PROCESS ORIENTED BASIS ESTIMATION IN PRESENCE OF NON-ORTHOGONAL BASIS ELEMENTS

by

Vivian Otero Padilla

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCES
in
INDUSTRIAL ENGINEERING

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2005

Approved by:

_____	_____
José R. Deliz Ph..D. Member, Graduate Committee	Date
_____	_____
Pedro Resto, Ph.D. Member, Graduate Committee	Date
_____	_____
David González Barreto, Ph.D. President, Graduate Committee	Date
_____	_____
Maria Medina M.S. Representative of Graduate Studies	Date
_____	_____
Agustín Rullán Ph.D. Chairperson of the Department	Date

ABSTRACT

Process Oriented Basis Representations (POBREP) is a multivariate Statistical Process Control (SPC) procedure with diagnosis capabilities developed by Barton and Gonzalez-Barreto (1996). Although this methodology is effective when orthogonal process-oriented basis (POB) is presented, it is diagnosis capabilities are at risk when the POB is not orthogonal. This research compared several methods to solve non-orthogonal POB's problem. Six scenarios with different Variance Inflation Factor (VIF) severity were created using the stencil printing process. Coefficients were estimated using five methods: Ordinary Least Square (OLS), Independent Subsets (IS), Simple Regression (SR), Ridge Regression (RR) and Constrained Solution Space (CSS). These methods were compared in terms of the lower Square Error (SE) and higher number of times the coefficient is between a confidence interval (Count). There were two comparable groups of results: (1) CSS and RR methods with lowest SE and highest Count and (2) OLS, IS and SR with higher SE and lower Count. The best method estimate POBREP coefficient in presence of non non-orthogonal basis elements is Constraint Space Solution.

RESUMEN

Representación de las Bases Orientadas al Proceso (POBREP) es una metodología de análisis multivariado desarrollada por Barton y González-Barreto (1996) que tiene la capacidad de diagnóstico. Esta metodología es efectiva cuando las bases orientadas al proceso (POB) son ortogonales, pero esta capacidad de diagnóstico se afecta cuando los POBs no son ortogonales. Esta investigación compara varios métodos que permiten resolver el problema de falta de ortogonalidad en los POBs. Seis escenarios con diferentes severidades de VIF fueron desarrollados utilizando el proceso de impresión de un esténcil. Los coeficientes fueron estimados usando cinco métodos: Minimizar Errores Cuadraros (OLS), Subgrupos Independientes (IS), Regresión Simple (SR), Regresión "Ridge" (RR) y Solución de Espacio Limitado (CSS). Estos métodos fueron comparados con el objetivo de minimizar los errores cuadrados (SE) y maximizar el numero de veces que el coeficiente se encuentran entre unos limites de confianza ("Count"). Hay dos grupos de resultados comparables: (1) CSS y RR con valores mínimos de SE y valores altos "Count" ,(2) OLS, IS y SR obtuvieron valores altos de SE y bajos de "Count". El método que mejor estima los coeficientes de POBREP en presencia de falta de ortogonalidad en los elementos de la base es Solución de Espacio Limitado.

[10]The fear of the LORD is the beginning of wisdom;

A good understanding have all those who do His commandments.

His praise endures forever

Psalm 111:10

To God for give me the knowledge and the strength, to Jiretly because she is a provision of God in my life and to my family for their unconditional support

ACKNOWLEDGEMENTS

I want to recognize the support during my graduate studies at the University of Puerto Rico to some people that collaborated with my research. Without their support it would be impossible for me to finish my work.

I want to express a sincere acknowledgement to my advisor, Dr. David Gonzalez Barreto because he gave me the opportunity to do research under his supervision, his encouragement, support, guidance, and transmitted knowledge for the completion of my work. Although the challenge increases; his positive attitude, flexibility and humanity gave me the motivation to reach this goal.

I would like to thank all my family, for their unconditional support, inspiration and love. Thanks Dad and Mom, you are my role models, your love for God inspired me to follow my way and to give all my heart in everything I do. Jiretly and Miguel you are my love, my motivation and the reason to reach my dreams.

TABLE OF CONTENTS

TABLE LIST

Tables **Page**

Appendix Tables

FIGURE LIST

Figures **Page**

Appendix Figures

1 INTRODUCTION

1.1 General Content

One of the principal objectives for all industry is to provide a quality product that meets and exceeds customer's expectations. Statistical process control techniques are very valuable to achieve this objective. Understanding and improving quality is a key factor leading to business success, growth and achievement of enhanced competitive positions. Among these activities, quality data collection and analysis is required. Thus industry is moving towards a data rich environment. A lot of money is invested in systems that have the ability to capture multivariate quality characteristics for 100% of manufactured products with in-line vision systems.

The growth of data-collection technology and the use of online computer for process monitoring have led to an increased interest in the simultaneous observation of several related quality characteristics or process variables (Lowry and Montgomery 1995). Multiple characteristics are being monitored simultaneously, consequently it is important to use and develop techniques for larger sets of multivariate process quality data. The research in multivariate process control methods began with Hotellings's (1947) work based on the multivariate normal distribution with known covariance matrix. Multivariate SPC methods objective is to: evaluate whether the process is operating under a stable behavior or not, provide an evaluation of the process capability, facilitate the detection and removal of special causes, and monitor the improvement effort results.

Multivariate statistical process control techniques monitor such multivariate data to provide a characterization of when irregular behaviors occur in the process, but do not provide any clues as to what caused those irregularities. This diagnosis is necessary for the operator or the engineer in order to determine the assignable cause and take the corrective actions.

1.2 Literature Review

Process Oriented Basis Representations (POBREP) is a process diagnostics methodology developed by Barton and Gonzalez-Barreto (1996). This methodology identifies the most likely causes of variation in product performance by linking patterns in multivariate data with pattern associated with certain kinds of production problems, this is the process-oriented basis. The multivariate quality vector can be represented as a linear combination of these basis elements,

$$x = Az \tag{1}$$

Multivariate Quality Vector x is a vector with one component for each product characteristic measured, for example a vector of repeated measurements over the surface of a part (dimensions, volumes, registration errors, etc.) Process Oriented Basis Vector (POB) is a specific vector A_j that links specific manufacturing problems with a pattern of errors in the quality vector. Process Oriented Basis Representation (POBREP) is the representation of a quality vector as a linear combination of the process oriented basis vectors. Basis elements with large coefficients suggest particular causes for process problems. This methodology was successfully applied by Gonzalez-Barreto (1996) in the

stencil printing process and by Espada-Colón (1998) in the component placement of printed circuit board.

1.3 Problem Statement

The orthogonallity of process-oriented basis elements plays an important role in the success of the POBREP methodology. When the basis elements are orthogonal, POBREP establishes a reliable and accurate link between multivariate quality vectors and potential process errors characterized by a process-oriented basis. The representation of the process oriented basis, the coefficients z values do not change when other orthogonal columns are added or deleted from process-oriented basis. Some problems arise when there is severe multicollinearity or dependency among the column vectors that model potential process problem. These problems occurs regard the reliability and explanatory power of the POBREP coefficients. Non-orthogonallity and dependency produce POBREP coefficients that are highly sensitive to small changes in multivariate process data; further, the coefficients can be too large on the average and may have wrong signs. The presence of multicollinearity on process-oriented basis coefficients may result in OLS (ordinary least square) estimates with high variance, which may be distant from the true values. This is not usually significant if the interest is to use the model, but it is of cardinal importance when the intent is to use the estimates for process diagnostics and control.

The POBREP strategies do not rely in the prediction capability, but it is interested in the meaning of the basis coefficients to diagnosis and control. Although the collinearity

problem is the same for Ordinary Least Squares regression and POBREP the interpretation of the results is different. Many procedures have been proposed for diagnosing the presence of collinearity and assessing its potential harm to regression estimates, but these proposed procedures to manage multicollinearity problem for normal regression not necessarily resolve the POBREP issue. This research will compare several methods that can be used to solve the presence of non-orthogonal basis elements in POBREP.

1.4 Research Motivation

The data rich environment provides the information required to characterize and improve the process, but the challenge consist in extracting meaningful information not only for process monitoring, but also for providing precise process diagnostic to help production personnel in the identification of the most likely process causes when irregularities are detected. This research is motivated by the need for a multivariate SPC procedure with diagnosis capabilities such as POBREP methodology to provide diagnosis when an out-of-control situation occurs. Although this methodology is effective when orthogonal process-oriented elements are presented, the diagnosis capabilities are at risk when the process-oriented basis is not orthogonal or even posses some dependencies among the basis elements. Many real problem process deviations may exist that could not be represented with an orthogonal basis. In order to address this possibility several methods to work with non-orthogonal process oriented basis are evaluated to deal with this issue.

1.5 Research Strategy

This research analyzed non-orthogonal process oriented basis (POB) as presented in Figure 1.1. The strategy is initiated by obtaining non-orthogonal process-oriented basis elements, potentially using the stencil printing process (Gonzalez-Barreto 1996). Variance Inflation Factor (VIF) by Marquardt (1970) will be used to assess multicollinearity severity. Three cases with different severity in multicollinearity were considered. One of them was when VIF $\leq$ 5 showing no serious multicollinearity problem, when 5 $\leq$ VIF < 10 a moderate multicollinearity, and when VIF > 10, a more serious multicollinearity problem. Two kinds of problems were analyzed: linear relation between two basis elements and linear relation between more than two basis elements. Table 1.1 shows six scenarios based on the severity of multicollinearity and the regressors involved in the relation.

POB coefficients were estimated per scenario using five methods. These methods were: Ordinary Least Square, Independent Subsets, Ridge Regression, Constrained Space Solution and Simple Regression. Each methodology detected the active basis for three cases: a) bias in one basis elements related, b) bias in two basis elements not related and c) bias in two basis elements related. These methods were compared in terms of the lower Square Error of the estimated coefficients z_i with the theoretical value and the maximum number of times z_i is contained in a confidence interval. Both measurements are presented in Chapter 3. Final recommendations were provided in terms of which strategy achieves better result in order to work with non-orthogonal POB's .

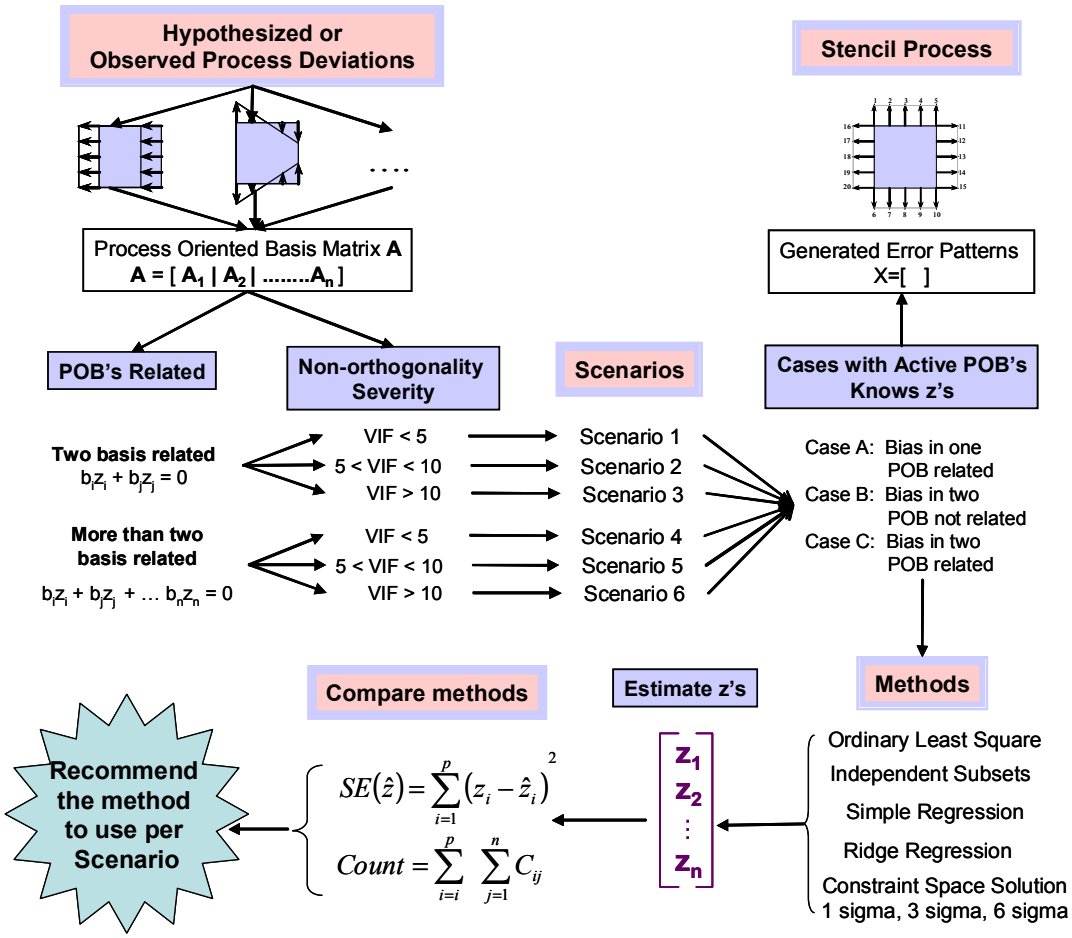


Figure 1.1: Strategy for non-orthogonal Scenarios.

Table 1.1: Strategy Matrix with non-orthogonal Scenarios.

Relation between Basis	Methods	Variance Inflation Factor		
		VIF < 5	5<VIF < 10	10< VIF
Two Basis Elements	Ordinary Least Square			
	Independent Subsets			
	Simple Regressions			
	Ridge Regression			
	Constrained space solution			
More than two Basis Elements	Ordinary Least Square			
	Independent Subsets			
	Simple Regressions			
	Ridge Regression			
	Constrained space solution			

1.5.1 Research Process Steps

1. Develop POB's with different non-orthogonallity problem.
    a. Two basis elements related
    b. More than two basis elements related
2. Develop scenarios with different non-orthogonallity severity.
    a. VIF ≤ 5
    b. $5 <$ VIF ≤ 10
    c. VIF > 10
3. Generate multiple vectors with known active patterns for different cases of severity.
    a. Case A : Bias in one basis element related
    b. Case B: Bias in two basis elements not related
    c. Case C: Bias in two basis elements related
4. Obtain z representations for each solution method
    a. Ordinary least Square
    b. Simple Regression
    c. Independent Subsets
    d. Ridge Regression
    e. Constrained Solution Space
5. Compare each method
    a. Estimated versus known z_i representations (SE).

b. Number of times z_i is in the known confidence interval (Count).

6. Select the best method per each scenario (severity), to minimize SE and maximize Count.

7. Recommend a procedure for dealing non-orthogonal POB's.

1.6 Organization of the Thesis

This research includes five major topics. In chapter one the problem under study and research motivation was presented. Second chapter consists of the review and analysis of background literature related to the problem area. It also provides a review of the POBREP methodology that will form the basis for some of the research that is described in the subsequent chapters. The third chapter includes stencil example scenarios, illustrates detailed description of the proposed methodology and explains the criteria for comparison POBREP solutions. The fourth chapter consists of the stencil example scenarios results for Scenario 1 (best case) and Scenario 6 (worst case). The same chapter presents result summary and recommendations per scenario. Finally, research contribution and future research are included in the fifth chapter.

2 BACKGROUND REVISION

2.1 Process-Oriented Basis Representation Methodology

Process Oriented Basis Representation (POBREP), developed by Barton and Gonzalez-Barreto (1996), is a process diagnostic methodology which identifies the most likely causes in product performance by linking patterns in multivariate performance data with patterns associated with certain kinds of production problems. For a single part, a multivariate quality vector $\boldsymbol{x}$ is defined as the set of m measured deviations from nominal. For example, in a component placement process a quality vector is represented by the coordinates (x_1, y_1) and (x_2, y_2) (see Figure 2.1). If displacement in X occurs the quality vector take values [1,0,1,0].

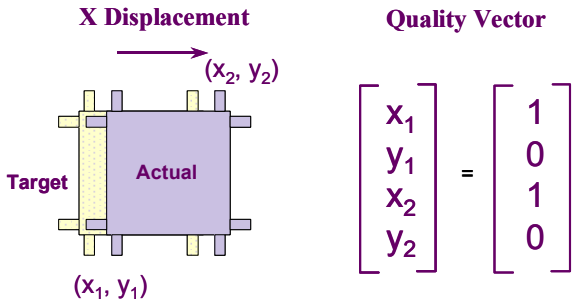


Figure 2.1: Quality vector in a placement example.

Linking process error with the resulting pattern of error over the surface of a manufactured part provides a way to diagnose observed error patterns in such parts, by representing the quality vector using a non-standard basis. It is possible to identify a pattern of errors for each potential cause of process bias or variability. Suppose that k

different patterns of interest can be identified for k different process causes, say $\mathbf{a_1}$, $\mathbf{a_2}$, . . . $\mathbf{a_k}$, where $\mathbf{a_i}$'s are $\boldsymbol{m}$ dimensional vectors. If the vectors $\mathbf{a_1}$, $\mathbf{a_2}$, . . . $\mathbf{a_k}$ are independent and $m = k$, then the cause related patterns provide an alternative basis for representing the same quality vector, and the representation of $\boldsymbol{x}$ in this basis is $\mathbf{x = z_1a_1 + z_2a_2 + \ldots z_ka_k}$. That is, $\boldsymbol{x}$ can be thought of as a weighted sum of characteristics patterns, where the amount of pattern $\mathbf{a_i}$ in $\boldsymbol{x}$ is indicated by the coefficient $\mathbf{z_i}$. The vector $\mathbf{z = (z_1, z_2, \ldots, z_k)'}$ can be found b solving the system of linear equations: $\mathbf{x = A\ z}$, where $\mathbf{A}$ is the matrix composed of column vectors $\mathbf{a_1}$, $\mathbf{a_2}$, . . . $\mathbf{a_k}$: $\mathbf{A = [a_1 \mid a_2 \mid \ldots \mid a_k]}$.

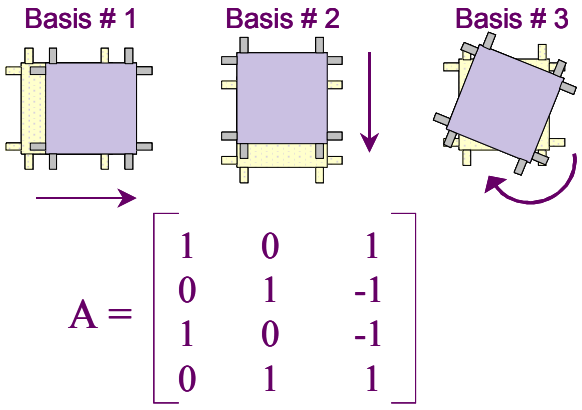


Figure 2.2: Process oriented basis for a placement example.

A is called a *process-oriented basis (POB).* For the component placement example the Figure 2.2 shows the possible POB. The POB presents X displacements, Y displacement and rotation. The observed quality vector $\boldsymbol{x}$ was decomposed into patterns corresponding to known causes. By solving the system of linear equations to find the z vector, a process-oriented basis representation is formed. The components of the z vector, z_i , i= 1, 2, . . . , k, are called *POBREP coefficients*. Using the process-oriented basis

representation z, diagnosis is possible: potential causes are associated with patterns ($\mathbf{a_i}$) having large positive or negative POBREP coefficients ($\mathbf{z_i}$). Figure 2.3 shows a graphic representation of the POBREP methodology for a placement example.

Note that in many cases it will not be necessary to construct a complete basis. This corresponds to a situation where k < m. When k < m, the process-oriented basis may not span the subspace that x lies in, hence there may be no exact solution to the linear system. In this situation, $\boldsymbol{x}$ can be represented as a linear combination of k basis elements and a residual vector in the following regression equation form:

$$\mathbf{x = Az + e} \qquad \textbf{(2)}$$

this can be solved by the ordinary least squares method (OLS). POBREP methodology differs from traditional regression context in that regression the equation is solved for many consecutive quality vectors, allowing the analysis of the behavior of z and e over time.

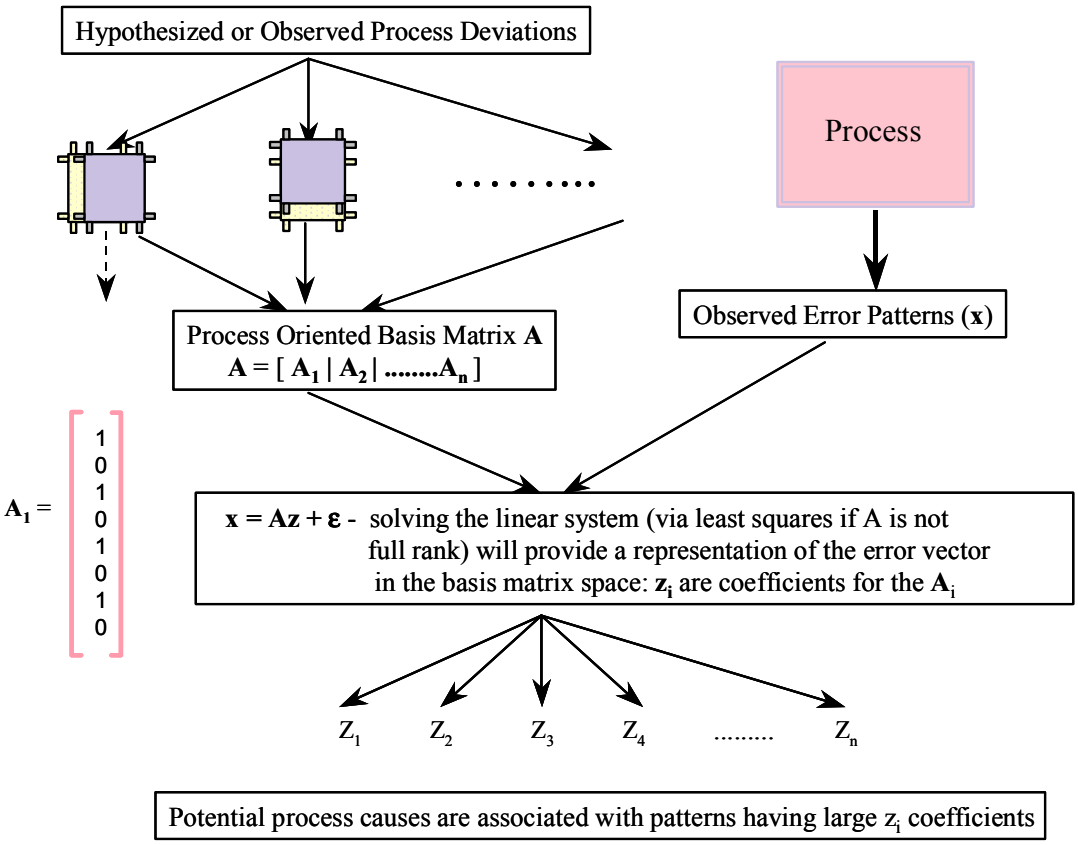


Figure 2.3: POBREP methodology for a placement example.

2.2 Multicollinearity

The purpose of a regression model is to identify the relative effects of the regressors variables, predict or estimate and select of an appropriate set of variables for the model. If there is no linear relationship between the regressors, they are said to be orthogonal. In regression when several regressors are highly correlated, this problem is called multicollinearity or collinearity. The multicollinearity is a serious problem that impacts the usefulness of a regression model, because it affects the ability to estimate regression coefficients adequately. Montgomery and Peck (1992) present four primary sources of multicollinearity:

- The data collection method employed
- Constraints on the model or in the population.
- Model specification
- An overdefined model

For the basis-oriented elements it is necessary to know and explain collinearity, because the coefficients are key to provide clues about what happens in the process. If these coefficients are over estimated or have the wrong sign the interpretation of what happens in the system is inappropriate.

2.2.1 Methods to diagnose multicollinearity

The usual approach to address multicollinearity in this context is to eliminate some of the regressors variables from consideration. Mason, Gunst, and Webster (1975) give three specific recommendations: 1) redefine the model in terms of a smaller set of regressors, 2) perform preliminary studies using only subsets of the original regressors, and 3) use principal components type regression methods to decide which regressors to remove from the model. The first two methods ignore the interrelationships between the regressors and consequently can lead to unsatisfactory results. Other approaches include collecting additional data and the use of estimation methods other than least squares that are specifically designed to combat the problem induced by multicollinearity.

Many procedures have been proposed for diagnosing the presence and the degree of multicollinearity. There are established mathematical tools to measure the level of non-orthogonallity among POB elements. These tools include correlation matrix, variance inflation, eigensystem analysis, and singular value decomposition.

The correlation matrix **Q** of A'A used (Webster, Gunst, Mason, 1975) is a very simple and helpful method for detecting dependencies between pairs of elements. Pair wise correlation close to +1 or –1 indicates near linear dependencies between the corresponding columns. When more than two elements of matrix ***A*** are involved in a linear dependency, there is no assurance that any of the pair wise correlations will be large.

Variance inflation factor (VIF) can be represented by the diagonal elements of $C = (Q)^{-1}$, C_{jj} can be written as : $C_{jj} = (1- R_j^2)^{-1}$ (Marquardt 1970), where R_j^2 is the coefficient of determination of z_i, this is the fraction of all variance in one $\boldsymbol{A_j}$ that can be predicted from the other $\boldsymbol{A_j}$. A general rule is that if any of the VIF's is greater than 4.0 one should suspect that multicollinearity might be a problem. If any of the VIF's exceeds 10, the associated regression coefficients are poorly estimated due to multicollinearity presence. There are instances where the maximum VIF is less than 10 and serious non-orthogonallity is present. For this reasons some data analysts prefer to use a maximum VIF of 4 or 5 as cutoff point (Marquardt, 1970) (Snee, 1973) (Belsley, Kuh, & Welsch, 1980).

In small models, the variables producing the non-orthogonallity are usually apparent. It may not be possible to determine in large models with several large VIF's which columns are involved in the relationship. There may be two or more subsets of columns of $\boldsymbol{A}$ that exhibit within-subset multicollinearity. In these situations, some analysts have found that an eigenvalue-eigenvector analysis of the correlation matrix is helpful. Marquardt has shown that there is a direct correspondence between the variance inflation factors and the eigenvalues and eigenvectors of the correlation matrix (Snee, 1973).

The characteristic roots or eigenvalues are used to measure the degree of multicollinearity among vectors in the matrix A'A. If there are one or more near linear dependencies in the matrix, then one or more of the characteristic roots will be small. Some analysts prefer to examine the *condition number* of A'A arguing that instead of

looking at small eigenvalues it is better to consider a ratio of the range of these roots. The condition number is defined as $k = \frac{\lambda_{\max}}{\lambda_{\min}}$ a measure of the spread in the eigenvalues spectrum of A'A. If the condition number is less than 100, there is no serious problem with multicollinearity. Condition numbers between 100 and 1000 indicate moderate to strong multicollinearity and if k exceeds 1000, severe multicollinearity is present (Montgomery and Peck 1992). The condition number helps to identify the presence of dependencies and decomposing A'A help to understand the elements involved in the dependency.

Belsley, Kuh and Welsch (1980) propose a similar approach for diagnosing multicollinearity, based on the singular value decomposition of A. Any n x p matrix A, considered here to be a matrix of n observations on p variant's, may be decomposed as: A_s = UDT' where U'U = T'T = I and D is a diagonal matrix with nonnegative elements μ_k, k = 1, 2, … , b called the singular values of A_s. The singular value decomposition (SVD) of matrix A_s is closely related to the familiar concepts of eigenvalues and eigenvectors, but some of the differences provides insight diagnostics. The degree of dependency relies on how small the minimum singular value is relative to the maximum value. This ratio is known as the *condition index,* which is similar to the condition number that provides a measure against which to measure the smallness. This index is defined by $\eta_k = \frac{\mu_{\max}}{\mu_k}$ $k = 1, \ldots, b$. The estimated variance of each coefficient provides the dependency relation. This is decomposed into a sum of terms each of which is

associated with a singular value. This suggests that an unusually high proportion of the variance of two or more coefficients concentrated in components with same small singular value provides evidence that the corresponding near dependency is causing problems. Belsley, Kuh and Welsh (1980) suggested regressing all but one of the columns involved in the dependency and using the remaining one as the response, in order to understand the relationship among these variables

The methods discussed were previously, used to detect and understand the nature of the dependencies when these exist. It is important to determine the methodology that helps dealing with this condition. There were other mathematical tools used to manage the presence of multicollinearity. Some of these tools were Ridge Regression, Principal Components, Latent Root, Constrained Space Solution, Independent Subsets, Simple Regressions and Stepwise Analysis.

2.2.2 Methods to manage multicollinearity

Hoerl and Kennard (1970) proposed the used of Ridge Regression to improve on the problem of high sensitivity. This approach abandons the usual least square solution and allows a small bias on the estimates to obtain a greatly reduced variance with a larger probability of being close to the true parameter value. When multicollinearity occurs, the variances are large, thus the estimates are far from the true values. Ridge Regression is an effective counter measure because it allows better interpretation of the regression coefficients by imposing some bias on the regression coefficients and shrinking their variances.

Principal Components is other method to obtain biased estimators of regression coefficients using the canonical form. The principal components regression approach combats multicollinearity by using less than the full set of principal components in the model. To obtain the principal component estimator, assume that the regressors are arranged in order of decreasing eigenvalues $\lambda_1 > \lambda_2 > ... \lambda_p > 0$. Suppose that the last s of these eigenvalues is approximately equal to zero. In principal components regression the principal components corresponding to near-zero eigenvalues are removed from the analysis and least squares applied to the remaining components. A simulation study by Gunst and Mason (1977) showed that principal components regression offers considerable improvement over least squares when the data are ill conditioned. They also point out that another advantage of principal components is that exact distribution theory and variable selection procedures are available (Montgomery and Peck 1992).

