]. Equation (2) does not

imply that a hypothesis consisting of non-estimable functions cannot be tested. However, it is reasonable to assume that a testable hypothesis is constructed by an estimable function for the following reason: If $H\beta = a$ is to be tested, then the case for full rank suggests that $H\beta^0 - a$ will be part of the test statistic that must be invariant to β^0 and this will be invariant if $H\beta$ is estimable.

Some application of estimability in state space model given by [17], and in factor analysis given by [18], application in global navigation satellite systems (GNSS) given by [19], and application in pharmaceutical models is given by [20].

2. THE METHOD

In linear model $Y = X\beta + \varepsilon$, here X is an $n \times p$ matrix, β is a $p \times 1$ parameter vector and Y is an $n \times 1$ observation vector. Let $A\beta$, where A is $s \times p$ matrix and if there is a matrix B that satisfies Equation (3) as follows:

$$E(B'Y) = A\beta \tag{3}$$

then $A\beta$ is said to be estimable. In other words, every function of observation is estimable [5, 16, 21, 22].

From model (1), $E(Y) = X\beta$, then from Equation (3), $A = BX$ or equivalently:

$$\rho \begin{bmatrix} A \\ X \end{bmatrix} = \rho(A) \tag{4}$$

with $\rho(.)$ denoting the rank of a matrix. Equation (4) is satisfied if and only if with generalized inverse [10, 11].

Theorem 1. [6]

If $a'\beta$ is estimable, then $a'\hat{\beta}$ is $\hat{\beta} = (X'X)^{-}X'Y$ which is invariant with respect to the choice of $(X'X)^{-}$.

Proof.

If $a'\beta$ is estimable, then $a' = b'X$ for some b. Hence

$$a'\hat{\beta} = a'(X'X)^{-}X'Y$$

$$a'\hat{\beta} = b'X(X'X)^{-}X'Y$$

is invariant of $a'\hat{\beta}$ is a consequence of the fact that it is invariant to the choice of generalized-inverse (g-inverse).

Milliken [3] discussed the concept of estimability of rank matrix involving the concept of generalized inverse matrix, in the sense of Moore-Penrose g-inverse [10]; that is, denotes the g-inverse of matrix A if it satisfies AA^{-} symmetric, $A^{-}A$ symmetric, $AA^{-}A = A$, and $A^{-}AA^{-} = A^{-}$. The estimability discussed the concerns of linear combinations of parameters in the linear model in Equation (1), and Equation (5) below is its normal equation:

$$X'X\hat{\beta} = X'Y \tag{5}$$

The Equation (6), $\hat{\beta}$, for Equation (5).

$$\hat{\beta} = X'Y + (I - X^{-}X)h \tag{6}$$

h is the $p \times 1$ vector. If $A\beta$ is an estimable linear combination, then it is known that the BLUE of the set with any solution of Equation (6). Then the estimability condition can be expressed in the form of the rank of the product matrix.

Theorem 2. [3]

For linear model (1) with rank matrix X of q, the linear combination $A\beta$ is estimable, where A is a $k \times p$ matrix of rank k, if and only if the rank of the matrix is $q - k$.

Proof.

For unbalanced designs, it is generally very difficult to determine whether the linear combination of β is estimable. The result of Theorem 1 is also difficult to verify, but the estimability condition can be formulated using the following matrix trace.

Theorem 3. [3]

For the conditions in Theorem 1, a linear combination $A\beta$ is estimable if and only if

$$tr[X(I - A^{-}A)\{X(I - A^{-}A)\}^{-}] = q - k \tag{7}$$

Proof.

Matrix in Equation (7) is idempotent, hence

$$\begin{aligned} \text{tr}\left[X(I - A^{-}A)\{X(I - A^{-}A)\}^{-}\right] &= \rho\big[X(I - A^{-}A) \\ &\quad \{X(I - A^{-}A)\}^{-}\big] \\ &= \rho\big[X(I - A^{-}A)\big] \\ &= q - k. \end{aligned}$$

Alaoof and Styan [7] explored the estimability characteristics based on X. Here we present two theorems that propose characteristics based on these two matrices.

Theorem 4. [7]

Estimability characteristics based on X. The vector $A\beta$ is estimable if $E(Y) = X\beta$ if and only if one of the following seven conditions is met.

