

# CHAPTER ONE

## INTRODUCTION

### 1.1 BACKGROUND OF THE STUDY

One of the assumptions made in estimating the parameters of the linear regression model using ordinary least square (OLS) method is that the explanatory variables  $X_i$ ,  $i = 1, 2, \dots, n$  are not linearly correlated.

However, in practice, this is violated because there is a degree of inter-correlation among the independent variables due to the interdependence of many factors. This intercorrelation is called multicollinearity (Frisch, 1934). In a situation where there is presence of multicollinearity, the estimate of the regression coefficient will be indeterminate and the variances of these estimates become inflated (Nduka 1999). Also, multicollinearity make the parameter estimate imprecise and unstable.

Furthermore, the estimated regression coefficients may become insignificant or have wrong signs and consequently will be sensitive to changes in the data. This is because when the independent variables are correlated, the estimated standard errors for the regression coefficients will be large, and value of t-statistics will be affected. The estimated regression coefficients with large standard errors will be unstable and an addition of more data points to the sample will cause a large change in the size of the regression coefficients and sometimes in the signs of the coefficients. When any of the coefficients changes sign from positive to negative or from negative to positive at model updating, the model will not produce a good result. Also, the values of regression coefficients may change radically on the removal or addition of a predictor variable in the equation, and the sign of the coefficients might even change and the sum of squares may also be affected.

Multicollinearity can be detected using any of the following tests:

(i) Bersley, Kuh, and Welch Test, (Bersley et al, 1980), (ii) Bunch-map test (Montgomery and Peck, 1992), (iii) Farrar-Glauber Test (Farrar and Glauber 1967), or (iv) Variance Inflation Factor (VIF) (Marquardt and Snee 1975) among others.

The problem of multicollinearity can be remedied using restricted least square methods, generalized least square methods, Theil and Golbeger mixed estimation techniques, principal component regression, and ridge regression techniques among others (Nduka, 1999).

Ridge regression has antecedents in the theory of numerical solution of non-linear least square problems. Levenberge (1944) suggested that when Taylor series approximation is applied to linearize the problem so that OLS can be applied, one should use the usual normal equation except for the coefficients of the principal diagonal, which increased by quantities proportional to the weighting factor. Hoerl (1962) discussed an adaptation of regression analysis method for examining a second order response surface called ridge regression that was proposed in 1959. Hoerl (1968) introduced the estimators for regression parameter which are now called ridge regression estimators, although the term was not used in their work. Hoerl and Kennard (1970a, 1970b and 1970c) introduced the estimators, proved several properties, defined the ridge trace, and illustrated applications of ridge regression methods. Marquardt (1970) provided the structure of ridge regression rules and their use in linear and non-linear regression estimation. Newhouse and Oman (1971) investigated mean square error of ridge estimators by Monte Carlo procedure. Mc Donald and Galarneau (1975) applied ridge regression method to a real life problem and Marquardt and Snee (1975) discussed the benefits of ridge regression in data analysis. Consequently, Hoerl et al (1975) proposed an adaptive rule for ridge regression estimation and found that the estimator has smaller mean square error than OLS estimator. Hockings (1976) employed variable selection methods to obtain ridge regression estimators.

Methods of choosing ridge estimator  $k$  from any data are in the literature ( see: Golstein and Smith (1974), Farebrother (1975), Guilkey and Murphy (1975), Obenchain (1975), Hoerl and Kennard (1975, 1976), Lawless and Wang (1976), Dempster et al (1977), Gibbons (1981), Saleh and Kibria, (1993), Crouze et al (1995), Kibria (2003), Pasha and Sha (2004), Muniz et al (2010), Dorugade and Kashid (2010), Li and Yang (2010) and many others. The central goal of their work is the minimization of mean square error of the ridge regression parameter.

## 1.2 THE RIDGE REGRESSION

In a multiple linear regression model, one considers the following model

$$Y = X\beta + \varepsilon \quad (1.1)$$

$$\text{where } X = \begin{pmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix} \text{ and } Y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$$

The matrix  $X$  contains the values of  $p$  predictor variables at each of  $n$  observations,  $Y$  is a vector of the observed values of the dependent variable,  $\beta$  is a vector of unknown coefficients and  $\varepsilon$  is a vector of experimental error with the properties  $E(\varepsilon)=0$  and  $E(\varepsilon'\varepsilon)=\sigma^2 I_n$ , where  $I_n$  is an identity matrix of order  $n$ . For convenience,  $X$  and  $Y$  are centred and scaled so that  $(X'X)$  has the form of a correlation matrix.

The ordinary least square (OLS) estimate of  $\beta$  is given by

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (1.2)$$

The estimate  $\hat{\beta}$  is chosen to minimize the residual sum of square

$$SSR(\beta) = (Y - X\hat{\beta})'(Y - X\hat{\beta}) \quad (1.3)$$

The properties of  $\hat{\beta}$  are that (i) it is unbiased, that is,  $E(\hat{\beta}) = \beta$  (ii) it has a minimum variance among all linear unbiased estimators (BLUE);  $Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ , which depends on the characteristics of the matrix,  $(X'X)$ . If this matrix is ill-conditioned (near dependency among various columns of  $X'X$  or  $\det(X'X) \approx 0$ ), the OLS method fails due to the problem of multicollinearity amongst others.

The idea in ridge regression method is to add a small positive number ( $k > 0$ ) to the diagonal elements of the matrix  $(X'X)$  in order to obtain a ridge estimator that shortens the length of the regressor vector.

Let  $\phi$  be the residual sum of square (SSR). Then

$$\phi = (Y - \hat{Y})'(Y - \hat{Y}) \quad (1.4)$$

Let  $\hat{Y} = X\hat{\beta}$  and let  $B$  be any estimate of the vector  $\beta$ , then

$$\phi = (Y - XB + XB - X\hat{\beta})'(Y - XB + XB - X\hat{\beta}) \quad (1.5)$$

$$= ((Y - XB) + X(B - \hat{\beta}))'((Y - XB) + X(B - \hat{\beta}))$$

$$= \left( (Y - XB)' + (B - \hat{\beta})' X' \right) ((Y - XB) + (B - \hat{\beta}))$$

$$= (Y - XB)'(Y - XB) + 2(B - \hat{\beta})' X'(Y - XB) + (B - \hat{\beta})' X'X(B - \hat{\beta}) \quad (1.6)$$

$$\text{But } 2((B - \hat{\beta})X'(Y - XB)) = 0$$

Then,

$$\phi = (Y - XB)'(Y - XB) + (B - \hat{\beta})' X'X(B - \hat{\beta}) \quad (1.7)$$

$$\phi_{\min} + \phi(B) \quad (1.8)$$

The value of  $\phi$  is the minimum value  $\phi_{\min}$  plus the value of the quadratic form in  $(B - \hat{\beta})$ . There is a continuum of values of  $B_0$  that will satisfy the relationship  $\phi_{\min} + \phi_0$ , where  $\phi_0 > 0$  is a fixed increment.

However, the average distances from  $\hat{\beta}$  to  $B$  is large if the eigenvalues of the matrix  $X'X$  are small (Hadi and Chatterjee, 2006). The more the ill-

conditioning of  $X'X$ , the more the length of  $\hat{\beta}$  is further from  $B$  without appreciable increase in the residual sum of squares value. It seems reasonable if one moves from the minimum sum of squares value to the direction that will shorten the length of the regression vector  $\hat{\beta}$ . That is, for a fixed value of  $\phi$  a single value of  $B$  is chosen to

Minimize  $B'B$

subject to

$$(B - \hat{\beta})' X'X (B - \hat{\beta}) = \phi_0, \quad \phi_0 > 0, \quad B \geq 0 \quad (1.9)$$

Converting equation (1.9) to unconstrained problem, we have

$$\text{Min } F = B'B + \frac{1}{k} \left[ (B - \hat{\beta})' X'X (B - \hat{\beta}) - \phi_0 \right] \quad (1.10)$$

where  $\frac{1}{k}$  is a multiplier. Then, a minimum value of  $B$  is attained if

$$\frac{\partial F}{\partial B} = 2B + \left( \frac{1}{k} \right) [2(X'X)B - 2(X'X)\hat{\beta}] = 0$$

$$2Bk + [2(X'X)B - 2(X'X)\hat{\beta}] = 0$$

$$Bk + [(X'X)B - (X'X)\hat{\beta}] = 0$$

$$Bk + (X'X)B = (X'X)\hat{\beta}$$

$$B[X'X + kI] = (X'X)(X'X)^{-1} X'Y$$

$$B = \frac{X'Y}{X'X + kI}$$

Hence,

$$B = \hat{\beta}(k) = [X'X + kI]^{-1} X'Y \quad (1.11)$$

where  $k$  is chosen in (1.11) to satisfy the constraint in (1.9) by a process known as ridge regression estimation and  $B$  is the ridge regression parameter designated as  $\hat{\beta}(k)$ . The constant  $k$  ( $k > 0$ ) is known as a "RIDGE" or biased parameter, and is estimated from the observed data (see Newhouse and Oman, 1971).

Hoerl and Kennard (1970a), in their pioneer work showed that as  $k$  increases from zero to infinity, the regression estimate tends to zero (this process is known as shrinkage).

These estimates  $\hat{\beta}(k)$ , for certain values of  $k$ , yield biased results and minimum mean square error  $MSE(\hat{\beta}(k))$ . However, the  $MSE(\hat{\beta}(k))$  depends on the unknown parameters  $k, \beta$  and  $\sigma^2$ .

### **1.3 STATEMENT OF THE PROBLEM**

The method of estimation of the biasing parameter  $k$  proposed by Khalaf and Shukur (2005) and Muniz et al (2010) used the maximum value of eigenvalues of the correlation matrix  $(\tilde{X}\tilde{X})$ . This method gives rise to the following problems:

- (i) if the eigenvalues are skewed positively or negatively the maximum value of eigenvalues may not be the true representation of the eigenvalues.
- (ii) the maximum eigenvalue chosen is an outlier or there may be other values that are outliers within the eigenvalues.

Hence, this work seeks to solve these problems by obtaining a true representation of the eigenvalues for estimating the biasing parameter  $k$ .

### **1.4 AIM AND OBJECTIVES OF THE STUDY**

The aim of this work is to obtain a true representation of the eigenvalues in the presence of skewed eigenvalues or outlier among the eigenvalues for estimating the biased parameter  $k$ .

While the objectives are:

- (i) to propose methods of estimating the biasing parameter  $k$  that will minimize the mean square error (MSE).

(ii) to compare the proposed estimators with the existing methods using mean square error (MSE) and prediction sum of square (PRESS).

### **1.5 SCOPE OF THE STUDY**

The Monte Carlo simulation method will be used to compare the performance of the proposed ridge regression estimators and some existing estimators using MSE and PRESS. A real life multicollinear data will also be used to investigate the validity of the proposed methods in terms of MSE and PRESS.

## CHAPTER TWO

### LITERATURE REVIEW

Several studies have been done on ridge regression parameter estimation and related topics in solving problems of multicollinearity in multiple regression. In this chapter, we shall discuss relevant literature on the ridge regression, ridge parameter estimation and multicollinearity.

Bata et al (2008) combined the criteria underlying the generalized ridge regression (GRR) and the Jackknife ridge regression (JRR) estimation to obtain a new estimate for the regression coefficients of linear regression when the response variables are collinear. The estimator is designated as the modified Jackknife ridge regression estimator (MJR). The performance of GRR and JRR estimations were equally studied using MSE criterion. It was observed that the MJR estimator has a smaller mean square error value than GRR and JRR estimators respectively.

Shi and Wang (1999) studied the local influence of observations in ridge regression model under the perturbation of variance of explanatory variables. Also they studied the effect of multicollinearity on the influential observation and a diagnostic for detecting the influential observations on the ridge parameter estimator. They observed that using the ridge estimator to replace least square estimate can reduce the effect of collinearity, and hence the influence of some observations can be reduced.

Cawley and Talbat (2002) proposed a novel reduced rank training algorithm for Kernel ridge regression models. This method demonstrates superior performance than that of least squares method. It also provides a plausible approach for large-scale regression problems as it is not necessary to store the entire Kernel matrix.



Newhouse and Oman (1971) stated that if the MSE is a function of  $\beta, \sigma^2, k$  and the explanatory variables are fixed, then the MSE is minimized when  $\beta$  is the normalized eigenvectors corresponding to the maximum eigenvalue of the matrix  $(X'X)$  subject to constraint  $\beta'\beta = 1$ . Hence, the coefficients  $\beta_1, \beta_2, \dots, \beta_p$  are the normalized eigenvectors corresponding to the maximum eigenvalues of the matrix  $(X'X)$  so that  $\beta'\beta = 1$ . The normalized eigenvectors corresponding to the minimum eigenvalue can as well be used, however the conclusion about the performance of the estimators in both cases do not differ greatly.

Kibria (2003) considered the relationship between  $\sigma^2$  and signal to noise ratio (variance of the random error) as

$$\rho = \frac{\beta' \beta}{\sigma^2} \quad (2.1)$$

Nordberg (1982) discussed the choice of optimal ridge estimator with respect to signal to noise ratio:

1. Suppose that the signal to noise ratio  $\rho = \frac{\beta' \beta}{\sigma^2}$  is small, then the optimal ridge estimator is very efficient compared to OLS estimator
2. Suppose that the signal-to-noise ratio is moderate and for large values of  $\sigma^2$  corresponding to large eigenvalues  $\lambda$ , of the matrix  $(X'X)$ , then the optimal ridge estimator is considerably more efficient than the ordinary least square estimator.
3. Suppose that the signal to noise ratio is moderate and for small eigenvalue, then optimal ridge estimator performed better than ordinary least square estimator.

Liu (2003) observed that the eigenvalues and eigenvectors of the correlation matrix indicate the degree of multicollinearity. An eigenvalue that approaches zero depicts a very strong linear dependency between regressors, while the

elements of the associated eigenvector display the weights of the corresponding regressor variables in the multicollinearity. Multicollinearity can be measured in terms of the ratio of the largest and the smallest eigenvalue. This quantity is the condition number of the correlation matrix:

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (2.2)$$

Large values of  $\kappa$  are indication of serious multicollinearity. A condition number of the correlation  $\kappa$  between 30 and 100 indicates a moderate correlation and a condition numbers greater than 100 suggest strong multicollinearity (Liu, 2003). The number of eigenvalues near zero indicates the number of collinearities detected among the regression variables.

Brown (1978) studied the ridge estimator  $\hat{\beta}(k) = (X'X + kI)^{-1}X'Y, k > 0$  without the usual assumption that the matrix  $(X'X)$  is of full rank. The estimator which had been studied in the literature is viewed as the limit of the ridge estimators as  $k \rightarrow 0^+$ . This concept was extended to linear models of arbitrary ranks. He further extended the works of Marquardt (1970) and Goldstein and Smith (1975) by examining the ridge estimator in the context of a linear model which is rank deficient.

Zang and Ibrahim (2005) conducted a simulation study on SPSS ridge regression in a situation where there is multicollinearity and when some variables are too important to be excluded from the analysis. They observed that the performance of the evaluated ridge regression estimator depends on the variance of the random error, the correlation among the explanatory variables and the unknown coefficient vector.

Cule and De Iorio (2013) stated that OLS is seldom appropriate in the context of genetic data due to high dimensionality of the data and the correlation structure

among the predictors. They adopted the ridge regression method to solve this problem by choosing the ridge parameter using the following algorithm:

1. Calculate the eigenvectors and eigenvalues of  $X'X$ :

$$X'X = Q\Lambda Q'$$

Here, columns of  $Q$  are the eigenvectors of  $X'X$  and  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{p-1}, \lambda_p)$  is a diagonal matrix with diagonal elements the eigenvalues of  $X'X$  in descending order.

2. Compute the principal components of  $X$  as  $Z = XQ$ , and the principal components regression (PCR) coefficients as

$$\hat{\alpha} = \Lambda^{-1}Z'Y$$

3. For  $r = 1, \dots, t$ , compute  $k_r$  as

$$k_r = \frac{r\hat{\sigma}^2}{\hat{\alpha}'_r\hat{\alpha}_r}$$

Where  $\hat{\alpha}_r$  is the  $r$ -length vector of the first  $r$  principal components regression coefficients,  $Z_r$  are the first  $r$  columns of  $Z$ , and

$$\hat{\sigma}_r^2 = \frac{(Y - Z_r\hat{\alpha}_r)'(Y - Z_r\hat{\alpha}_r)}{n - r} \quad (2.3)$$

Among possible  $r$ , choose  $r^*$  as the value of  $r$  that minimizes the quantity  $q$

$$q = r - \sum_j^t \frac{\lambda_j^2}{(\lambda_j + k)^2} \quad (2.4)$$

4. Denote  $k_r$  using the chosen value of  $r$  as  $k_{r^*}$  and use this in fitting the ridge estimates:

$$\hat{\beta}(k_{r^*}) = (X'X + k_{r^*}I)^{-1}X'Y \quad (2.5)$$

The motivation behind this method is to choose a shrinkage parameter that performs better than principal component regression model with the same degrees of freedom. This proposed method was applied to data on genetic variant and the result has the advantage of improving interpretability of the fitted coefficients by relating genetic variants to phenotype.

Shen et al (2013) proposed an algorithm using generalized ridge regression to solve problems associated with genome-wide association studies and genomic selection when fitting models with large number of parameters. They also proposed heteroscedastic effect model. The efficiency of the model for different data sizes were evaluated via simulation. The algorithm was implemented in R-  
Package and the result gives better prediction than Ordinary Ridge Regression.

Adnan et al (2006) compared the mean square errors of different methods of handling multicollinearity problems using Monte Carlo simulation. They observed that ridge regression and partial least square regression are generally effective in handling multicollinearity problem in specific cases of low and moderate number of regressors, and for high number of regressors the ridge regression performs better.

Collaghan and Chen (2008) presented a method for identifying coefficient estimates that are ill-conditioned. They used eigenvalues and condition indexes from a hypothetical regression model to illustrate the best practice for collinearity diagnostics. They observed that collinearity has to be detected and dealt with in any multiple regression models to avoid wrong interpretation of the regression parameters.

According to Neter et al (1996), ridge regression estimates tend to be stable in the sense that they are usually affected by small changes in the data on which the fitted regression is based. In contrast, ordinary least squares estimates may be highly unstable when the independent variables are highly multicollinear.

Also, the ridge estimated regression function at times provides good estimates of mean responses or predictions of new observations for levels of the independent variables outside the region of the observations on which the regression function is based. In contrast, the estimated regression functions based on ordinary least squares perform quite poorly in such instances. Of course, any estimation or prediction outside the region of the observations should always be made with great caution. Major limitation of ridge regression is that ordinary inference procedures are not applicable and exact distributional properties are not known. Another limitation is that the choice of the biasing constant  $k$  is subjective. While formal methods have been developed for making this choice, these methods have their own limitations.

Rawlings et al (1988) showed that multicollinearity does not affect the precision of the estimated responses (and predictions) at the observed points in the explanatory variable  $X$ , rather it causes variance inflation of estimated responses at other points. He further suggested that ridge regression affects the multicollinearity by reducing the apparent magnitude of the correlations.

According to Myers (1990), ridge regression is the most popular estimation procedure for solving multicollinearity problems. Under the condition of multicollinearity, a huge price is paid for the unbiasedness property that one achieves by using ordinary least squares. Biased estimation is used to attain a substantial reduction in variance with an accompanied increase in stability of the regression coefficients. The coefficients become biased and, the reduction in variance is of greater magnitude than the bias induced in the estimators.

Draper and Smith (1981) stated that the use of ridge regression is perfectly sensible in circumstances that large regression coefficients are unrealistic from a practical point of view. However, they advised against the indiscriminate use of ridge regression unless its limitations are fully appreciated.

Ryan (1997) states that for all practical purposes, the ordinary least square (OLS) estimator is biased, and that ridge regression permits the use of a set of regressors that may seem appropriate in reducing multicollinearity in regression analysis.

Khalaf (2012) proposed a new estimator of the ridge parameter  $k$  by using the residual sum of squares (SSR) to estimate the variance of the random error as follows

$$\hat{\sigma}^{*2} = \frac{SSR}{n - p + 2} \quad (2.6)$$

He further suggested that dividing the SSR with  $n-p+2$  rather than  $n-p$  as suggested by Rao (1973) yields an estimate of  $\sigma^2$  with smaller MSE

Furthermore, using equation (2.6) in Hoerl and Kennard (1970b) and Hoerl et al (1975) yields the following estimators

$$k_{(1)}^* = \frac{\hat{\sigma}^{*2}}{\hat{\beta}_j^2} \quad (2.7)$$

$$k_{(2)}^* = \frac{p\hat{\sigma}^{*2}}{\hat{\beta}'\hat{\beta}} \quad (2.8)$$

The performance of these proposed estimators were encouraging when compared with the methods proposed by Hoerl and Kennard (1970b) and Hoerl et al (1975) from the MSE point of view.

In another development, Khalaf (2012) proposed another new estimator by modifying Hoerl and Kennard (1970b) by adding the reciprocal of the mean of the maximum and the minimum eigenvalues to the estimator of  $k$  as proposed by Hoerl and Kennard (1970a) as follows

$$\begin{aligned}\hat{k}_{(GK)} &= \frac{\hat{\sigma}^2}{\max(\hat{\alpha}_i^2)} + \frac{1}{\frac{1}{2}(\lambda_{\max} + \lambda_{\min})} \\ &= \frac{\hat{\sigma}^2}{\max(\hat{\alpha}_i^2)} + \frac{2}{(\lambda_{\max} + \lambda_{\min})}\end{aligned}\tag{2.9}$$

where  $\lambda_{\max}$  is the maximum eigenvalue and  $\lambda_{\min}$  is the minimum eigenvalue. He further investigated the performance of the proposed estimators using the Monte Carlo simulation, where levels of correlation, number of explanatory variables and sample sizes were varied and each combination was replicated 5,000 times. The performance of the proposed estimators was evaluated using mean square error criteria.

Alkassab and Qwaider (2010) compared the performance of least squares method and ridge regression method using Monte Carlo Simulation; they created fourteen independent variables with four different values of correlation between these variables and compared them using the MSE criterion. The result showed that the ridge regression method performed better than the least square method.

Lipovesky (2009) studied and compared the behavior of the coefficients of OLS regression with the coefficients regularized by the one-parameter ridge (Ridge-1) and the two-parameter ridge (Ridge-2) and observed that:

- The ridge models are not prone to multicollinearity.
- The fit quality of Ridge-2 does not decrease with the profile parameter increase, but the ridge-1 model converges to a solution proportional to the coefficients of paired correlation between the dependent variables and prediction.
- The correlation-regression (CORE) model suggests meaningful coefficients and net effects for the individual impact of the predictors,

high quality model fit, and convenient analysis and interpretation of the regression.

Perperoglou (2013) suggested that to account for dynamic behaviour of fixed covariates, penalized Cox model can be employed in models with interactions of the covariates and known time function. He discussed some of the penalised model with time varying effect such as ridge penalty on time-varying effects, fixed time-varying effect and dynamic ridge penalty on time-varying effects. Consequently, he suggested using flexible time function such as B-spline to constraint the behaviour of the models by adding proper penalties

Yan (2008) proposed a modified non-linear generalised ridge regression (MNGRR) to model highly non-linear system. In practice MNGRR is applied to develop Naphtha 95% cut off point sensor due to the existence of highly non-linear relationship between process variable and naphtha 95% cut off point in atmospheric distillation unit. The fact remains that few on-line hardware sensors are available. The result obtained showed that the performance of MNGRR is better than that of linear regression, non-linear ordinary least square regression and non-linear traditional ridge regression.

Nagaraj and Thiagarasu (2014) introduced directed ridge regression to predicted two or more variables which are highly correlated in high dimensional document space. Directed ridge regression is a statistical technique to estimate the relationship among the variables based on the eigenvalues to find the similarities between the documents. This alters the diagonal of the eigenvalues. Using experimental results to compare the correlation preserving index with the directed ridge regression, they observed that the directed ridge regression achieves efficient document clustering.

Nja (2013) proposed an nth-order Jackknife ridge estimator using canonical parameters transformation. He computed the parameters, biases and variance of



the estimators to show their behaviour and strength. Furthermore, he compared the  $n$ th-order Jackknife ridge regression estimator with generalised ridge regression, Jackknife ridge, second-order Jackknife ridge regression methods and observed that higher order Jackknife ridge regression estimators are superior to lower order Jackknife ridge estimators in terms of smaller bias.

Ishwaran et al (2010) showed that the estimator of weighted generalised ridge regression (WGRR) offers a unique advantage in correlated high-dimensional problems and can be effectively obtained using Bayesian Spike and Slab models which are effective for prediction. Also in a sparse variable selection a generalization of the elastic net can be used in tandem with these Bayesian estimates to solve the problem of multicollinearity and select the variables correctly.

Dorugade (2014) proposed a ridge estimator by modifying the work of Alkhamisi and Shurkur (2007) and obtained an improved ridge regression estimator with smaller MSE. This was illustrated via simulation studies and numeric example with only 4 explanatory variables.

Sinan and Genc (2012) compared the commonly used methods for choosing the ridge parameter using an actual data set obtained from General Directorate of Turkish Highways and Turkish Statistical Institute. They obtained the ridge parameters using the ridge trace, ordinary ridge estimator and iterative method for ordinary ridge estimation and compared their performances using MSE. They observed that ridge trace method is more useful than other methods.

Kalatzis et al (2008) investigated the investment decision considering the presence of financial constraint of 377 large Brazilian firms from 1997 to 2004 using panel data. Using Bayesian econometric model classified by the firm capital intensity and considering ridge regression for the problem of

multicollinearity among the model variables observed that the presence of multicollinearity provides important changes in the regression parameter.

Wu and Liu (2014) considered several estimators for estimating the stochastic restricted ridge regression. This method considered was compared in a simulation study with other existing methods; the result showed that the stochastic restricted ridge estimator performed better than other existing ones in terms of smaller MSE.

Cheng and Wu (2006) observed that when dealing with the predicting problem from automobile market, that the result from partial least square regression (PLSR) appear to be unstable which is due to the impact of the relevant information contained in the explanatory variables that is irrelevant to the response variable. In view of this, they proposed a method known as modified partial least square regression (MPLSR) to solve this problem of instability. A Monte Carlo experiment was employed to compare the proposed methods and the existing ones. The result showed that MPLSR is the most stable and accurate method especially when the ratio of the number of observation and the number of explanatory is low.

El-Dereny and Rashwan (2011) considered many different methods of ridge regression estimation to solve multicollinearity problem. The methods considered include ordinary ridge regression, generalised ridge regression and directed ridge regression. Data simulation was used to make comparisons between these method and OLS. The result showed that the ridge regression method is better than OLS in the presence of multicollinearity.

However, none of these methods considered the situation of skewness of eigenvalues and presence of outlier among the eigenvalues of a correlation matrix  $(\tilde{X}\tilde{X})$  in estimation of ridge regression parameter  $k$  that minimizes the mean square error. Hence, this work contributes to the existing literature by proposing a model that considers skewness and outlier among the eigenvalues of the correlation matrix in estimating the ridge parameter.

## CHAPTER THREE

### METHODOLOGY

#### 3.0 INTRODUCTION

In this chapter, we shall look at the general form of linear regression model, ridge regression estimation and methods of estimating the ridge parameter  $k$ . We shall also present the proposed methods of estimating the ridge parameter  $k$

#### 3.1 GENERAL FORM OF MULTIPLE LINEAR REGRESSION MODELS

Linear regression model is a model in which the regression function is linear in parameters and is given as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (3.1)$$

where  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  are referred to as the regression coefficients and  $\varepsilon$  is the random disturbance or error term.

Suppose we have  $n$  observations of a response variable  $Y$  and  $p$  explanatory variables  $X_1, X_2, \dots, X_p$ , it is convenient to arrange the data in an array as shown in Table 3.1.

**Table 3.1 Array of data consisting of n-observations of a response variable on p-explanatory variables**

Observation Number	Y	X <sub>1</sub>	X <sub>2</sub>	.....	X <sub>p</sub>
1	y <sub>1</sub>	x <sub>11</sub>	x <sub>12</sub>	.....	x <sub>1p</sub>
2	y <sub>2</sub>	x <sub>21</sub>	x <sub>22</sub>	.....	x <sub>2p</sub>
.	.	.	.	.....	.
.	.	.	.	.....	.
.	.	.	.	.....	.
.	.	.	.	.....	.
n	y <sub>n</sub>	x <sub>n1</sub>	x <sub>n2</sub>	.....	x <sub>np</sub>

Accordingly, we write

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (3.2a)$$

where  $y_i$  represents the  $i$ th value of the response variable  $Y$ .  $x_{i1}, x_{i2}, \dots, x_{ip}$  represent values of the predictor variables for the  $i$ th unit ( $i$ th row in Table 3.1), and  $\varepsilon_i$  represents the error in the approximation of  $y_i$ .

The standard multiple regression model in matrix notation is given as

$$Y = X\beta + \varepsilon, \quad (3.2b)$$

where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \cdot & \cdot & x_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix}$$

### 3.2 THE ORDINARY LEAST SQUARE (OLS) METHOD OF ESTIMATING REGRESSION PARAMETERS

Ordinary least squares (OLS) is a technique that minimizes the sum of squared distance (sum of square error term) between the observed values and the predicted value ( $y - \hat{y}$ ), if the predictor variables are correlated, the model coefficient of the OLS is indeterminate, so the values of the dependent variable cannot be predicted, therefore a regularization parameter is added to the OLS to obtain the estimate of the model parameter.

OLS technique is an unbiased linear estimation technique of any data that are linearly related. It also has the smallest variance among the unbiased estimators (BLUE). However, this is true under certain assumptions which are summarized as follows:

### Assumptions about the Predictor Variables

- The predictor variables  $X_1, X_2, \dots, X_p$  are non-random, that is the values  $x_{1j}, x_{2j}, \dots, x_{nj}$ ;  $j = 1, 2, \dots, p$  are assumed fixed.
- The values  $x_{1j}, x_{2j}, \dots, x_{nj}$ ;  $j = 1, 2, \dots, p$  are measured without error. The errors in measurement will affect the residual variance, the multiple correlation coefficients, and the individual estimate of the regression coefficients.
- The predictor variables  $X_1, X_2, \dots, X_p$  are assumed to be linearly independent of each other. This assumption is needed to guarantee the uniqueness of the least squares solution. If this assumption is violated, the problem is termed collinearity problem (Neter et al 1996)

The first two of the above assumptions about the predictors cannot be validated, so they do not play a major role in the analysis. However, they do influence the interpretation of the regression results.

Violation of any of these assumptions leads to a biased estimator. In practice, it is very difficult to satisfy the above assumptions especially the predictor variables  $X_1, X_2, \dots, X_p$  are assumed to be linearly independent of each other.

The OLS estimation of parameters  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  is given as follows:

$$\hat{Y} = X\hat{\beta} + \varepsilon, \quad (3.3)$$

where

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)' \quad (3.4)$$

$$X = (1, X_1, X_2, \dots, X_p) \quad (3.5)$$

and  $\varepsilon \sim N(0, \sigma^2)$ .  $\hat{\beta}$  is the estimate of of the parameter  $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)'$

In OLS regression, the parameters of the model are determined by minimizing the error between the observed and the predicted values as follows:

Let  $B$  be the set of all possible vector  $\beta$ . The objective of OLS is to find a vector  $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$  from  $B$  that minimizes the sum of squared residuals

$$S(\beta) = \sum_{i=1}^k \varepsilon_i^2 = \varepsilon' \varepsilon = (y - X\beta)'(y - X\beta) \quad (3.6)$$

Given  $y' = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$  and  $X = \begin{pmatrix} 1 & x_{11} & \cdot & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & \cdot & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \cdot & \cdot & \cdot & x_{np} \end{pmatrix}$

A minimum sum of squares residual will always exist since  $S(\beta)$  is a real valued, convex, differential function. If we expand  $S(\beta)$  in (3.6) we obtain

$$S(\beta) = y'y + \beta'X'X\beta - 2\beta'X'y \quad (3.7)$$

We obtain the derivatives of  $S(\beta)$  with respect to  $\beta$  as follows

$$\frac{\partial S(\beta)}{\partial \beta} = 2X'X\beta - 2X'Y \quad (3.8)$$

Equating (3.8) to zero yields the normal equation as follows

$$X'X\hat{\beta} = X'Y \quad (3.9)$$

If  $X'X$  is of full rank then  $X'X$  is non-singular and the unique solution of (3.9) is

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (3.10)$$

The variance-covariance of  $\hat{\beta}$  is obtained by substituting  $Y = X\beta + \varepsilon$  in equation (3.10) to obtain

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X'(X\beta + \varepsilon) \\ &= (X'X)^{-1} (X'X)\beta + (X'X)^{-1} X'\varepsilon \\ &= \beta + (X'X)^{-1} X'\varepsilon \end{aligned} \quad (3.11)$$

Therefore

$$\hat{\beta} - \beta = (X'X)^{-1} X'\varepsilon$$

By definition  $\text{Var}(\hat{\beta}) = E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right]$

$$\begin{aligned}
&= E\left\{\left((X'X)^{-1}X'\varepsilon\right)\left((X'X)^{-1}X'\varepsilon\right)\right\} \\
&= E\left\{\left((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right)\right\}
\end{aligned} \tag{3.12}$$

Since in matrix notation  $(AB)' = B'A'$  and also  $X$  is non stochastic, on taking expectation of (3.12) we have

$$\begin{aligned}
\text{Cov}(\hat{\beta}) &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\
&= (X'X)^{-1}(X'X)\sigma^2I(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned} \tag{3.13}$$

This is on the assumption that  $E(\varepsilon\varepsilon') = \sigma^2I$

When  $X'X$  is deficient in rank due to linear dependence among the explanatory variables leading to singularity of  $X'X$ , canonical form is frequently used (Yan and Su, 2009). This can be stated as follows:

If  $(X'X)_{p \times p}$  is symmetric, using spectral decomposition we have

$$X'X = D\Lambda D'$$

Where  $D = (d_1, \dots, d_p)$  is the orthogonal matrix of the standardized eigenvectors  $d_1, d_2, \dots, d_p$  of  $X'X$  and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  is the diagonal matrix of eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_p$  of  $X'X$  and  $DD' = I$ . The model

$Y = X\beta + \varepsilon_i$  can be re-written in canonical form as

$$\begin{aligned}
Y &= XDD'\beta + \varepsilon_i \\
&= X^*\alpha + \varepsilon_i
\end{aligned} \tag{3.14}$$

where  $X^* = XD$ ,  $\alpha = D'\beta$  and  $X^{*'}X^* = D'X'XD = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  so that the columns of the vector  $X^*$  are orthogonal.

The OLS estimate of  $\hat{\beta}$  in canonical form in model (3.14) is

$$\hat{\beta} = \Lambda^{-1} X' Y \quad (3.15)$$

The covariance matrix of  $\hat{\beta}$  is

$$V(\hat{\beta}) = \sigma^2 \Lambda^{-1} \quad (3.16)$$

### 3.3 THE RIDGE REGRESSION ESTIMATION

Ridge Regression, also known as Regularized Least Square Method is a technique intended to overcome the ill-conditioned situation in ordinary least squares (Draper and Smith 1998). The ill-condition occur when the correlation(s) between the explanatory variables in the design matrix is high, which implies that  $(X'X)^{-1}$  is rank deficient giving rise to unstable parameter estimates, typically with large standard error. Therefore, in ridge regression, additional information is added to remove the ill-conditioning. From Bayesian point of view, it is similar to placing a prior knowledge about the model coefficient of the form  $\beta \sim N(0, \lambda^{-1})$  and computing the mode of the posterior. Ridge regression estimation is obtained by the structural risk minimization of the error function that is both the sum of square error and the weights. This is known as zero order Tikhonov regularization (Rao and Toutemberg; 1995).

Now, given the scalar mean dispersion error (MDE)

$$M(\hat{\beta}, \beta) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \quad (3.17)$$

Then, the covariance matrix of an estimator  $\hat{\beta}$  is

$$V(\hat{\beta}) = E\left\{(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'\right\} \quad (3.18)$$

If  $E(\hat{\beta}) = \beta$ ,  $\hat{\beta}$  is an unbiased estimator of  $\beta$  otherwise  $\hat{\beta}$  is a biased estimator.

The difference between  $E(\hat{\beta})$  and  $\beta$  is called the bias and it is given as

$$Bias(\hat{\beta}, \beta) = E(\hat{\beta}) - \beta$$



If  $\hat{\beta}$  is unbiased, obviously the  $Bias(\hat{\beta}, \beta) = 0$

Hence, equation (3.17) is decomposed as follows

$$\begin{aligned} M(\hat{\beta}, \beta) &= E\left(\left(\hat{\beta} - E(\hat{\beta})\right) + \left(E(\hat{\beta}) - \beta\right)\right)\left(\left(\hat{\beta} - E(\hat{\beta})\right) + \left(E(\hat{\beta}) - \beta\right)\right)' \\ &= V(\hat{\beta}) + \left(Bias(\hat{\beta}, \beta)\right)\left(Bias(\hat{\beta}, \beta)\right)' \end{aligned} \quad (3.19)$$

Equation (3.19) shows that the scalar MDE of an estimator is the sum of the covariance matrix and the squared bias.

The scalar MDE of equation (3.17) is interpreted as the mean Euclidean distance between vector  $\hat{\beta}$  and  $\beta$  (Birkes and Dodge 1993). In the case of rank  $X'X = k < p$ , the OLS estimate has the minimum-variance property in the class of linear unbiased estimators.

We obtain the scalar mean dispersion error (MDE) of  $\hat{\beta}$  using the canonical form of equations (3.15) and (3.16) as follows (Rao and Toutenberg, 1995)

Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  denote the eigenvalues of the matrix  $(X'X)$ , then

$$tr\{M(\hat{\beta}, \beta)\} = tr\{V(\hat{\beta})\} = \sigma^2 \sum_{i=1}^p \lambda_i^{-1} \quad (3.20)$$

Hence, multicollinearity means a global unfavourable distance to the real parameter vector. Hoerl and Kernard (1970a) used the interpretation of the scalar MDE as basis for the definition of the ridge estimate as:

$$\hat{\beta}(k) = (X'X + kI)^{-1} X'Y \quad (3.21)$$

with the non-stochastic quantity  $k \geq 0$ , being the control parameter. For  $k = 0$ ,  $\hat{\beta}(0) = \hat{\beta}$  is the ordinary least square (OLS) estimate.

### 3.4 METHODS OF RIDGE REGRESSION PARAMETER ESTIMATION

Ridge regression provides alternative estimation method that may be used when the predictor variables are highly collinear. The estimators produced are biased

and tend to have a smaller MSE than the OLS estimation (Hoerl and Kennard, 1970b).

Consider the multiple regression

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon_i \quad (3.22)$$

Using the unit length scaling shown below:

$$\begin{aligned} \tilde{Y} &= \frac{Y - \bar{y}}{L_y}, \\ \tilde{X}_j &= \frac{X_j - \bar{x}_j}{L_j}, j = 1, 2, \dots, p \end{aligned} \quad (3.23)$$

where  $\bar{y}$  is the mean of  $Y$ ,  $\bar{x}_j$  is the mean of  $X_j$ , and

$$L_y = \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}, \text{ and } L_j = \sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, i = 1, 2, \dots, n \quad (3.24)$$

$$\text{such that } \sum_{i=1}^n x_{ij}^2 = 1, j = 1, 2, \dots, p$$

we obtain the following model

$$\tilde{Y} = \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \dots + \beta_p \tilde{X}_p + \varepsilon' \quad (3.25)$$

When the regressors are centered and scaled, the resulting constant  $\beta_0 = \bar{y}$  are not affected by the collinearity (Chatterjee and Hadi, 2006)

Equation (3.25) is written in matrix form as

$$\tilde{Y} = \tilde{X}\beta + \varepsilon \quad (3.26)$$

where

$$\tilde{Y} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}_n \end{bmatrix} \quad \tilde{X} = \begin{bmatrix} \tilde{x}'_{11} & \tilde{x}_{12} & \cdot & \cdot & \cdot & \tilde{x}_{1p} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdot & \cdot & \cdot & \tilde{x}_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \tilde{x}_{n1} & \tilde{x}_{n2} & \cdot & \cdot & \cdot & \tilde{x}_{np} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

$\tilde{Y}$  is  $n \times 1$  column vector,  $\tilde{X}$  is  $n \times p$  matrix,  $\beta$  is  $p \times 1$  column vector and  $\varepsilon$  is  $n \times 1$  column vector of error terms.

In order to study the special nature of the matrix  $\tilde{X}\tilde{X}$  and the OLS normal equation when the variables had been transformed, there is need to decompose the correlation matrix containing all the pairwise correlation coefficients among the response  $Y$  and predictor  $X_1, X_2, \dots, X_p$  into two matrices.

The first matrix denoted by  $r_{\tilde{X}\tilde{X}}$  is the correlation matrix of the  $X$  variables. The elements of  $r_{\tilde{X}\tilde{X}}$  are the coefficients of simple correlation between all pairs of the  $X$  – variables.

Thus,

$$r_{\tilde{X}\tilde{X}} = \begin{pmatrix} 1 & r_{12} & \cdot & \cdot & \cdot & r_{1p} \\ r_{21} & 1 & \cdot & \cdot & \cdot & r_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{p1} & r_{p2} & \cdot & \cdot & \cdot & 1 \end{pmatrix} \quad (3.27)$$

Here  $r_{ij}$  is the coefficient of correlation between  $X_i$  and  $X_j$ . The correlation matrix  $r_{\tilde{X}\tilde{X}}$  is symmetric.

The second matrix,  $r_{\tilde{Y}\tilde{X}}$ , is a vector containing the coefficients of simple correlation between the response variable  $Y$  and each of the  $X$  variables, denoted by

$$r_{\tilde{Y}\tilde{X}} = \begin{pmatrix} r_{y1} \\ r_{y2} \\ \cdot \\ \cdot \\ \cdot \\ r_{yp} \end{pmatrix} \quad (3.28)$$

The transformed variables in the correlation matrix  $\tilde{X}\tilde{X}$  are obtained as follows:

(i) in the upper left corner of the correlation matrix  $\tilde{X}\tilde{X}$

$$\tilde{X}_{ii}^2 = \frac{\sum_{j=1}^n (\tilde{X}_{ij} - \bar{X}_i)^2}{\sqrt{\sum_{j=1}^n (\tilde{X}_{ij} - \bar{X}_i)^2}} = 1 \quad \text{for } i=1, 2, \dots, p \quad (3.29)$$

(ii) in the first row, for the second column of the matrix  $\tilde{X}\tilde{X}$ , we have

$$\tilde{X}_{11}\tilde{X}_{12} = \frac{\sum_{j=1}^n (\tilde{X}_{11} - \bar{X}_1)(\tilde{X}_{12} - \bar{X}_2)}{\sqrt{\sum_{j=1}^n (\tilde{X}_{11} - \bar{X}_1)^2 \sum_{j=1}^n (\tilde{X}_{12} - \bar{X}_2)^2}} = r_{12} \quad (3.30)$$

$r_{12}$  is the coefficient of correlation between  $X_1$  and  $X_2$

Consequently

$$\tilde{X}_{ij}\tilde{X}_{jk} = \frac{\sum_{j=1}^n (\tilde{X}_{ij} - \bar{X}_j)(\tilde{X}_{jk} - \bar{X}_k)}{\sqrt{\sum_{j=1}^n (\tilde{X}_{ij} - \bar{X}_j)^2 \sum_{k=1}^n (\tilde{X}_{jk} - \bar{X}_k)^2}} = r_{jk} \quad \text{for } j \neq k \quad (3.31)$$

The ridge regression coefficients in correlation form is given as:

$$\begin{aligned} \beta(k) &= (\tilde{X}\tilde{X} + kI)^{-1} \tilde{X}\tilde{Y} \\ (\tilde{X}\tilde{X} + kI)\beta(k) &= \tilde{X}\tilde{Y} \end{aligned} \quad (3.32)$$

and is presented in form of matrices of correlation coefficients as follows:

$$\begin{bmatrix} 1 & r_{12} & \cdot & \cdot & \cdot & r_{1p} \\ r_{21} & 1 & \cdot & \cdot & \cdot & r_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{p1} & r_{p2} & \cdot & \cdot & \cdot & 1 \end{bmatrix} + \begin{bmatrix} k & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & k & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & k \end{bmatrix} = \begin{bmatrix} (1+k) & r_{12} & \cdot & \cdot & \cdot & r_{1p} \\ r_{21} & (1+k) & \cdot & \cdot & \cdot & r_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{p1} & r_{p2} & \cdot & \cdot & \cdot & (1+k) \end{bmatrix}$$

$$\begin{bmatrix} (1+k) & r_{12} & \cdot & \cdot & \cdot & r_{1p} \\ r_{21} & (1+k) & \cdot & \cdot & \cdot & r_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{p1} & r_{p2} & \cdot & \cdot & \cdot & (1+k) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix} = \begin{bmatrix} r_{y1} \\ r_{y2} \\ \cdot \\ \cdot \\ \cdot \\ r_{yp} \end{bmatrix} \quad (3.33)$$

The equations for the ridge regression coefficients based on equation (3.33) is as follows

$$\begin{aligned}
(1+k)\beta_1 + r_{12}\beta_2 + \dots + r_{1p}\beta_p &= r_{1y} \\
r_{12}\beta_1 + (1+k)\beta_2 + \dots + r_{2p}\beta_p &= r_{2y} \\
\text{.....} & \\
\text{.....} & \\
r_{p1}\beta_1 + r_{p2}\beta_2 + \dots + (1+k)\beta_p &= r_{py}
\end{aligned}
\tag{3.34}$$

Where  $r_{ij}$  is the coefficient of correlation between  $i$ th and  $j$ th predictor variable and  $r_{iy}$  is the correlation between the  $i$ th predictor variable and the response variable  $\tilde{Y}$ . The solution  $(\beta_1, \dots, \beta_p)$  of Equation (3.34) is the set of estimated ridge regression coefficient. The important thing that distinguishes ridge regression and OLS is  $k$ , the biasing parameter. Thus, increase in  $k$  causes increase in bias of the estimates.