Other procedure that follows the same philosophy as the principal components method is the Latent Root Regression Analysis developed by Hawkins (1973) and Webster et al. (1974). The procedure forms estimators from the eigenvalues (or latent root) of the correlation matrix of regressor of response variables, Gunst et al. (1976) and Gunst and Mason (1977) indicates that latent root regression may provide considerable improvement in mean Square Error over least squares. Gunst (1979) points out that latent root regression can produce regression coefficients that are very similar to those found by principal components, particularly when there are only one or two strong multicollinearity in $\boldsymbol{A}$.

Principal component and Latent Root require a matrix **A**=[Y : X] would be a matrix of values of responses and regressor variables which were normalized but not centered. This new matrix does not have the meaning of the original matrix of POB's. These methods eliminate the small eigenvalue changing the meaning of the POB as physical patterns. Because these methods change the physical pattern meaning, they were not included in this research.

Birgoren (1997) developed a general constrained space solution (CSS) strategy for solving the multicollinearity problem. This technique restricts the solution space for the regression coefficients, forcing the solution to be consistent with the physical properties of the monitored process. For many processes, there is a highest attainable level for the magnitude of each process error which is very consistent with a process-oriented basis element. The CSS technique uses theses highest errors levels to bind the POBREP coefficients from above or from below, or both; hence it imposes inequality constraints on the feasible space for the POBREP coefficients.

Independent Subsets is a method for constructing several models when the POB are non-orthogonal. If influential groups of largest size *m* are suspected, there are $\sum_{i=1}^{m}\binom{n}{i}$ such cases for which calculations are required. The real challenge is to determine and create the Independent Subsets for the process oriented basis elements. The objective is to understand the linear dependencies using Singular Value Decomposition methodology to form independent sets of basis elements to be analyzed separately using the OLS.

Simple Regression is a suggested method when multicollinearity problem is detected. The idea behind the Simple Regression is perform an individual regression to each POB to determine which pattern arises. The contribution of this method is the individual projection of POB that can provide accurate coefficient estimators with out the impact of other non-orthogonal basis elements.

Another common approach to select a subset of variables from a complex model is stepwise regression. A stepwise regression is a procedure to examine the impact of each variable to the model step by step. The variable that cannot contribute much to the variance explained would be thrown out. There are several versions of stepwise regression such as forward selection, backward elimination, and stepwise. If eliminated elements of the basis are possible without affecting the diagnosis capabilities of POBREP, it is a sensible thing to do. However, in many instances this might not be a good solution since all basis elements might be of interest for performing different diagnostics.

3 Non-Orthogonal Basis Elements

3.1 Introduction

The Process-Oriented Basis Representative (POBREB) methodology relies on the estimation of the basis coefficients $\mathbf{z_i}$. Using POBREP diagnosis is possible, because potential causes are associated with patterns ($\mathbf{A_i}$) having large positive or negative POBREP coefficients ($\mathbf{z_i}$). This methodology uses ordinary least squares (OLS) as the solution method for the linear system:

$$\mathbf{z = (A' A)^{-1}A'x} \qquad \mathbf{(3)}$$

The multivariate quality vector can be represented as a linear combination of these basis elements.

$$\mathbf{X = Az+\varepsilon} \qquad \mathbf{\varepsilon \sim N_b(0,\sigma^2 I)} \qquad \mathbf{(4)}$$

The standard error for the z coefficients is given by $\sqrt{\hat{\sigma} * c_{jj}}$ where sigma square is the means Square Error and c_{jj} is the j^{th} diagonal element of $\mathbf{(A'A)^{-1}}$. The correlation matrix **A'A** is nearly a diagonal matrix if the column vectors are orthogonal to each other. The orthogonallity of process-oriented basis elements plays an important role in the success of the POBREP methodology. When the basis elements are orthogonal, POBREP establishes a reliable and accurate link between multivariate quality vectors and potential process errors characterized by a process-oriented basis. The values of coefficients z_i do not change when other orthogonal columns are added or deleted from process-oriented basis. But in some cases the process-oriented basis elements could not

be orthogonal. When the basis elements are not orthogonal different problems arise regarding the reliability and explanatory power of the POBREP coefficients.

Consider the following non-orthogonal matrix A and quality vector x.

$$A = \begin{bmatrix} 1 & 100 & 20 \\ 1 & 200 & 30 \\ 1 & 150 & 40 \end{bmatrix} \qquad x = \begin{bmatrix} 25 \\ 37 \\ 35 \end{bmatrix}$$

After solving this regression problem using OLS the following are the coefficients estimation: $x = 15.00 - 0.37z_1 + 2.33z_2$. Individual regressions are performed on z_1 and z_2, the coefficient estimation is: $x = 17.30 + 0.10z_1$ and $x = 15.20 + 0.53z_2$. Different coefficients are obtained from the first regression in comparison to the individual ones, z_1 changes the sign and z_2 reduce significantly the value. That is a consequence of the non-orthogonal matrix ***A***. Coefficients in POBREP establish what happens in the process, if these are over estimated or have wrong sign the interpretation could be incorrect.

Multicollinearity refers to linear dependencies among the column vectors of the matrix ***A*.** When dependencies holds exactly, **Det(A) = 0** and this matrix is not invertible. The implementation of the POBREP methodology depends on the degree of orthogonallity among the columns of ***A*.** Three cases are discussed below:

- Process-oriented basis elements are orthogonal and linearly independent
- Process-oriented basis elements are non-orthogonal and linearly independent.
- Process-oriented basis elements non-orthogonal and linearly dependent.

3.2 Matrix A Orthogonal and Linearly Independent

When $\boldsymbol{A}$ is a complete orthogonal basis, the coefficients estimates will be reduced to the solution $\mathbf{z} = \mathbf{A}^{-1}\mathbf{x}$. For some cases it will not be possible or practical to construct a complete basis. In this case, the number of basis elements will be less than the number of measurements in the error quality vector. This incomplete basis case is solved using the Ordinary Least Squares (OLS) solution.

3.3 Matrix A Non-Orthogonal and Linearly Independent

The first step will be to assess if the basis elements are non-orthogonal and/or linearly dependent. In this research, the severity of the multicollinearity problem is based on the Marquardt VIF factor. For each elements j in **A** the VIF is calculated. If $\text{Max}\{\text{VIF}\} \leq 5.0$ the multicollinearity problem will not be considered severe. In this case the estimates for the **z** coefficients are obtained with the least squares method, for the case for the matrix $\boldsymbol{A}$ full rank and orthogonal. If any of the VIF's are greater than 10, the corresponding least squares are likely to be so poorly estimated that a modification of the model or estimation criterion may be required (Snee, 1973). Least square estimates for the **z** coefficients when strong collinearity is present have minimum variance in the class of unbiased linear estimators, but this variance may be large. These large variances results in two practical difficulties when severe multicollinearity is present: (1) The estimators can be very unstable, that is, sensitive to small perturbations in the data; (2) the estimators tend to give results for the coefficients that are too large in magnitude, either positive or negative.

Several methods for managing these problems were discussed in chapter two of this research.

3.4 Matrix A are Linearly Dependent

When an exact linear dependency exists among the elements of A**,** solution can not be obtained including all basis elements since **A'A** cannot be inverted. Understanding the nature of dependencies becomes fundamentals in dealing with this problem, using the same methods mentioned in section 3.3.

3.5 Stencil Printing Operation

An example for this research is the stencil printing operation in electronics manufacturing, the case study obtained from Gonzalez-Barreto (1996). This operation uses a vision system to take measurement of solder paste volume at several locations along a rectangular region on which an integrated circuit will be mounted later. In this process, solder paste is applied by squeegee to all pads on the board. Due to the dimension of the leads for fine-pitch components, solder paste volume on each pad is considered critical for this operation. This measured is related to two of the most common problems in printed circuit board assembly: 1) excess solder may cause short, 2) lack solder might results in an open. This component is measured at 20 different locations on a single part, five measurements are recorded per side as presented in Figure 3.1. Outward arrows represent a positive deviation (excess volume), and inward arrows represent a negative deviation (insufficient volume).

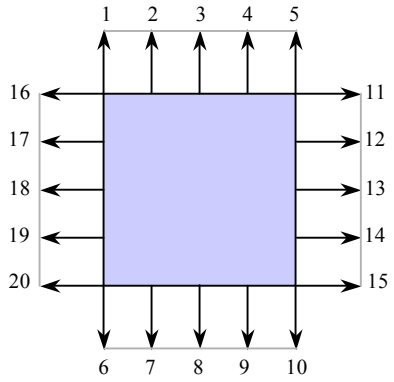


Figure 3.1: Multiple dimension quality characteristic (n = 20 space) corresponding to too much solder paste uniformly at all 20 locations

3.5.1 Generating Process Oriented Basis (POB) Elements

Process-oriented basis elements (POBs) for the stencil printing operation were defined with the help of knowledgeable process engineers. Figure 3.2 shows the specified POBs with their corresponding causes. For this case study, four process oriented basis elements that cause process variation for the stencil printing operation was identified. The following problems in the stencil printing operation are represented by four basis elements: Poor board alignment on the horizontal axis, insufficient paste, poor board alignment on the vertical axis, and squeegee pivot problem. In this operation, lack or excess of solder paste volume at each printing location around the rectangular printing area causes serious quality problems, and each of the four problems gives rise to a distinct pattern of deviations in the amount of solder paste around the rectangular area. Table 3.1 shows the corresponding four vectors for the elements describe above. Gonzalez-Barreto (1996) used two additional elements that showed a non-linear behavior, but in order to simplify this analysis these two POBs will not be included.

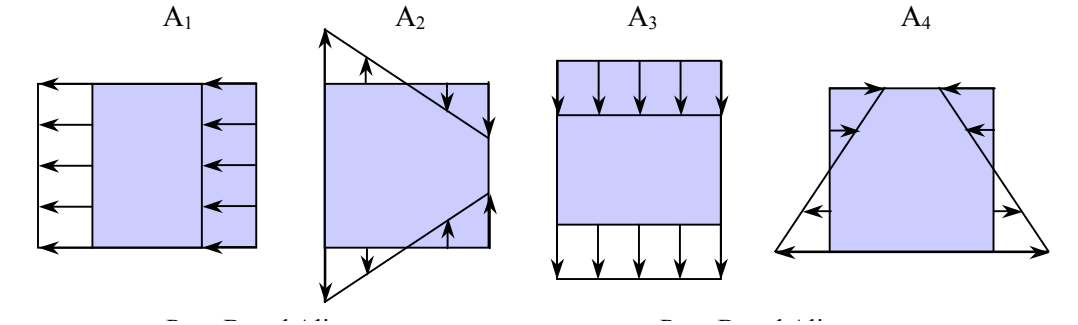


Figure 3.2: Process Oriented Basis for the stencil printing process

Table 3.1: Process Oriented Basis for the stencil printing process

Position	A_1	A_2	A_3	A_4
1	0	1	1	0
2	0	0.5	1	0
3	0	0	1	0
4	0	-0.5	1	0
5	0	-1	1	0
6	0	1	-1	0
7	0	0.5	-1	0
8	0	0	-1	0
9	0	-0.5	-1	0
10	0	-1	-1	0
11	-1	0	0	1
12	-1	0	0	0.5
13	-1	0	0	0
14	-1	0	0	-0.5
15	-1	0	0	-1
16	1	0	0	1
17	1	0	0	0.5
18	1	0	0	0
19	1	0	0	-0.5
20	1	0	0	-1

The basis elements in Table 3.1 are orthogonal, but not always the POBs for a given process will be orthogonal. Figure 3.3 presents non-orthogonal basis elements for the stencil printing process. Process experts can provide expected patterns based on well-understood physical phenomena. Eight additional process oriented basis elements that caused process variation in stencil printing operation were identified for the stencil printing operation. These problems were: inconsistence in squeegee pivot, insufficient paste, and insufficient pressure in one side of the board, snap- off distance, poor board alignment, and stencil problem. Table 3.2 shows the corresponding eight vectors for the elements describe in Figure 3.3.

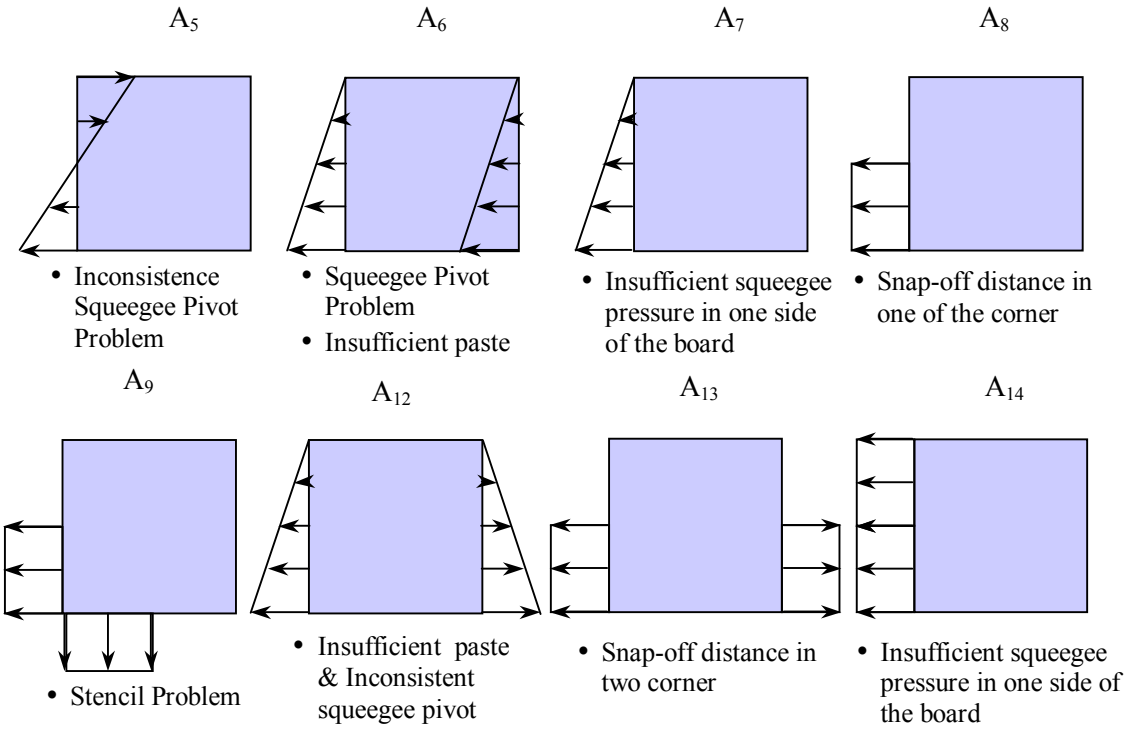


Figure 3.3: Non-Orthogonal Process Oriented Basis for the stencil

Table 3.2: Non-Orthogonal Process Oriented Basis for the stencil printing process

Position	A_5	A_6	A_7	A_8	A_9	A_{12}	A_{13}	A_{14}
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0
7	0	0	0	0	1	0	0	0
8	0	0	0	0	1	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0
12	0	-0.25	0	0	0	0.25	0	0
13	0	-0.5	0	0	0	0.5	1	0
14	0	-0.75	0	0	0	0.75	1	0
15	0	-1	0	0	0	1	1	0
16	-1	0	0	0	0	0	0	1
17	-0.5	0.25	0.25	0	0	0.25	0	1
18	0	0.5	0.5	1	1	0.5	1	1
19	0.5	0.75	0.75	1	1	0.75	1	1
20	1	1	1	1	1	1	1	1

3.5.2 Generating and Validating Quality Vector

In order to compare the methods when multicollinearity is presented among the POB's elements, data vectors are simulated. To present different scenarios of non-orthogonal POBREP basis, data with known structure was generated using Matlab® program (see Appendix A1). For this case only patterns of deviations from the target (bias) were generated. The quality vector was generated using random normal variables.

$$\vec{x} = \sum_{j=1}^{b} \left(\left(\sigma_j \times N[0,1] + \gamma_j \right) \vec{A}_j \right) + \vec{Z} \qquad \textbf{(5)}$$

where,

$\vec{x}$ = quality vector

$N[0,1]$ = random normal variable

σ_j = standard deviation for basis element j

$\vec{A}_j$ = POB element j

γ_j = offset for basis element j

b = # of basis elements under consideration

$\vec{Z} \sim N_b(0, sigma)$

sigma = standard deviation for the error

To validate the generator 60 samples were created. Table 3.3 shows data for the first 30 samples with no bias associated, and for the last 30 samples were created with bias in the first basis elements. Figure 3.4 present data generated using equation 5 when there is not bias in the basis elements, no trend is detected and all the boxes cover zero. The boxplot presented in Figure 3.5 shows the bias in the first basis element. In Figure 3.1 the basis elements 1 presented low value in 11-15 and high values in16 -20, in the same

way the boxplots appear in Figure 3.5. From the manufacturing standpoint this means that poor board alignment or insufficient paste is presented.

Table 3.3: Baseline Case

Elements	First 30 Samples		Last 30 Samples	
	Bias	**Std.Dev.**	**Bias**	**Std.Dev.**
1	0	0.2887	2	0.2887
2	0	0.2887	0	0.2887
3	0	0.2887	0	0.2887
4	0	0.2887	0	0.2887

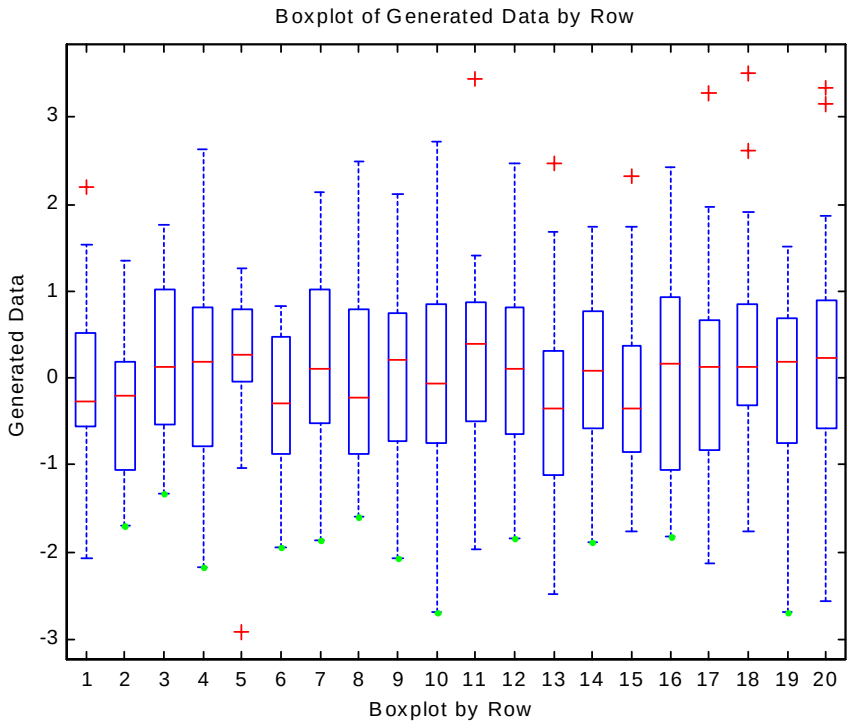


Figure 3.4: Box plots for generate data when no basis element is present

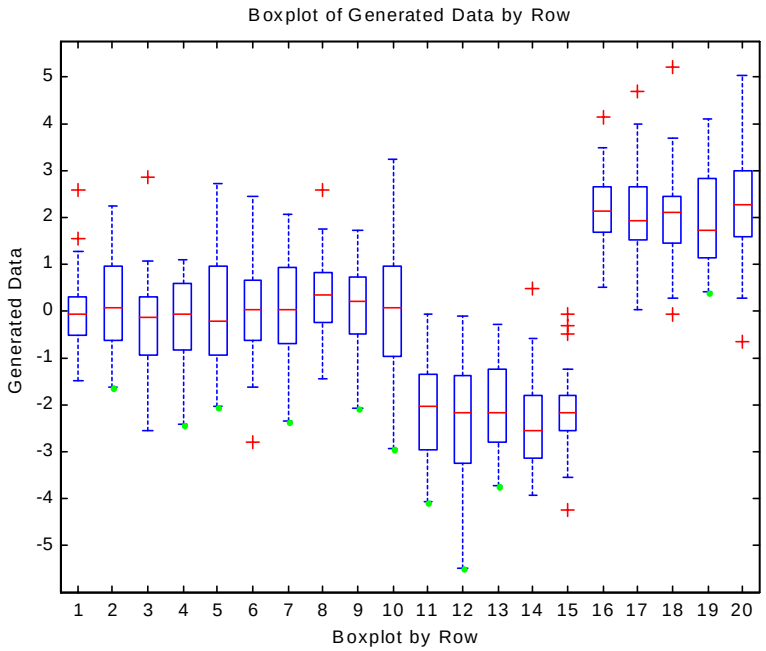


Figure 3.5: Box plots for data generated when basis element 1 offset

Figure 3.6 presents the z representations for each 4 basis with no basis elements active for the first 30 samples and the last 30 with first basis (represented by z_1=2) with the expected bias for the first basis element. z_1 presents the largest coefficients in the last 30 samples and all other representations lie around zero, indicating nonrelevance of those basis elements. This generator was validated for each basis and including more than one basis showing good results. The procedure explained above showed the functionality of the data of the data vector generator.

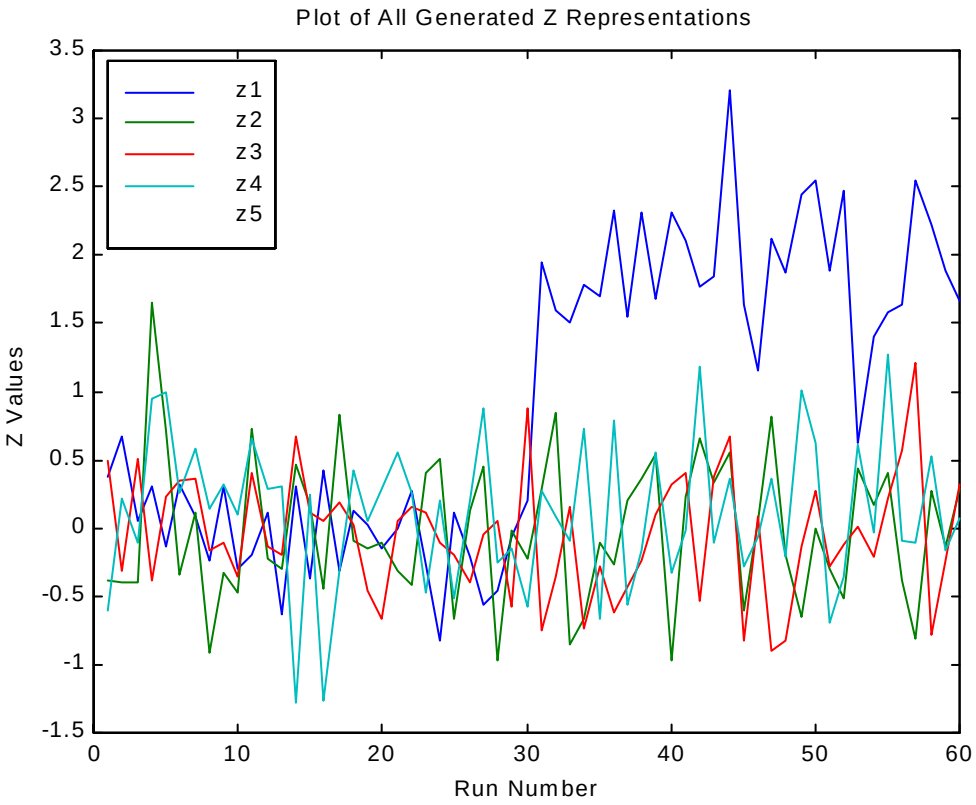


Figure 3.6: zi representations when basis element 1 experience offset.

3.5.3 Diagnosis of Multicollinearity in the Process Oriented Basis Elements

To provide diagnosis about the condition of the ***A*** matrix the Correlation Matrix, the Variance Inflation Factor, Eigenvalue System, and Singular Value Decomposition of the POB elements were evaluated. A program prepared in Matlab® facilitated the evaluation of these methods. In this work, the Variance Inflation Factor by Marquardt (1970) will be used to assess multicollinearity severity and Singular Value Decomposition (Belsley, Kuh, Welsch (1980)) to understand the relation between the basis and the contribution to the variability.

In POBREP one question is if the matrix ***A*** is orthogonal. Consider matrix ***A*** with 7 basis elements from stencil printing problem, $A = [A_1|A_2|A_3|A_4|A_7|A_{12}]$, the first four basis are orthogonal taken from Table 3.1, and the last two non-orthogonal basis taken from Table 3.2. If we observe the matrix **A'A** is evident that the matrix is not orthogonal, given the nonzero off-diagonal elements**.**

$$A'A = \begin{bmatrix} 1.00 & 0 & 0 & 0 & 0.63 & 0 \\ 0 & 1.00 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.00 & -0.44 & -.71 \\ 0.63 & 0 & 0 & -0.44 & 1.00 & 0.63 \\ 0 & 0 & 0 & -0.70 & 0.63 & 1.00 \end{bmatrix}$$

To understand the severity of the multicollinearity issue we use the Variance Inflation Factor (VIF). For the previous matrix ***A*** VIF = [3 1 1 2 5 4]. The maximum Variance Inflation Factor is 5, this indicate that the multicollinearity is not severe because VIF is less than 5.

Using the stencil case study some of these basis elements are grouped in order to construct cases with different multicollinearity severities. The severity of the problem will be classified as follows: VIF $\leq$ 5 not serious, 5 < VIF $\leq$ 10 moderate and VIF > 10 as more serious. Also, relationships that involve two basis elements and more than two basis elements will be incorporated into the study. All this strategy was presented in Figure 1.1. Table 3.4 shows the six cases with their respective POBs that will be considered in this study. Each scenario was evaluated using three cases: A, B and C. Case A adds bias to one of the basis element in the relation, Case B activates two basis elements not related and Case C activates two basis elements involved in the relation. A total of 18 simulations were ran considering six scenarios and three cases.

Table 3.4: Scenarios for the stencil printing process

Relation between Basis	***Variance Inflation Factor***		
	VIF $\leq$ 5	**5 < VIF $\leq$ 10**	**VIF > 10**
Two Basis Elements	**Scenario 1** A=[A_1|A_2|A_3|A_4|A_7|A_{12}] VIF_{max}=5 Relation: A_7 and A_{12} Case A: z_7=2 Case B: z_7=2, z_2=3 Case C: z_7=2, z_{12}=3	**Scenario 2** A=[A_1|A_2|A_3|A_4|A_7|A_8] VIF_{max}=10 Relation: A_7 and A_8 Case A: z_7=2 Case B: z_7=2, z_2=3 Case C: z_7=2, z_8=3	**Scenario 3** A=[A_1|A_2|A_3|A_4|A_5|A_6|A_7|A_9] VIF_{max}=15.5 Relation: A_6 and A_7 Case A: z_7=2 Case B: z_7=2, z_2=3 Case C: z_7=2, z_6=3
More than two Basis Elements	**Scenario 4** A=[A_2|A_3|A_4|A_7|A_{14}] VIF_{max}=5 Relation: A_4, A_7 and A_{14} Case A: z_7=2 Case B: z_7=2, z_2=3 Case C: z_7=2, z_{14}=3	**Scenario 5** A=[A_1|A_2|A_3|A_4|A_7|A_{12}|A_{13}] VIF_{max}=10 Relation: A_7, A_{12} and A_{13} Case A: z_7=2 Case B: z_7=2, z_2=3 Case C: z_7=2, z_{14}=3	**Scenario 6** A=[A_1|A_2|A_3|A_4|A_7|A_8|A_{12}|A_{13}] VIF_{max}=20 Relation: A_7, A_8, A_{12} and A_{13} Case A: z_7=2 Case B: z_7=2, z_2=3 Case C: z_7=2, z_{12}=3

Appendix B presents all the VIF per each case. Scenario 1 has a weak multicollinearity since VIF = 5. In Scenario 2 the higher Variance Inflation Factor is 10, indicating a moderate multicollinearity. Scenario 3 shows a severe multicollinearity with a maximum VIF = 15.5. The next three cases are when there are more that two basis elements related. The maximum VIF in Scenario 4 is 5, there is not severe multicollinearity. Scenario 5 shows a moderate multicollinearity with a maximum VIF = 10. The worst case is Scenario 6 with VIF = 20 as maximum value.

To understand the relation among the basis element and the contribution with the variability, Singular Value Decomposition (SVD) was used to decompose the variance. In the following paragraphs a description of the SVD method used to identify relationship between POB elements will be presented.

The rescaled matrix $\mathbf{A_s}$ can be decomposed as:

$$\mathbf{A_s = UDT'} \qquad \textbf{(6)}$$

Where $\mathbf{U'U = T'T = I}$ and **D** is a diagonal matrix with nonnegative elements μ_k, k=1, …, b called the singular values of $\mathbf{A_s}$.

The degree of dependency relies on how small the minimum singular value is relative to the maximum value. This ratio is known as the *condition index* which is similar to the condition number that provides a benchmark against which to measure the smallness. This index is defined by

$$\eta_k = \frac{\mu_{\max}}{\mu_k} \quad k = 1, \ldots, b \qquad \textbf{(7)}$$

According to Belsley, Kuh, Welsch, (1980), if the condition index is greater than 10 the relation exist. The nature of the dependency can be obtained by looking at the variance of the estimated representations. The variance-covariance matrix of the least square estimator $\mathbf{z=(A_s' A_s)^{-1}A'x}$ is $\mathbf{\sigma^2(A_s' A_s)^{-1}}$. Using the SVD, $\mathbf{A_s = UDT'}$ the variance of z is

$$T(z) = \sigma^2 \sum_j \frac{t_{kj}^2}{\mu_j^2} \tag{8}$$

where the μ_j's are the singular values and the t_{kj}'s are obtained from the T matrix. Note that this decomposes the variance of the components, each associated with one and only one of b singular values (or eigenvalues μ_j^2). Since the singular values appear in the denominator, those components associated with near dependencies, small μ_j, will be large relative to other components. This suggests that an unusually high proportion of the variance of two or more coefficients concentrated in components with same small singular value provides evidence that the corresponding near dependency is causing problems.

