A PENALTY FUNCTION IN BINARY LOGISTIC REGRESSION

Piyada Phrueksawatnon

A Dissertation Submitted in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Statistics)
School of Applied Statistics
National Institute of Development Administration
2018

A PENALTY FUNCTION IN BINARY LOGISTIC REGRESSION

Piyada Phrueksawatnon

School of Applied Statistics

Professor..Major Advisor
(Jirawan Jitthavech, Ph.D.)

Associate Professor..Co-Advisor
(Vichit Lorchirachoonkul, Ph.D.)

The Examining Committee Approved This Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Dortor of Philosophy (Statistics).

Associate Professor..Committee Chairperson
(Supol Durongwatana, Ph.D.)

Professor..Committee
(Samruam Chongcharoen, Ph.D.)

Professor..Committee
(Jirawan Jitthavech, Ph.D.)

Associate Professor..Committee
(Vichit Lorchirachoonkul, Ph.D.)

Assistant Professor..Dean
(Pramote Luenam, Ph.D.)

June 2019

ABSTRACT

Title of Dissertation	A Penalty Function in Binary Logistic Regression
Author	Miss Piyada Phrueksawatnon
Degree	Doctor of Philosophy (Statistics)
Year	2018

An algorithm is proposed to determine the logistic ridge parameter minimizing the MSE of the estimated parameter estimators, together with a theorem on the upper-bound of the optimal logistic ridge parameter to facilitate the nonlinear optimization. A simulation is used to evaluate the relative efficiencies of the proposed estimator and other six well-known ridge estimators with respect to the maximum likelihood estimator. The simulation results confirm that the relative efficiency of the proposed estimator is highest among other well-known estimators. Finally, a real-life data set is used to repeat the evaluation and the conclusion is the same as in the simulation.

ACKNOWLEDGEMENTS

First and foremost, I offer my sincerest gratitude to my advisor Professor Dr.Jirawan Jitthavech who has supported me throughout this study and my co-advisor Associate Professor Dr.Vichit Lorchirachoonkul who has provided much assistance so that I can overcome the difficulties in my dissertation – without both of them this dissertation would not have been completed.

I am grateful to my dissertation committee members: Associate Professor Dr.Supol Durongwatana, Professor Dr.Samruam Chongcharoen, Professor Dr.Jirawan Jitthavech, and Associate Professor Dr.Vichit Lorchirachoonkul for their suggestions. I feel delighted to extend gratitude towards all of my friends, especially for whom at the Graduate School of Applied Statistics, National Institute of Development Administration (NIDA), for making my time full of happiness and energy. I gratefully acknowledge the School of Applied Statistics, NIDA for providing the facilities, and the Ministry of Science and Technology, Thailand and the University of Payao for financial support.

Finally, I would like to thank my family for their love and support; their love is always inside me.

Piyada Phrueksawatnon

May 2019

TABLE OF CONTENTS

LIST OF TABLES

Tables **Page**

LIST OF FIGURES

Figures **Page**

CHAPTER 1

INTRODUCTION

This chapter is organized as follows. The background on multicollinearity is introduced in Section 1.1, and the objective, scope, and usefulness of the study are presented in Sections 1.2–1.4.

1.1 Background

The problem of multicollinearity (near-multicollinearity in this dissertation) occurs when the explanatory variables are highly correlated. This problem is common in applied research and leads to high variance and unstable parameter estimates when estimating both ordinary least squares (OLS) in linear regression and the maximum likelihood (ML) estimator in logistic regression. Both $\mathbf{X'X}$ in linear regression and $\mathbf{X'VX}$ in logistic regression are ill-conditioned matrices that are near singularity, and they directly affect the performance of the estimator. Moreover, they result in undesirable asymptotic properties in a logistic regression such as large variances, which can cause a lack of statistical significance in the test for an individual predictor even when the overall model is strongly significant (e.g. Schaefer, 1986; Marx and Smith, 1990; Mansson and Shukur, 2011; Ogoke, Nduka and Nja, 2013).

Penalized regression methods are the most effective and popular to remedy the multicollinearity problem. The common concept of these is a tradeoff between the variances and biases of the parameter estimates whereby the penalization yields regression coefficients with lower variances but higher biases than in the unpenalized model. Ridge regression is a very popular method but offers more efficient estimates that may be biased (De Grange, Fariña and De Dios Ortúzar, 2015). Moreover, an important obstacle in ridge regression is the selection of a ridge estimator which does not have an exact criterion, and so many researchers have proposed methods to estimate the ridge parameter.

In this dissertation, finding the optimal value of the ridge parameter without approximation is the goal, thus an extended study of the penalized function of a binary logistic regression in the presence of multicollinearity is of interest. This leads to the construction of a new estimator by applying penalization using a penalized ML estimator instead of the standard method when the data are multicollinear among the explanatory variables.

1.2 Objectives of the Study

In this study, a penalty function in binary logistic regression is considered with the following objectives as:

1) To propose a new estimator in binary logistic regression in present of multicollinearity
2) To investigate the properties of the proposed estimator

1.3 Scope of the Study

In this study, the proposed estimator is derived based on multiple logistic regression with a binary outcome in presence of the multicollinearity under the following scopes:

1) Assume that some explanatory variables have high correlation levels, while the other explanatory variables are independent or have low correlation levels.
2) The data are assumed to have no missing values

1.4 Usefulness of the Study

The proposed estimator can be applied in many fields such as clinical trials, medicine, biomedicine, biostatistics, health sciences, social sciences, finance, economics, engineering, and even politics for classification problems or for predicting the probability of an interesting event. For example, in a clinical trial, logistic regression can be utilized to classify the type of cancer cells based on genetic data, while in politics, it might be used to predict whether a voter will vote for an interested political party based on personal information such as age, income, gender, abode, voting in previous elections, and so on.

CHAPTER 2

LITERATURE REVIEW

In this study, the area of interest is a binary logistic regression model in the presence of multicollinearity. This chapter starts with a review of the development of the binary logistic model and its assumptions (Section 2.1) followed by a review of the estimating parameter for the logistic regression model (Section 2.2). Next, a review of the problem of multicollinearity in logistic regression is provided in Section 2.3, and finally, a review of penalty regression is presented in Section 2.4.

2.1 The Development of the Logistic Regression Model

Logistic regression is commonly used to model the probability of a binary outcome when given explanatory variables of interest: these can be either categorical or continuous. In addition to the features of the dependent variable, the difference between logistic and linear regression can both be reflected in a model and in the assumptions. This provides the conditional mean of the regression model in the range 0 and 1 for exhibiting a change in the conditional mean per unit in an explanatory variable. Hence, it is widely applied in many fields, especially in clinical trials, medicine, health sciences, etc. because a model with an S-shaped curve could be used to describe the combined effect of several risk factors on the risk of, for instance, contracting a disease. The maximum likelihood (ML) method is used to estimate parameters in the linear predictor (e.g. Hosmer and Lemeshow, 1989, 2000: 4-10). For logistic regression modeling, binary logistic regression where the response data can take one of two possible values (0 or 1) is used in the present study.

Consider the linear regression model

$$y_i = \mathbf{w}_i'\boldsymbol{\beta} + \varepsilon_i, \tag{2.1}$$

where $\mathbf{w}_i' = \left(\mathbf{1}, \mathbf{w}_{i1}, \mathbf{w}_{i2}, \dots, \mathbf{w}_{ip}\right)$; $\mathbf{w}_{ij}$ is a $p \times 1$ vector of the centered and scaled explanatory variables, where $w_{ij} = \dfrac{x_{ij} - \bar{x}_j}{SS_j^{1/2}}$; $i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$; $SS_j = \sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2$; x_{ij} is the observed value in unit i of explanatory variable j ; $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \beta_2, \dots, \beta_p\right)'$ is a $p \times 1$ standardized regression parameter vector; y_i is the response in unit i; and ε_i is the error for unit i, $\varepsilon_i \sim N(0, \sigma^2)$ .Moreover, Schaefer (1979:3) mentioned "without loss of generality".

The expectation of y_i in (2.1) is

$$E(\mathrm{y}_i) = \mathbf{w}_i'\boldsymbol{\beta}. \tag{2.2}$$

In the case where dependent variable y_i is a dummy variable with only two possible values (0 or 1), the expectation of y_i becomes

$$\begin{aligned} E(y_i) &= 1 \cdot P(y_i = 1) + 0 \cdot P(y_i = 0), \\ &= P(y_i = 1). \end{aligned} \tag{2.3}$$

From (2.1) and (2.3), this implies that

$$E(y_i \mid \mathbf{w}_i) = \mathbf{w}_i'\boldsymbol{\beta} = P(y_i = 1 \mid \mathbf{w}_i). \tag{2.4}$$

Equation (2.4) is called the "linear probability model" and is defined in terms of a linear predictor. The variance of y_i can be written as

$$
\begin{aligned}
Var(y_i) &= E[y_i - E(y_i)]^2, \\
&= \left[1 - P(y_i = 1)\right]^2 P(y_i = 1) + \left[-P(y_i = 1)\right]^2 P(y_i = 0), \\
&= \left[1 - P(y_i = 1)\right]^2 P(y_i = 1) + \left[P(y_i = 1)\right]^2 \left[1 - P(y_i = 1)\right], \\
&= P(y_i = 1)\left[1 - P(y_i = 1)\right]\left[1 - P(y_i = 1) + P(y_i = 1)\right], \\
&= P(y_i = 1)\left[1 - P(y_i = 1)\right]. \qquad (2.5)
\end{aligned}
$$

Since a regression model has the important assumption of ε_i and if a regression analysis is applied with dependent variable y_i is a dummy variable (0 or 1), then ε_i can be either of two values. When considering (2.1), if $y_i = 1$, then $\varepsilon_i = 1 - \mathbf{w}_i'\boldsymbol{\beta}$ with $P(y_i = 1)$, and if $y_i = 0$, then $\varepsilon_i = -\mathbf{w}_i'\boldsymbol{\beta}$ with $P(y_i = 0)$. Therefore,

$$
\begin{aligned}
E(\varepsilon_i) &= \left(-\mathbf{w}_i'\boldsymbol{\beta}\right) P(y_i = 0) + \left(1 - \mathbf{w}_i'\boldsymbol{\beta}\right) P(y_i = 1), \\
&= \left(-\mathbf{w}_i'\boldsymbol{\beta}\right)\left[1 - P(y_i = 1)\right] + \left(1 - \mathbf{w}_i'\boldsymbol{\beta}\right) P(y_i = 1), \\
&= -P(y_i = 1)\left[1 - P(y_i = 1)\right] + \left[1 - P(y_i = 1)\right] P(y_i = 1), \\
&= 0, \qquad (2.6)
\end{aligned}
$$

and

$$
\begin{aligned}
Var(\varepsilon_i) &= \left(-\mathbf{w}_i'\boldsymbol{\beta}\right)^2 P(y_i = 0) + \left(1 - \mathbf{w}_i'\boldsymbol{\beta}\right)^2 P(y_i = 1), \\
&= \left[-P(y_i = 1)\right]^2 \left[1 - P(y_i = 1)\right] + \left[1 - P(y_i = 1)\right]^2 P(y_i = 1), \\
&= P(y_i = 1)\left[1 - P(y_i = 1)\right]\left[P(y_i = 1) + 1 - P(y_i = 1)\right], \\
&= P(y_i = 1)\left[1 - P(y_i = 1)\right]. \qquad (2.7)
\end{aligned}
$$

From (2.6) and (2.7), it can be concluded that the error has a distribution with zero mean and variance $P(y_i = 1)\left[1 - P(y_i = 1)\right]$ in which its variance depends on the values of the explanatory variables. Thus, when assuming that ε_i is not true, which results in the estimator, then the error assumption is not true. Therefore, when the outcome variable is dichotomous in logistic regression, the error term violates the assumptions of homoscedasticity and normality. Using the OLS method yields an unbiased estimator, but the estimated variance of the estimator is not the smallest. Thus, the standard errors in the presence of heteroscedasticity will be incorrect and any test for significance will be invalid. Therefore, we need to fit the regression model when the dependent variable is dichotomous to find the predicted values of the response $(\hat{\mathbf{y}})$ necessary using the linear probability model in (2.4). The value of $\mathbf{w}_i'\boldsymbol{\beta}$ is the probability estimate of $P(y_i = 1)$ (i.e. $P(y_i = 1) = \pi(\mathbf{w}_i)$), which is in the interval 0 and 1 $\left(0 < P(y_i = 1) = \pi(\mathbf{w}_i) < 1\right)$. However, the estimate of $\mathbf{w}_i'\boldsymbol{\beta}$ might be below 0 or above 1 depending on the range of values of the explanatory variables, i.e. $-\infty < \mathbf{w}_i'\boldsymbol{\beta} < \infty$. To improve this problem, $\pi(\mathbf{w}_i)$ needs to be transformed into odds and the natural log of the odds for eliminating the ceiling of 1 and the floor of 0. To consider the interval of probability $\pi(\mathbf{w}_i)$, $0 < \pi(\mathbf{w}_i) < 1$ by taking the odds, we have $0 < \dfrac{\pi(\mathbf{w}_i)}{1-\pi(\mathbf{w}_i)} < \infty$, and then by taking the natural log of the odds, $-\infty < \ln\left[\dfrac{\pi(\mathbf{w}_i)}{1-\pi(\mathbf{w}_i)}\right] < \infty$. Hence, we can obtain $-\infty < \ln\left[\dfrac{\pi(\mathbf{w}_i)}{1-\pi(\mathbf{w}_i)}\right] < \infty$ and $-\infty < \mathbf{w}_i'\boldsymbol{\beta} < \infty$.