### 3.4.1 THE VARIANCE OF THE RIDGE REGRESSION PARAMETER

When the variables for the regression parameter are centred and scaled then

$$E(\beta(k) - \beta)' (\beta(k) - \beta) = E(\beta(k)' \beta(k)) - (\beta' \beta)
\tag{3.35}$$

But

$$\begin{aligned}
\hat{\beta}(k) &= (\tilde{X}'\tilde{X} + kI)^{-1} \tilde{X}'\tilde{Y} \\
&= W\tilde{X}'\tilde{Y},
\end{aligned}
\tag{3.36}$$

where

$$W = (\tilde{X}'\tilde{X} + kI)^{-1}$$

Since  $\tilde{X}'\tilde{Y} = (\tilde{X}'\tilde{X})\hat{\beta}$  from Equation (3.9) and substituting it in equation (3.36) we have

$$\begin{aligned}
\hat{\beta}(k) &= (\tilde{X}'\tilde{X} + kI)^{-1} (\tilde{X}'\tilde{X})\hat{\beta} \\
(\tilde{X}'\tilde{X} + kI)\hat{\beta}(k) &= (\tilde{X}'\tilde{X} + kI)(\tilde{X}'\tilde{X} + kI)^{-1} (\tilde{X}'\tilde{X})\hat{\beta}
\end{aligned}$$

$$\begin{aligned}
(\tilde{X}\tilde{X} + kI)\hat{\beta}(k) &= I(\tilde{X}\tilde{X})\hat{\beta} \\
(\tilde{X}\tilde{X})^{-1}(\tilde{X}\tilde{X} + kI)\hat{\beta}(k) &= \hat{\beta} \\
(I_p + k(\tilde{X}\tilde{X})^{-1})\hat{\beta}(k) &= \hat{\beta} \\
\hat{\beta}(k) &= (I_p + k(\tilde{X}\tilde{X})^{-1})^{-1} \hat{\beta} \\
&= Z\hat{\beta} \tag{3.37}
\end{aligned}$$

where  $Z = (I_p + k(\tilde{X}\tilde{X})^{-1})^{-1}$

The variance of the ridge regression parameter is

$$\begin{aligned}
\text{Var}(\hat{\beta}(k)) &= Z(\tilde{X}\tilde{X})^{-1} \tilde{X}' \text{Var}(Y) (\tilde{X}\tilde{X})^{-1} Z' \\
&= \sigma^2 Z(\tilde{X}\tilde{X})^{-1} Z' \tag{3.38}
\end{aligned}$$

Let  $\xi(W) = \frac{1}{\lambda_j + k}$  be the eigenvalues of W, and its trace =  $\sum_{j=1}^p \frac{1}{(\lambda_j + k)}$  the sum of

the diagonal elements and  $\xi(Z) = \frac{\lambda_j}{\lambda_j + k}$  be the eigenvalues of Z and its trace

$\sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)}$ , using spectral decomposition of matrices where  $\lambda_j$  is the eigenvalues

of the matrix  $(\tilde{X}\tilde{X})$ , Equation (3.35) becomes

$$E(\hat{\beta}(k) - \beta)'(\hat{\beta}(k) - \beta) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} \tag{3.39}$$

This is a decreasing function of  $k$  and it shows the effect of ridge parameter on the total variance of the regression coefficients. For OLS where  $k = 0$

$$\text{Var}(\beta(0)) = \sigma^2 \sum_{j=1}^p \left( \frac{1}{\lambda_j} \right) \tag{3.40}$$

### 3.4.2 THE TOTAL MEAN SQUARE ERROR OF THE RIDGE REGRESSION PARAMATER

The MSE of the biased estimator is obtained as follows:

$$\begin{aligned}\hat{\beta}(k) &= (\tilde{X}\tilde{X} + kI)^{-1} \tilde{X}\tilde{Y} \\ &= (\tilde{X}\tilde{X} + kI)^{-1} \tilde{X}\tilde{X}(\tilde{X}\tilde{X})^{-1} \tilde{X}\tilde{Y}\end{aligned}\quad (3.41)$$

Substituting  $\hat{\beta} = (\tilde{X}\tilde{X})^{-1} \tilde{X}\tilde{Y}$  in Equation (3.41), we have

$$\begin{aligned}\hat{\beta}(k) &= \left[ I + k(\tilde{X}\tilde{X})^{-1} \right]^{-1} \tilde{X}\tilde{X}\hat{\beta} \\ &= Z\hat{\beta}\end{aligned}\quad (3.42)$$

Substituting Equation (3.42) in Equation (3.35), we have

$$\begin{aligned}MSE(\beta(k)) &= E\left[ (\hat{\beta}(k) - \beta)' (\hat{\beta}(k) - \beta) \right] \\ &= E\left[ (Z\hat{\beta} - \beta)' (Z\beta - \beta) \right] \\ &= E\left[ (Z\hat{\beta} - Z\beta)' (Z\hat{\beta} - Z\beta) \right] + E\left[ \hat{\beta}'Z'Z\beta - \beta'Z\hat{\beta} - \hat{\beta}'Z'\beta + \hat{\beta}\beta \right] \\ &= E\left[ (Z\hat{\beta} - Z\beta)' (Z\hat{\beta} - Z\beta) \right] + \left[ \hat{\beta}'Z'Z\beta - 2\beta'Z\beta + \beta'\beta \right] \\ &= E\left[ (\hat{\beta} - \beta)' Z'Z(\hat{\beta} - \beta) \right] + (Z\beta - \beta)' (Z\beta - \beta)\end{aligned}\quad (3.43)$$

Consider the expected value of a quadratic form  $x'Ax$ . Where A is a symmetric matrix. Here  $E(x) = 0$  since  $\hat{\beta}$  is assumed to be unbiased. Then  $E(x'Ax) = \text{trace}(AV)$  where  $V = \text{var}(x)$ . Thus the sum of the variances of all  $\hat{\beta}_j(k)$  is the sum of the diagonal elements of Equation (3.38) which is the trace of that matrix. Hence evaluating Equation (3.43) gives

$$\begin{aligned}MSE(\beta(k)) &= \sigma^2 \text{tr}\left[ (\tilde{X}\tilde{X})^{-1} Z'Z \right] + \beta'(Z - I)' (Z - I)\beta \\ &= \sigma^2 \left[ \text{Trace}(\tilde{X}\tilde{X} + kI)^{-1} - k \text{Trace}(\tilde{X}\tilde{X} + kI)^{-2} \right] + \beta'(Z - I)' (Z - I)\beta \\ &= \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + \beta'(Z - I)' (Z - I)\beta\end{aligned}\quad (3.49)$$

For two square matrices  $A$  and  $B$ ,  $B(I + AB)^{-1} = I - A(B^{-1} + A)^{-1}$  (Rao and Toutenberg 1995). Now, let  $A = kI$  and  $B = (\tilde{X}\tilde{X})^{-1}$ , we obtain  $[I + k(\tilde{X}\tilde{X})^{-1}]^{-1} = I - k(\tilde{X}\tilde{X} + kI)^{-1}$ . Since  $[I + k(\tilde{X}\tilde{X} + kI)^{-1}][\tilde{X}\tilde{X} + kI]^{-1} = (\tilde{X}\tilde{X} + kI)^{-1} - k(\tilde{X}\tilde{X} + kI)^{-2}$ , we then obtain the trace of the difference of these two matrices, and the trace of a difference of two matrices is the difference of the two traces (Liptchuz and Lipson 2009). Since the trace of a matrix is sum of the eigenvalues,  $\lambda_j$ , the trace of an inverse of a matrix is the sum of  $\lambda_j^{-1}$ , but the eigenvalues of  $(\tilde{X}\tilde{X} + kI)$  are  $\lambda_j + k$ , thus, the trace of the first term is evaluated as:

$$tr\left(\left(I + k(\tilde{X}\tilde{X})^{-1}\right)^{-1}(\tilde{X}\tilde{X} + kI)^{-1}\right) = \sum_{j=1}^p \frac{1}{(\lambda_j + k)} - k \sum_{j=1}^p \frac{1}{(\lambda_j + k)^2} = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} \quad (3.50)$$

Simplifying the second term of Equation (3.49) by using the equivalent expression for  $Z$ , we have

$$\beta'(Z - I)'(Z - I)\beta = k^2 \beta'(\tilde{X}\tilde{X} + kI)^{-2} \beta \quad (3.51)$$

Since  $X'X$  is a positive definite symmetric matrix, there exists an orthogonal matrix  $D$  such that  $D'CD = \Lambda$ , where  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  and  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  are the eigenvalues of  $C = \tilde{X}\tilde{X}$ . Let  $\alpha = D'\beta$ , then the orthogonal version of the multiple regression of Equation (3.26) becomes

$$Y = X^* \alpha + \varepsilon \quad (3.52)$$

Where  $X^* = \tilde{X}D$ , and  $\alpha = D'\beta$  since  $D'D = DD' = I$  ( $p \times p$  orthogonal matrix),  $\alpha$  is  $p \times 1$  vector and  $\beta$  is also a  $p \times 1$  vector of regression coefficients. Using the second term of Equation (3.48) as in Equation (3.50) we have

$$k^2 \beta'(X'X + kI)^{-2} \beta = \sum_{j=1}^p \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2} \quad (3.53)$$

Then, the total mean square error (TMSE) of the regression coefficient  $\hat{\beta}(k)$  in the presence of multicollinearity when the biased parameter is involved using Equations (3.50) and (3.53) is:



$$\text{TMSE}(k) = E\left\{\left(\hat{\beta}(k) - \beta\right)' \left(\hat{\beta}(k) - \beta\right)\right\} = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^p \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2} \quad (3.54)$$

Where  $\alpha_j$  is the  $j$ th element of the vector  $\alpha = D'\beta$ . Instead of choosing only a single value of  $k$  we consider separate ridge parameters for each of the regression coefficient making the ridge regression parameter a vector denoted by  $k$ , therefore

$$\text{TMSE}(k_i) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_i)^2} + \sum_{j=1}^p \frac{k_i^2 \alpha_j^2}{(\lambda_j + k_i)^2} \quad (3.55)$$

where  $k = (k_1, \dots, k_p)$ ,  $j = 1, 2, \dots, p$

As  $k$  continues to increase without bound, all the regression estimates tend towards zero. This is because the ridge method tends to shrink the estimates of the regression coefficients towards zero which is sometimes generically referred to as shrinkage estimator (Chatterjee and Hadi, 2006). The idea of ridge regression is to pick value  $k$  for which the reduction in the variance is not exceeded by the increase in bias.

We now attempt to select the value of  $k$  and to obtain the corresponding values of regression coefficients. If multicollinearity occurs, the ridge regression estimators vary as  $k$  increases from zero until stability is achieved. However, the value of  $k$  is directly related to the amount of bias introduced, hence the smallest value of  $k$  for which the stability occurs is selected.

### 3.5 METHODS OF ESTIMATING THE RIDGE PARAMETER $k$

Several authors have suggested several methods of choosing the value of  $k$  and some of them are described below

#### 3.5.1 THE RIDGE TRACE

This is the graph of the values of the parameter estimate  $\hat{\beta}_j(k)$  against the values of  $k$  for the range between 0 and 1. The general idea is to use the values

of  $k$  at which the parameter estimate tend to stabilize. The behaviour of  $\hat{\beta}_j(k)$  as a function of  $k$  is easily observed from the ridge trace. The value of  $k$  selected is the smallest value for which all the coefficients  $\hat{\beta}_j(k)$  are stable. (Hoerl and Kennard, 1970a, Chatterjee and Hadi 2006, Ryan 1997, Myers 1990 and Birkes and Dodge 1993). In addition, at the selected value of  $k$ , the residual sum of square should remain close to its minimum value and the variance inflation factor  $Vif_{(j)}$  (the  $j$ th factor of the diagonal element of the matrix  $(\tilde{X}\tilde{X} + kI)^{-1} \tilde{X}\tilde{X}(\tilde{X}\tilde{X} + kI)^{-1}$ ) should be less than 10. (see Figure 3.1)

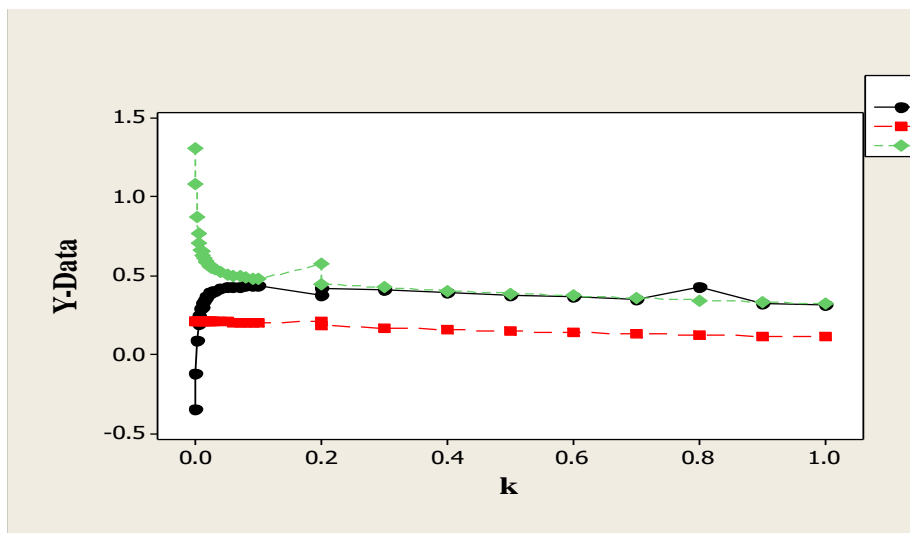


Figure3.1. RIDGE TRACE(see Chatterjee and Hadi 2006)

10

### 3.5.2 METHOD OF HOERL AND KENNARD (1970a)

Hoerl and Kennard (1970a) showed that the value of  $k$  that will minimize equation (3.54) the total mean square error (MSE) of the ridge regression coefficient is

$$k_j = \frac{\sigma^2}{\alpha_j^2} \quad j = 1, 2, \dots, p \quad (3.56)$$

Where  $\sigma^2$  represents the error variance of the model in Equation (3.25),  $\alpha_j$  is the  $j$ th element of  $\alpha$ . However, the optimal value of  $k_j$  depends on the unknown  $\sigma^2$  and  $\alpha_j^2$ , and they must be obtained from the observed data. They further suggested to replace  $\sigma^2$  and  $\alpha_j^2$  by their corresponding unbiased estimators.

That is

$$\hat{k}_j = \frac{\hat{\sigma}^2}{\hat{\alpha}_j^2} \quad (3.57)$$

where  $\hat{\sigma}^2 = \frac{\sum \varepsilon_i \varepsilon_i}{(n-p)}$  represents the residual mean square estimate which is unbiased for  $\sigma^2$  and  $\alpha_j$  is the  $j$ th element of  $\hat{\alpha}$  ( $p \times 1$  vector)

Hoerl and Kennard (1970b) suggested  $k$  to be a single value that can be obtained from the equation (3.57) as follows

$$\hat{k} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2} \quad (3.58)$$

where  $\hat{\alpha}_{\max}$  is the maximum value of  $\hat{\alpha}$ .

### 3.5.3 THE METHOD OF FIXED POINT

Hoerl et al (1975) suggested estimating the value of  $k$  by using the harmonic mean of  $\alpha_j^2$  of equation (3.57) as follows

$$k = \frac{p\hat{\sigma}^2}{\sum_{j=1}^p \hat{\alpha}_j^2} \quad (3.59)$$

### 3.5.4 THE METHOD OF LAWLESS and WANG (1976)

Lawless and Wang (1976) adopts Bayesian approach and assumes that  $\hat{\beta}(k)$  has a prior distribution that is multivariate normal with mean 0 and variance-covariance matrix  $\sigma_\beta^2 I_p$ , that is  $\hat{\beta}(k) \sim N_p(0, \sigma_\beta^2 I_p)$ . The Baye's estimator for  $\beta(k)$  is  $\hat{\beta}(k)$  where the bias of the total mean square error, which is the measure of goodness of an estimator, is as follows

$$\tilde{\alpha}_{B,j} = \frac{\lambda_j}{\lambda_j + \frac{\sigma^2}{\sigma_\alpha^2}} \hat{\alpha}_j \quad (j=1,2,\dots,p) \quad (3.60)$$

Since  $\sigma^2$  and  $\sigma_\alpha^2$  are unknown, they are to be estimated from the sample

(Note: if  $\sigma^2$  is estimated by  $\hat{\sigma}^2$  and  $\sigma_\alpha^2$  by  $\frac{1}{p} \sum_{j=1}^p \alpha_j^2$ , the method of Hoerl et al

(1975) will be obtained, but  $\frac{1}{p} \sum_{j=1}^p \alpha_j^2$  is not an ideal estimate of  $\sigma_\alpha^2$  according to

Lawless and Wang 1976)

Now in this setting, if  $\alpha \sim N_p(0, \sigma^2 I_p)$ , that is introducing the eigenvalues of the design matrix in the estimation of the ridge parameter  $k$ , then,

$$E\{\lambda_j \hat{\alpha}_{ji}^2\} = \sigma^2 + \lambda_j \sigma_\alpha^2 \quad (3.61)$$

And

$$E\left\{\sum_{i=1}^p \frac{\lambda_j \alpha_j}{\sigma^2}\right\} = p + \frac{\sum_{j=1}^p \lambda_j \sigma_\alpha^2}{\sigma^2} \quad (3.62)$$

Since  $\tilde{X}\tilde{X}$  is a correlation matrix, trace  $(\tilde{X}\tilde{X}) = \sum_{j=1}^p \lambda_j = p$ , unconditionally

$$E\left\{\sum_{j=1}^p \frac{\lambda_j \hat{\alpha}_j^2}{p\sigma^2}\right\} - 1 = \frac{\sigma_\alpha^2}{\sigma^2} \quad (3.63)$$

But because  $\sigma_\alpha^2$  is much larger than  $\sigma^2$ ,  $\frac{\sum_{j=1}^p \lambda_j \hat{\alpha}_j^2}{p\sigma^2}$  is used to provide a reasonable of

$\frac{\sigma_\alpha^2}{\sigma^2}$  given as

$$\frac{\sigma_\alpha^2}{\sigma^2} = \frac{1}{p\hat{\sigma}^2} \sum_{j=1}^p \lambda_j \hat{\alpha}_j^2 \quad (3.64)$$

Hence, using Equation (3.64) in (3.60) the estimator  $k$  is

$$k = \frac{\sigma_\alpha^2}{\sigma^2} = \frac{p\hat{\sigma}^2}{\sum_{j=1}^p \lambda_j \hat{\alpha}_j^2} \quad (3.65)$$

where  $\lambda_j$  is the eigenvalue of  $\tilde{X}\tilde{X}$

### 3.4.5 THE METHOD OF KHALAF AND SHUKUR (2005)

Khalaf and Shurkur (2005) modified Lawless and Wang (1975) to obtain an improved method of estimating  $k$  that gives a smaller TMSE as follows:

Assume that the model  $Y = X\beta + \varepsilon$  is taken to be in canonical form so that  $XX = C$  where  $C$  is a  $p \times p$  diagonal matrix with elements  $\lambda_j (> 0)$ . The least square estimate of the  $j$ th element of  $\beta$  (say  $\hat{\beta}_j$ ) is

$$\hat{\beta}_j = \frac{X_j'Y}{\lambda_j} \quad j = 1, 2, \dots, p \quad (3.66)$$

where  $X_j$  is the  $j$ th column vector of  $X$ . The version of the generalized ridge regression estimator suggested by Hoerl and Kennard (1970 a, b) is written as

$$\tilde{\beta}_j = \left( \frac{\lambda_j}{\lambda_j + \hat{k}} \right) \hat{\beta}_j \quad (3.67)$$

where  $\hat{\beta}_j = T_p^{-1}X_jY$  and  $T_p = XX$ . Hoerl and Kennard (1970) showed that a sufficient condition that  $\tilde{\beta}_j$  will have smaller MSE than  $\hat{\beta}_j$  is that

$$0 < \hat{k} < k^* \quad (3.68)$$

where  $k^* = \frac{\hat{\sigma}^2}{\alpha_{\max}^2}$ ,  $\alpha_{\max}^2$  is the largest value of  $\alpha = D\beta$ , (see Equation (3.51)).

Here  $\hat{\sigma}^2$  which is the usual estimate of  $\sigma^2$  defined by

$$\hat{\sigma}^2 = \frac{\left[ (Y - X\hat{\beta})'(Y - X\hat{\beta}) \right]}{n - p - 1}$$

With reference to Equation (3.68) as Hoerl and Kennard estimator, and, instead of estimating  $\sigma^2$  and  $\alpha_j^2$  separately, the new estimator that estimates the optimal shrinkage parameter  $k^*$  as a single quantity is

$$k^* = \frac{\lambda_{\max} \hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2} \quad (3.69)$$

where  $\lambda_{\max}$  is the maximum eigenvalues of the matrix  $\tilde{X}\tilde{X}$  and  $0 < k^* < \infty$ .

The idea was accomplished by adding the amount  $\frac{\hat{\sigma}^2}{\lambda_{\max}}$  to the denominator of

Hoerl and Kennard (1970a) which is a function of the correlation between the independent variables. This amount varies with the size of the sample used ( the

number of observation) and also to keep this kind of variation fixed,  $\frac{\hat{\sigma}^2}{\lambda_{\max}}$  is

multiplied by the number of degrees of freedom (n-p-1) to separate out the variation that only depends on the strength of the multicollinearity.

Muniz et al (2010) modified Hoerl and Kennard (1970a) and Kalaf and Shurkur (2005) to obtain the following estimators:

$$(a) \hat{K}_1 = \max\left(\frac{1}{q_j}\right) \quad (3.70)$$

$$(b) \hat{K}_2 = \max(q_j) \quad (3.71)$$

$$(c) \hat{K}_3 = \left(\prod_{j=1}^p \frac{1}{q_j}\right)^{\frac{1}{p}} \quad (3.72)$$

$$(d) \hat{K}_4 = \left(\prod_{j=1}^p q_j\right)^{\frac{1}{p}} \quad (3.73)$$

$$(e) \hat{K}_5 = \text{median}\left(\frac{1}{q_j}\right) \quad (3.74)$$

$$\text{where } q_j = \frac{\lambda_{\max} \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_{\max} \alpha_j^2}, \quad j = 1, 2, \dots, p$$

### 3.6 THE PROPOSED METHODS OF ESTIMATING THE RIDGE PARAMETER $k$

Lawless and Wang (1976) introduced the eigenvalues of  $\tilde{X}\tilde{X}$  in the estimation of  $k$  as proposed by Hoerl, et al (1975) and obtained an improved TMSE. Also

Khalaf and Shurkur (2005) combined the methods of Lawless and Wang (1976) and Hoerl and Kennard (1970a) and obtained an improved estimate of  $k$  which

produces smaller TMSE. In their own setting, they used the maximum

eigenvalue of  $\tilde{X}\tilde{X}$  without taking into consideration whether the eigenvalues are skewed either to the left or to the right, or if the maximum eigenvalue is an outlier among the eigenvalues.

### 3.6.1 WHEN THE EIGENVALUES OF THE CORRELATION MATRIX $\tilde{X}\tilde{X}$ ARE SKEWED

Skewness is the degree of asymmetry or departure from symmetry of a distribution or a set of data. (Spiegel and Stephen 2008). For a skewed set of data (eigenvalues)  $\lambda_j$ , the median tends to lie on the same side of the longer tail.

The skewness of a set of data  $\lambda_1, \lambda_2, \dots, \lambda_n$  can be estimated as follows:

$$J = \frac{\sum_{j=1}^p (\lambda_j - \bar{\lambda})^3}{ps^3} \quad (3.75)$$

where  $\bar{\lambda}$  is the mean of the eigenvalues,  $s$  is the standard deviation of the eigenvalues and  $p$  is the number of explanatory variables. Positive value of skewness indicates that the data are skewed right; this means that the right tail is long relative to the left tail. Negative value of skewness indicates that the data are skewed left; this means that the left tail is long relative to the right tail (Doane and Seward, 2011).

Considering the situation where the eigenvalues are skewed positively and the maximum  $\lambda_{\max}$  is at the longer tail of the distribution, then  $\lambda_{\max}$  does not represent the distribution of the eigenvalues. In view of this, the centre of the distribution is the median which gives a true representation of the eigenvalue (Szekely and Mori 2001, Von Hippel 2005).

Now, assuming that the model in Equation (3.26) is observed to be in canonical form

$$Y = X^* \alpha + \varepsilon \quad (3.76)$$

where  $X^* = XD$ , and  $\alpha = D'\beta \tilde{X}\tilde{X}$  is positive definite symmetric since  $D'D = DD' = I$  ( $p \times p$  orthogonal matrix),  $\alpha$  is  $p \times 1$  vector and  $\beta$  is also a  $p \times 1$  vector of regression coefficients and  $D'\tilde{X}\tilde{X}D = \text{diag}(\lambda_1, \dots, \lambda_p)$  where  $(\lambda_1, \dots, \lambda_p)$  are the eigenvalues of  $\tilde{X}\tilde{X}$ .

$D$  is the  $p \times p$  matrix of the eigenvectors corresponding to the eigenvalues of  $\tilde{X}\tilde{X}$ .

In a situation where the explanatory variables are fixed, then the MSE is minimized when  $\beta$  is normalized eigenvectors corresponding to the largest eigenvalues of  $\tilde{X}\tilde{X}$  subject to the constraint that  $\beta'\beta = 1$  (Newhouse and Oman 1971), and the columns of the eigenvectors are orthogonal (Ayres, 1962).

The performance of the maximum and minimum eigenvalues in the estimation of the ridge parameter do not differ as attested by Kibria (2003) and Pasha and Sha (2004). Consequently, Khalaf and Shurkur (2005) used the maximum eigenvalue to estimate the ridge parameter  $k$  and obtain a smaller MSE. According to Newhouse and Oman (1971), using the smallest eigenvalue will equally result to smaller MSE. Therefore in a case of skewed eigenvalues, their performance may not be the same. Thus, in order to overcome this, the median of the skewed set of eigenvalues is introduced in Equation (3.69) to obtain the weights:

$$w_j = \frac{\hat{\lambda}_{med} \hat{\sigma}^2}{(n-p) + \hat{\lambda}_{med} \alpha_j^2} \quad j = 1, 2, \dots, p \quad (3.77)$$

A set of data with a skewness value of 2 is exponentially distributed (Stuart and Ord, 1994). Taking the eigenvalues as a set of data, with coefficient of skewness 2, then the set of eigenvalues is exponentially distributed. Evidently, according to Ross (2009), the median of a continuous random variable  $X$ , with cumulative distribution function  $F(x)$  is the value  $m$  such that

$$P\{X > m\} = P\{X < m\} = \frac{1}{2} \quad (3.78)$$



If  $X$  is exponential with parameter  $\lambda$ . Then  $X$  has a density function

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases} \quad (3.79)$$

This is a steadily decreasing function and the median is as in Equation (3.78).

Then the median is the cumulative density function  $F(x)$  at  $m = \frac{1}{2}$  obtained as:

$$\begin{aligned} P(x < m) &= \int_0^m \lambda e^{-\lambda x} dx = \frac{1}{2} \\ &= -e^{-\lambda x} \Big|_0^m = \frac{1}{2} \\ &= 1 - e^{-\lambda m} = \frac{1}{2} \\ e^{-\lambda m} &= -\frac{1}{2} \end{aligned} \quad (3.80)$$

Obtaining the natural log of (3.80) and solving for  $m$  we have

$$\begin{aligned} -\lambda m &= \ln 2 \\ m &= \frac{\ln 2}{\lambda} \end{aligned}$$

This is a situation where the parameter of the exponential distribution is  $\lambda$ . In a case where the parameter is  $\frac{1}{\lambda}$  the median is

$$m = \lambda \ln 2$$

Since  $\lambda$  is the mean of the exponential distribution and for eigenvalues of the matrix  $\tilde{X}\tilde{X}$  the mean is 1. Thus,

$$m = \hat{\lambda}_{med} = \ln 2 \quad (3.81)$$

Therefore, substituting  $\hat{\lambda}_{med} = \ln 2$  in Equation (3.77) to obtain

$$w_j = \frac{\ln 2(\hat{\sigma}^2)}{(n-p)\hat{\sigma}^2 + \ln 2(\hat{\alpha}_j^2)} \quad j = 1, 2, \dots, p \quad (3.82)$$

In the light of Muniz et al (2010), we propose the following estimators for the ridge parameter  $k$

$$(a) \hat{K}_1^* = \max \left( \frac{1}{w_j} \right) \quad (3.83)$$

$$(b) \hat{K}_2^* = \max(w_j) \quad (3.84)$$

$$(c) \hat{K}_3^* = \left( \prod_{j=1}^p \frac{1}{w_j} \right)^{\frac{1}{p}} \quad (3.85)$$

$$(d) \hat{K}_4^* = \left( \prod_{j=1}^p w_j \right)^{\frac{1}{p}} \quad (3.86)$$

$$(e) \hat{K}_5^* = \text{median} \left( \frac{1}{w_j} \right) \quad (3.87)$$

$$(f) \hat{K}_6^* = \text{median}(w_j) \quad (3.88)$$

### 3.6.2 WHEN THERE ARE OUTLIERS AMONG THE EIGENVALUES OF THE CORRELATION MATRIX $\tilde{X}\tilde{X}$

We start with the singular value decomposition of  $X = UDV'$ , where U and V are orthonormal matrices and D is a diagonal matrix containing singular values in descending order of magnitude. We present k as  $k = XX' = U\Lambda U'$  where  $\Lambda = D^2$ . Alternatively, we construct a matrix  $k^* = XX = V\Lambda V'$  with the same eigenvalue as k. by the Courant-Fischer Min-Max Theorem (McDiarmid, 1989) and the definition of a projection operator, the first and or the (largest) eigenvalue of k is given by

$$\begin{aligned} \lambda_1 &= \max_{0 \neq \alpha \in R^n} \frac{\alpha' k^* \alpha}{\alpha' \alpha} \\ &= \max_{0 \neq \alpha \in R^n} \frac{\alpha' X X \alpha}{\alpha' \alpha} \end{aligned}$$

$$\begin{aligned}
&= \max_{0 \neq \alpha \in R^n} \frac{\|X\alpha\|^2}{\alpha'\alpha} = \max_{0 \neq \alpha \in R^n} \sum_{i=1}^n \|P_\alpha(x_i)\|^2 \\
&= \sum_{j=1}^n \|x_j\|^2 - \min_{0 \neq \alpha \in R^n} \sum_{i=1}^n \|\bar{P}_\alpha(x_i)\|^2
\end{aligned} \tag{3.89}$$

Where  $P_\alpha(x)$  is the projection of  $x$  onto the space spanned by  $\alpha$ , and  $\bar{P}_\alpha(x)$  is the projection of  $x$  onto the space perpendicular to  $\alpha$  (Izenman and Shen, 2008).

Equation (3.89) suggests that the first eigenvector can be characterized as the direction for which the residual sum of square is minimized. Applying the same line of reasoning to the general form of the courant-fischer Min-Max theorem, the  $m$ th eigenvalue of  $k$  can be expressed as

$$\lambda_m(k) = \max_{\dim T=m} \min_{0 \neq \alpha \in T} \sum_{i=1}^n \|P_\alpha(x_i)\|^2 \tag{3.90}$$

Which implies that if  $\alpha^m$  is the  $m$ th eigenvector of  $k^*$ , then

$$\lambda_m(k) = \sum_{i=1}^n \|P_{\alpha^m}(x_i)\|^2 \tag{3.91}$$

Consequently, if  $T_m$  is the space spanned by the first  $m$  eigenvectors, then

$$\sum_{j=1}^m \lambda_j(k) = \sum_{i=1}^n \|P_{T_m}(x_i)\|^2 = \sum_{i=1}^n \|x_i\|^2 - \sum_{i=1}^n \|\bar{P}_{T_m}(x_i)\|^2 \tag{3.92}$$

By induction over the dimension of  $T$ , it readily follows that we can characterize the sum of the first  $m$  and the sum of the last  $(n-m)$  eigenvalues by,

$$\sum_{j=1}^m \lambda_j(k) = \max_{\dim(T)=m} \sum_{i=1}^n \|P_T(x_i)\|^2 = \sum_{i=1}^n \|x_i\|^2 - \min_{\dim(T)=m} \sum_{i=1}^n \|\bar{P}_T(x_i)\|^2 \tag{3.93}$$

And

$$\sum_{j=m+1}^n \lambda_j(k) = \sum_{i=1}^n \|x_i\|^2 - \sum_{j=1}^m \lambda_j(k) = \min_{\dim(T)=m} \sum_{i=1}^n \|\bar{P}_T(x_i)\|^2 \tag{3.94}$$

respectively (Izenman and Shen, 2008). Hence, when  $m=n-1$ , it implies that the subspace spanned by the last eigenvector is characterized as that for which the residual sum of square is minimized.

Outliers among the eigenvalues are extreme eigenvalues which can create difficulties and it can be detected using Dixon's Q-test (Deon and Dixon 1951).

In statistics, Dixon's Q-test or simply Q-test is used for identification and rejection of outliers. To apply Q-test in "bad data", arrange the data in order of increasing values and obtain Q as follows

$$Q = \frac{gap}{range} \quad (3.95)$$

Where  $gap$  in absolute value is the difference between the outlier in question and the closest value to it and  $range$  is the difference between the maximum observation and the minimum observation. If  $Q > Q_T$  where  $Q_T$  is the critical value corresponding to the sample size and confidence level, then reject the questionable point. (Rorabacher 1991)

Once the maximum eigenvalue is an outlier, following Newhouse and Oman (1971), it will not give the same result as the minimum eigenvalue, and then the geometric mean of the eigenvalues is used as follows

$$\bar{\lambda}_g = \left( \prod_{j=1}^p \lambda_j \right)^{\frac{1}{p}} \quad (3.96)$$

Substituting (3.96) in Khalaf and Shurkur (2005) we have

$$v_j = \frac{\bar{\lambda}_g (\hat{\sigma}^2)}{(n-p)\hat{\sigma}^2 + \bar{\lambda}_g \hat{\alpha}_j^2}, \quad j=1,2,\dots,p \quad (3.97)$$

Substituting  $v_j$  in Muniz et al (2010) the following estimators were obtained

$$(f) \hat{k}_7^* = \max \left( \frac{1}{v_j} \right) \quad (3.98)$$

$$(g) \hat{k}_8^* = \max(v_j) \quad (3.99)$$

$$(h) \hat{k}_9^* = \left( \prod_{j=1}^p \frac{1}{v_j} \right)^{\frac{1}{p}} \quad (3.100)$$

$$(i) \hat{k}_{10}^* = \left( \prod_{j=1}^p v_j \right)^{\frac{1}{p}} \quad (3.101)$$

$$(j) \hat{k}_{11}^* = \text{median} \left( \frac{1}{v_j} \right) \quad (3.102)$$

$$(k) \hat{k}_{12}^* = \text{median}(v_j) \quad (3.103)$$

## **CHAPTER FOUR**

### **DATA ANALYSIS AND DISCUSSIONS**

#### **4.0 INTRODUCTION**

In this chapter, we present the application of the proposed method of ridge regression estimation to a real life problem. Also a Monte Carlo simulation will be used to evaluate the performance of the OLS, the proposed estimators and other existing ridge regression estimators, using computer programme in R.

#### **4.1 MONTE CARLO SIMULATION**

Here we give a brief description of the factors that vary in our simulation study.

##### **4.1.1 The Number of Explanatory Variables ( $p$ )**

In most simulation studies, (Alkhamisi and Shukur (2008), Muniz and Kibria (2009), and Muniz et al, (2010)), the proposed ridge estimators is calculated using a fairly low number of explanatory variables (two and four are the most commonly values of  $p$ ). Hence, there is a need to investigate the validity of the proposed estimators with more variables to see which ridge estimator is most preferred with respect to MSE and PRESS respectively. Hence the number of explanatory variables  $p = 3,4,5,6,7,8,9$  shall be considered in the simulation study.

##### **4.1.2 The Sample Size ( $n$ )**

Increase in the sample size  $n$  has a positive effect both on MSE and PRESS. Since the number of explanatory variables is quite high, we shall fix the number of degrees of freedom instead of the number of observations. Otherwise many combinations of different  $n$  and  $p$  will not be possible to estimate the PRESS since the degrees of freedom will exceed the number of explanatory variables. In this work we shall consider  $n = 10, 30, 50$  and  $80$  respectively.

### 4.1.3 The Strength of Correlation among the Explanatory Variables ( $\rho$ )

The factor that affects the properties of different estimation methods is the degree of correlation between the explanatory variables. So we shall use  $\rho = 0.30, 0.50, 0.70, 0.90$  and  $0.99$  as the strength of correlation in the following data generating method for the explanatory variables:

$$x_{ij} = (1 - \rho^2)^{1/2} z_{ij} + \rho z_{ij} \quad (4.1)$$

where  $\rho$  represents the correlation between the explanatory variables,  $z_{ij}$  are generated using the standard normal distribution. Furthermore, the  $n$  observations for the dependent variable are generated by using the following equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip} + e_i, \quad i = 1, 2, \dots, n \quad (4.2)$$

Where  $e_i \sim N(0, \sigma^2)$  and  $\beta_0 = 0$ .  $\beta_j$  are chosen according to Newhouse and Oman (1971) proposal.

### 4.1.4 Variability of the Random Error ( $e_i$ )

In most studies (Kibria 2003, Manson et al 2010), it has been shown that increasing the variance of the error term  $e_i$  has a negative impact on the MSE and PRESS especially for the OLS estimator. In view of this, we shall use  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$  and  $1.0$ .

### 4.1.5 RIDGE ESTIMATORS CONSIDERED IN THE SIMULATION

We shall compare the following existing ridge estimators with the proposed ridge estimators

(i)  $k_0$  = method of ordinary least square estimator

(ii)  $k_1 = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2}$

(iii)  $k_2 = k = \frac{p \hat{\sigma}^2}{\sum_{j=1}^p \hat{\alpha}_j^2}$

$$(iv) k_3 = \frac{p\sigma^2}{\sum_{j=1}^p \lambda_j \hat{\alpha}_j^2}$$

$$(v) k_4 = \frac{\lambda_{\max} \hat{\sigma}}{(n-p)\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2}$$

$$\text{For } q_j = \frac{\lambda_{\max} \hat{\sigma}}{(n-p)\hat{\sigma}^2 + \lambda_{\max} \alpha_j^2}, \quad j = 1, 2, \dots, p$$

$$(vi) k_5 = \max\left(\frac{1}{q_j}\right)$$

$$(vii) k_6 = \max(q_j)$$

$$(viii) k_7 = \left(\prod_{j=1}^p \frac{1}{q_j}\right)^{\frac{1}{p}}$$

$$(ix) k_8 = \left(\prod_{j=1}^p q_j\right)^{\frac{1}{p}}$$

$$(x) K_9 = \text{median}\left(\frac{1}{q_j}\right)$$

### The Proposed Ridge Estimators:

$$(xi) K_{10} = \max\left(\frac{1}{w_j}\right)$$

$$(xii) K_{11} = \max(w_j)$$

$$(xiii) K_{12} = \left(\prod_{j=1}^p w_j\right)^{\frac{1}{p}}$$

$$(xiv) K_{13} = \left(\prod_{j=1}^p \frac{1}{w_j}\right)^{\frac{1}{p}}$$

$$(xv) K_{14} = \text{median}\left(\frac{1}{w_j}\right)$$

$$(xvi) K_{15} = \text{median}(w_j)$$

where  $w_j = \frac{\ln 2(\hat{\sigma}^2)}{(n-p)\hat{\sigma}^2 + \ln 2(\hat{\alpha}_j^2)}$  ,  $j = 1, 2, \dots, p$

$$(xvii) K_{16} = \max\left(\frac{1}{v_j}\right)$$

$$(xviii) K_{17} = \max(v_j)$$

$$(xix) K_{18} = \left(\prod_{j=1}^p v_j\right)^{\frac{1}{p}}$$

$$(xx) K_{19} = \left(\prod_{j=1}^p \frac{1}{v_j}\right)^{\frac{1}{p}}$$

$$(xxi) K_{20} = \text{median}\left(\frac{1}{v_j}\right)$$

$$(xxii) K_{21} = \text{median}(v_j)$$

where  $v_j = \frac{\bar{\lambda}_g(\hat{\sigma}^2)}{(n-p)\hat{\sigma}^2 + \hat{\lambda}_g \hat{\alpha}_j^2}$  ,  $j = 1, 2, \dots, p$

## 4.2 METHOD OF COMPARING ESTIMATORS

### 4.2.1 PREDICTED ERROR SUM OF SQUARE (PRESS) STATISTIC

Predicted error sum of square (PRESS) statistic is an interesting and very important criterion, which can be used as a form of data validation in generating prediction errors

Consider a set of data in which we withhold or set aside the first observation from the sample and we use the remaining  $(n-1)$  observations to estimate the coefficients for a particular candidate model. The first observation is then replaced and the second observation withheld with coefficients estimated again. We remove each observation one at a time, and thus the candidate fit n-times.



The deleted response is estimated each time, resulting in  $n$  predicted error or PRESS residuals  $y_i - \hat{y}_{(i)} = e_{(i)}$  ( $i = 1, \dots, n$ ). These PRESS residuals are true prediction errors with  $\hat{y}_{(i)}$  being independent of  $y_i$ . The prediction  $\hat{y}_{(i)}$  is the regression function evaluated at  $X = x_i$  but  $y_i$  was set aside and not used in obtaining the coefficients. Notionally, we have

$$\hat{y}_{(i)} = x_i' \hat{\beta}_{(i)} \quad (4.3)$$

where  $\hat{\beta}_{(i)}$  is the set of coefficients computed without the use of  $i$ th observation.

Then,

$$\hat{\beta}_{(i)} = (X_{(i)}' X_{(i)})^{-1} X_{(i)}' y_{(i)} \quad (4.4)$$

where  $X_{(i)}$  and  $y_{(i)}$  are the  $X$  and  $y$  vector with the  $i$ th observation withheld and PRESS is defined as

$$\text{PRESS} = \sum_{i=1}^n e_{(i)}^2 = \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2 \quad (4.5)$$

Thus the  $i$ th PRESS residual is as follows

$$\begin{aligned} e_{(i)} &= y_i - \hat{y}_{(i)} \\ &= y_i - x_i' \hat{\beta}_{(i)} \\ &= y_i - x_i' (X_{(i)}' X_{(i)})^{-1} X_{(i)}' y_{(i)} \end{aligned} \quad (4.6)$$

But

$$(X_{(i)}' X_{(i)})^{-1} = (X' X)^{-1} + \frac{(X' X)^{-1} x_i x_i' (X' X)^{-1}}{1 - h_{ii}} \quad (4.7)$$

where  $h_{ii} = x_i' (X' X)^{-1} x_i$  using Equation (4.7) in Equation (4.6), we have

$$e_{(i)} = y_i - x_i' \left[ (X'X)^{-1} + \frac{(X'X)^{-1} x_i x_i' (X'X)^{-1}}{1 - h_{ii}} \right] X_{(i)}' y_i \quad (4.8)$$

$$\begin{aligned} e_{(i)} &= y_i - x_i' (X'X)^{-1} X_{(i)}' y_{(i)} - \frac{x_i' (X'X)^{-1} x_i x_i' (X'X)^{-1} X_{(i)}' y_{(i)}}{1 - h_{ii}} \\ &= \frac{(1 - h_{ii}) y_i - (1 - h_{ii}) x_i' (X'X)^{-1} X_{(i)}' y_{(i)} - h_{ii} x_i' (X'X)^{-1} X_{(i)}' y_{(i)}}{1 - h_{ii}} \\ &= \frac{(1 - h_{ii}) y_i - x_i' (X'X)^{-1} X_{(i)}' y_{(i)}}{1 - h_{ii}} \end{aligned} \quad (4.9)$$

but

$$X'y = X_{(i)}' y_{(i)} + x_i' y_i$$

$$X_{(i)}' y_{(i)} = X'y - x_i' y_i \quad (4.10)$$

Substituting Equation (4.10) in Equation (4.9) we have

$$\begin{aligned} e_{(i)} &= \frac{(1 - h_{ii}) y_i - x_i' (X'X)^{-1} (X'y - x_i' y_i)}{1 - h_{ii}} \\ e_{(i)} &= \frac{(1 - h_{ii}) y_i - x_i' (X'X)^{-1} X'y + x_i' (X'X)^{-1} x_i' y_i}{1 - h_{ii}} \\ e_{(i)} &= \frac{(1 - h_{ii}) y_i - x_i' \hat{\beta} + h_{ii} y_i}{1 - h_{ii}} \\ e_{(i)} &= \frac{y_i - x_i' \hat{\beta}}{1 - h_{ii}} \end{aligned} \quad (4.11)$$

where  $y_i - x_i' \hat{\beta}$  is the ordinary residual  $e_i$  for the least square fit to all  $n - 1$  observations so that the  $i$ th PRESS residual is

$$e_{(i)} = \frac{e_e}{1-h_{ii}} \quad (4.12)$$

Since PRESS is the sum of square of the PRESS residual, then, it becomes

$$\text{PRESS} = \sum_{i=1}^n \left( \frac{e_i}{1-h_{ii}} \right)^2 \quad (4.13)$$

### **4.3 SIMULATION RESULTS OF MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE n =10**

The simulated values of MSE for n=10 (small sample size) are presented in Tables 4.1(a) – 4.1(g) for  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$ , number of explanatory variable  $p = 3, 4, 5, 6, 7, 8, \text{ and } 9$  and  $\rho = 0.3, 0.5, 0.7, 0.9, 0.99$ . The values of the MSEs of the proposed ridge estimators and some existing estimators are also presented in Figures 4.1 – 4.7 (see Appendix 2).

### **4.4 DISCUSSION OF SIMULATION RESULTS OF MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE n = 10**

In Table 4.1(a) (Appendix 2), as the value of standard deviation  $\sigma$  increases, the value of the MSE for the ridge estimators  $k_4, k_7, k_9, k_{12}$  and  $k_{21}$  show a little increase. Also in a severe multicollinearity, there is high value of MSE for all the ridge estimators especially that of the  $k_0$  (OLS). Figure 4.1 also showed that increase in the standard deviation  $\sigma$  does not affect the values of MSE for the ridge estimators  $k_1, k_2, k_8, k_{10}, k_{14}, k_{16}$  and  $k_{20}$ , but there is reduction in MSE for the estimators  $k_{14}$  and  $k_{20}$  as  $\sigma$  increases.

In Table 4.1(b), the MSE of the ridge estimators  $k_1, k_2, k_3, k_4, k_6, k_9, k_{12}$  and  $k_{21}$  increases respectively as the standard deviation  $\sigma$  increases, whereas there is a

decrease in MSE of the ridge estimators  $k_{10}$ ,  $k_{13}$ ,  $k_{14}$ ,  $k_{16}$ ,  $k_{18}$  and  $k_{20}$  as the standard deviation  $\sigma$  increases. Also, in the presence of severe collinearity  $\rho = 0.99$ , the ridge estimators  $k_{18}$  showed the smallest MSE for all the values of the standard deviation when the number of explanatory variables  $p = 4$

When the number of explanatory variables  $p = 5$ , there is an increase in MSE for the following ridge estimators –  $k_4$ ,  $k_7$ ,  $k_9$ ,  $k_{15}$ , and  $k_{21}$  when the standard deviation  $\sigma$  increases. Also, the ridge estimators;  $k_1$ ,  $k_5$ ,  $k_{16}$ ,  $k_{18}$  and  $k_{20}$  recorded smaller values of MSE in a severe multicollinearity than other estimators. (see Table 4.1(c) and Figure 4.3)

When  $p = 6$  for the values of standard deviation  $\sigma$  and correlation coefficient  $\rho$  as shown in Table 4.1(d) and Figure 4.4, there is an increase in the values of MSE respectively as the value of the standard deviation  $\sigma$  increases, but for the ridge estimators  $k_9$ ,  $k_{11}$ ,  $k_{12}$ ,  $k_{15}$ ,  $k_{17}$ ,  $k_{19}$ , and  $k_{21}$  the level of increase in value is more than that of the estimators  $k_1$ ,  $k_4$ ,  $k_5$ ,  $k_6$ ,  $k_7$ ,  $k_8$ ,  $k_9$  and  $k_{18}$ . Furthermore, when  $p = 7$ , except for  $k_0$ , all the estimators yielded small MSE. (see Table 4.1(e) and Figure 4.5)

When  $p = 8$  as presented in Table 4.1(f) and Figure 4.6, there is no pattern of movement among the MSE of the ridge estimators under consideration but they all have smaller MSE with  $k_{20}$  having the smallest value of MSE as the standard deviation  $\sigma$  increases when compared to  $k_0$ .

Table 4.1(g) and Figure 4.7 showed simulated values of MSE when  $p = 9$  at different values of standard deviation  $\sigma$  and correlation coefficient  $\rho$ . For  $k_0$  there is a very high value of MSE which continues to increase as the value of  $\sigma$  increases.

Consequently,  $n = 10$  (small sample size) showed large values of MSE when the number of explanatory variables  $p$  is large. In a severe multicollinearity, there

is increase in MSE as the value of  $\sigma$  increases. Also the ridge estimators  $k_1, k_2, k_8, k_{14}, k_{16}$  and  $k_{20}$  showed decrease in MSE as the  $\sigma$  and  $p$  increases respectively.

#### **4.5 SIMULATION RESULTS OF MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 30$**

Tables 4.2(a) – 4.2(g) summarised the simulated MSE values for  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$ , number of explanatory variable  $p = 3, 4, 5, 6, 7, 8$ , and  $9$  and  $\rho = 0.3, 0.5, 0.7, 0.9, 0.99$  for the sample size  $n = 30$ . The values of the MSEs of the ridge estimators are also presented in Figures 4.8 – 4.14 (see Appendix 3).

#### **4.3.4 DISCUSSION OF SIMULATION RESULTS OF MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 30$**

When  $p = 3$ , Table 4.2(a) and Figure 4.8 showed the simulated values of MSE for different values of standard deviation  $\sigma$  and correlation coefficient  $\rho$ . The ridge estimators  $k_1, k_3$  and  $k_2$  showed smaller MSE when  $\sigma$  is small and large MSE when  $\sigma$  is large, where as  $k_{10}, k_{14}$  and  $k_{20}$  showed smaller MSE in the presence of large standard deviation  $\sigma$ . Furthermore, the MSE of the ridge estimators  $k_0, k_9, k_{11}, k_{12}, k_{17}, k_{19}$ , and  $k_{21}$  increases as the values of  $\sigma$  increases respectively.

The simulated values of MSE when the number of explanatory variables  $p = 4$  are presented in Table 4.2(b) and the graph of the MSE is presented in Figure 4.9. As the value of the standard deviation  $\sigma$  increases, the MSE for the ridge estimators  $k_0, k_9, k_{12}, k_{19}$  and  $k_{21}$  also increases respectively. Moreover, when  $\sigma$  increases, there is a steady decrease in MSE of the ridge estimators  $k_{10}, k_{14}$  and

$k_{20}$ . But for ridge estimators  $k_1$ ,  $k_2$  and  $k_3$ , there is an increase in MSE from  $\sigma = 0.1$  to  $\sigma = 0.4$  and the maintained almost the same values at higher values of  $\sigma$ .

Table 4.2(c) and Figure 4.10 showed the simulated values of MSE when  $p = 5$ ,  $\sigma = 0.1$  to  $1.0$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$ . The MSE values showed a steady increase for  $k_0, k_9, k_{15}, k_{19}$  and  $k_{21}$  when  $\sigma$  increases. Moreover, the ridge estimators  $k_8, k_{12}, k_{18}$  and  $k_{20}$  showed smaller values of MSE all through the values of  $\sigma$ .

Furthermore, Table 4.2(d) and Figure 4.11 showed the simulated values of MSE when the explanatory variable  $p = 6$  for the values of  $\sigma$  and  $\rho$ . When  $\sigma$  increases, the MSE values of  $k_0, k_9, k_{12}$  and  $k_{21}$  increases steadily, while those of  $k_{16}, k_{18}$  and  $k_{20}$  decreases also in severe multicollinearity, the MSE of these ridge estimators  $k_{20}, k_8, k_{16}$  and  $k_{18}$  remains small.

The simulated values of MSE for the values of  $\sigma$  and  $\rho$  when the number of explanatory variables  $p = 7$  are presented in Table 4.2(e) and Figure 4.12. The values obtained showed that the ridge estimators  $k_9, k_{19}$  and  $k_{21}$  produce higher MSE values whereas those of  $k_8, k_{15}$  and  $k_{20}$  showed a decrease in the value of MSE as the values of  $\sigma$  increases respectively.

Moreover, when the explanatory variable  $p$  is 8 as presented in Table 4.2(f) and Figure 4.13, the MSE of  $k_1$  showed a sharp peak at  $\sigma = 0.6$  and at  $\sigma = 0.9$ . also,  $k_0$  (OLS) showed large MSE as the values of  $\sigma$  increases. When  $p = 9$ , as shown in Tables 4.2(g) and Figure 4.14, the pattern of movement of MSE for different ridge estimators  $k$  remained the same when compared to other values of the number of explanatory variables, but the estimators  $k_{14}$  and  $k_{20}$  still maintained smaller values of MSE when  $\sigma$  is large.

Consequently, when the sample size  $n = 30$ , there is increase in the values of MSE as the value of the standard deviation  $\sigma$  increases for some ridge estimators, whereas for some ridge estimators there is a decrease in the values of MSE as the value of the standard deviation  $\sigma$  increases. But the estimator  $k_{20}$

maintained almost the same small value of MSE for both small and large values of  $\sigma$  even in the presence of severe multicollinearity. It was also observed that  $k_0$ ,  $k_9$ ,  $k_{19}$  and  $k_{21}$  depicted sharp increase as  $\sigma$  increases.

#### **4.7 SIMULATION RESULTS OF MEAN SQUARE ERROR MSE OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 50$**

Tables 4.3(a) – 4.3(g) showed the simulated MSE values for  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$ , number of explanatory variable  $p = 3, 4, 5, 6, 7, 8$ , and  $9$  and  $\rho = 0.3, 0.5, 0.7, 0.9, 0.99$  when the sample size  $n = 50$ . The MSE values of the estimators are also plotted and presented in Figures 4.15 – 4.21 (see Appendix 4)

#### **4.8 DISCUSSION OF SIMULATION RESULTS OF MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 50$**

When the number of explanatory variables is equal to 3, the simulated MSE values are presented in Table 4.3(a) and Figure 4.15. The ridge regression estimator  $k_3$  showed smaller MSE while the estimators  $k_{12}$ ,  $k_{15}$  and  $k_{21}$  showed continuous increase in MSE as the value of  $\sigma$  increases, but  $k_{20}$  showed a decrease in MSE as  $\sigma$  increases.