Define the k j^{th} variance decomposition proportion as the proportion of the variance of the k^{th} basis element associated with the j^{th} component of its decomposition. These proportions are calculated as

$$\phi_{kj} = \frac{t_{kj}^2}{\mu_j^2} \tag{9}$$

$$\phi_{kj} = \sum_{j=1}^{b} \phi_{kj}, k = 1,...,b \tag{10}$$

Then the variance-decomposition proportions are

$$\pi_{kj} = \frac{\phi_{kj}}{\phi_k}, k, j = 1,...,b \tag{11}$$

Weak dependencies are associated with condition indexes around 5 or 10, whereas moderate to strong relations are associated with condition indexes of 30 to 100 (Belsley, Kuh and Welsch (1980)). In this research the involvement of a basis element in a relationship is established when the condition index exceeds five (5) and the variance decomposition proportion is greater than 0.1.

Table 3.5 presents one condition index 6.8, indicating there is a relation between the basis. Variance proportion shows the poor alignment and insufficient paste, 0.63 and .24 (greater than 0.10) are the related basis elements corresponding to A_7, A_{12} and A_{13}.

Table 3.5: Singular Value Analysis for Scenario 1

Condition Index	**Variance Proportion**					
	A_1	**A_2**	**A_3**	**A_4**	**A_7**	**A_{12}**
1.0000	0.0271	0	0	0.0006	0.0017	0.0003
1.0398	0	0	0.1000	0	0	0
1.1992	0.0028	0	0	0.0343	0.0016	0.0120
1.4705	0	0.2000	0	0	0	0
2.2904	0.0019	0	0	0.1138	0.0114	0.0488
6.8100	0.0683	0	0	0.0012	**0.6253**	**0.2389**

Similar analysis was performed for each scenario and case (see Appendix B) in Table 3.4, there is a summary of the relations per scenario. In Scenario 2 the relation is

between A_7 and A_8. Scenario 3 the relation is with A_6 and A_7. The basis elements A_4, A_7 and A_{14} are related in Scenario 4. Scenario 5 shows the dependency between basis A_7, A_{12} and A_{13}. In Scenario 6 the relation is between A_7, A_8, and A_{12}.

There are three different cases per scenario that consist of a total of 18 configurations to analyze (see Table 3.4). In order to limit the number of configurations, the scenarios only included deviation from the target (bias) while the variation remained constant.

3.5.4 Dealing with Multicollinearity in the Process Oriented Basis (POB) Elements

The methods evaluated in this research when multicollinearity is presented within the POB's are: Ordinary Least Square (OLS), Independent Subsets (IS), Simple Regressions (SR), Ridge Regression (RR) and Constrained Space Solution (CSS).

3.5.4.1 Independent Subsets

Independent Subsets is a methodology to create several orthogonal subsets with POB's. The objective is to understand the linear dependencies that exist in matrix $\boldsymbol{A}$ using Singular Value Decomposition methodology (as presented in section 3.5.3) to form independent sets of basis elements to be analyzed separately using the OLS.

The methodology is summarized as follows:

1. Obtain the singular value decomposition of $\mathbf{A_s}$, and from this calculate:
    a. The conditions indexes η_k
    b. The matrix of variance-decomposition proportions π
2. Determine the number and relative strengths of the near dependencies by the condition indexes exceeding 5.
3. Determine the involvement for the columns with variance-decomposition proportions pi greater than 0.1 associated with the condition indexes exceeding the threshold value.
4. Create subsets of $\boldsymbol{A_s}$ for each independent basis elements.
5. Perform an OLS per each Independent Subset.

6. If a basis is presented in more than one subset, the mean of the z representations in m subsets is considered the estimated POB coefficient.

For the Scenario 1 steps 1 to 3 were determined in section 3.5.3, z_7 and z_{12} were the basis elements related. Two subsets were created, $A_{S1} = [A_1 \mid A_2 \mid A_3 \mid A_4 \mid A_7]$ and $A_{S2} = [A_1 \mid A_2 \mid A_3 \mid A_4 \mid A_{12}]$. To estimate the coefficient OLS was obtained per each subset. For those elements that appear in more than one subset the mean of all estimates was used as the POB coefficient. Consider Scenario 6, Table 3.6 shows the condition index greater than 5 is η_{13} = 19.07 with variance proportion 0.59, 0.32, 0.27 and 0.15 (greater than 0.10). Indicating a relation between basis elements A_7, A_8, A_{12} and A_{13}. Four subsets were created, $A_{S1} = [A_1 \mid A_2 \mid A_3 \mid A_4 \mid A_7]$, $A_{S2} = [A_1 \mid A_2 \mid A_3 \mid A_4 \mid A_8]$, $A_{S3} = [A_1 \mid A_2 \mid A_3 \mid A_4 \mid A_{12}]$ and $A_{S4} = [A_1 \mid A_2 \mid A_3 \mid A_4 \mid A_{13}]$.

Table 3.6: Singular Value Analysis for Scenario 6

Condition Index	**Variance Proportion**							
	A_1	**A_2**	**A_3**	**A_4**	**A_7**	**A_8**	**A_{12}**	**A_{13}**
1.000	0.0033	0	0	0.0042	0.00030	0.0006	0.0007	0.0013
1.007	0.228	0	0	0.0033	0	0	0.0004	0.0008
1.2338	0	0	0.1000	0	0	0	0	0
1.7449	0	0.2000	0	0	0	0	0	0
2.2904	0.0015	0	0	0.1376	0.0002	0.0007	0.00009	0.0026
5.0479	0.0639	0	0	0.0015	0.0159	0.0479	0.0088	0.0147
7.1044	0.0009	0	0	0.0032	0.0321	0.0148	0.0922	0.0671
19.0764	0.0077	0	0	0.0002	**0.5915**	**0.3281**	**0.2720**	**0.1516**

3.5.4.2 Simple Regression

The idea behind Simple Regression is to perform an individual regression to each POB, as presented in equation 12 to determine which pattern arises. The contribution of this method is the individual projection of POB that can potentially provide accurate coefficient estimators without the impact of other non-orthogonal basis elements.

$$z_i = \left(A_i' A_i\right)^{-1} A_i' x \qquad \textbf{(12)}$$

3.5.4.3 Ridge Regression

When multicollinearity occurs, the variances are large and thus far from the true value. Hoerl and Kennard introduced Ridge Regression methodology to deal in the presence of severe multicollinearity. Ridge Regression is an effective counter measure because it allows better interpretation of the regression coefficients by shrinking their variances [Morris, 1982; Pagel & Lunneberg, 1985]. This methodology allows a small bias on the estimates to obtain a greatly reduced variance with a larger probability of being close to true parameter value [Neter, Wasserman, Kunter, 1990]. The **z** estimate is unbiased but imprecise, while estimate **z** using Ridge Regression will result in a more precise estimate with a small bias.

The z coefficients, $\mathbf{z_r}$ are found from a modified version of the least square solutions. Matrix $\mathbf{(A'A+kI)^{-1}, k \geq 0}$ is used instead of the usual $\mathbf{(A'A)^1}$ Note that when k=0 the coefficients provided by the ridge procedure are the least square estimators.

The mean Square error of the z_r ridge estimator is

$$MSE(z_r) = Variance(z_r) + Bias(z_r) \quad \textbf{(13)}$$

The challenge in Ridge Regression is to choose the value of k such that the reduction in the variance term is greater than the increase in the squared bias. Hoerl and Kennard proved that there exists a nonzero value of k for which the MSE of $\mathbf{z_r}$ is less than the variance of the least squares estimator **z**. There is some controversy in the literature as to how determine such a value of k. The ridge trace is an inspection method to determine an appropriate value of k proposed by Hoerl and Kennard. There are other methods suggested by McDonald and Galarneau (1975), Mallows (1973), Waha, Golub and Health (1979), and others. There is no assurance that any of these procedures will produce similar choice for k. Furthermore, there is no guarantee that these methods are superior to straightforward inspection of the ridge trace. In this research ridge trace is used as the method to estimate k the bias parameter.

The ridge trace is a plot of the elements of $\mathbf{z_r}$ versus k, k in the interval [0,1]. Marquardt and Snee (1975) suggest using up to about 25 values of k, spaced approximately logarithmically over the interval [0,1]. If there are severe multicollinearity, the regression coefficients will be unstable in the trace. At some value of k, the ridge estimates $\mathbf{z_r}$ will stabilize. The intention is to select the smallest k, biasing parameter, at which $\mathbf{z_r}$ are stable. Figure 3.7 shows ridge trace for Scenario 6A, by inspection selecting k = 0.2 provides a stable z coefficient. Table F.6 in Appendix F shows k values selected per scenario.

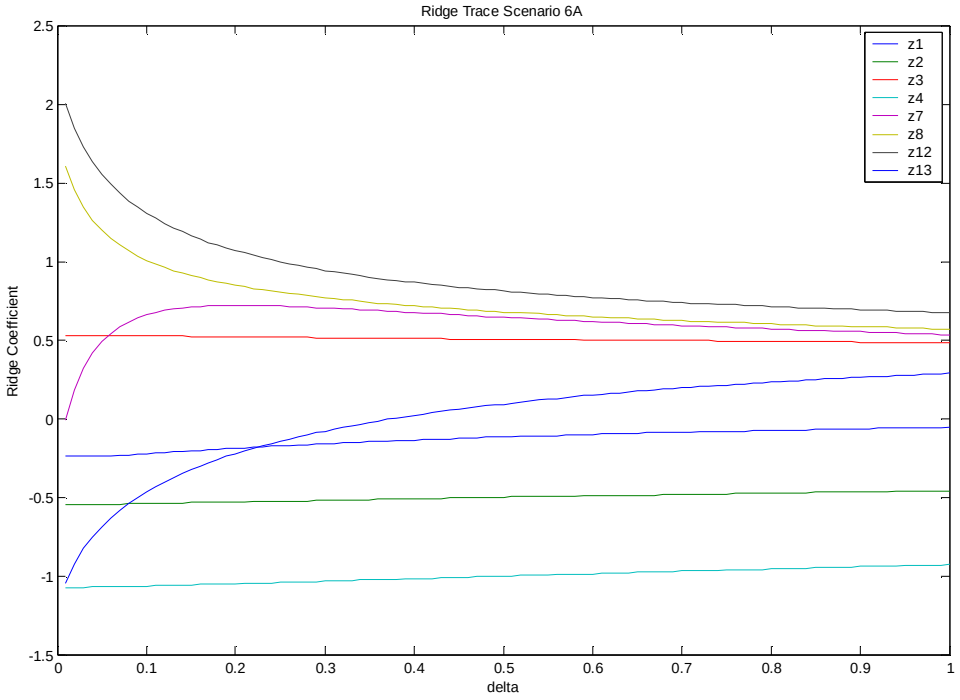


Figure 3.7: Ridge Trace for Scenario 6A

3.5.4.4 Constrained Space Solution

The Constrained Solution Space (CSS) technique restricts the solution space for the regression coefficients, forcing the solution to be consistent with the physical properties of the monitored process.

Common cause variation always exists, and it might contribute to a quality vector in such a way that the least squares solution to $x = Az + e$ produce a z vector that is outside the feasible space of z_o. POBREP coefficient outside the feasible space of z_o is considered unrealistically high in magnitude, and can be avoided by solving a constrained least squares problem: the CSS technique basically proposes solving a constrained least

squares problem to obtain the POBREP coefficients. When a quality vector is decomposed into known patterns using constraints on the magnitude of the coefficients, the contribution of each pattern will be calculated with respect to limits based on the physical process specifications, hence the strength of each problem will be constrained in a realistic way. Consequently, the residual vector ***e*** will contain the real amount process error that cannot be explained by the basis, and patterns in e will reveal potential missing basis elements.

In many processes is possible to specify the highest level a process error can attain for each process problem; these levels can be imposed on the associated process-oriented basis components z_a as lower and upper bounds: $l \leq z_a \leq u$ where l and u are m-dimensional lower and upper bound vectors. The lower and upper bounds can be easily obtained if the basis elements are scaled. Therefore, it will suffice to specify bounds l and u in actual measurement units, which is an easy task for a process engineer.

Let denote these constraints by $l_z \leq z \leq u_z$, where l_z and u_z are k-dimensional lower and upper bound vectors for z. The constrained space defined in this way is the feasible space for the POBREP coefficients, hence the least squares solution should be solved with respect to these constraints. Since the CSS technique involves solving a constrained least squares problem, the solution requires solving a constrained quadratic optimization problem as equation 14 shows. The strategy is to calculate the unconstrained POBREP solution for a new quality vector, and check the feasibility of the solution. If the solution is infeasible, then a problem with bounds will be solved.

$$\min \| \ x\text{-}Az \|^2 \qquad \text{s.t. } l_z \leq z \leq u_z \tag{14}$$

Using Matlab the constrained least squares problem in (14) was solved to obtain $\boldsymbol{z_c}$ by a quadratic programming algorithm. In this case the variation was in constraint as follows:

$$\begin{aligned} & l_z \leq z_i \leq u_z \\ & l_z = z_i - c\sigma \\ & u_z = z_i + c\sigma \qquad c = 1, \quad 2, \quad \ldots \end{aligned} \tag{15}$$

To reduce complexity in CSS, this research considers and compares CSS at 1 sigma, 3 sigma and 6 sigma.

3.6 Criteria for comparing POBREP solutions

POBREP establishes a reliable link between multivariate quality vectors and potential process errors when POB's are orthogonal. Coefficients developed from non-orthogonal data will be too large in absolute value and will often have wrong signs, because strong multicollinearity may result in large variances. Six cases are presented in Table 3.4 to compare estimation methods for POB coefficient.

The best method that deals with non-orthogonallity should reduce the variability and reach the theoretical POB coefficients. These methods should show their strength to deal with non-orthogonal elements, when there are different relationship and severities among the relationships. The data generated with known structure facilitated this comparison. The measures used in this research to compare the methods will be labeled Square Error and Count. These measurements pretend to assess the ability of the method to detect basis elements activity in presence of relationships among matrix ***A*** basis element. Research had been conducted to examine the effectiveness of biased estimators and to attempt to determine which procedures perform best. Gunst, Webster and Mason (1976) used the Square Errors (SE), as presented in equation 16, to compare methodologies accuracy.

$$SE(\hat{z}) = \sum_{i=1}^{p} (z_i - \hat{z}_i)^2 \qquad \textbf{(16)}$$

Each methodology was compared in terms of the lower Square Error of the estimated coefficients z_i.

To measure if the coefficients reach the target for known bias, consider a measure Count. Count is another useful statistics in comparing the methods by the number of times z_i is contained in the following confidence interval.

$$LCI_i = z_i - 1.645 \times \frac{\sigma}{\sqrt{n}}$$
$$UCI_i = z_i + 1.645 \times \frac{\sigma}{\sqrt{n}} \quad \textbf{(17)}$$

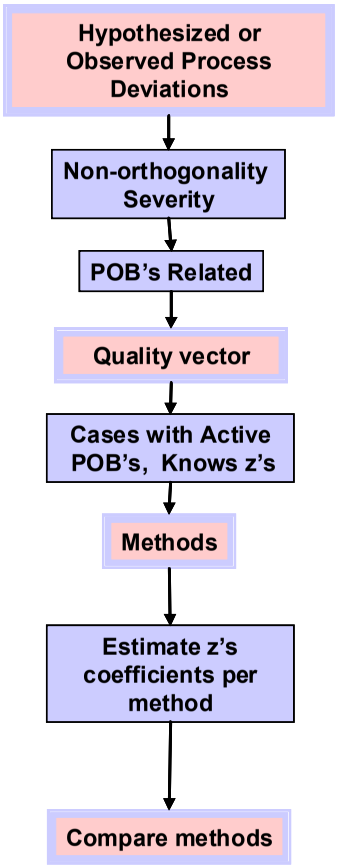
$$Count = \sum_{i=i}^{p} \sum_{j=1}^{n} C_{ij} \qquad C_{ij} = \begin{cases} 1 & LCI_i < \hat{z}_{ij} < UCI_i \\ 0 & \hat{z}_{ij} < LCI_i \\ 0 & \hat{z}_{ij} > UCI_i \end{cases} \quad \textbf{(18)}$$

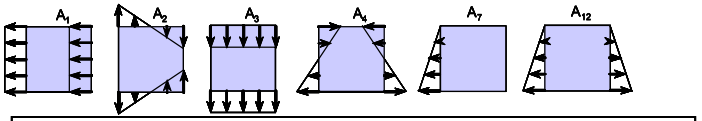
This interval use knows $\hat{z}$, and assumes normality due $z_{\alpha/2} = 1.645$ at 90 % confidence and use the standard deviation of quality vector. The suffix i is the basis and j is the run number. There is a confidence interval per basis and it compares z coefficient against each confidence interval. Each methodology was compared in terms of the higher number of estimated coefficients z_i included in the interval. The method which posses the minimum SE and maximum Count will be the best to deal with non-orthogonallity.

4 Methods Comparison using Stencil Printing Scenarios

Since the simulation, estimation and analysis for each case in each scenario will follow the same procedure, details for graphical and analytical results will be provided in this chapter for Scenario 1, lowest severity, and Scenario 6, highest severity. Details for all other four scenarios are included in Appendix C to Appendix F. Figure 4.1 shows Scenario 1 strategy: (1) observed process deviation and create POB matrix $\boldsymbol{A}$, (2) identify non-orthogonal severity using VIF, $VIF_{max}=5$, (3) determine POB's related, A_7 and A_{12}, (4) generate error pattern with known bias, Case A: bias in one POB related $z_7=2$, Case B: bias in two POB not related $z_7=2$, $z_2=3$, Case C: bias in two POB related $z_7=2$, $z_{12}=3$, (5) estimate $\boldsymbol{z}$ coefficients using seven methods, (6) compare each method using Square Error (SE) and Count measure.

To present the results some graphs were created: run chart per method and a run chart presents the Square Error per method. Scenarios 1 and 6 were explained in detail. The same results for Scenarios 2 to 5 were included in Appendix C-F.





Process Oriented Basis Matrix **A:** $\mathbf{A} = [\, \mathbf{A}_1 \mid \mathbf{A}_2 \mid \mathbf{A}_3 \mid \mathbf{A}_4 \mid \mathbf{A}_7 \mid \mathbf{A}_{12} \,]$

VIF < 5 VIF = [3 1 1 2 5 4]

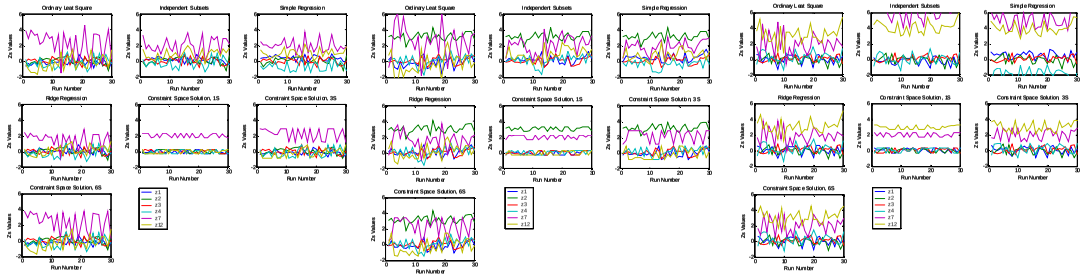
Two basis related $3A_{i7} + 2A_{j12} = 0$

Generate Error Pattern X=[ ]

Case A: Bias in one POB related $z_7 = 2$

Case B: Bias in two POB not related $z_7 = 2, z_2 = 3$

Case C: Bias in two POB related $z_7 = 2, z_{12} = 3$



Methodology	A	B	C	Mean
	SE	SE	SE	
Ordinary Least Square	4.53	7.30	3.96	5.26
Independent Subsets	3.15	4.04	21.93	9.71
Simple Regression	3.27	3.07	16.07	7.47
Ridge Regression	2.27	3.15	2.64	2.68
Constraint Space Solution 1S	0.36	0.37	0.36	0.37
Constraint Space Solution 3S	1.71	1.92	1.70	1.78
Constraint Space Solution 6S	3.10	3.35	2.67	3.04

Methodology	A	B	C	Mean
	Count	Count	Count	
Ordinary Least Square	169	164	168	167
Independent Subsets	176	167	143	162
Simple Regression	175	173	130	159
Ridge Regression	175	174	175	175
Constraint Space Solution 1S	180	180	180	180
Constraint Space Solution 3S	180	180	180	180
Constraint Space Solution 6S	169	180	180	176

Figure 4.1: Scenario 1 Strategy and Results

4.1 Results Scenario 1

Scenario 1A as presented in Figure 4.2 shows the methodology comparison for this POB's. POB's values for z_1 to z_3 were close to 0, with the exception of z_4 with values below the target. All the methods provided the expected results for the first four POB's, since these POB's were orthogonal. POB's z_7 and z_{12} presented high variability and high values due to the multicollinearity problem.

Ordinary Least Square (OLS) presented high variability and values greater than 2 for z_7. Independent Subsets (IS) showed high variability in z_7, but shift z_{12} above the target. Simple Regression (SR) showed reduction in the variability for z_7, but z_{12} was above the target and z_4 was below the target. Ridge Regression (RR) reduced variability, but stayed always below the target for z_7. The z_{12} and z_4 in RR reduced variability and it was in the target. The Constraint Space Solution (CSS) presented high variability for z_7, but controlled well the variability of the other POB's. This methodology showed better results when constraint space was 3S and 1S.

The Square Error (Figure 4.3) showed similar results for all the methods, but RR and CSS had the lowest values in average as presented in Table 4.1. CSS at 6S showed highest mean SE than RR, but lowest standard deviation. SE for CSS at 6S showed similar results as IS. Table 4.2 shows Count measure with better results for CSS at 1S and 3S, but very similar results for RR, IS, SR and OLS.

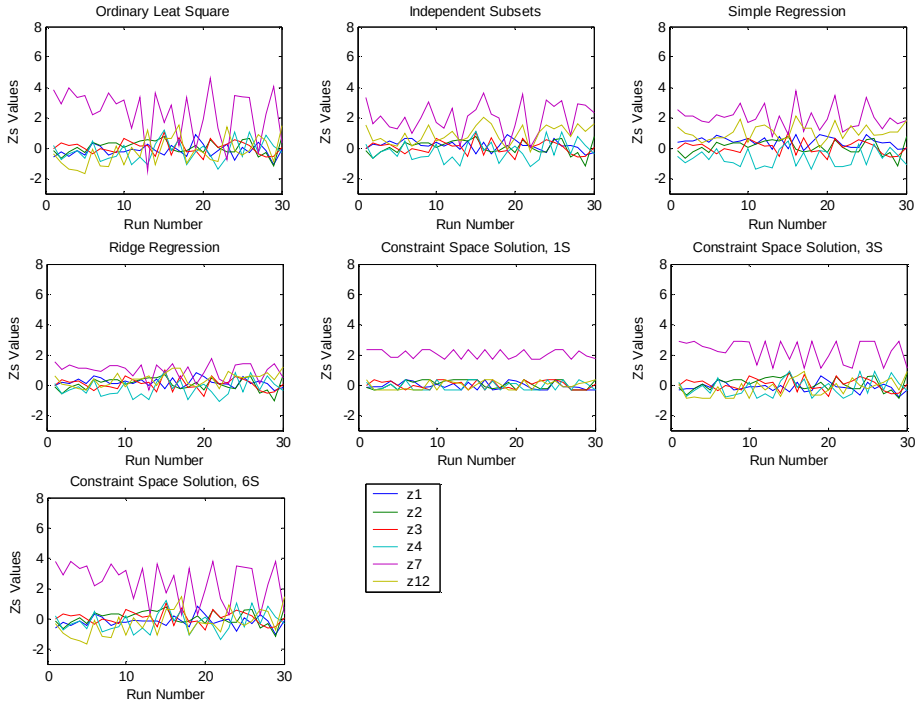


Figure 4.2: POB's values per methods in Scenario 1A

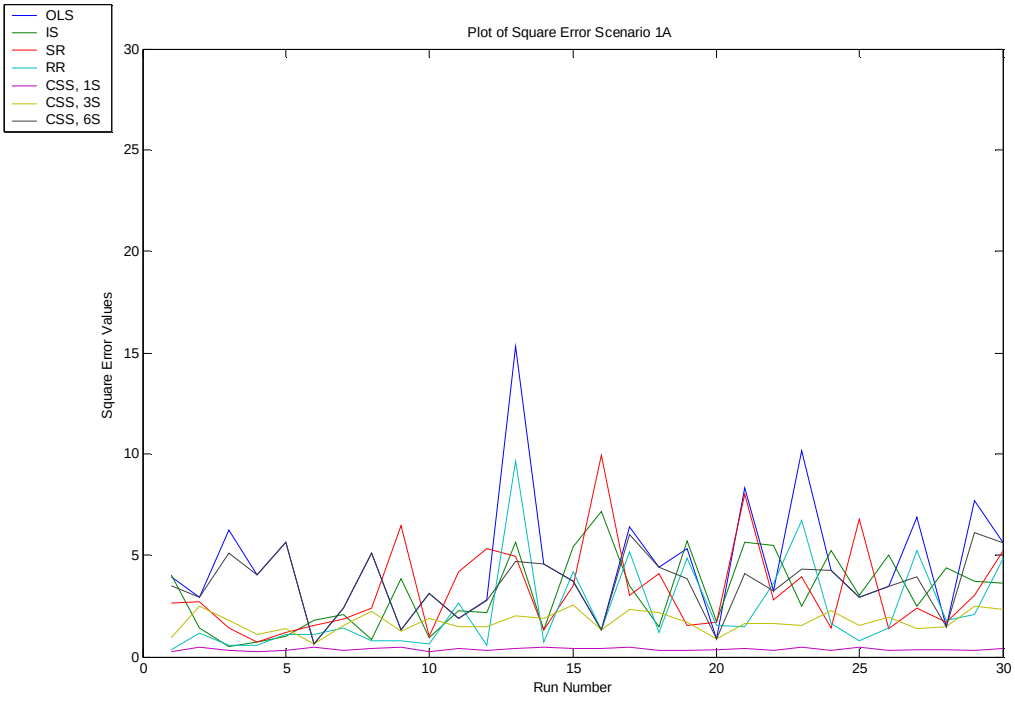


Figure 4.3: Square Error Methods Comparison for Scenario 1

Table 4.1: Square Error measure Scenario 1A

Methodology	Square Error	
	Mean	Std. Dev.
Ordinary Least Square	4.53	3.06
Independent Subsets	3.15	1.87
Simple Regression	3.27	2.26
Ridge Regression	2.27	2.03
Constraint Space Solution 1S	0.36	0.07
Constraint Space Solution 3S	1.71	0.51
Constraint Space Solution 6S	3.10	1.16

Table 4.2: Count measure Scenario 1A

Methodology	z_1	z_2	z_3	z_4	z_7	z_{12}	Total
Ordinary Least Square	30	30	30	30	20	29	169
Independent Subsets	30	30	30	30	27	29	176
Simple Regression	30	30	30	30	29	26	175
Ridge Regression	30	30	30	30	25	30	175
Constraint Space Solution 1S	30	30	30	30	30	30	180
Constraint Space Solution 3S	30	30	30	30	30	30	180
Constraint Space Solution 6S	30	30	30	30	20	29	169

Scenario 1B in which the offset was for z_7=2 and z_2=3 (not related variables) presented all the methods estimating z_2 on target (Figure 4.4). But also there was an increment in z_7 and z_{12} variability in comparison with Scenario 1A. All the methods were very similar detecting bias and variability as shows Figure 4.4. Ordinary Least Square (OLS) presented high variability for z_7 and z_{12}. Independent Subsets (IS) showed less variability in z_7 in comparison with OLS but z_7 and z_{12} were not on target. Simple Regression (IS) showed for z_7 reduction in the variability, but z_{12} was above the target and z_4 was below the target. Ridge Regression (RR) showed results similar to IS and SR, but RR controlled better z_{12} than the other methods. Constraint Space Solution at 6S (CCS 6S) presented high variability for z_7, but controlled well the variability of the other POB's. This methodology showed better results when constrain space at 3S and 1S.

Square Error (Figure 4.5) presented better results in CSS at 1S and 3S. Table 4.4 shows lower mean SE for SR and RR than CSS at 6S, but CSS at 6S had lowest standard deviation. OLS showed highest mean and standard deviation. Table 4.2 shows Count measure with better results for CSS at 1S and 3S, but very similar results for RR and SR.