(1.1) $A = BX$ for some matrix B.

(1.2) $\rho[X'A'] = \rho(X)$.

(1.3) $\rho[X(I - A^{-}A)] = \rho(X) - \rho(A)$ for some g-inverse A.

(1.4) $AX^{-}X = A$ for some g-inverse X.

(1.5) AX_1^{-} is invariant for every least squares X_1^{-}.

(1.6) $\rho(AX_1^{-})$ is invariant for every least squares X_1^{-}.

(1.7) $\rho(AX_1^{-}) = \rho(A)$ is invariant for every solution of least squares.

Theorem 5 [7]

Estimability characteristics based on X. The vector $A\beta$ is estimable if $E(Y) = X\beta$ if and only if one of the following nine conditions is met.

(1.1) $A\hat{\beta}$ is invariant for every $\hat{\beta}$ that satisfied $X'\hat{X}\hat{\beta} = X'Y$,

(1.2) $\rho[X'A'] = \rho(X'X)$,

(1.3) $\rho[X(I - A^{-}A)] = \rho(X) - \rho(A)$ for some g-inverse A,

(1.4) $A(X'X)^{-}(X'X) = A$ for some g-inverse $X'X$,

(1.5) $A(X'X)^{-}A'$ is invariant for every g-inverse $X'X$,

(1.6) $\rho(A(X'X)^{-}A')$ is invariant for every g-inverse $X'X$, $X'X$,

(1.7) $\rho(A(X'X)^{-}A') = \rho(A)$ is invariant for every g-inverse $X'X$,

(1.8) $$\rho\left(\begin{bmatrix} X'X & A' \\ A & 0 \end{bmatrix}\right) = \rho(X'X) + \rho(A(X'X)^{-}A')$$ for some g-inverse $X'X$.

(1.9) $$\rho\left(\begin{bmatrix} (X'X)^{-} + (X'X)^{-}A'S^{-}A(X'X)^{-} & -(X'X)^{-}A'S^{-} \\ -S^{-}A(X'X)^{-} & S^{-} \end{bmatrix}\right) = \rho\left(\begin{bmatrix} X'X & A' \\ A & 0 \end{bmatrix}\right)^{-} \quad 1.9$$

For some g-inverse $(X'X)^{-}$ and S^{-} where $S = -A(X'X)^{-}A'$.

For special cases where the design has several missing observations and the model has restrictions or conditions on its parameters, see [9] who presents estimability criteria for split plot designs with several missing observations and restrictions on its parameters.

Connected Design and Estimability.

Let the design:

Design I. Column Trt

X			X
	X		X
		X	X
		X	X

Connected

Design II. Column Trt

X		X	X
	X		X
X		X	X
	X		X
X		X	X

Not connected

Note: "X" represents the observation of a combination of treatments [2].

Another concept related to the estimability function is the connectedness of the two-way treatment structure. If it is assumed that the row and column levels of the treatments do not interact, then the combined treatment can be modeled

$$\mu_{ij} = \mu + \tau_i + \beta_j \qquad i = 1, 2, \dots, b;\ j = 1, 2, \dots, r.$$

A two-way treatment structure is said to be connected if and only if the data occurring in the cells of the two-way treatment structure is such that

$$\beta_j - \beta_{j*} \text{ and } \tau_i - \tau_{i*},$$

estimable for every $i \neq i^*$ and $j \neq j^*$.

For design I is connected, while design II is not connected. As an illustration, from design I,

$$\beta_1 - \beta_2 = (\mu + \tau_1 + \beta_1) - (\mu + \tau_1 + \beta_5) + (\mu + \tau_2 + \beta_5) - (\mu + \tau_2 + \beta_2).$$

It is a linear combination of cell means. So $\beta_1 - \beta_2$ is estimable. Meanwhile, from design II, we cannot obtain $\beta_1 - \beta_2$, so at II, $\beta_1 - \beta_2$ is not estimable.