There is a function for linking between the distribution of $\mathbf{y}$ and the linear predictor in logistic regression and is known as the “logit link”:

$$\operatorname{logit}(\pi(\mathbf{w}_i)) = \ln\left[\frac{\pi(\mathbf{w}_i)}{1-\pi(\mathbf{w}_i)}\right]. \tag{2.8}$$

Therefore,

$$\begin{aligned} \ln\left[\frac{\pi(\mathbf{w}_i)}{1-\pi(\mathbf{w}_i)}\right] &= \mathbf{w}_i'\boldsymbol{\beta}, \qquad (2.9) \\ \frac{\pi(\mathbf{w}_i)}{1-\pi(\mathbf{w}_i)} &= \exp(\mathbf{w}_i'\boldsymbol{\beta}), \end{aligned}$$

$$1+\frac{\pi(\mathbf{w}_i)}{1-\pi(\mathbf{w}_i)} = 1+\exp(\mathbf{w}_i'\boldsymbol{\beta}),$$

$$\frac{1}{1-\pi(\mathbf{w}_i)} = 1+\exp(\mathbf{w}_i'\boldsymbol{\beta}),$$

$$1-\pi(\mathbf{w}_i) = \frac{1}{1+\exp(\mathbf{w}_i'\boldsymbol{\beta})},$$

$$\pi(\mathbf{w}_i) = \frac{\exp(\mathbf{w}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{w}_i'\boldsymbol{\beta})}.$$

The equation (2.9) is called “log odds” or “logit”.The logistic regression model is defined as

$$P(\mathrm{y}_i=1\,|\,\mathbf{w_i}) = \frac{\exp(\mathbf{w}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{w}_i'\boldsymbol{\beta})} = \pi(\mathbf{w}_i),\quad 0<\pi(\mathbf{w_i})<1,\quad \mathrm{i}=1,2,\ldots,\mathrm{n}. \tag{2.10}$$

Hereinafter, the assumptions of the logistic regression model (Midi, Sarkar and Rana, 2010; Jirawan Jitthavech, 2015: 326-327) are:

1) y_i is Bernoulli distribution, $y_i \in \{0,1\}$, $i=1,2,\ldots,n.$
2) y_i, $i=1,2,\ldots,n$ is independent.
3) w_{ij}, $i=1,2,\ldots,n$, $j=1,2,\ldots,p$ are not linear combinations of each other.
4) Errors are Bernoulli distributed.
5) No important variables are omitted.
6) No extraneous variables are included.
7) The explanatory variables are measured without error.

2.2 The Estimating Parameter for the Logistic Regression Model

As discussed earlier, the error term has neither a normal distribution nor equal variances for the explanatory variable values, so the parameter estimates from OLS would give inefficient estimates. Instead of the OLS method, an ML estimator can be used to estimate the regression coefficient in a logistic regression model. The objective of this estimation is to find a set value for parameter $\boldsymbol{\beta}$ that maximizes the likelihood function. In a very general sense, the ML method yields values for the unknown parameters which maximize the probability of obtaining the observed set of data (Hosmer and Lemeshow, 2000: 8). The likelihood function for Equation (2.5) can be written as

$$L(\boldsymbol{\beta}|\, y_i) = \prod_{i=1}^{n} \pi(w_i)^{y_i}[1-\pi(w_i)]^{1-y_i} \ , \quad i = 1, 2, \ldots, n. \tag{2.11}$$

However, for convenience in mathematical calculations, it is easier to work with the logarithm of Equation (2.11), and so the log-likelihood function can be defined as

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \{ y_i \ln \pi(w_i) + (1-y_i)\ln[1-\pi(w_i)] \}, \\ &= \sum_{i=1}^{n} \{ y_i \ln \pi(w_i) + \ln[1-\pi(w_i)] - y_i \ln[1-\pi(w_i)] \}, \\ &= \sum_{i=1}^{n} \left\{ y_i \ln\left[\frac{\pi(w_i)}{1-\pi(w_i)} \right] + \ln[1-\pi(w_i)] \right\}, \\ &= \sum_{i=1}^{n} \{ y_i (\mathbf{w}_i'\boldsymbol{\beta}) - \ln[1+\exp(\mathbf{w}_i'\boldsymbol{\beta})] \}. \end{aligned} \tag{2.12}$$

Equation (2.12) is differentiated with respect to $\beta_0, \beta_1, \ldots, \beta_p$, and the first derivative of (2.12) is in the form

$$\begin{aligned} l'(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} &= \sum_{i=1}^{n} \left\{ y_i w_{ij} - \frac{\exp(\mathbf{w}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{w}_i'\boldsymbol{\beta})} \cdot w_{ij} \right\}, \\ &= \sum_{i=1}^{n} y_i w_{ij} - \sum_{i=1}^{n} \frac{\exp(\mathbf{w}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{w}_i'\boldsymbol{\beta})} \cdot w_{ij}, \\ &= \sum_{i=1}^{n} y_i w_{ij} - \sum_{i=1}^{n} \pi(\mathbf{w}_i) \cdot w_{ij}, \end{aligned}$$

$$l'(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_{ij} \left[y_i - \pi(\mathbf{w}_i) \right]. \tag{2.13}$$

This implies the matrix form

$$l'(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{W}' \left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w}) \right], \tag{2.14}$$

where

$$\mathbf{W} = \begin{bmatrix} 1 & w_{11} & w_{12} & \cdots & w_{1p} \\ 1 & w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & w_{n1} & w_{n1} & \cdots & w_{np} \end{bmatrix} = \left[\mathbf{1} \quad \mathbf{w}_1 \quad \mathbf{w}_2 \quad \ldots \quad \mathbf{w}_n \right]',$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\pi}(\mathbf{w}_i) = \begin{bmatrix} \pi(\mathbf{w}_1) \\ \pi(\mathbf{w}_2) \\ \vdots \\ \pi(\mathbf{w}_n) \end{bmatrix}, \quad \mathbf{V} = diag\left(\pi(\mathbf{w}_i)\left(1 - \pi(\mathbf{w}_i)\right)\right).$$

Moreover, to find the optimal $\boldsymbol{\beta}$, the derivative equations are set to zero, thus the likelihood score equations can be expressed as

$$\mathbf{W}' \left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w}) \right] = \mathbf{0}. \tag{2.15}$$

The second derivative of (2.12) is in the form

$$\begin{aligned}
l''(\boldsymbol{\beta}) &= \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_j \partial\beta_k} \\
&= -\sum_{i=1}^{n} \frac{\partial}{\partial\beta_k}\left\{w_{ij}\left[y_i - \pi(\mathbf{w}_i)\right]\right\}, \\
&= -\sum_{i=1}^{n} w_{ij}\left\{\frac{\left[1+\exp(\mathbf{w}_i'\boldsymbol{\beta})\right]\exp(\mathbf{w}_i'\boldsymbol{\beta})w_{ik} - \exp(\mathbf{w}_i'\boldsymbol{\beta})\exp(\mathbf{w}_i'\boldsymbol{\beta})w_{ik}}{\left[1+\exp(\mathbf{w}_i'\boldsymbol{\beta})\right]^2}\right\}, \\
&= -\sum_{i=1}^{n} w_{ij}w_{ik}\left\{\frac{\left[1+\exp(\mathbf{w}_i'\boldsymbol{\beta}) - \exp(\mathbf{w}_i'\boldsymbol{\beta})\right]\exp(\mathbf{w}_i'\boldsymbol{\beta})}{\left[1+\exp(\mathbf{w}_i'\boldsymbol{\beta})\right]^2}\right\}, \\
&= -\sum_{i=1}^{n} w_{ij}w_{ik}\left\{\frac{\exp(\mathbf{w}_i'\boldsymbol{\beta})}{\left[1+\exp(\mathbf{w}_i'\boldsymbol{\beta})\right]^2}\right\}, \\
&= -\sum_{i=1}^{n} w_{ij}w_{ik}\left\{\pi(\mathbf{w}_i)\left[1-\pi(\mathbf{w}_i)\right]\right\}. \qquad (2.16)
\end{aligned}$$

Equation (2.16) implies the matrix form

$$l''(\boldsymbol{\beta}) = \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = -\mathbf{W}'\mathbf{V}\mathbf{W}, \qquad (2.17)$$

where V is a diagonal matrix with elements $v_i = \pi(\mathbf{w}_i)\left[1-\pi(\mathbf{w}_i)\right]$, $i = 1, 2, \ldots, n$.

The most common technique for estimating parameter $\boldsymbol{\beta}$ is the ML method. From normalized Equation (2.15), the closed form of estimate $\boldsymbol{\beta}$ cannot be found, but the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ can be obtained by using an iteratively reweighted least-squares algorithm. The value of $\hat{\boldsymbol{\beta}}$ at the $(t+1)^{st}$ iteration in the Newton-Raphson method (Hosmer and Lemeshow, 1989) is given by the value of $\hat{\boldsymbol{\beta}}$ at the t^{th} iteration as

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{t+1} &= \hat{\boldsymbol{\beta}}^{t} + \left[-l''(\boldsymbol{\beta})\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{t}}\right]^{-1}\left(l'(\boldsymbol{\beta})\big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{t}}\right),\\ &= \hat{\boldsymbol{\beta}}^{t} + \left(\mathbf{W}'\hat{\mathbf{V}}^{t}\mathbf{W}\right)^{-1}\mathbf{W}'\left[\mathbf{y}-\hat{\boldsymbol{\pi}}^{t}(\mathbf{w})\right],\\ &= \left(\mathbf{W}'\hat{\mathbf{V}}^{t}\mathbf{W}\right)^{-1}\left[\left(\mathbf{W}'\hat{\mathbf{V}}^{t}\mathbf{W}\right)\hat{\boldsymbol{\beta}}^{t} + \mathbf{W}'\left(\mathbf{y}-\hat{\boldsymbol{\pi}}^{t}(\mathbf{w})\right)\right],\\ &= \left(\mathbf{W}'\hat{\mathbf{V}}^{t}\mathbf{W}\right)^{-1}\mathbf{W}'\hat{\mathbf{V}}^{t}\left[\mathbf{W}\hat{\boldsymbol{\beta}}^{t} + \left(\hat{\mathbf{V}}^{t}\right)^{-1}\left(\mathbf{y}-\hat{\boldsymbol{\pi}}^{t}(\mathbf{w})\right)\right],\end{aligned} \quad (2.18)$$

where t denotes the iteration step, $\mathbf{W} = [\mathbf{1} \quad \mathbf{w}_1 \quad \mathbf{w}_2 \quad \dots \quad \mathbf{w}_n]'$ is an $n \times p$ observation matrix in which the i^{th} row is $\mathbf{w}_i$, $\hat{\mathbf{V}}^t$ is an $n \times n$ diagonal matrix in which the diagonal element is $\hat{\boldsymbol{\pi}}^t(\mathbf{w})\left[1-\hat{\boldsymbol{\pi}}^t(\mathbf{w})\right]$, and $\hat{\boldsymbol{\pi}}^t(\mathbf{w})$ is an $n \times 1$ vector of the i^{th} element of the estimated $\boldsymbol{\pi}(\mathbf{w})$ value at the t^{th} iteration. Obviously, $\hat{\boldsymbol{\beta}}^{t+1}$ in (2.18) can be written as

$$\hat{\boldsymbol{\beta}}^{t+1} = \left(\mathbf{W}'\hat{\mathbf{V}}^{t}\mathbf{W}\right)^{-1}\mathbf{W}'\hat{\mathbf{V}}^{t}\hat{\mathbf{z}}^{t}, \quad (2.19)$$

where $\hat{\mathbf{z}}^t = \mathbf{W}\hat{\boldsymbol{\beta}}^t + \left(\hat{\mathbf{V}}^t\right)^{-1}\left[\mathbf{y}-\hat{\boldsymbol{\pi}}^t(\mathbf{w})\right]$.

When convergence is obtained, $\hat{\boldsymbol{\beta}}^{t+1}$ becomes the ML estimator $\hat{\boldsymbol{\beta}}_{ML}$ (Schaefer, 1979; Hosmer and Lemeshow, 2000; Rashid and Shifa, 2009; Okeh and Oyeka, 2013; Al Turk and Alsomahi, 2014):

$$\hat{\boldsymbol{\beta}}_{ML} = \lim_{t\to\infty}(\mathbf{W}'\hat{\mathbf{V}}^{t}\mathbf{W})^{-1}\mathbf{W}'\hat{\mathbf{V}}^{t}\hat{\mathbf{z}}^{t} = (\mathbf{W}'\hat{\mathbf{V}}\mathbf{W})^{-1}\mathbf{W}'\hat{\mathbf{V}}\hat{\mathbf{z}}. \quad (2.20)$$

In the studies by Schaefer (1979) and Lee and Silvapulle (1988), the authors found that the category explanatory variables might result in the MLE not existing (if there is a perfect explanatory variable of the outcome). To overcome this problem, the model can be rebuilt without this category of explanatory variable, which results in the

MLE always existing. Moreover, there should be enough reason to exclude this variable in the model.