Table 4.3(b) showed the simulated values of MSE for all the ridge regression estimators when the number of explanatory variable  $p$  is 4. The estimators  $k_0$ ,  $k_9$ ,  $k_{12}$ ,  $k_{15}$  and  $k_{21}$  showed a steady and sharp increase in MSE as the value of  $\sigma$  increases (see Figure 4.16)

Furthermore, for  $p=5$ , the values of MSE were presented in Table 4.3 (c) and Figure 4.17 showed that the estimators,  $k_1$ ,  $k_3$ ,  $k_{12}$  and  $k_{20}$  have smaller MSE

than other estimators, where as  $k_9$ ,  $k_{15}$  and  $k_{21}$  showed a steady increase in MSE as the standard deviation  $\sigma$  increases.

Also the simulated values of MSE for  $p=6$  showed that  $k_1$  maintained an outstanding movement on the graph as the value of  $\sigma$  increases, but  $k_2$ ,  $k_3$ ,  $k_{18}$  and  $k_{20}$  have smaller MSE than other estimators (see Figure 4.18 and Table 4.3(d) ).

The simulated values of MSE when the number of explanatory variables is 7 for the values of the standard deviation  $\sigma$  and the correlation coefficient  $\rho$  showed that increase in the value of  $\sigma$  increases the values of MSE for the following ridge regression estimators  $k_9$ ,  $k_{15}$ ,  $k_{19}$  and  $k_{21}$ . Moreover the MSE values of the following ridge estimators  $k_2$ ,  $k_8$  and  $k_{20}$  maintained a steady value as the standard deviation  $\sigma$  increases (see Table 4.3(e) and Figure 4.19).

When the number of explanatory variable  $p$  is 8, the simulated values of MSE are presented in Table 4.3(f) and Figure 4.20. Also, when  $p$  is 9, the simulated MSE values are presented in Table 4.3(g) and Figure 4.21 for all the proposed and existing ridge regression estimators under study. In Figure 4.20, the MSE value for the ridge estimator  $k_1$  shows a steady increase as the standard deviation  $\sigma$  increases where as in  $p=9$  the value of  $k_1$  increases for  $\sigma = 0.1$  to 0.5 and later moved approximately in a straight line.

In summary, when the sample size is 50, there is sharp increase in the simulated values MSE for  $k_0$ ,  $k_1$ ,  $k_9$ ,  $k_{19}$  and  $k_{21}$  as the values of  $\sigma$  changes. Consequently there is a decrease in the simulated values of MSE for  $k_8$  and  $k_{20}$  as the value of  $\sigma$  increases. The values of MSE for the ridge regression estimators  $k_1$ ,  $k_2$ ,  $k_3$ ,  $k_4$  and  $k_{18}$  increase from zero to one when  $\sigma=3$  and run in approximately straight line from 0.4 to 1.0. However,  $k_{14}$  and  $k_{20}$  maintained smaller values of MSE throughout the values of  $p$  and in severe multicollinearity.



#### **4.9 SIMULATION RESULTS OF MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 80$**

The simulated values of MSE of the proposed estimators and some existing ones for sample size  $n = 80$  were presented in Tables 4.4(a) – 4.4(g) and Figures 4.22 - 4.28. The simulation is done for various values of the variance of random error  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$ , correlation coefficient  $\rho = 0.3, 0.5, 0.7, 0.9, 0.99$  and number of explanatory variable  $p = 3, 4, 5, 6, 7, 8$ , and 9 (see Appendix 5).

#### **4.10 DISCUSSION OF SIMULATION RESULTS OF MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 80$**

The simulated values of MSE for the number of explanatory variables  $p=3$  are shown in Table 4.4(a) and Figure 4.22. Observations from Table 4.4(a) and Figure 4.22 showed that the ridge estimators,  $k_{15}$  and  $k_{21}$  increases respectively as the value of the standard deviation  $\sigma$  increases. Moreover, the values of the MSE for the ridge estimators,  $k_8$  and  $k_{20}$  decrease as  $\sigma$  increases and those of  $k_1$ ,  $k_2$  and  $k_3$  increase at a smaller rate as  $\sigma$  increases.

Table 4.4(b) and Figure 4.23 showed the simulated values of MSE for the ridge estimators under consideration when the number of explanatory variables  $p = 4$ . The observations made include smaller MSE for  $k_1$ ,  $k_2$ ,  $k_3$ ,  $k_8$  and  $k_{20}$  with  $k_3$  having the smallest MSE when compared to other ridge regression estimators in the presence of severe multicollinearity.

Table 4.4(c) and Figure 4.24 showed the simulated values of MSE for the ridge estimators when  $p = 5$  for different values of standard deviation  $\sigma$  and correlation coefficient  $\rho$ . At each value of  $\sigma$ , there is increase in MSE as the values of  $\rho$  increases from 0.3 to 0.99 among all the ridge estimators except for

$k_8$ ,  $k_{18}$  and  $k_{20}$  where the value of their MSE decreases as the correlation coefficient  $\rho$  increases.

When the number of explanatory variables  $p = 6$ , the simulated MSE for different ridge regression estimators were presented in Table 4.4(d) and Figure 4.25. The ridge estimators  $k_1$ ,  $k_6$ ,  $k_9$ ,  $k_{15}$ ,  $k_{17}$  and  $k_{21}$  showed continuous increase in the value of MSE from approximately 0.005 to 7.0 as the  $\sigma$  increases while the values of the MSE for ridge estimators  $k_7$ ,  $k_8$ ,  $k_{13}$  and  $k_{20}$  decreases on the average approximately from 0.7 to 0.4. Also the values of MSE for  $k_1$ ,  $k_2$ ,  $k_3$  and  $k_4$  increase on the average approximately from 0.05 to 5.0.

Furthermore, the simulated values of MSE when the number of explanatory variables  $p = 7$  for  $\sigma$  and  $\rho$  are presented in Table 4.4(e) and Figure 4.26 and that of  $p=8$  are presented in Tables 4.4(f) and Figure 4.27. For smaller variance of the random error  $\sigma$  some of the proposed estimators produce smaller values of MSE. Consequently, as the value of  $\sigma$ , increases, the MSE of the following ridge estimators  $k_0$ ,  $k_1$ ,  $k_4$ ,  $k_6$ ,  $k_7$ ,  $k_9$ ,  $k_{11}$ ,  $k_{12}$ ,  $k_{15}$ ,  $k_{17}$ ,  $k_{19}$  and  $k_{21}$  tend to increase whereas those of  $k_{10}$ ,  $k_{13}$ ,  $k_{14}$ ,  $k_{16}$ ,  $k_{18}$  and  $k_{20}$  tend to decrease as  $\sigma$  increases. Moreover, the MSE of these ridge estimators,  $k_1$ ,  $k_2$ , and  $k_3$  showed a little increase in the values of MSE as  $\sigma$  increases.

The simulated values of MSE for the number of explanatory variable  $p = 9$  are presented in Table 4.4(g) and Figure 4.28. As  $\sigma$  increases, the MSE values for  $k_8$ ,  $k_{10}$ ,  $k_{13}$ ,  $k_{14}$ ,  $k_{16}$ ,  $k_{18}$  and  $k_{20}$  decrease whereas the MSE of the other ridge estimators tend to increase with increase in  $\sigma$  except for the estimators  $k_1$ ,  $k_2$  and  $k_3$  which increase in a smaller quantity than others.

#### **4.11 SIMULATION RESULT OF PREDICTION SUM OF SQUARES (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 10$**

The values of PRESS for different ridge regression estimators when the sample size  $n = 10$  were presented in Tables 4.5(a) – 4.5(g) and Figures 4.29 to 4.35. These values are for the values of the correlation coefficient  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$ , standard deviation  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ , and  $1.0$  and the number of explanatory variables  $p = 3, 4, 5, 6, 7, 8$ , and  $9$  (see Appendix 6).

#### **4.12 DISCUSSION OF SIMULATION RESULTS OF PREDICTION SUM OF SQUARE (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 10$**

When the number of explanatory variables  $p = 3$  presented in Table 4.5(a) and Figure 4.29 showed an increase in PRESS of the estimators  $k_0, k_1, k_2, k_3, k_4, k_6, k_7, k_{12}, k_{19}$  and  $k_{21}$  as the standard deviation  $\sigma$  increases, whereas for the ridge estimators  $k_5, k_8, k_{10}, k_{13}, k_{14}, k_{16}, k_{18}$  and  $k_{20}$  the values of PRESS decrease. In a severe multicollinearity, there is increase in the value of PRESS for  $k_0$ , but for other estimators, as  $\rho$  increases, the PRESS value reduces for the ridge estimators  $k_{13}, k_{18}$  and  $k_{20}$  respectively.

Moreover, when the number of explanatory variables  $p$  is  $4$  and  $5$  respectively, there is increase in PRESS values for the ridge estimators  $k_1, k_2, k_3, k_4, k_{11}, k_{12}$  and  $k_{21}$  while the ridge estimators  $k_5, k_8, k_{10}, k_{18}$  and  $k_{20}$  showed decrease in PRESS as the value of  $\sigma$  increases. (see Tables 4.5(b) and (c) and Figures 4.30 and 4.31)

As the number of explanatory variables,  $p$ , increases, especially when  $p = 6$ , there is reduction in the value of PRESS for  $k_5, k_{13}, k_{16}, k_{18}$  and  $k_{20}$ . when the variance of the random error increases.  $K_8$  maintained approximately the same

value as  $\sigma$  increases for all the values of the correlation coefficient. Also,  $k_2$  showed a steady and sharp increase as the  $\sigma$  increases (see Table 4.5(d) and Figure 4.31)

When  $p = 7$ , the graph of PRESS presented in Figure 4.33 showed that some of the values of  $k$  depicts reduction in PRESS as  $\sigma$  increases and in severe multicollinearity for  $k_5, k_{14}, k_{16}, k_{18}$  and  $k_{20}$  whereas  $k_1, k_7, k_8$  and  $k_9$ . Also, the values of PRESS for  $k_{10}, k_{13}, k_{18}$  and  $k_{20}$  showed almost the same value throughout the values of  $\sigma$  but it becomes smaller at higher collinearity.(see Table 4.5 (e))

Generally, as the number of explanatory variables increases, with increase in the standard deviation  $\sigma$ , the PRESS values for  $k_5, k_{10}, k_{13}$ , and  $k_{18}$  reduces drastically while those of  $k_1, k_2$  and  $k_7$ , increases drastically equally. But there is reduction in PRESS as the value of the correlation coefficient increases when the sample size,  $n$ , is small ( $n = 10$ ).

#### **4.13 SIMULATION RESULTS OF PREDICTION SUM OF SQUARE (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 30$**

Tables 4.6(a) – 4.6(g) showed the simulated value of PRESS when the sample size  $n = 30$  for the standard deviation  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ , and  $1.0$  and correlation coefficient  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$ . The values of their PRESS are also presented graphically in Figures 4.36 – 4.42 (see Appendix 7).

#### **4.14 DISCUSSION OF SIMULATION RESULTS OF PREDICTION SUM OF SQUARES (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 30$**

Table 4.6(a) and Figure 4.36 showed the simulated values of PRESS for the number of explanatory variables  $p = 3$ . The values of PRESS increase steadily as the standard deviation  $\sigma$  increases for the ridge estimators  $k_1, k_2, k_3, k_6$  and  $k_{21}$ , whereas  $k_5, k_8, k_{10}, k_{13}, k_{16}, k_{18}$  and  $k_{20}$  showed almost the same values of PRESS as the standard deviation  $\sigma$  increases.

The simulated values of PRESS for correlation coefficient  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$  and  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9,$  and  $1.0$  are presented in Table 4.6(b) and Figure 4.37. The results showed the same pattern of movement when  $p = 3$ . This implies that when the number of explanatory variable  $p$  is small, all the ridge estimators gave the same PRESS values

In Table 4.6(c) and Figure 4.48, the simulated PRESS values when  $p = 5$  are presented. It was observed from Table 4.6(c) and Figure 4.48 that the ridge estimators  $k_5, k_8$  and  $k_{10}$ , showed approximately the same values of PRESS as the  $\sigma$  increases. Then, for  $k_1, k_2, k_3, k_9, k_{15}, k_{17}$  and  $k_{21}$ , there is sharp increase in the values of PRESS as the standard deviation  $\sigma$  increases. They also showed higher value of PRESS when  $\rho = 0.3$  (small multicollinearity). But  $k_{12}, k_{13}, k_{16}, k_{18}$  and  $k_{20}$  showed the same moderate values of PRESS and decrease continuously as the standard deviation  $\sigma$  increases even in the presence of multicollinearity.

The PRESS values obtained when  $p = 6$  showed the same pattern of movement when  $p = 5$  except for that of  $k_2$  which recorded higher PRESS values when  $\sigma$  is  $0.9$  and  $1.0$ . (see Table 4.6(d) and Figure 4.39).

The simulated values of PRESS for the values of  $\sigma$  and  $\rho$  when the number of explanatory variables  $p = 7$  presented in Table 4.6 (e) and Figure 4.40 showed that the ridge estimators  $k_0$ ,  $k_9$  and  $k_{21}$  have almost the same PRESS as  $\sigma$  increases. But the ridge estimators  $k_5$ ,  $k_{10}$ ,  $k_{13}$ ,  $k_{16}$ ,  $k_{18}$  and  $k_{20}$  showed a decrease in PRESS values while  $k_8$  showed approximately the same value as  $\sigma$  increases.

The PRESS values for  $p = 8$  and  $p = 9$  are presented in Tables 4.6(f) and 4.6(g) and in Figures 4.41 and 4.42 respectively. The PRESS values for  $k_2$  and  $k_3$  showed an outstanding steady increase than the other estimators, but  $k_5$  showed an outstanding decrease.  $k_8$ ,  $k_{16}$ ,  $k_{18}$  and  $k_{20}$  maintained approximately the same values of PRESS as  $\sigma$  increases.

Consequently, when the number of explanatory variables increases for a small sample size, there is steady increase in the values of PRESS of the estimators  $k_0$ ,  $k_1$ ,  $k_2$ ,  $k_3$ ,  $k_{17}$ ,  $k_{19}$  and  $k_{21}$ , but as the number of explanatory variable,  $p$ , increases, the value of PRESS of the estimators  $k_5$ ,  $k_8$ ,  $k_{13}$ ,  $k_{16}$ ,  $k_{18}$  and  $k_{20}$  decreases continuously, or sometimes maintained approximately the same small values in the presence of multicollinearity. Hence,  $k_8$ ,  $k_{18}$  and  $k_{20}$  are not much affected by the variance of the random error  $\sigma$

#### **4.15 SIMULATION RESULTS OF PREDICTION SUM OF SQUARES (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 50$**

The simulated values of PRESS when the sample size  $n = 50$  for correlation coefficient  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$  and the standard deviation  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$  are presented in Tables 4.7(a) – 4.7(g) and Figures 4.43 – 4.49 (see Appendix 8)

#### **4.16 DISCUSSION OF SIMULATION RESULTS OF PREDICTION SUM OF SQUARE (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 50$**

Table 4.7(a) and Figure 4.43 showed the simulated values of PRESS for  $p = 3$  when  $n = 50$ . There is increase in the values of PRESS for the ridge estimators  $k_0, k_1, k_2, k_6, k_{11}, k_{19}$  and  $k_{21}$  as  $\sigma$  increases. More so, the PRESS values for the ridge estimators  $k_8, k_{12}, k_{16}, k_{18}$  and  $k_{20}$  are approximately the same as  $\sigma$  increases for different values of multicollinearity.

When the number of explanatory variables is 4 or 5, the simulated PRESS values are presented in Tables 4.7(b) and 4.7(c) respectively. The values of PRESS for the ridge estimators  $k_1, k_2, k_3, k_7, k_{11}, k_{19}$  and  $k_{21}$  increase from a very small value when  $\sigma$  increases to higher values,. Moreover, the ridge estimators  $k_8, k_{12}, k_{13}, k_{14}, k_{16}, k_{18}$  and  $k_{20}$  maintained approximately the same values of PRESS as the standard deviation  $\sigma$  increases. (also see Figures 4.44 and 4.45)

The PRESS values for  $p = 6$  are presented in Table 4.7(d) and Figure 4.46. It is observed from Table 4.7(d) that as the standard deviation  $\sigma$  increases, the PRESS values for the estimators  $k_{18}$  and  $k_{20}$  and the PRESS values for  $k_1, k_2, k_3, k_4, k_{11}, k_{15}$  and  $k_{19}$  increases. Consequently when  $p = 7$ , the simulated PRESS values showed an increase for the ridge estimators  $k_1, k_2, k_3, k_4, k_7, k_9, k_{19}$  and  $k_{20}$  respectively. Hence, they have smaller values of PRESS when  $\sigma$  is small and larger values of PRESS when  $\sigma$  is large. (see Table 4.7 (e) and Figure 4.47)

When the number of explanatory variables  $p = 8$  and  $p = 9$ , the simulated values of PRESS are presented in Table 4.7(f) and 4.7(g) and Figures 4.48 and 4.49 respectively. Tables 4.7 (f), (g) show increase in the simulated values of PRESS

as  $\sigma$  increases for the ridge estimators  $k_1, k_2, k_3, k_4, k_7, k_9, k_{11}, k_{19}$  and  $k_{21}$ , and a little decrease of PRESS for the ridge estimators  $k_8, k_{13}, k_{16}, k_{18}$  and  $k_{20}$ .

In summary, for large sample size  $n = 50$ , and for increase in the number of explanatory variables  $p$  and the standard deviation  $\sigma$  the simulated values of PRESS showed continuous increase even in the presence of multicollinearity for the ridge estimators  $k_1, k_2, k_3, k_4, k_6, k_{11}, k_{12}, k_{15}, k_{17}, k_{19}$  and  $k_{21}$ . The ridge estimators  $k_5, k_8, k_{10}, k_{13}, k_{14}, k_{16}, k_{18}$  and  $k_{20}$  showed smaller steadily decreasing values of PRESS as  $\sigma$  and the number of explanatory variables  $p$  increases respectively in the presence of multicollinearity.

#### **4.17 SIMULATION RESULTS OF PREDICTION SUM OF SQUARES (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 80$**

The simulated PRESS values for the ridge estimators when the sample size,  $n = 80$  are presented in Tables 4.8(a) – 4.8(g) and Figures 4.50 -4.56 for the correlation coefficient  $\rho = 0.3, 0.5, 0.7, 0.9$ , and  $0.99$ , the standard deviation  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$  and  $1.0$  and the number of explanatory variables  $p = 3, 4, 5, 6, 7, 8$ , and  $9$  (see Appendix 9).

#### **4.18 DISCUSSION OF SIMULATION RESULTS OF PREDICTION SUM OF SQUARES (PRESS) OF THE PROPOSED ESTIMATORS WHEN THE SAMPLE SIZE $n = 80$**

When the number of explanatory variable  $p = 3$ , the simulated PRESS values are presented in Table 4.8(a) and Figure 4.50. Observations made from Table 4.8(a) is that as the variance of the random error  $\sigma$  increases, the press values for  $k_1, k_2, K_3, k_4, k_7, k_{11}, k_{19}$  and  $k_{21}$  increases respectively, where as the estimators  $k_5, k_8, k_{12}, k_{16}, k_{18}$  and  $k_{20}$ , maintained the same PRESS values and decreases respectively as the value of  $\sigma$  increases, but  $k_8$  and  $k_{20}$  showed smaller values PRESS than the other estimators.



When the number of explanatory variables  $p = 4$ , the simulation result showed a continuous and sharp increase in the values of PRESS for the ridge estimators  $k_1, k_2, k_3$  and  $k_{21}$ , where as the ridge estimators  $k_5, k_8, k_{14}, k_{18}$  and  $k_{20}$  exhibited small decrease in PRESS values when the standard deviation increases.(see Table 4.8(b) and Figure 4.51)

The simulated values of PRESS are presented in Table 4.8(c) and Figure 4.52 for the number of explanatory variables  $p = 5$  and Table 4.8(d) and Figure 4.53 when  $p = 6$ . Tables 4.8(c), (d) showed increase in PRESS values for the ridge estimators  $k_1, k_2, k_3, k_4, k_7, k_{19}$ , and  $k_{21}$  from 0.05 to approximately 0.5 on the average. However, the PRESS values when  $p = 7$  showed decrease in PRESS value for the ridge estimators  $k_5, k_8, k_{12}, k_{18}$  and  $k_{20}$ . (Table 4.8(e) and Figure 4.54)

Furthermore, the simulated PRESS values when  $p = 8$  and when  $p = 9$  are presented in Tables 4.8(f) and 4.8(g) and in Figures 4.55 and 4.56 respectively. Tables 4.8(f), (g) approximately have the same observations in such a way that the simulated PRESS values for  $k_1, k_2, k_3, k_4, k_{11}, k_{15}, k_{19}$  and  $k_{21}$  increase as the standard deviation  $\sigma$  increases while those of  $k_5, k_8, k_{12}, k_{16}, k_{18}$  and  $k_{20}$  decrease in PRESS value when the standard deviation  $\sigma$  decreases and in the presence of multicollinearity.

#### **4.19 SUMMARY OF DISCUSSIONS ON THE MSE AND PRESS VALUES OF THE PROPOSED ESTIMATORS**

The simulated results were presented in Tables 4.1 to 4.8 and Figures 4.1 to 4.56. In Tables 4.1 to 4.4 and Figures 4.1 to 4.28, we presented the simulated MSE values for number of sample size  $n = 10, 30, 50$  and  $80$ , for values of  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$  and  $1.0$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$  when  $p = 3, 4, 5, 6, 7, 8$ , and  $9$ . Furthermore, the simulated PRESS values were presented in Tables 4.5 to 4.8 and Figures 4.29 to 4.56 for sample size  $n$

=10, 30, 50 and 80, for values of  $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$  and  $1.0$ , and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$  when  $p = 3, 4, 5, 6, 7, 8,$  and  $9$ .

The dominant effect on the MSE and PRESS is the variance of the error term  $\sigma$  (standard deviation) of the dependent variable which leads to a higher MSE and PRESS. Increasing the strength of the correlation has little impact on the PRESS of the ridge estimators, where as an increase in the strength of the correlation increases the MSE of the ridge estimators.

Comparing the MSE with the value of PRESS under different ridge estimators and keeping the variance of the random error  $\sigma$ , number of explanatory variables  $p$  and sample size  $n$  fixed, the expected relationship between PRESS and MSE is evident, since a smaller MSE leads to a smaller PRESS. But ridge estimators produced smaller MSE and PRESS even when the number of explanatory variables increases. Based on the PRESS and MSE criteria, we observed that estimators  $k_1, k_2, k_3, k_4, k_7, k_{10}, k_{11}, k_{12}, k_{15}, k_{17}, k_{19},$  and  $k_{21}$  performed better when compared to other ridge estimators.

From the result of the simulation studies, for  $\rho=0.9$  and  $\rho=0.99$ , the ridge regression has shown itself to be better in terms of yielding a lower MSE than the method of OLS. However, where  $\rho = 0.3$  and  $\rho = 0.5$ , the result of the simulation studies coincide. This indicates that ridge regression is very effective when multicollinearity is present that is when the explanatory variables are highly correlated.

The result of the simulation studies on the proposed estimators are in line with the works of Hoerl et al (1975), Khalaf and Shukur (2008) and Khalaf et al (2010) but there were substantial reduction in MSE of the proposed estimators in cases of severe multicollinearity.

Therefore, the number of explanatory variables, standard deviation, amount of correlation have significant effect on the values of MSE and PRESS for various

ridge regression estimators. The ridge regression yields better estimates than the methods of OLS. Of all the estimators,  $k_{11}$  and  $k_{19}$  performed better in terms of smaller MSE and PRESS.

## **4.20 DATA COLLECTION AND ANALYSIS**

### **4.20.1 DATA COLLECTION**

To illustrate the performance of the estimators, we considered economic indicator data from the Central Bank of Nigeria (CBN) Statistical Bulletin 2010 edition.(see Appendix II) The data consist of Gross Domestic Product, ( $Y$ ) as the dependent variable and 10 (ten) independent variables namely:

Total money supply( $X_1$ ), Credit to private sector( $X_2$ ), Exchange rate ( $X_3$ ) External reserve ( $X_4$ ), Agricultural Loan( $X_5$ ), Foreign trade ( $X_6$ ), Oil import ( $X_7$ ), Non oil import ( $X_8$ ), Oil export ( $X_9$ ) and Non-oil export ( $X_{10}$ )

### **4.20.2 DATA ANALYSIS**

We considered the linear regression model for the data:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \varepsilon_i$$

Since the data are not of the same unit of measurement, we centre and scale the data using unit length scaling method described in section 3.4 and the result is presented in Table 4.9.

**TABLE 4.1 CENTERED AND SCALED DATA FOR ANALYSIS**

GDP	MONEY SUPPLY	CREDIT PRIV SECTOR	EXCHANGE RATE	EXTERNAL RESERV	AGRIC LOAN	FOREIGN TRADE	OIL IMPORT	NON OIL IMPORT	OIL EXPORT	NON OIL EXPORT
-0.062	-0.1075	-0.09409	-0.176936	-0.08808	-0.10463	-0.13431	-0.1191	-0.12918	-0.12927	-0.11678
-0.062	-0.10743	-0.09395	-0.176735	-0.13108	-0.10493	-0.13448	-0.119	-0.12945	-0.12942	-0.11704
-0.062	-0.10733	-0.09388	-0.176572	-0.13613	-0.10458	-0.13457	-0.119	-0.12968	-0.12947	-0.11686
-0.062	-0.10721	-0.09382	-0.176442	-0.13779	-0.10547	-0.13457	-0.119	-0.12991	-0.12938	-0.11696
-0.062	-0.10709	-0.09378	-0.176031	-0.13238	-0.10397	-0.13448	-0.1191	-0.12989	-0.12924	-0.11649
-0.062	-0.10702	-0.09363	-0.172439	-0.12461	-0.10212	-0.13462	-0.1188	-0.13014	-0.1294	-0.11639
-0.062	-0.10682	-0.09322	-0.166072	-0.08185	-0.09954	-0.13346	-0.1179	-0.12893	-0.1283	-0.11341
-0.062	-0.10626	-0.09279	-0.164418	-0.07233	-0.09828	-0.1333	-0.1177	-0.12856	-0.12829	-0.11228
-0.061	-0.10586	-0.09258	-0.155316	-0.10008	-0.09746	-0.13204	-0.1173	-0.12749	-0.12682	-0.11191
-0.061	-0.1055	-0.09236	-0.153256	-0.10367	-0.09982	-0.12971	-0.1168	-0.1258	-0.12396	-0.11134
-0.06	-0.10433	-0.09181	-0.147289	-0.09551	-0.10108	-0.12777	-0.1161	-0.12053	-0.12339	-0.1087
-0.059	-0.10246	-0.09065	-0.123734	-0.11043	-0.10062	-0.12296	-0.1116	-0.11528	-0.1187	-0.10954
-0.058	-0.09964	-0.08585	-0.108582	-0.08551	-0.10117	-0.12172	-0.1034	-0.11517	-0.11801	-0.10812
-0.056	-0.09625	-0.08472	-0.109108	-0.04973	-0.09946	-0.00637	-0.1029	-0.11567	-0.11874	-0.10745
-0.049	-0.09319	-0.08218	-0.109108	-0.11146	-0.09479	-0.07559	-0.0595	-0.05565	-0.07844	-0.07436
-0.043	-0.09023	-0.07811	-0.109108	-0.11778	-0.0901	-0.06978	-0.0571	-0.08057	-0.05856	-0.07393
-0.043	-0.08671	-0.07271	-0.109108	-0.07113	-0.08883	-0.06226	-0.0553	-0.04568	-0.06265	-0.06305
-0.043	-0.08281	-0.07023	-0.109108	-0.05125	-0.09085	-0.07965	-0.0518	-0.04784	-0.09007	-0.0539
-0.04	-0.07547	-0.06472	0.116625	-0.08072	-0.08852	-0.06352	-0.0381	-0.04918	-0.06503	-0.08108
-0.03	-0.06246	-0.05783	0.14663	-0.05388	-0.07969	-0.03282	-0.0346	-0.03497	-0.02337	-0.07114
-0.029	0.553826	-0.04152	0.177993	-0.02228	-0.05159	-0.02251	-0.0284	0.009763	-0.02786	-0.0652
-0.014	-0.02963	-0.03002	0.20677	-0.04211	-0.02686	-0.02143	0.01928	0.013512	-0.03842	0.05919
-0.003	-0.00644	-0.01848	0.233505	-0.05328	-0.01822	0.04529	0.03352	0.079992	0.036071	0.059273
0.0176	0.002886	0.004121	0.246716	-0.00129	0.052154	0.09493	0.0026	0.07844	0.119029	0.093825
0.9535	0.029272	0.033085	0.242403	0.14289	0.125873	0.21564	0.18593	0.120387	0.266005	0.080115
0.066	0.089833	0.064516	0.231257	0.29737	0.218985	0.22911	0.1528	0.169813	0.268805	0.131644
0.0809	0.159651	0.161088	0.222273	0.39074	0.231447	0.29155	0.17481	0.263315	0.319777	0.254061
0.1122	0.310342	0.387757	0.199106	0.54456	0.40715	0.40102	0.41145	0.345968	0.419743	0.344631
-0.043	0.38476	0.541082	0.295816	0.3826	0.53181	0.33474	0.2878	0.375542	0.317378	0.421653
-0.041	0.467053	0.611252	0.300268	0.29619	0.48519	0.52963	0.67423	0.612817	0.459979	0.621555

Hence, the transformed model becomes:

$$\tilde{Y} = \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \beta_3 \tilde{X}_3 + \beta_4 \tilde{X}_4 + \beta_5 \tilde{X}_5 + \beta_6 \tilde{X}_6 + \beta_7 \tilde{X}_7 + \beta_8 \tilde{X}_8 + \beta_9 \tilde{X}_9 + \beta_{10} \tilde{X}_{10} + \varepsilon'$$

The OLS method is applied to the data in Table 4.9 and the following estimates are obtained using computer programme in R (see Appendix XI) and the results are presented in Tables 4.2, 4.3 and 4.4.

**Table 4.2 REGRESSION COEFFICIENTS AND THEIR VARIANCE INFLATION FACTORS (VIF)**

Predictor	Coef	SE coeff	tvalue	Pr(> t )	VIF
Constant	0	0.02208	0	1	
money	0.1105	0.2961	0.37	0.713	5.998
cred.	-0.173	1.328	-0.13	0.898	120.598
exchange	0.2884	0.3088	0.93	0.362	6.523
external	-1.1082	0.5152	-2.15	0.045	18.155
agric	0.575	1.509	0.38	0.707	155.735
foreign	0.41	1.109	0.37	0.715	84.11
oil	1.4578	0.85	1.72	0.103	49.418
nonoilimport	-3.905	2.033	-1.92	0.07	282.606
oilexport	2.952	1.387	2.13	0.047	131.644
nonoilexport	-0.228	1.571	-0.15	0.886	168.873

The regression equation is:

$$\hat{Y} = 0.000 + 0.111X_1 - 0.173X_2 + 0.288X_3 - 1.108X_4 + 0.575X_5 + 0.41X_6 + 1.458X_7 - 3.905X_8 + 2.952X_9 - 0.228X_{10}$$

**Table 4.3 ANALYSIS OF VARIANCE (ANOVA) TABLE FOR REGRESSION**

Source	DF	SS	MS	F	P
Regression	10	0.72221	0.07222	4.94	0.001
Residual error	19	0.27779	0.01462		
Total	29	1			

S = 0.120916 R-Sq = 72.2% R-Sq(adj) = 57.6% PRESS = 49956.6 R-Sq(pred) = 0.00%

**Table 4.4 ANALYSIS OF VARIANCE TABLE FOR REGRESSION COEFFICIENTS**

Source	DF	seq SS	Adj. SS	Adj MS	F	P-value
<b>Regression</b>	10	0.72221	0.722206	0.072221	5.19959	0.00086
<b>money</b>	1	0.01977	0.002037	0.002037	0.14663	0.705817
<b>cred.</b>	1	0.00249	0.000248	0.000248	0.01783	0.895123
<b>exchange</b>	1	0.18919	0.012748	0.012749	0.91784	0.349482
<b>external</b>	1	0.07249	0.067642	0.067642	4.86996	0.039182
<b>agric</b>	1	0.06037	0.002123	0.002123	0.15284	0.699971
<b>foreign</b>	1	0.11848	0.002002	0.002002	0.14413	0.708213
<b>oil</b>	1	0.05044	0.043003	0.043003	3.09606	0.093771
<b>nonoilimport</b>	1	0.13302	0.053959	0.053959	3.88482	0.062722
<b>oilexport</b>	1	0.07564	0.066179	0.066179	4.76461	0.041144
<b>nonoilexport</b>	1	0.00031	0.000309	0.000309	0.02221	0.883016
<b>Error</b>	20	0.27779	0.277794	0.01389		
<b>Total</b>	30	1				

Table 4.2 shows the values of the regression coefficient and their corresponding variance inflation factors (VIF). We observed from Table 4.2 at 1% level of significance, that none of the regression coefficients is significant because all

the p-values are more than 0.01, while at 5% level of significance, only the variable – oil export is significant

Furthermore, the regression coefficients of credit to private sector, external reserve, non-oil import and non-oil export show negative impact on GDP (the dependent variable), whereas others which are positive show that as they increases the GDP also increases.

The analysis of variance table for regression model presented in Table 4.3 shows that the model is adequate for prediction since the p-value = 0.001 which is less than 0.01 (1%) .This can be further explained by the value of the  $R^2 = 0.722$  which shows that 10-variable regression model explains 72.2% of the variation in Gross Domestic Product (GDP).

Here, we diagnose the data on explanatory variables for the presence of multicollinearity using the off diagonal elements of the correlation matrix, variance inflation factor and eigenvalues of the correlation matrix ( $\tilde{X}\tilde{X}$ ) respectively. The off diagonal elements of the correlation matrix ( $\tilde{X}\tilde{X}$ ) are shown below:

$$\begin{bmatrix} 1.00 & 0.795 & 0.722 & 0.731 & 0.784 & 0.776 & 0.779 & 0.815 & 0.753 & 0.777 \\ 0.795 & 1.00 & 0.681 & 0.859 & 0.970 & 0.917 & 0.942 & 0.952 & 0.885 & 0.969 \\ 0.722 & 0.681 & 1.00 & 0.728 & 0.751 & 0.827 & 0.765 & 0.823 & 0.835 & 0.781 \\ 0.731 & 0.859 & 0.728 & 1.00 & 0.937 & 0.932 & 0.866 & 0.889 & 0.744 & 0.878 \\ 0.784 & 0.970 & 0.751 & 0.937 & 1.00 & 0.958 & 0.937 & 0.960 & 0.951 & 0.968 \\ 0.776 & 0.917 & 0.827 & 0.932 & 0.958 & 1.00 & 0.966 & 0.979 & 0.988 & 0.963 \\ 0.779 & 0.942 & 0.765 & 0.866 & 0.937 & 0.966 & 1.00 & 0.980 & 0.946 & 0.971 \\ 0.815 & 0.952 & 0.823 & 0.889 & 0.960 & 0.979 & 0.980 & 1.00 & 0.961 & 0.991 \\ 0.753 & 0.885 & 0.835 & 0.744 & 0.951 & 0.988 & 0.946 & 0.961 & 1.00 & 0.941 \\ 0.777 & 0.969 & 0.781 & 0.878 & 0.968 & 0.963 & 0.971 & 0.991 & 0.941 & 1.00 \end{bmatrix}$$

A very simple measure of multicollinearity is the inspection of the off-diagonal elements  $r_{ij}$  in  $\tilde{X}\tilde{X}$  . If the regressors  $\tilde{x}_i$  and  $\tilde{x}_j$  are nearly dependent, then  $|r_{ij}|$  will be near unity. Evidently, the matrix of correlation shows that there is presence of multicollinearity among the factors because some of the factors

recorded high correlation coefficient of 0.991 for  $X_8X_{10}$  and 0.988 for  $X_6X_9$  . Thus, it is not surprising as these variables are economic variables and they are highly correlated. Thus, the correlation matrix indicates that there are linear dependences of variables (regressors).

Also, the variances of the estimated regression coefficients are large, (see Table 4.2) which shows that the estimated regression coefficients are not stable.

The eigenvalues of the correlation matrix  $\tilde{X}\tilde{X}$  are  $\lambda_1= 8.934482$ ,  $\lambda_2=0.408675$ ,  $\lambda_3= 0.33299$ ,  $\lambda_4 = 0.193776$ ,  $\lambda_5= 0.078547$ ,  $\lambda_6 = 0.019136$ ,  $\lambda_7= 0.017502$ ,  $\lambda_8= 0.009329$ ,  $\lambda_9= 0.003579$  and  $\lambda_{10}= 0.001983$ .

The eigenvalues are used to measure the extent of the multicollinearity in the data. If there are near-linear dependences of variables, then the value of one or more eigenvalues will be small. These small eigenvalues imply that there are near-linear dependences among the columns of  $\tilde{X}$  . In this setting, we examine the condition number of the matrix  $\tilde{X}\tilde{X}$

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$$

which defines the measure of spread in the eigenvalues spectrum of  $\tilde{X}\tilde{X}$  . Generally, if the condition number is less than 100, there is no serious problem with multicollinearity. Condition number between 100 and 1000 imply strong multicollinearity and if it exceeds 1000, severe multicollinearity is indicated.

There are three small eigenvalues which is an indicator that the data is ill conditioned. The condition number is 4505 which is more than 1000. This indicates the presence of severe multicollinearity.

The variance inflation factor (VIF) as proposed by Marquardt (1963) is a factor that measures the combined effect of the dependencies among the regressors on



the variance. Large value of VIF indicates the presence of multicollinearity. If the value of VIF exceeds 10, it is an indicator that the associated regression coefficients are poorly estimated because of multicollinearity. From Table 4.2, the maximum VIF is 282.606 which is an evidence of the presence of multicollinearity in the data. Furthermore, the VIF of other variables are large. This shows that eight regressors are involved in the multicollinearity.

The eigenvectors (D) of the matrix of the explanatory variables corresponding to the eigenvalues of  $(\tilde{X}\tilde{X})$  are shown below:

$$D = \begin{pmatrix} d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 & d_8 & d_9 & d_{10} \\ -0.2796 & 0.4402 & 0.7901 & 0.2529 & 0.1560 & 0.0729 & 0.0585 & 0.0169 & -0.0592 & -0.5487 \\ -0.3186 & -0.3124 & 0.2749 & -0.2251 & -0.4322 & -0.0735 & -0.3210 & -0.2302 & 0.5122 & -0.2533 \\ -0.2791 & 0.7673 & -0.3819 & -0.1737 & -0.3256 & -0.1412 & -0.1748 & -0.0142 & -0.0005 & -0.0252 \\ -0.311 & -0.1619 & -0.1959 & 0.742 & -0.0722 & -0.5293 & 0.0384 & 0.0186 & 0.0298 & -0.0098 \\ -0.3271 & -0.2203 & 0.0203 & 0.1188 & -0.4420 & 0.4374 & -0.2078 & 0.1874 & -0.4581 & 0.3938 \\ -0.3302 & -0.0331 & -0.1841 & 0.0309 & 0.2875 & 0.2449 & 0.0826 & -0.8276 & -0.1387 & 0.0342 \\ -0.3249 & -0.1391 & 0.0088 & -0.3176 & 0.5264 & -0.3334 & -0.5274 & 0.1977 & -0.2594 & 0.0346 \\ -0.3316 & -0.0233 & 0.0020 & -0.2398 & 0.1134 & -0.1041 & 0.4262 & 0.1337 & 0.4094 & 0.6642 \\ -0.3265 & -0.0013 & -0.2855 & 0.1736 & 0.2964 & 0.5405 & -0.0030 & 0.3895 & 0.3497 & -0.3574 \\ -0.3280 & -0.1554 & 0.0087 & -0.3180 & -0.1473 & -0.1541 & 0.5930 & 0.1311 & -0.3834 & -0.4536 \end{pmatrix}$$

where the eigenvectors  $d_i$  corresponds to the eigenvalues  $\lambda_i$  for  $i = 1, 2, \dots, 10$ . The vector  $\alpha_j$  is obtained as  $D'\beta$ , where  $\beta$  is the value of the regression coefficients obtained using OLS method as presented in Table 4.2 and D is the matrix of the eigenvectors corresponding to the eigenvalues of the correlation matrix  $(\tilde{X}\tilde{X})$ . Therefore

$$\hat{\alpha}_j = (-0.10273, 0.28273, -0.75665, 0.33407, 1.17513, 2.46951, -2.69405, 0.67104, -1.30587, -3.21296)$$

The eigenvalues of the correlation matrix  $(\tilde{X}\tilde{X})$

$$8.934481565, 0.408675312, 0.332990495, 0.193776346, 0.078546823, 0.019136476, 0.017501792, 0.009328713, 0.003579253, 0.001983224$$

have the skewness value of 2.23 which shows that the eigenvalues are exponentially distributed, and the maximum eigenvalue 8.934481565 is an

outlier. Since there is an outlier, implementing the proposed estimator in the ridge regression estimation, to obtain the values of ridge regression coefficients, the mean square error (MSE) and the prediction sum of square (PRESS) of each value of k as shown in Table 4.5. For easy comparison, the values of the ridge regression coefficients, the MSE and the PRESS of the existing estimators are also presented in Table 4.5.

**TABLE 4.5 RIDGE REGRESSION COEFFICIENTS, MEAN SQUARE ERROR AND PREDICTED SUM OF SQUARE FOR DIFFERENT VALUES OF RIDGE PARAMETERS (k)**

	$k_0$	$k_1$	$k_2$	$k_3$	$k_4$	$k_5$	$k_6$	$k_7$	$k_8$	$k_9$	$k_{10}$
k	0	5E-04	0.0019	0.0718	0.0005	1993.6	0.234	<b>0.0067</b>	190.65	0.0063	771.5
PRESS	3.2236	2.536	1.8983	0.7272	2.5368	0.9993	0.8949	<b>1.464</b>	0.9936	1.4876	10.744
MSE	34.9666	9.1988	9.64723	3.32994	<b>6.7103</b>	<b>3.137</b>	3.22648	<b>2.1329</b>	3.1189	<b>8.1594</b>	27.989
b1	0.1105	0.111	0.1108	0.1206	0.1106	7E-05	-0.1539	<b>-0.0914</b>	0.0007	-0.087	0.0002
b2	-0.1728	-0.17	-0.173	-0.163	-0.173	7E-05	-0.2708	<b>-0.4448</b>	0.0007	-0.442	0.0002
b3	0.2884	0.289	0.2891	0.3172	0.2886	0.0002	0.2199	<b>0.2122</b>	0.002	0.2134	0.0005
b4	-1.1082	-1.11	-1.108	-1.084	-1.108	0.0002	0.0742	<b>-0.7475</b>	0.0017	-0.765	0.0004
b5	0.575	0.575	0.5755	0.5942	0.5751	0.0001	-0.0046	<b>0.7135</b>	0.0013	0.7279	0.0003
b6	0.4103	0.411	0.411	0.4374	0.4105	0.0002	0.2093	<b>0.4844</b>	0.0019	0.4834	0.0005
b7	1.4578	1.458	1.4584	1.4807	1.458	0.0002	0.1569	<b>1.0422</b>	0.0016	1.0588	0.0004
b8	-3.905	-3.9	-3.905	-3.886	-3.905	0.0001	-0.0966	<b>-1.5677</b>	0.0013	-1.617	0.0003
b9	2.9516	2.952	2.9524	2.9831	2.9518	0.0002	0.3874	<b>1.8701</b>	0.0022	1.8996	0.0006
b10	-0.2283	-0.23	-0.228	-0.213	-0.228	0.0001	-0.2074	<b>-1.1569</b>	0.0011	-1.157	0.0003

$k_0 \equiv$  OLS,  $k_1 \equiv$  Hoerl and Kernard (1970),  $k_2 \equiv$  Hoerl et al (1975),  $k_3 \equiv$  Lawless and Wang (1976),  $k_4 \equiv$  Khalaf and Shurkur (2005),  $k_5 - k_9 \equiv$  Muniz et al (2010),  $k_{10} - k_{21} \equiv$  the proposed estimators

**TABLE 4.5 (cont) RIDGE REGRESSION COEFFICIENTS, MEAN SQUARE ERROR AND PREDICTED SUM OF SQUARE FOR DIFFERENT VALUES OF RIDGE PARAMETERS (k)**

	<b>k<sub>11</sub></b>	<b>k<sub>12</sub></b>	k <sub>13</sub>	k <sub>14</sub>	k <sub>15</sub>	k <sub>16</sub>	k <sub>17</sub>	k <sub>18</sub>	k <sub>19</sub>	k <sub>20</sub>	<b>k<sub>21</sub></b>
k	<b>0.0324</b>	<b>0.0044</b>	229.66	217.27	0.0053	2337.8	<b>0.0029</b>	687.72	0.0015	534.91	<b>0.0019</b>
PRESS	<b>1.028</b>	<b>1.6273</b>	0.9947	0.9944	1.5528	0.9994	<b>1.7983</b>	0.9981	2.1015	0.9976	<b>1.977</b>
MSE	<b>7.1641</b>	<b>11.023</b>	27.9788	27.9764	<b>12.019</b>	27.99	<b>8.9083</b>	27.9878	<b>8.8645</b>	27.9872	<b>8.8415</b>
b <sub>1</sub>	<b>-0.167</b>	<b>-0.0637</b>	0.0006	0.0006	-0.076	6E-05	<b>-0.036</b>	0.0002	0.0097	0.0003	<b>-0.008</b>
b <sub>2</sub>	<b>-0.481</b>	<b>-0.4224</b>	0.0006	0.0006	-0.433	6E-05	<b>-0.397</b>	0.0002	-0.3455	0.0003	<b>-0.3673</b>
b <sub>3</sub>	<b>0.2106</b>	<b>0.2215</b>	0.0017	0.0018	0.2171	0.0002	<b>0.2319</b>	0.0006	0.249	0.0007	<b>0.2423</b>
b <sub>4</sub>	<b>-0.244</b>	<b>-0.8519</b>	0.0014	0.0015	-0.809	0.0001	<b>-0.929</b>	0.0005	-1.0125	0.0006	<b>-0.9844</b>
b <sub>5</sub>	<b>0.2403</b>	<b>0.7914</b>	0.0011	0.0012	0.762	0.0001	<b>0.8259</b>	0.0004	0.8106	0.0005	<b>0.8253</b>
b <sub>6</sub>	<b>0.4343</b>	<b>0.4756</b>	0.0016	0.0017	0.4801	0.0002	<b>0.4646</b>	0.0005	0.4469	0.0007	<b>0.4537</b>
b <sub>7</sub>	<b>0.5877</b>	<b>1.1459</b>	0.0013	0.0014	1.1018	0.0001	<b>1.2289</b>	0.0005	1.3279	0.0006	<b>1.2934</b>
b <sub>8</sub>	<b>-0.603</b>	<b>-1.9027</b>	0.0011	0.0012	-1.752	0.0001	<b>-2.235</b>	0.0004	-2.7555	0.0005	<b>-2.5533</b>
b <sub>9</sub>	<b>1.1044</b>	<b>2.0586</b>	0.0019	0.002	1.9768	0.0002	<b>2.2241</b>	0.0006	2.4599	0.0008	<b>2.3706</b>
b <sub>10</sub>	<b>-0.775</b>	<b>-1.1316</b>	0.0009	0.0009	-1.151	9E-05	<b>-1.051</b>	0.0003	-0.8478	0.0004	<b>-0.9353</b>

$k_0 \equiv$  OLS,  $k_1 \equiv$  Hoerl and Kernard (1970),  $k_2 \equiv$  Hoerl et al (1975),  $k_3 \equiv$  Lawless and Wang (1976),  $k_4 \equiv$  Khalaf and Shurkur (2005),  $k_5 - k_9 \equiv$  Muniz et al (2010),  $k_{10} - k_{21} \equiv$  the proposed estimators

#### **4.21 DISCUSSION ON THE COMPARISON OF PROPOSED AND EXISTING ESTIMATORS USED IN REAL LIFE STUDY**

The Gross Domestic Product (GDP) of Nigeria is regressed on 10 economic predictor variables (total money supply, credit to private sector, exchange rate, external reserve, agricultural loan, foreign trade, oil import, non oil import, oil export and non oil export). Evidently, all the indicators of multicollinearity were present in the problem (except exchange rate and money supply).

We observed that all the estimators performed better than OLS in terms of MSE. Furthermore, when the estimators are compared using PRESS statistic, some estimators exhibited smaller PRESS values than the OLS. Amongst the

proposed estimators:  $k_{11}$ ,  $k_{12}$ ,  $k_{15}$ ,  $k_{17}$ ,  $k_{19}$ , and  $k_{21}$ ;  $k_7$  and  $k_9$  of Muniz et al (2010) showed smaller MSE than the other estimators. The estimators  $k_{11}$ ,  $k_{17}$ ,  $k_{19}$  and  $k_{21}$  showed smaller PRESS than the OLS. However, estimator  $k_{11}$  performed better than all the estimators in terms of smaller MSE and PRESS with value of  $k_{11}=0.0324$ . Hence, in a situation where the eigenvalues are skewed, the estimator  $k_{11}$  performs better than the other estimators, but where there is presence of an outlier among the eigenvalues of the matrix of the explanatory variables the estimator  $k_{17}$  does better than other estimators.

## CHAPTER FIVE

### SUMMARY, CONCLUSION AND RECOMMENDATION

#### 5.0 INTRODUCTION

This chapter gives the summary, conclusion and recommendations on the study. The contribution to knowledge and suggestion for further studies are also highlighted.