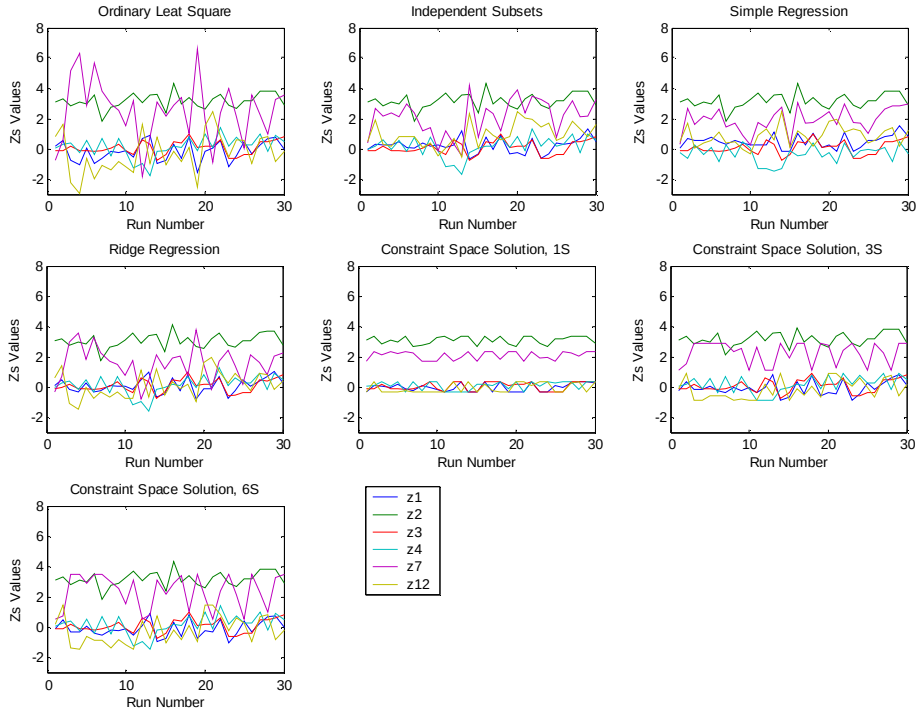


Figure 4.4: POB's values per methods in Scenario 1B

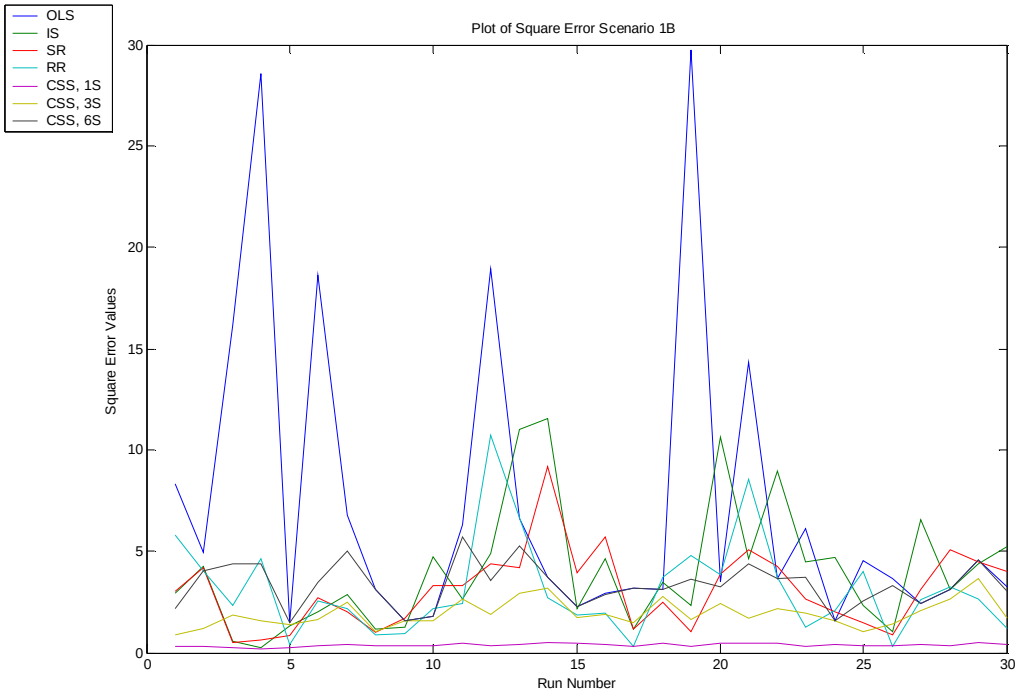


Figure 4.5: Square Error Methods Comparison for Scenario 1B

Table 4.3: Square Error measure Scenario 1B

Methodology	Square Error	
	Mean	Std. Dev.
Ordinary Least Square	7.30	7.68
Independent Subsets	4.04	3.05
Simple Regression	3.07	1.89
Ridge Regression	3.15	2.38
Constraint Space Solution 1S	0.37	0.08
Constraint Space Solution 3S	1.92	0.66
Constraint Space Solution 6S	3.35	1.09

Table 4.4: Count measure Scenario 1B

Methodology	z_1	z_2	z_3	z_4	z_7	z_{12}	Total
Ordinary Least Square	30	30	30	29	20	25	164
Independent Subsets	30	30	30	29	25	23	167
Simple Regression	30	30	30	30	27	26	173
Ridge Regression	30	30	30	30	25	29	174
Constraint Space Solution 1S	30	30	30	30	30	30	180
Constraint Space Solution 3S	30	30	30	30	30	30	180
Constraint Space Solution 6S	30	30	30	30	30	30	180

Bias in Scenario 1C were in z_7=2, z_{12}=3 two related variables. Figure 4.6 shows the Independent Subset (IS) and Simple Regression (SR) increase the value of z_7 and z_{12} above the target. Simple Regression also increased the value for z_4. Constraint Space Solution (CSS) at 6S, Ordinary Least Square (OLS) and Ridge Regression (RR) presented similar results. As in previous scenarios CSS showed the best results when the space was constraint at 3S and 1S.

The Square Error (SE) is showed in Figure 4.7 CSS and RR had in average the lowest values (Table 4.5). CSS at 3S and 1S showed better results than other methods. SE for CSS at 6S showed similar results as RR. OLS showed better results than IS and SR. Table 4.6 shows the Count with better results for CSS and RR. Simple Regression and IS had the lowest Count since their ability to detect z_7 on target was 2 and 0.

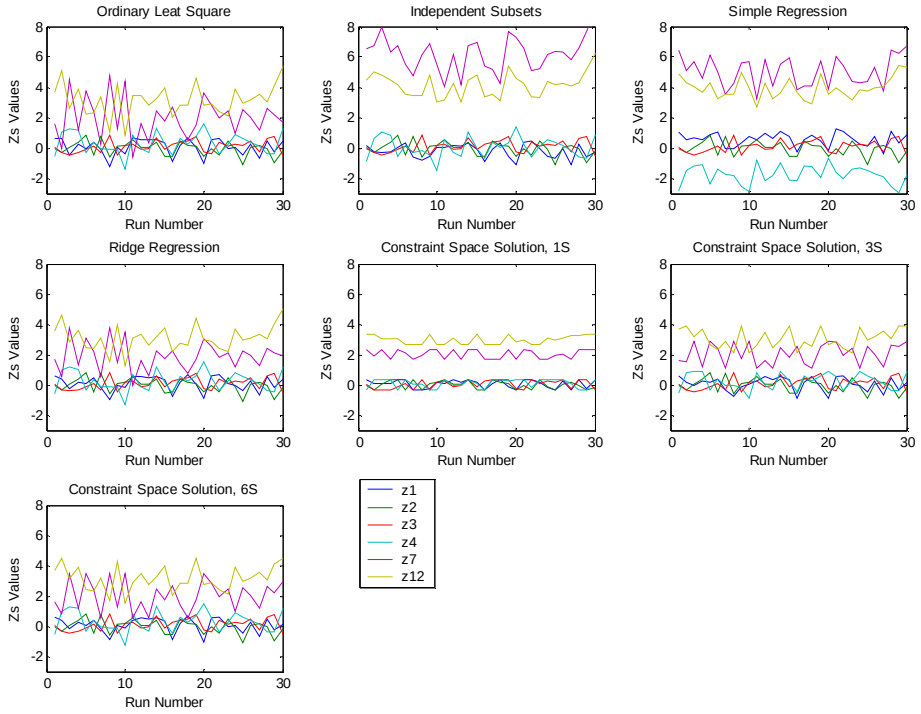


Figure 4.6: Methodology for Scenario 1C

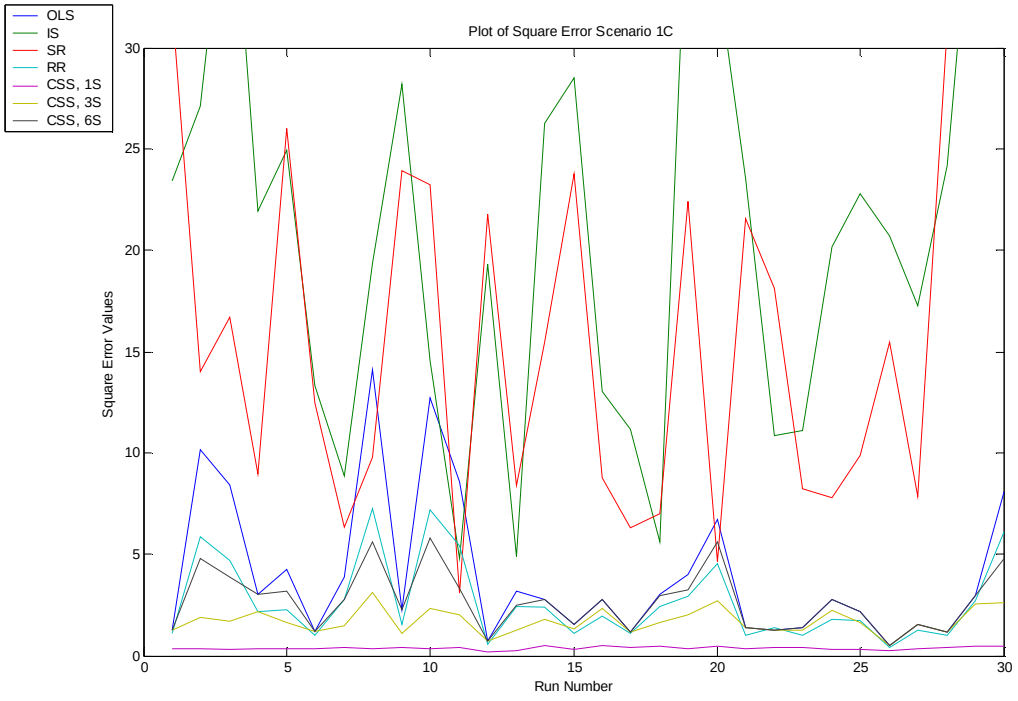


Figure 4.7: Square Error Methods Comparison for Scenario 1C

Table 4.5: Square Error measure Scenario 1C

Methodology	Square Error	
	Mean	Std. Dev.
Ordinary Least Square	3.96	3.64
Independent Subsets	21.93	12.08
Simple Regression	16.07	9.27
Ridge Regression	2.64	1.99
Constraint Space Solution 1S	0.36	0.08
Constraint Space Solution 3S	1.70	0.61
Constraint Space Solution 6S	2.67	1.49

Table 4.6: Count measure Scenario 1C

Methodology	z_1	z_2	z_3	z_4	z_7	z_{12}	Total
Ordinary Least Square	30	30	30	30	22	26	168
Independent Subsets	30	30	30	30	0	23	143
Simple Regression	30	30	30	13	2	25	130
Ridge Regression	30	30	30	30	27	28	175
Constraint Space Solution 1S	30	30	30	30	30	30	180
Constraint Space Solution 3S	30	30	30	30	30	30	180
Constraint Space Solution 6S	30	30	30	30	30	30	180

Table 4.7 and 4.8 presents Count and SE measures for Scenario 1. Constraint Space Solution (CSS) reported the highest Count number follow for Ridge Regression (RR) and Ordinary Least Square (OLS). Independent Subset (IS) and Simple Regression (SR) although showed good numbers in Case A and B, did not detect the bias when two related basis elements were involved. Square Error (SE) presented similar results as Count. The lowest SE was CSS at 1S and 3S. RR had lower SE results than CSS at 6S. OLS had better results than IS and SR due in Scenario 1C IS and SR had higher SE. The method to reduce SE with highest values of Count were CSS and RR.

Table 4.7: Count measure Scenario 1 Summary

Methodology	A	B	C	Mean
	Count	Count	Count	
Ordinary Least Square	169	164	168	167
Independent Subsets	176	167	143	162
Simple Regression	175	173	130	159
Ridge Regression	175	174	175	175
Constraint Space Solution 1S	180	180	180	180
Constraint Space Solution 3S	180	180	180	180
Constraint Space Solution 6S	169	180	180	176

Table 4.8: Mean Square Error measure Scenario 1 Summary

Methodology	A	B	C	Mean
	MSE	MSE	MSE	
Ordinary Least Square	4.53	7.30	3.96	5.26
Independent Subsets	3.15	4.04	21.93	9.71
Simple Regression	3.27	3.07	16.07	7.47
Ridge Regression	2.27	3.15	2.64	2.68
Constraint Space Solution 1S	0.36	0.37	0.36	0.37
Constraint Space Solution 3S	1.71	1.92	1.70	1.78
Constraint Space Solution 6S	3.10	3.35	2.67	3.04

4.2 Results Scenario 6

Scenario 6 is the worst case scenario due VIF was above 10 and there were more that two variables related. In Scenario 6A, Figure 4.8 shows z_{12}, z_{13}, z_7 and z_8 as the POB's involved in the relation, in this one the offset was only in z_7=2. POB's included in the relations presented high variability. Ordinary Least Square (OLS) had the highest variability and did not detect the offset POB. Independent Subset (IS) showed less variability than OLS but the detection was not clear and z_{12} was above the target. Simple Regression (SR) showed less variability than IS and OLS. Ridge Regression (RR) had the lowest variability but the offset in z_7 was not evident. Constraint Space Solution (CSS) although with high variability in z_7 detected the offset. As presented in Figure 4.8 the offset was nearer to the target and with low variability when the constraint was 3S or 1S.

Figure 4.9 shows Square Error (SE) for CSS, RR and SR with the lowest values. Square Error results for SR and RR was influenced due low variability in all POB's, but there was not detection as show Figure 4.8. OLS had the highest SE mean and standard deviation, followed for IS. CSS at 6S showed highest mean SE than RR, but lowest standard deviation. The SE was reduced when CSS at 3S and 1S. Table 4.10 shows Count measure with better results for CSS, than RR and SR

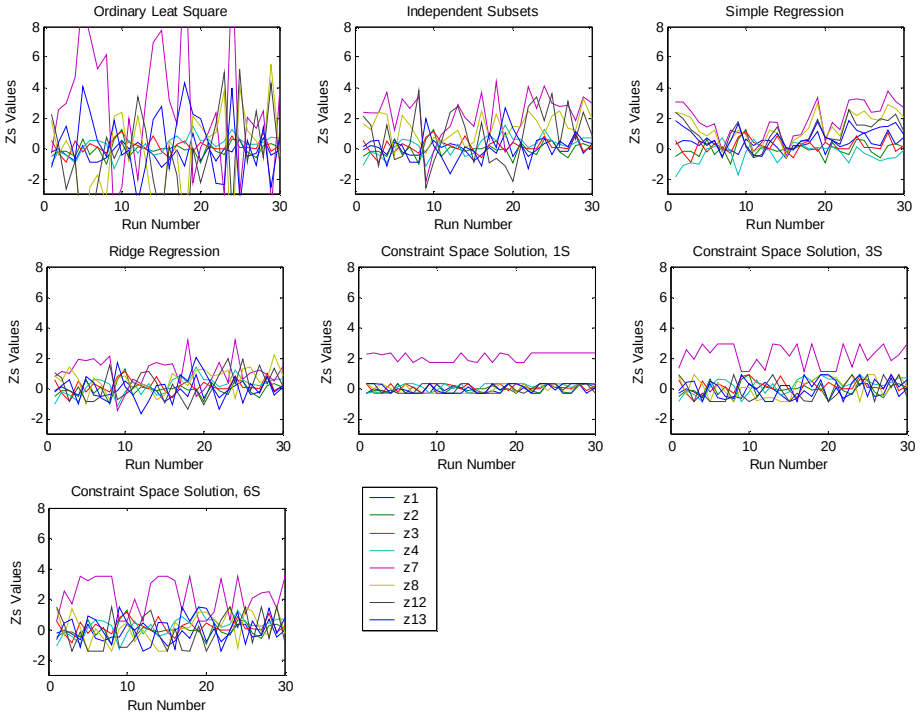


Figure 4.8: POB's values per methods in Scenario 6A

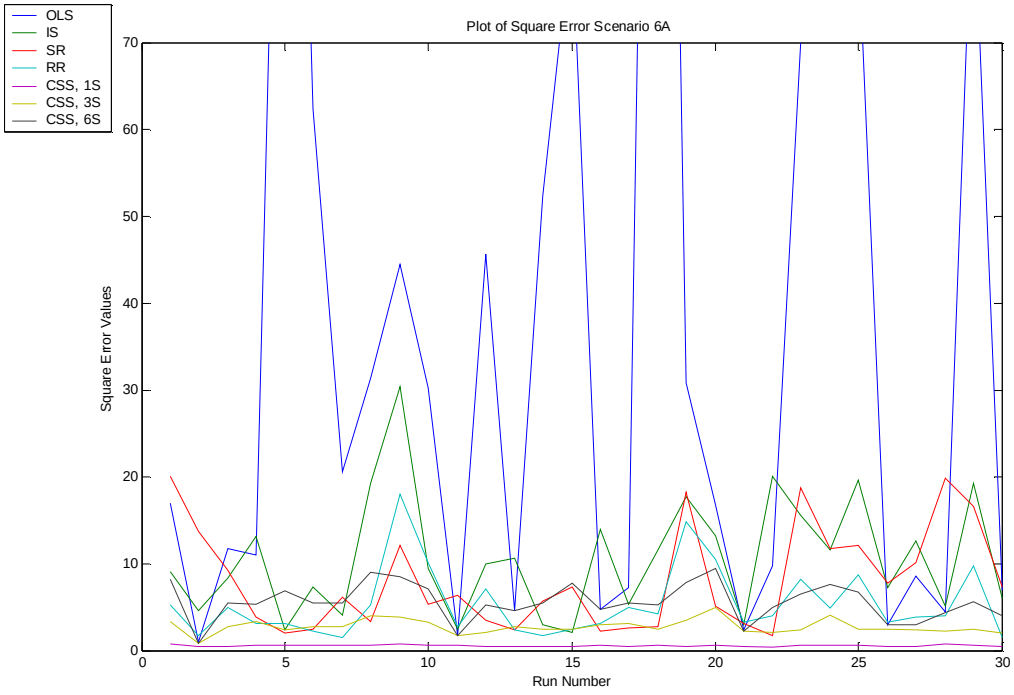


Figure 4.9: Square Error Methods Comparison for Scenario 6A

Table 4.9: Square Error measure Scenario 6A

Methodology	Square Error	
	Mean	Std. Dev.
Ordinary Least Square	44.64	58.86
Independent Subsets	10.56	6.77
Simple Regression	8.08	5.89
Ridge Regression	5.31	4.00
Constraint Space Solution 1S	0.52	0.08
Constraint Space Solution 3S	2.72	0.79
Constraint Space Solution 6S	5.57	2.11

Table 4.10: Count measure Scenario 6A

Methodology	z_1	z_2	z_3	z_4	z_7	z_8	z_{12}	z_{13}	Total
Ordinary Least Square	30	30	30	30	11	14	14	15	174
Independent Subsets	30	30	30	30	22	19	17	27	205
Simple Regression	30	30	30	28	26	18	22	26	210
Ridge Regression	30	30	30	30	24	29	28	27	228
Constraint Space Solution 1S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 3S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 6S	30	30	30	30	30	30	30	30	240

In Scenario 6B the offset was in $z_2 = 3$ and $z_7 = 2$. All methods estimated z_2 as expected (Figure 4.10). But the related variables presented high variability and off target. Variability increased in comparison to Scenario 6A. Ordinary Least Square (OLS) presented high variability and not detection. Independent Subsets (IS) showed all the variables related to z_7 shifted. Simple Regression (SR) increased variability but again all the variables related to z_7 were shifted. Ridge Regression (RR) showed low variability, but z_7 was confounded with other POB. Constrain Space Solution (CSS) at 6S reported similar results as RR. CSS at 3S and 1S showed better results than other methods.

Square Errors (SE) showed similar results as previous scenarios. Figure 4.11 shows as the better methods: RR, SR and CSS. A notable reduction in SE occurred when CSS was 3S or 1S (Table 4.11). OLS had the highest SE mean and standard deviation, followed for IS. CSS at 6S showed highest mean SE than RR, but lowest standard deviation. Table 4.12 shows the Count with better results for CSS, them RR and SR.

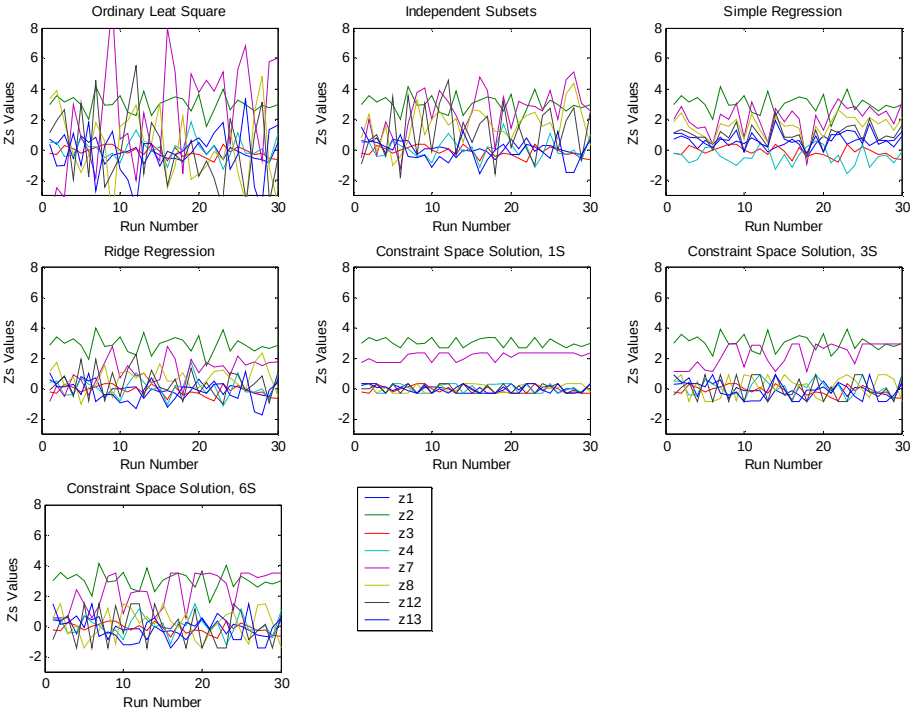


Figure 4.10: POB's values per methods in Scenario 6B

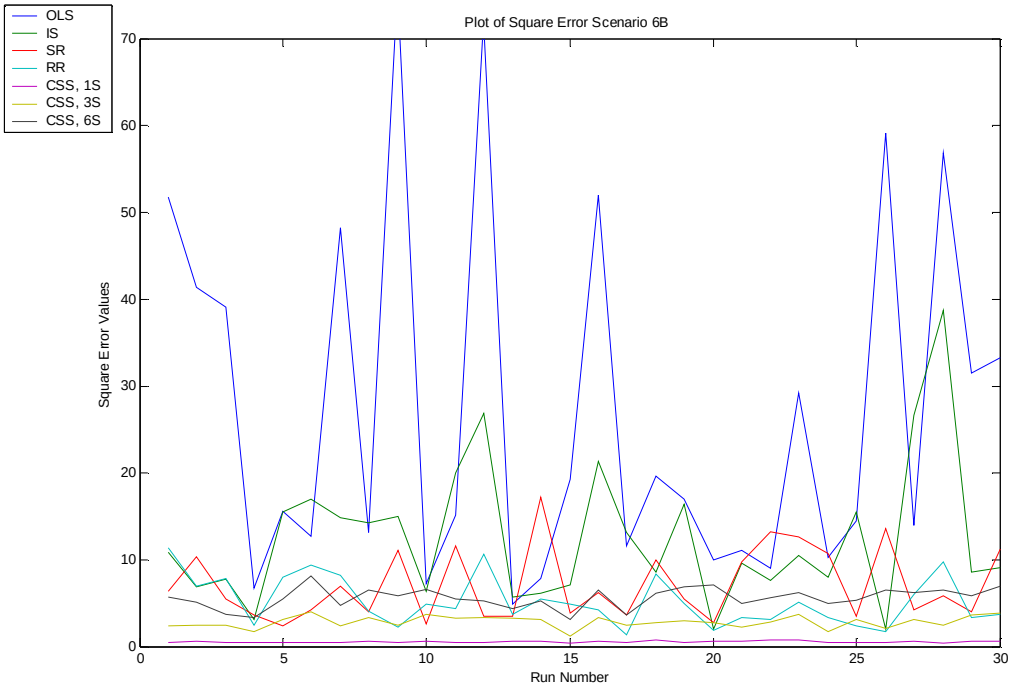


Figure 4.11: Square Error Methods Comparison for Scenario 6B

Table 4.11: Square Error measure Scenario 6B

Methodology	Square Error	
	Mean	Std. Dev.
Ordinary Least Square	26.96	20.95
Independent Subsets	12.45	8.12
Simple Regression	7.07	4.09
Ridge Regression	5.19	2.80
Constraint Space Solution 1S	0.50	0.09
Constraint Space Solution 3S	2.80	0.67
Constraint Space Solution 6S	5.55	1.17

Table 4.12: Count measure Scenario 6B

Methodology	z_1	z_2	z_3	z_4	z_7	z_8	z_{12}	z_{13}	Total
Ordinary Least Square	30	30	30	29	7	15	12	22	175
Independent Subsets	30	30	30	29	18	15	15	30	197
Simple Regression	30	30	30	30	28	16	27	28	219
Ridge Regression	30	30	30	29	25	27	27	29	227
Constraint Space Solution 1S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 3S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 6S	30	30	30	30	30	30	30	30	240

Bias in Scenario 6C were z_7=2, z_8=3 two related variables. Figure 4.12 presents Ordinary Least Square (OLS) with high variability and Simple Regression (SR) with small variability, but, SR shifted z_1, z_4, z_8, z_{12}, and z_{13}. Independent Subsets (IS) shifted z_7 and increased variability. Ridge Regression (RR) showed low variability with better bias identification. Constraint Space Solution at 6S (CSS) showed the offset but had high variability in z_7, z_8. As in the other scenarios CSS showed the best results when the space was constrained at 3S and 1S.

Figure 4.13 shows Square Error (SE) for RR and CSS with the lowest values. Table 4.13 shows CSS at 6S with higher SE mean than RR, but lowest standard deviation. The SE was reduced for CSS at 3S and 1S. OLS and SR had highest SE mean and standard deviation than IS. Table 4.14 shows Count with better results for CSS and RR. SR and IS had the lowest Count due their poor ability to detect z_7 on target..

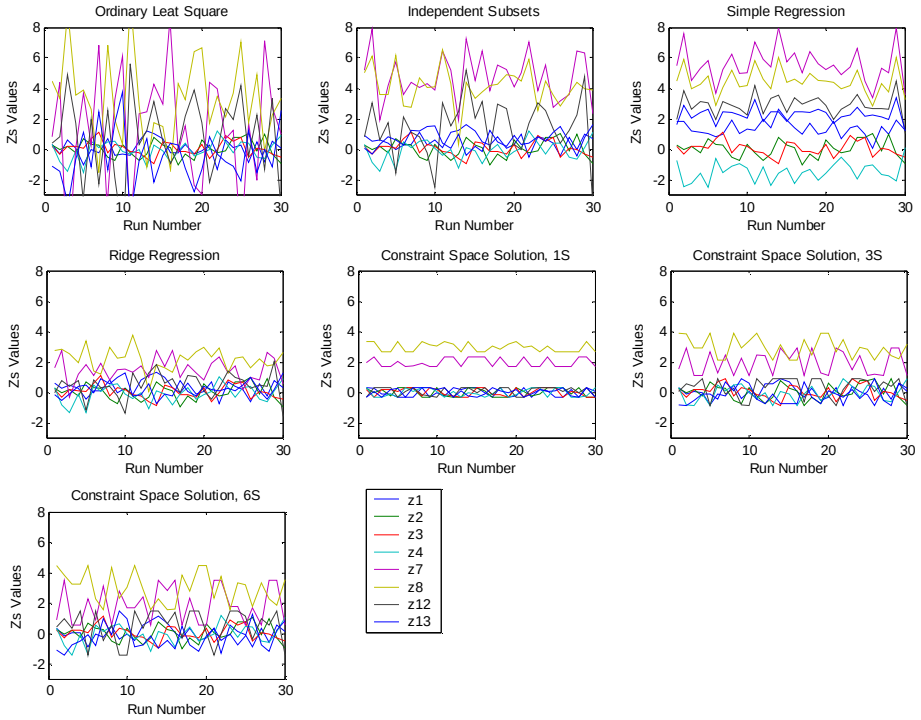


Figure 4.12: POB's values per methods in Scenario 6C

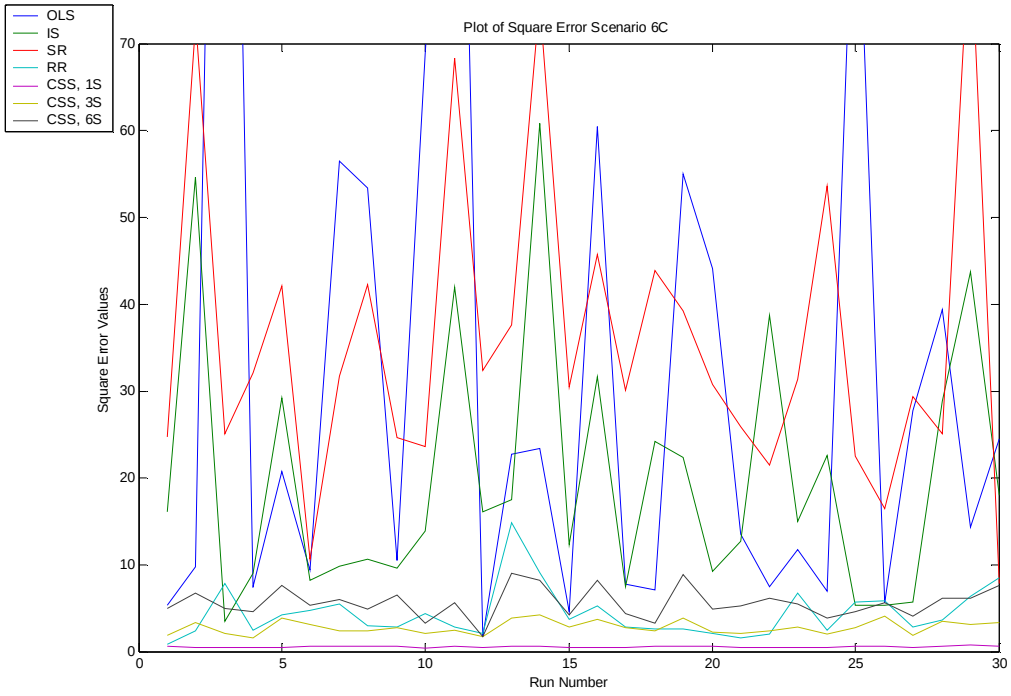


Figure 4.13: Square Error Methods Comparison for Scenario 6C

Table 4.13: Square Error measure Scenario 6C

Methodology	Square Error	
	Mean	Std. Dev.
Ordinary Least Square	35.13	43.59
Independent Subsets	20.10	15.01
Simple Regression	35.99	18.54
Ridge Regression	4.37	2.90
Constraint Space Solution 1S	0.50	0.08
Constraint Space Solution 3S	2.75	0.75
Constraint Space Solution 6S	5.55	1.72

Table 4.14: Count measure Scenario 6C

Methodology	z_1	z_2	z_3	z_4	z_7	z_8	z_{12}	z_{13}	Total
Ordinary Least Square	30	30	30	30	11	18	10	21	180
Independent Subsets	30	30	30	30	7	21	13	30	191
Simple Regression	21	30	30	19	2	22	1	3	128
Ridge Regression	30	30	30	30	28	26	28	30	232
Constraint Space Solution 1S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 3S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 6S	30	30	30	30	30	30	30	30	240

Count and Square Error (SE) measures for Scenario 6 were presented in Table 4.15 and 4.16. Constraint Space Solution (CSS) reported the highest Count number follow for Ridge Regression (RR) and Independent Subset (IS). Simple Regression (SR) although showed good numbers in Case A and B, did not detect the bias when two related basis elements were involved. Ordinary Least Square (OLS) had the lowest Count value due the highest variability in the POB's. The lowest SE was CSS at 1S and 3S. RR had lowest results than CSS at 6S and IS had better results than OLS and SR. SR showed the

highest SE value in Scenario 6C, increasing their overall mean. The method recommended when VIF > 10 with more than two related variable is CSS due reduce SE and show highest Count.