Under certain regular conditions (Cox and Hinkley, 1974; Rashid, 2008; Rashid and Shifa, 2009), as n increases, $\hat{\boldsymbol{\beta}}_{ML}$ asymptotically approaches $\boldsymbol{\beta}$ and is distributed as $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}\right) \sim N\left(\mathbf{0}, (\mathbf{W}'\hat{\mathbf{V}}\mathbf{W})^{-1}\right)$ (Schaefer, 1979; Lee and Silvapulle, 1988; Marx, 1988; Duffy and Santner, 1989; Akay, 2014). The asymptotic mean square error (MSE) of $\hat{\boldsymbol{\beta}}_{ML}$ is defined as

$$MSE(\hat{\boldsymbol{\beta}}_{ML}) = E\left[(\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})\right] = tr\left(\mathbf{W}'\hat{\mathbf{V}}\mathbf{W}\right)^{-1} = \sum_j \frac{1}{\lambda_j}, \quad (2.21)$$

where $\lambda_j \geq 0,\ j = 0, 1, 2, \ldots, p,$ is the j^{th} eigenvalue of the semi-definite matrix $\mathbf{W}'\hat{\mathbf{V}}\mathbf{W}$ (Schaefer, 1979; Marx, 1988; Marx and Smith, 1990).

2.3 The Problem of Multicollinearity in Logistic Regression

In the presence of multicollinearity, two or more explanatory variables are highly correlated, thus multicollinearity can be defined as the nearly linear dependence of the column of $\mathbf{W}$ which violates the assumption of the logistic regression (2.10). In a logistic regression model, Schaefer (1979) demonstrated that multicollinearity affects the $\mathbf{W}'\mathbf{V}\mathbf{W}$ matrix in the same way as in the $\mathbf{W}'\mathbf{W}$ matrix in multiple linear regression.

Theorem 2.1: If matrix $\mathbf{W}$ is near singularity, then $\mathbf{W}'\mathbf{V}\mathbf{W}$ is an ill-conditioned matrix where $\mathbf{V}$ is nonsingular.

Proof. Let $\boldsymbol{\Sigma}$ be a variance-covariance matrix of $\mathbf{Y}$ and assume that $\mathbf{u}$ is any non-zero $n \times 1$ column vector, $\mathbf{u} \in R^n$. By definition, $\boldsymbol{\Sigma} = E\left[\left(\mathbf{Y} - E[\mathbf{Y}]\right)\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\right]$, then

$$\mathbf{u}'\boldsymbol{\Sigma}\mathbf{u} = E\left[\left\{\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\mathbf{u}\right\}'\left\{\left(\mathbf{Y} - E[\mathbf{Y}]\right)'\mathbf{u}\right\}\right] = E\left[s^2\right] \geq 0, \quad s = \left(\mathbf{Y} - E[\mathbf{Y}]\right)'\mathbf{u}\cdot$$

Let $\mathbf{V} = diag(\mathbf{\Sigma}) = diag\left(\pi(\mathbf{w}_i)\left(1-\pi(\mathbf{w}_i)\right)\right)$; $0 < \pi(\mathbf{w}_i) < 1$. Since $\mathbf{\Sigma}$ is positive definite, then $\mathbf{V}$ is positive definite as well as being nonsingular, thus it can be written as $\mathbf{V} = \mathbf{V}^{\frac{1}{2}}\mathbf{V}^{\frac{1}{2}}$. Consider $\mathbf{W}'\mathbf{V}\mathbf{W} = \mathbf{W}'\mathbf{V}^{\frac{1}{2}}\mathbf{V}^{\frac{1}{2}}\mathbf{W} = \mathbf{L}'\mathbf{L}$, where $\mathbf{L} = \mathbf{V}^{-\frac{1}{2}}\mathbf{W}$, and we have $r(\mathbf{W}) = r(\mathbf{L})$. When two or more explanatory variables in a logistic regression model are highly correlated, we have $r(\mathbf{W}) < p$, and thus $r(\mathbf{L}) < p$. Hence, $\left(\mathbf{L}'\mathbf{L}\right)^{-1}$ does not exist and $\mathbf{L}'\mathbf{L}$ is called an ill-conditioned matrix.

Since the semi-definite matrix $\mathbf{W}'\mathbf{V}\mathbf{W}$ is not full rank, there is near singularity in the $\mathbf{W}'\mathbf{V}\mathbf{W}$ matrix (Marx and Smith, 1990; Vágó and Kemény, 2006), resulting in the problem in the inverse matrix $(\mathbf{W}'\mathbf{V}\mathbf{W})^{-1}$ because of $\det\left(\mathbf{W}'\mathbf{V}\mathbf{W}\right) = \prod_{j=0}^{p} \lambda_j \cong 0$ and $\lambda_j \cong 0$ for some j. This may induce imprecision in the MLEs since $\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{W}'\hat{\mathbf{V}}\mathbf{W})^{-1}\mathbf{W}'\hat{\mathbf{V}}\hat{\mathbf{z}}$. The covariance estimate of regression coefficients are inflated, leading to large standard errors which affect the confidence intervals and hypothesis testing (Hoerl and Kennard, 1970). If the degree of multicollinearity becomes more severe, there are common characteristics in the identification of multicollinearity (e.g. Schaefer, 1979; Schaefer, Roi and Wolfe, 1984; Schaefer, 1986):

i) The subsidiary or the auxiliary regression (the regression of each explanatory variable (w_i) on the remaining explanatory variables with computing corresponding R^2 (Gujarati and Porter, 2010)), R_j^2, tends to one for some j.

ii) The sum of squared residuals from the regression model in (i) tends to zero.

iii) The smallest eigenvalue tends to zero.

If there are one or more near-linear dependences in the data, then one or more eigenvalues in $\mathbf{W}'\hat{\mathbf{V}}\mathbf{W}$ will be small, which means that the variance of the corresponding regression coefficient will be large. Therefore, when the degree of multicollinearity is more intense, one or more of the estimates will be unstable and the estimates may not reflect the true effect of the explanatory variables. In practice, multicollinearity inflates the estimated variances of the ML estimator and thus can cause precision problems when identifying the effects of the explanatory variables

(Schaefer, Roi and Wolfe, 1984). Ryan (1997) selected an indicator of multicollinearity in a logistic regression. If the explanatory variables are all continuous, then pairwise correlation and variance inflation factors may be used, but if some explanatory variables are not continuous, one possible check for multicollinearity in qualitative variables is the kappa measure of agreement.

2.4 Penalized Regression

In the presence of multicollinearity or when having a large number of predictors, the ML method often realizes unstable estimates and inaccurate variances of the logistic regression associated with parameter estimates and certain prediction regions. There are several corrections for dealing with the multicollinearity problem. Penalized regression methods are one of the most effective and popular methods to remedy the multicollinearity problem. The common concept of penalized regression is a tradeoff between the variances and biases of the parameter estimates. The penalization yields regression coefficients with lower variances than in an unpenalized model, but it gives high biases in the regression coefficients. Hence, for fitting the logistic regression, the model is penalized by adding a penalty function to its likelihood function. The penalty term provides biased penalized estimates but can also substantially reduce the variance. Moreover, depending on the form of the penalty, it allows us to carry out variable selection as well as shrinkage of the estimates (Crotty and Barker, 2014). There are two ways that penalization methods can assist as variable selection and shrinkage. The penalty criterion is in the form

$$l^{p}(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - kP(\boldsymbol{\beta}), \tag{2.22}$$

where $P(\cdot)$ is a penalty function on the parameter and k is penalty parameter. The estimate of $\hat{\boldsymbol{\beta}}$ is the maximization of Equation (2.22):

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \left\{ l^{p}(\boldsymbol{\beta}) \right\}.$$

A proper penalty function should produce an estimator having three properties accordingly avoiding excessive bias (unbiasedness), forcing sparse solutions to reduce the model complexity (sparsity), and satisfying certain conditions to produce a continuous model (stability). These conditions imply that the penalty function must be singular at the origin and nonconvex over $(0,\infty)$ (Antoniadis and Fan, 2001). Moreover, the penalty function should be selected based on principles that can solve the optimization problem easily.

Many penalty functions have been proposed in several articles. For example, the L_1 penalty, $P(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$, results in a LASSO (least absolute shrinkage and selection operator) which can generate sparse models that are easily interpretable. However, when the explanatory variables comprise a category, the LASSO solution can exhibit an undesirable feature by selecting only dummy variables instead of whole factors (Meier, van de Geer and Bühlmann, 2008; Makalic and Schmidt, 2011). Moreover, the empirical observations indicate that if there are many high correlations between explanatory variables, the prediction performance of LASSO is dominated by ridge regression (Tibshirani, 1996). Next, the L_2 penalty (or quadratic penalty), $P(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_2^2$, yields ridge-type regression, which is nevertheless a popular method. Therefore, in this work, the focus is on ridge regression in a logistic regression model, the details of which are explained next.

2.4.1 Logistic Ridge Regression

Ridge regression in multiple regressions for improving the problem of multicollinearity was first proposed by Hoerl and Kennard (1970). Next, Schaefer, Roi and Wolfe (1984) derived a ridge-type estimator by applying the idea of the ridge parameter by Hoerl and Kennard (1970) in multiple linear regression to logistic regression. They defined logistic ridge regression (LRR) as an estimator that minimizes the length of the estimate of $\boldsymbol{\beta}$. The ridge logistic estimator obtained by Schaefer et al. (1984) is defined as

$$\hat{\boldsymbol{\beta}}_{LRR}(k) = (\mathbf{W}'\mathbf{V}\mathbf{W} + k\mathbf{I}_p)^{-1}(\mathbf{W}'\mathbf{V}\mathbf{W})\hat{\boldsymbol{\beta}}_{ML}. \tag{2.23}$$

However, in Equation (2.23), $\mathbf{V}$ depends on the unknown parameter $\boldsymbol{\beta}$, so in any application, they proposed using

$$\hat{\boldsymbol{\beta}}_{LRR}(k) = (\mathbf{W}'\hat{\mathbf{V}}\mathbf{W} + k\mathbf{I}_p)^{-1}(\mathbf{W}'\hat{\mathbf{V}}\mathbf{W})\hat{\boldsymbol{\beta}}_{ML}, \tag{2.24}$$

where $\hat{\mathbf{V}}$ is an estimate of $\mathbf{V}$ using $\hat{\boldsymbol{\beta}}_{ML}$ and the ridge parameter $k \geq 0$ determines the amount of shrinkage. The present work references the logistic ridge estimator of Schaefer et al. (1984) because the MSE of the estimator is of interest.

Later, Le Cessie and Houwellingen (1992) applied ridge regression to logistic regression for correcting parameter estimates and decreasing prediction errors. This method adds an L_2 penalty parameter to the likelihood function such that coefficients are shrunk individually according to the variance of each explanatory variable. By following the concept of the penalized ridge regression, the penalized log-likelihood becomes a combination of the log-likelihood $l(\boldsymbol{\beta})$ in Equation (2.11) and a penalty function of L_2 norm of $\boldsymbol{\beta}$ expressed as (Duffy and Santner, 1989; Le Cessie and Van Houwelingen, 1992)

$$l_{LRR}(\boldsymbol{\beta}) \;=\; l(\boldsymbol{\beta}) - \frac{k}{2}\boldsymbol{\beta}'\boldsymbol{\beta}, \tag{2.25}$$

where k is penalty parameter (called "ridge parameter").

The logistic ridge estimator can by found by maximizing Equation (2.25):

$$\hat{\boldsymbol{\beta}}_{LRR} = \arg\max_{\boldsymbol{\beta}} \left(l(\boldsymbol{\beta}) - \frac{k}{2}\boldsymbol{\beta}'\boldsymbol{\beta} \right). \tag{2.26}$$

The first derivative of (2.25) is in the form

$$l'_{LRR}(\boldsymbol{\beta}) = \frac{\partial l_{LRR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right] - k\boldsymbol{\beta}. \tag{2.27}$$

The second derivative of (2.25) is in the form

$$l''_{LRR}(\boldsymbol{\beta}) = \frac{\partial^2 l_{LRR}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\mathbf{W}'\mathbf{V}\mathbf{W} - k\mathbf{I}, \tag{2.28}$$

where $\mathbf{I}$ is a $(p\times 1)\times(p\times 1)$ identity matrix.

The Newton-Raphson method is the conventional iterative method to solve (2.25) by updating the parameter vector in the Newton-Raphson step; Lee and Van Houwelingen (1992) proposed a first approximation for the ridge logistic estimator as

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{(1)}_{LRR}(k) &= \boldsymbol{\beta}_0 + \left[-l''_{LRR}(\boldsymbol{\beta})\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right]^{-1} \left(l'_{LRR}(\boldsymbol{\beta})\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right), \\
&= \boldsymbol{\beta}_0 + \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} + k\mathbf{I} \right)^{-1} \left\{ \mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right] - k\boldsymbol{\beta}_0 \right\}, \\
&= \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} + k\mathbf{I} \right)^{-1} \left\{ \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} + k\mathbf{I} \right)\boldsymbol{\beta}_0 + \mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right] - k\boldsymbol{\beta}_0 \right\}, \\
&= \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} + k\mathbf{I} \right)^{-1} \left\{ \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} \right)\boldsymbol{\beta}_0 + \mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right] \right\}, \\
&= \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} + k\mathbf{I} \right)^{-1} \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} \right) \left\{ \boldsymbol{\beta}_0 + \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} \right)^{-1} \mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right] \right\}, \\
&= \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} + k\mathbf{I} \right)^{-1} \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} \right) \hat{\boldsymbol{\beta}},
\end{aligned} \tag{2.29}$$

where $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W} \right)^{-1} \mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right]$ with the real parameter value $\boldsymbol{\beta}_0$ and the $\mathbf{V}$ matrix evaluated at $\boldsymbol{\beta}_0$. Replacing $\mathbf{W}'\mathbf{V}(\boldsymbol{\beta}_0)\mathbf{W}$ by its estimate $\mathbf{W}'\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W}$ in (2.29) yields the logistic ridge estimator of Schaefer et al. (1984) in (2.24); note that if the ML estimator is infinite, this estimator does not exist (Lee and Van Houwelingen, 1992; Özkale, 2016). The logistic ridge estimator of Schaefer et al. (1984) is mentioned in the present work.