Testable Hypothesis

In testing the hypothesis $H_0 : H\beta = a$ with the alternative $H_a : H\beta \neq a$, $H\beta$ must be an estimable function of the parameters. Seely [15] illustrates the problem that arises if we do not restrict the hypothesis testing problem to parametric functions. This can result in the distribution classes for H_0, P_0 and the distribution classes for H_a, P_a being non-exclusive. In the hypothesis $H_0 : H\beta = a$ with the alternative $H_a : H\beta \neq a$, if $H\beta$ is an estimable function, the hypothesis is called a testable hypothesis. Searle et al [16] showed that if $H\beta$ is not estimable, the numerator of the sum of squares in the F-ratio will not be well-defined.

The definition of a testable hypothesis is that a testable hypothesis does not provide much information. By testable hypothesis, we mean a hypothesis that can be expressed in the form of an estimable function. It seems reasonable to assume that a testable hypothesis is one constructed by an estimable function for the following reason: If $H\beta - a$ is to be tested, then the results

from the case of a linear model with full column rank suggest that $H\beta_0 - a$ is part of the test statistic, which must of course be invariant with respect to β_0, and this will be invariant only if $H\beta$ is estimable.

To construct a hypothesis test, we will present Milliken's [3] approach. For linear model (1), we consider the null hypotheses which given in Equation (8) as follow:

$$H_0 : H\beta = 0 \quad \text{with the alternative} \quad H_a : H\beta \neq 0 \tag{8}$$

With H being a $k \times p$ matrix with rank k and the linear combination $H\beta$ is estimable. So

$$\rho[X(I - H^+H)] = q - k.$$

With the Principle of Conditional Error [2] we can use it to calculate the sum of the squares of the hypothesis error as follows. Equation (9) is the model for the null hypothesis in Equation (8).

$$Y = X(I - H^+H)\beta + e \tag{9}$$

The sum of the squared errors of model (9) is

$$SS_{ER} = Y'[I - X(I - H^+H)(X(I - H^+H))^+]Y.$$

The sum of squares of the linear model errors is

$$SSE = Y'[I - XX^+]Y.$$

With the Principle of Conditional Error, the sum of squares due to the null hypothesis can be formulated as follows,

$$SSH_0 = SS_{ER} - SSE = Y'[XX^+ - X(I - H - H)(X(I - H^+H))^+]Y.$$

And with the Principle Conditional Error, the statistical test is

$$F = \frac{SSH_0/k}{SSE/(n-k)}.$$

With F distributed F with degrees of freedom k and $n - k$.

3. RESULTS AND DISCUSSION

Application of a testable hypothesis on unbalanced data. As an illustration, suppose we have the following data:

X	X	
	X	X
X	X	
X	X	X

With X indicating existing observations, the Equation (10) is the model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \tag{10}$$

Where Y_{ij} is the observation in the first row and the second column j; α_i the i^{th} row effect; β_j is the j^{th} effect of column, and ε_{ij} is random error. The data above, in the form of a linear model, can be written as follows:

$$Y = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \varepsilon$$

Or it can be written as Equation (11) as follows

$$Y = X\beta + \varepsilon \tag{11}$$

The matrix X is not fully column-ranked, with parameter β being $\beta' = (\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \beta_4)$. Using elementary row operations, the echelon matrix of X is,

$$X * = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \varepsilon$$

So the parameter function is estimable:

$$\mu+\alpha_3+\beta_4-\alpha_1-\alpha_2, \quad \alpha_1-\alpha_3, \quad \alpha_2-\alpha_3, \quad \beta_1-\beta_4, \quad \beta_2-\beta_4, \quad \beta_3-\beta_4.$$

Based on this estimable parameter function, we can construct several testable hypotheses, including:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

This hypothesis in the form of the above parameter function is equivalent to the hypothesis

$$H_0 : \beta_1 - \beta_4 = 0, \quad \beta_2 - \beta_4 = 0, \text{ and } \beta_3 - \beta_4 = 0$$

or

$$\text{or } H_0 : \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \beta = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

or $H_0 : H\beta = 0$, where $H\beta$ is an estimable parameter function. So this hypothesis is a testable hypothesis. To test this hypothesis, we can use the procedure above or the Principle of Conditional Error. We can also test the hypothesis

$$H_0 : \ \alpha_1 = \alpha_2 = \alpha_3$$

This hypothesis in the form of an estimable parameter function can be expressed in the following hypothesis form:

$$H_0 : \ \alpha_1 - \alpha_3 = 0 \text{ and } \alpha_2 - \alpha_3 = 0,$$

or

$$H_0 : \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \beta = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or $H_0 : \ H_1\beta = 0$, where $H_1\beta$ is an estimable parameter function, so it is a testable hypothesis.