Table 4.15: Count measure Scenario 6 Summary

Methodology	A	B	C	Mean
	Count	Count	Count	
Ordinary Least Square	174	175	180	176
Independent Subsets	205	197	191	198
Simple Regression	210	219	128	186
Ridge Regression	228	227	232	229
Constraint Space Solution 1S	240	240	240	240
Constraint Space Solution 3S	240	240	240	240
Constraint Space Solution 6S	240	240	240	240

Table 4.16: Mean Square Error measure Scenario 6 Summary

Methodology	A	B	C	Mean
	MSE	MSE	MSE	
Ordinary Least Square	44.64	26.96	35.13	35.58
Independent Subsets	10.56	12.45	20.10	14.37
Simple Regression	8.08	7.07	35.99	17.05
Ridge Regression	5.31	5.19	4.37	4.96
Constraint Space Solution 1S	0.52	0.50	0.50	0.51
Constraint Space Solution 3S	2.72	2.80	2.75	2.75
Constraint Space Solution 6S	5.57	5.55	5.55	5.56

4.3 Summary Results Scenarios 2 to 5

Scenario 2, 3, 4 and 5 (Appendix C-F) showed similar results as Scenario 1 and 6. Ordinary Least Square (OLS) increased variability of POB's as scenarios raise VIF. Independent Subsets (IS) and Simple Regression (SR) controlled the variability but offset the related POB's. Ridge Regression (RR) reduced variability but in some cases was under the target. Constraint Space Solution (CSS) at 6S was better than RR in terms of Count measure and similar in terms of SE. CSS reduced the variability as constraint the space at 1S or 3S. The lowest SE was and highest Count measure was CSS at 1S and 3S.

4.4 Scenarios Comparison

This section compared scenarios using Square Errors (SE) and Count measure as decision variables. To compare between scenarios the Square Error and Count measure were standardized, using the number of POB per scenario. The methodologies were: Ordinary Least Square (OLS), Independent Subset (IS), Simple Regression (SR), Ridge Regression (RR) and Constraint Space Solution at 1S (CSS 1S), at 3S (CSS 3S) and 6S (CSS 6S). The comparison will determine which of these procedures were the most adequate to solve non-orthogonal problem in the Process Oriented Basis.

Figure 4.14 and Figure 4.15 present SE Boxplots and Count Boxplots per scenario. These Boxplots show individual results per scenarios. Table 4.17 and Table 4.18 tabulate the Mean Square Error (MSE) and Count Percentage (Count %) measure for all the scenarios. The following statements present the interpretation of the Boxplots and tabulate data. OLS showed high variability in SE and lower detection in Counts measures as VIF increased. This was expected since in presence of multicollinearity OLS estimates of POB's will have high variance and be distant from the true values. A certain amount of non-orthogonallity can be tolerated, in Scenario 1 and 4, OLS showed lower SE than SR and IS. IS presented their lowest SE result when VIF was moderate ($5 \leq VIF < 10$). In comparison with OLS and SR, IS had higher Count measure and lower SE values. This method regress a subset of independents POB reducing variability, but for the related basis elements the estimation was in some cases off target. SR increases

SE and Count measure as the scenarios increased in VIF. This method failed consistently estimating POB coefficients on target.

The best results were presented using RR or CSS. RR had the lowest SE, but not the highest Count values compared with CSS at 6S. SE results for RR were influenced due low variability in all POB's. The low Count can be explained by the bias allowed in RR to obtain a variance reduction in the estimates. CSS reduced the variability and estimated on target POB's given that this method uses a function to minimize the variance as explained in section 3.5. This procedure could become complex if there is not feasible solution or more than one feasible solution exists. The research Scenarios considered there was always an unique feasible solution. CSS showed the lowest SE and highest Count measure when the constraint space was below ± 3S. CSS within ± 6S presented similar results that RR.

Table 4.19 shows Count measure only for POB's related, this percentage permit identify how methods estimate only non-orthogonal basis. This measure provides similar conclusion than in Table 4.18, but magnify the difference between methods.

Although in Scenario 1 RR and CSS reduced the SE is not feasible to invest time and resources using these methods, since OLS could provide good estimates. For Scenario 2, OLS was not an alternative, again RR and CSS were the best alternative. Scenario 3 showed how a severe VIF affect the results for OLS, IS and SR. RR and CSS again present better SE and Count values. Scenario 4 showed better results for RR and CSS, but as in Scenario 1 OLS provided good results. For Scenario 5 CSS was the best

alternative. Scenario 6 was the worst case due VIF > 10 and there was more than one relation, in this case OLS showed higher SE results than IS and SR and CSS was the best alternative.

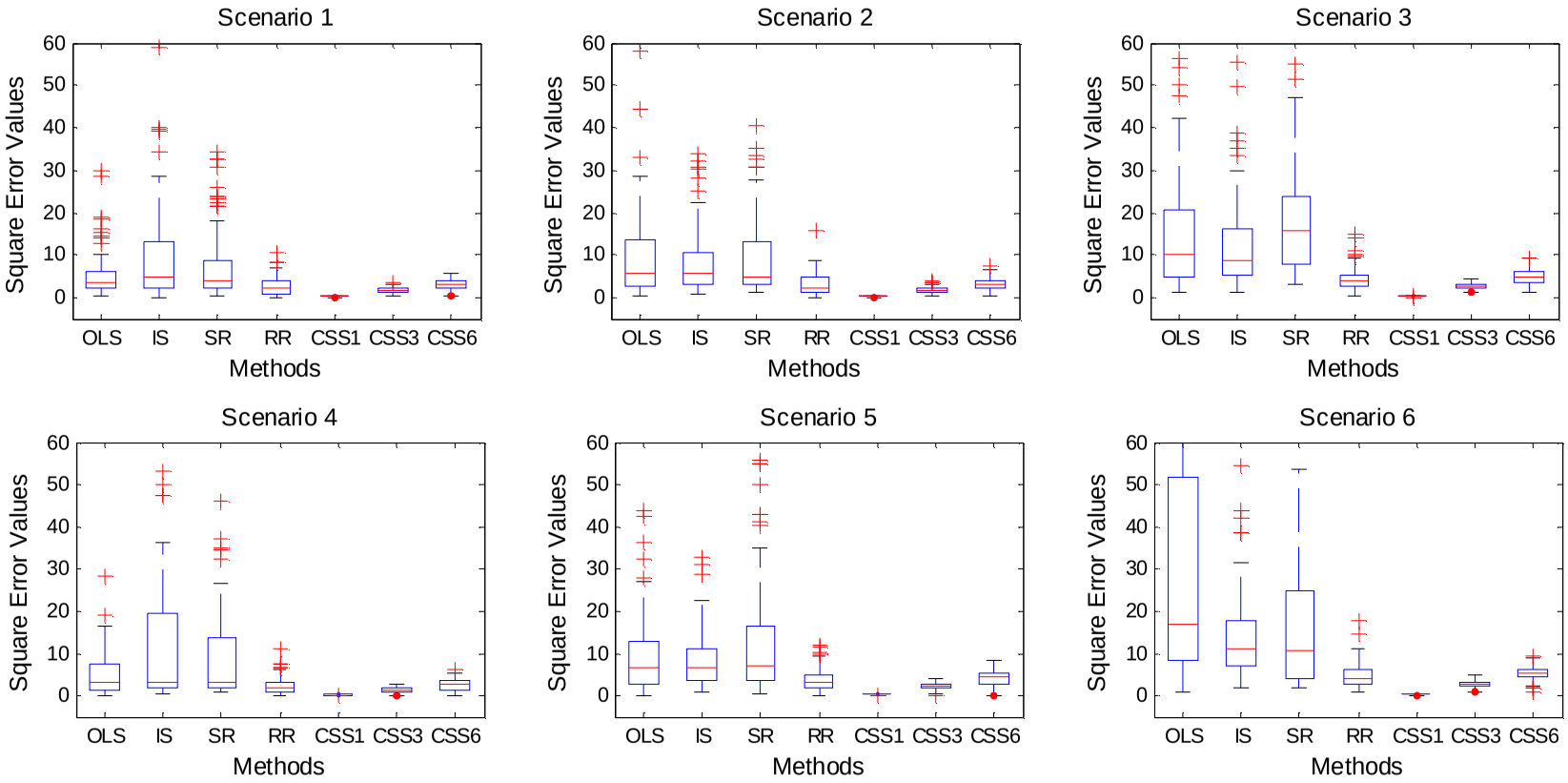


Figure 4.14: Square Error Methods Comparison for All Scenarios

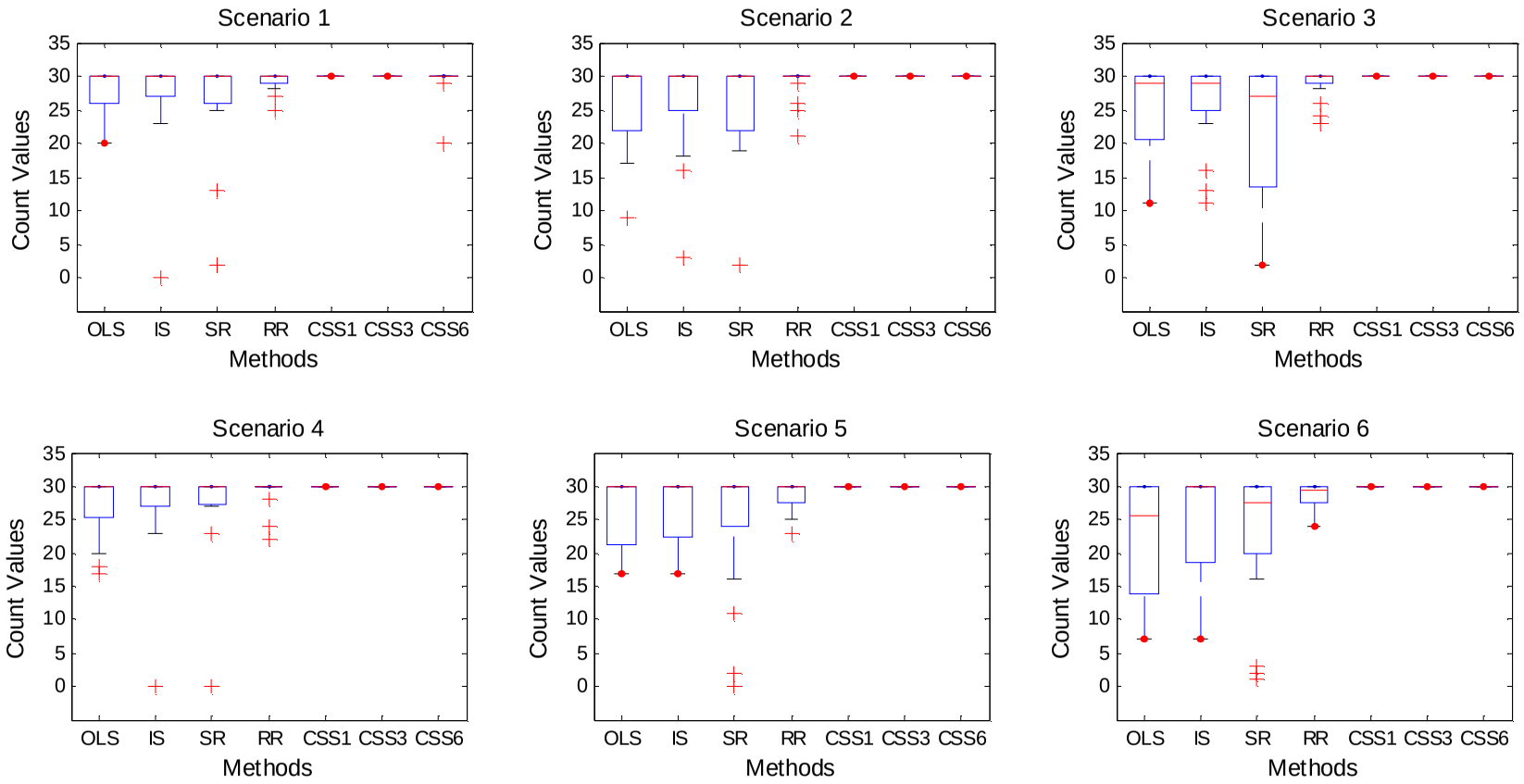


Figure 4.15: Count measure Methods Comparison for All Scenarios

Table 4.17: Standardize Mean Square Error measure Summary Scenarios

Methodology	Scenario					
	1	2	3	4	5	6
Ordinary Least Square	2.63	5.12	5.66	3.24	4.11	13.34
Independent Subsets	4.85	4.36	4.92	6.60	3.66	5.39
Simple Regression	3.74	4.78	6.81	5.23	5.59	6.39
Ridge Regression	1.34	1.58	1.67	1.49	1.63	1.86
Constraint Space Solution 1S	0.18	0.19	0.19	0.19	0.18	0.19
Constraint Space Solution 3S	0.89	0.90	1.00	0.86	0.96	1.03
Constraint Space Solution 6S	1.52	1.63	1.87	1.53	1.81	2.08

Table 4.18: Count % measure Summary Scenarios

Methodology	Scenario					
	1	2	3	4	5	6
Ordinary Least Square	92.8%	86.7%	84.6%	89.3%	87.1%	73.5%
Independent Subsets	90.0%	87.2%	86.7%	89.6%	88.4%	82.4%
Simple Regression	88.5%	86.1%	71.3%	89.6%	82.4%	77.4%
Ridge Regression	97.0%	96.5%	96.3%	96.4%	95.7%	95.4%
Constraint Space Solution 1S	100%	100%	100%	100%	100%	100%
Constraint Space Solution 3S	100%	100%	100%	100%	100%	100%
Constraint Space Solution 6S	100%	100%	100%	100%	100%	100%

Table 4.19: Count % measure Summary Scenarios for POB related

Methodology	Scenario					
	1	2	3	4	5	6
Ordinary Least Square	79%	60%	48%	74%	78%	47%
Independent Subsets	71%	62%	56%	74%	80%	65%
Simple Regression	75%	67%	46%	74%	69%	61%
Ridge Regression	91%	89%	88%	91%	93%	91%
Constraint Space Solution 1S	100%	100%	100%	100%	100%	100%
Constraint Space Solution 3S	100%	100%	100%	100%	100%	100%
Constraint Space Solution 6S	94%	100%	100%	100%	100%	100%

Figure 4.16 and 4.17 present Square Error (SE) and Count results per scenarios using each method. There were two comparable groups of results: (1) Constraint Space Solution (CSS) and Ridge Regression (RR) methods with lowest SE and highest Count measure and (2) OLS, IS and SR with higher SE and lower Count measure. CSS at 1S and 3S had the lowest SE and the highest Count measure. CSS at 6S showed similar results than RR for SE and better results for Count measure than RR. OLS had lower SE for Scenario 1 and 4 than SR or IS. OLS, IS and SR had similar SE results in Scenario 2. Scenario 3, 5 and 6 showed better SE results for IS than OLS and SR. OLS had higher Count measure for Scenario 1 than SR or IS and similar Count results in Scenario 2, 3, 4 and 5. In Scenario 6 IS showed higher Count results than OLS and SR. For lower VIF OLS was the best alternative, IS showed better results for moderate to severe VIF.

For all Scenarios there were not significant differences between RR and CSS. To decide which method use, evaluate if your process has physical or knows constraints to set up the CSS procedure. Although RR presented low variability for all the POB's, does not always provide a good identification of the offset POB in comparison with CSS. Remember the objective of POBREP is to detect POB offset to understand the process issues, if RR is not effective doing this, CSS will be the best alternative. POBREP is interested in the meaning of the basis elements coefficients for diagnosis and control. After comparing several procedures using SE and Count the most adequate procedure to estimate the Process Oriented Basis representation in presence of non-orthogonal basis elements was Constraint Space Solution at 1S.

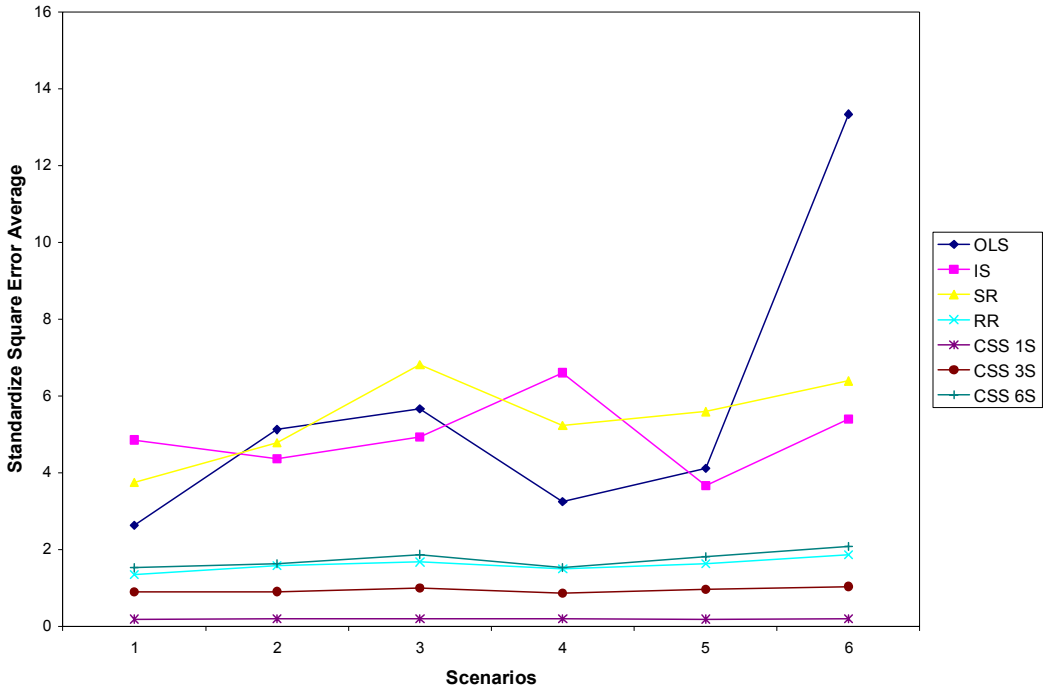


Figure 4.16: Square Error Methods Comparison for All Scenarios

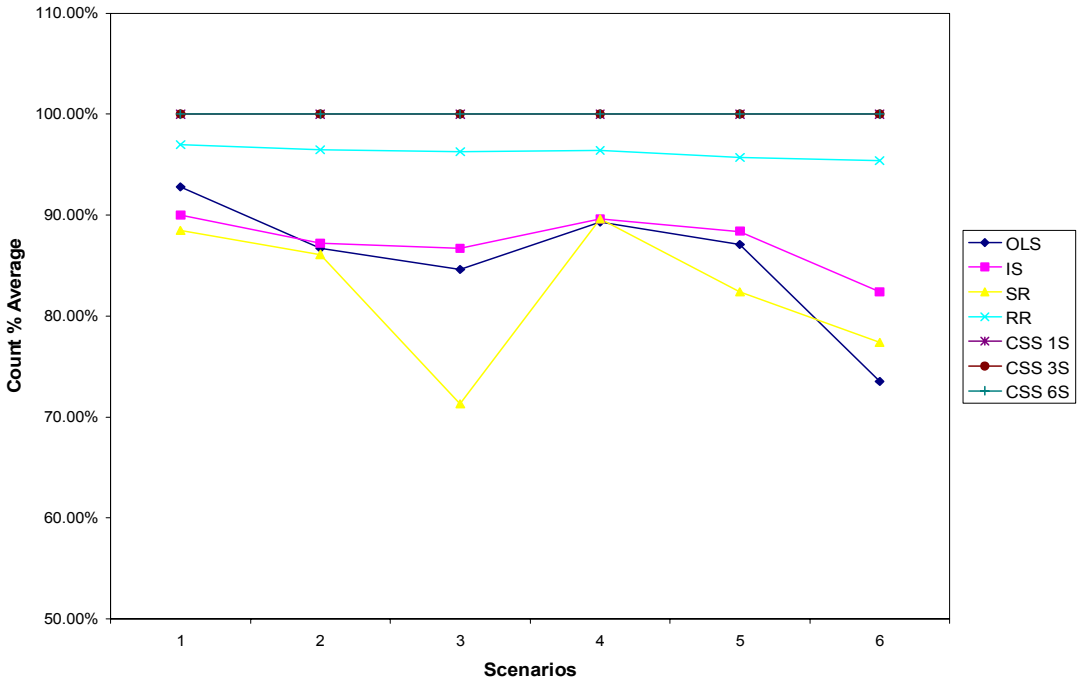


Figure 4.17: Count % measure Methods Comparison for All Scenarios

5 CONCLUSIONS AND FUTURE WORK

5.1 Conclusion

POBREP is a useful methodology to provide diagnosis in a multivariate process, but when the basis is not orthogonal may result in large variances and covariance for the least squares estimators of the regression coefficients. To assure POBREP capabilities remains, this research recommends method to deal with non-orthogonal POB basis estimating coefficients for process diagnostics and control.

In this work, the performances of several methods were evaluated for their capabilities to deal with non-orthogonal basis. The method to manage non-orthogonallity should reduce the variability and reach the theoretical POB coefficients, reducing the Square Error (SE) and maximizing Count measure. After evaluating six scenarios with different severity in terms of VIF and POB's related two comparable groups of results were detected: (1) Constraint Space Solution (CSS) and Ridge Regression (RR) methods with lowest SE and highest Count measure and (2) OLS, IS and SR with higher SE and lower Count measure. RR and CSS were the best methods to address the non-orthogonal problem. CSS exhibit significant improvements in the estimate of POBs when the constraint was below ± 3 sigma. In terms of the other methods, when there is not severe multicollinearity was better to use OLS than SR or IS. If the multicollinearity is moderate to severe IS was a better alternative than OLS or SR.

In conclusion the most adequate procedure to estimate the Process Oriented Basis representation in presence of non-orthogonal basis elements was Constraint Space Solution at 1S.

5.2 Future Research

This research only evaluates cases with bias in the POB's but with constant standard deviation. Evaluate how these methodologies react to changes in variance will be helpful to understand their power to estimate POB coefficients.

Constraint Space Solution (CSS) was the best method to estimate POBREP coefficients when the constraints were below ± 3 sigma. Other alternatives to challenge this methodology will be defining the constraints as a percentage of the target value or if to establish physical process constraints. For CSS there are three possibilities feasibility cases, evaluation of a case with more than one feasible solution could be interesting in order to identify the complexity of these methods.

Other interesting aspect will be to apply this methodology with other physical case that reveals a multicollinearity severity represented by VIF > 100. Although Ridge Regression and Constraint Space Solution will be providing good results, it will be interesting to evaluate Independent Subsets results, since VIF severity increase this method was providing better results.

REFERENCES

1. Belsey, D.A., Kuh, E., and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.

2. Birgoren, B. (1997). "Multivariate statistical process control for quality diagnostics and applications to process oriented basis representations", Ph.D. Thesis, Department of Industrial and Manufacturing Engineering, Pennsylvania State University, State College, Pennsylvania.

3. Espada-Colón, H., 1998, *Process Diagnostic in SMT Component Placement using POBREP*, Master Thesis, Department of Industrial Engineering, University of Puerto Rico Mayagüez Campus, May 1998.

4. Gunst, R.F. and Mason, R.L. (1977). "Biased estimation in regression: An evaluation using mean squared error," *J. Am. Statistics*, 72, 616-628.

5. Gunst, R.F. (1979). "Similarities among least squares, principal component, and latent root regression estimators," presented at the Washington D.C., Joint Statistical Meetings.

6. Gonzalez-Barreto, D.R. (1996). "Process Oriented Basis Representations for Multivariate Process Diagnostics and Control", Ph.D. Thesis, Department of Industrial and Manufacturing Engineering, Pennsylvania State University, State College, Pennsylvania.

7. Hawkins, D.M. (1973). "On the investigation of alternative regressions by principal components analysis," *Appl, Statistics,* 22, 275-286.

8. Hoerl, A.E., and Kennard, R.W. (1970a). "Ridge Regression: Biased estimation for nonorthogonal problems," *Technometrics,* 12, 55-67.

9. Hoerl, A.E., and Kennard, R.W. (1970b). "Ridge Regression: Applications to nonorthogonal problems," *Technometrics,* 12, 69-82

10. Hotelling, H., Multivariate Quality Control- Illustrated b the Air Testing of Sample Bombsights, in *Techniques of Statistical Analysis*, edited by C. Eisenhart, M. W. Hastay, and W. A. Wallis, McGraw-Hill, New York; 1947, pp. 111-184.

11. Lowry, C.A. and Montgomery, D.C. (1995). "A Review of Multivariate Control Charts". *IIE Transactions*, 27, pp.800-810.

12. Marquardt, D.W. (1970). "Generalized Inverses, Ridge Regression, biased linear estimation, and nonlinear estimation," *Technometrics*, 12, 591-612.

13. Mason, R.L., Gunst, R.F. and Webster, J.T. (1975). "Regression analysis and problems of multicollinearity," *Common. Statistics*, 4(3), 277-292.

14. Montgomery, D.C. and Peck. E. A. (1992). *Introduction to Linear Regression Analysis,* John Wiley and Sons, New York, NY.

15. Snee, R.D. (1973). "Some aspects of nonorthogonal data analysis, Part 1. Developing predicting equations," *Journal of Quality Technology,* 5, 67-79.

16. Walpole, R.E., and Myers, R.H. (1989). *Probability and Statistics for Engineers and Scientists.* New York: MacMillan.

17. Webster, J.T., Gunst, R.F., and Mason, R.L. (1974). "Latent Root regression analysis," *Technometrics*, 16, 513-522.

18. Webster, J.T., Gunst, R.F., and Mason, R.L. (1976). "A Comparison of Least Squares and Latent Root Regression Estimators," *Technometrics*, 18, 75-83.