Conventionally, the parameter vector is estimated in two stages to satisfy (2.25). The first stage is to propose a closed-form estimator to approximate the logistic ridge parameter k that fulfills some of the criteria to reduce the total variance of the parameter estimator. The second stage is to estimate the parameter vector $\boldsymbol{\beta}$ in accordance with (2.26) based on the approximate value of the logistic ridge parameter k in the first stage.

The MSE of the LRR estimator is derived as

$$
\begin{aligned}
MSE(\hat{\boldsymbol{\beta}}_{LRR}) &= E\left[(\hat{\boldsymbol{\beta}}_{LRR} - \boldsymbol{\beta})' (\hat{\boldsymbol{\beta}}_{LRR} - \boldsymbol{\beta}) \right], \\
&= E\left[(\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})' \mathbf{A}' \mathbf{A} (\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}) \right] + \boldsymbol{\beta}' (\mathbf{A} - \mathbf{I})' (\mathbf{A} - \mathbf{I}) \boldsymbol{\beta}, \\
&= \sum_j \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \boldsymbol{\beta}' \left(\mathbf{W}' \hat{\mathbf{V}} \mathbf{W} + k\mathbf{I} \right)^{-2} \boldsymbol{\beta}', \\
&= \phi_1(k) + \phi_2(k), \qquad (2.30)
\end{aligned}
$$

where $\mathbf{A} = \left(\mathbf{W}' \hat{\mathbf{V}} \mathbf{W} + k\mathbf{I} \right)^{-1} \mathbf{W}' \hat{\mathbf{V}} \mathbf{W}$ and λ_j is the j^{th} element of the eigenvalues in $\mathbf{W}' \hat{\mathbf{V}} \mathbf{W}$, and the ridge parameter $k \geq 0$ determines the amount of shrinkage. The MSE of the LRR estimator in Equation (2.30) is established from the two important parts: $\phi_1(k)$ and $\phi_2(k)$. The total variance in (2.30), $\phi_1(k)$, is a continuous monotone decreasing function of k, while the second term in (2.30), $\phi_2(k)$, is the squared bias of the LRR estimator.

2.4.2 The Selection of Ridge Parameter

Estimating the ridge parameter is an essential problem for ridge regression, and plenty of researchers have proposed various techniques to achieve this. Nevertheless, the ridge regression method does not provide a unique estimate of the estimator to solve the multicollinearity problem since there are no definite fixed rules to select the ridge parameter. Consequently, to find the optimal value of ridge parameter k without estimation is the main focus in this dissertation.

The main problem of ridge regression is to find the optimal value of ridge parameter k, which is of interest in this work. To achieve this, a compromise between the two ideas of fitting the model of dependent variables on the explanatory variables and shrinking the regression coefficients is sought. The ridge parameter controls the amount of shrinkage, so a larger ridge parameter is shrunk more when the sum of squares of the coefficients is small. Thus, if k approaches infinity, all estimated coefficients tend toward zero. Although it has been pointed out in a large number of studies that the value of ridge parameter k varies in the interval $[0,\infty)$, ridge regression cannot improve the estimated regression coefficients for all cases. To find the value of k, Conniffe and Stone (1973) believed that a suitable value provides estimates of the regression coefficients that are stable with reasonable absolute values and the correct sign. More specifically, a value of k is sought such that the logistic ridge estimator has a lower mean square error (MSE) than the original estimator. Since the MSE of the ridge estimator is a function of variance (the decreasing function of k) and the squared bias (the increasing function of k). Therefore, a value of k must be chosen such that the variance term of the estimator is greater than the increase in the squared bias term. However, the choice of the optimum k is not well-defined.

There have been several researches that focused on different ways to examine the logistic ridge parameter in logistic regression (e.g. Schaefer et. al., 1984; Schaefer, 1986; Lee and Silvapulle, 1988; Le Cessie and Van Houwelingen, 1992; Kibria, Mansson and Shukur, 2012; Özkale and Arıcan, 2016; Asar, 2017; Asar, Arashi and Wu, 2017). However, they struggled to find a closed-form solution for the optimal ridge parameter k based on the data. In the past, since computers were limited when processing large amounts of data, it was necessary to approximate the value of k. However, nowadays, computers have much better performance and science has progressed accordingly, and so finding the closed form of k has become less important. Therefore, finding the optimal real value of the ridge parameter in logistic regression can be carried out without approximating its value, which is the objective of the present study.

It can be seen that as the asymptotic variance decreases, the squared bias becomes large when k increases. Therefore, the objective of logistic regression is to choose a value of k such that the reduction in the variance term is greater than the increase in the squared bias. In multiple regression models, the MSE of ridge regression varies depending on the value of k having the range $0<k<k_{max}$, which affects the MSE of ridge regression less than that of OLS (Hoerl and Kennard, 1970). Several researches have mainly focused on different ways to examine the ridge parameter (e.g.

Hoerl and Kennard, 1970; Kibria, 2003; Khalaf and Shukur, 2005; Alkhamisi, Khalaf and Shukur, 2006; Muniz and Kibria, 2009; Muniz, Kibria, Mansson and Shukur, 2012). Later, Schaefer et al. (1984) first extended the ridge regression in a logistic regression and demonstrated that LRR outperforms ML when the explanatory variables are multicollinearity, which was later supported by Schaefer (1986) and Mansson and Shukur (2011).

Many studies have investigated ridge parameter k under multiple linear regression models (e.g. Hoerl and Kennard, 1970; Hoerl, Kennard and Baldwin 1975; Kibria, 2003; Khalaf and Shukur, 2005; Alkhamisi, Khalaf and Shukur, 2006; Muniz et al., 2012; Dorugade, 2014).

Mansson and Shukur (2011) investigated a suitable value for ridge parameter k; they advised that the ridge parameter based on Kibria (2003) might be the best option when the degree of correlation between the explanatory variables is $0.75 \le \rho < 0.95$, while the ridge parameter based on Muniz and Kibria (2009) might be proper when the degree of correlation between the explanatory variables is high ($0.95 \le \rho < 1$). They subsequently proposed a generalized M (GM) estimator.

However, Muniz and Kibria (2009) found that besides the ridge parameter affecting MSE, other factors influence the estimated MSE in multiple linear regressions, namely the correlation between the explanatory variables, sample size, the standard deviation of the errors, and the number of replications. When the standard deviation of the errors increases, the higher correlation between the explanatory variables brings about an increase in MSE whereas an increase in sample size causes a decrease. Later on, a number of researchers studied the determinants impacting the estimate of MSE in logistic regression models consisting of the degree of correlation between the explanatory variables, the value of the intercept, the number of observations, and the number of explanatory variables (e.g. Mansson and Shukur, 2011; Kibria et al., 2012).

In summary, six estimators of the logistic ridge parameter are compared with the proposed value of the ridge parameter

1) The HK estimator (Hoerl and Kennard, 1970)

$$\hat{k}_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2} \tag{2.31}$$

2) The HKB estimator (Hoerl et al., 1975)

$$\hat{k}_{HKB} = \frac{(p+1)\hat{\sigma}^2}{\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}} \tag{2.32}$$

3) The SRW1 and SRW2 estimators (Schaefer et al., 1984)

$$\hat{k}_{SRW1} = \frac{1}{\hat{\alpha}_{\max}^2} \tag{2.33}$$

$$\hat{k}_{SRW2} = \frac{(p+1)}{\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}} \tag{2.34}$$

4) The GM estimator (Kibria, 2003)

$$\hat{k}_{GM} = \frac{\hat{\sigma}^2}{\left(\prod_{j=0}^{p} \hat{\alpha}_j^2\right)^{1/(p+1)}} \tag{2.35}$$

5) The WA estimator (Wu and Asar, 2016)

$$\hat{k}_{WA} = \frac{(p+1)}{\sum_{j=0}^{p} \alpha_j^2 \Big/ \left[1 + \left(1 + \lambda_j \alpha_j^2\right)^{1/2}\right]} \tag{2.36}$$

In these equations, $\hat{\sigma}^2$ is the variance of the residual in the model and $\hat{\alpha}_j$ and $\hat{\alpha}_{\max}$ are the respective j^{th} and maximum elements of vector $\hat{\boldsymbol{\alpha}} = \boldsymbol{\gamma}'\hat{\boldsymbol{\beta}}_{ML}$, where $\boldsymbol{\gamma}$ is the orthogonal transformation such that $\mathbf{W}'\mathbf{V}\mathbf{W} = \boldsymbol{\gamma}'\boldsymbol{\Lambda}\boldsymbol{\gamma}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues of $\mathbf{W}'\mathbf{V}\mathbf{W}$ (Schaefer, 1979; Schaefer et al., 1984; Marx, 1988; Kibria et al., 2012).

CHAPTER 3

THE PROPOSED ESTIMATOR

In this chapter, an estimator for logistic regression in the face of multicollinearity is proposed in Section 3.1, and the bounds of the ridge parameter are derived and discussed in Section 3.2.

3.1 The Proposed Estimator

An iterative algorithm for determining the optimal ridge parameter k is proposed using a criterion to minimize the MSEs of the estimated coefficients in the logistic regression model. In the neighborhood of convergence, the vector of the first derivatives of the penalized log-likelihood function in (2.25) with respect to $\boldsymbol{\beta}$ approaches zero. Expanding the vector of the first derivatives about the population parameter vector $\boldsymbol{\beta}$ as a Taylor Series yields a first order approximation as

$$\frac{\partial l_{LRR}(\hat{\boldsymbol{\beta}}_{LRR})}{\partial \boldsymbol{\beta}} \approx l'_{LRR}(\boldsymbol{\beta})\,|_{\boldsymbol{\beta}} + \left(\hat{\boldsymbol{\beta}}_{LRR} - \boldsymbol{\beta}\right) l''_{LRR}(\boldsymbol{\beta})\,|_{\boldsymbol{\beta}},$$

$$= \left\{\mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right] - k\boldsymbol{\beta}\right\}\Big|_{\boldsymbol{\beta}} - \left(\mathbf{W}'\mathbf{V}\mathbf{W} + k\mathbf{I}\right)\Big|_{\boldsymbol{\beta}} \left(\hat{\boldsymbol{\beta}}_{LRR} - \boldsymbol{\beta}\right).$$

Therefore,

$$\frac{\partial l_{LRR}(\hat{\boldsymbol{\beta}}_{LRR})}{\partial \boldsymbol{\beta}} \approx \left\{\mathbf{W}'\left[\mathbf{y} - \boldsymbol{\pi}(\mathbf{w})\right] - k\boldsymbol{\beta}\right\}\Big|_{\boldsymbol{\beta}} - \left(\mathbf{W}'\mathbf{V}\mathbf{W} + k\mathbf{I}\right)\Big|_{\boldsymbol{\beta}} \left(\hat{\boldsymbol{\beta}}_{LRR} - \boldsymbol{\beta}\right) = 0 \tag{3.1}$$

After a few manipulations in (3.1), it can be shown that

$$\left\{\mathbf{W}'\left[\mathbf{y}-\boldsymbol{\pi}(\mathbf{w})\right]-k\boldsymbol{\beta}\right\}\Big|_{\boldsymbol{\beta}}-\left(\mathbf{W}'\mathbf{V}\mathbf{W}+k\mathbf{I}\right)\Big|_{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}}_{LRR}-\boldsymbol{\beta}\right) = 0,$$

$$\mathbf{W}'\left(\mathbf{y}-\boldsymbol{\pi}(\mathbf{w})\right)|_{\boldsymbol{\beta}} -k\boldsymbol{\beta}-\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)\left(\hat{\boldsymbol{\beta}}_{LRR}-\boldsymbol{\beta}\right) = 0,$$

$$\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)\left(\hat{\boldsymbol{\beta}}_{LRR}-\boldsymbol{\beta}\right) = \mathbf{W}'\left(\mathbf{y}-\boldsymbol{\pi}(\mathbf{w})\right)|_{\boldsymbol{\beta}} +k\boldsymbol{\beta},$$

$$\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)\hat{\boldsymbol{\beta}}_{LRR} = \mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\boldsymbol{\beta}+\mathbf{W}'\left(\mathbf{y}-\boldsymbol{\pi}(\mathbf{w})\right)\Big|_{\boldsymbol{\beta}}$$

Thus,

$$\hat{\boldsymbol{\beta}}_{LRR} = \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\left[\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\boldsymbol{\beta}+\mathbf{W}'\left(\mathbf{y}-\boldsymbol{\pi}(\mathbf{w})\right)|_{\boldsymbol{\beta}}\right], \tag{3.2}$$

and then

$$\hat{\boldsymbol{\beta}}_{LRR} = \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\hat{\boldsymbol{\beta}}_{ML}. \tag{3.3}$$

The asymptotic bias and variance of $\hat{\boldsymbol{\beta}}_{LRR}$ in matrix form in (3.3) can be expressed as

$$\begin{aligned}
Bias(\hat{\boldsymbol{\beta}}_{LRR}) &= E\left(\hat{\boldsymbol{\beta}}_{LRR}-\boldsymbol{\beta}\right),\\
&= E\left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\hat{\boldsymbol{\beta}}_{ML}\right]-\boldsymbol{\beta},\\
&= \left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\right]E\left(\hat{\boldsymbol{\beta}}_{ML}\right)-\boldsymbol{\beta},\\
&= \left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\right]\boldsymbol{\beta}-\boldsymbol{\beta},\\
&= \left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}-\mathbf{I}\right]\boldsymbol{\beta},\\
&= \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\left[\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}-\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)\right]\boldsymbol{\beta},\\
&= -k\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\boldsymbol{\beta},
\end{aligned} \tag{3.4}$$