#### 5.1 SUMMARY OF FINDINGS

Problems involving eigenvalues of the matrix of the explanatory variables in estimating the ridge parameters for solving multicollinearity problem in multiple regressions have been considered in this research work. This is a situation where the eigenvalues of the explanatory variable are skewed or among the eigenvalues, there is an outlier. Methods of estimating the ridge parameter were proposed taking into consideration the skewness and the outlier among the eigenvalues. A computer program in R was written for the implementation of the proposed methods and its comparison with some existing methods via Monte Carlo simulation. Mean square error (MSE) and prediction sum of square (PRESS) were used as criteria for comparing the estimators. The comparison was made under the same sample size  $n$ , the random error  $\sigma$ , number of explanatory variable  $p$ , and correlation coefficient  $\rho$ .

From the results obtained from this simulation study, it is evident that the performance of ridge regression depends on the random error and the correlation among the explanatory variables. Also increase in the standard deviation of the random error,  $\sigma$  increases the mean square error and PRESS of an estimator.

The ridge estimator studied have shown to be better than OLS method in all the cases in terms of smaller mean square error. They exhibit a substantial reduction

in MSE in all the cases and PRESS in some of the cases. Also the simulation study has shown that for  $\rho=0.9$  and  $\rho = 0.99$ , (a severe multicollinearity) ridge regression yields minimum mean square error where as in a situation of low multicollinearity ( $\rho =0.3$  and  $\rho =0.5$ ) the results coincide with OLS. This indicates that ridge regression is efficient when multicollinearity is high and present that is when the explanatory variables are highly collinear.

The ridge regression was also applied to an economic data on Nigerian Economic indicators using GDP as dependent variable and ten other explanatory variables obtained from CBN Statistical Bulletin (2010) to illustrate in real life situation the ridge regression estimation. The data were highly collinear, as high as  $r_{8,10} = 0.99$ . from the estimated ridge regression coefficients obtained, we observed that the coefficient of  $b_1(k)$  change sign from positive to negative and that some of the ridge regression parameter estimates recorded high values of  $b_j(k)$  while some showed reduction in values of  $b_j(k)$

The eigenvalues of the matrix  $(\tilde{X}\tilde{X})$  is skewed with a skewness value of 2.23. Further observations were that some of the estimators recorded high values of  $k$  ( $k_5, k_{19}$ ) and some recorded smaller values of  $k$  ( $k_{11}, k_{12}$  and  $k_{15}$ ). These smaller values of  $k$  yielded minimum MSE and smaller PRESS. Amongst all the estimators considered,  $k_{11}$  is the best in terms of minimum mean square and smaller PRESS with the value of  $k_{11}=0.0324$ .

Furthermore, Q-test for detection of outlier in the eigenvalues of the matrix  $(\tilde{X}\tilde{X})$ , the maximum eigenvalue is found to be an outlier. In view of this, the method of geometric mean is used in the eigenvalues to obtain the ridge parameter and their corresponding ridge regression coefficients. The performances of these estimators are encouraging when compared with the existing ones. We also observed that amongst the proposed estimators,  $k_{17}, k_{19}$ ,

and  $k_{21}$  performed well in terms of minimum mean square error and smaller PRESS, and  $k_{17}$  with the value of  $k_{17} = 0.003$  seems to be the best.

## 5.2 CONCLUSION

The following are the conclusions made on this research work:

1. The proposed estimators yielded smaller mean square error and predicted sum of square than the existing estimators.
2. In the simulation study, increasing the correlation between the explanatory variables has negative effect on the MSE and PRESS.
3. Increasing the number of regressors ( $p$ ) has positive effect on the mean square error and Prediction sum of square.
4. When the sample size increases, the mean square error decreases even when the correlation between the explanatory variables is large.
5. The performance of the estimator depends on the error variance of the distribution.
6. The proposed estimators  $k_{11}$ ,  $k_{12}$ ,  $k_{15}$ ,  $k_{17}$ ,  $k_{19}$  and  $k_{21}$  are better estimators than the existing ones. Amongst these proposed estimators,  $K_{11} = \max(w_j)$  has a smallest MSE and PRESS respectively, in all the simulation studies.
7. Based on the real life application, we observed that
  - (i) when the ridge regression was introduced, the regression coefficients  $b_1$  changed sign due to the presence of multicollinearity.
  - (ii) the size of the regression coefficients were either reduced or enlarged.

In conclusion, when the eigenvalues of the matrix of the explanatory variables are skewed, the proposed ridge estimator  $k_{11}$  performed better than existing ones

in terms of PRESS and MSE. Whereas when there is an outlier among the eigenvalues of the matrix of the explanatory variables the proposed ridge estimator  $k_{17}$  performed better than other existing ridge estimators in terms of PRESS and MSE.

### 5.3 RECOMMENDATIONS

Based on the findings from this work, we make the following recommendations:

- When the eigenvalues of a correlation matrix are skewed,  $K_{11} = \max(w_j)$  will be used to estimate the ridge parameter.
- In the presence of outlier among the eigenvalues of the correlation matrix,  $K_{17} = \max(v_j)$  will be used to estimate the ridge parameter.
- Since in both cases,  $K_{17}$  exhibits smaller mean square error (MSE) and prediction sum of square (PRESS), it will be used to estimate the ridge parameter.



## REFERENCES

- Adnan N, Ahmad M. H and Adnan R. (2006), A Comparative Study on Some Methods for Handling Multicollinearity Problems, *Mathematika*, **22**, 109-119.
- Alkassab M. M and Qwaider O. Q (2010), A Comparison Between Unbiased Ridge and Least Square Regression Methods Using Simulation Technique, *Journal of Modern Applied Statistical Methods*, vol. **9**, (2), pp 488-494.
- Alkhamisi, M., Khalaf, S. and Shukur, G. (2006), Some Modifications for Choosing Ridge Parameters, *Communications in Statistics – Theory and Methods*, **35**, (11), 2005-2020.
- Alkhamisi M. and Shukur G (2007), Developing Ridge Parameters for SUR Models, *Communications in Statistics –Theory and Methods*, **37**, (4) , pp 544-564
- Ayres F. (1962), *Theory and Problems of Matrices*, Schaum Outline Series, McGraw Hill New York.
- Bata F.M., Ramanathan T. V. and Gore S. D (2008), The Efficiency of Modified Jackknife and Ridge –Type Regression Estimators, *A Comparison Surveys in Mathematics and its Applications*, **3**, 111-122.
- Belsley, David. A., Edwin. Kuh, and Welsch, Roy. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.*, John Wiley and Sons New York.
- Birkes D. and Dodge Y. (1993), *Alternative Methods of Regression* John Wiley and Sons, Canada.
- Brown, K. G. (1978), On Ridge Estimation in Rank Deficient Models, *Communications in Statistics: Theory and Methods*, **7**, (2), 187-192
- Cawley G. C. and Talbot L. C (2002), Reduced Rank Kernel Ridge Regression, *Neural Processing Letters*, **16**, 293-302
- Central Bank of Nigeria (2010), *CBN Statistical Bulletin*, Vol. 21, Central Bank of Ngeria.
- Callaghan, K. and Chen, J. (2008), Revisiting the Collinear Data Problem: An Assessment of Estimators “ill-Conditioning” in Linear Regression, *Practical Assessment, Research and Evaluation*, Vol.13, No 5
- Chatterjee, S. and Hadi, A. S. (2006), *Regression Analysis by Example*, Wiley – Interscience, Canada.
- Cheng B. and Wu X. (2006), A Modified Partial Least Square Regression (MPLSR) Method in Prediction, *Journal of Data Science*, **4**, 257-274.

- Crouse, R.H., Jin, C. and Hanumara, R. C., (1995), Unbiased Ridge Estimation with Prior Information and Ridge Trace, *Communications in Statistics – Theory and Methods*, Vol 24, Pp2341 – 2354.
- Cule E. and De Iorio M. (2013), Ridge Regression in Prediction Problems: Automatic Choice of the Ridge Parameter, *Genetic Epidemiology*, **37**, 704–714, Wiley.
- Dempster, A. P., Schatzoff, M., and Wermuth, N. (1977), A Simulation Study of Alternatives to Ordinary Least Squares, *Journal of the American Statistical Association*, **72**, 77-91.
- Deon R. B. and Dixon W. J. (1951), Simplified Statistics for Small Numbers of Observations, *Annal. Chem.* **23**, (4), 636 – 638.
- Doane D. P and Seward L. E (2011), Measuring Skewness: A Forgotten Statistic? *Journal of statistics Education*, **19** (2), 1-18.
- Dorugade, A. V. and Kashid, D. N. (2010), Alternative Methods for Choosing Ridge Parameter for Regression, *Applied Mathematical Sciences*, **4**, (9), 447-456.
- Dorugade A. V (2014), On Comparison of Some Ridge Parameters in Ridge Regression, *Journal of Applied Statistics*, **15**, (1), 33 – 45.
- Draper N. R. and Smith H. (1998), *Applied Regression Analysis*, Wiley International, New York.
- El-Dereny M and Rashwan N. I. (2011), Solving Multicollinearity Problem Using Ridge Regression Models, *International Journal of Contemporary Mathematical Sciences*, **6**, (12), 585 – 600.
- Farrar and Glauber (1967), Multicollinearity in Regression Analysis, *Review of Economics and Statistics*, **49**, 92-107.
- Farebrothers C. (1975), Principle Component Estimators and Minimum Mean Square Criteria in Regression Analysis, *Review of Econometrics and Statistics*, **54**, 332-356.
- Frisch, R. (1934), *Statistical Confluence Analysis by Means of Complete Regression System*, Publication **5**, (Oslo: University Institute of Economics).
- Gibbons D. G. (1981), A Simulation Study of Some Ridge Estimators, *Journal of the American Statistical Association*, **76**, 131-139.
- Goldstein, M. and Smith, A. F. M. (1974), Ridge-Type Estimators for Regression Analysis, *Journal of the Royal Statistical Society, Series B (Methodological)*, **36** (2), 284-291.

- Guilkey, D.K and Murphy J.L. (1975), Directed Ridge Regression Techniques in Cases of Multicollinearity, *Journal of American Statistical Association*, **70**, 767-775.
- Hocking, R. R., Speed, F. M. and Lynn, M. J. (1976), A Class of Biased Estimators in Linear Regression, *Technometrics*, **18**, (4), 425-437.
- Hoerl A. E. (1968), On Regression Analysis and Biased Estimation (Abstract), *Technometrics*, **10**, 422-433.
- Hoerl A. E. and Kennard, R. W. (1970a), Ridge regression: Biased Estimation for Non-Orthogonal Problems, *Technometrics*, **12**, 55 – 67.
- Hoerl A. E. and Kennard, R. W. (1970b), Ridge regression: An Applications for Non-orthogonal Problems, *Technometrics*, **12**, 69 – 82.
- Hoerl A. E and Kennard R. W. (1970c), Erratum to Ridge Regression: Application to Non-orthogonal Problems, *Technometrics*, **12**, 723.
- Hoerl, A. E. and Kennard, R. W. (1976), Ridge regression: Iterative Estimation of Biasing Parameter, *Communications in Statistics- Theory and Methods*, **5**, 77-88.
- Hoerl A. E., Kennard R. W. and Baldwin, K. F. (1975), Ridge Regression: Some Simulations, *Communications in Statistics, Simulation and Computation*, **4**, 105-123.
- Ishwaran H., Kogalur U. B. and Rao, J. S. (2010), Spikeslab: Prediction and Variable Selection using Spike and Slab Regression, *The R Journal* **2** (2) 68 – 73
- Izenman A. J. and Shen Y (2008), *Outlier Detection Using the Smallest Kernel Principal Components*, Research gate, Temple University, New York.
- Kalatzis A. E. G., Basseto C. F. and Azzoni C. R. (2008), Multicollinearity and Financial Constraint in Investment Decision: A Bayesian Ridge Regression, *Proceeding of International Conference on Applied Economics*, 437-447.
- Khalaf G. (2012a), A Proposed Ridge Parameter to Improve the Least Square Estimation, *Journal of Modern Applied Statistical Methods*, **11**, (2), 443-449.
- Khalaf G. (2012b), Improved Estimator in The Presence Of Multicollinearity, *Journal of Modern Applied Statistical Methods*, **11**, (1), 152-157.
- Khalaf G. and Shukur G. (2005), Choosing Ridge Regression Problems for Regression Problems, *Communications in Statistics – Theory and Methods*, **34**, 1177-1182.

- Kibria B. M. G (2003), Performance of some New Ridge Regression Estimators Communications in Statistics – Simulation and Computation, **32**, (2), 419-435.
- Lawless J. F. and Wang, P. (1976), A simulation Study of Ridge and Other Regression Estimators, Communications in Statistics A, **5**, 307 – 323.
- Levenberge K., (1944), A Method for the Solution of Certain Non-Linear Problems in Least Square, Quarterly Journal of Applied Mathematics, **2**, 164-168.
- Li, Y. and Yang, H. (2010), A New Stochastic Mixed Ridge Estimator in Linear Regression Model, Statistical Papers, **51**, 315-523.
- Lipovesky S. (2009), Multiple Regression in Pair Correlation Solution, Journal of Modern Applied Statistical Methods, **8**, (1), 22-131.
- Liptshutz S. and Lipson M. (2009), Linear Algebra, Schaum’s Outlines Series, 4<sup>th</sup> edition, Mc Graw Hill, USA.
- Liu, K. (2003), Using Liu-type Estimator to combat Collinearity, Communications in Statistics: Theory and Methods, **32**, 1009 – 1020.
- Mansson, K., Shukur, G. and Kibria, B. M. G. (2010), A Simulation Study of Some Ridge Regression Estimators under Different Distributional Assumptions, Communications in Statistics – Simulations and Computations, **39**, (8), 1639 -1670.
- Marquardt D. W. (1970), Generalized Inverses, Ridge Regression Biased Linear Estimation and Non-Linear Estimation, Technometrics, **12**, 591 – 612.
- Marquart , D. W and Snee, R. D. (1975), Ridge Regression in Practice, The American Statistical Association, **29**, 3-19.
- Mc Diarmid C. (1989), On the Method of Bounded Differences, Surveys in Combinatorics, Cambridge University Press, 148-188.
- McDonald G. C. and Galarneau D. I. (1975), A Monte Carlo Evaluation of Some Ridge-Type Estimators, Journal of the American Statistical Association, **70**, 407 – 416.
- Montgomery D. C. and Peck E. A. (1992), Introduction to Linear Regression Analysis, John Wiley and Sons, NewYork.
- Muniz, G. and Kibria, B. M. G. (2009), On Some Ridge Regression Estimators: An Empirical Comparison, Communications in Statistics- Simulations and Computation, **38**, 621 – 630.
- Muniz, G., Kibria, B. M. G. and Shukur, G. (2010), On Developing Ridge Regression Parameters: A Graphical Investigation, Working Paper.

- Myers R. H. (1990), Classical and Modern Regression with Applications, 2<sup>nd</sup> Edition, PWS-KENT Publishers, California, USA.
- Nagaraj, R. and Thiagarasu, V. (2014), Collinearity Similarity Measure Based on Document Clustering with Directed Ridge Regression, Indian Journal of Science and Technology, **7**, (5), 692 – 697.
- Neter J, Kutner M. A. Nachtshein and Wasserman W. (1996), Applied Linear Statistical Models, Mac-Graw Hill, USA.
- Newhouse, J. P. and Oman, S. D. (1971), An Evaluation of Ridge Estimators, Rand Corporation, **15**, (1), P-716-PR.
- Nja M.E (2013), A Proposed Nth-order Jackknife Ridge Estimator for Linear Regression Design, Journal of Mathematical Theory and Modelling, **3**, (8), 2224 – 2258.
- Nduka E.C (1999), Principles of Applied Statistics – Regression and Correlation Analysis, Cristal Publishers, Okigwe.
- Nordberg, L. (1982), A Procedure for Determination of A Good Ridge Parameter in Linear Regression, Communications in Statistics, A, **(11)**, 285-309.
- Obenchain, R., L. (1975), Ridge Analysis Following a Preliminary Test of the Shrunken Hypothesis, Technometrics, **17**, 431 – 441.
- Pasha, G. R. and Ali Sha M. A. (2004), Application of Ridge Regression to Multicollinear Data, Journal of Research (Science), Bahauddin Zakariya University, Multan Pakistan, 97-106
- Perperoglou A. (2013), Cox Model with Dynamic Ridge Penalties on Time-Varying Effect of the Covariates, Statistics in Medicine, Wiley.
- Rao C.R (1973), Linear Statistical Inference and its Applications, John Wiley, New York.

- Rao C. R and Toutenberg H. (1995), Linear Models: Least Square and Alternatives, Springer-Verlag, New York.
- Rawlings, J. O., Pantula, S. G. and Dickey, D. A. (1988), Applied Regression Analysis: A Research Tool, Springer –Verlag, New York 2<sup>nd</sup> Edition.
- Rorabacher, D. B. (1991), Statistical Treatment of Rejection of Deviant Values: Critical Values of Dixon Q Parameter and Related Subrange Ratios at The 95% Confidence Level, *Annal. Chem.*, **63**, (2), 139 – 146.
- Ross S. M (2009), Introduction to Probability and Statistics for Engineers and Scientists, 4<sup>th</sup> Ed., Elsevier Academic Press USA.
- Ryan (1997), Modern Regression Methods, Wiley International, New York.
- Saleh, A. K. and Kibria, B. M. G. (1993), Performance of Some New Preliminary Test Ridge Regression Estimators and Their Properties, *Communications in Statistics- Theory and Methods*, **22**, 2747 – 2764.
- Shen, X., Alam M., Fikse F. and Ronnegard L. (2013), A Novel Generalized Ridge Regression Method for Quantitative Genetics, *Genetics*, **193**, (4), 1255-1268.
- Shi L. and Wang, X. (1999), Local Influence in Ridge Regression, *Computation Statistics and Data Analysis*, **31**, 341-353.
- Sinan A. and Genc A. (2012), Comparing the Most Commonly used Classical Methods for Determining the Ridge Parameter in Ridge Regression, *Journal of Selcuk University Natural and Applied Sciences*, **1**, (2), 13 – 24.
- Spiegel L. and Stephen M. L (2008), Linear Algebra, Schaum Outline Series, USA.
- Stuart, A. and Ord, A. (1994) Kendall's Advanced Theory of Statistics, Vol. 1, 6<sup>th</sup> edition, Edward Arnold, London.
- Szekely, G. J and Mori, T. F. (2001), A Characteristic Measure of Asymmetry and its Application for Testing Diagonal Symmetry, *Communications in Statistics – Theory and Methods*, **30**,(8), 1633-1639.

- von Hippel P. T. (2005), Mean, Median and Skew: Correcting a Textbook Rule, Journal of Statistics Education, **13**, (2), 136-141.
- Wethrill, F. G (1986), Evaluation of Ordinary Ridge Regression, Bulletin of Mathematical Statistics, **18**, 1-35.
- Wu J. and Liu C. (2014), Performance of Some Stochastic Restricted Ridge Regression in Linear Regression Model, Journal of Applied Mathematics, [Http://dx.doi.org/10.1155/2014/508793](http://dx.doi.org/10.1155/2014/508793).
- Yan X. (2008), Modified Non-Linear Generalized Ridge Regression and its Application to Develop Naphtha Cut-off Point Soft Sensor, Journal of Computer and Chemical Engineering, **32**, 608 – 621.
- Yan X. and Su X. (2009), Linear Regrssion Analysis: Theory and Computing, World Scientific Publishing Coy Ltd, USA.
- Zang J. and Ibrahim, M. (2005), A simulation Study on SPSS Ridge Regression and Ordinary Least Square Regression Procedures for Multicollinearity Data, Journal of Applied Statistics, **32**, 571- 586.

## APPENDIX I

### COMPUTER PROGRAMME IN R FOR THE MONTE CARLO SIMULATION STUDY

```
n=c(10,30,50,80)
p=c(3,4,5,6,7,8,9)
sigma=c(0.8,0.9,1.0)

rho=c(0.3,0.5,0.8,0.9,0.99)
for (a in 1:1)
{
n1=n[a]
for (b in 1:1)
{
p1=p[b]
for (c in 1:3)
{
sigma1=sigma[c]
output4=matrix(0,5,5)
for (d in 1:5)
{
rho1=rho[d]
R=2000
Beta_k=matrix(0,R,p1);Beta=matrix(0,R,p1)
PRESS=0; MSE=0; IBeta=matrix(0,R,p1)
for (r in 1:R)
{
#To generate the explanatory variables
x=matrix(0,n1,p1)
z=matrix(0,n1,p1)
for (i in 1:n1){
z[i,]=rnorm(p1)
x[i,]=c(sqrt(1-rho1^2)*z[i,]+rho1*z[i,])
}
x #To view the generated matrix
#Centre and Scale
L=0
for (j in 1:p1) L[j]= sqrt(sum((x[,j]-mean(x[,j]))^2))
X=matrix(0,n1,p1)
for (j in 1:p1) X[,j]=(x[,j]-mean(x[,j]))/L[j]
X
D=t(X)%*%X
D
```



```

E=eigen(D)
Q=E$vectors
Lambda=E$values
Beta=E$vectors[,1]
Beta
#t(Beta)%*%Beta
#To generate the dependent variable
y=0
for (i in 1:n1) y[i]=sum(Beta*x[i,])+rnorm(1,mean=0,sd=sigma1)
y
#Centre and Scale y
Ly=sqrt(sum((y-mean(y))^2))
Y=(y-mean(y))/Ly
Y
#Fit the regression model
reg=lm(Y~X-1)
#Obtain the alpha_j's
bbeta=reg$coef
alpha=t(Q)%*%bbeta
sigma_hat=sigma1 #Use other values later
q=0
for (j in 1:p1)
q[j]=(max(Lambda)*(sigma1^2))/((n1-
p1)*(sigma1^2)+max(Lambda)*(alpha[j]^2))
#Formula 6
k=max(q)
#Obtain the ridge regression coefficients
I=diag(p1)
IBeta[r,]=c((solve(t(X)%*%X+k*I))%*%(t(X)%*%Y))
#Obtain the MSE
u=uu=0
for (f in 1:p1){ u[f]=Lambda[f]/((Lambda[f]+k)^2)
uu[f]=(alpha[f]^2)/((Lambda[f]+k)^2)
}
MSE[r]=c((sigma_hat^2)*sum(u)+(k^2)*sum(uu))
#To obtain PRESS
Hat=X%*%solve(t(X)%*%X+k*I)%*%t(X)
Y_hat=X%*%IBeta[r,]
e=Y-Y_hat
ee=0
for(i in 1:n1)
{
ee[i]=e[i]/(1-Hat[i,i])
}

```

```

}
PRESS[r]=c(sum(ee^2))
}
#For Summary
cc=mean(PRESS)
output4[d,1:5]=c(round(cc,4),n1,p1,sigma1,rho1)
}#rho
print(output4)
}#sigma
}#p
}#n
#At the end, do in R console, File -> Save to file -> Documents -> Filename.txt
#To save the file as a .txt file. When you have done this, delete all the text in the
#output and paste the data in a spreadsheet to make calculations easier.
#####
#####
THIS PART IS FOR FORMULARS FOR CALCULATING K
#####
#####
k=0
#Formula 1
k=(sigma1^2)/max(alpha^2)
#Formula 2
k=(p1*(sigma1^2))/(t(alpha)%*%alpha)
k=c(k)
#Formula 3
w=0
for (j in 1:p1) w[j]=Lambda[j]*alpha[j]
k=(p1*(sigma1^2))/sum(w)
#Formula 4
k=(max(Lambda)*(sigma1^2))/((n1-
p1)*(sigma1^2)+max(Lambda)*max(alpha^2))
q=0
for (j in 1:p1)
q[j]=(max(Lambda)*(sigma1^2))/((n1-
p1)*(sigma1^2)+max(Lambda)*(alpha[j]^2))
#Formula 5
k=max(1/q)
#Formula 6
k=max(q)
#Formula 7
k=prod(q)^(1/p1)
#Formula 8

```

```

k=median(1/q)
#Formula 9
k=median(q)
w=0
for (j in 1:p1) w[j]=(log(2)*(sigma1^2))/((n1-
p1)*(sigma1^2)+log(2)*(alpha[j]^2))
#Formula 10
k=max(1/w)
#Formula 11
k=max(w)
#Formula 12
k=prod(w)^(1/p1)
#Formula 13
k=prod(1/w)^(1/p1)
#Formula 14
k=median(1/w)
#Formula 15
k=median(w)
lambdag=prod(Lambda)^(1/p1)
v=0
for (j in 1:p1) v[j]=(lambdag*(sigma1^2))/((n1-
p1)*(sigma1^2)+lambdag*(alpha[j]^2))
#Formula 16
k=max(1/v)
#Formula 17
k=max(v)
#Formula 18
k=prod(1/v)^(1/p1)
#Formula 19
k=prod(v)^(1/p1)
#Formula 20
k=median(1/v)
#Formula 21
k=median(v)

```



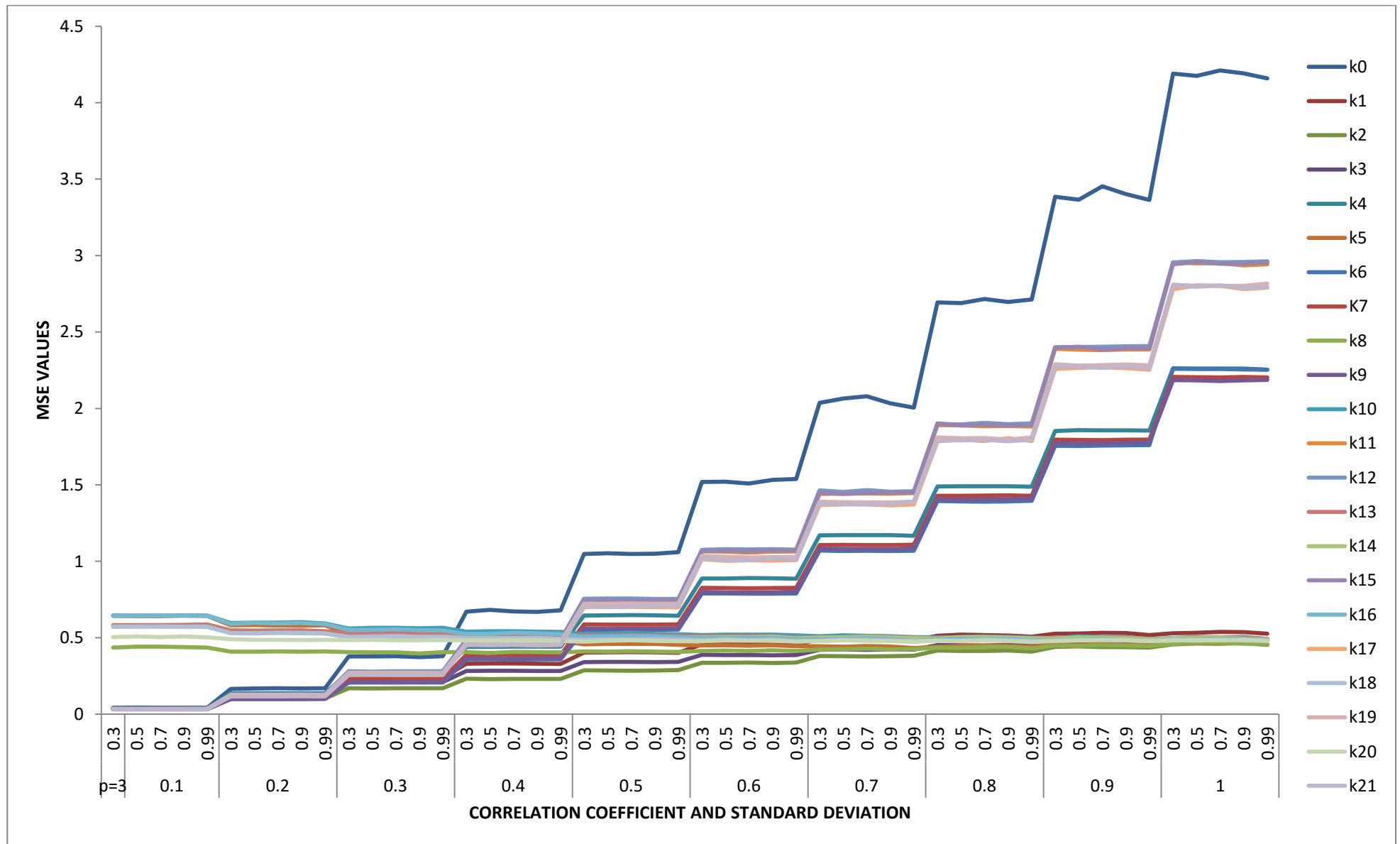


Figure 4.1 Graph of Simulated values of MSE for  $n=10$ ,  $p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.1(b) Simulated MSE values for  $n=10$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



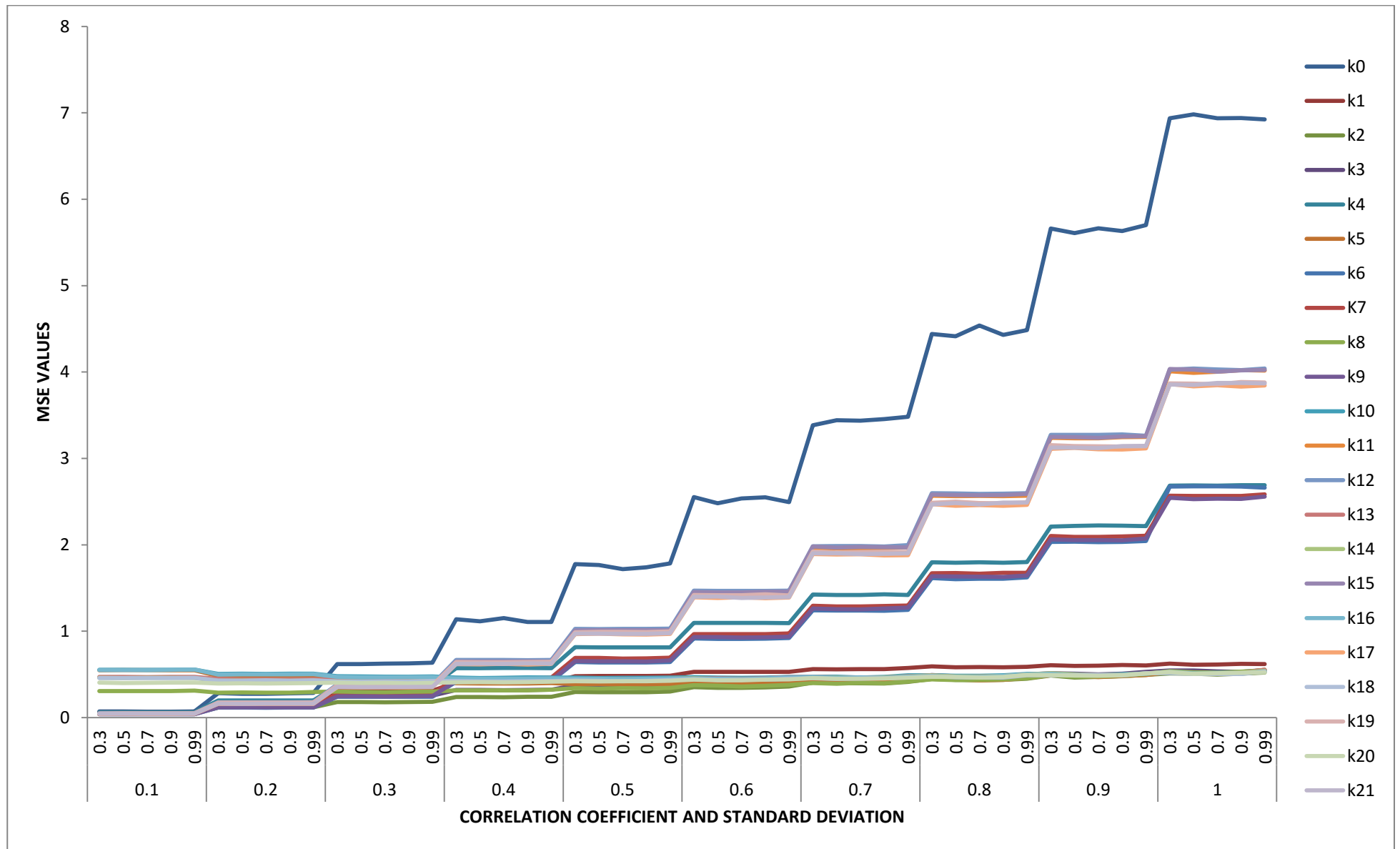


Figure 4.2 Graph of Simulated values of MSE for  $n=10, p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.1(c) Simulated MSE values for  $n=10, p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





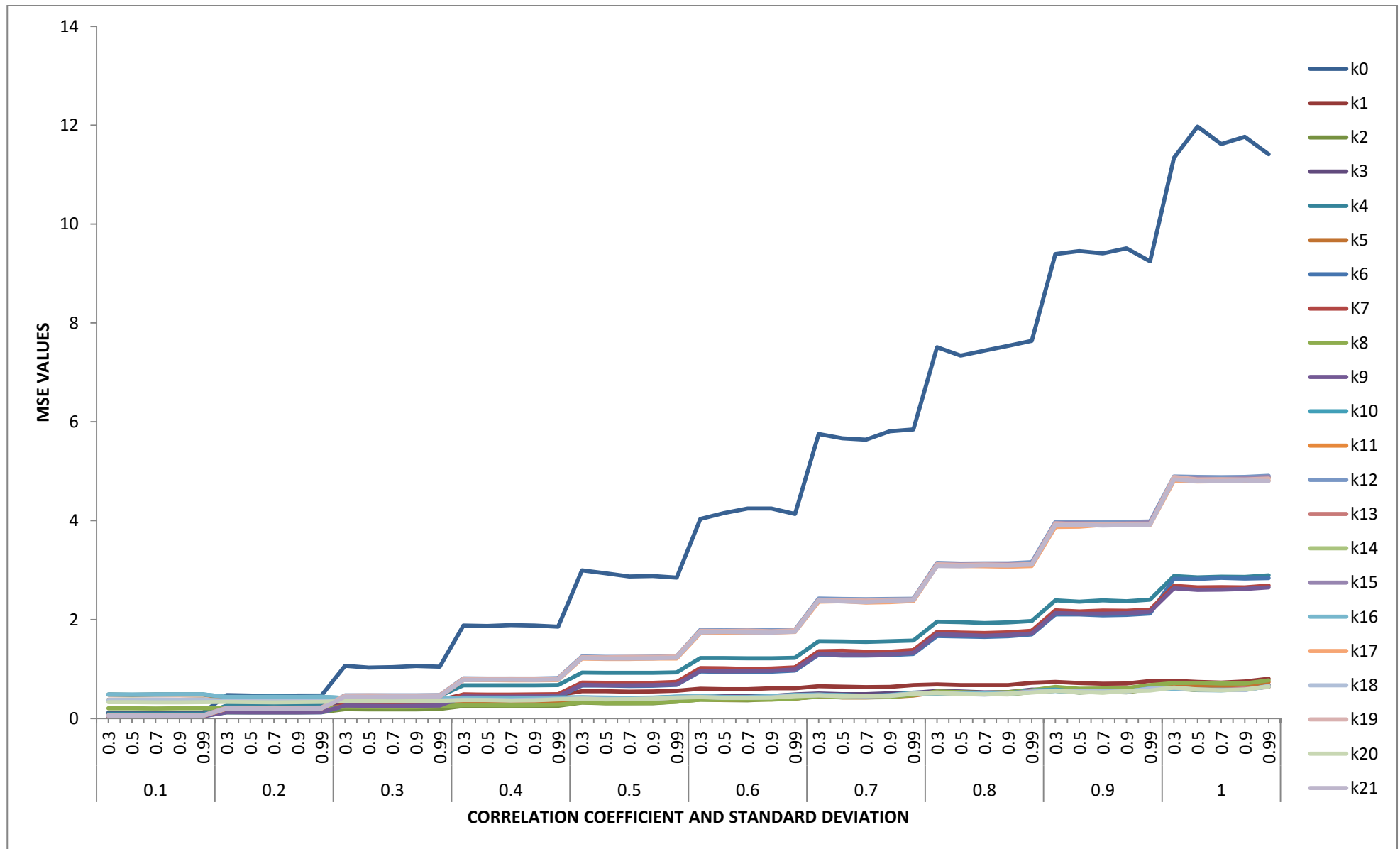


Figure 4.3 Graph of Simulated values of MSE for  $n=10, p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.1(d) Simulated MSE values for  $n=10, p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



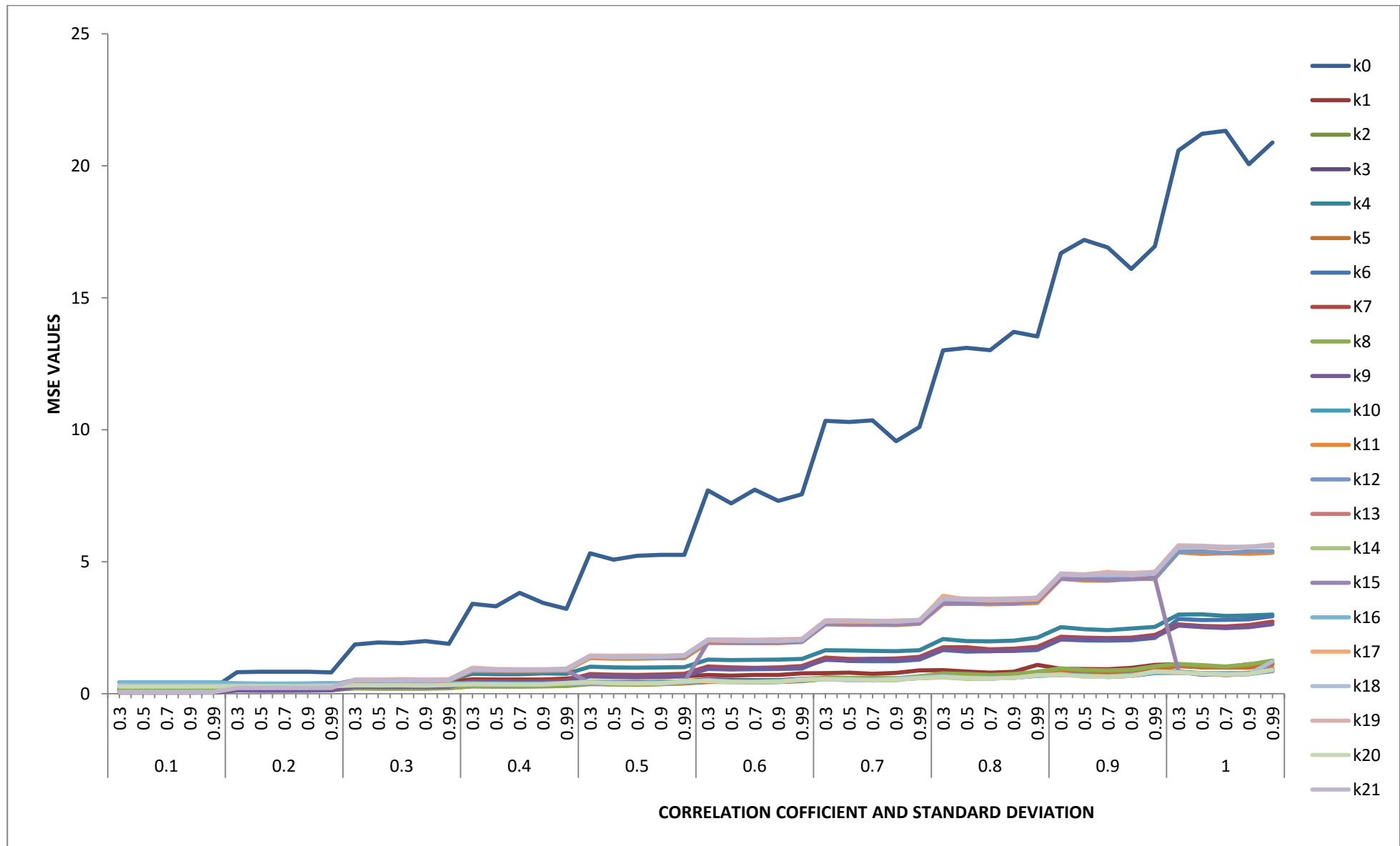


Figure 4.4 Graph of Simulated values of MSE for  $n=10, p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.1(e) Simulated MSE values for  $n=10, p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



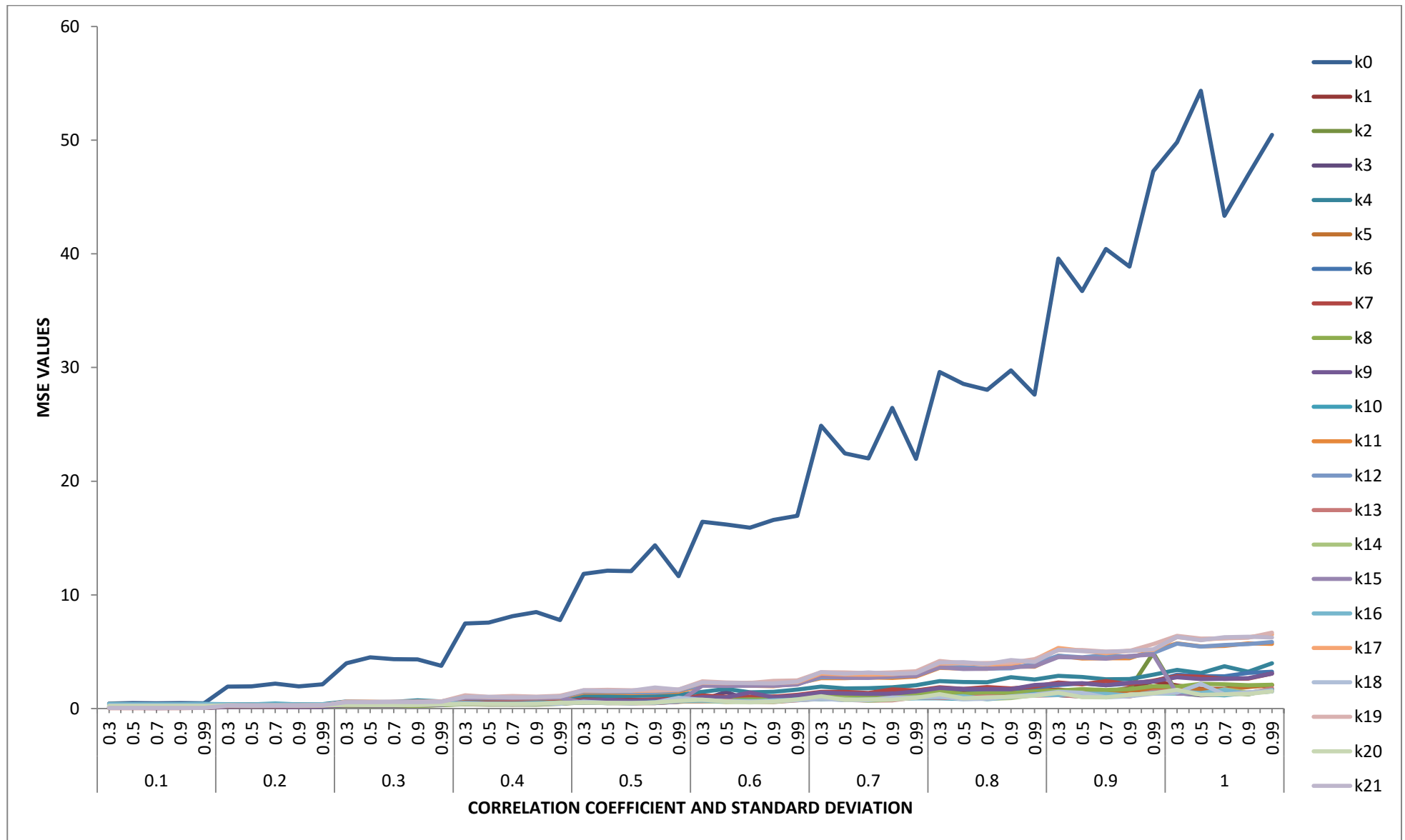


Figure 4.5 Graph of Simulated values of MSE for  $n=10$ ,  $p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.1(f) Simulated MSE values for  $n=10$ ,  $p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



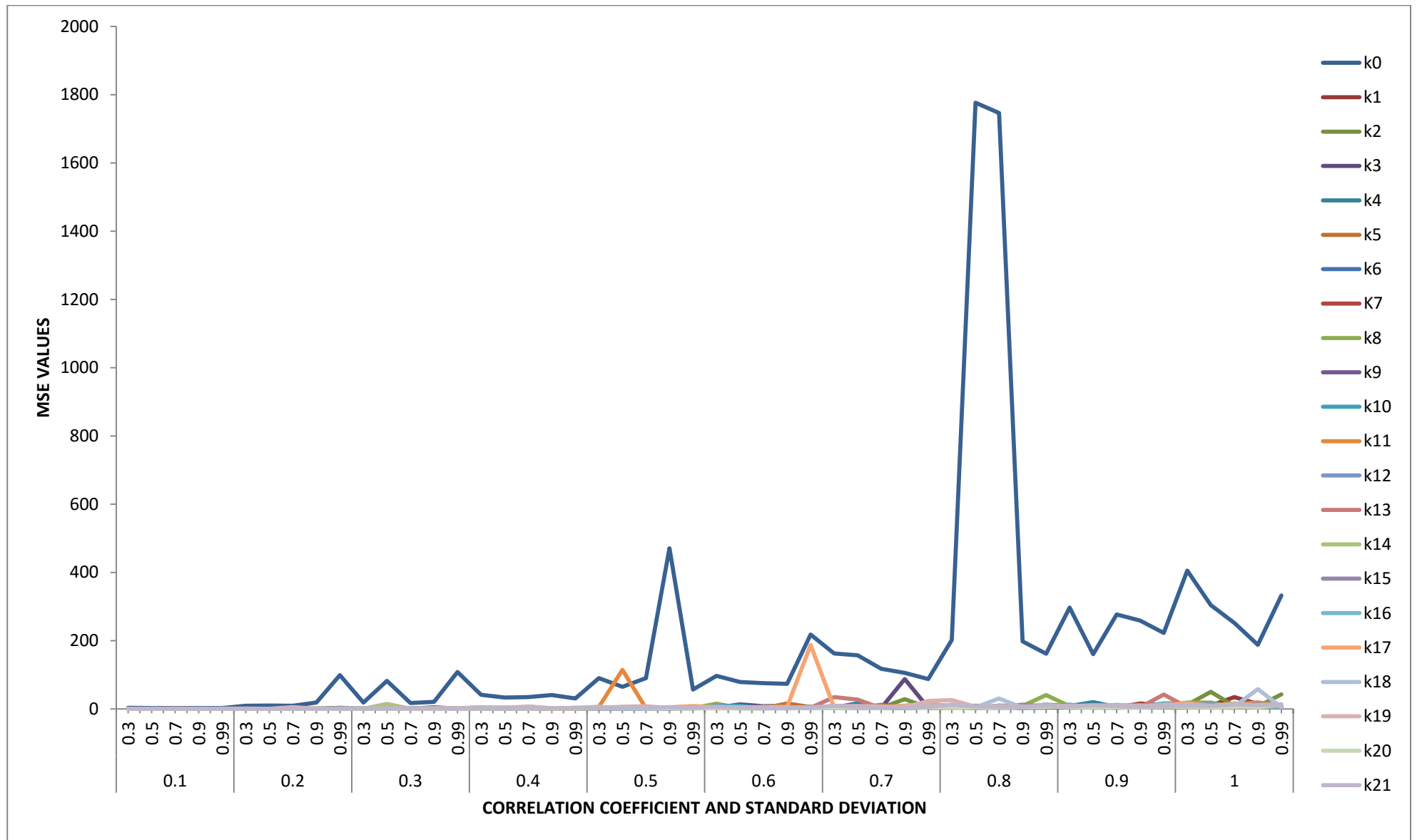


Figure 4.6 Graph of Simulated values of MSE for  $n=10, p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.1(g) Simulated MSE values for  $n=10, p=9$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





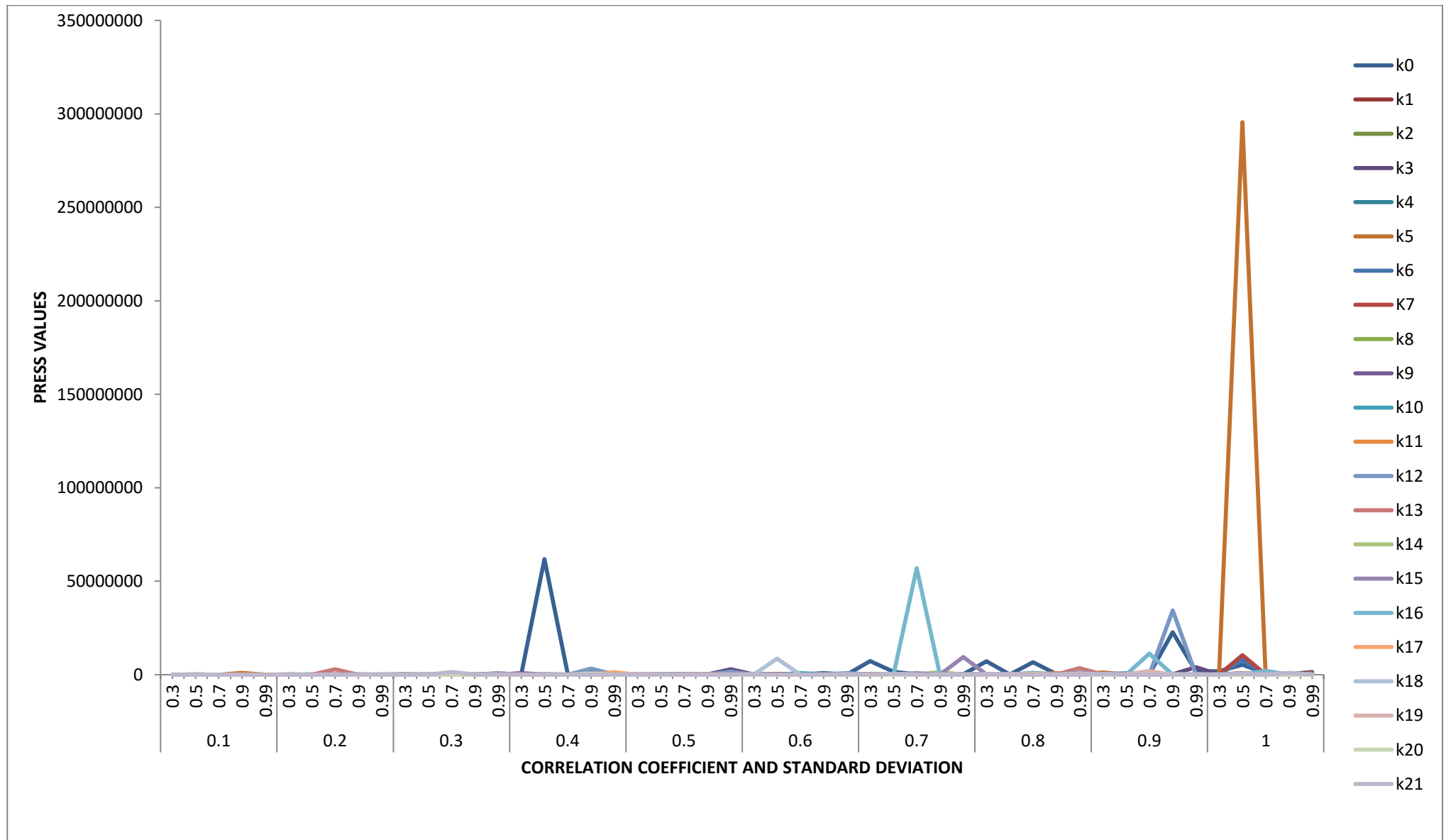


Figure 4.7 Graph of Simulated values of MSE for  $n=10, p=9$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