APPENDIX A MATLAB® SCRIPTS

APPENDIX A1 VECTOR GENERATOR

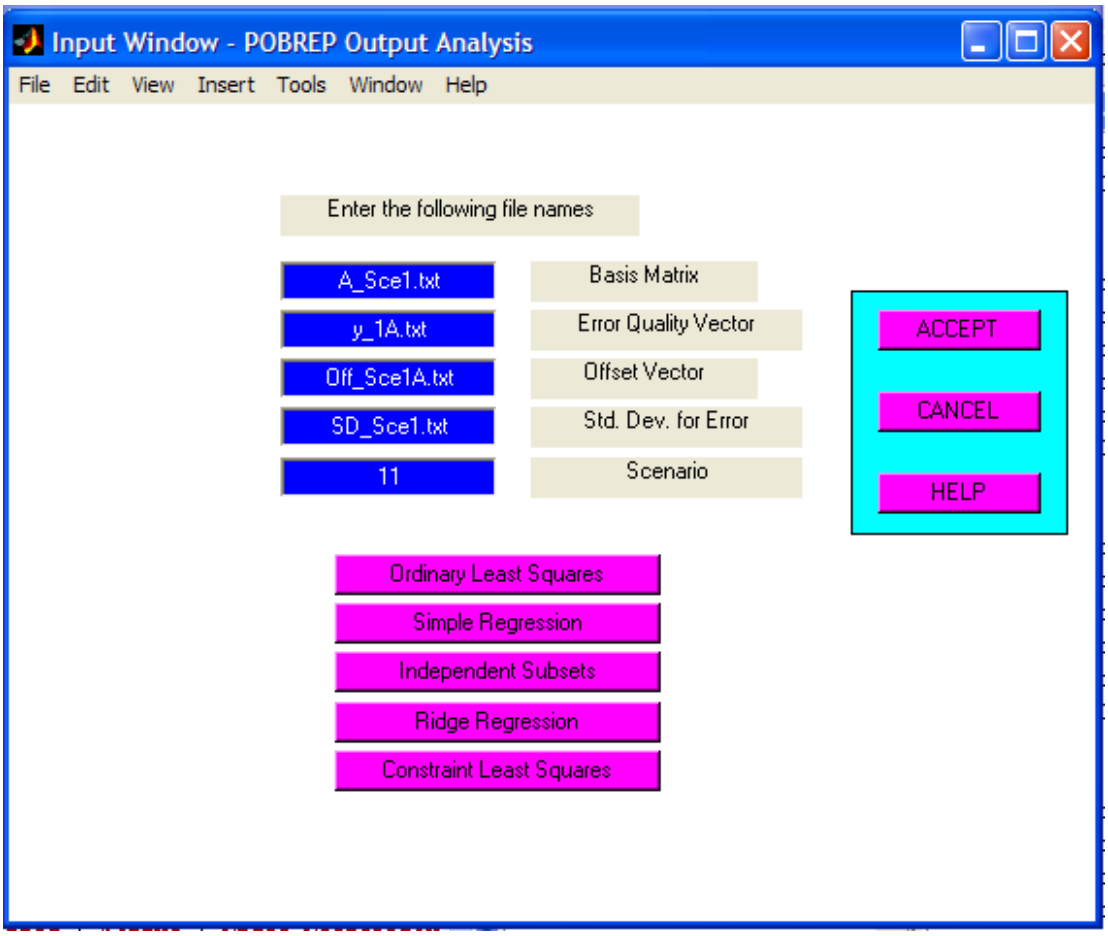
```
%Script Name :  genepr.m
%************************************************
% Script to generate a multivariate quality vector
% to be analyzed through the use of POBREP
%************************************************

clc
load(filen1)
load(filen2)
load(filen3)

[amat] = eval(strrep(filen1,'.txt',''));
[offset] = eval(strrep(filen2,'.txt',''));
[stddev] = eval(strrep(filen3,'.txt',''));

%
[nr,nbase] = size(amat);
vlngth = length(amat);
%Generate Multivariate Quality Vector
%
rand('normal')
for k=1:nobs
      for l = 1:nbase
      ranno(l) = rand(1,1);
      end
            for m = 1:vlngth
            sumbas = 0.0;
                  for n = 1:nbase
                  rancom = (ranno(n)*stddev(n)+offset(n))*amat(m,n);
                  sumbas = sumbas + rancom;
                  end
             yvector(m,k) = sumbas + rand(1,1) * sigma;
             end
end
```

APPENDIX A2 MATLAB WINDOW



APPENDIX A3 ORDINARY LEAST SQUARE

```
%Script Name :  pobrep.m
%************************************************
load(filen1)
load(filen2)
load(filen3)
[amat] = eval(strrep(filen1,'.txt',''));
[offset] = eval(strrep(filen3,'.txt',''));
[yvector] = eval(strrep(filen2,'.txt',''));
[nr,nbase] = size(amat)
[nry,nbasey]=size(yvector)
Off = repmat(offset,nbasey,1)

        f2 = figure
        for i =1:nbasey
                [b] = amat\yvector(:,i);
                zcoef2(:,i) = b;
                plot(zcoef2')
                xlabel('Run Number')
                ylabel('Z Values')
                legend('z1','z2','z3')
                title('Plot of All Generated Z Representations')
 end
 p_but = uicontrol(gcf,'Style','Push','position', [.89 .94 .10 .05],....
                'units','normalized','background','m',....
                'String','CLOSE','foreground','b',....
                'Callback','close(f2)');

  Error=sum((Off'-zcoef2).^2)
```

APPENDIX A4 INDEPENDENT SUBSETS

```
%Script Name :  Independent subsets
%***********************************************

load(filen1)
load(filen2)
load(filen3)
[amat] = eval(strrep(filen1,'.txt',''));
[offset] = eval(strrep(filen3,'.txt',''));
[yvector] = eval(strrep(filen2,'.txt',''));
[nr,nbase] = size(amat);
[nry,nbasey]=size(yvector);
vlngth = length(amat);
Off = repmat(offset,nbasey,1)

X=(1:nbasey)
z1=[];
z2=[];
z3=[];
z4=[];
z5=[];
z6=[];
z7=[];
z8=[];
zt=[];

zac=[ ];
bac=[ ];
zacc=[ ];
bacc=[ ];
sub=0;
corrida=0;
entrar=0;
de=0;
in=0;
damat=[ ];
dz=[ ];
iamat=[ ];
iz=[ ];
depend=0

[V,D] = eig(corrcoef(amat));
VIF=diag(inv(corrcoef(amat)));
[u,s,v] = svd(amat);
sin=diag(s);


%Calcular condition index and variance
```

```
for i=1:nbase
   miu(i)=(max(sin))/sin(i);
end

for i=1:nbase
  for j=1:nbase
    pi(i,j)=(v(j,i)^2/sin(i)^2)/(VIF(j,:));
    if and( pi(i,j)>.1, miu(:,i)>5)
       depend=1;
   end
  end
end


for i=1:nbase
       de=0;
       in=0;
       damat=[ ];
       dz=[ ];
       iamat=[ ];
       iz=[ ];
      for j=1:nbase
         if and( pi(i,j)>.1, miu(:,i)>5)
           de=de+1;
           damat(:,de)=amat(:,j);
           dz(:,de)=j;
           [dnr,dnbase] = size(damat);
           entrar=1;
         else
           in=in+1;
           iamat(:,in)=amat(:,j);
           iz(:,in)=j;
         end
      end
%Si no hay variables dependientes
      if de==0
        dnbase=1;
        if and(corrida==0,depend==0)
            entrar=1;
        end
        corrida=corrida+1
      end

      %Crear matrix entrando dependientes una por una

   if entrar==1
      for m=1:dnbase
         subamat=[ ];
         sub=sub+1;
         if de<nbase
            subamat=[iamat];
                  z=[iz];
```

```
end
if de>0
   addcol=damat(:,m);
   addj=dz(:,m);
   subamat=[subamat addcol];
   z=[z, addj]
end
[snr,snbase] = size(subamat);
eval(['zs',int2str(sub),' = z']);
zac=[zac,z];
eval(['subamats',int2str(sub),' = subamat']);
for num=1:30
    y=[yvector(:,num)];
   b = subamat\y;
   eval(['zcoef',int2str(sub),' = b']);
   bac=[bac,b];
end
        for j=1:snbase
            if z(:,j)==1
               z1=[z1
               bac(j,:)];
               [zr,znbase] = size(z1);
               if zr > 1
                  zm1=mean(z1);
               else
                  zm1=z1;
               end
               zt(1,:)=zm1;
            end
            if z(:,j)==2
               z2=[z2
               bac(j,:)];
               [zr,znbase] = size(z2);
               if zr > 1
                  zm2=mean(z2);
               else
                  zm2=z2;
               end
               zt(2,:)=zm2;
            end
            if z(:,j)==3
               z3=[z3
               bac(j,:)];
               [zr,znbase] = size(z3);
               if zr > 1
                  zm3=mean(z3);
               else
                  zm3=z3;
               end
               zt(3,:)=zm3;
            end
            if z(:,j)==4
```

```
    z4=[z4
    bac(j,:)];
    [zr,znbase] = size(z4);
    if zr > 1
       zm4=mean(z4) ;
    else
       zm4=z4;
    end
    zt(4,:)=zm4;
end
if z(:,j)==5
    z5=[z5
    bac(j,:)];
    [zr,znbase] = size(z5);
    if zr > 1
       zm5=mean(z5);
    else
       zm5=z5;
    end
    zt(5,:)=zm5;
end
if z(:,j)==6
    z6=[z6
    bac(j,:)];
    [zr,znbase] = size(z6);
     if zr > 1
       zm6=mean(z6);
    else
       zm6=z6;
    end
    zt(6,:)=zm6;
end
if z(:,j)==7
    z7=[z7
    bac(j,:)];
    [zr,znbase] = size(z7);
    if zr > 1
       zm7=mean(z7);
    else
       zm7=z7;
    end
    zt(7,:)=zm7;
end
if z(:,j)==8
    z8=[z8
    bac(j,:)];
    [zr,znbase] = size(z8);
    if zr > 1
       zm8=mean(z8);
    else
       zm8=z8;
    end
```

```
                    zt(8,:)=zm8;
                  end
                end
        subamat=[ ];
        z=[ ];
        bac=[];
      end
   end
end

f1=figure
title('Plot of Generated Z Representations')

for i=1:nbase
   subplot(3,3,i)
   [nrr,nbasee] = size(eval(['z',int2str(i)]))
   if nrr==1
     plot(zt(i,:))
                    xlabel('Run Number')
                    axis([0 30 -2 6])
                    ylabel('Zs Values')
      title('Independent Subsets ')
  else
     errorbar(X,zt(i,:),(zt(i,:)-min(eval(['z',int2str(i)]))),(max(eval(['z',int2str(i)]))-zt(i,:)))
                    xlabel('Run Number')
                    axis([0 30 -2 6])
                    ylabel('Zs Values')
title('Independent Subsets ')
  end
          title(['Z ',num2str(i),' Values'])
   p_but = uicontrol(gcf,'Style','Push','position', [.89 .94 .10 .05],....
                    'units','normalized','background','m',....
                    'String','CLOSE','foreground','b',....
                    'Callback','close(f1)');
end
Error=sum((Off'-zt).^2)
f2=figure
plot(zt')
xlabel('Run Number')
ylabel('Z Values')
legend('z1','z2','z3')
title('Plot of All Generated Z Representations')
p_but = uicontrol(gcf,'Style','Push','position', [.89 .94 .10 .05],....
                    'units','normalized','background','m',....
                    'String','CLOSE','foreground','b',....
                    'Callback','close(f2)');
```

APPENDIX A5 SIMPLE REGRESSION

```
Script Name :  Simple_Regression.m
%************************************************
% Script to determine Pobrep coefficient using
% Simple Regression to each Basis Element
%************************************************

load(filen1)
load(filen2)
load(filen3)
[amat] = eval(strrep(filen1,'.txt',''));
[offset] = eval(strrep(filen3,'.txt',''));
[yvector] = eval(strrep(filen2,'.txt',''));
[nr,nbase] = size(amat)
[nry,nbasey]=size(yvector)
Off = repmat(offset,nbasey,1)
vlngth = length(amat)

%Generate Regression and Graphics

po=1;
f2=figure
for j=1:nbase

  for i=1:nbasey

        [b] = amat(:,j)\yvector(:,i);
        zcoef2(po,i) = b;
        plot(zcoef2')
        legend('z1','z2','z3')
        xlabel('Run Number')
        ylabel('Z Values')
        title('Plot of All Generated Z Representations with Simple Regression')
  end
  po=po+1;
end

p_but = uicontrol(gcf,'Style','Push','position', [.89 .94 .10 .05],....
                  'units','normalized','background','m',....
                  'String','CLOSE','foreground','b',....
                  'Callback','close(f2)');

Error=sum((Off'-zcoef2).^2)
```

APPENDIX A6 RIDGE REGRESSION

Script Name : RR.m

```
%*********************************************
%* Script to calculate ridge coefficients and *
%* produce ridge trace for each coefficient.  *
%* User must specify X matrix,  a vector of   *
%* deltas for the biased estimates and the    *
%* vector of responses y.                     *
%*********************************************
%
load(filen1)
load(filen2)
load(filen3)
[amat] = eval(strrep(filen1,'.txt',''));
[offsett] = eval(strrep(filen3,'.txt',''));
[ridvec0] = eval(strrep(filen2,'.txt',''));
load offset.txt
[nr,nbase] = size(amat);
[nry,nbasey]=size(ridvec0)
Off = repmat(offsett,nbasey,1)
rbeta=[];

[offset]=offset;
x = amat;
xtx = x'*x;
d = length(xtx);
inc = offset;
k = length(offset);

f4=figure
for i= 1:k
offs = inc(i)*eye(d);
xtxd = xtx + offs;
y = ridvec0(:,1);
xty = x'*y;
xtxdi = inv(xtxd);
beta = xtxdi*xty;
rbeta = [rbeta beta];
orbeta = inv(xtx)*xty;
end
plot(inc,rbeta')
xlabel('delta')
ylabel ('Ridge Coefficient')
      legend('z1','z2','z3','z4','z7','z8','z12','z13')
title (['Ridge Trace Example 1'])
     p_but = uicontrol(gcf,'Style','Push','position', [.89 .94 .10 .05],....
                       'units','normalized','background','m',....
                       'String','CLOSE','foreground','b',....
                       'Callback','close(f4)');
```

```
[row,ylength] = size(ridvec0);
kr = input('Enter the number of k: ');
f3=figure
for i =1:ylength
        offsetf = kr*eye(d);
y = ridvec0(:,i);
beta1 = inv(xtx + offsetf)*x'*y;
zcoef(:,i) = beta1;
plot(zcoef')
legend('z1','z2','z3','z4','z7','z8','z12','z13')
        xlabel('Run Number')
        ylabel('Z Values')
        title('Plot of All Generated Z Representations')
end

 p_but = uicontrol(gcf,'Style','Push','position', [.89 .94 .10 .05],....
                    'units','normalized','background','m',....
                    'String','CLOSE','foreground','b',....
                    'Callback','close(f3)');
Error=sum((Off'-zcoef).^2)
```

APPENDIX A7 CONSTRAINT SPACE SOLUTION

```
%Script Name :  cls.m
%************************************************
load(filen1)
load(filen2)
load(filen3)
load(filen4)

[amat] = eval(strrep(filen1,'.txt',''));
[offset] = eval(strrep(filen3,'.txt',''));
[yvector] = eval(strrep(filen2,'.txt',''));
[stddev] = eval(strrep(filen4,'.txt',''));
 [nr,nbase] = size(amat)
[nry,nbasey]=size(yvector)
[nr,nbase] = size(amat);
Off = repmat(offset,nbasey,1)

por=(1:1:8)
[pnr,pnbase] = size(por);
noffset=offset+.1
f2 = figure
for d=1:pnbase
   subplot(3,3,d)
   zcoef1=[];
for i =1:nbasey
      LB=offset-por(:,d)*stddev';
      UB=offset+por(:,d)*stddev';
       [b] = lsqlin(amat,yvector(:,i),[],[],[],[],LB,UB);
       zcoef1(:,i) = b;
       plot(zcoef1')
       axis([0 30 -2 6])
       title([int2str(por(:,d)),' Sigma'])
       xlabel('Run Number')
        ylabel('Zs Values')
end
 if d==1
    zcoef2=zcoef1;
    Error2=sum((Off'-zcoef2).^2)
    elseif d==3
       zcoef3=zcoef1;
       Error3=sum((Off'-zcoef3).^2)
    elseif d==6
       zcoef4=zcoef1;
       Error4=sum((Off'-zcoef4).^2)
    f4=figure
          plot(zcoef1')
           axis([0 30 -2 6])
title('Plot of All Generated Z Representations with CLS, 6S')
                    xlabel('Run Number')
                    ylabel('Zs Values')
```

```
    end
 end
end
end
legend('z1','z2','z3')
  p_but = uicontrol(gcf,'Style','Push','position', [.89 .94 .10 .05],....
                        'units','normalized','background','m',....
                        'String','CLOSE','foreground','b',....
                        'Callback','close(f2)');
end
Error=sum((Off'-zcoef1).^2)
```

APPENDIX B DIAGNOSIS MULTICOLLINEARITY FOR SCENARIOS 2 - 6

Table B. 1: Variance Inflation Factor for the stencil printing process Scenarios

Relation between Basis		Variance Inflation Factor		
		VIF ≤ 5	**5 < VIF ≤ 10**	**VIF > 10**
Two Basis Elements		***Scenario 1***	***Scenario 2***	***Scenario 3***
	z_1	3	2	4.2
	z_2	1	1	1.6
	z_3	1	1	2.2
	z_4	2	1	2
	z_5			5.2
	z_6			**15.5**
	z_7	**5**	**10**	**11.9**
	z_8		**8.5**	
	z_9			5.6
	z_{12}	**4**		
	z_{13}			
	z_{14}			
More than two Basis Elements		***Scenario 4***	***Scenario 5***	***Scenario 6***
	z_1	1	3	3
	z_2	1	1	1
	z_3	2	1	1
	z_4		2	2
	z_5			
	z_6			
	z_7	**5**	**5**	**20**
	z_8			**17**
	z_9			
	z_{12}			**16**
	z_{13}		**10**	**14**
	z_{14}	**4**	**7**	

Table B. 2: Singular Value Analysis for Scenario 2

Condition Index	Variance Proportion					
	A_1	A_2	A_3	A_4	A_7	A_8
1.0000	0.0305	0	0	0.0010	0.0008	.0014
1.1047	0	0	0.1000	0	0	.0000
1.4491	0.0098	0	0	0.0867	0.0009	.0016
1.5623	0	0.2000	0	0	0	.0000
2.6436	0.0403	0	0	0.0863	0.0117	.0289
10.3669	0.0028	0	0	0.0038	**0.5733**	**.3603**

Table B. 3: Singular Value Analysis for Scenario 3

Condition Index	Variance Proportion							
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_9
1.0000	0.0055	0.0002	0.0023	0.0013	0.0003	0.0003	0.0007	0.0031
1.2163	0.0036	0.0001	0.0334	0	0	0.0001	0.0002	0.0005
1.4272	0.009	0.0006	0.0025	0.0267	0.0047	0.0001	0.0003	0.0006
1.7014	0.0003	0.098	0.0021	0.004	0.0004	0	0.0001	0.0019
2.4847	0.0143	0.0222	0.007	0.0669	0	0.0009	0.0034	0.0172
4.0508	0	0.025	0.0145	0.0708	0.0721	0.0059	0.0067	0.052
5.2998	0.0119	0.012	0.0077	0.0302	0.1419	0.0139	0.0445	0.0302
12.1380	0.003	0.0009	0.0007	0.0001	0.0113	**0.3916**	**0.2752**	0.0028

Table B. 4: Singular Value Analysis for Scenario 4

Condition Index	Variance Proportion				
	A_2	A_3	A_4	A_7	A_{14}
1	0	0.1	0	0	0
1.2272	0	0	0.0112	0.0077	0.0224
1.4142	0.2	0	0	0	0
1.4142	0	0	0.08	0	0.01
6.5187	0	0	**0.1088**	**0.6323**	**0.2176**

Table B. 5: Singular Value Analysis for Scenario 5

Condition Index	Variance Proportion						
	A_1	A_2	A_3	A_4	A_7	A_{12}	A_{13}
1.0000	0.0020	0	0	0.0076	0.0016	0.0019	0.0042
1.1137	0.0277	0	0	0.0015	0.0004	0.0003	0.0007
1.1356	0	0.0000	0.1000	0	0	0	0
1.6059	0	0.2000	0	0	0	0	0
2.1454	0.0002	0	0	0.1379	0.0012	0.0022	0.0074
5.9954	0.0368	0	0	0.0008	**0.3287**	0.0247	**0.1122**
8.6097	0.0332	0	0	0.0022	**0.3081**	**0.3309**	**0.1135**

APPENDIX C RESULTS SCENARIO 2

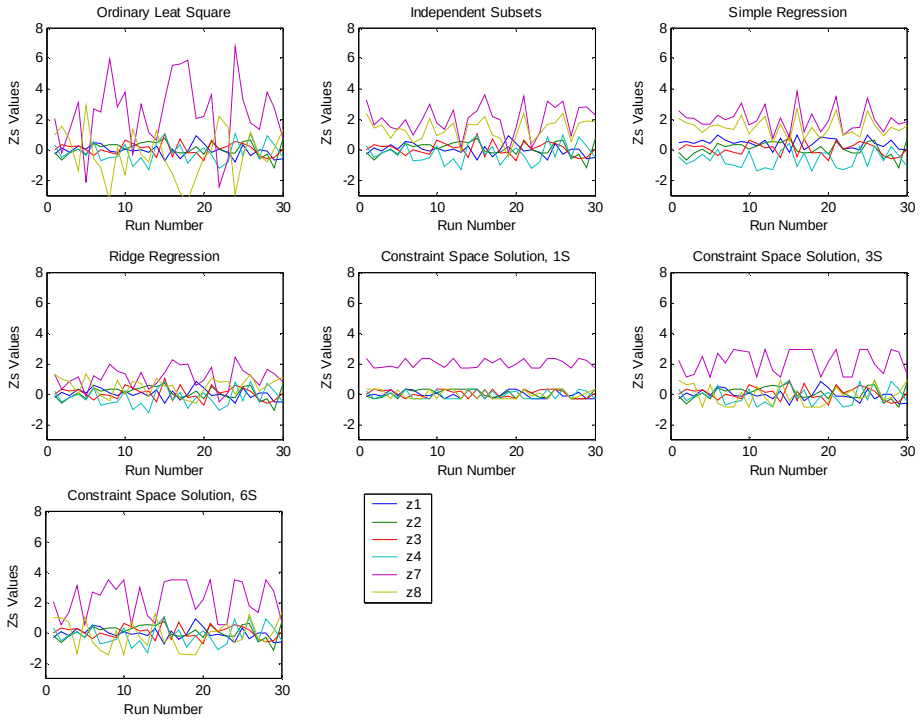


Figure C. 1: Methodology for Scenario 2A

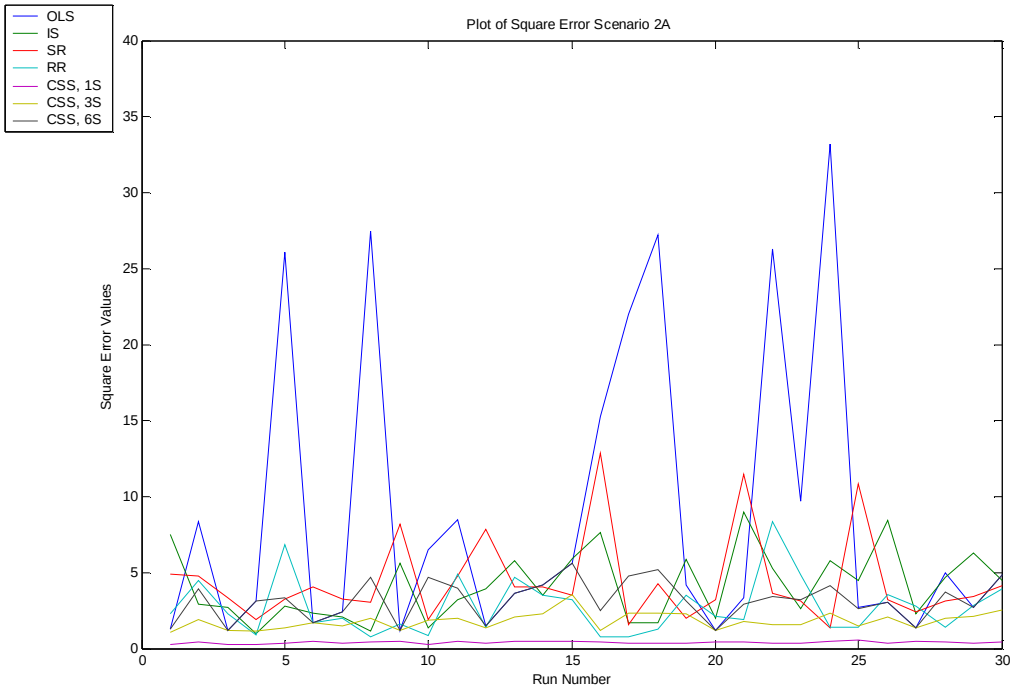


Figure C. 2: Square Error Methods Comparison for Scenario 2A

Table C. 1: Square Error measure Scenario 2A

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	8.80	4.18	9.86
Independent Subsets	4.11	3.68	2.27
Simple Regression	4.42	3.46	2.90
Ridge Regression	2.73	2.17	1.85
Constraint Space Solution 1S	0.38	0.36	0.07
Constraint Space Solution 3S	1.78	1.79	0.55
Constraint Space Solution 6S	3.15	3.17	1.30

Table C. 2: Count measure Scenario 2A

Methodology	z_1	z_2	z_3	z_4	z_7	z_8	Total
Ordinary Least Square	30	30	30	30	18	22	160
Independent Subsets	30	30	30	30	27	18	165
Simple Regression	30	30	30	30	29	22	171
Ridge Regression	30	30	30	30	25	30	175
Constraint Space Solution 1S	30	30	30	30	30	30	180
Constraint Space Solution 3S	30	30	30	30	30	30	180
Constraint Space Solution 6S	30	30	30	30	30	30	180

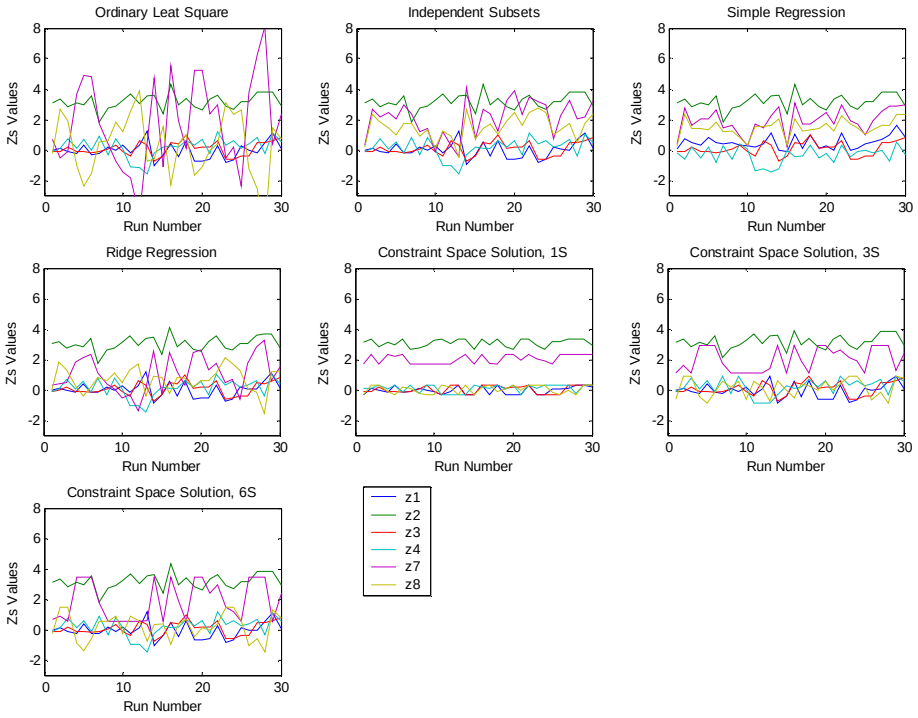


Figure C. 3: Methodology for Scenario 2B

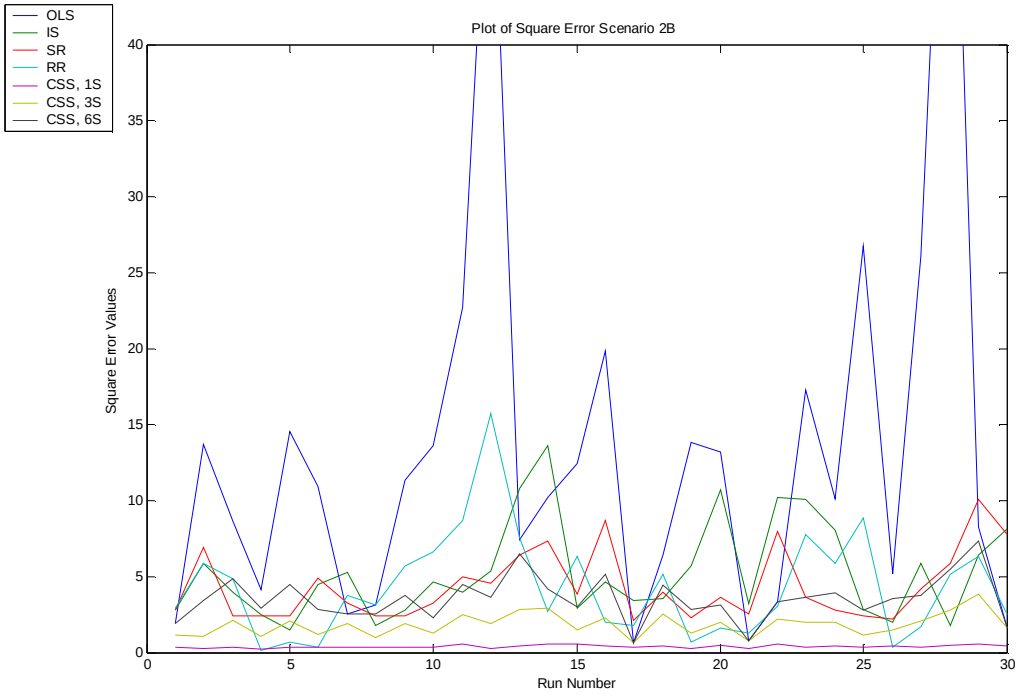


Figure C. 4: Square Error Methods Comparison for Scenario 2B

Table C. 3: Square Error measure Scenario 2B

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	13.81	10.53	14.97
Independent Subsets	5.27	4.52	3.17
Simple Regression	4.33	3.59	2.26
Ridge Regression	4.28	3.42	3.40
Constraint Space Solution 1S	0.36	0.34	0.09
Constraint Space Solution 3S	1.81	1.91	0.74
Constraint Space Solution 6S	3.50	3.45	1.47