And

$$
\begin{aligned}
&Var(\hat{\boldsymbol{\beta}}_{LRR}) \\
&\quad = Var\left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\hat{\boldsymbol{\beta}}_{ML}\right], \\
&\quad = \left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\right]Var(\hat{\boldsymbol{\beta}}_{ML})\left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\right]', \\
&\quad = \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}, \\
&\quad = \left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}. \qquad (3.5)
\end{aligned}
$$

The scalar form of (3.5) can be written as

$$
\begin{aligned}
Var(\hat{\boldsymbol{\beta}}_{LRR}) &= tr\left[\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\right], \\
&= tr\left[\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-2}\right], \\
&= \sum_j \frac{\lambda_j}{(\lambda_j+k)^2}. \qquad (3.6)
\end{aligned}
$$

Subsequently, the asymptotic MSE of $\hat{\boldsymbol{\beta}}_{LRR}$ is given by

$$
\begin{aligned}
MSE(\hat{\boldsymbol{\beta}}_{LRR}) &= E\left[\left(\hat{\boldsymbol{\beta}}_{LRR}-\boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}}_{LRR}-\boldsymbol{\beta}\right)\right], \\
&= tr\left[Var(\hat{\boldsymbol{\beta}}_{LRR})\right]+\left[Bias(\hat{\boldsymbol{\beta}}_{LRR})\right]^2, \\
&= \sum_j \frac{\lambda_j}{(\lambda_j+k)^2}+\left[-k\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\boldsymbol{\beta}\right]'\left[-k\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-1}\boldsymbol{\beta}\right], \\
&= \sum_j \frac{\lambda_j}{(\lambda_j+k)^2}+k^2\boldsymbol{\beta}'\left(\mathbf{W}'\mathbf{V}(\boldsymbol{\beta})\mathbf{W}+k\mathbf{I}\right)^{-2}\boldsymbol{\beta}. \qquad (3.7)
\end{aligned}
$$

In practice, since the parameter $\boldsymbol{\beta}$ and $\mathbf{V}(\boldsymbol{\beta})$ are unknown, the parameters $\boldsymbol{\beta}$ and $\mathbf{V}(\boldsymbol{\beta})$ are replaced by their estimates $\hat{\boldsymbol{\beta}}_{ML}$ and $\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML})$, respectively, in (3.4) and (3.6), and following the decomposition of the symmetric matrix $\left(\mathbf{W}'\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W}+k\mathbf{I}\right)$, $MSE(k\,|\,\hat{\boldsymbol{\beta}}_{LRR})$ becomes the approximate asymptotic MSE of $\hat{\boldsymbol{\beta}}_{LRR}$, which can be expressed as

$$
\begin{aligned}
MSE(k\,|\,\hat{\boldsymbol{\beta}}_{ML}) &= \sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^2} + k^2\hat{\boldsymbol{\beta}}'_{ML}\left(\mathbf{W}'\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W}+k\mathbf{I}\right)^{-2}\hat{\boldsymbol{\beta}}_{ML}, \\
&= \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^2} + k^2\hat{\boldsymbol{\beta}}'_{ML}\left(\boldsymbol{\gamma}\boldsymbol{\Lambda}\boldsymbol{\gamma}'+k\mathbf{I}\right)^{-2}\hat{\boldsymbol{\beta}}_{ML}, \\
&= \sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^2} + k^2\hat{\boldsymbol{\beta}}'_{ML}\boldsymbol{\gamma}\left(\boldsymbol{\Lambda}+k\mathbf{I}\right)^{-2}\boldsymbol{\gamma}'\hat{\boldsymbol{\beta}}_{ML}, \\
&= \sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^2} + k^2\sum_j \frac{\alpha_j^2}{(\lambda_{ML_j}+k)^2},
\end{aligned}
\tag{3.8}
$$

where λ_{ML_j} is the j^{th} eigenvalue of $\mathbf{W}'\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W}$, $\hat{\boldsymbol{\alpha}}=\boldsymbol{\gamma}'\hat{\boldsymbol{\beta}}_{ML}$ and $\boldsymbol{\gamma}$ and $\boldsymbol{\Lambda}$ are the eigenvector and the diagonal matrix of the eigenvalues of $\mathbf{W}'\mathbf{V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W}$, respectively.

The minimum of (3.8) is unique since the variance is a monotonically decreasing function of k and the squared bias is an monotonically increasing function of k. The iterative Nelder-Mead algorithm one of the popular methods in nonlinear programming (Nelder and Mead, 1965; Baeyens, Herreros and Perán, 2016) is used to search for the value minimizing $MSE(k\,|\,\hat{\boldsymbol{\beta}}_{ML})$ in (3.8). At iteration $t+1$, k^{t+1} is given by

$$
k^{t+1} = \arg\min_k \left[= \sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^2} + k^2\sum_j \frac{\alpha_j^2}{(\lambda_{ML_j}+k)^2} \right]. \tag{3.9}
$$

Upon convergence, the ridge parameter k^{t+1} approaches the minimizing value k_{opt},

$$k_{opt} = \arg\min_{k} MSE(k \mid \hat{\boldsymbol{\beta}}_{ML}), \tag{3.10}$$

and the proposed parameter $\hat{\boldsymbol{\beta}}_{LRR}(k_{opt})$ can be derived from (3.3) as

$$\hat{\boldsymbol{\beta}}_{LRR}(k_{opt}) = \left\{\mathbf{W'V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W} + k_{opt}\mathbf{I}\right\}^{-1} \mathbf{W'V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W}\hat{\boldsymbol{\beta}}_{ML}\,. \tag{3.11}$$

The maximum likelihood estimate ($\hat{\boldsymbol{\beta}}_{ML}$) is obtained using the iteratively reweighted least squares algorithm based on (2.20).

The general iterative algorithm can be summarized as follows:

1) Assume $t = 0$ and $k^t = 0,$ and determine an initial LRR estimate of the population parameter vector $\boldsymbol{\beta}_{LRR}(k^t) = \hat{\boldsymbol{\beta}}_{ML}$ and a termination criterion.

2) Compute the eigenvalues λ_j^t of $\mathbf{W'V}(\hat{\boldsymbol{\beta}}_{ML})\mathbf{W} + k^t$ I.

3) Compute k^{t+1} in (3.9) by using the Nelder-Mead algorithm.

4) If the termination criterion is satisfied, the $k_{opt} = k^{t+1}$, compute $\hat{\boldsymbol{\beta}}_{LRR}(k_{opt})$ in (3.11), and terminate the algorithm; else $t = t + 1$ and go to step 2.

3.2 The Bounds of the Ridge Parameter

In this work, the ridge parameter *k* is determined such that the MSE in (3.7) is minimized. Due to the high nonlinearity of *k* in $MSE(k \mid \hat{\boldsymbol{\beta}}_{ML})$, the optimum value of *k* is obtained using an iterative algorithm. The following theorem is developed to provide an approximate upper bound of the ridge parameter for the search for k_{opt}.

Theorem 3.1: Let $c = p \Big/ \sum_j \lambda_{ML_j} \alpha_j^2$. If $k_{opt} = \arg\min MSE(k \mid \hat{\boldsymbol{\beta}}_{ML})$, then

$$0 < k_{opt} \le c\left[2 + \frac{f_2}{2f_1}\left(1 - \sqrt{1 + \frac{4f_1 f_3}{f_2^2}}\right)\right] \text{ if } 0 < \frac{\lambda_{ML_{\max}}}{c} < 2(2^{1/3} - 1), \tag{3.12}$$

$$0 < k_{opt} < c\left(1 + \frac{1}{\sqrt{1 + \frac{4f_1 f_3}{f_4^4}}}\right) \text{ if } 2(2^{1/3} - 1) \le \frac{\lambda_{ML_{\max}}}{c} < 1.6550, \tag{3.13}$$

$$0 < k_{opt} \le c + \sqrt{3c\lambda_{ML_{\max}}} \text{ if } \frac{\lambda_{ML_{\max}}}{c} \ge 1.6550, \tag{3.14}$$

where

$$f_1 = \frac{3\lambda_{ML_{\max}}}{8c}\left(1 + \frac{\lambda_{ML_{\max}}}{2c}\right)\left(1 + \frac{\lambda_{ML_{\max}}}{c}\right), \tag{3.15}$$

$$f_2 = 1 + \frac{3\lambda_{ML_{\max}}}{4c}\left(1 + \frac{\lambda_{ML_{\max}}}{2c}\right)^2, \tag{3.16}$$

$$f_3 = 2 - \left(1 + \frac{\lambda_{ML_{\max}}}{2c}\right)^3, \tag{3.17}$$

$$f_4 = \frac{\lambda_{ML_{\max}}}{4c}\left[\left(\frac{\lambda_{ML_{\max}}}{c}\right)^2 + 5\left(\frac{\lambda_{ML_{\max}}}{c}\right) + 8\right]. \tag{3.18}$$

Proof. The first derivative of $MSE(k\,|\,\hat{\boldsymbol{\beta}}_{ML})$ in (3.8) with respect to k can be expressed as

$$\frac{\partial MSE(k\,|\,\hat{\boldsymbol{\beta}}_{ML})}{\partial k} = -2\sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^3} + 2k\sum_j \frac{\alpha_j^2}{(\lambda_{ML_j}+k)^2} - 2\mathrm{k}^2\sum_j \frac{\alpha_j^2}{(\lambda_{ML_j}+k)^3},$$

$$= -2\sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^3} + \sum_j \frac{\alpha_j^2}{(\lambda_{ML_j}+k)^3}\left[2k\left(\lambda_{ML_j}+k\right)-2k^2\right],$$

$$= -2\sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^3} + 2k\sum_j \frac{\lambda_{ML_j}\alpha_j^2}{(\lambda_{ML_j}+k)^3}. \tag{3.19}$$

From (3.19), the sufficient condition that $MSE(k\,|\,\hat{\boldsymbol{\beta}}_{ML})$ is an increasing function of k is defined as

$$k\sum_j \frac{\lambda_{ML_j}\alpha_j^2}{(\lambda_{ML_j}+k)^3} \quad > \quad \sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^3}. \tag{3.20}$$

The upper bound for the right-hand side (RHS) of (3.20) can be approximated as

$$\sum_j \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^3} \quad < \quad \sum_j \frac{\lambda_{ML_j}}{k^3} = \frac{p}{k^3}. \tag{3.21}$$

For the left-hand side (LHS) of (3.20), the j^{th} element of the diagonal matrix is

$$\lambda_{ML_{\max}}+k \quad > \quad \lambda_{ML_j}+k, \text{ for all } j\text{'s},$$

$$(\lambda_{ML_{\max}}+k)^3 \quad > \quad (\lambda_{ML_j}+k)^3, \text{ for all } j\text{'s},$$

$$\frac{1}{(\lambda_{ML_{\max}}+k)^3} \quad < \quad \frac{1}{(\lambda_{ML_j}+k)^3}, \text{ for all } j\text{'s},$$

$$\frac{\lambda_{ML_j}}{(\lambda_{ML_{\max}}+k)^3} < \frac{\lambda_{ML_j}}{(\lambda_{ML_j}+k)^3}, \text{ for all } j\text{'s},$$

$$\frac{\lambda_{ML_j}\alpha_j^2}{(\lambda_{ML_{\max}}+k)^3} < \frac{\lambda_{ML_j}\alpha_j^2}{(\lambda_{ML_j}+k)^3}, \text{ for all } j\text{'s}.$$

Therefore, the lower bound for the LHS in (3.20) can be written as

$$k\sum_j \frac{\lambda_{ML_j}\alpha_j^2}{(\lambda_{ML_{\max}}+k)^3} < k\sum_j \frac{\lambda_{ML_j}\alpha_j^2}{(\lambda_{ML_j}+k)^3}. \tag{3.22}$$

Subsequently, the sufficient condition in (3.20) becomes

$$\frac{p}{k^3} < \frac{k}{(\lambda_{ML_{\max}}+k)^3}\sum_j \lambda_{ML_j}\alpha_j^2,$$

$$\frac{k^4}{(\lambda_{ML_{\max}}+k)^3} > \frac{p}{\sum_j \lambda_{ML_j}\alpha_j^2} = c. \tag{3.23}$$

Rearranging the inequality in (3.23) can be carried out as follows:

$$k^4 > c(\lambda_{ML_{\max}}+k)^3,$$

$$k^{4/3} > c^{1/3}(\lambda_{ML_{\max}}+k),$$

$$k^{4/3} - c^{1/3}k > c^{1/3}\lambda_{ML_{\max}},$$

$$k\left(k^{1/3}-c^{1/3}\right) > c^{1/3}\lambda_{ML_{\max}}. \tag{3.24}$$

This implies that the necessary condition for $MSE(k\,|\,\hat{\boldsymbol{\beta}}_{ML})$ to be an increasing function of k is $k > c$. Let $k = c + \delta$, where $\delta > 0$. By replacing k in the terms for c and δ into (3.24), we obtain

$$(c+\delta)\left[(c+\delta)^{1/3}-c^{1/3}\right] \;>\; c^{1/3}\lambda_{ML_{\max}}\,, \tag{3.25}$$