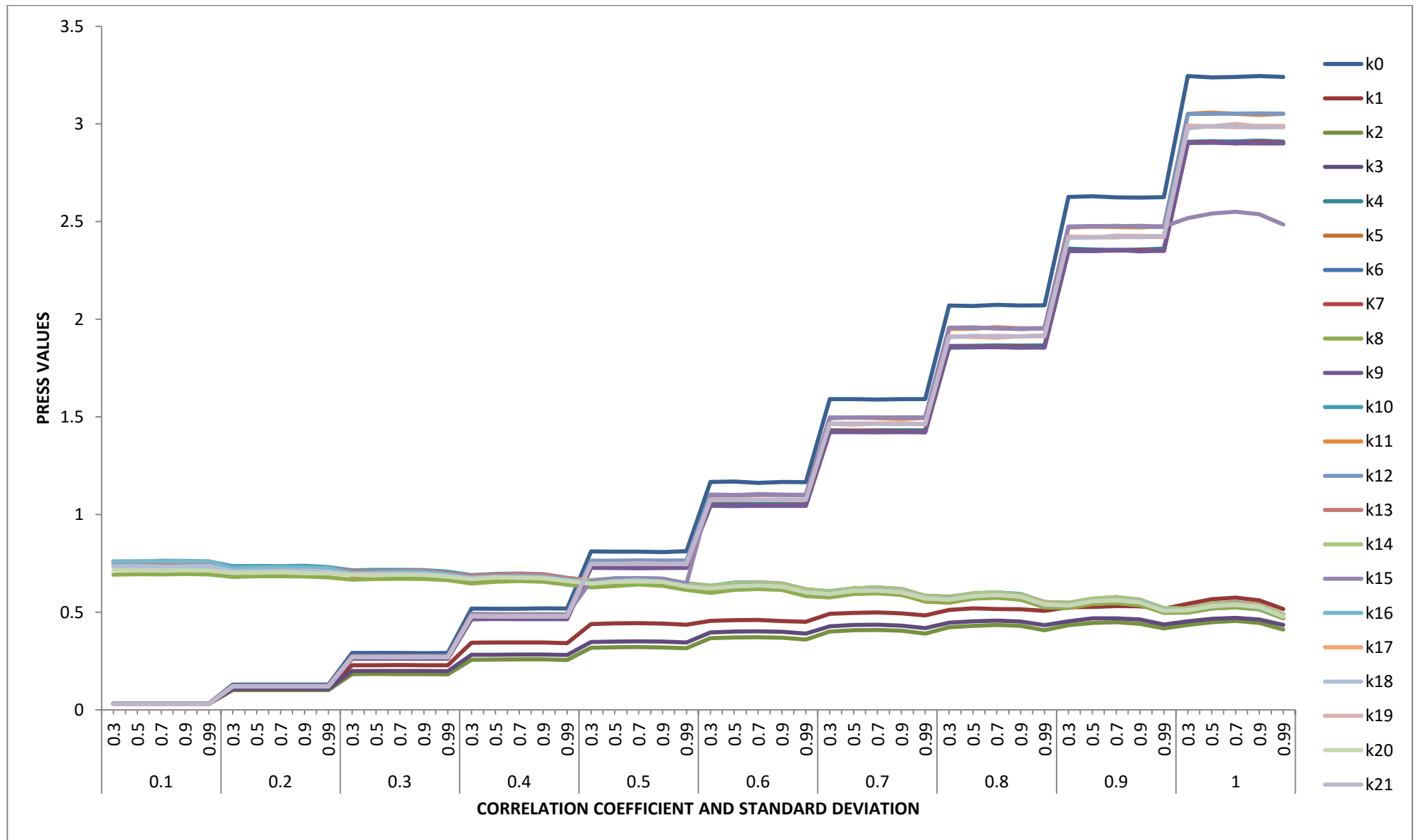


Figure 4.8 Graph of Simulated values of MSE for  $n=30, p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



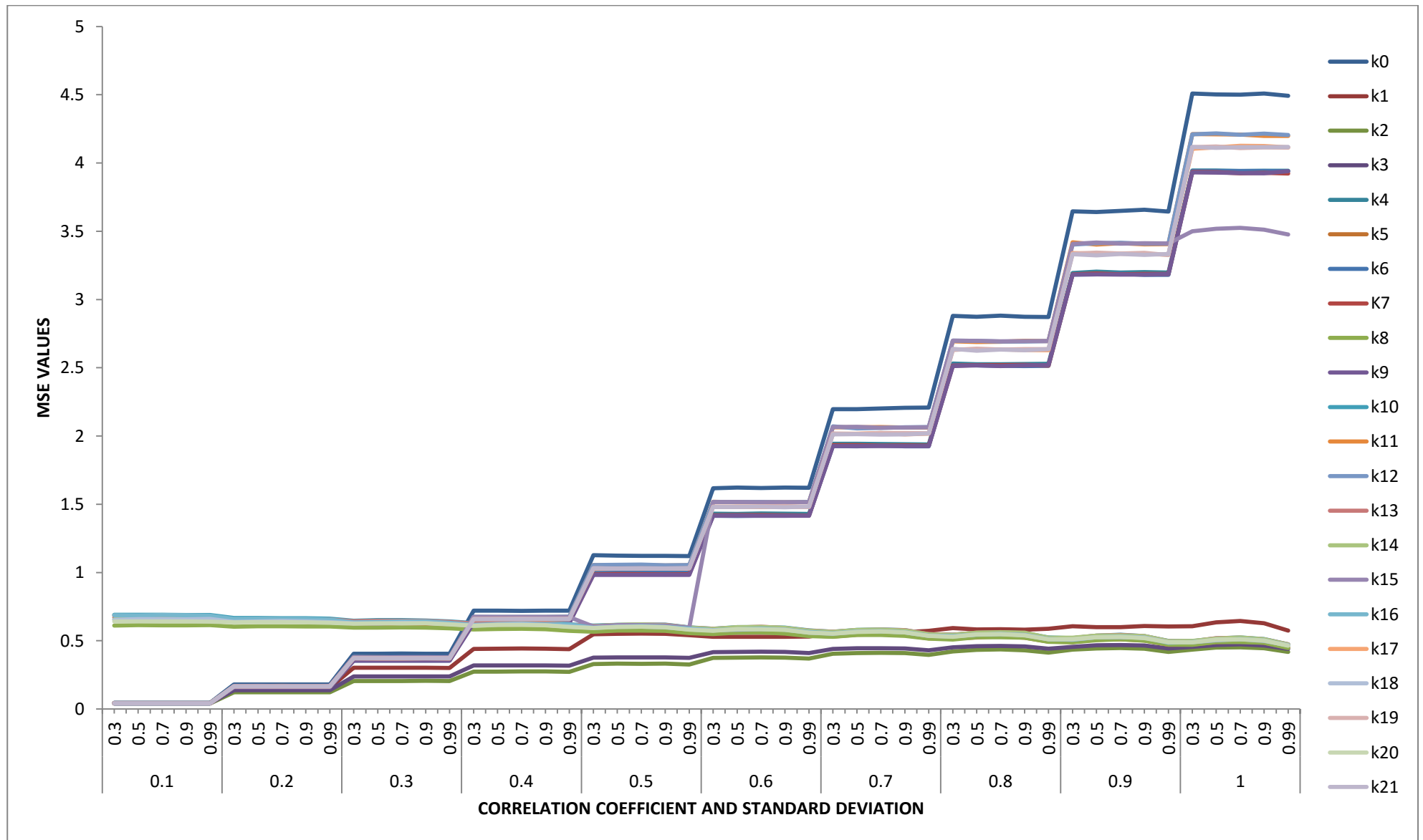


Figure 4.9 Graph of Simulated values of MSE for  $n=30, p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



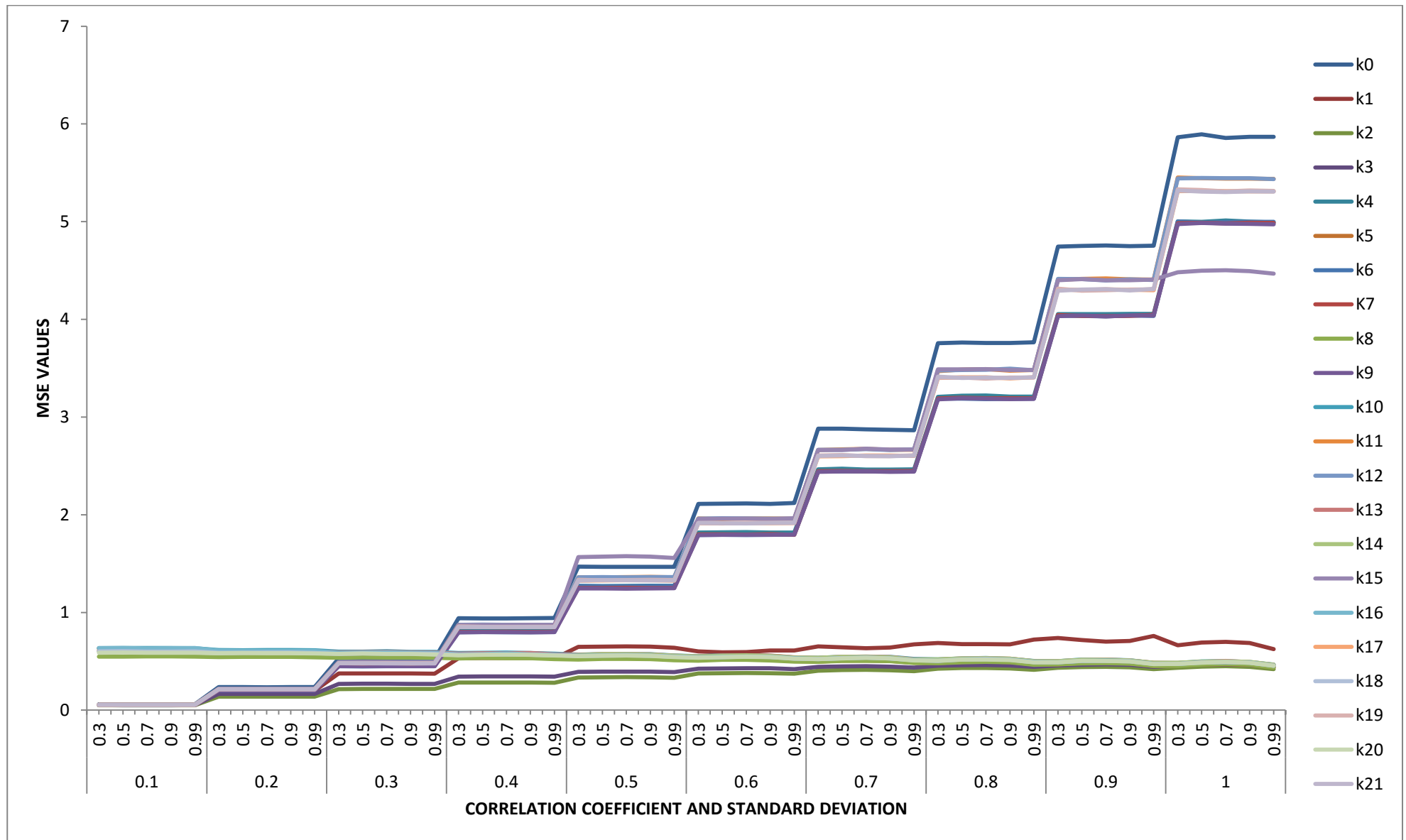


Figure 4.10 Graph of Simulated values of MSE for  $n=30, p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





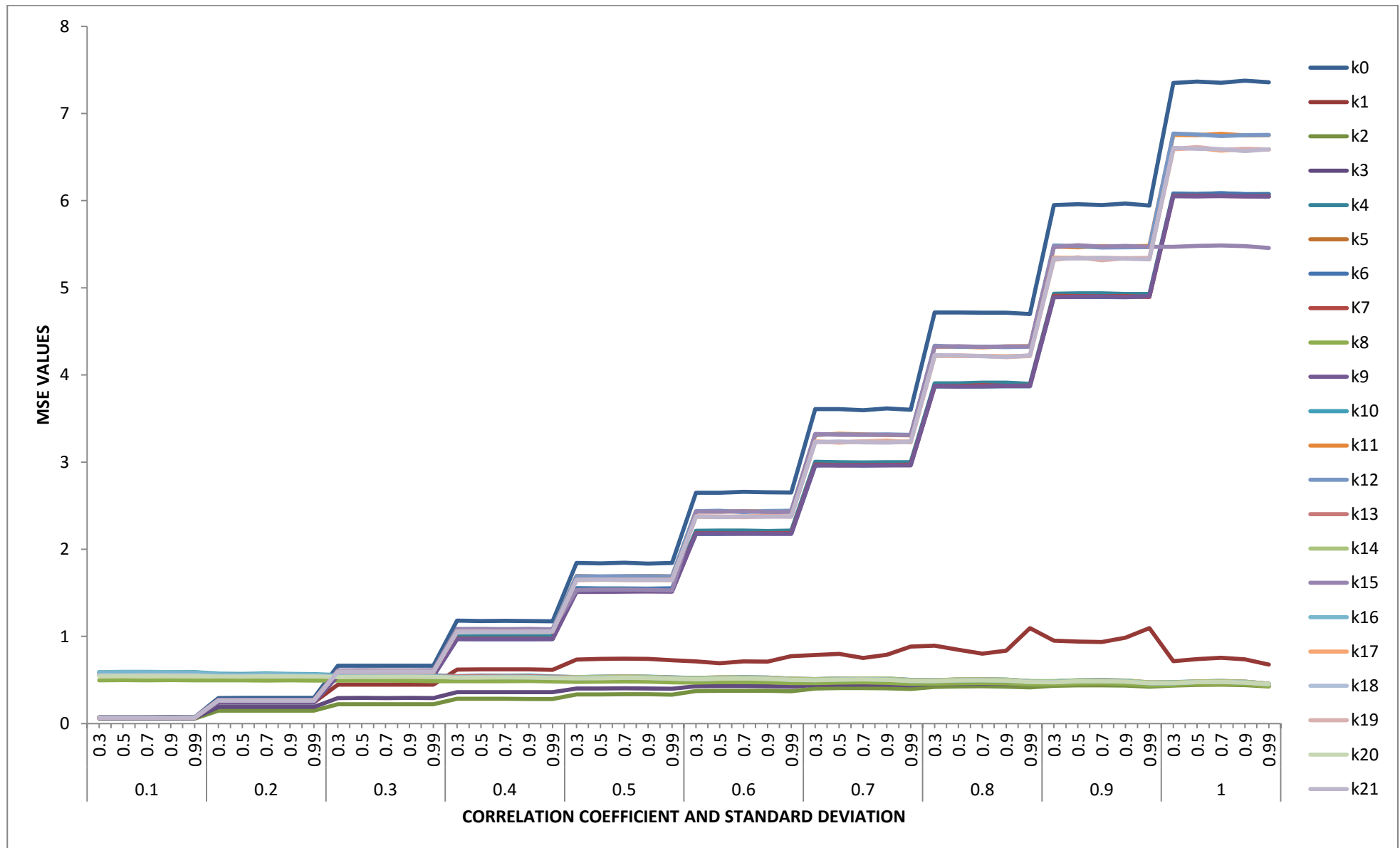


Figure 4.11 Graph of Simulated values of MSE for  $n=30$ ,  $p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



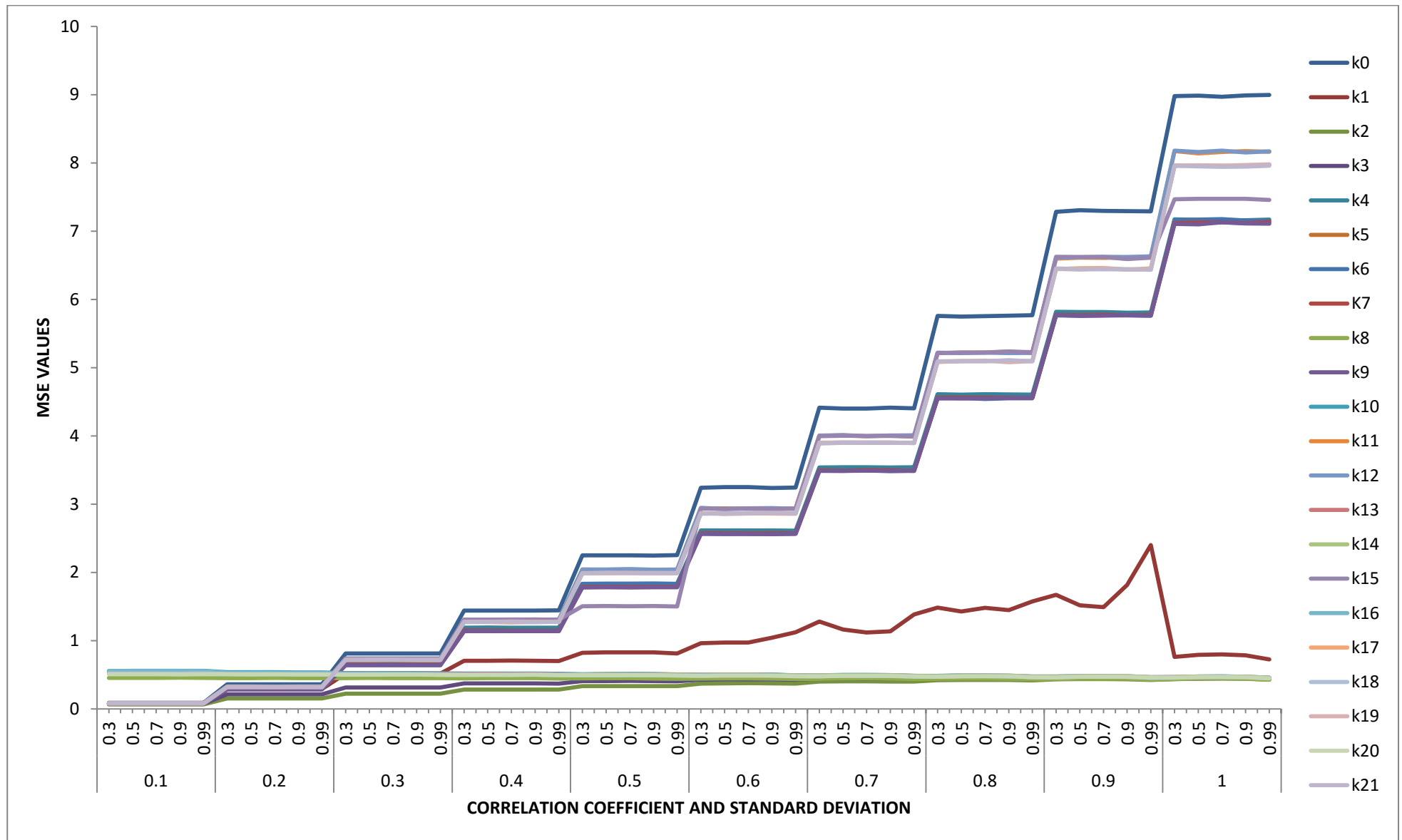


Figure 4.12 Graph of Simulated values of MSE for  $n=30$ ,  $p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



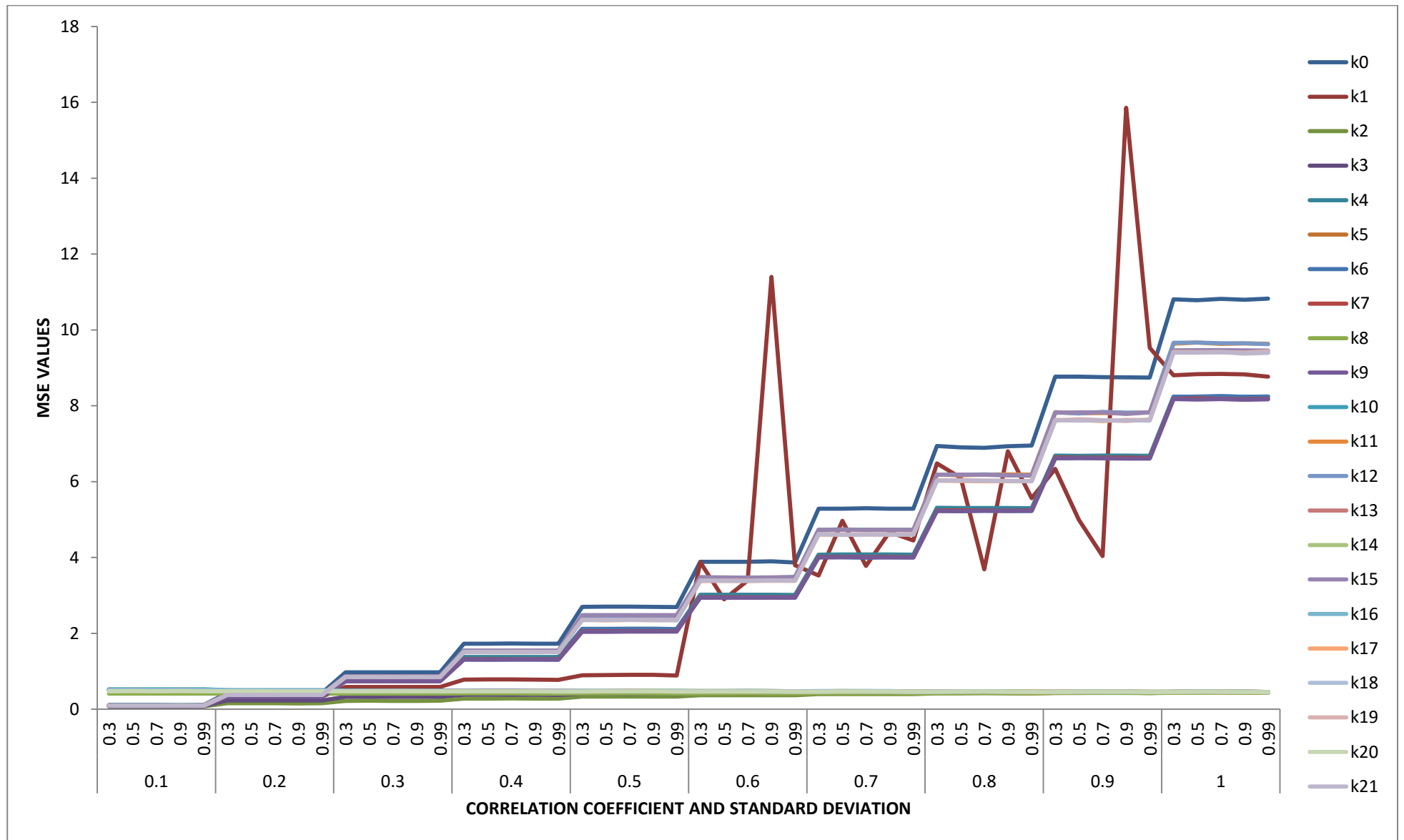
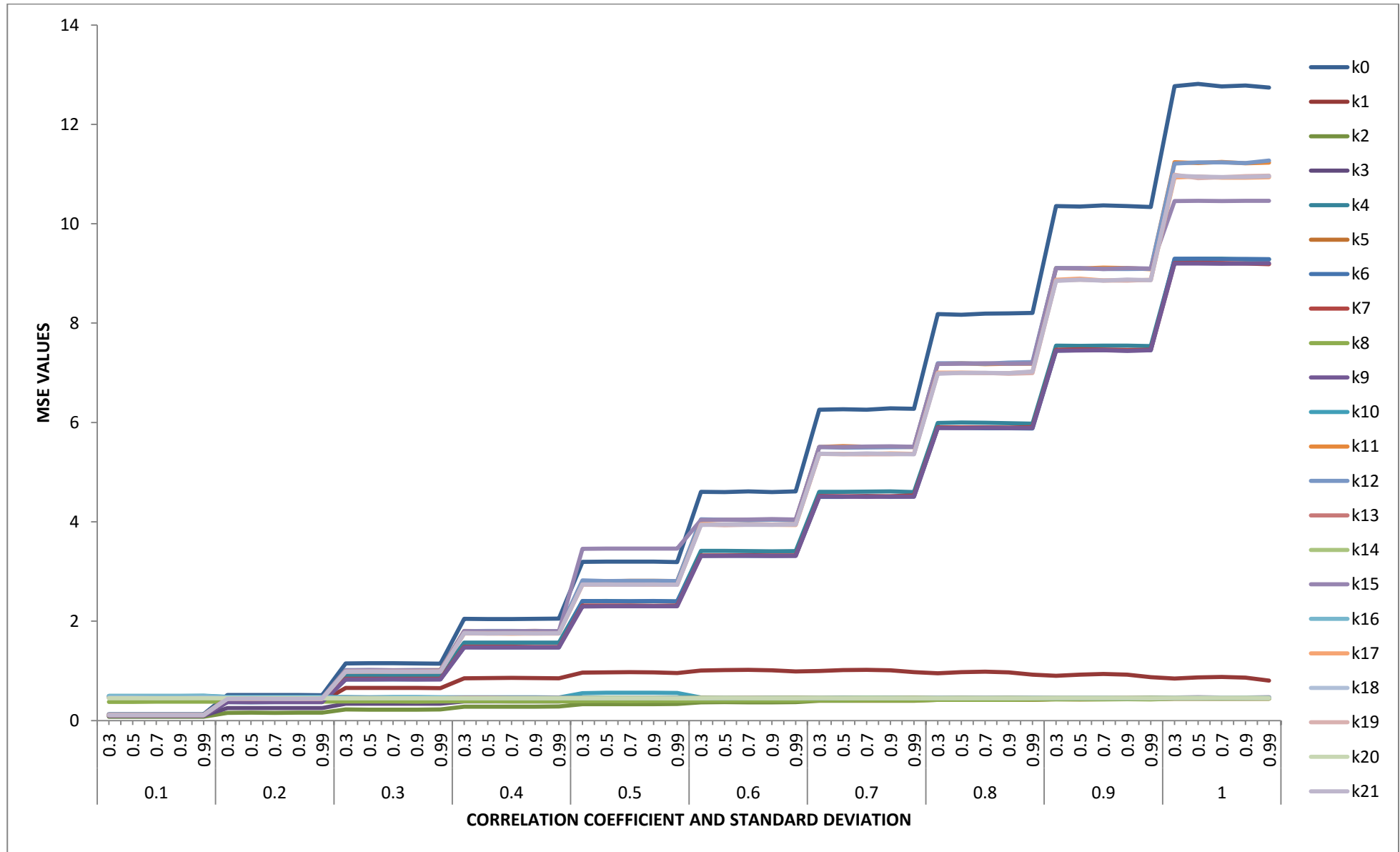


Figure 4.13 Graph of Simulated values of MSE for  $n=30, p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





**Figure 4.14** Graph of Simulated values of MSE for  $n=30$ ,  $p=9$   $\sigma = 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





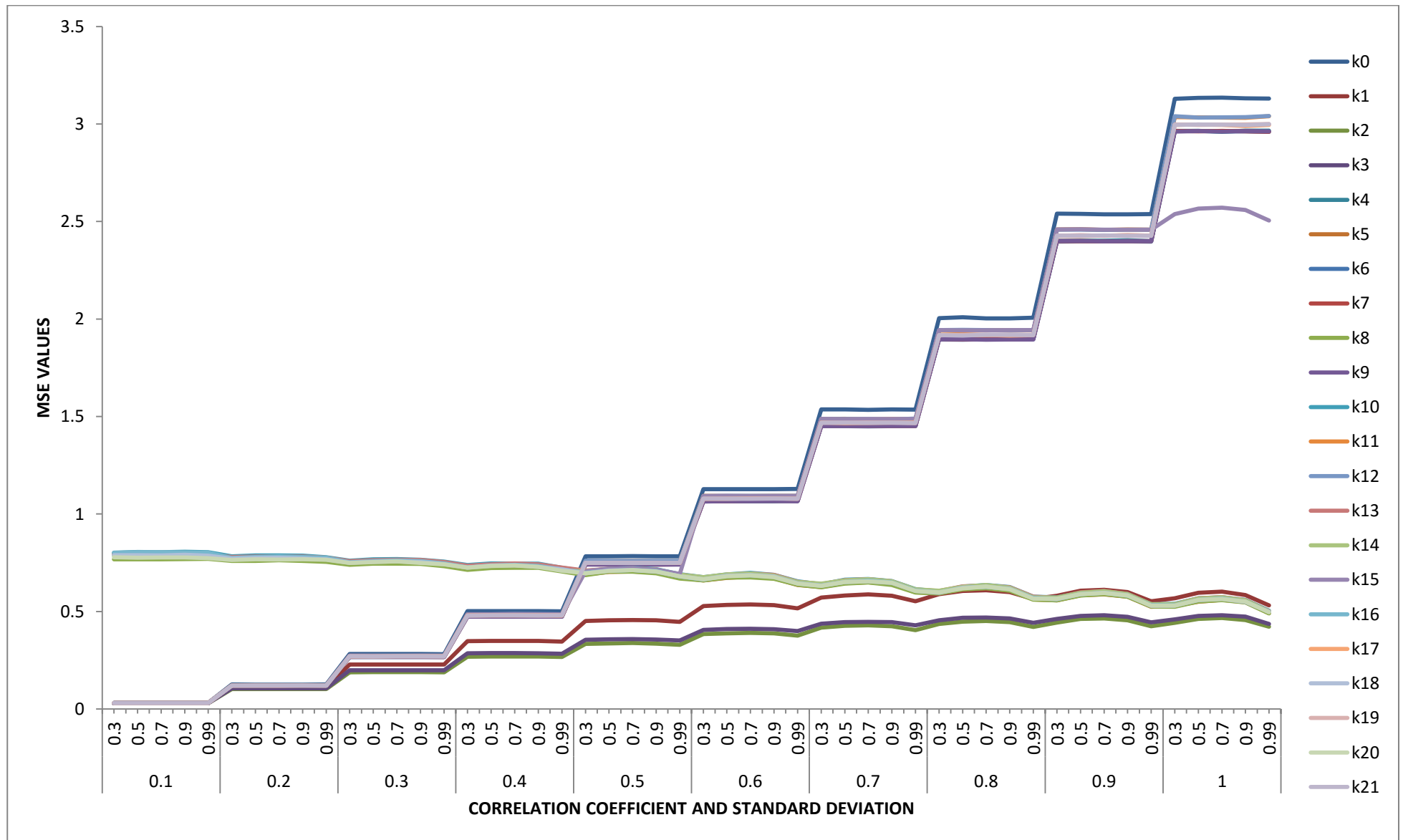


Figure 4.15 Graph of Simulated values of MSE for  $n=50, p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



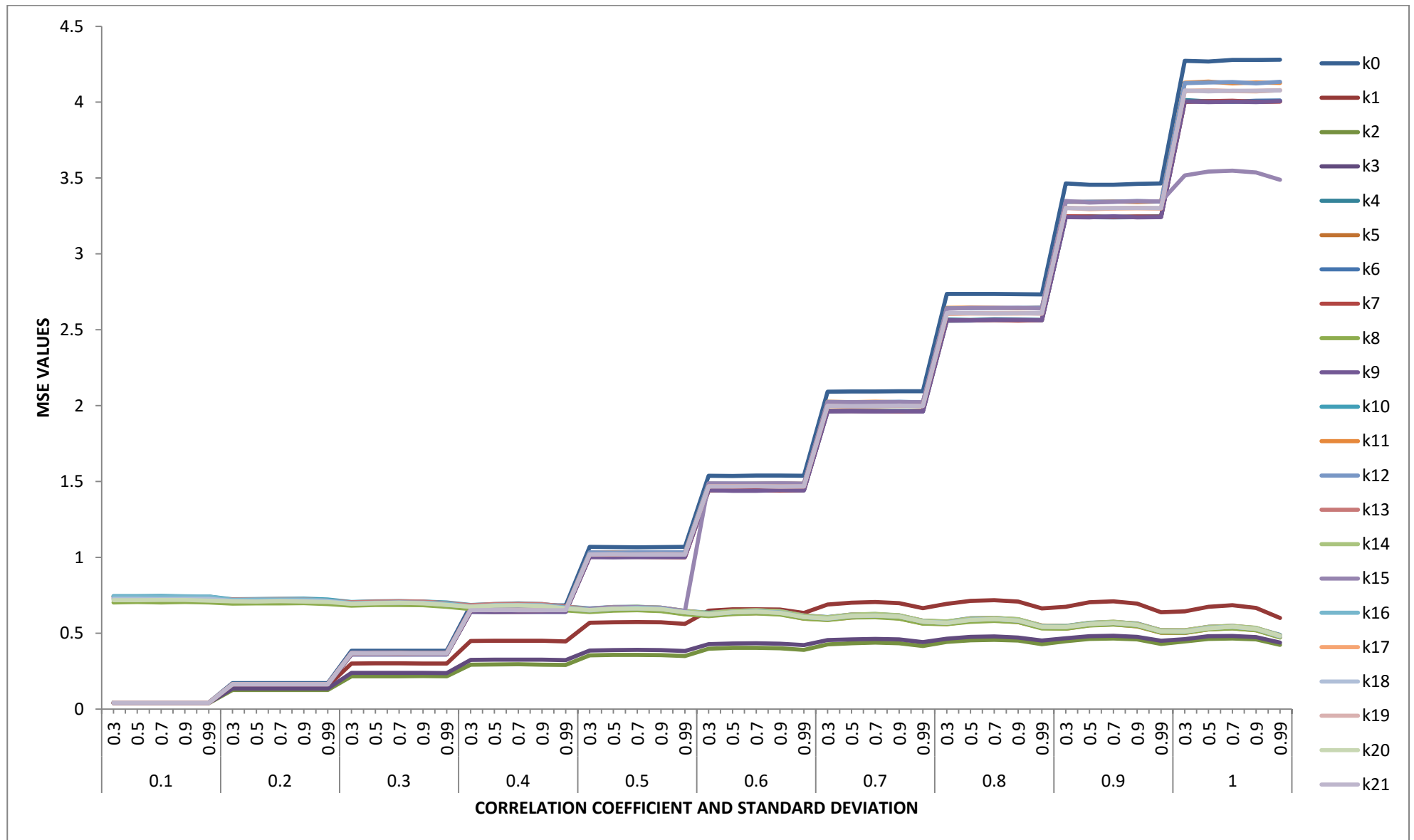


Figure 4.16 Graph of Simulated values of MSE for  $n=50, p=4 \sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



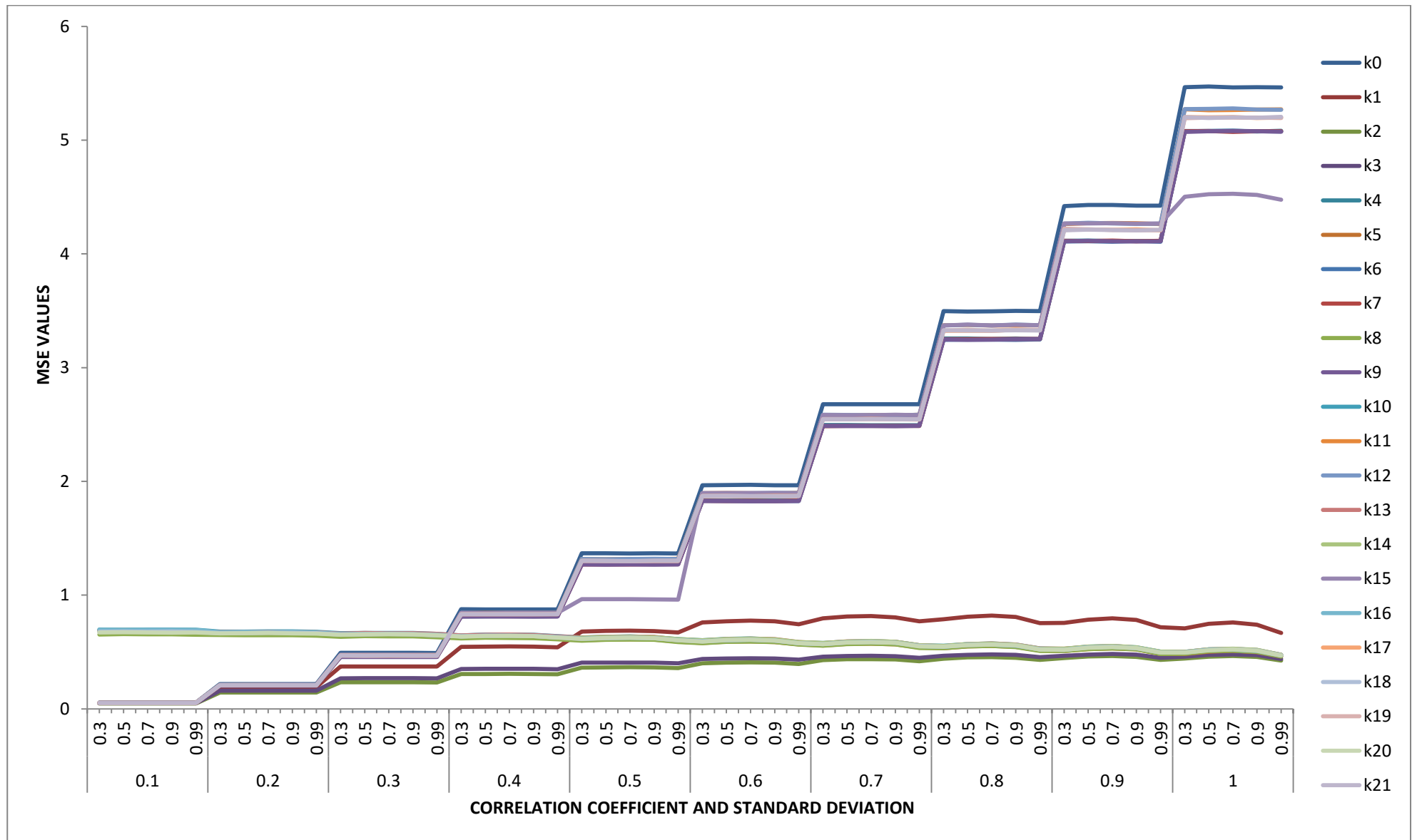


Figure 4.17 Graph of Simulated values of MSE for  $n=50$ ,  $p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



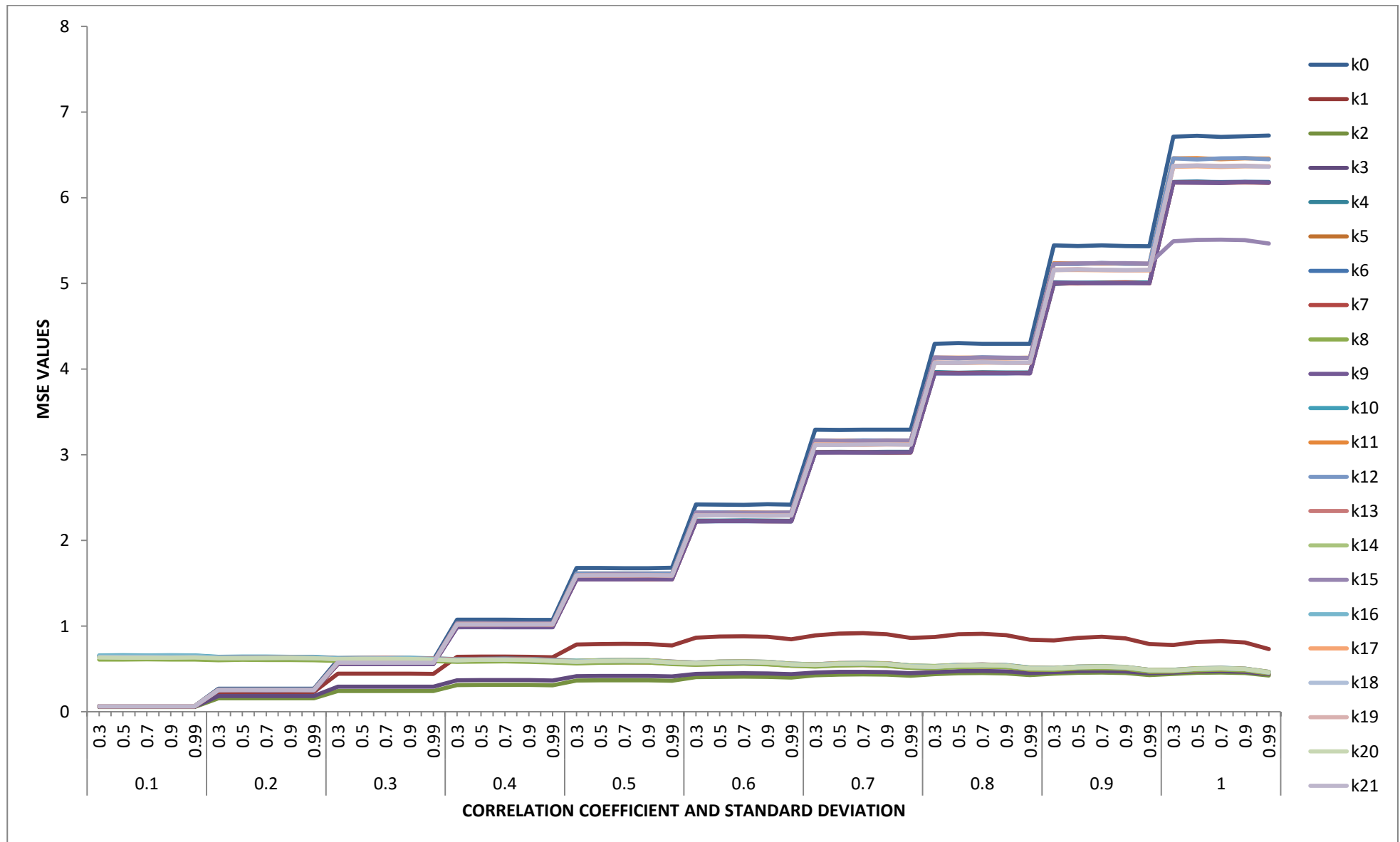


Figure 4.18 Graph of Simulated values of MSE for  $n=50$ ,  $p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





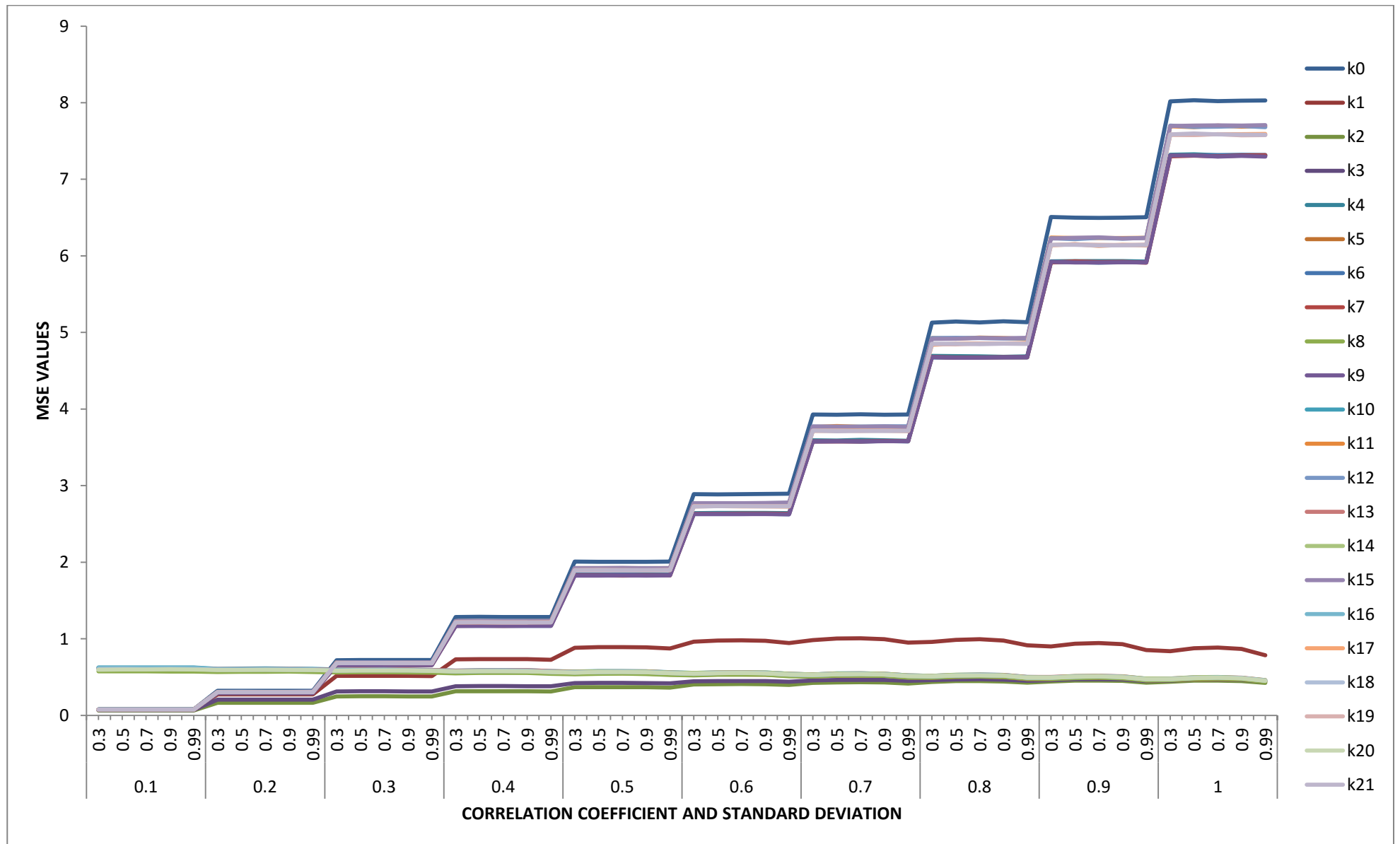


Figure 4.19 Graph of Simulated values of MSE for  $n=50$ ,  $p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



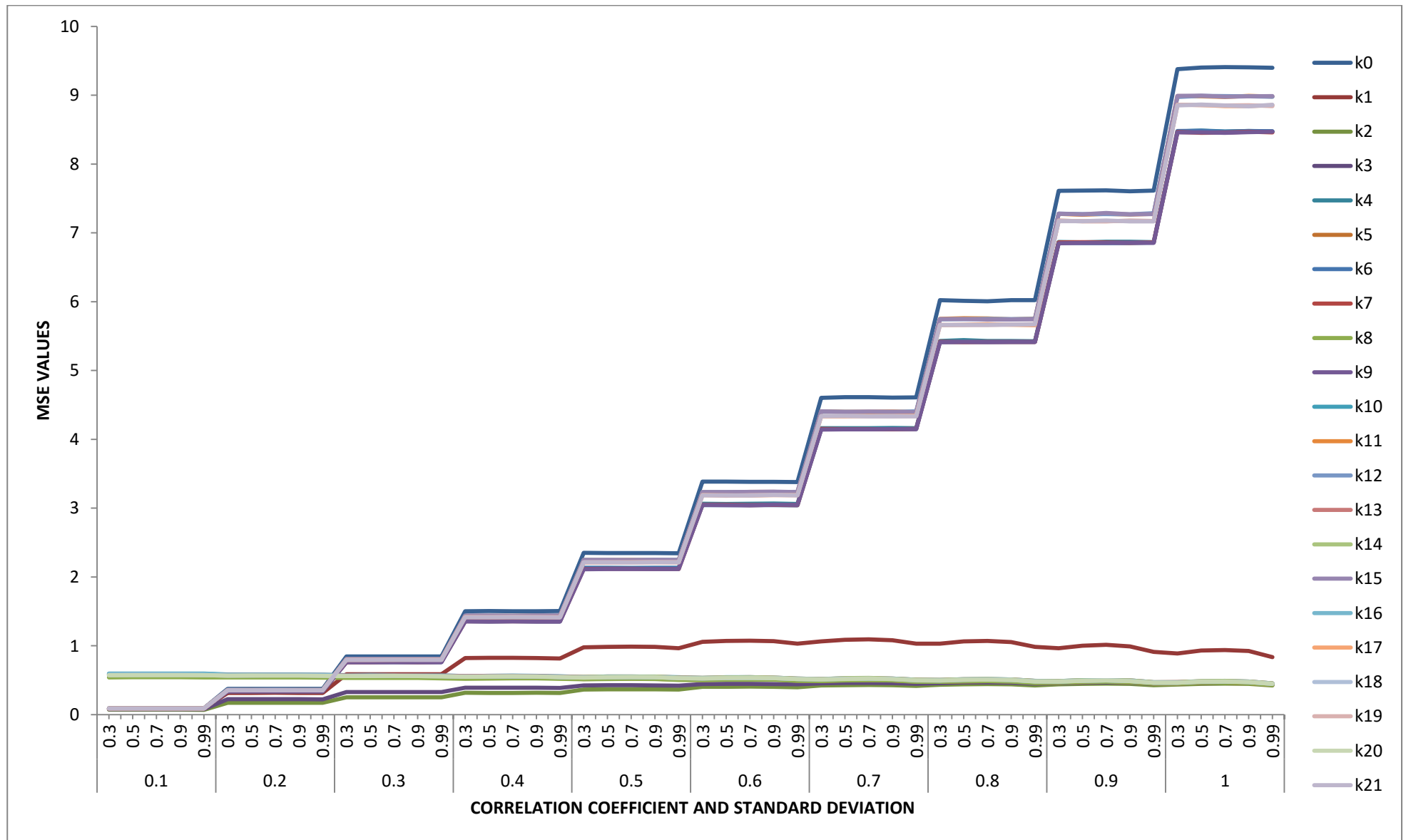


Figure 4.20 Graph of Simulated values of MSE for  $n=50$ ,  $p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