4. CONCLUSIONS

The concept of estimability developed primarily in the periods of 1960s to 1970s, with numerous research findings on its characteristics and testability. The concept of estimability is particularly useful in addressing cases of missing data that lead to unbalanced data. The study and application of estimability to messy data is also very useful in identifying estimable parameters and testability of parameters. The concept of estimability continues to evolve, and its application in research across various fields continues to expand.

5. ACKNOWLEDGMENT

The authors would like to thank Qatar University, Lampung University, and International Islamic University Malaysia (IIUM) for their support in the research collaboration and publication among the universities. The authors would also like to thank the reviewers for their feedback and suggestions for improving this paper.

REFERENCES

[1] RC Bose. The Fundamental Theorem of Linear Estimation. In *Proc. 31st Indian Scientific Congress (1944)*, pages 2–3, 1944.

[2] George A Milliken and Dallas E Johnson. *Analysis of Messy Data, Volume I: Designed Experiments*. Taylor & Francis, 2001.

[3] George A Milliken. New Criteria for Estimability for Linear Models. *The Annals of Mathematical Statistics*, 42(5):1588–1594, 1971.

[4] Shayle Robert Searle, Charles E McCulloch, and John M Neuhaus. *Generalized, Linear and Mixed Models*. Wiley Hoboken, NJ, USA, 2001.

[5] George Seber. *The Linear Model and Hypothesis*. Springer, 2015.

[6] Andre I Khuri. *Linear Model Methodology*. Chapman and Hall/CRC, 2009.

[7] IS Alalouf and George PH Styan. Characterizations of Estimability in the General Linear Model. *The Annals of Statistics*, 7(1):194–200, 1979.

[8] Jan R Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 2019.

[9] Mustofa Usman. *Testing Hypothesis in Split Plot Design When Some Observations Are Missing*. PhD thesis, Kansas State University, 1985. Unpublished Dissertation.

[10] Franklin A Graybill. Theory and Application of the Linear Model. 1976.

[11] Shayle Robert Searle. *Linear Models*. John Wiley and Sons, New York, 1971.

[12] NR Mohan Madhyastha, Sreenivasan Ravi, and AS Praveena. *A First Course in Linear Models and Design of Experiments*. Springer, 2020.

[13] Jerzy K Baksalary and R Kala. Extensions of Milliken's estimability criterion. *The Annals of Statistics*, 4(3):639–641, 1976.

[14] RK Elswick Jr, Chris Gennings, Vernon M Chinchilli, and Kathryn S Dawson. A Simple Approach for Finding Estimable Functions in Linear Models. *The American Statistician*, 45(1):51–53, 1991.

[15] Justus Seely. Estimability and Linear Hypotheses. *The American Statistician*, 31(3):121–123, 1977.

[16] Shayle R Searle, F Michael Speed, and George A Milliken. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician*, 34(4):216–221, 1980.

[17] Saang-Yoon Hyun and Kyuhan Kim. An Evaluation of Estimability of Parameters in the State-Space Non-Linear Logistic Production Model. *Fisheries Research*, 245:106135, 2022.

[18] Yunxiao Chen, Xiaoou Li, and Siliang Zhang. Structured Latent Factor Analysis for Large-Scale Data: Identifiability, Estimability, and Their Implications. *Journal of the American Statistical Association*, 115(532):1756–1770, 2020.

[19] A Khodabandeh and PJG Teunissen. Integer Estimability in GNSS Networks. *Journal of Geodesy*, 93(9):1805–1819, 2019.

[20] Iman Moshiritabrizi, Kaveh Abdi, Jonathan P McMullen, Brian M Wyvratt, and Kimberley B McAuley. Parameter Estimation and Estimability Analysis in Pharmaceutical Models with Uncertain Inputs. *AIChE Journal*, 70(1):e18168, 2024.

[21] Ravindra B Bapat. *Linear Algebra and Linear Models*. Springer, 2000.

[22] Alan Agresti. *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons, 2015.