<u>Case</u>: $0<\dfrac{\delta}{c}<1$. The inequality in (3.25) can be written in term of $\dfrac{\delta}{c}$ as

$$(c+\delta)\left[(c+\delta)^{1/3}-c^{1/3}\right] \;>\; c^{1/3}\lambda_{ML_{\max}}\,,$$

$$c\left(1+\frac{\delta}{c}\right)\left[c^{1/3}\left(1+\frac{\delta}{c}\right)^{1/3}-c^{1/3}\right] \;>\; c^{1/3}\lambda_{ML_{\max}}\,,$$

$$\left(1+\frac{\delta}{c}\right)\left[\left(1+\frac{\delta}{c}\right)^{1/3}-1\right] \;>\; \frac{\lambda_{ML_{\max}}}{c}\,. \tag{3.26}$$

Let $\dfrac{\delta}{c}=1-\Delta_1$. The condition $0<\dfrac{\delta}{c}<1$ results in $0<\Delta_1<1$. We can replace $\dfrac{\delta}{c}$ in terms of Δ_1 into (3.26) as

$$\left(1+1-\Delta_1\right)\left[\left(1+1-\Delta_1\right)^{1/3}-1\right] \;>\; \frac{\lambda_{ML_{\max}}}{c}\,,$$

$$\left(2-\Delta_1\right)\left[\left(2-\Delta_1\right)^{1/3}-1\right] \;>\; \frac{\lambda_{ML_{\max}}}{c}\,,$$

$$2\left(1-\frac{\Delta_1}{2}\right)\left[2^{1/3}\left(1-\frac{\Delta_1}{2}\right)^{1/3}-1\right] \;>\; \frac{\lambda_{ML_{\max}}}{c}\,,$$

$$\left[2^{1/3}\left(1-\frac{\Delta_1}{2}\right)^{1/3}-1\right] \;>\; \frac{\lambda_{ML_{\max}}}{2c\left(1-\dfrac{\Delta_1}{2}\right)}\,,$$

$$2\left(1-\frac{\Delta_1}{2}\right) \;>\; \left[1+\frac{\lambda_{ML_{\max}}}{2c\left(1-\dfrac{\Delta_1}{2}\right)}\right]^3. \tag{3.27}$$

When considering the RHS of (3.27), its expanded form is

$$
\begin{aligned}
RHS &= \left[1 + \frac{\lambda_{ML_{\max}}}{2c\left(1-\frac{\Delta_1}{2}\right)} \right]^3 , \\
&= 1 + 3\left[\frac{\lambda_{ML_{\max}}}{2c\left(1-\frac{\Delta_1}{2}\right)} \right] + 3\left[\frac{\lambda_{ML_{\max}}}{2c\left(1-\frac{\Delta_1}{2}\right)} \right]^2 + \left[\frac{\lambda_{ML_{\max}}}{2c\left(1-\frac{\Delta_1}{2}\right)} \right]^3 , \\
&= 1 + \frac{3}{2}\frac{\lambda_{ML_{\max}}}{c}\left(\frac{1}{1-\frac{\Delta_1}{2}} \right) + \frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{c} \right)^2 \left(\frac{1}{1-\frac{\Delta_1}{2}} \right)^2 + \frac{1}{8}\left(\frac{\lambda_{ML_{\max}}}{c} \right)^3 \left(\frac{1}{1-\frac{\Delta_1}{2}} \right)^3 , \\
&= 1 + \frac{3}{2}\frac{\lambda_{ML_{\max}}}{c}\left(1-\frac{\Delta_1}{2}\right)^{-1} + \frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{c} \right)^2 \left(1-\frac{\Delta_1}{2}\right)^{-2} + \frac{1}{8}\left(\frac{\lambda_{ML_{\max}}}{c} \right)^3 \left(1-\frac{\Delta_1}{2}\right)^{-3} .
\end{aligned}
\tag{3.28}
$$

Expanding the Taylor Series of $\left(1-\frac{\Delta_1}{2}\right)^{-r}, r = 1, 2, 3$ gives the second order of $\frac{\Delta_1}{2}$:

$$
\left(1-\frac{\Delta_1}{2}\right)^{-1} \approx 1 + \frac{\Delta_1}{2} + \left(\frac{\Delta_1}{2}\right)^2 = 1 + \frac{1}{2}\Delta_1 + \frac{1}{4}\Delta_1^2 , \tag{3.29}
$$

$$
\left(1-\frac{\Delta_1}{2}\right)^{-2} \approx 1 + 2\left(\frac{\Delta_1}{2}\right) + 3\left(\frac{\Delta_1}{2}\right)^2 = 1 + \Delta_1 + \frac{3}{4}\Delta_1^2 , \tag{3.30}
$$

$$
\left(1-\frac{\Delta_1}{2}\right)^{-3} \approx 1 + 3\left(\frac{\Delta_1}{2}\right) + 6\left(\frac{\Delta_1}{2}\right)^2 = 1 + \frac{3}{2}\Delta_1 + \frac{3}{2}\Delta_1^2 . \tag{3.31}
$$

Replacing (3.29), (3.30), and (3.31) in (3.28) results in

$$
\begin{aligned}
RHS = {} & 1 + \frac{3}{2}\frac{\lambda_{ML\max}}{c}\left(1 + \frac{1}{2}\Delta_1 + \frac{1}{4}\Delta_1^2\right) + \frac{3}{4}\left(\frac{\lambda_{ML\max}}{c}\right)^2\left(1 + \Delta_1 + \frac{3}{4}\Delta_1^2\right) \\
& + \frac{1}{8}\left(\frac{\lambda_{ML\max}}{c}\right)^3\left(1 + \frac{3}{2}\Delta_1 + \frac{3}{2}\Delta_1^2\right), \\
= {} & 1 + 3\left(\frac{\lambda_{ML\max}}{2c}\right) + 3\left(\frac{\lambda_{ML\max}}{2c}\right)^2 + \left(\frac{\lambda_{ML\max}}{2c}\right)^3 \\
& + \left[\frac{3}{2}\left(\frac{\lambda_{ML\max}}{2c}\right) + 3\left(\frac{\lambda_{ML\max}}{2c}\right)^2 + \frac{3}{2}\left(\frac{\lambda_{ML\max}}{2c}\right)^3\right]\Delta_1 \\
& + \left[\frac{3}{4}\left(\frac{\lambda_{ML\max}}{2c}\right) + \frac{9}{4}\left(\frac{\lambda_{ML\max}}{2c}\right)^2 + \frac{3}{2}\left(\frac{\lambda_{ML\max}}{2c}\right)^3\right]\Delta_1^2, \\
= {} & \left(1 + \frac{\lambda_{ML\max}}{2c}\right)^3 + \frac{3}{2}\left(\frac{\lambda_{ML\max}}{2c}\right)\left[1 + 2\left(\frac{\lambda_{ML\max}}{2c}\right) + \left(\frac{\lambda_{ML\max}}{2c}\right)^2\right]\Delta_1 \\
& + \frac{3}{4}\left(\frac{\lambda_{ML\max}}{2c}\right)\left[1 + 3\left(\frac{\lambda_{ML\max}}{2c}\right) + 2\left(\frac{\lambda_{ML\max}}{2c}\right)^2\right]\Delta_1^2, \\
= {} & \left(1 + \frac{\lambda_{ML\max}}{2c}\right)^3 + \frac{3\lambda_{ML\max}}{4c}\left(1 + \frac{\lambda_{ML\max}}{2c}\right)^2\Delta_1 \\
& + \frac{3\lambda_{ML\max}}{8c}\left(1 + \frac{\lambda_{ML\max}}{2c}\right)\left(1 + \frac{\lambda_{ML\max}}{c}\right)\Delta_1^2.
\end{aligned}
$$

The inequality in (3.27) can be expressed as

$$
\begin{aligned}
2\left(1 - \frac{\Delta_1}{2}\right) > {} & \left(1 + \frac{\lambda_{ML\max}}{2c}\right)^3 + \frac{3\lambda_{ML\max}}{4c}\left(1 + \frac{\lambda_{ML\max}}{2c}\right)^2\Delta_1 \\
& + \frac{3\lambda_{ML\max}}{8c}\left(1 + \frac{\lambda_{ML\max}}{2c}\right)\left(1 + \frac{\lambda_{ML\max}}{c}\right)\Delta_1^2,
\end{aligned}
$$

$$2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \quad > \quad \Delta_1+\frac{3\lambda_{ML_{\max}}}{4c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2\Delta_1 + \frac{3\lambda_{ML_{\max}}}{8c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_1^2,$$

$$2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \quad > \quad \left[1+\frac{3\lambda_{ML_{\max}}}{4c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2\right]\Delta_1 + \frac{3\lambda_{ML_{\max}}}{8c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_1^2. \tag{3.32}$$

The inequality in (3.32) becomes

$$f_1\Delta_1^2 + f_2\Delta_1 \quad < \quad f_3. \tag{3.33}$$

where f_1, f_2 and f_3 are defined in (3.15), (3.16) and (3.17) respectively.

The inequality (3.33) can be derived as

$$\Delta_1^2 + \frac{f_2}{f_1}\Delta_1 \quad < \quad \frac{f_3}{f_1},$$

$$\left(\Delta_1 + \frac{f_2}{2f_1}\right)^2 \quad < \quad \frac{f_2^2}{4f_1^2} + \frac{f_3}{f_1},$$

$$\Delta_1 + \frac{f_2}{2f_1} \quad < \quad \sqrt{\frac{f_2^2}{4f_1^2} + \frac{f_3}{f_1}},$$

$$\Delta_1 \quad < \quad \sqrt{\frac{f_2^2}{4f_1^2} + \frac{f_3}{f_1}} - \frac{f_2}{2f_1},$$

$$\Delta_1 \quad < \quad \sqrt{\frac{f_2^2}{4f_1^2}\left(1+\frac{4f_1f_3}{f_2^2}\right)} - \frac{f_2}{2f_1}.$$

Thus, the upper bound of Δ_1 can be written as

$$\Delta_1 < \frac{f_2}{2f_1}\left(\sqrt{1+\frac{4f_1 f_3}{f_2^2}}-1\right). \tag{3.34}$$

A positive value for f_3 is required to fulfill the assumption that $\Delta_1 > 0$, which leads to the conditions for the inequality as follows:

$$\begin{aligned}
2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 &> 0,\\
\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 &< 2,\\
1+\frac{\lambda_{ML_{\max}}}{2c} &< 2^{1/3},\\
\frac{\lambda_{ML_{\max}}}{c} &< 2(2^{1/3}-1).
\end{aligned} \tag{3.35}$$

From the definition of Δ_1 and the inequality in (3.34), we can obtain

$$\begin{aligned}
\frac{\delta}{c} &= 1-\Delta_1,\\
\frac{\delta}{c} &= 1-\frac{f_2}{2f_1}\left(\sqrt{1+\frac{4f_1 f_3}{f_2^2}}-1\right),\\
\delta &= c\left[1-\frac{f_2}{2f_1}\left(\sqrt{1+\frac{4f_1 f_3}{f_2^2}}-1\right)\right].
\end{aligned}$$

Consequently, the upper bound of k, k_{ub}, to search for the optimal k_{opt} is given by

$$
\begin{aligned}
k_{ub} &= c+\delta, \\
&= c + c\left[1 - \frac{f_2}{2f_1}\left(\sqrt{1+\frac{4f_1f_3}{f_2^2}}-1\right)\right], \\
&= c\left[2 - \frac{f_2}{2f_1}\left(\sqrt{1+\frac{4f_1f_3}{f_2^2}}-1\right)\right], \\
&= c\left[2 + \frac{f_2}{2f_1}\left(1-\sqrt{1+\frac{4f_1f_3}{f_2^2}}\right)\right].
\end{aligned} \tag{3.36}
$$

which is for $0 < \frac{\lambda_{ML_{\max}}}{c} < 2(2^{1/3}-1)$. This is because

<u>Case</u>: $\frac{\delta}{c} > 1$. Inequality (3.25) can be rearranged in terms of $\frac{c}{\delta} < 1$ as

$$
\begin{aligned}
(c+\delta)\left[(c+\delta)^{1/3} - c^{1/3}\right] &> c^{1/3}\lambda_{ML_{\max}}, \\
\delta\left(1+\frac{c}{\delta}\right)\left[\delta^{1/3}\left(1+\frac{c}{\delta}\right)^{1/3} - c^{1/3}\right] &> c^{1/3}\lambda_{ML_{\max}}, \\
\delta^{1/3}\left(1+\frac{c}{\delta}\right)^{1/3} - c^{1/3} &> \frac{c^{1/3}\lambda_{ML_{\max}}}{\delta\left(1+\frac{c}{\delta}\right)}, \\
\delta^{1/3}\left(1+\frac{c}{\delta}\right)^{1/3} &> c^{1/3} + \frac{c^{1/3}\lambda_{ML_{\max}}}{\delta\left(1+\frac{c}{\delta}\right)}, \\
\left(1+\frac{c}{\delta}\right)^{1/3} &> \left(\frac{c}{\delta}\right)^{1/3}\left[1+\frac{\lambda_{ML_{\max}}}{\delta\left(1+\frac{c}{\delta}\right)}\right],
\end{aligned}
$$

$$1+\frac{c}{\delta} \;>\; \frac{c}{\delta}\left[1+\frac{\lambda_{ML_{\max}}}{\delta\left(1+\frac{c}{\delta}\right)}\right]^3, \tag{3.37}$$

$$1+\frac{c}{\delta} \;>\; \frac{c}{\delta}\left[1+\frac{\left(\lambda_{ML_{\max}}+c\right)}{\delta}\right]^3\cdot\frac{1}{\left(1+\frac{c}{\delta}\right)^3}, \tag{3.38}$$

$$1+\frac{c}{\delta} \;>\; \frac{c}{\delta}\left[1+\frac{c}{\delta}\left(\frac{\lambda_{ML_{\max}}}{c}+1\right)\right]^3\cdot\frac{1}{\left(1+\frac{c}{\delta}\right)^3}.$$