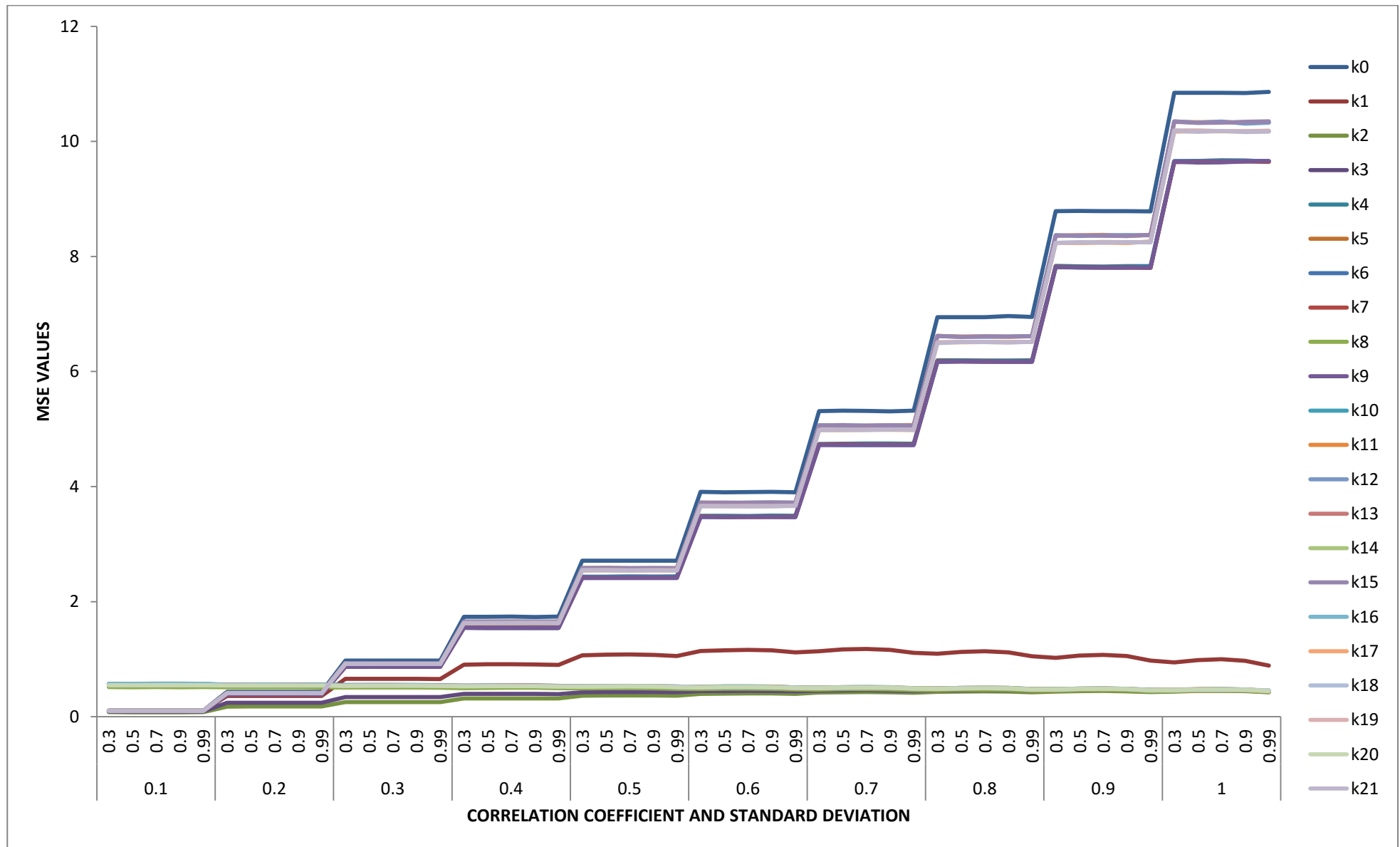


Figure 4.21 Graph of Simulated values of MSE for  $n=50, p=9 \sigma = 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



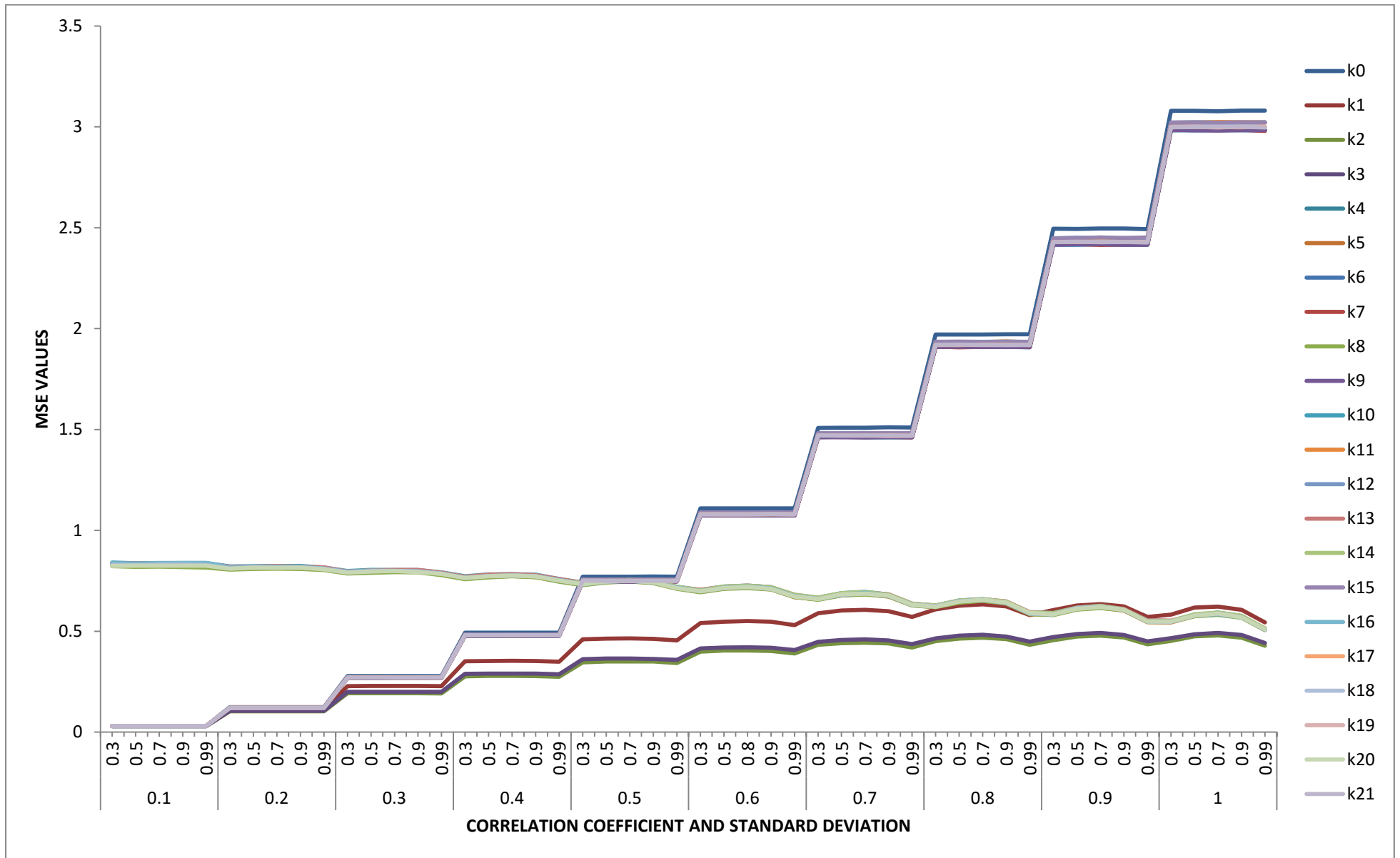


Figure 4.22 Graph of Simulated values of MSE for  $n=80$ ,  $p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.4(b) Simulated MSE values for  $n=80$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





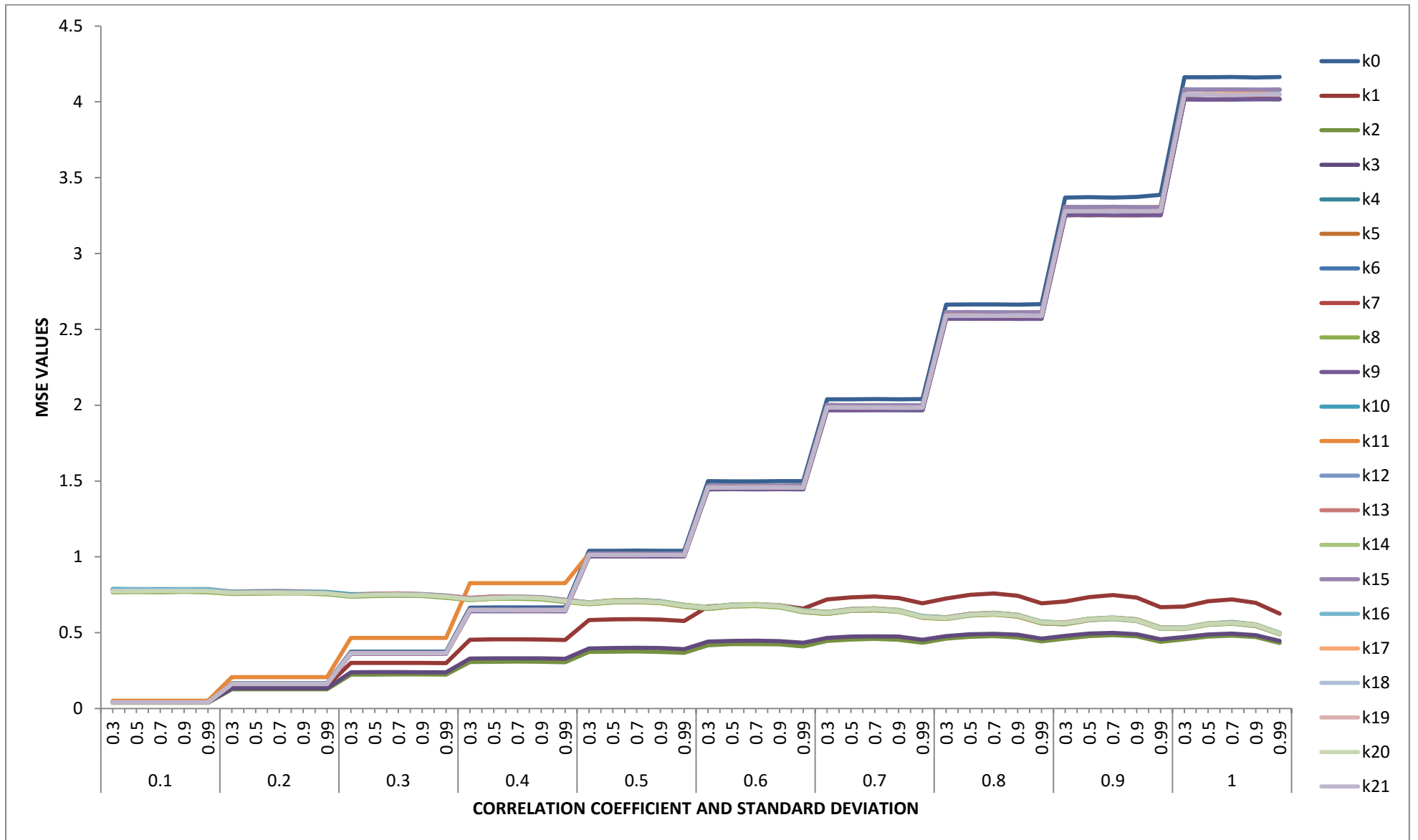


Figure 4.23 Graph of Simulated values of MSE for  $n=80$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



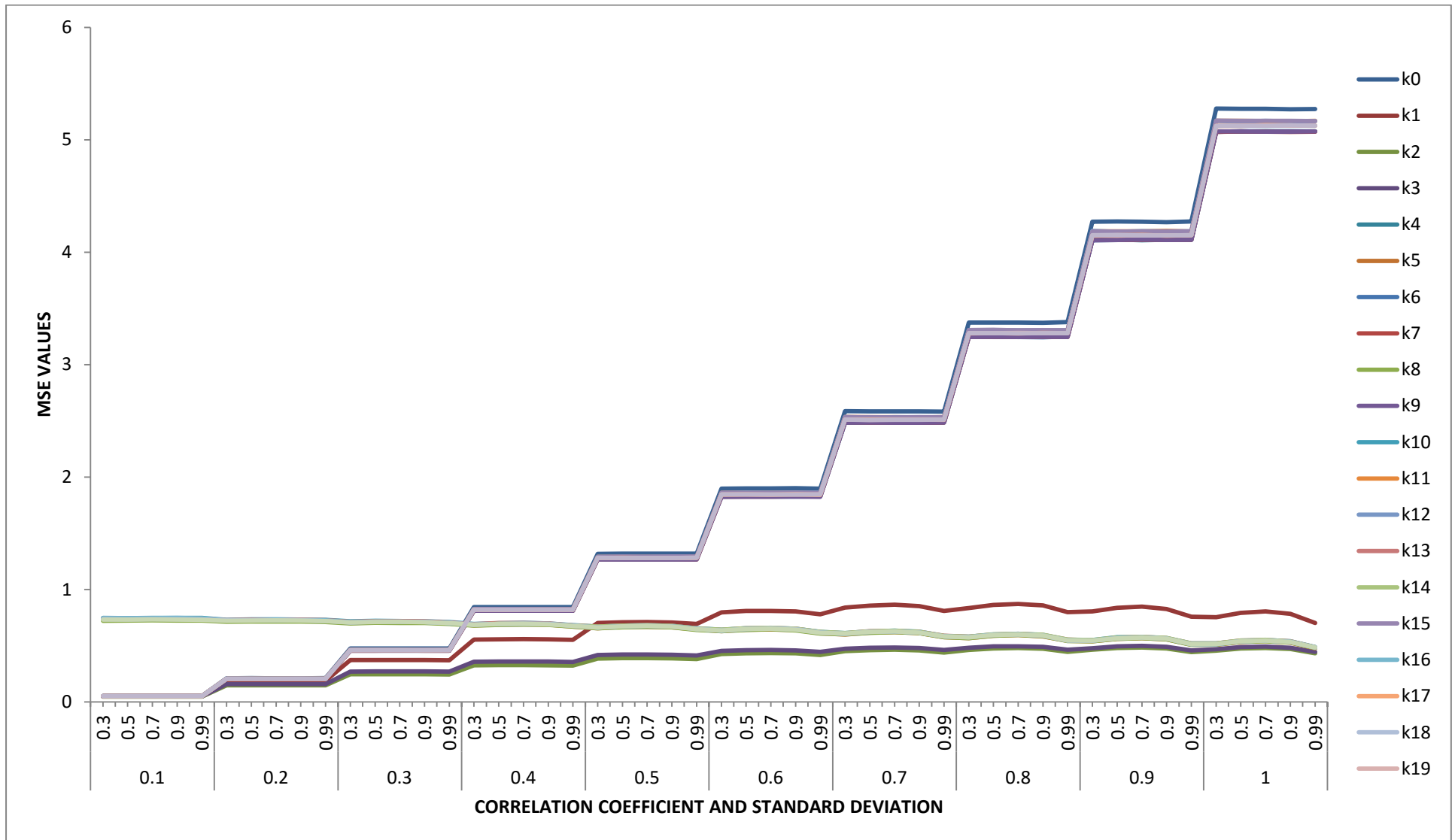


Figure 4.24 Graph of Simulated values of MSE for  $n=80$ ,  $p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



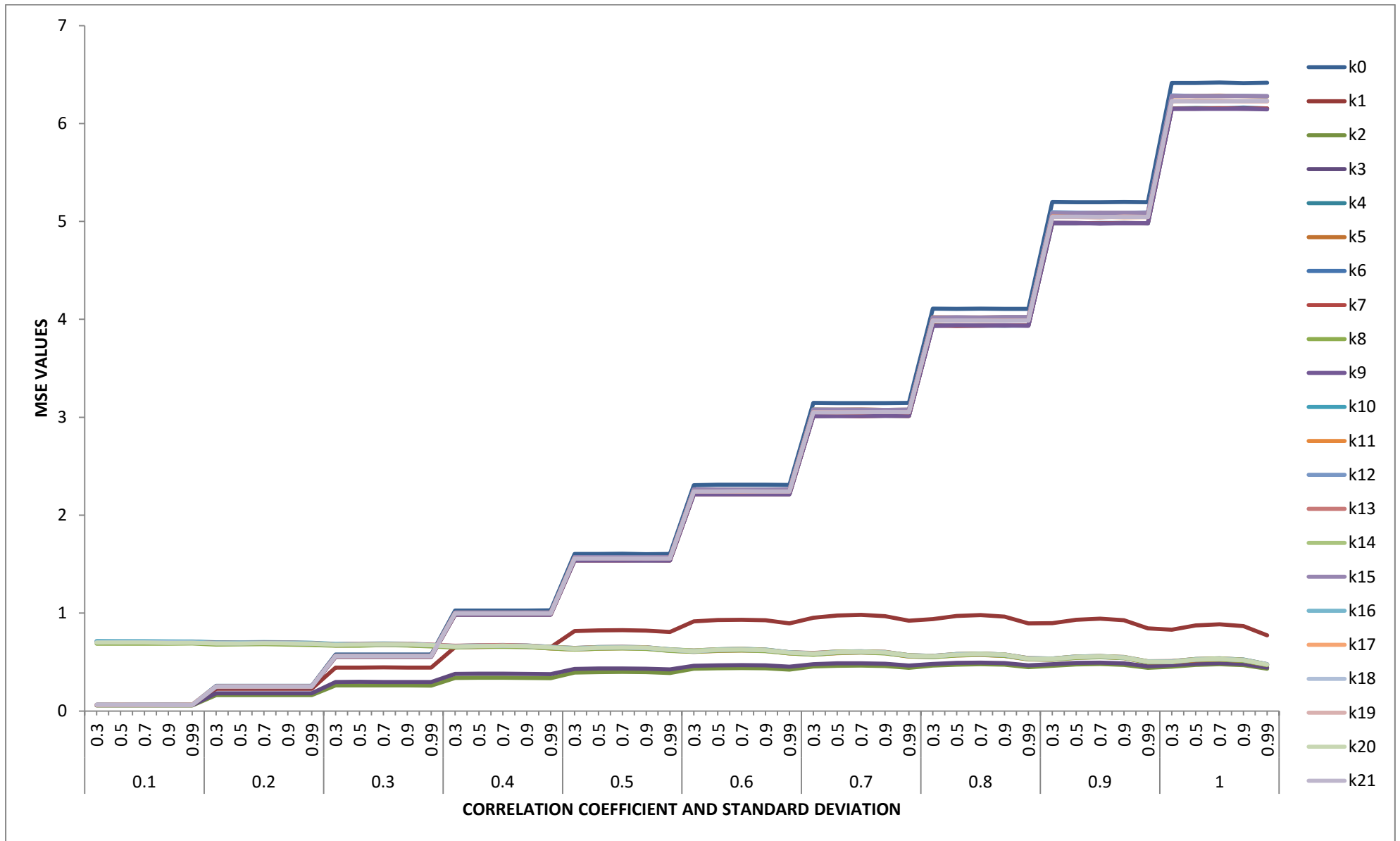


Figure 4.25 Graph of Simulated values of MSE for  $n=80$ ,  $p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



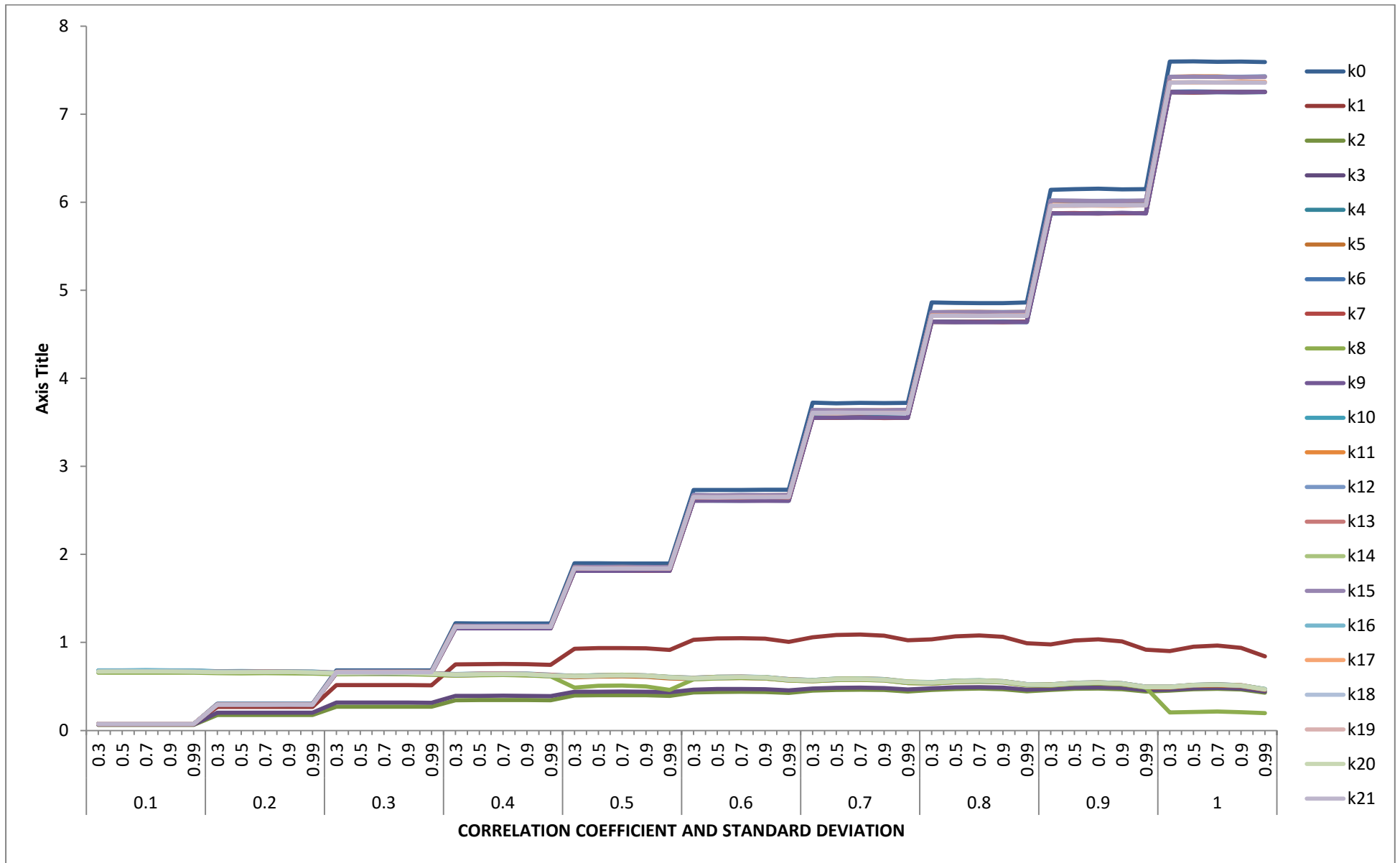


Figure 4.26 Graph of Simulated values of MSE for  $n=80$ ,  $p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.4(f) Simulated MSE values for  $n=80$ ,  $p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





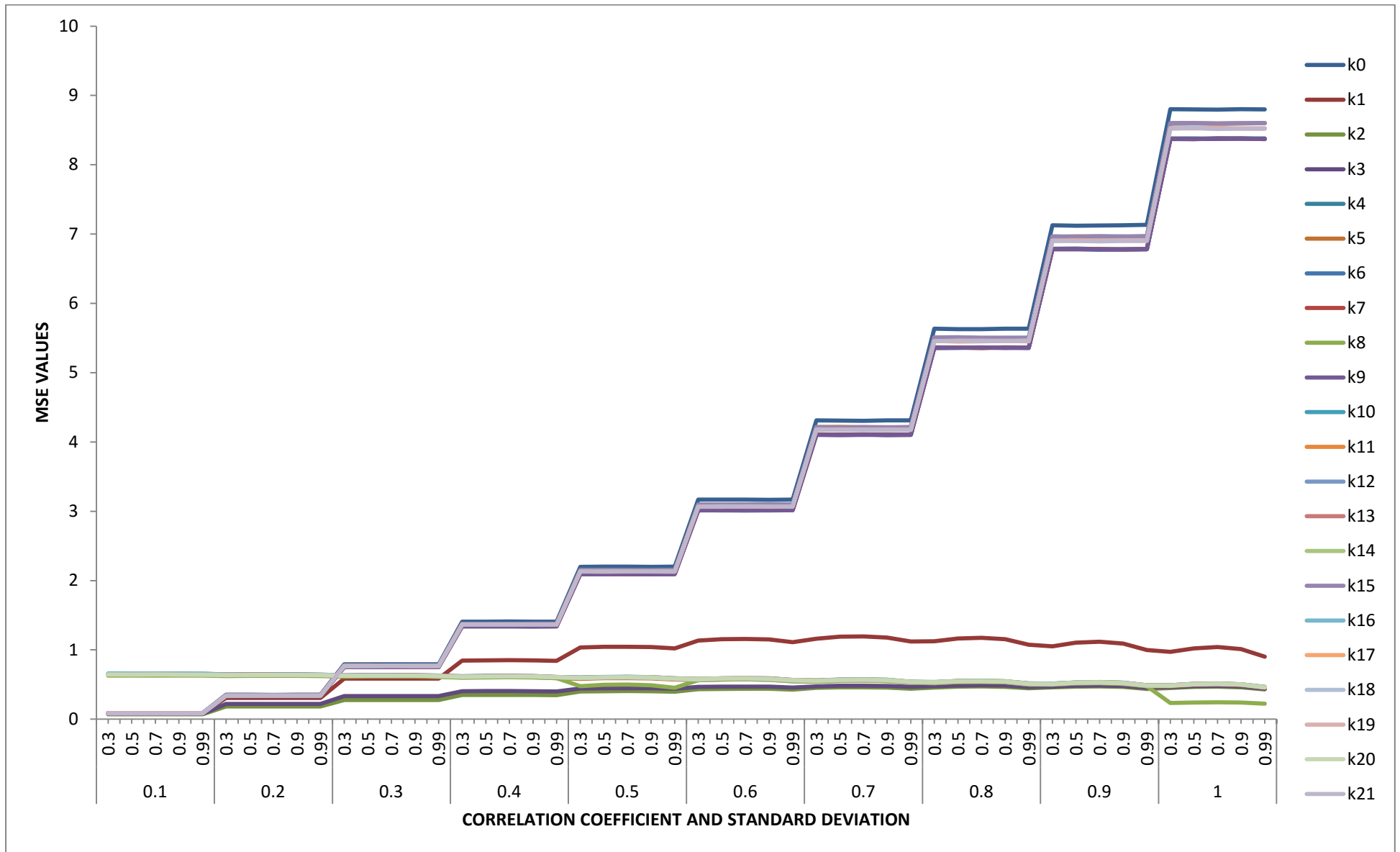


Figure 4.27 Graph of Simulated values of MSE for  $n=80$ ,  $p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



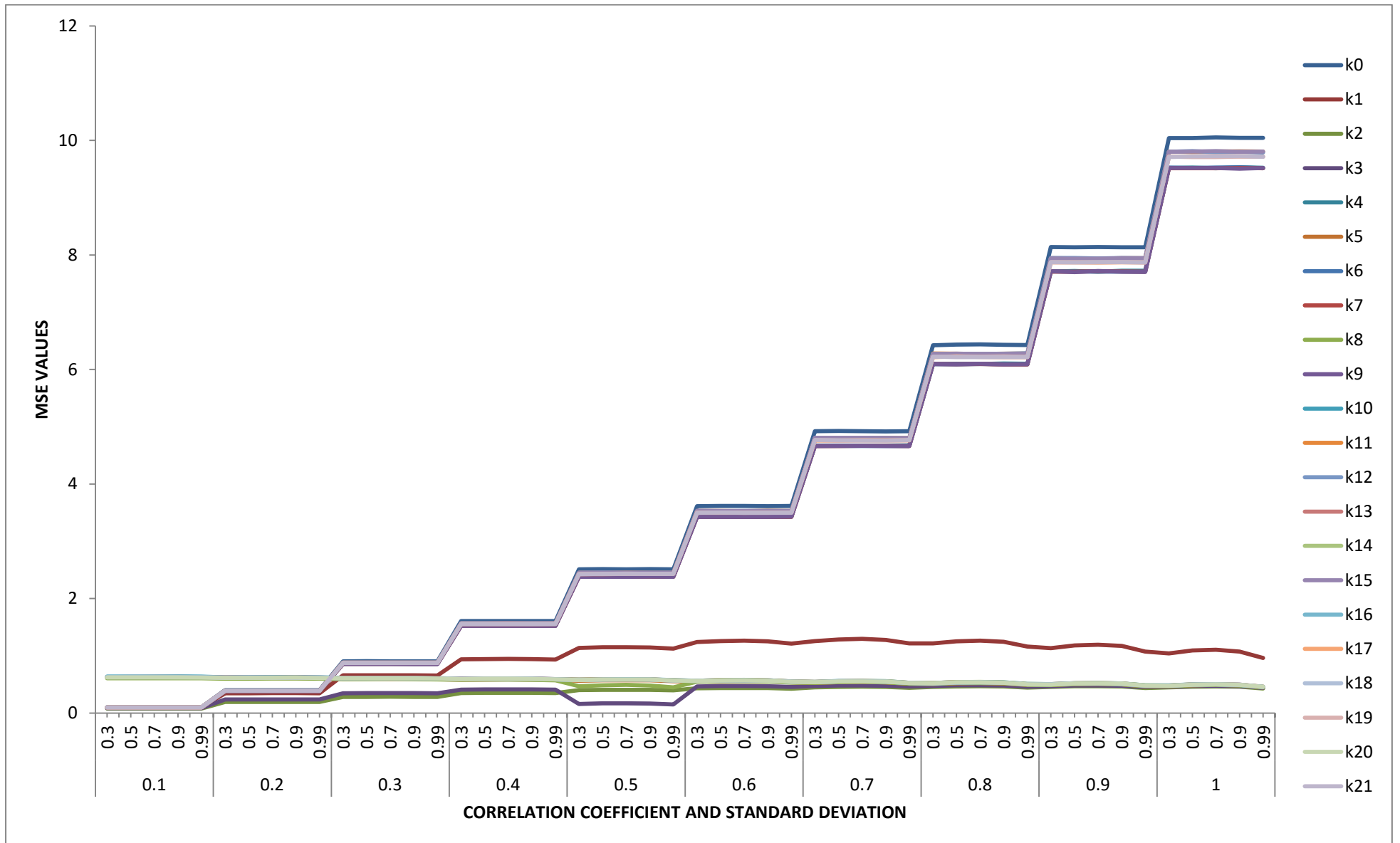


Figure 4.28 Graph of Simulated values of MSE for  $n=80$ ,  $p=9$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



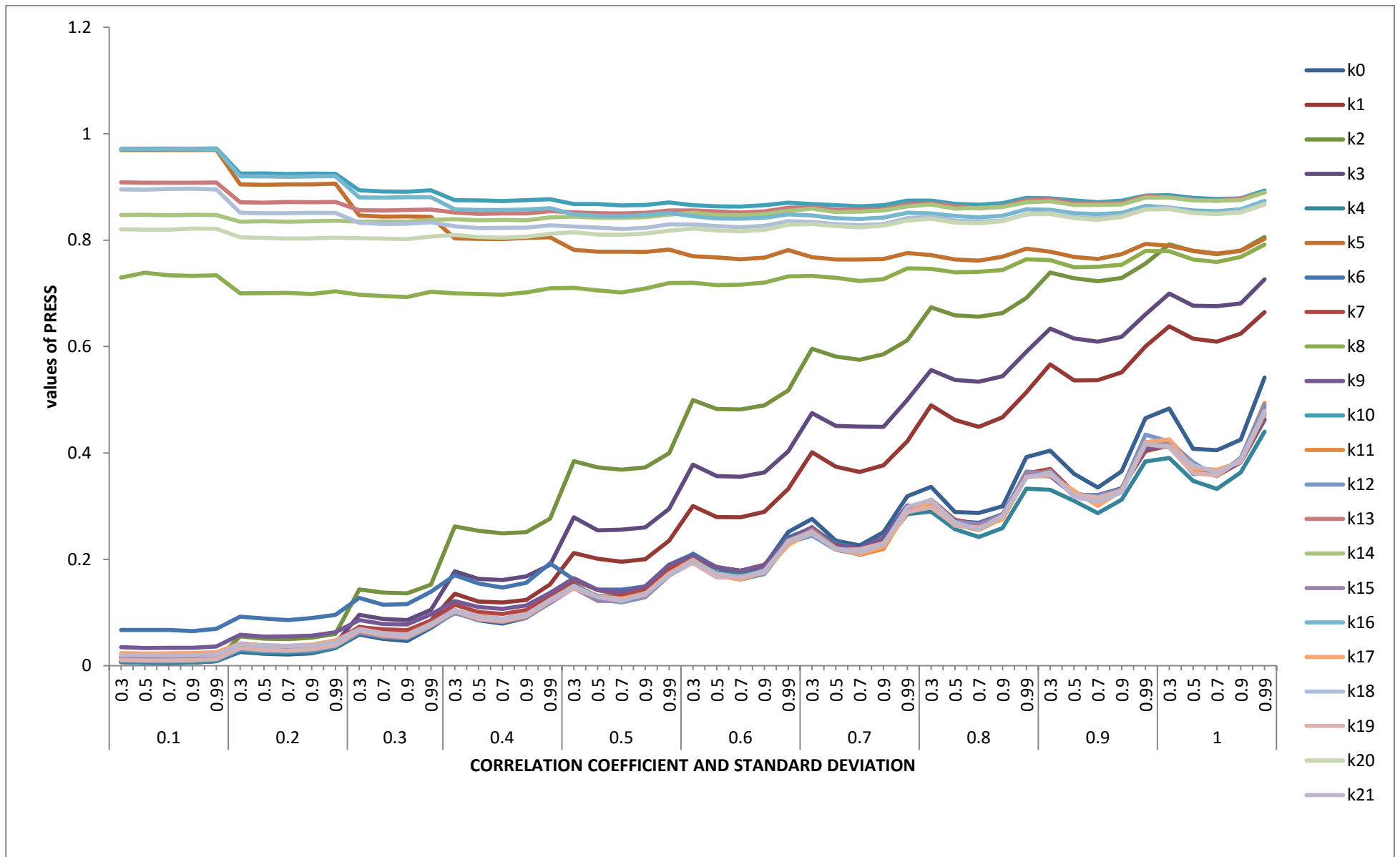


Figure 4.29 Graph of Simulated values of PRESS for  $n=10, p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



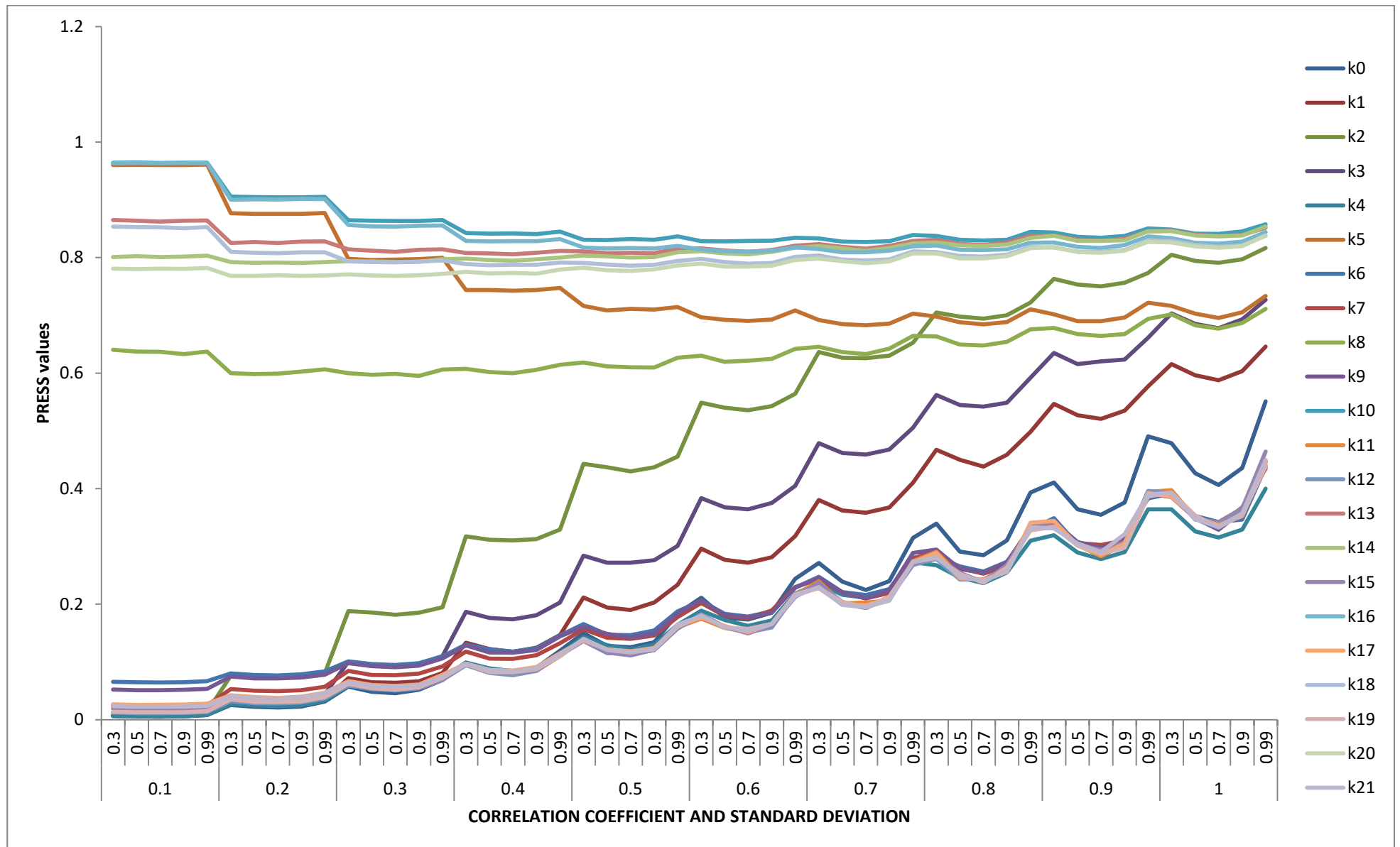


Figure 4.30 Graph of Simulated values of PRESS for  $n=10$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





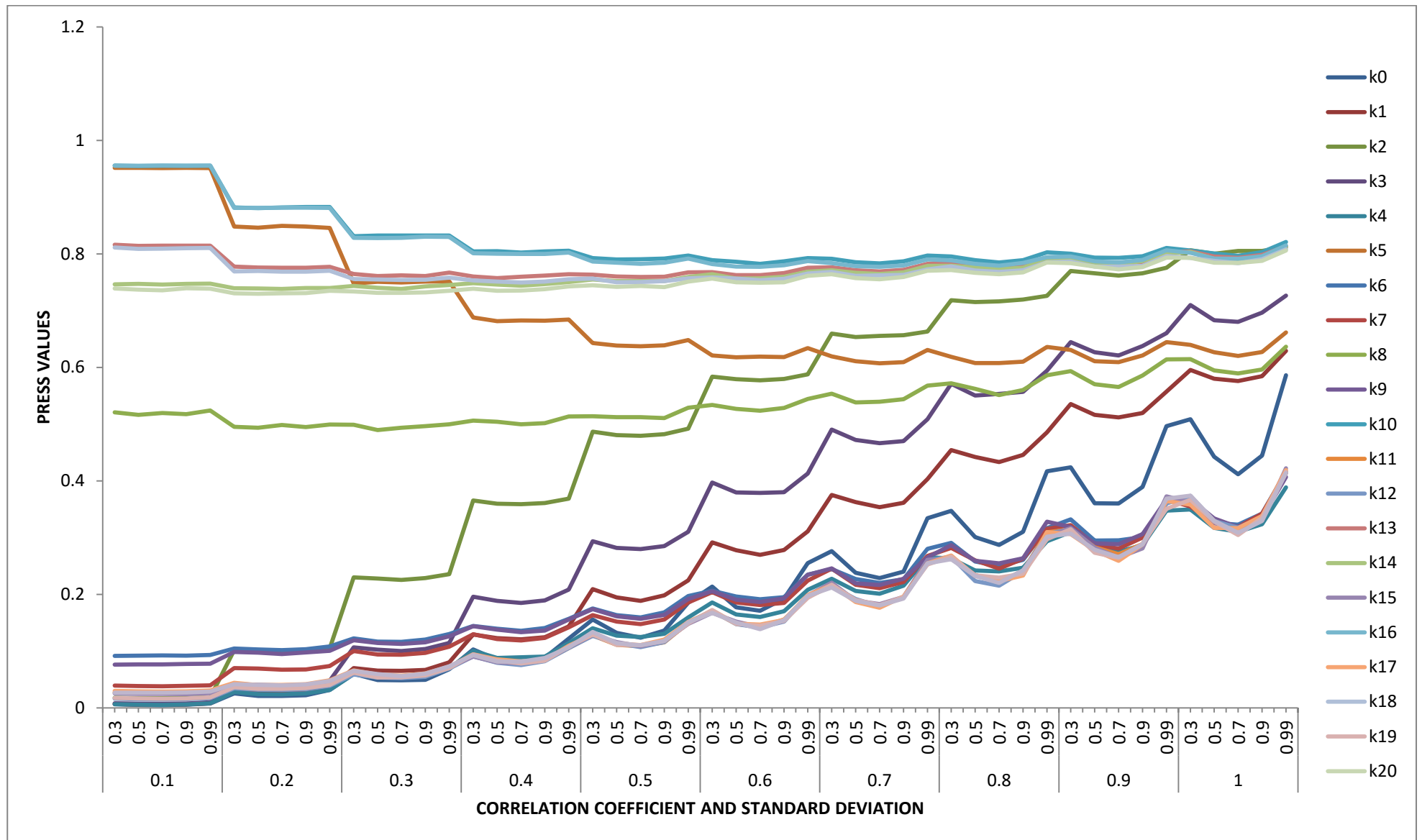


Figure 4.31 Graph of Simulated values of PRESS for  $n=10$ ,  $p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



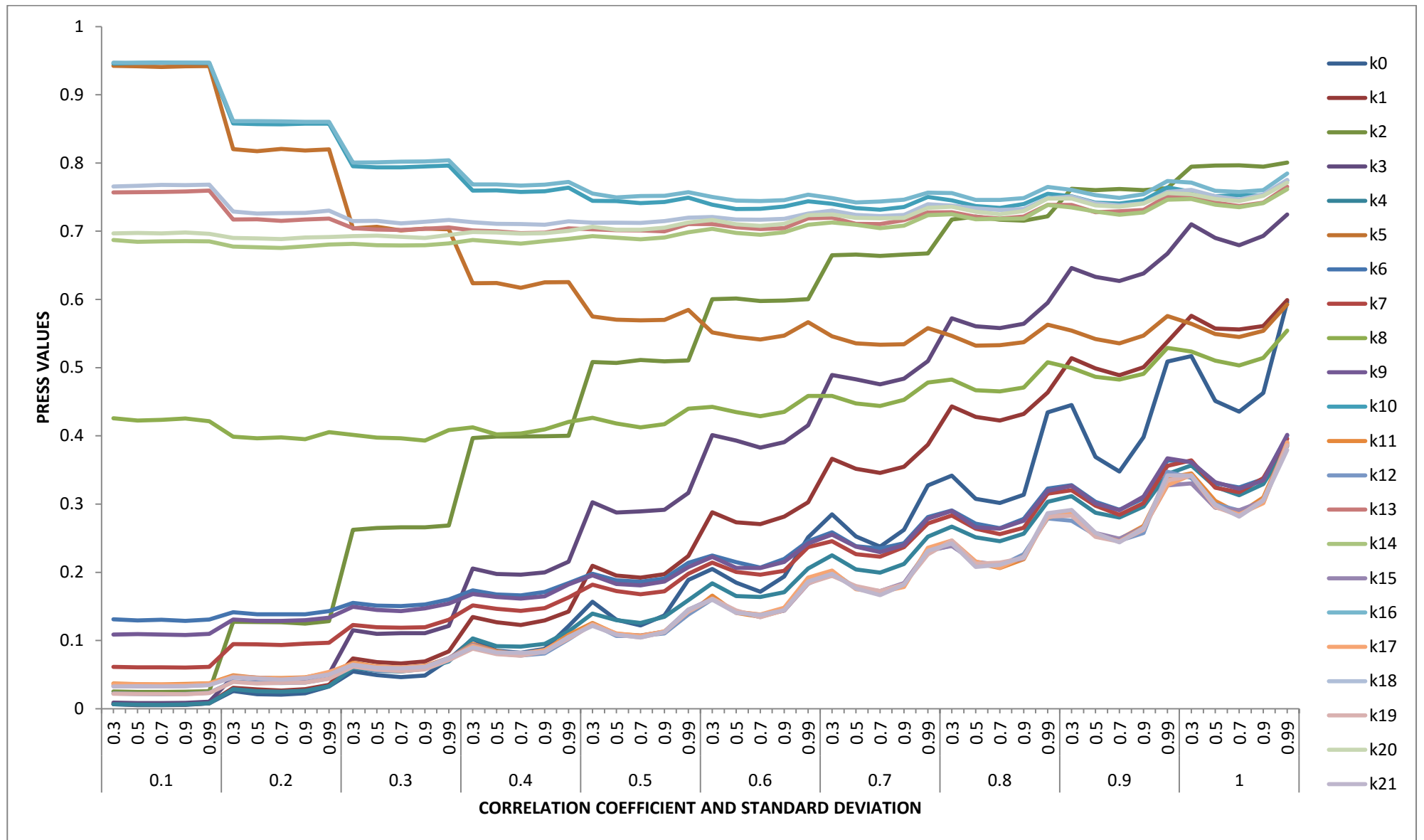


Figure 4.32 Graph of Simulated values of PRESS for  $n=10, p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



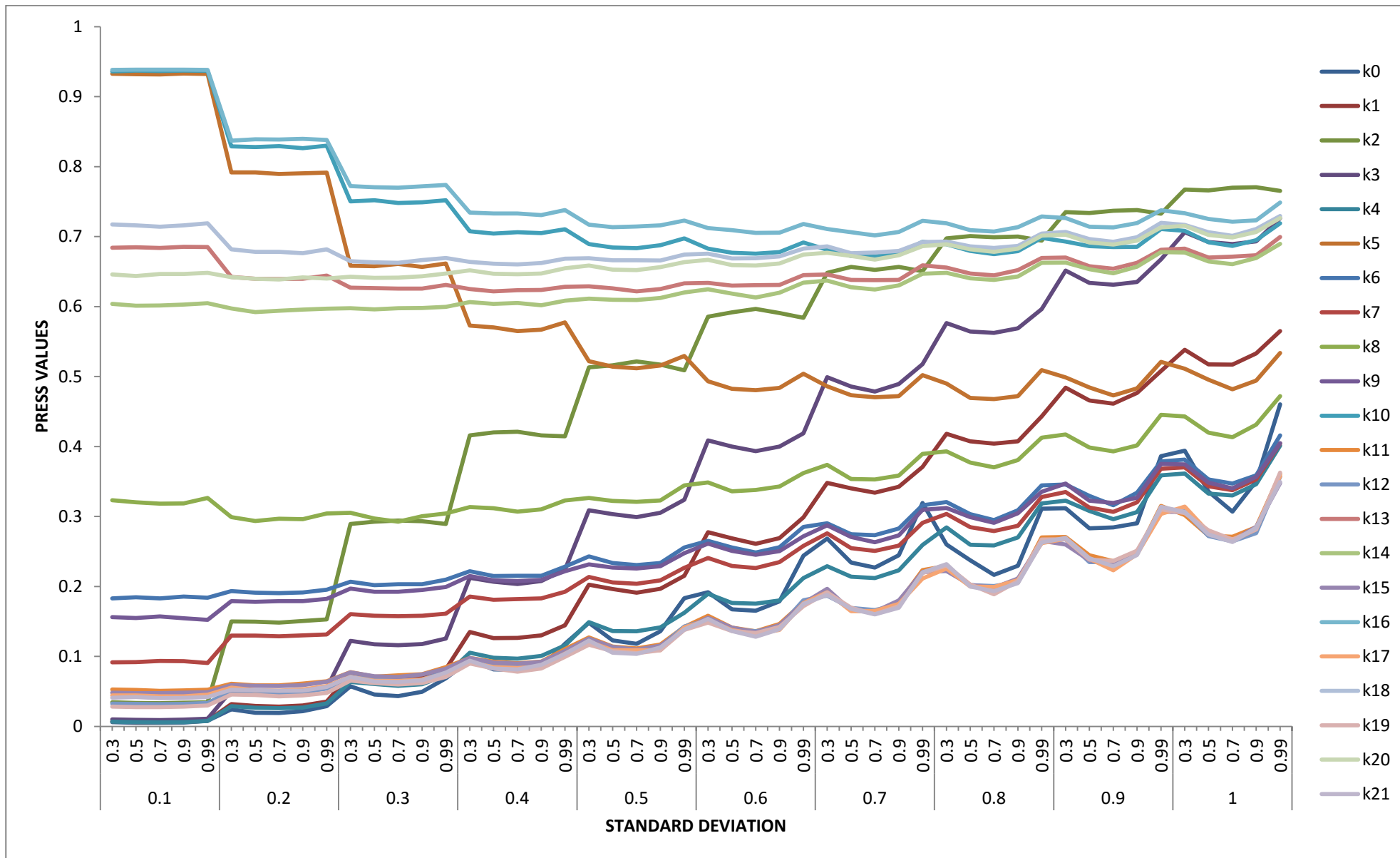


Figure 4.33 Graph of Simulated values of PRESS for  $n=10, p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