Table C. 4: Count measure Scenario 2B

Methodology	z_1	z_2	z_3	z_4	z_7	z_8	Total
Ordinary Least Square	30	30	30	30	9	17	146
Independent Subsets	30	30	30	30	25	16	161
Simple Regression	30	30	30	30	28	21	169
Ridge Regression	30	30	30	30	21	26	167
Constraint Space Solution 1S	30	30	30	30	30	30	180
Constraint Space Solution 3S	30	30	30	30	30	30	180
Constraint Space Solution 6S	30	30	30	30	30	30	180

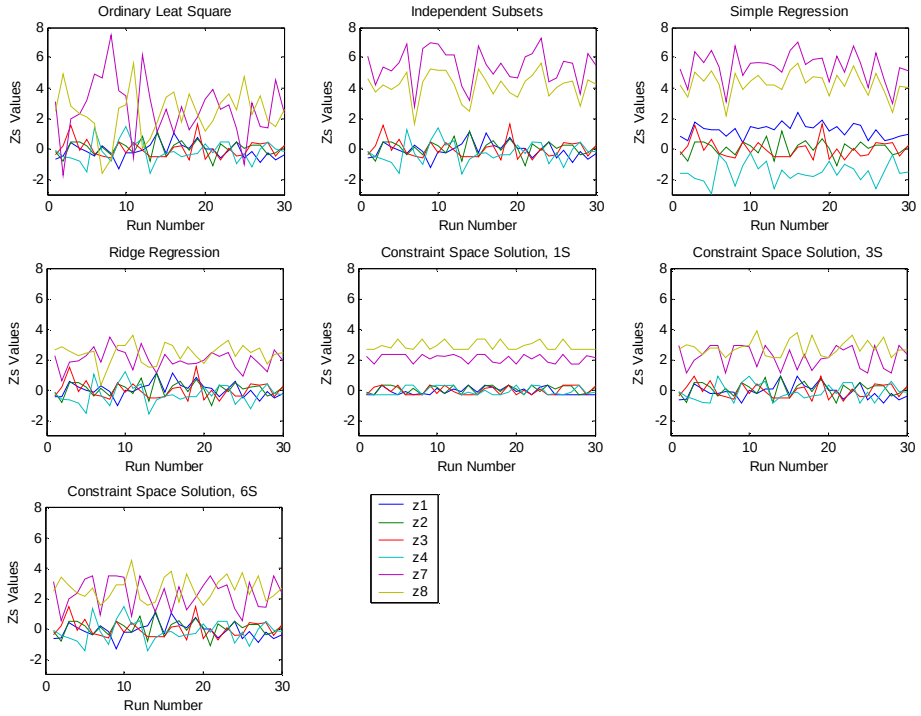


Figure C. 5: Methodology for Scenario 2C

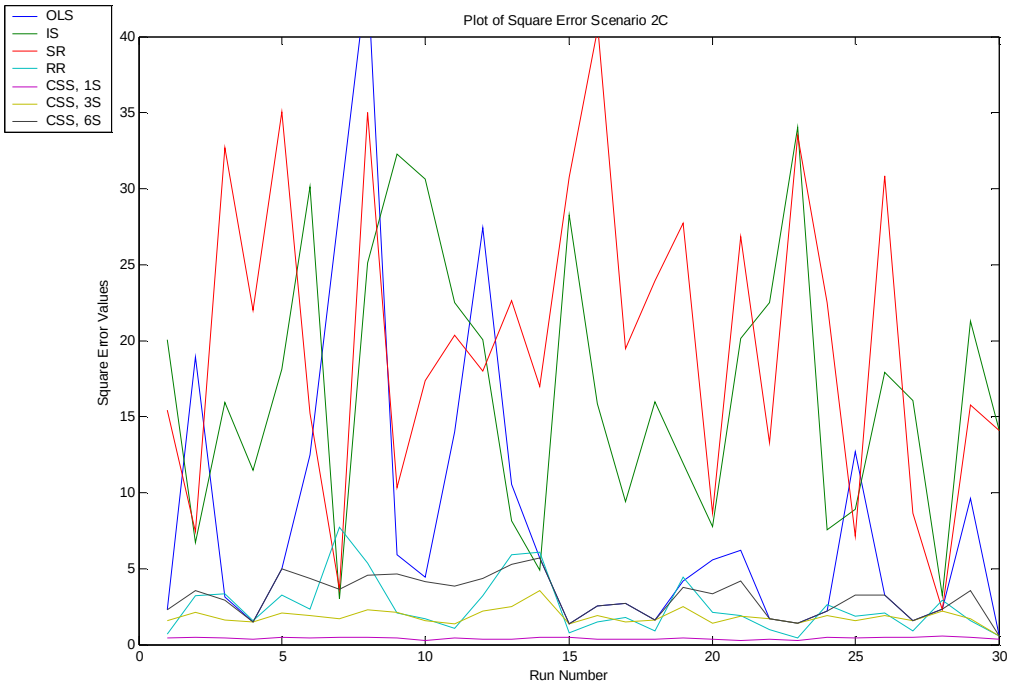


Figure C. 6: Square Error Methods Comparison for Scenario 2C

Table C. 5: Square Error measure Scenario 2C

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	8.11	4.29	10.01
Independent Subsets	16.77	16.02	8.84
Simple Regression	19.91	18.70	10.18
Ridge Regression	2.47	1.94	1.81
Constraint Space Solution 1S	0.39	0.39	0.07
Constraint Space Solution 3S	1.80	1.71	0.52
Constraint Space Solution 6S	3.14	3.28	1.32

Table C. 6: Count measure Scenario 2C

Methodology	z_1	z_2	z_3	z_4	z_7	z_8	Total
Ordinary Least Square	30	30	30	30	20	22	162
Independent Subsets	30	30	30	29	3	23	145
Simple Regression	25	30	30	19	2	19	125
Ridge Regression	30	30	30	30	30	29	179
Constraint Space Solution 1S	30	30	30	30	30	30	180
Constraint Space Solution 3S	30	30	30	30	30	30	180
Constraint Space Solution 6S	30	30	30	30	30	30	180

Table C. 7: Count measure Scenario 2 Summary

Methodology	A	B	C	Mean
	Count	Count	Count	
Ordinary Least Square	160	146	162	156
Independent Subsets	165	161	145	157
Simple Regression	171	169	125	155
Ridge Regression	175	167	179	174
Constraint Space Solution 1S	180	180	180	180
Constraint Space Solution 3S	180	180	180	180
Constraint Space Solution 6S	180	180	180	180

Table C. 8: Square Error measure Scenario 2 Summary

Methodology	A	B	C	Mean
	MSE	MSE	MSE	
Ordinary Least Square	8.80	13.81	8.11	10.24
Independent Subsets	4.11	5.27	16.77	8.71
Simple Regression	4.42	4.33	19.91	9.56
Ridge Regression	2.73	4.28	2.47	3.16
Constraint Space Solution 1S	0.38	0.36	0.39	0.37
Constraint Space Solution 3S	1.78	1.81	1.80	1.80
Constraint Space Solution 6S	3.15	3.50	3.14	3.27

APPENDIX D RESULTS SCENARIO 3

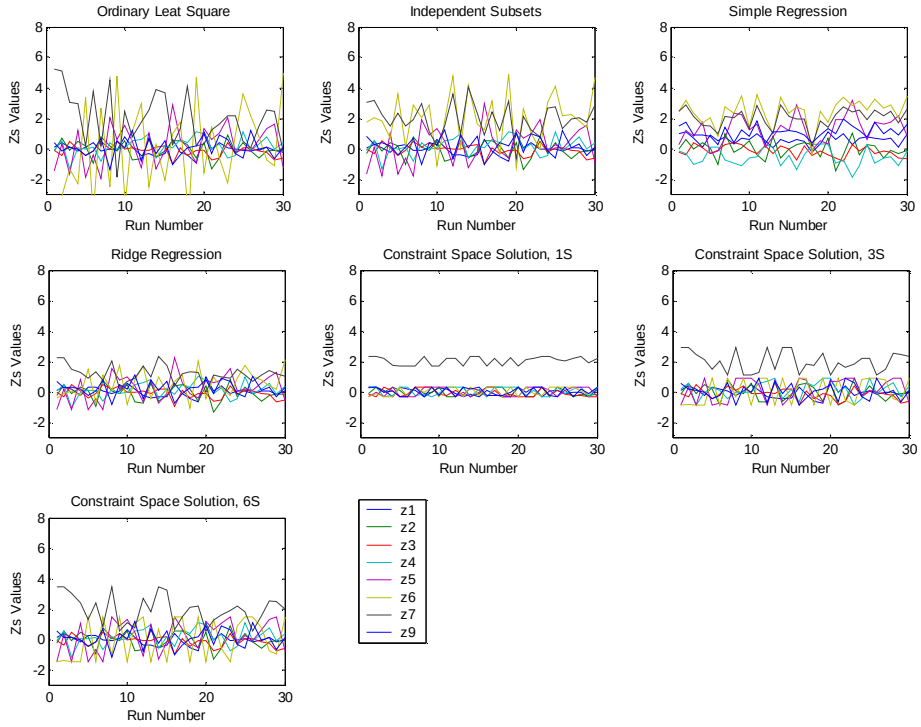


Figure D. 1: Methodology for Scenario 3A

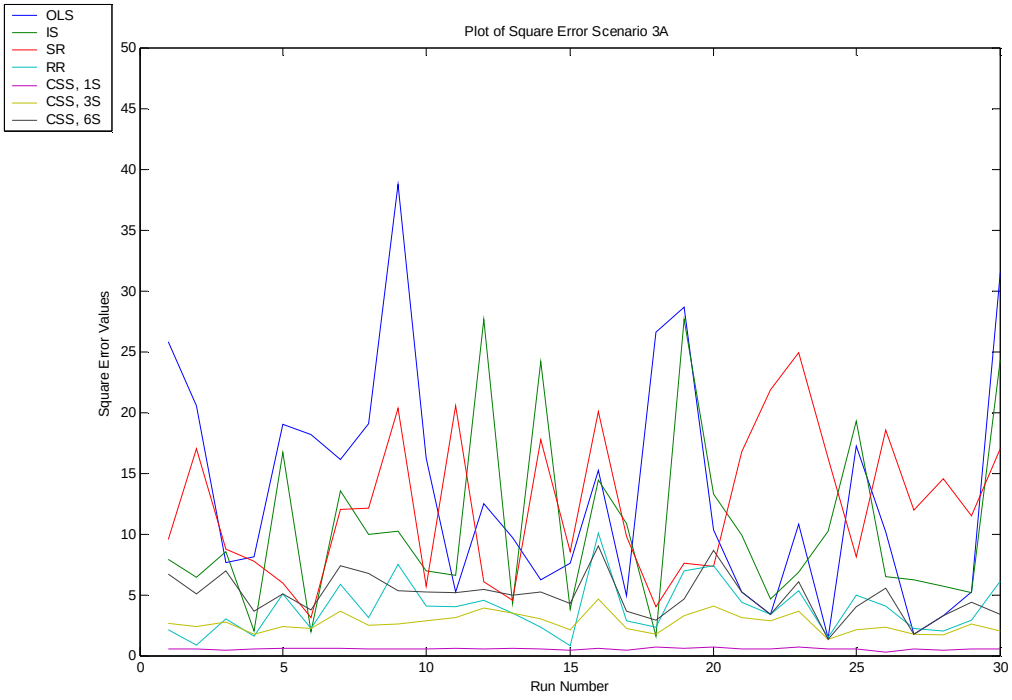


Figure D. 2: Square Error Methods Comparison for Scenario 3A

Table D. 1: Square Error measure Scenario 3A

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	13.52	10.52	9.54
Independent Subsets	10.55	8.24	7.46
Simple Regression	12.30	11.69	6.05
Ridge Regression	3.87	3.39	2.18
Constraint Space Solution 1S	0.50	0.50	0.09
Constraint Space Solution 3S	2.66	2.51	0.77
Constraint Space Solution 6S	4.91	5.07	1.76

Table D. 2: Count measure Scenario 3A

Methodology	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_9	Total
Ordinary Least Square	30	30	30	30	24	11	15	30	200
Independent Subsets	30	30	30	30	23	11	27	30	211
Simple Regression	30	30	30	29	19	4	29	26	197
Ridge Regression	30	30	30	30	29	26	28	30	233
Constraint Space Solution 1S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 3S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 6S	30	30	30	30	30	30	30	30	240

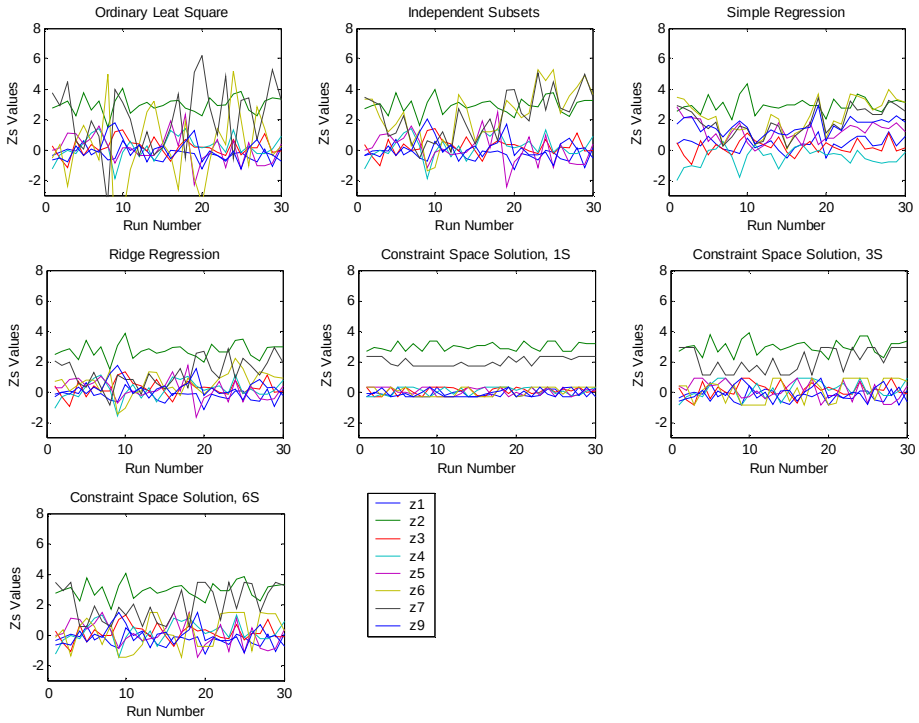


Figure D. 3: Methodology for Scenario 3B

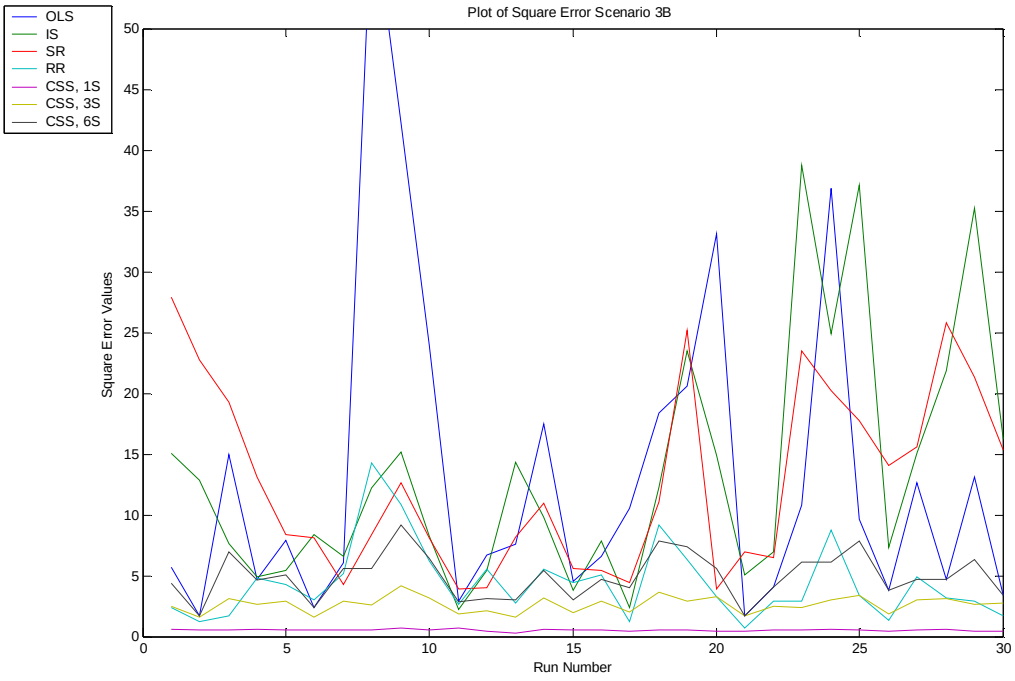


Figure D. 4: Square Error Methods Comparison for Scenario 3B

Table D. 3: Square Error measure Scenario 3B

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	13.33	7.74	13.95
Independent Subsets	13.36	11.00	9.94
Simple Regression	12.72	10.98	7.53
Ridge Regression	4.39	3.31	3.06
Constraint Space Solution 1S	0.49	0.49	0.08
Constraint Space Solution 3S	2.61	2.67	0.66
Constraint Space Solution 6S	4.90	4.69	1.88

Table D. 4: Count measure Scenario 3B

Methodology	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_9	Total
Ordinary Least Square	30	30	30	29	27	17	15	29	207
Independent Subsets	30	30	30	29	27	11	23	28	208
Simple Regression	30	30	30	28	21	10	28	17	194
Ridge Regression	30	30	30	30	28	29	24	29	230
Constraint Space Solution 1S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 3S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 6S	30	30	30	30	30	30	30	30	240

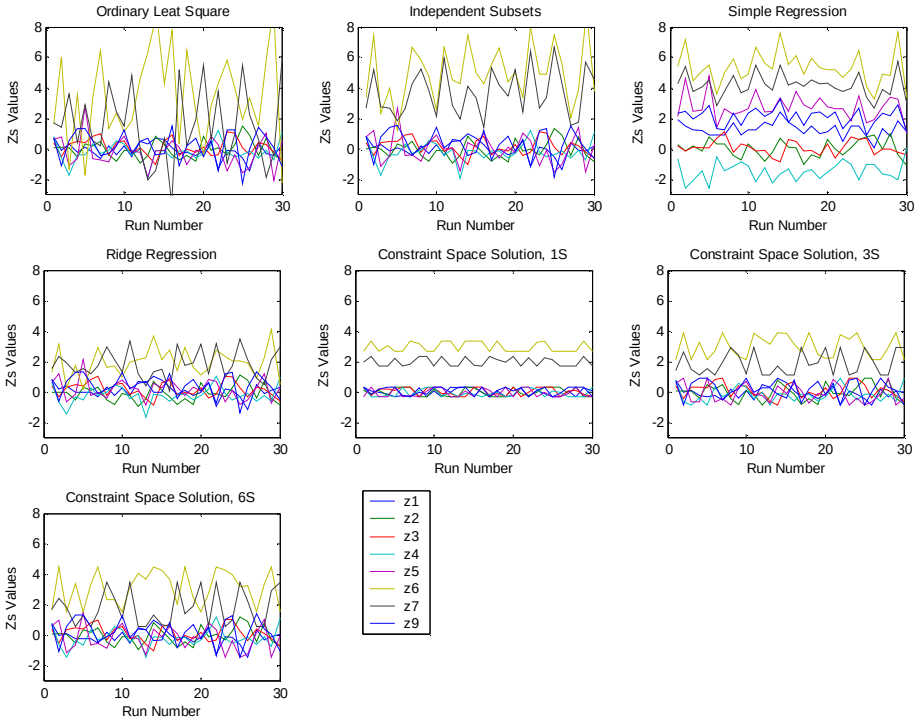


Figure D. 5: Methodology for Scenario 3C

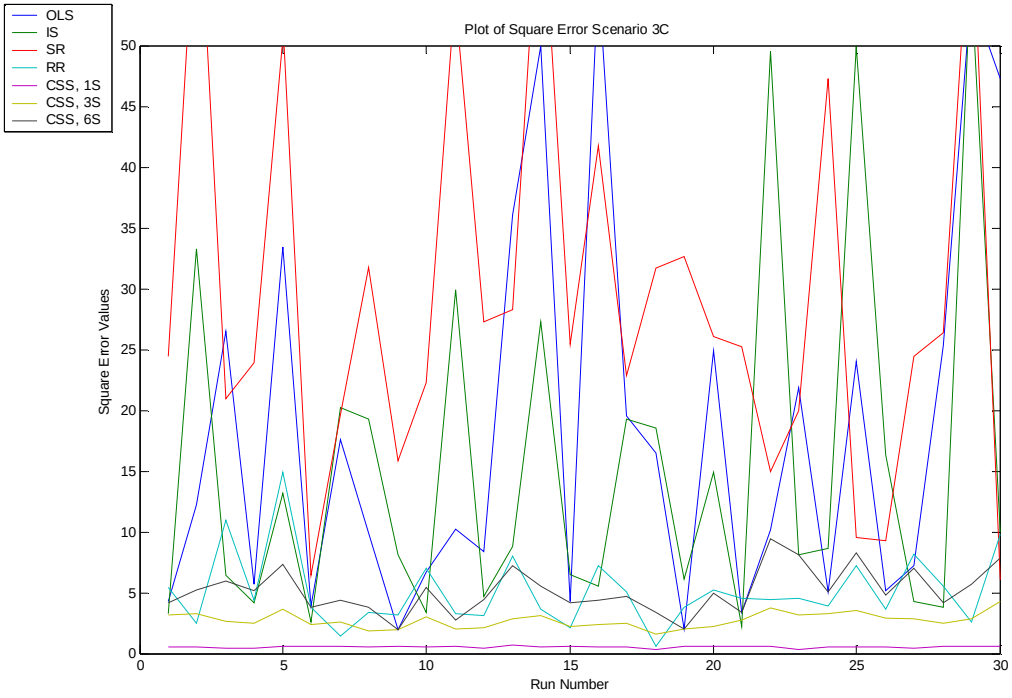


Figure D. 6 Square Error Methods Comparison for Scenario 3C

Table D. 5: Square Error measure Scenario 3C

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	18.45	11.19	16.38
Independent Subsets	15.43	8.71	14.84
Simple Regression	29.43	25.33	16.66
Ridge Regression	5.09	4.38	3.01
Constraint Space Solution 1S	0.51	0.52	0.08
Constraint Space Solution 3S	2.71	2.65	0.62
Constraint Space Solution 6S	5.13	4.85	1.84

Table D. 6: Count measure Scenario 3C

Methodology	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_9	Total
Ordinary Least Square	30	30	30	28	27	14	14	29	202
Independent Subsets	30	30	30	28	29	13	16	29	205
Simple Regression	24	30	30	18	2	6	5	7	122
Ridge Regression	30	30	30	29	29	23	29	30	230
Constraint Space Solution 1S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 3S	30	30	30	30	30	30	30	30	240
Constraint Space Solution 6S	30	30	30	30	30	30	30	30	240

Table D. 7: Count measure Scenario 3 Summary

Methodology	A	B	C	Mean
	Count	Count	Count	
Ordinary Least Square	200	207	202	203
Independent Subsets	211	208	205	208
Simple Regression	197	194	122	171
Ridge Regression	233	230	230	231
Constraint Space Solution 1S	240	240	240	240
Constraint Space Solution 3S	240	240	240	240
Constraint Space Solution 6S	240	240	240	240

Table D. 8: Square Error measure Scenario 3 Summary

Methodology	A	B	C	Mean
	MSE	MSE	MSE	
Ordinary Least Square	13.52	13.33	18.45	15.10
Independent Subsets	10.55	13.36	15.43	13.11
Simple Regression	12.30	12.72	29.43	18.15
Ridge Regression	3.87	4.39	5.09	4.45
Constraint Space Solution 1S	0.50	0.49	0.51	0.50
Constraint Space Solution 3S	2.66	2.61	2.71	2.66
Constraint Space Solution 6S	4.91	4.90	5.13	4.98

APPENDIX E RESULTS SCENARIO 4

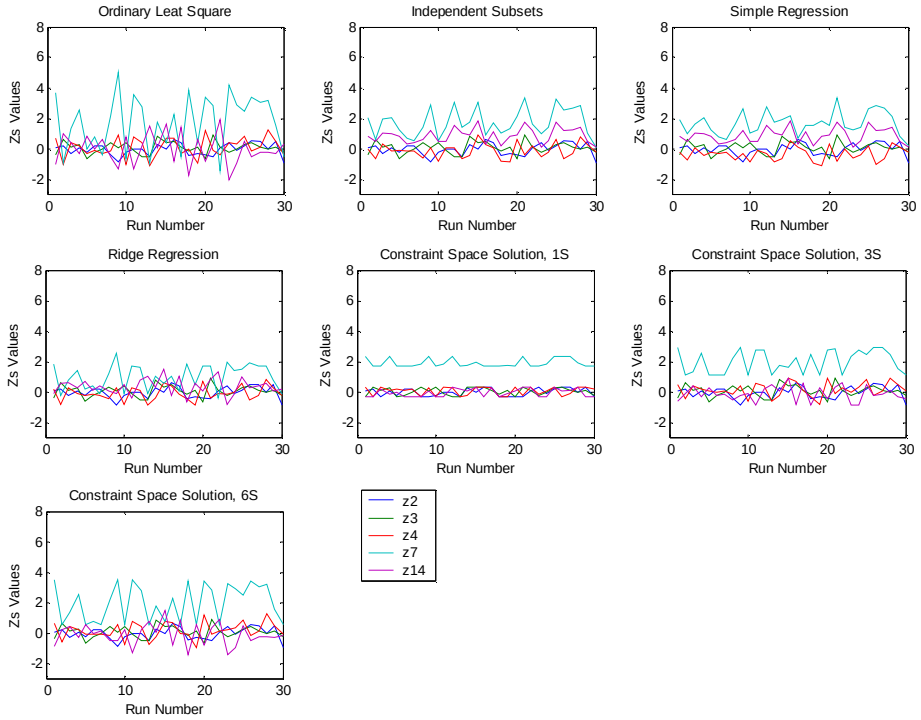


Figure E. 1: Methodology for Scenario 4A

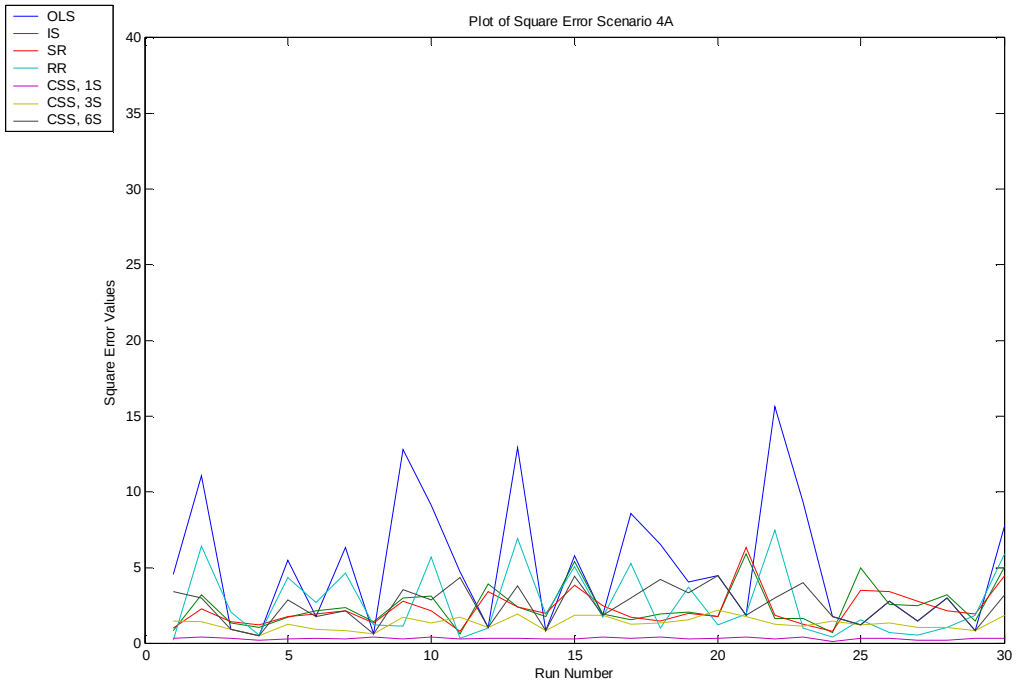


Figure E. 2: Square Error Methods Comparison for Scenario 4A

Table E. 1: Square Error measure Scenario 4A

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	4.96	4.26	4.23
Independent Subsets	2.41	1.96	1.40
Simple Regression	2.25	1.93	1.19
Ridge Regression	2.63	1.74	2.25
Constraint Space Solution 1S	0.30	0.32	0.07
Constraint Space Solution 3S	1.30	1.29	0.42
Constraint Space Solution 6S	2.51	2.85	1.24

Table E. 2: Count measure Scenario 4A

Methodology	z_2	z_3	z_4	z_7	z_{14}	Total
Ordinary Least Square	30	30	30	18	27	135
Independent Subsets	30	30	30	29	27	146
Simple Regression	30	30	30	29	27	146
Ridge Regression	30	30	30	22	30	142
Constraint Space Solution 1S	30	30	30	30	30	150
Constraint Space Solution 3S	30	30	30	30	30	150
Constraint Space Solution 6S	30	30	30	30	30	150