Expanding the Taylor Series of $\left(1+\frac{c}{\delta}\right)^{-3}$ to the second order of $\frac{c}{\delta}$ gives

$$1+\frac{c}{\delta} \;>\; \frac{c}{\delta}\left[\begin{array}{l}1+\frac{3c}{\delta}\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)+3\left(\frac{c}{\delta}\right)^2\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^2\\ +\left(\frac{c}{\delta}\right)^3\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^3\end{array}\right]\left[1-\frac{3c}{\delta}+6\left(\frac{c}{\delta}\right)^2\right],$$

$$> \;\frac{c}{\delta}\left[\begin{array}{l}1+\frac{3c}{\delta}\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)+3\left(\frac{c}{\delta}\right)^2\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^2+\left(\frac{c}{\delta}\right)^3\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^3\\ -\frac{3c}{\delta}-9\left(\frac{c}{\delta}\right)^2\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)-9\left(\frac{c}{\delta}\right)^3\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^2\\ -3\left(\frac{c}{\delta}\right)^4\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^3+6\left(\frac{c}{\delta}\right)^2+18\left(\frac{c}{\delta}\right)^3\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\\ +18\left(\frac{c}{\delta}\right)^4\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^2+6\left(\frac{c}{\delta}\right)^5\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^3\end{array}\right],$$

$$
\begin{aligned}
1 &> \frac{c}{\delta}\left[\begin{array}{l} \frac{3c}{\delta}\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)-\frac{3c}{\delta}-3\left(\frac{c}{\delta}\right)^2\left(1+\frac{\lambda_{ML_{\max}}}{c}\right) \\ +6\left(\frac{c}{\delta}\right)^2+3\left(\frac{c}{\delta}\right)^2\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)^2 \end{array}\right], \\
&> \frac{c}{\delta}\left[\frac{3c}{\delta}\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)-\frac{3c}{\delta}\right], \\
&> 3\left(\frac{c}{\delta}\right)^2\left[1+\frac{\lambda_{ML_{\max}}}{c}-1\right], \\
&> \frac{3c\lambda_{ML_{\max}}}{\delta^2}, \\
\delta^2 &> 3c\lambda_{ML_{\max}}, \\
\delta &> \sqrt{3c\lambda_{ML_{\max}}}.
\end{aligned}
$$

Therefore, $$\delta > \sqrt{3c\lambda_{ML_{\max}}}, \qquad \text{for } \frac{c}{\delta} < 1. \tag{3.39}$$

The upper bound of k to search for the optimal k_{opt} in the case of $\frac{\delta}{c} >> 1$ is equal to

$$k_{ub} \approx c + \sqrt{3c\lambda_{ML_{\max}}}. \tag{3.40}$$

Note that the approximation of the upper bound of k in (3.34) may be invalid when $\frac{\delta}{c}$ is slightly greater than 1. Consider the case where $\frac{c}{\delta} = 1 - \Delta_2$, in which Δ_2 is slightly greater than zero, then $\frac{1}{\delta} = \frac{1-\Delta_2}{c}$. If substituting $\frac{c}{\delta}$ in terms of Δ_2 in (3.37), we can show that

$$1+\left(1-\Delta_2\right) \;>\; c\left(\frac{1-\Delta_2}{c}\right)\left[1+\frac{\lambda_{ML_{\max}}}{\left(1+1-\Delta_2\right)}\left(\frac{1-\Delta_2}{c}\right)\right]^3,$$

$$2-\Delta_2 \;>\; \left(1-\Delta_2\right)\left[1+\frac{\lambda_{ML_{\max}}\left(1-\Delta_2\right)}{c\left(2-\Delta_2\right)}\right]^3,$$

$$2-\Delta_2 \;>\; (1-\Delta_2)\left[1+\frac{\lambda_{ML_{\max}}(1-\Delta_2)}{2c\left(1-\frac{\Delta_2}{2}\right)}\right]^3. \tag{3.41}$$

Considering the RHS of Inequality (3.41):

$$RHS \;=\; (1-\Delta_2)\left[1+\frac{\lambda_{ML_{\max}}(1-\Delta_2)}{2c\left(1-\frac{\Delta_2}{2}\right)}\right]^3,$$

$$=\; (1-\Delta_2)\left[1+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\frac{(1-\Delta_2)}{\left(1-\frac{\Delta_2}{2}\right)}+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2\frac{(1-\Delta_2)^2}{\left(1-\frac{\Delta_2}{2}\right)^2}+\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3\frac{(1-\Delta_2)^3}{\left(1-\frac{\Delta_2}{2}\right)^3}\right].$$

Expanding the Taylor Series of $\left(1-\frac{\Delta_2}{2}\right)^{-r}$, $r=1,2,3$, to the second order of Δ_2 gives

$$RHS = (1-\Delta_2)\begin{bmatrix} 1+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)(1-\Delta_2)\left(1+\frac{1}{2}\Delta_2+\frac{1}{4}\Delta_2^2\right) \\ +3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2(1-\Delta_2)^2\left(1+\Delta_2+\frac{3}{4}\Delta_2^2\right) \\ +\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3(1-\Delta_2)^3\left(1+\frac{3}{2}\Delta_2+\frac{3}{2}\Delta_2^2\right) \end{bmatrix},$$

$$RHS = (1-\Delta_2)\begin{bmatrix} 1+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{1}{2}\Delta_2+\frac{1}{4}\Delta_2^2-\Delta_2-\frac{1}{2}\Delta_2^2-\frac{1}{4}\Delta_2^3\right) \\ +3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2(1-2\Delta_2+\Delta_2^2)\left(1+\Delta_2+\frac{3}{4}\Delta_2^2\right) \\ +\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3(1-3\Delta_2+3\Delta_2^2-\Delta_2^2)\left(1+\frac{3}{2}\Delta_2+\frac{3}{2}\Delta_2^2\right) \end{bmatrix},$$

$$= (1-\Delta_2)\begin{bmatrix} 1+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1-\frac{1}{2}\Delta_2-\frac{1}{4}\Delta_2^2-\frac{1}{4}\Delta_2^3\right) \\ +3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2\left(1-\Delta_2-\frac{1}{4}\Delta_2^2-\frac{1}{2}\Delta_2^3+\frac{3}{4}\Delta_2^4\right) \\ +\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3\left(1-\frac{3}{2}\Delta_2-\Delta_2^3+3\Delta_2^4-\frac{3}{2}\Delta_2^5\right) \end{bmatrix},$$

$$= (1-\Delta_2)\begin{bmatrix} 1+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2+\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \\ +\Delta_2\left[-\frac{3}{2}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)-3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2-\frac{3}{2}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3\right] \\ +\Delta_2^2\left[-\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)-\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2\right] \\ +\Delta_2^3\left[-\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)-\frac{3}{2}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2-\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3\right] \\ +\Delta_2^4\left[\frac{9}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2+3\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3\right]-\frac{3}{2}\Delta_2^5\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \end{bmatrix}. \qquad (3.42)$$

Abandoning the higher order than the second order of Δ_2 in (3.42) arrives at

$$RHS \approx (1-\Delta_2)\left[\begin{array}{l}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3-\frac{3}{2}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left[1+2\left(\frac{\lambda_{ML_{\max}}}{2c}\right)+\left(\frac{\lambda_{ML_{\max}}}{2c}\right)^2\right]\Delta_2\\ -\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left[1+\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\right]\Delta_2^2\end{array}\right],$$

$$= \left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3-\frac{3}{2}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2\Delta_2-\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\Delta_2^2$$

$$-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3\Delta_2+\frac{3}{2}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2\Delta_2^2+\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\Delta_2^3,$$

$$\approx \left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2\left[\frac{3}{2}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)+\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\right]\Delta_2$$

$$+\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left[2\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)-1\right]\Delta_2^2,$$

$$= \left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2$$

$$+\frac{3}{4}\left(\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2^2.$$

Therefore, Inequality (3.41) can be rearranged as

$$2-\Delta_2 \;>\; \left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3-\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2\Delta_2$$
$$+\frac{3}{8}\frac{\lambda_{ML_{\max}}}{c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2^2,$$

$$2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \;>\; \frac{3}{8}\frac{\lambda_{ML_{\max}}}{c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2^2$$
$$-\left[\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^2-1\right]\Delta_2,$$

$$2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \;>\; \frac{3}{8}\frac{\lambda_{ML_{\max}}}{c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2^2$$
$$-\left[\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}+\frac{\lambda_{ML_{\max}}^2}{4c^2}\right)-1\right]\Delta_2,$$

$$2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \;>\; \frac{3}{8}\frac{\lambda_{ML_{\max}}}{c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2^2$$
$$-\left[\left(1+\frac{\lambda_{ML_{\max}}}{c}+\frac{\lambda_{ML_{\max}}}{c}+\frac{\lambda_{ML_{\max}}^2}{c^2}+\frac{\lambda_{ML_{\max}}^2}{4c^2}+\frac{\lambda_{ML_{\max}}^3}{4c^3}\right)-1\right]\Delta_2,$$

$$2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \;>\; \frac{3}{8}\frac{\lambda_{ML_{\max}}}{c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2^2$$
$$-\left[\frac{2\lambda_{ML_{\max}}}{c}+\frac{5}{4}\left(\frac{\lambda_{ML_{\max}}}{c}\right)^2+\frac{\lambda_{ML_{\max}}^3}{4c^3}\right]\Delta_2,$$

$$2-\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)^3 \;>\; \frac{3}{8}\frac{\lambda_{ML_{\max}}}{c}\left(1+\frac{\lambda_{ML_{\max}}}{2c}\right)\left(1+\frac{\lambda_{ML_{\max}}}{c}\right)\Delta_2^2$$
$$-\frac{\lambda_{ML_{\max}}}{4c}\left[\left(\frac{\lambda_{ML_{\max}}}{c}\right)^2+5\left(\frac{\lambda_{ML_{\max}}}{c}\right)+8\right]\Delta_2. \qquad (3.43)$$

The inequality in (3.43) becomes

$$f_1\Delta_2^2 - f_4\Delta_2 \;<\; f_3, \tag{3.44}$$

where f_1, f_3, and f_4 are defined in (3.15), (3.17), and (3.18), respectively. Inequality (3.44) can be derived as

$$\begin{aligned}
f_1\Delta_2^2 - f_4\Delta_2 \;&<\; f_3, \\
\Delta_2^2 - \frac{f_4}{f_1}\Delta_2 \;&<\; \frac{f_3}{f_1}, \\
\left(\Delta_2 - \frac{f_4}{2f_1}\right)^2 \;&<\; \frac{f_4^2}{4f_1^2} + \frac{f_3}{f_1}, \\
\left(\Delta_2 - \frac{f_4}{2f_1}\right)^2 \;&<\; \frac{f_4^2}{4f_1^2}\left(1 + \frac{f_1 f_3}{f_4^2}\right), \\
\Delta_2 - \frac{f_4}{2f_1} \;&<\; \pm\frac{f_4}{2f_1}\sqrt{1 + \frac{f_1 f_3}{f_4^2}}, \\
\Delta_2 \;&<\; \frac{f_4}{2f_1} \pm \frac{f_4}{2f_1}\sqrt{1 + \frac{f_1 f_3}{f_4^2}}, \\
\Delta_2 \;&<\; \frac{f_4}{2f_1}\left(1 \pm \sqrt{1 + \frac{f_1 f_3}{f_4^2}}\right).
\end{aligned} \tag{3.45}$$

From the definition of Δ_2 and the inequality (3.45), two cases can be considered.

First, in the case where, $\Delta_2 < \frac{f_4}{2f_1}\left(1 + \sqrt{1 + \frac{f_1 f_3}{f_4^2}}\right)$ and $\frac{f_4}{2f_1}\left(1 + \sqrt{1 + \frac{f_1 f_3}{f_4^2}}\right) > 0$, these conditions are invalid.