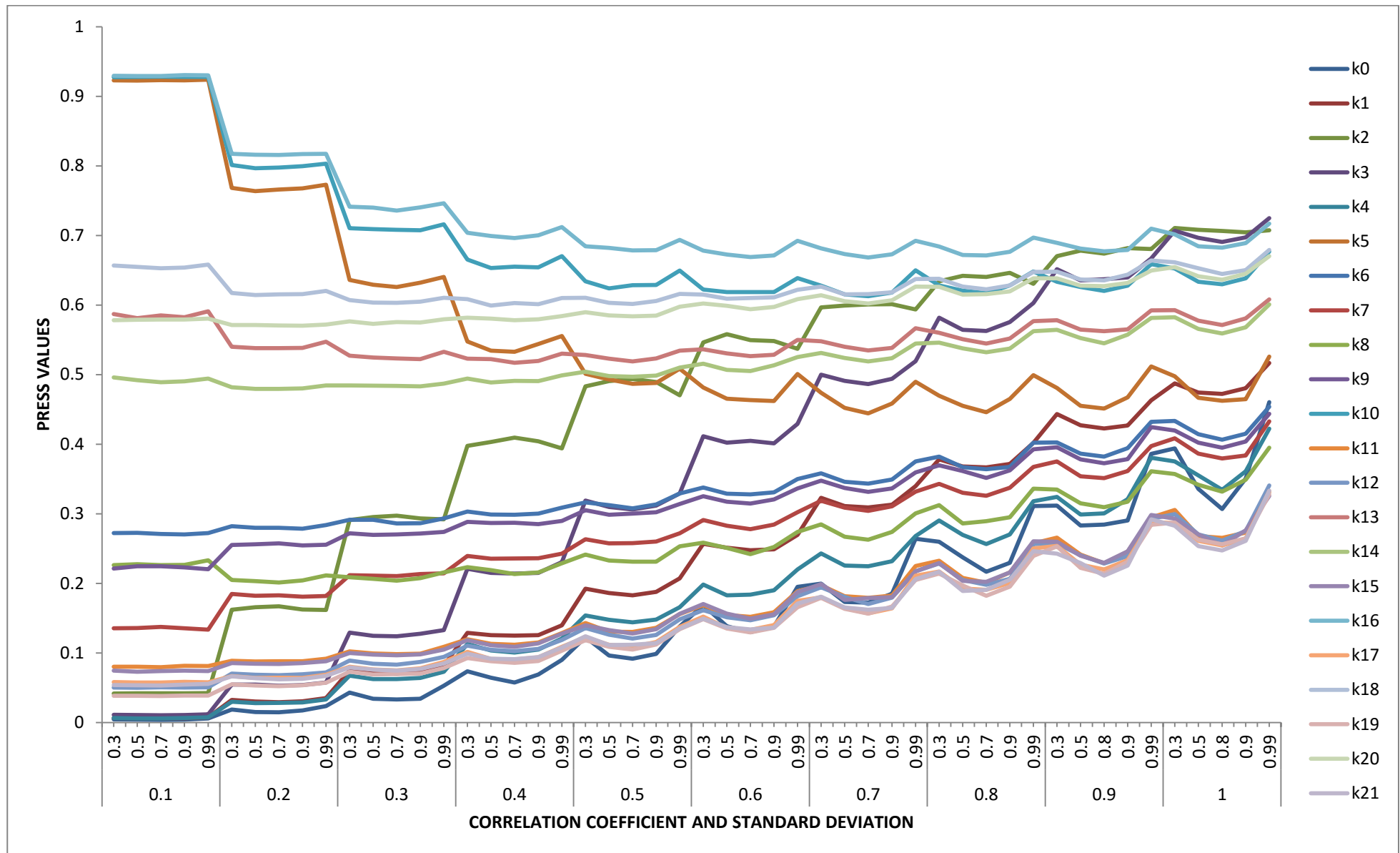


Figure 4.34 Graph of Simulated values of PRESS for  $n=10$ ,  $p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





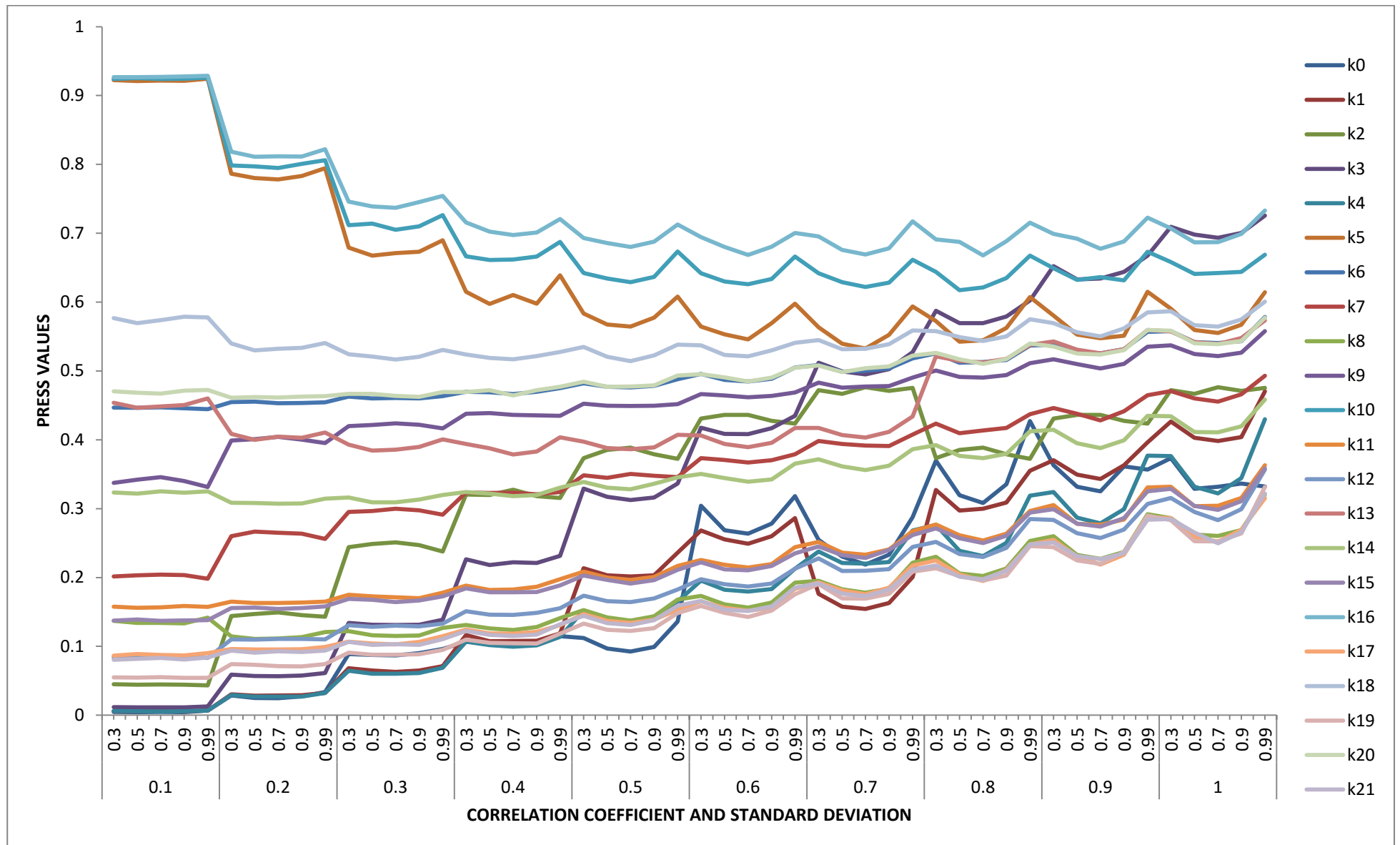


Figure 4.35 Graph of Simulated values of PRESS for  $n=10, p=9$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



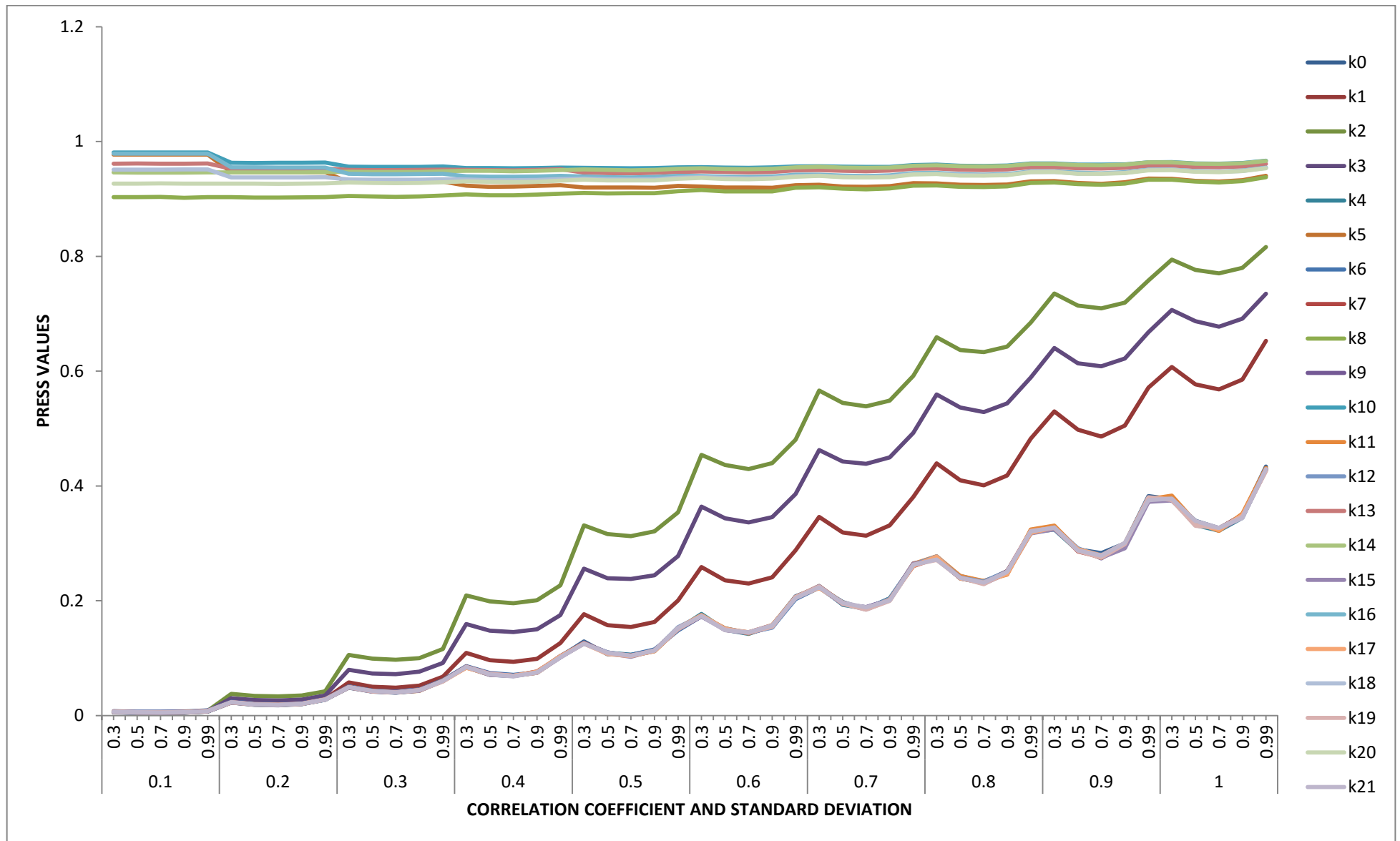


Figure 4.36 Graph of Simulated values of PRESS for  $n=30$ ,  $p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



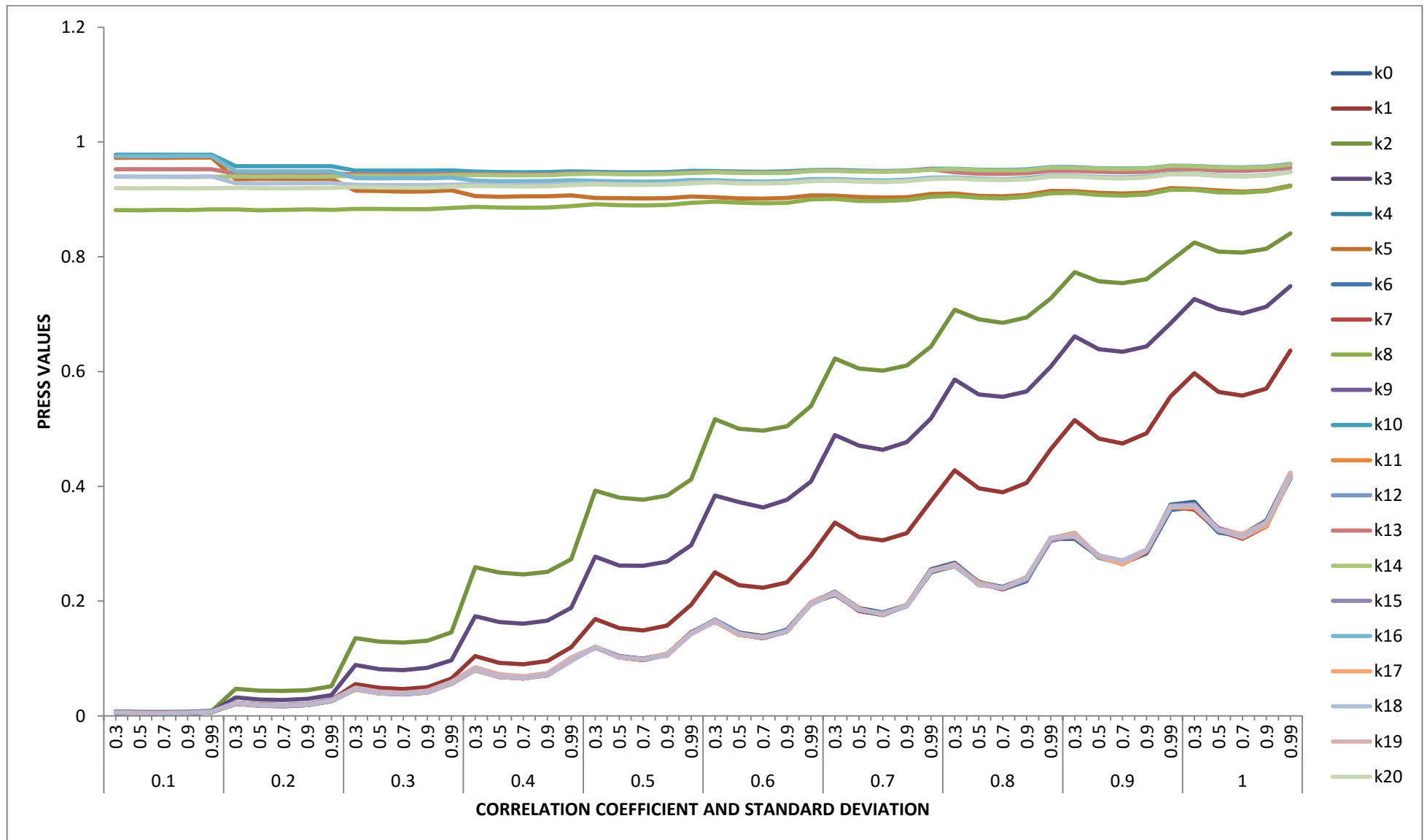


Figure 4.37 Graph of Simulated values of PRESS for  $n=30$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



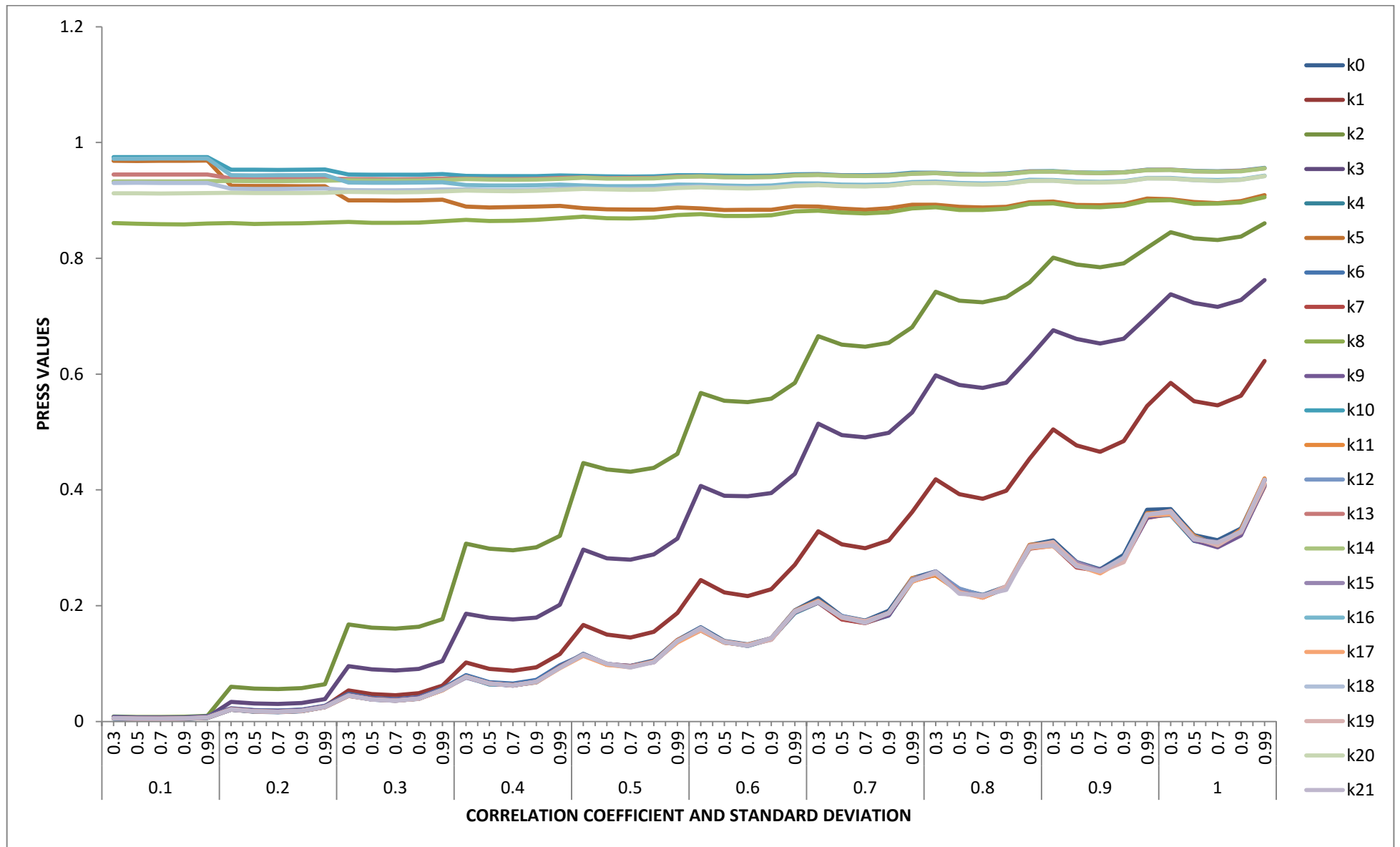


Figure 4.38 Graph of Simulated values of PRESS for  $n=30, p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





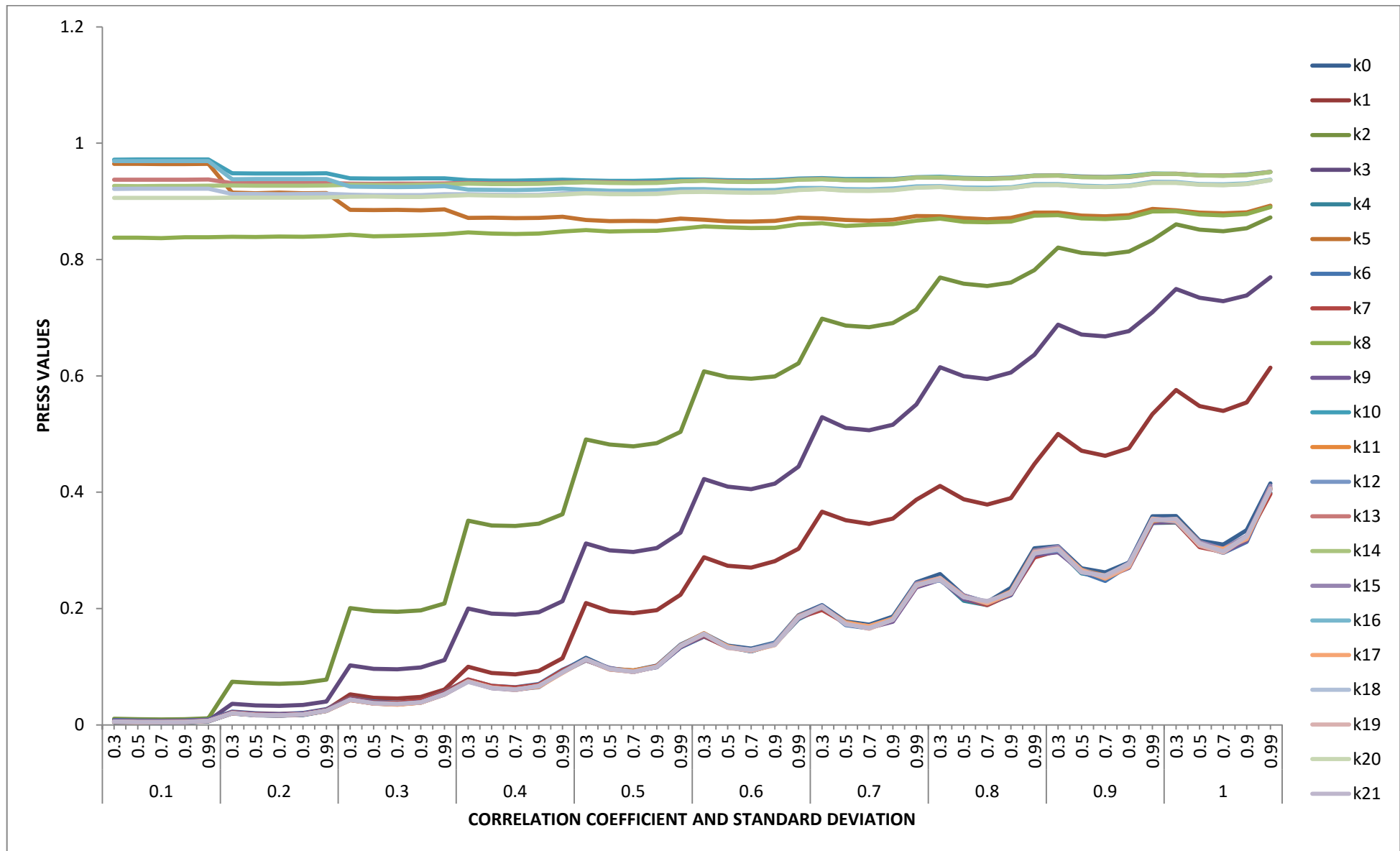


Figure 4.39 Graph of Simulated values of PRESS for  $n=30$ ,  $p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



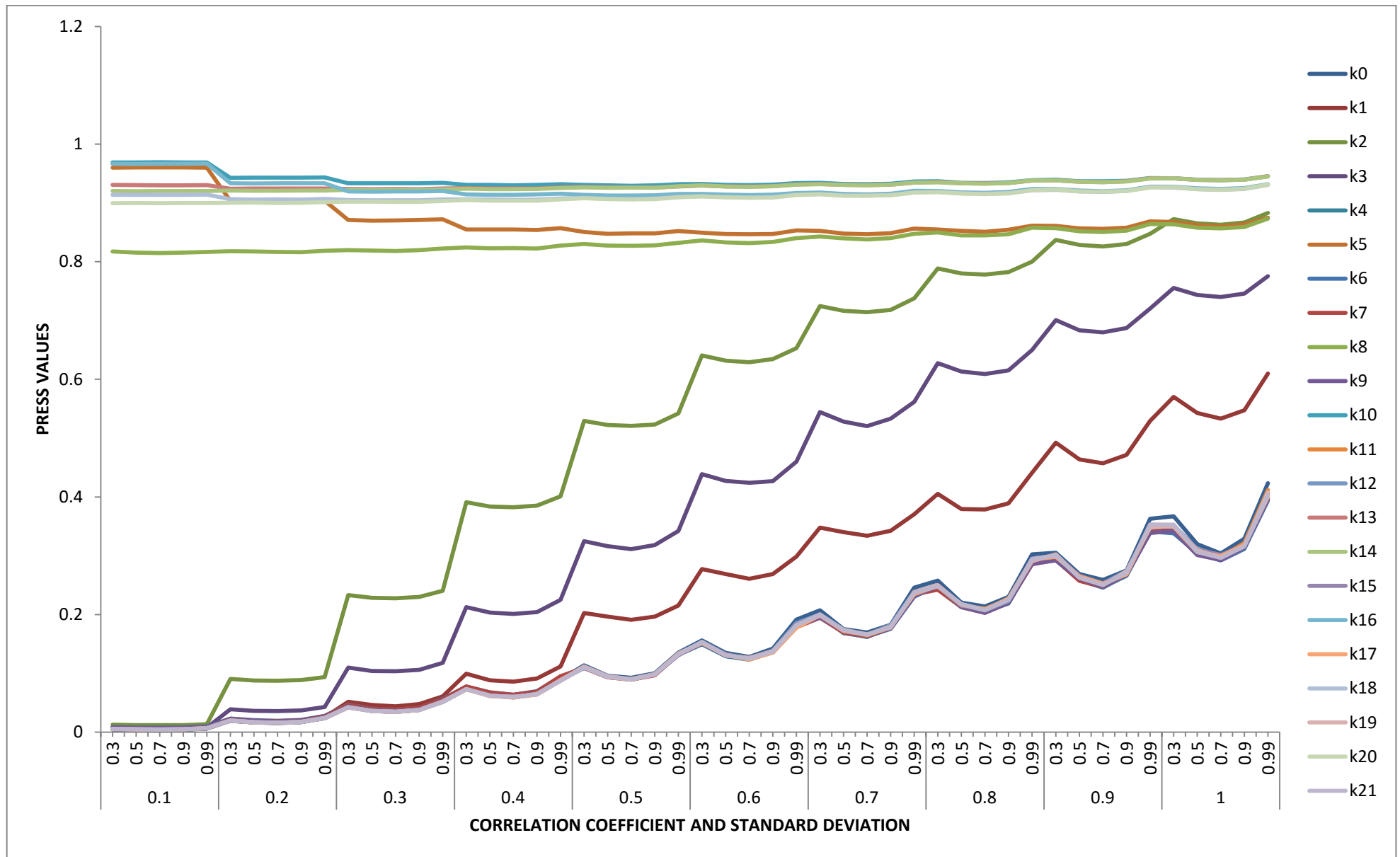


Figure 4.40 Graph of Simulated values of PRESS for  $n=30$ ,  $p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



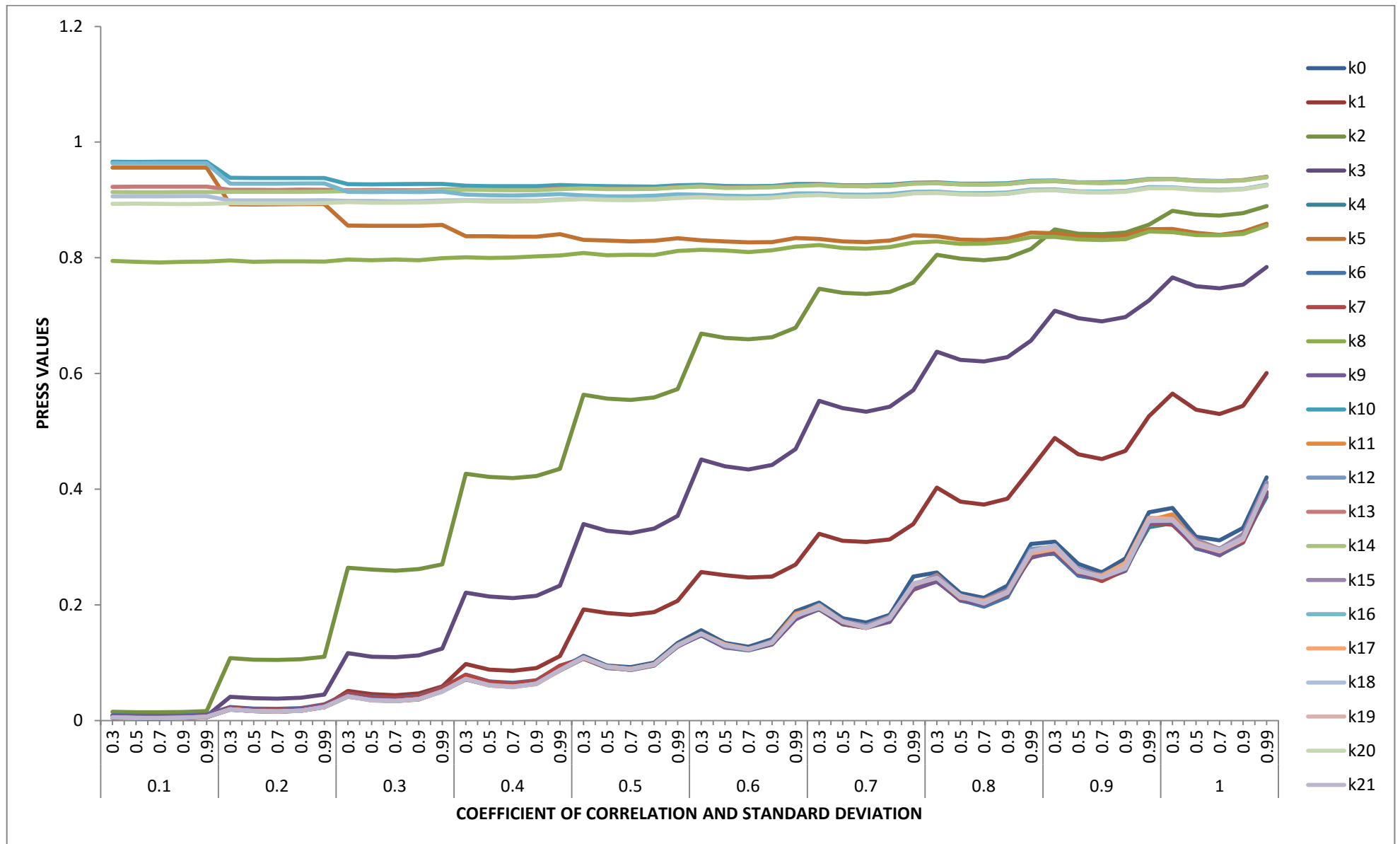


Figure 4.41 Graph of Simulated values of PRESS for  $n=30$ ,  $p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



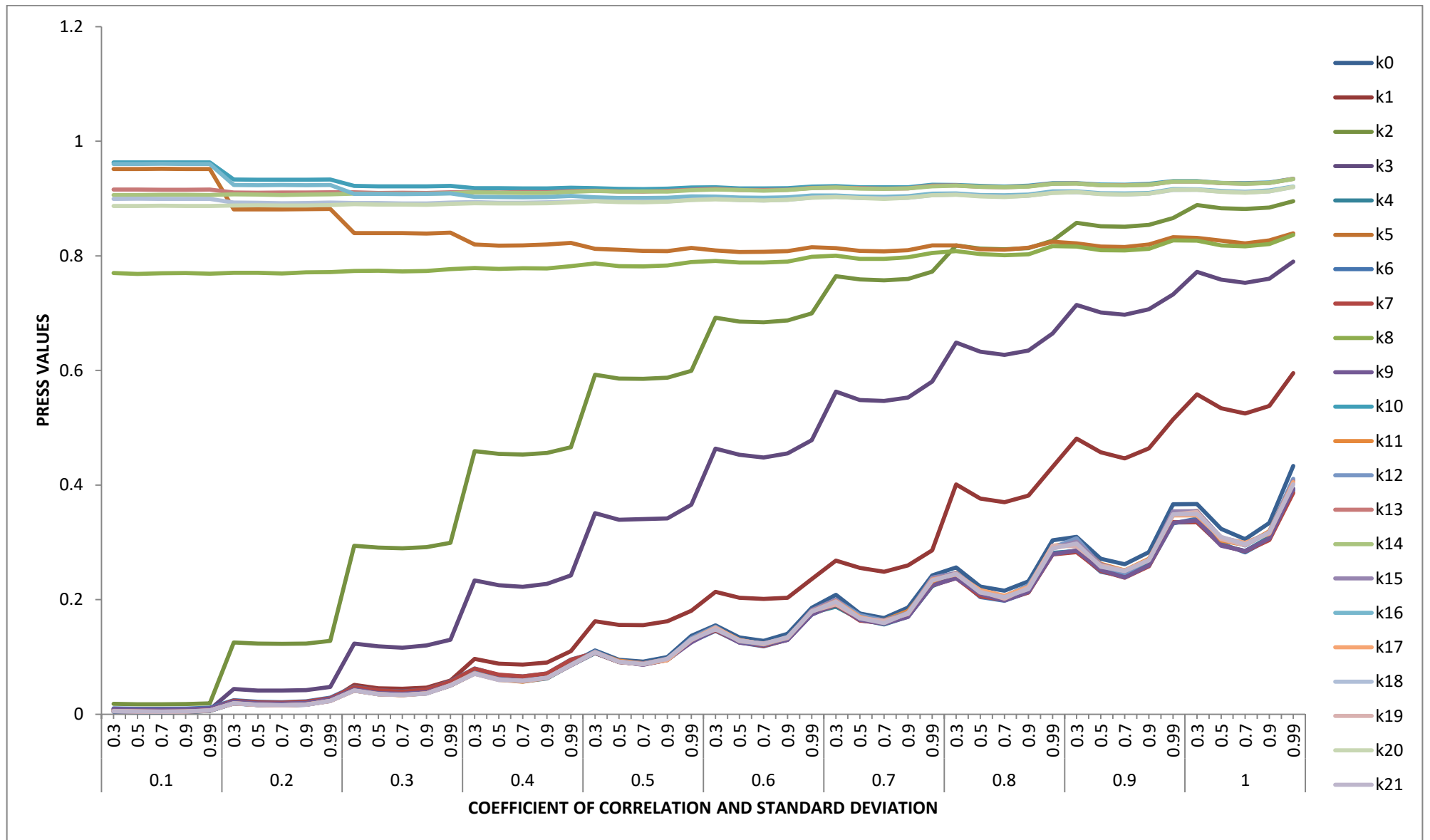


Figure 4.42 Graph of Simulated values of PRESS for  $n=30$ ,  $p=9$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





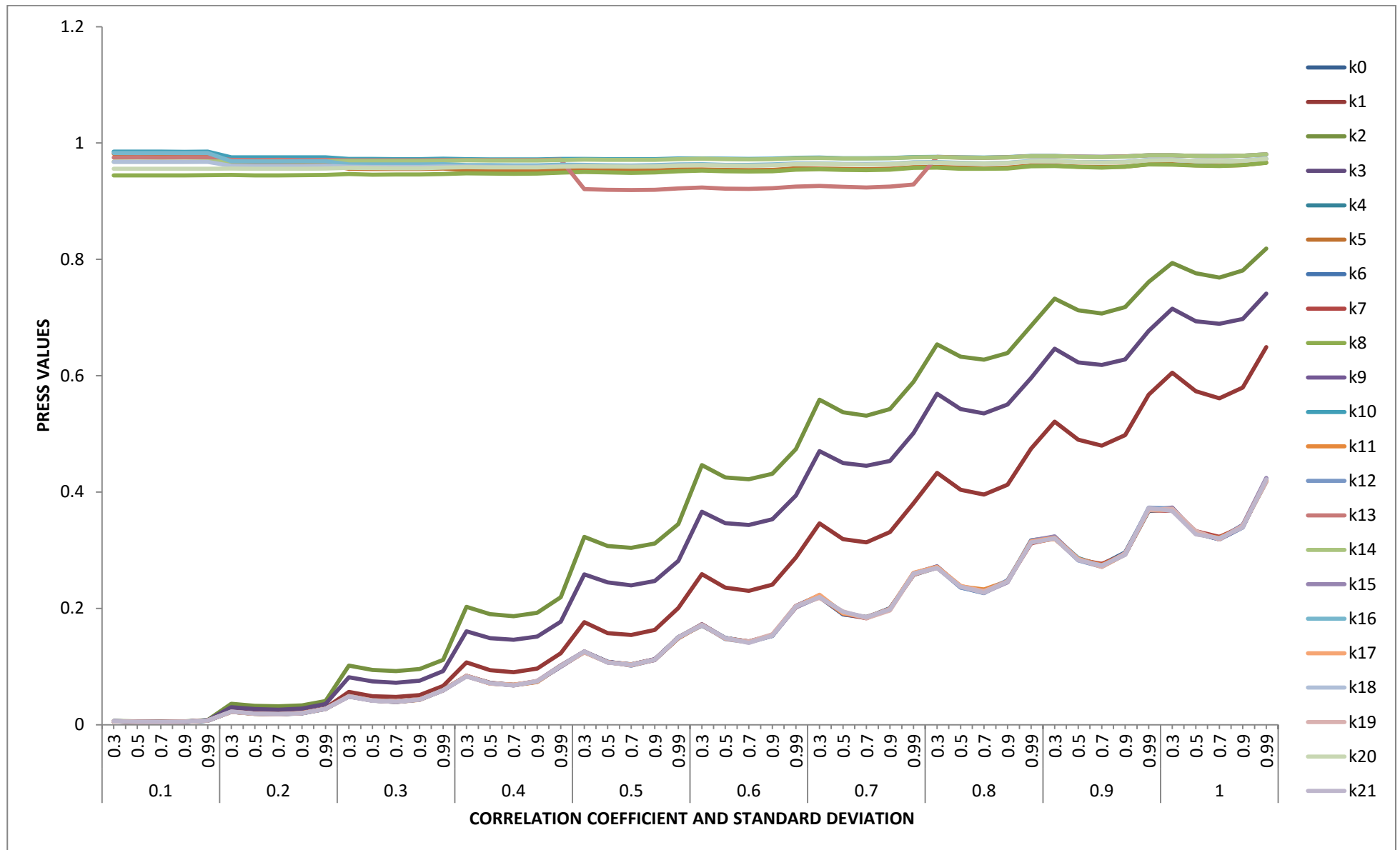


Figure 4.43 Graph of Simulated values of PRESS for  $n=50$ ,  $p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



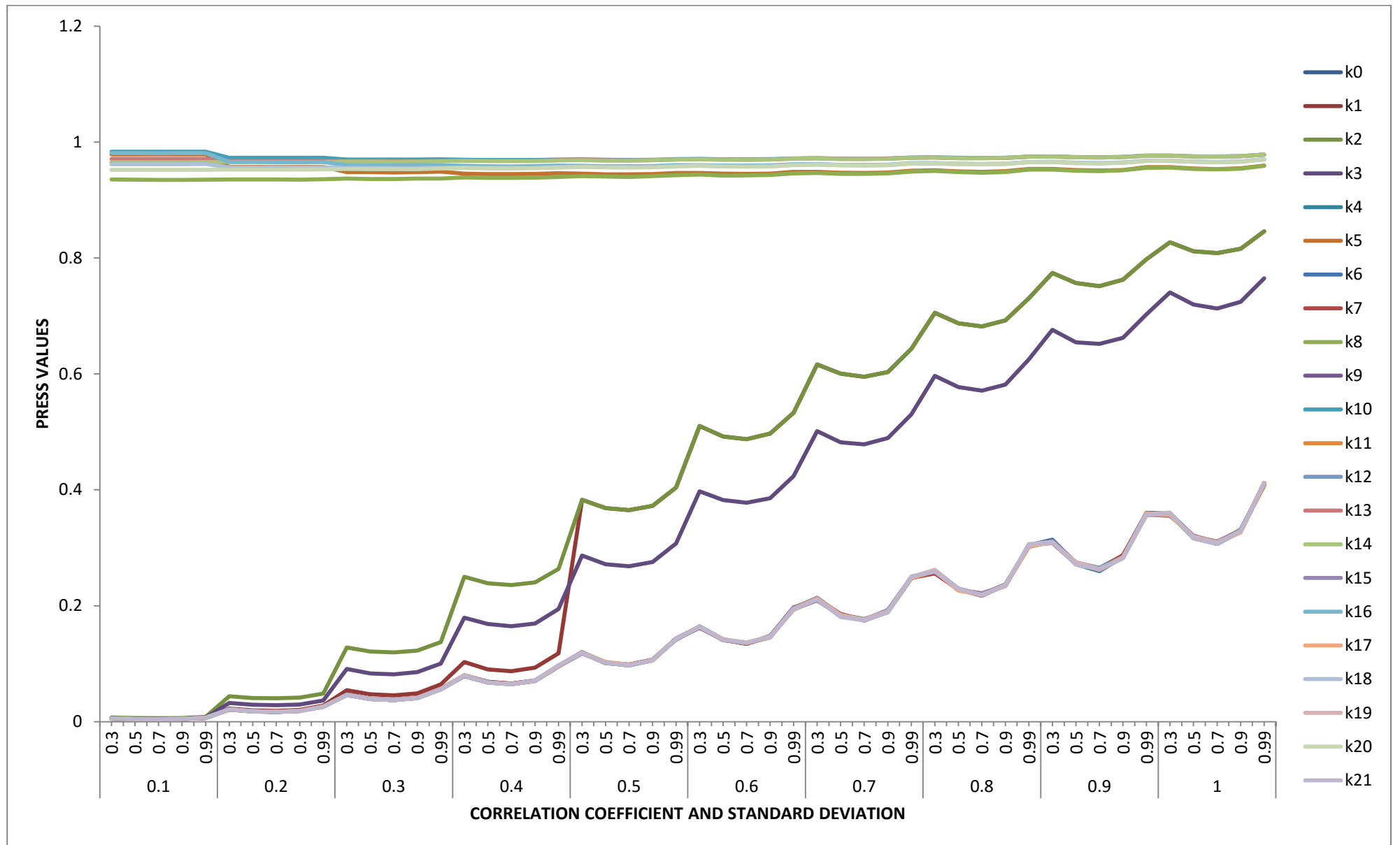


Figure 4.44 Graph of Simulated values of PRESS for  $n=50$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



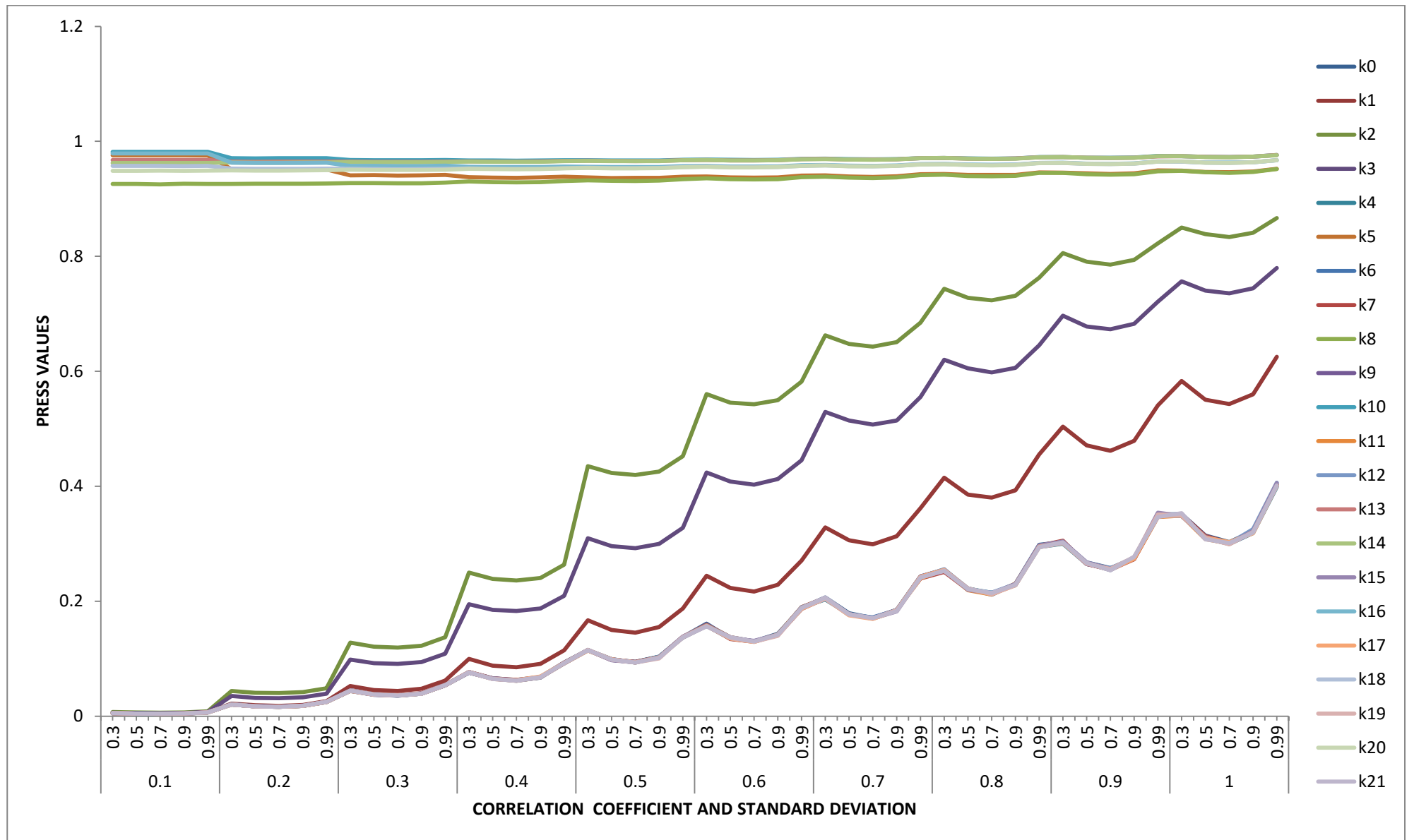


Figure 4.45 Graph of Simulated values of PRESS for  $n=50$ ,  $p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



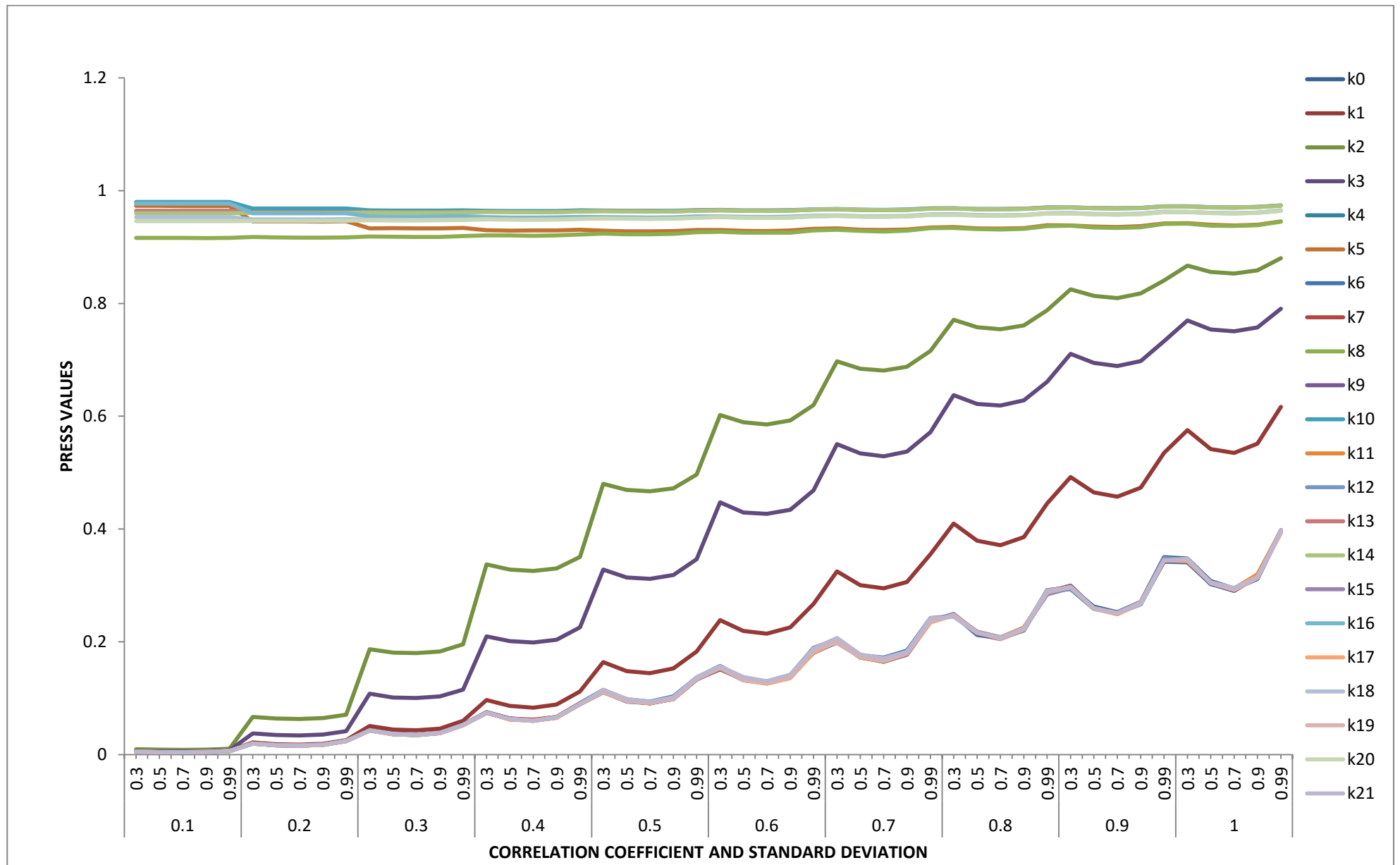


Figure 4.46 Graph of Simulated values of PRESS for  $n=50$ ,  $p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