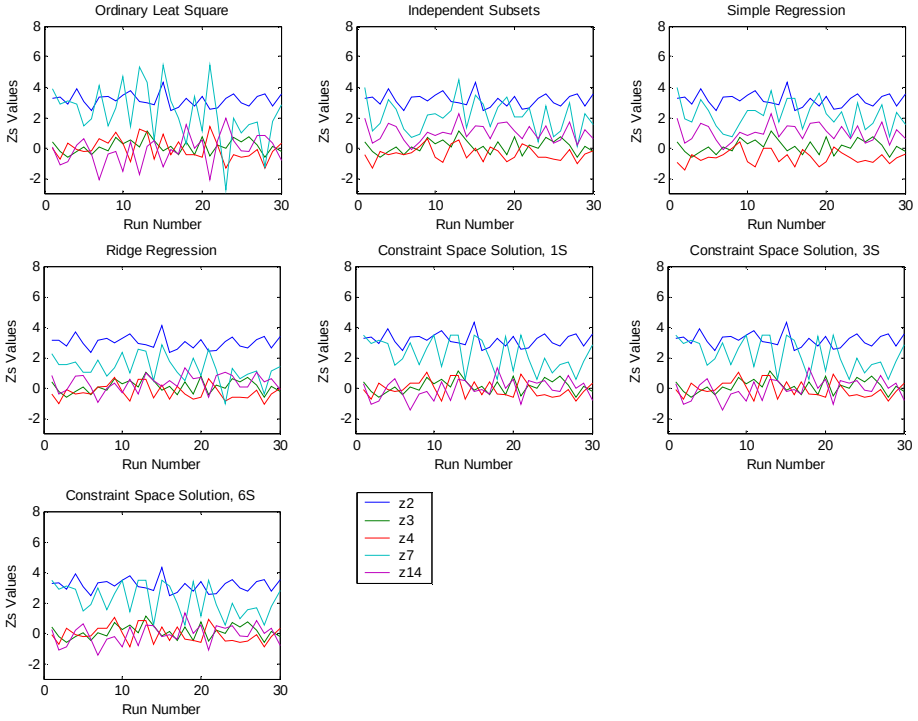


Figure E. 3: Methodology for Scenario 4B

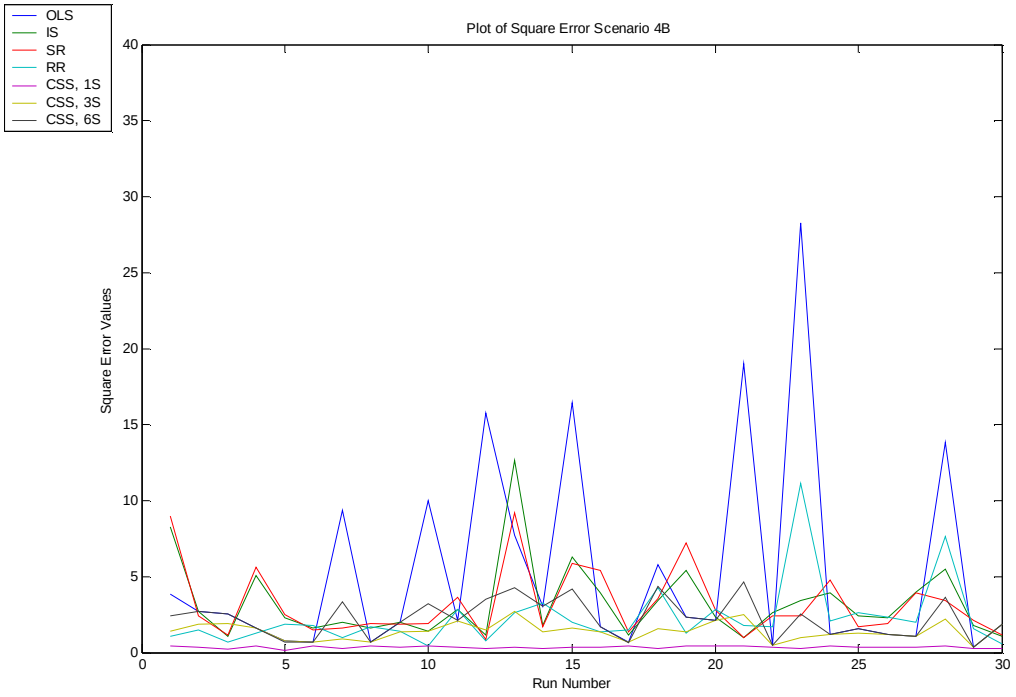


Figure E. 4: Square Error Methods Comparison for Scenario 4B

Table E. 3: Square Error measure Scenario 4B

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	5.34	2.12	6.85
Independent Subsets	3.20	2.37	2.50
Simple Regression	3.17	2.36	2.27
Ridge Regression	2.27	1.70	2.15
Constraint Space Solution 1S	0.32	0.33	0.08
Constraint Space Solution 3S	1.38	1.35	0.58
Constraint Space Solution 6S	2.20	2.12	1.25

Table E. 4: Count measure Scenario 4B

Methodology	z_2	z_3	z_4	z_7	z_{14}	Total
Ordinary Least Square	30	30	30	20	26	136
Independent Subsets	30	30	30	27	27	144
Simple Regression	30	30	30	28	27	145
Ridge Regression	30	30	30	28	30	148
Constraint Space Solution 1S	30	30	30	30	30	150
Constraint Space Solution 3S	30	30	30	30	30	150
Constraint Space Solution 6S	30	30	30	30	30	150

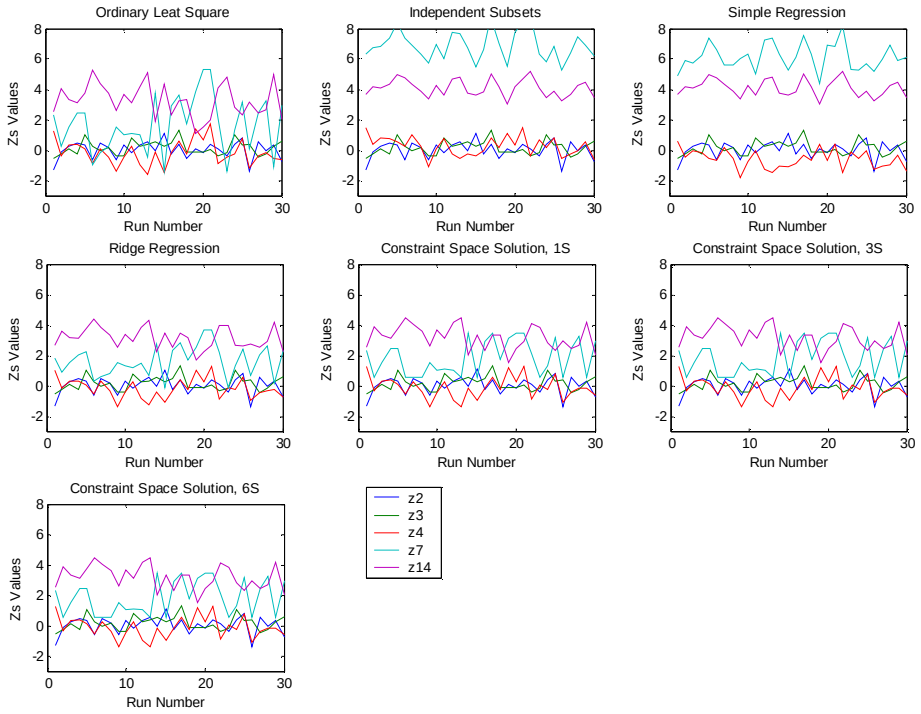


Figure E. 5: Methodology for Scenario 4C

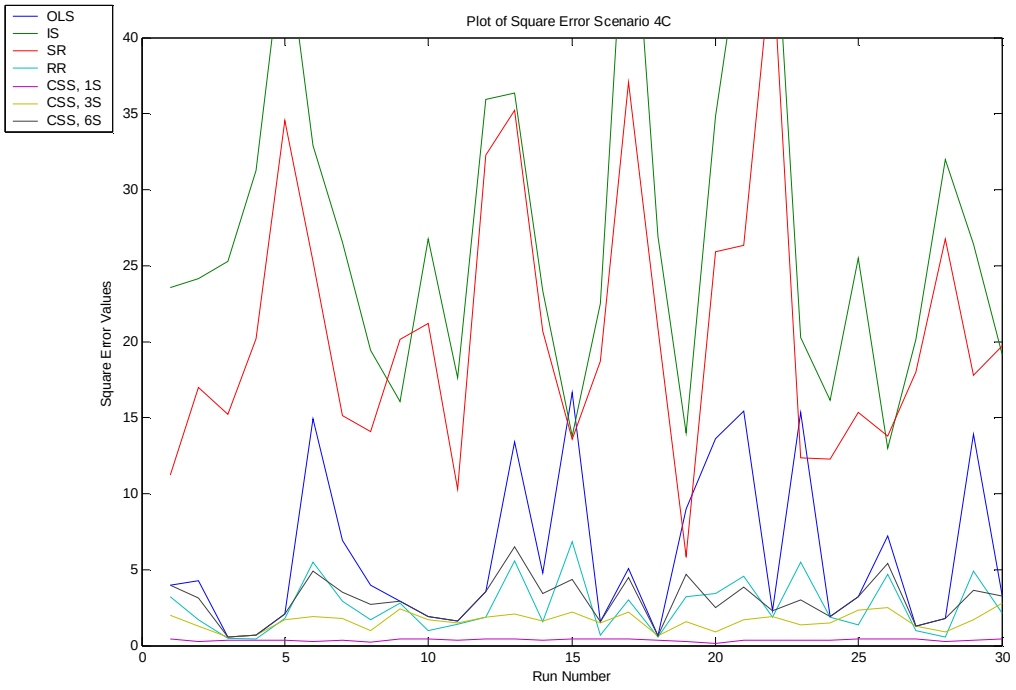


Figure E. 6: Square Error Methods Comparison for Scenario 4C

Table E. 5: Square Error measure Scenario 4C

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	5.89	3.73	5.35
Independent Subsets	27.37	25.36	10.97
Simple Regression	20.73	19.19	9.08
Ridge Regression	2.58	1.84	1.80
Constraint Space Solution 1S	0.33	0.35	0.07
Constraint Space Solution 3S	1.61	1.66	0.57
Constraint Space Solution 6S	2.94	3.03	1.45

Table E. 6: Count measure Scenario 4C

Methodology	z_2	z_3	z_4	z_7	z_{14}	Total
Ordinary Least Square	30	30	29	17	25	131
Independent Subsets	30	30	30	0	23	113
Simple Regression	30	30	29	0	23	112
Ridge Regression	30	30	30	24	30	144
Constraint Space Solution 1S	30	30	30	30	30	150
Constraint Space Solution 3S	30	30	30	30	30	150
Constraint Space Solution 6S	30	30	30	30	30	150

Table E. 7: Count measure Scenario 4 Summary

Methodology	A	B	C	Mean
	Count	Count	Count	
Ordinary Least Square	135	136	131	134
Independent Subsets	146	144	113	134
Simple Regression	146	145	112	134
Ridge Regression	142	148	144	145
Constraint Space Solution 1S	150	150	150	150
Constraint Space Solution 3S	150	150	150	150
Constraint Space Solution 6S	150	150	150	150

Table E. 8: Square Error measure Scenario 4 Summary

Methodology	A	B	C	Mean
	MSE	MSE	MSE	
Ordinary Least Square	4.96	5.34	5.89	5.40
Independent Subsets	2.41	3.20	27.37	11.00
Simple Regression	2.25	3.17	20.73	8.72
Ridge Regression	2.63	2.27	2.58	2.49
Constraint Space Solution 1S	0.30	0.32	0.33	0.32
Constraint Space Solution 3S	1.30	1.38	1.61	1.43
Constraint Space Solution 6S	2.51	2.20	2.94	2.55

APPENDIX F RESULTS SCENARIO 5

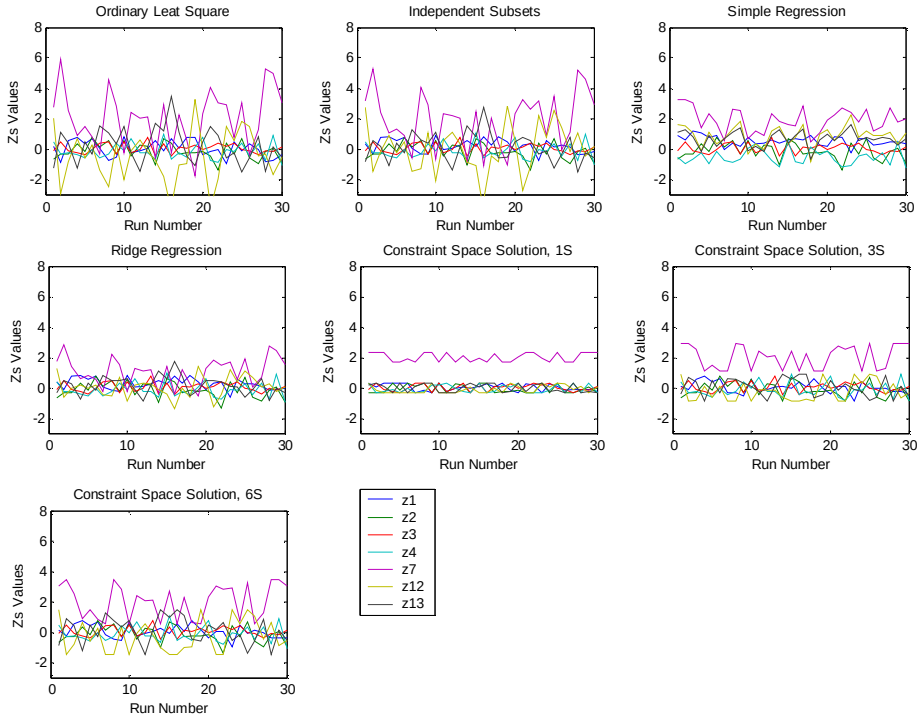


Figure F. 1: Methodology for Scenario 5A

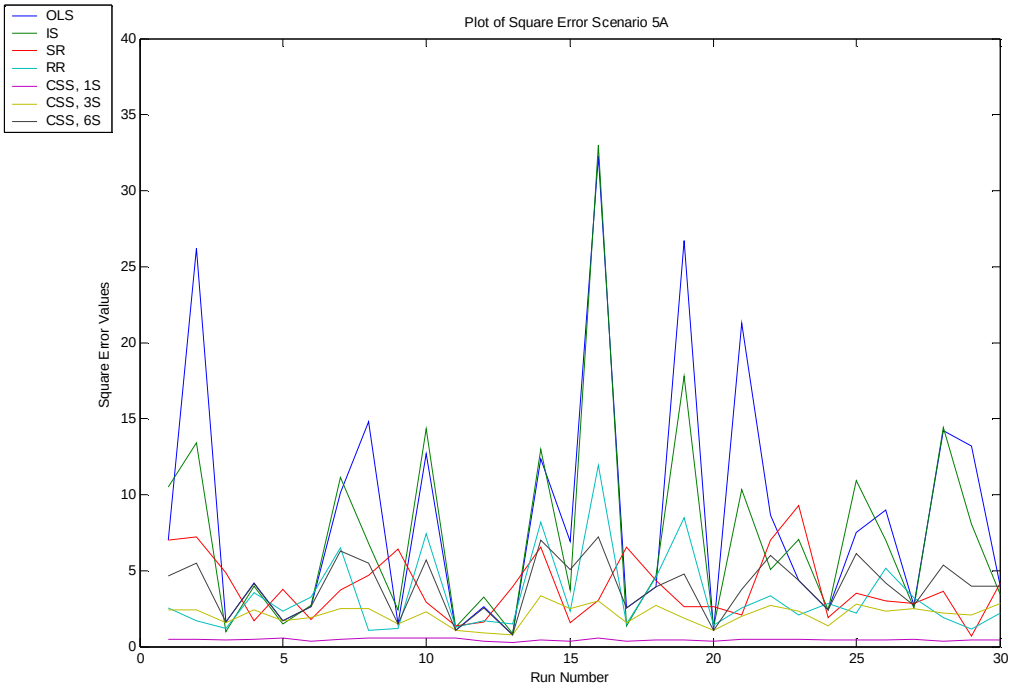


Figure F. 2: Square Error Methods Comparison for Scenario 5A

Table F. 1: Square Error measure Scenario 5A

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	8.64	5.62	8.45
Independent Subsets	7.27	4.81	6.87
Simple Regression	3.85	3.54	2.14
Ridge Regression	3.32	2.31	2.65
Constraint Space Solution 1S	0.42	0.43	0.08
Constraint Space Solution 3S	2.08	2.28	0.65
Constraint Space Solution 6S	3.91	4.04	1.86

Table F. 2: Count measure Scenario 5A

Methodology	z_1	z_2	z_3	z_4	z_7	z_{12}	z_{13}	Total
Ordinary Least Square	30	30	30	30	19	19	28	186
Independent Subsets	30	30	30	30	20	23	29	192
Simple Regression	30	30	30	30	30	28	30	208
Ridge Regression	30	30	30	30	23	30	29	202
Constraint Space Solution 1S	30	30	30	30	30	30	30	210
Constraint Space Solution 3S	30	30	30	30	30	30	30	210
Constraint Space Solution 6S	30	30	30	30	30	30	30	210

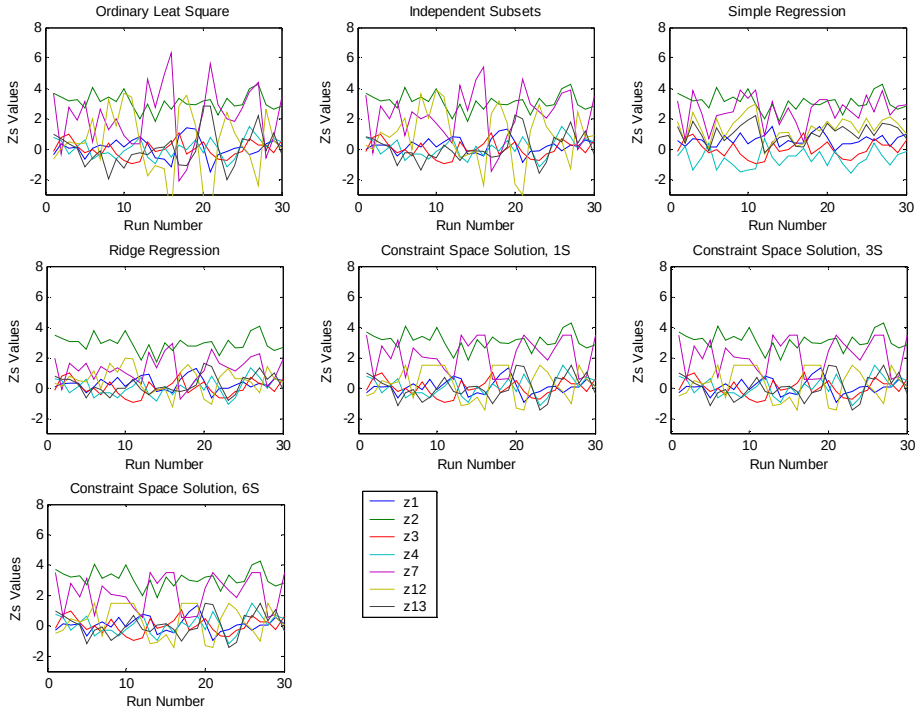


Figure F. 3: Methodology for Scenario 5B

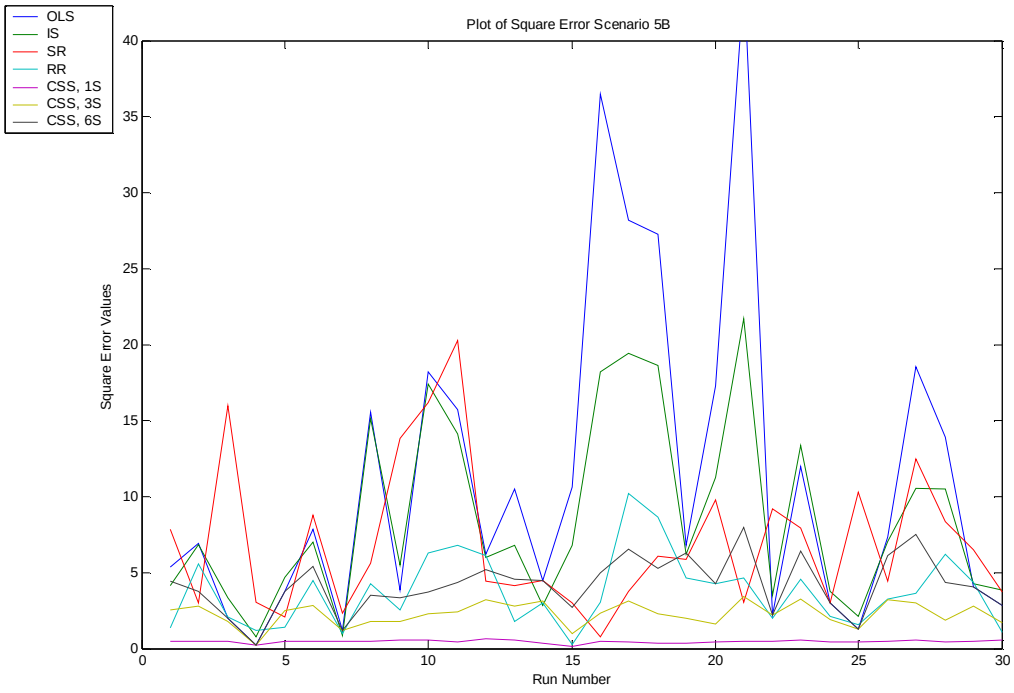


Figure F. 4: Square Error Methods Comparison for Scenario 5B

Table F. 3: Square Error measure Scenario 5B

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	11.22	7.13	10.80
Independent Subsets	8.52	6.73	6.03
Simple Regression	6.98	5.70	4.79
Ridge Regression	3.72	3.46	2.38
Constraint Space Solution 1S	0.44	0.47	0.10
Constraint Space Solution 3S	2.24	2.29	0.77
Constraint Space Solution 6S	4.17	4.29	1.84

Table F. 4: Count measure Scenario 5B

Methodology	z_1	z_2	z_3	z_4	z_7	z_{12}	z_{13}	Total
Ordinary Least Square	30	30	30	30	17	17	25	179
Independent Subsets	30	30	30	30	20	17	27	184
Simple Regression	30	30	30	30	26	16	25	187
Ridge Regression	30	30	30	30	25	28	30	203
Constraint Space Solution 1S	30	30	30	30	30	30	30	210
Constraint Space Solution 3S	30	30	30	30	30	30	30	210
Constraint Space Solution 6S	30	30	30	30	30	30	30	210

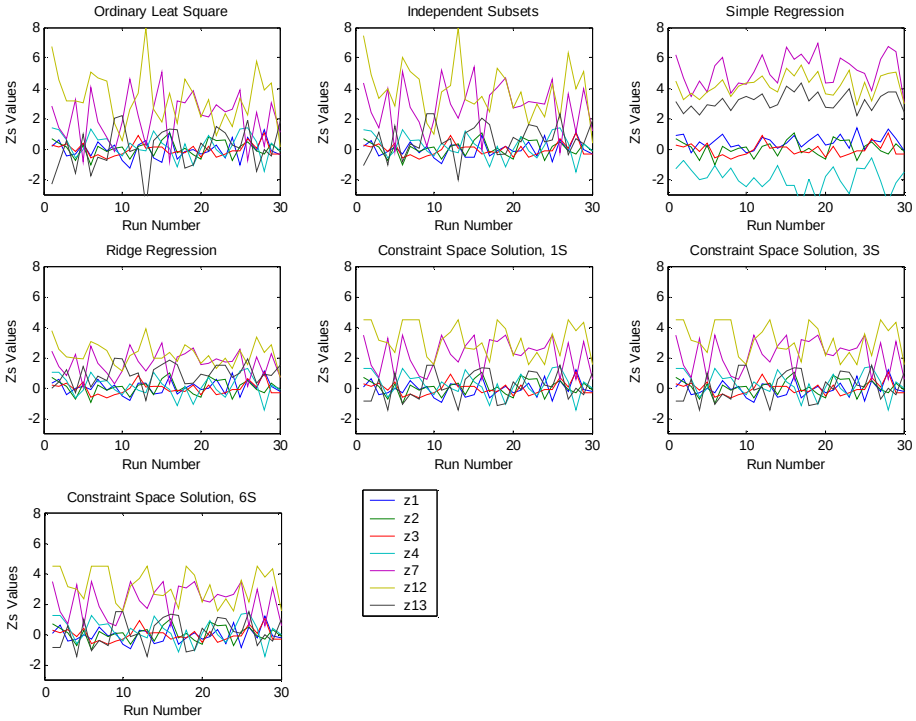


Figure F. 5: Methodology for Scenario 5C

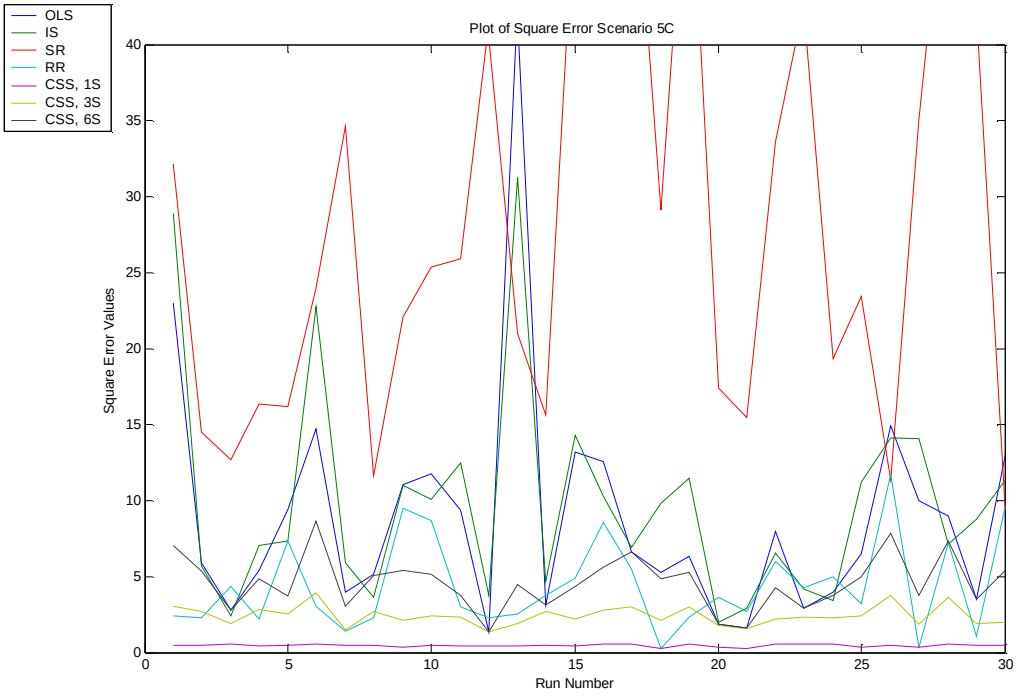


Figure F. 6: Square Error Methods Comparison for Scenario 5C

Table F. 5: Square Error measure Scenario 5C

Methodology	Square Error		
	Mean	Median	Std. Dev.
Ordinary Least Square	8.93	6.56	8.03
Independent Subsets	9.83	8.05	7.13
Simple Regression	28.29	24.67	14.20
Ridge Regression	4.36	3.37	2.98
Constraint Space Solution 1S	0.43	0.44	0.08
Constraint Space Solution 3S	2.39	2.34	0.64
Constraint Space Solution 6S	4.57	4.64	1.77

Table F. 6: Count measure Scenario 5C

Methodology	z_1	z_2	z_3	z_4	z_7	z_{12}	z_{13}	Total
Ordinary Least Square	30	30	30	30	19	22	23	184
Independent Subsets	30	30	30	30	17	20	24	181
Simple Regression	30	30	30	11	2	21	0	124
Ridge Regression	30	30	30	30	26	26	26	198
Constraint Space Solution 1S	30	30	30	30	30	30	30	210
Constraint Space Solution 3S	30	30	30	30	30	30	30	210
Constraint Space Solution 6S	30	30	30	30	30	30	30	210

Table F. 7: Count measure Scenario 5 Summary

Methodology	A	B	C	Mean
	Count	Count	Count	
Ordinary Least Square	186	179	184	183
Independent Subsets	192	184	181	186
Simple Regression	208	187	124	173
Ridge Regression	202	203	198	201
Constraint Space Solution 1S	210	210	210	210
Constraint Space Solution 3S	210	210	210	210
Constraint Space Solution 6S	210	210	210	210

Table F. 8: Square Error measure Scenario 5 Summary

Methodology	A	B	C	Mean
	MSE	MSE	MSE	
Ordinary Least Square	8.64	11.22	8.93	9.60
Independent Subsets	7.27	8.52	9.83	8.54
Simple Regression	3.85	6.98	28.29	13.04
Ridge Regression	3.32	3.72	4.36	3.80
Constraint Space Solution 1S	0.42	0.44	0.43	0.43
Constraint Space Solution 3S	2.08	2.24	2.39	2.24
Constraint Space Solution 6S	3.91	4.17	4.57	4.21

Table F. 9: Ridge Regression k estimation per Scenario

Methodology	A	B	C
Scenario 1	.2	.25	.3
Scenario 2	.2	.3	.4
Scenario 3	.2	.3	.4
Scenario 4	.3	.3	.3
Scenario 5	.3	.35	.4
Scenario 6	.2	.25	.3