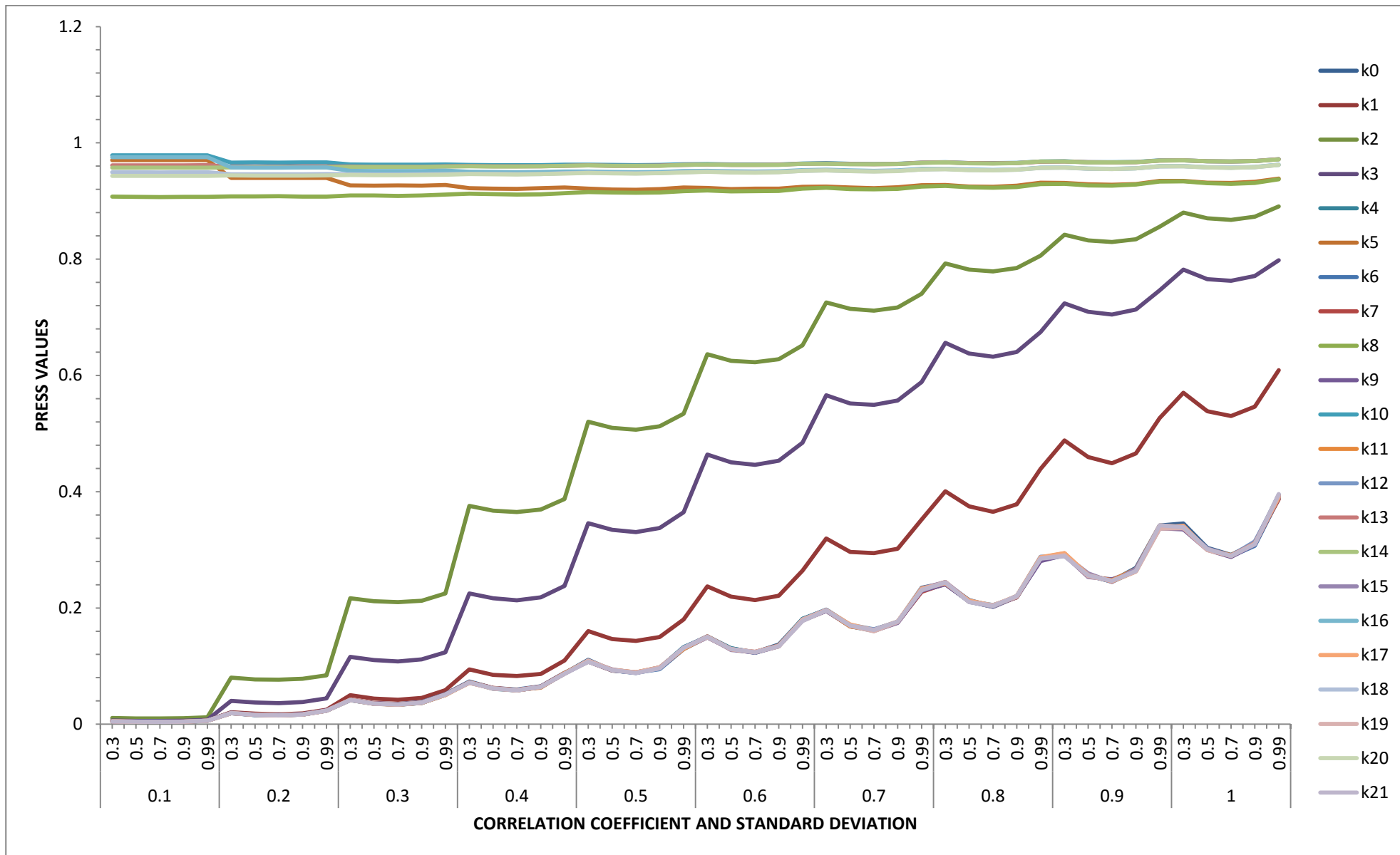


Figure 4.47 Graph of Simulated values of PRESS for  $n=50$ ,  $p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



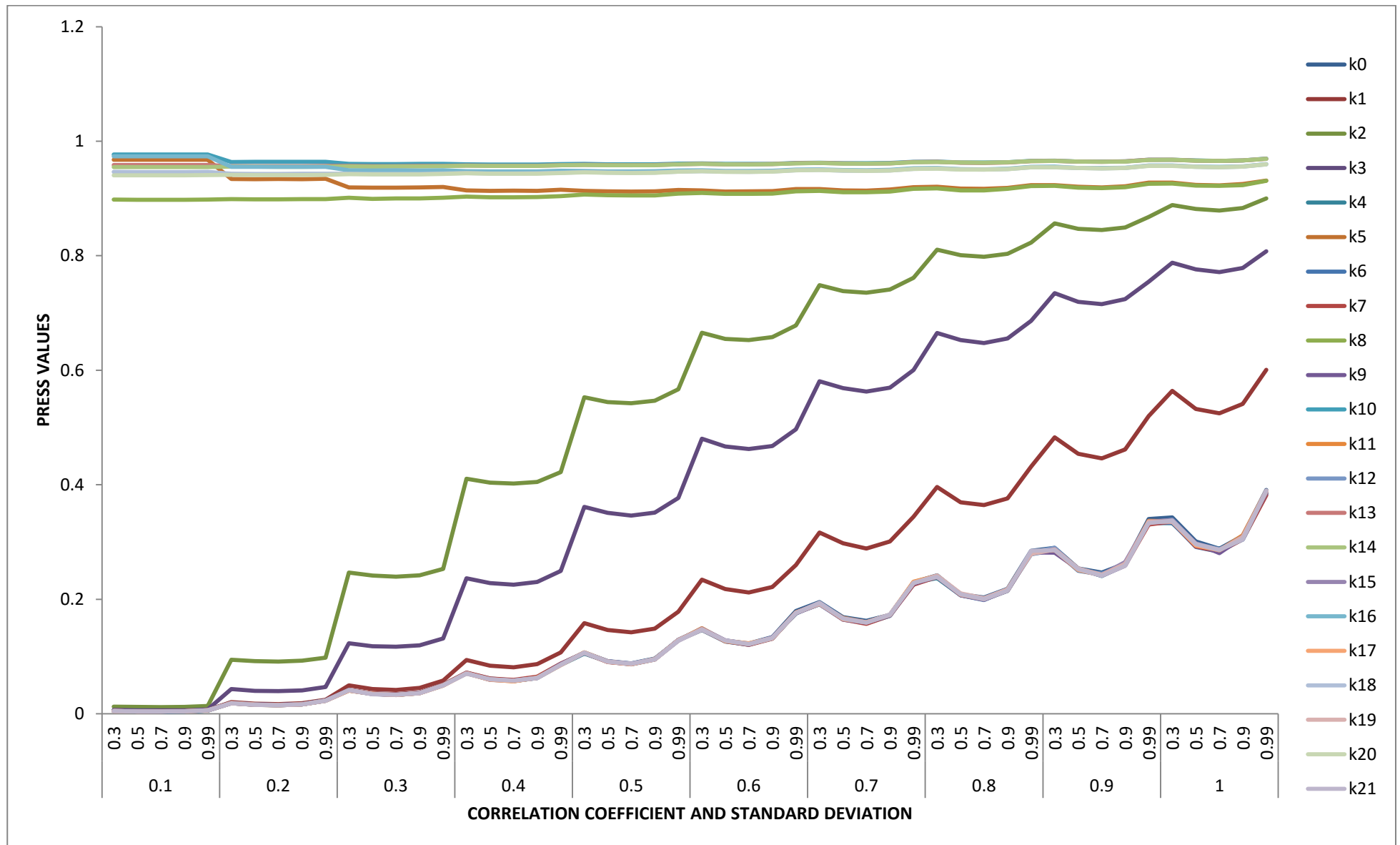


Figure 4.48 Graph of Simulated values of PRESS for  $n=50, p=8 \sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



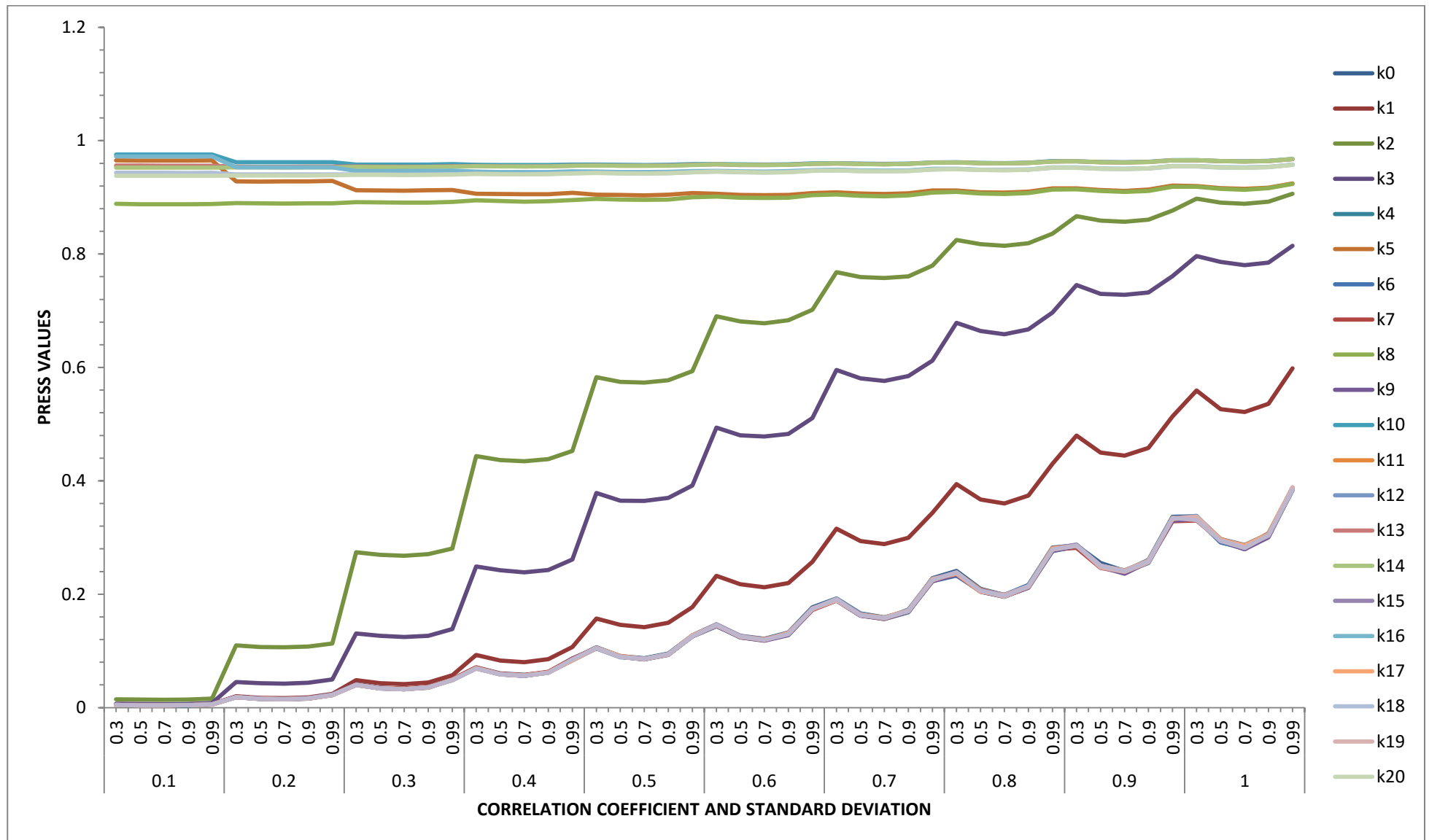


Figure 4.49 Graph of Simulated values of PRESS for  $n=50$ ,  $p=9$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



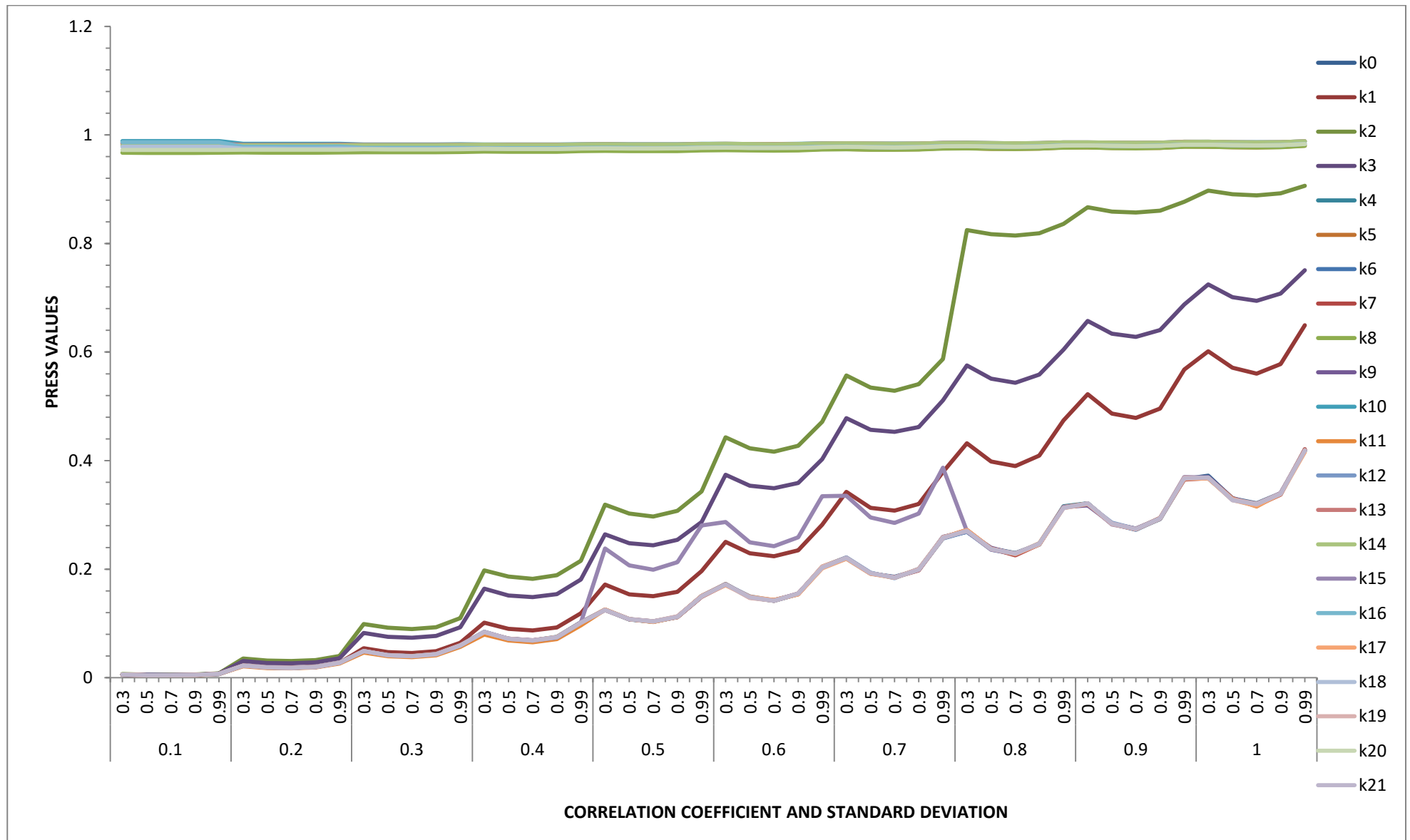


Figure 4.50 Graph of Simulated values of PRESS for  $n=80$ ,  $p=3$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

Table 4.8(b) Simulated PRESS for  $n=80$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





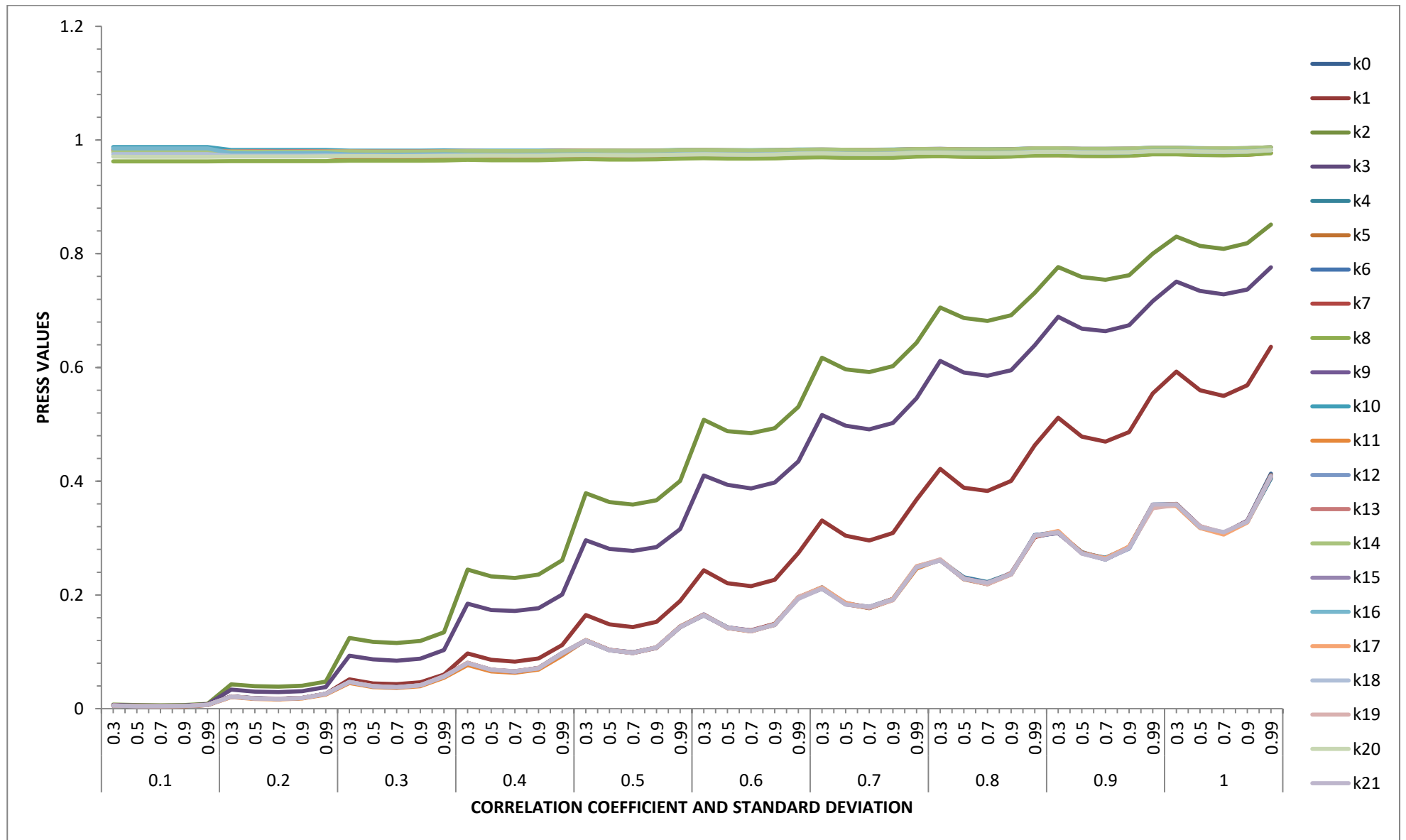


Figure 4.51 Graph of Simulated values of PRESS for  $n=80$ ,  $p=4$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



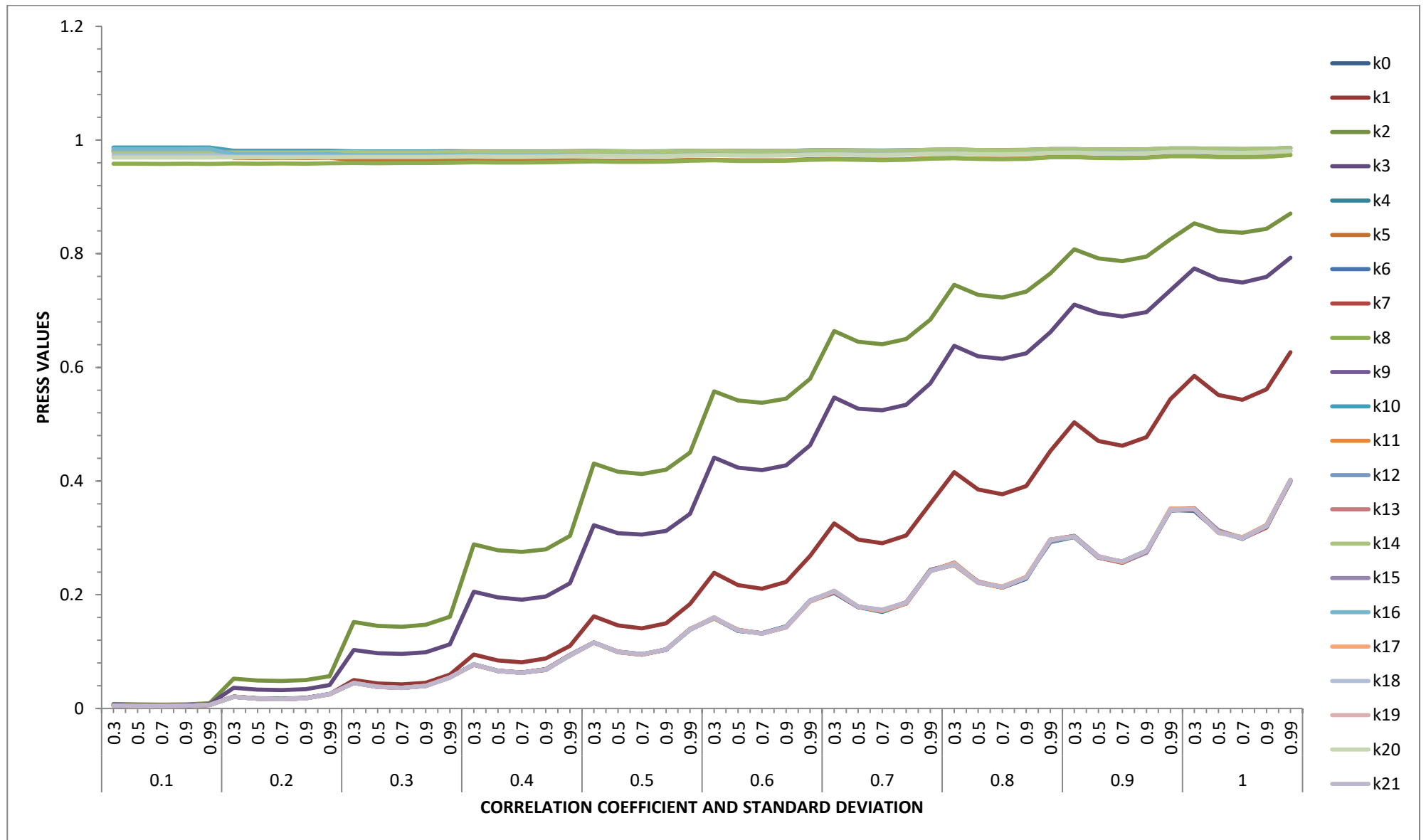


Figure 4.52 Graph of Simulated values of PRESS for  $n=80$ ,  $p=5$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



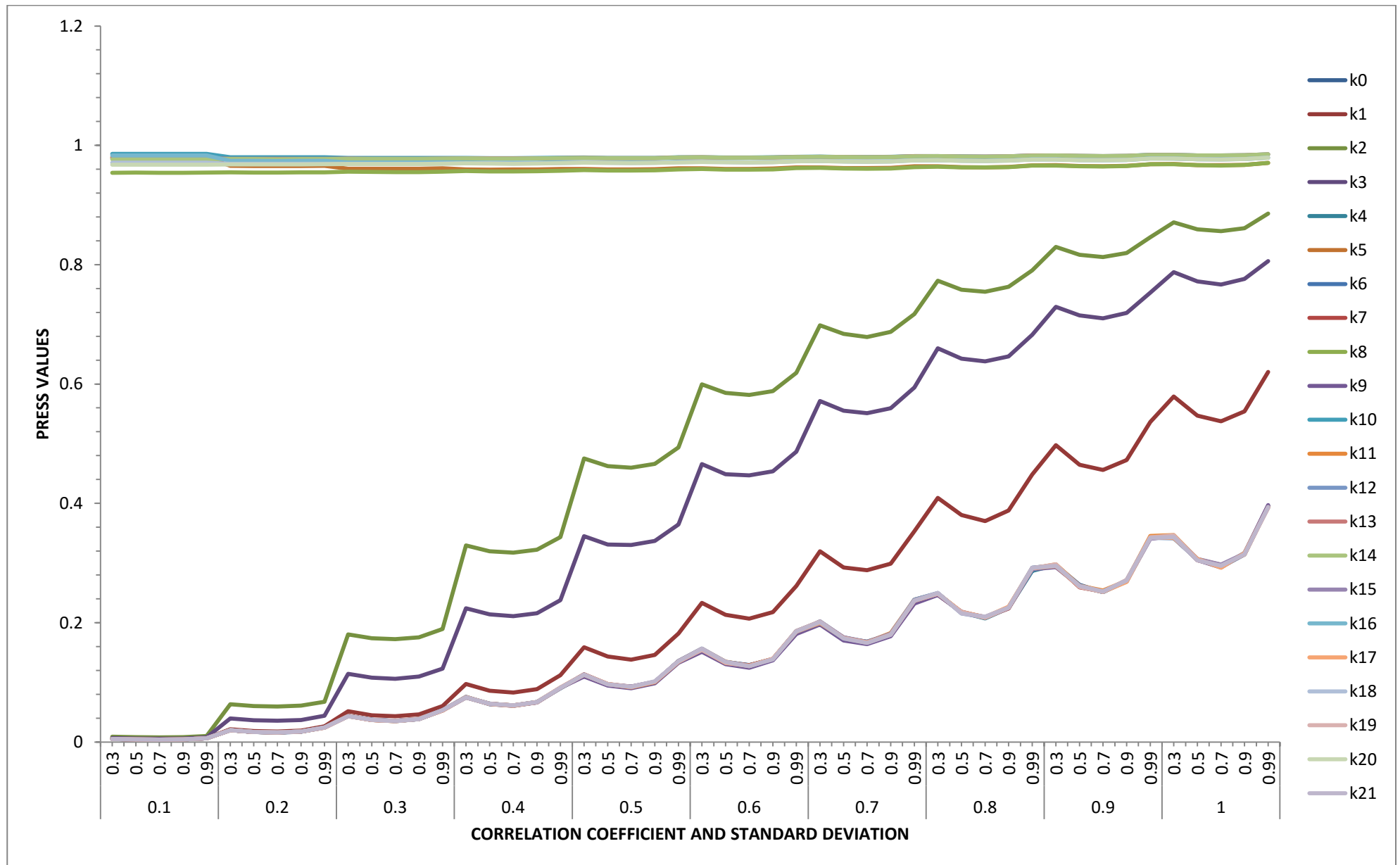


Figure 4.53 Graph of Simulated values of PRESS for  $n=80, p=6$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



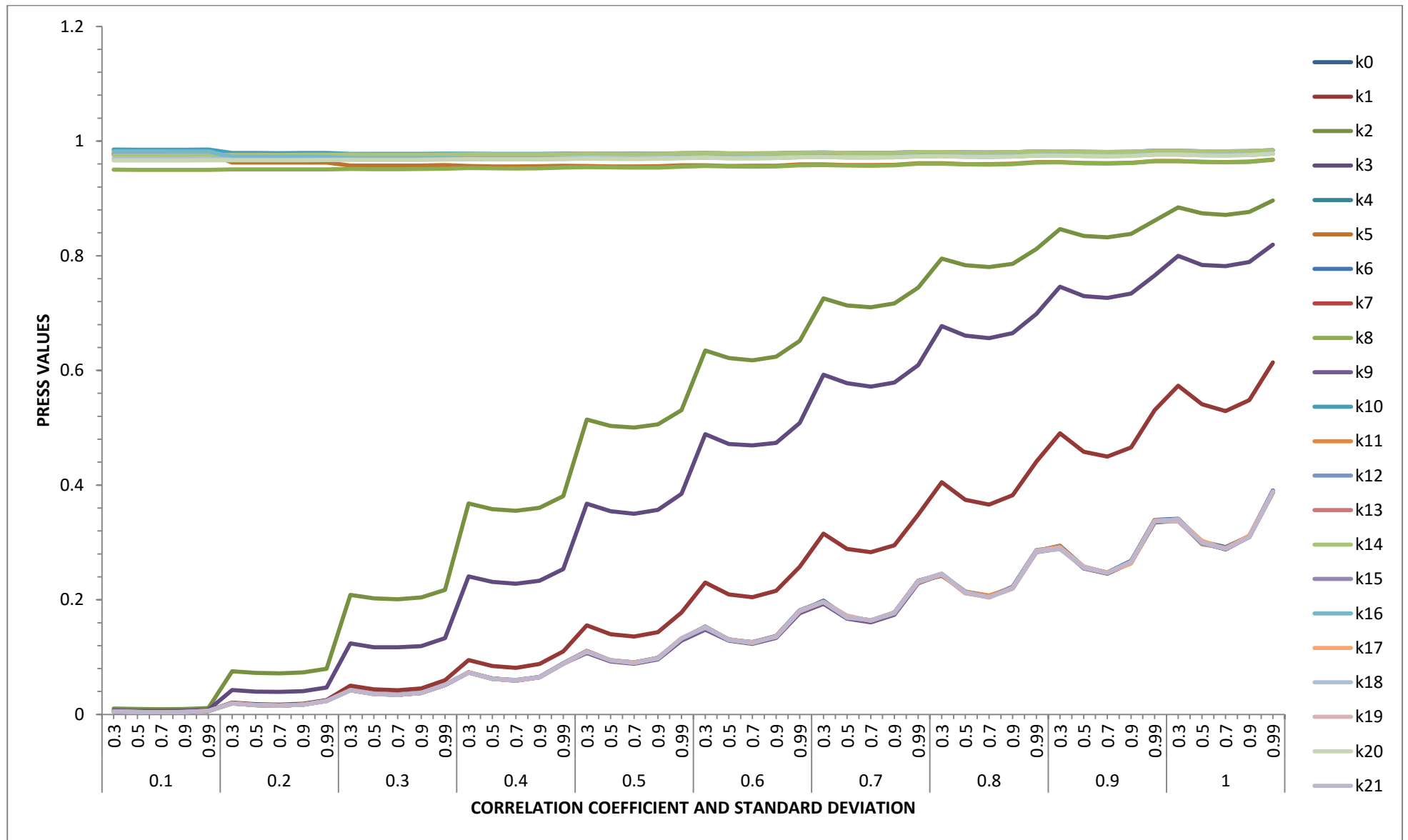


Figure 4.54 Graph of Simulated values of PRESS for  $n=80$ ,  $p=7$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$





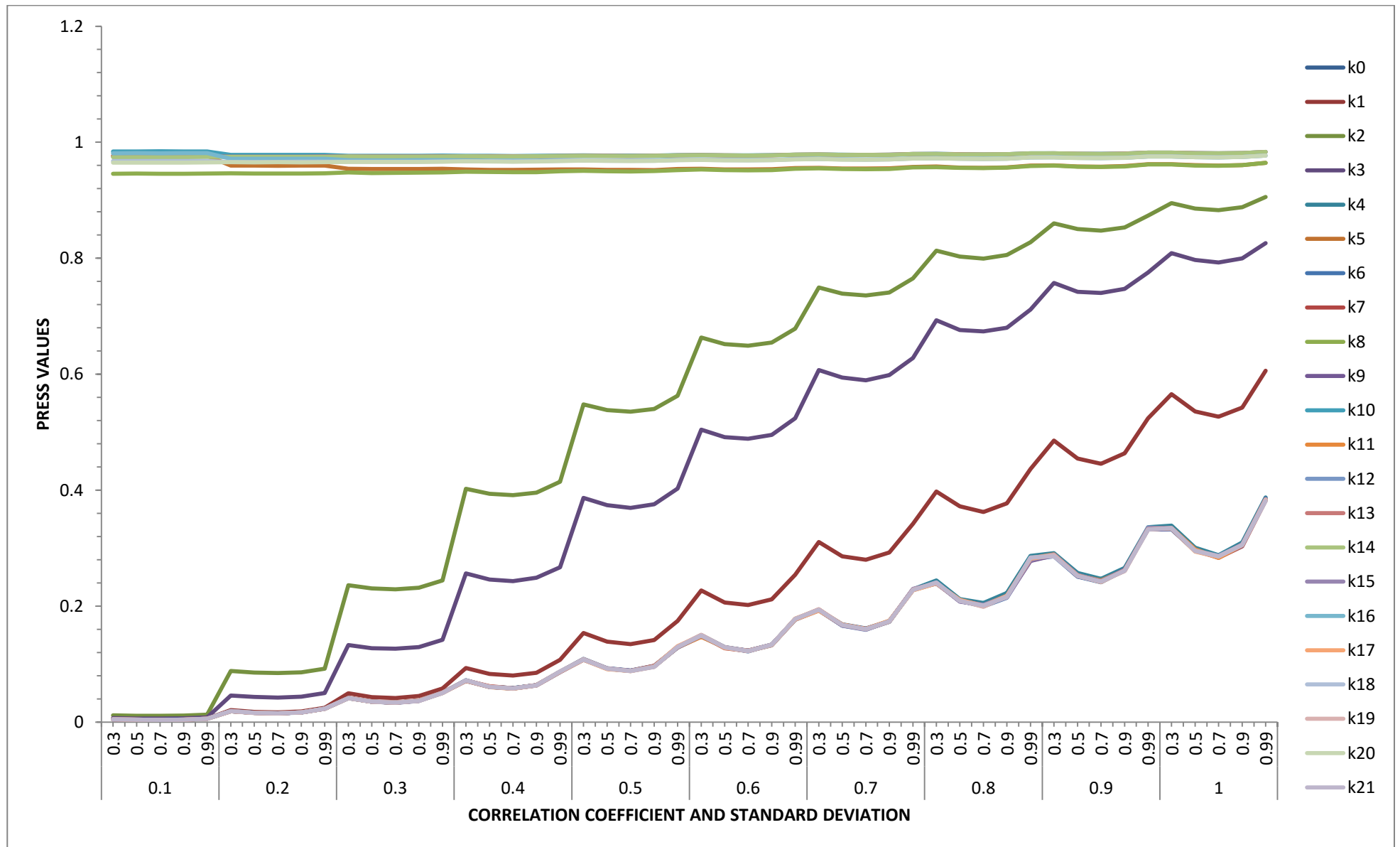


Figure 4.55 Graph of Simulated values of PRESS for  $n=80$ ,  $p=8$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$



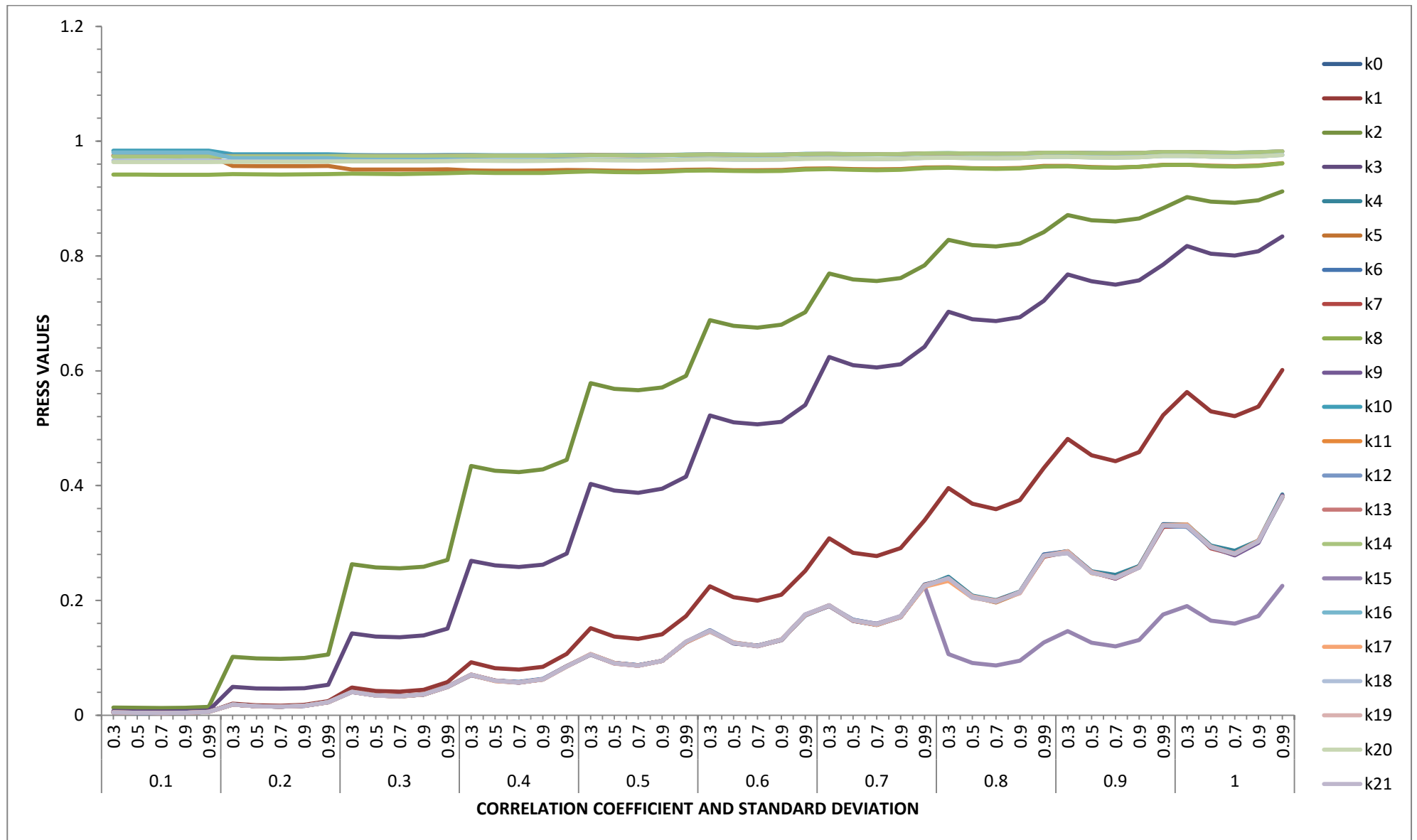


Figure 4.56 Graph of Simulated values of PRESS with  $n=80$ ,  $p=9$   $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$  and  $\rho = 0.3, 0.5, 0.7, 0.9$  and  $0.99$

**APPENDIX X**  
**TABLE NIGERIAN GDP AND OTHER ECONOMIC VARIABLES**

<b>GDP</b>	<b>MONEY SUPPLY</b>	<b>CRED PRIV. SECTOR</b>	<b>EXCHANGE RATE</b>	<b>EXTERNAL RESERV</b>	<b>AGRICLOAN</b>	<b>FOREIGN TRADE</b>	<b>OIL IMPORT</b>	<b>NONOIL IMPORT</b>	<b>OILEXPORT</b>	<b>NONOIL EXPort</b>
51732	14471	8570	0.61	56195	35642	23863	120	12720	10681	343
53659	15787	10668	0.673	12324	31764	18977	226	10545	8003	203
57963	17688	11668	0.724	7171	36308	16406	172	8732	7201	301
64326	20106	12463	0.765	5480	24655	16266	282	6896	8841	247
73542	22299	13070	0.894	10998	44244	18783	52	7011	11224	497
74542	23806	15247	2.021	18922	68417	14904	914	5070	8369	552
111913	27574	21083	4.018	62554	102153	48222	3170	14692	28209	2152
147941	38357	27326	4.537	72267	118611	52639	3803	17643	28435	2757
228451	45903	30403	7.392	43953	129300	88831	4672	26189	55017	2954
281550	52857	33548	8.038	40293	98494	155604	6073	39645	106627	3260
329071	75401	41352	9.909	48620	82107	211024	7772	81716	116858	4677
555446	111112	58123	17.298	33392	88032	348763	19562	123590	201384	4227
715242	165339	127118	22.051	58824	80846	384400	41136	124493	213779	4991
945557	230293	143424	21.886	95329	103186	3688480	42350	120439	200710	5349
2008564	289091	180005	21.886	32345	164162	1705789	155826	599302	927565	23096
2799036	345854	238597	21.886	25896	225503	1872170	162179	400448	1286216	23328
2906625	413280	316207	21.886	73492	242038	2087379	166903	678814	1212499	29163
2816406	488146	351956	21.886	93777	215697	1589275	175854	661565	717787	34070
3312241	628952	431168	92.693	63709	246083	2051486	211662	650854	1169477	19493
4717332	878457	530373	102.105	91089	361450	2930746	220818	764205	1920900	24823
4909526	12699322	764962	111.943	123330	728545	3226134	237107	1121074	1839945	28009
7128203	1508173	930494	120.97	103104	1051590	3256873	361710	1150985	1649446	94732
8742647	1952922	1096536	129.356	91702	1164460	5168122	398922	1681313	2993110	94776
11673602	2131820	1421664	133.5	144753	2083745	6589827	318115	1668931	4489472	113309
1.48E+08	2637914	1838390	132.147	291849	3046739	10047391	797299	2003557	7140579	105956
18709786	3799538	2290618	128.651	449473	4263060	10433200	710683	2397836	7191086	133595
20874172	5138701	3680090	125.833	544732	4425862	12221711	768227	3143726	8110500	199258
25424948	8029089	6941383	118.566	701675	6721075	15357293	1386730	3803073	9913651	247839
2896746	9456480	9147417	148.902	536428	8349509	13458920	1063544	4038990	8067233	289153
3124539	11034941	10157021	150.298	448268	7740508	19041169	2073579	5931795	10639417	396377

## APPENDIX XI

### COMPUTER PROGRAMME IN R FOR THE ESTIMATION OF REGRESSION PRARMETERS USING OLS AND RIDGE REGRESSION ESTIMATORS, MSE AND PRESS OF THE ESTIMATORS

GDP=c(51731.81,53658.95,57963.31,64326.36,73542.06,74542.06,111912.92,147941.12,228451.46,281550.26,329070.74,555445.51,715241.91,945557.02,2008564.02,2799036.11,2906624.88,2816405.99,3312240.87,4717332.08,4909526.48,7128203.11,8742646.65,11673602.22,147735323.15,18709786.49,20874172.36,25424947.72,2896746.21,3124538.98)

MONEYSUPPLY=c(14471.17,15786.74,17687.93,20105.94,22299.24,23806.40,27573.58,38356.80,45902.88,52857.03,75401.18,111112.31,165338.75,230292.60,289091.07,345853.96,413280.13,488145.79,628952.16,878457.27,12699321.61,1508172.91,1952922.28,2131820.08,2637913.73,3799538.05,5138700.94,8029088.61,9456480.31,11034940.93)

#### CREDITTOPRIVATESECTOR

=c(8570.05,10668.34,11668.04,12462.93,13070.34,15247.45,21082.99,27326.42,30403.22,33547.70,41352.46,58122.95,127117.71,143424.21,180004.96,238596.56,316207.08,351956.19,431168.36,530373.30,764961.52,930493.93,1096535.57,1421664.03,1838389.93,2290617.76,3680090.19,6941383.41,9147417.17,10157021.18)

EXCHANGERATE=c(0.61,0.6729,0.7241,0.7649,0.8938,2.0206,4.0179,4.5367,7.3916,8.0378,9.9095,17.2984,22.0511,21.8861,21.8861,21.8861,21.8861,21.8861,92.6934,102.1052,111.9433,120.970,129.356,133.500,132.147,128.651,125.833,118.566,148.9017,150.298)

EXTERNALRESERV=c(56194.8,12324.3,7171.4,5479.7,10997.7,18922.05,62554.27,72266.81,43953.22,40293.2,48620.04,33391.95,58824.15,95329.02,32345.01,25895.59,73492.11,93776.73,63709.2,91089.2,123329.83,103104.08,91701.66,144753.06,291849.31,449473.06,544731.68,701674.6,536428.25,448268.46)

AGRICLOAN=c(35642.4,31763.9,36307.5,24654.9,44243.6,68417.4,102152.5,118611.0,129300.3,98494.4,82107.4,88031.8,80845.8,103186.0,164162.1,225502.5,242038.2,215697.2,246082.5,361450.4,728545.4,1051589.8,1164460.4,2083744.7,3046738.5,4263060.3,4425861.8,6721074.6,8349509.3,7740507.6)

#### FOREIGNTRADE

=c(23862.9,18976.9,16406.2,16266.3,18783.4,14904.2,48222.3,52638.5,88831.4,155604.0,211023.6,348762.5,384399.5,3688480.0,1705789.1,1872170.0,2087379.3,1589275.4,2051485.5,2930745.7,3226134.2,3256873.0,5168121.7,6589826.8,10047391.1,10433200.0,12221711.0,15357292.7,13458920.0,19041168.8)

OILIMPORT=c(119.8,225.5,171.6,282.4,51.8,913.9,3170.1,3803.1,4671.6,6073.1,7772.2,19561.5,41136.1,42349.6,155825.9,162178.7,166902.5,175854.2,21

```

1661.8,220817.69,237106.83,361710.0,398922.31,318114.72,797298.94,71068
3,768226.84,1386729.93,1063544.18,2073579.03)
NONOILIMPORT=c(12719.8,10545,8732.1,6895.9,7010.8,5069.7,14691.6,176
42.6,26188.6,39644.8,81716.0,123589.7,124493.3,120439.2,599301.8,400447.
9,678814.1,661564.5,650853.9,764204.7,1121073.5,1150985.33,1681312.96,1
668930.55,2003557.39,2397836.32,3143725.79,3803072.68,4038990.2,593179
5.19)
OILEXPORT=c(10680.5,8003.2,7201.2,8840.6,11223.7,8368.5,28208.6,28435.
4,55016.8,106626.5,116858.1,201383.9,213778.8,200710.2,927565.3,1286215.
9,1212499.4,717786.5,1169476.9,1920900.4,1839945.25,1649445.82,2993109.
95,4489472.19,7140578.92,7191085.64,8110500.38,9913651.13,8067233.0,10
639417.3)
NONOILEXPORT=c(342.8,203.2,301.3,247.4,497.1,552.1,2152,2757.4,2954.4
,3259.6,4677.3,4227.3,4991.3,5349,23096.1,23327.5,29163.3,34070.2,19492.9,
24822.9,28008.6,94731.85,94776.44,113309.4,105955.9,133595,199257.9,247
839,289152.6,396377.2)
y=cbind(GDP)
y
x=cbind(MONEYSUPPLY,CREDITTOPRIVATESECTOR,EXCHANGERAT
E,EXTERNALRESERV,AGRICLOAN,FOREIGNTRADE,OILIMPORT,NON
OILIMPORT,OILEXPORT,NONOILEXPORT)
x
X=matrix(0,30,10)
L=0
for (j in 1:10) L[j]= sqrt(sum((x[,j]-mean(x[,j]))^2))
for (j in 1:10) X[,j]=(x[,j]-mean(x[,j]))/L[j]
X
Ly=sqrt(sum((y-mean(y))^2))
Y=(y-mean(y))/Ly
Y
B=solve((t(X)%*%X))%*%(t(X)%*%Y)
B
reg=lm(Y~X-1)
bbeta=reg$coef
bbeta
anova(reg)
summary(reg)
Q=cbind(resid(reg))
Q
Beta_r=cbind(bbeta) #The regression coefficients
Beta_r
A=solve(t(X)%*%X)
A

```

```

DD=eigen(t(X)%*%X)
DD
D=DD$vectors #eigenvectors of X'X
D%*%t(D) #Just to confirm that this is equal to I_p
t(D)%*%(t(X)%*%X)%*%D
for (j in 1:10)
lambda_j=DD$values
lambda_j #lambda here is the same as the diagonal element of
t(D)%*%(t(X)%*%X)%*%D
M=median(lambda_j)
M
alpha_j=t(D)%*%bbeta
alpha_j
n=30
p=10
sigma_hat=0.072 #Use other values later
T_j=alpha_j^2
T_j
max(T_j)
S=0
for (j in 1:p) S_j=((sigma_hat^2)*((prod(lambda_j))^(1/p)))/(((n-
p)*(sigma_hat^2))+((prod(lambda_j))^(1/p))*(T_j)))
S_j
#Formula 21
k=median(S_j)
k
I=diag(p)
W=k*I
W
Z=(A+W)
Z
Beta_k=Z%*%(t(X)%*%Y)
Beta_k
frac1_j=frac2_j=0
for(j in 1:10){
frac1_j=lambda_j/((lambda_j+k)^2)
sum(frac1_j)
frac2_j=((alpha_j)^2)/((lambda_j+k)^2)
frac2_j
}
mse=((sigma_hat^2)*sum(frac1_j))+((k^2)*sum(frac2_j))
mse
#To obtain PRESS

```

```

Hat=X%*%solve(t(X)%*%X)%*%t(X)
Hat
V=cbind(1-diag(Hat))
V
Y_hat=X%*%Beta_k
Y_hat
e=Y-Y_hat
e
PRESS=sum((e/V)^2)
PRESS
MSE
k=0
#Formula 1
k=(sigma1^2)/max(alpha^2)
#Formula 2
k=(p*(sigma_hat^2))/(t(alpha_j)%*%alpha_j)
k
#Formula 3
k=(p*(sigma_hat^2))/sum(lambda_j*(T_j))
k
#Formula 4
k=(max(lambda_j)*(sigma_hat^2))/((n-
p)*(sigma_hat^2)+max(lambda_j)*max(T_j))
k
q=0
for (j in 1:p) q_j=(max(lambda_j)*(sigma_hat^2))/((n-
p)*(sigma_hat^2)+(max(lambda_j)*(T_j)))
q_j
#Formula 5
k=max(1/q_j)
k
#Formula 6
k=max(q_j)
k
#Formula 7
k=(prod(1/q_j))^(1/p)
k
#Formula 8
k=median(1/q_j)
k
#formula 9
k=median(q_j)
k

```



```

L=log(2)
L
k
w=0
for (j in 1:p) w_j=(L*(sigma_hat^2))/((n-p)*(sigma_hat^2)+(L*(T_j)))
w_j
#Formula 10
k=max(1/w_j)
k
#Formula 11
k=max(w_j)
k
#Formula 12
k=(prod(1/w_j))^(1/p)
k
#Formula 13
k=(prod(w_j))^(1/p)
k
#Formula 14
k=median(1/w_j)
k
#Formula 15
k=median(w_j)
k
S=0
for (j in 1:p) S_j=((sigma_hat^2)*((prod(lambda_j))^(1/p)))/(((n-
p)*(sigma_hat^2))+((prod(lambda_j))^(1/p))*(T_j)))
S_j
#Formula 16
k=max(1/S_j)
k
#Formula 17
k=max(S_j)
k
#Formula 18
k=(prod(1/S_j))^(1/p)
k
#Formula 19
k=(prod(S_j))^(1/p)
k
#Formula 20
k=median(1/S_j)
k

```

#Formula 21  
 $k = \text{median}(S_j)$   
k