Second, in the case where $\Delta_2 < \frac{f_4}{2f_1}\left(1-\sqrt{1+\frac{f_1 f_3}{f_4^2}}\right)$ and $\frac{f_4}{2f_1}\left(1-\sqrt{1+\frac{f_1 f_3}{f_4^2}}\right) > 0$, obviously $1+\frac{f_1 f_3}{f_4^2} > 0$, then $f_3 > -\frac{f_4^2}{4f_1}$. Consider

$$\frac{f_4}{2f_1} = \frac{\frac{\lambda_{ML\max}}{4c}\left[\left(\frac{\lambda_{ML\max}}{c}\right)^2 + 5\left(\frac{\lambda_{ML\max}}{c}\right) + 8\right]}{2\cdot\frac{3\lambda_{ML\max}}{8c}\left(1+\frac{\lambda_{ML\max}}{2c}\right)\left(1+\frac{\lambda_{ML\max}}{c}\right)},$$

$$\frac{f_4}{2f_1} = \frac{\left(\frac{\lambda_{ML\max}}{c}\right)^2 + 5\left(\frac{\lambda_{ML\max}}{c}\right) + 8}{3\left(1+\frac{3\lambda_{ML\max}}{2c}+\frac{\lambda_{ML\max}^2}{2c^2}\right)},$$

$$= \frac{2\left(\frac{\lambda_{ML\max}}{c}\right)^2 + 10\left(\frac{\lambda_{ML\max}}{c}\right) + 16}{3\left(\frac{\lambda_{ML\max}}{c}\right)^2 + 9\left(\frac{\lambda_{ML\max}}{c}\right) + 6} > 1.$$

Thus, it can be concluded that

$$0 < \Delta_2 < \frac{f_4}{2f_1}\left(1-\sqrt{1+\frac{4f_1 f_3}{f_4^2}}\right). \tag{3.46}$$

When f_3 approaches zero from the negative direction, Δ_2 approaches zero from the opposite direction. On the other hand, from (3.17), Δ_2 approaches zero when $\frac{\lambda_{ML\max}}{c} \to 2(2^{1/3}-1)$. In this case, the upper bound of k to search for the optimal k_{opt} is defined as

$$k_{ub} \approx c\frac{2-\frac{f_4}{2f_1}\left(1-\sqrt{1+\frac{4f_1f_3}{f_4^2}}\right)}{1-\frac{f_4}{2f_1}\left(1-\sqrt{1+\frac{4f_1f_3}{f_4^2}}\right)}. \quad (3.47)$$

From (3.35) and (3.47), we find that the upper bound in (3.47) is less than the one in (3.35) for $2(2^{1/3}-1)<\frac{\lambda_{ML\max}}{c}<1.6550$, which corresponds to $1<\frac{\delta}{c}<3.5719$, but is greater than one in (3.35) for $\frac{\lambda_{ML\max}}{c}>1.6550$ or $\frac{\delta}{c}>3.5719$, as shown in Figure 3.1. Therefore, $2(2^{1/3}-1)<\frac{\lambda_{ML\max}}{c}<1.6550$, the upper bound of k in (3.47), is used to search for the optimal k_{opt}; otherwise, the upper bound of k in (3.35) is used instead.

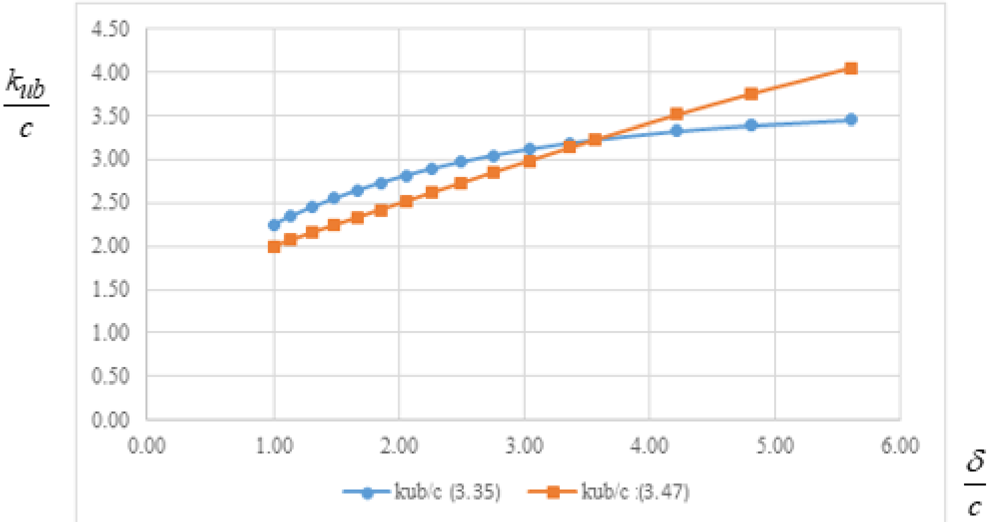


Figure 3.1 Comparison of $\frac{k_{ub}}{c}$ in (3.35) and (3.47) as a function of $\frac{\delta}{c}$.

CHAPTER 4

SIMULATION STUDY

The main focus of this work is to study the effect of multicollinearity on ML and LRR estimators. Therefore, a simulation study was conducted to investigate the efficiency of the proposed ridge estimator and compare it with seven well-known ridge parameter estimators. Since $MSE\left(k \mid \hat{\boldsymbol{\beta}}_{ML}\right)$ is a function of one parameter only, the iterative Nelder-Mead (1965) algorithm was used to search for k_{opt}. Afterward, the capability of the proposed estimator (k_{opt}) and the other ridge regression estimator were compared via their MSE. The details and results of the simulation study are contained in Sections 4.1 – 4.2. Moreover, results with a real-life data example are presented in Section 4.3.

4.1 Details of the Simulation Study

Each explanatory variable ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$) was generated for 10,000,000 observations using a uniform distribution. The relevant explanatory variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$ were in the form of $\mathbf{x}_1 \sim U(10,18)$, $\mathbf{x}_2 \sim U(15,40)$, $\mathbf{x}_3 \sim U(30,50)$, $\mathbf{x}_4 \sim U(2,6)$ and $\mathbf{x}_5 \sim U(0.5,1.5)$. Correlated explanatory variable data was created by applying Spearman's correlation. The process of simulation was as follows:

1) Given the desired correlation matrix $\mathbf{R}^S$, compute the j^{th} element in the adjusted correlation matrix $\mathbf{R}^{adj}$ (Hotelling and Pabst, 1936) as

$$r_{ij}^{adj} = 2\sin\left(\frac{\pi r_{ij}^S}{6}\right). \tag{4.1}$$

2) Perform the Cholesky decomposition of the adjusted correlation matrix as

$$\mathbf{R}^{adj} = \mathbf{L}\mathbf{L}', \tag{4.2}$$

where $\mathbf{L}$ is a lower triangular matrix.

3) Generate correlated standard normal random numbers $\mathbf{r}_c$ as

$$\mathbf{r}_c = \mathbf{L}\mathbf{r} \tag{4.3}$$

where $\mathbf{r}$ comprises the standard normal random numbers.

4) Create the correlated uniform random numbers as

$$\mathbf{u} = (\mathbf{b} - \mathbf{a})F(\mathbf{r}_c) + \mathbf{a} \tag{4.4}$$

where $\mathbf{a}$ and $\mathbf{b}$ are the lower and upper bound vectors of the uniform numbers.

Thus, theoretical correlation matrices of the explanatory variables were created for two cases as follows:

First, for three explanatory variables:

$$\mathbf{R}^S = \begin{bmatrix} 1 & \rho_{12} & 0 \\ \rho_{12} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ where } \rho_{12} = 0.90,\ 0.95,\ 0.99.$$

Second, for five explanatory variables:

$$\mathbf{R}^S = \begin{bmatrix} 1 & \rho_{12} & 0 & 0 & 0 \\ \rho_{12} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \rho_{34} & 0 \\ 0 & 0 & \rho_{34} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where i) $\rho_{12} = 0.90$ and $\rho_{34} = 0.90$,

ii) $\rho_{12} = 0.99$ and $\rho_{34} = 0.90$,

and iii) $\rho_{12} = 0.99$ and $\rho_{34} = 0.99$.

Samples with sample size (n) 100, 200, 500, and 1,000 were randomly selected from the population using a simple random sampling method comprising 500 replications. The explanatory variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$ for each dataset were then standardized by using unit length scaling as $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4, \mathbf{w}_5$. The parameter values of $\boldsymbol{\beta}$ were set as $\beta_0 = 0.3$, $\beta_1 = 2$, $\beta_2 = 1$, $\beta_3 = -1.5$ $\beta_4 = 2.5$, and $\beta_5 = -1.2$. For $\boldsymbol{\beta}$, the dependent variable y_i was generated such that it was Bernoulli distributed with the probability in (2.8). After generating $\mathbf{X}$ and $\mathbf{y}$, $\hat{\boldsymbol{\beta}}_{ML}$ was computed by using the SAS 9.4 logistic regression program (PROC LOGISTIC). This method was repeated to estimate the ridge parameter for each method.

4.2 The Results of the Simulation Study

In this section, the results of the simulation study are presented. The estimators were compared based on the MSE criterion. The performance of the logistic ridge estimator in (3.10) was evaluated through the MSE in Equation (3.7) directly:

$$MSE = \sum_j \frac{\lambda_{ML_j}}{\left(\lambda_{ML_j} + k\right)^2} + k^2 \sum_j \frac{\alpha_j^2}{\left(\lambda_{ML_j} + k\right)^2}. \tag{4.5}$$

Moreover, the MSE of each LRR estimator will be compared to the MSE of MLE and be presented as relative efficiency (RE):

$$RE = \frac{MSE(ML)}{MSE(LRR)} \times 100. \tag{4.6}$$

In addition, the deviance was used as a goodness of fit criterion by considering the error component of the fitted logistic regression model and its aggregate statistics. It is defined as

$$DEV = \sum_{i=1}^{n} d_i^2, \tag{4.7}$$

where the deviance residuals are $d_i = \sqrt{2\,|\ln(\pi_i(\hat{\boldsymbol{\beta}}))\,|}$ for $y_i = 1$ and $d_i = \sqrt{2\,|\ln(1-\pi_i(\hat{\boldsymbol{\beta}}))\,|}$ for $y_i = 0$ (Marx and Smith, 1990; Hosmer, Taber and Lemeshow, 1991).

The results summarized from 500 replications in each simulation cases are shown in Appendix A. The ridge parameters in Appendix A are the medians since the estimates are highly skewed to the right. The medians of seven ridge parameters in Table 4.1 and 4.2 decrease rapidly from the order of hundredths when the correlation coefficient is 0.90 to the order of thousandths when the correlation coefficient increases to 0.99. In some simulation cases, the ridge parameters are greater than unity (see details in Appendix B and C). From the simulation results in Appendix A, it can be concluded that the relative efficiency of the proposed estimator, k_{opt}, with respect to the ML estimator in both cases of three and five explanatory variables is higher than other six well-known estimators in all simulation cases. The relative efficiencies of the proposed estimator, k_{opt}, and other six well-known estimators are summarized in Table 4.1 and 4.2 in the case of three and five explanatory variables respectively. Among the top three highest efficient estimators are k_{opt}, k_{SRW1} and k_{HKB}. The ridge parameter decreases as the degree of multicollinearity increases and, as expected, the ridge parameter is rather stable when the sample size increases. It can be seen from the tables in Appendix A that the deviance of the LRR model is only slightly higher than the deviance of the ML model. The MSEs of the estimated coefficients of the correlated variables in the LRR models, except the LRR model associated with k_{GM}, decreases significantly from the ML case. It should be noted that the efficiency of k_{GM} is lower than the ML estimator in some simulation data sets when the multicollinearity is severe. This is the case of over penalization with the high value of ridge parameter.

The distributions of ridge parameters and k_{ub} are shown in Appendix B and Appendix C. It can be seen that the distributions are very skewed to the right with, in most cases, the maximum value is much greater than unity and the median is much less than unity. The distribution of k_{ub} is not skewed since the mean is approximately equal to the median in all cases as shown in Table 4.3 and 4.4. The upper-bound k_{ub} increases as the sample size increases but is rather stable as the degree of multicollinearity increases.

Table 4.1 The Relative Efficiencies of k_{opt}, k_{HK}, k_{HKB}, k_{SRW1}, k_{SRW2}, k_{GM} and k_{WA} in the Case of Three Explanatory Variables

Sample size	$\rho = 0.90$		$\rho = 0.95$		$\rho = 0.99$	
	k_{opt}	***RE***	k_{opt}	***RE***	k_{opt}	***RE***
100	0.0474	225.80	0.0240	252.12	0.0051	269.98
200	0.0684	253.14	0.0254	259.37	0.0049	273.26
500	0.0570	243.34	0.0270	255.50	0.0053	308.65
1000	0.0618	246.26	0.0240	252.38	0.0050	278.42
	k_{HK}	***RE***	k_{HK}	***RE***	k_{HK}	***RE***
100	0.0150	189.46	0.0096	216.64	0.0022	240.20
200	0.0196	211.49	0.0111	227.39	0.0023	246.06
500	0.0195	208.20	0.0123	229.65	0.0026	277.29
1000	0.0200	211.57	0.0110	225.19	0.0025	255.35
	k_{HKB}	***RE***	k_{HKB}	***RE***	k_{HKB}	***RE***
100	0.0405	218.28	0.0290	243.22	0.0083	256.41
200	0.0561	243.45	0.0346	247.62	0.0085	256.40
500	0.0542	232.80	0.0365	241.94	0.0099	287.87
1000	0.0554	235.17	0.0326	238.39	0.0093	256.62
	k_{SRW1}	***RE***	k_{SRW1}	***RE***	k_{SRW1}	***RE***
100	0.0358	221.03	0.0217	249.54	0.0050	269.73
200	0.0438	245.86	0.0235	257.06	0.0049	273.03
500	0.0409	236.56	0.0252	253.52	0.0053	308.56
1000	0.0408	239.74	0.0225	249.77	0.0050	278.17
	k_{SRW2}	***RE***	k_{SRW2}	***RE***	k_{SRW2}	***RE***
100	0.0943	185.34	0.0678	197.14	0.0190	195.83
200	0.1219	210.64	0.0744	202.28	0.0188	199.30
500	0.1063	201.28	0.0727	199.66	0.0198	248.75
1000	0.1117	202.43	0.0660	196.89	0.0183	204.02
	k_{GM}	***RE***	k_{GM}	***RE***	k_{GM}	***RE***
100	0.2329	108.88	0.1971	116.17	0.1406	95.91
200	0.2868	132.99	0.2331	117.13	0.1516	99.47
500	0.3165	125.05	0.2819	112.46	0.1632	192.46
1000	0.3136	126.67	0.2388	110.82	0.1635	102.36
	k_{WA}	***RE***	k_{WA}	***RE***	k_{WA}	***RE***
100	0.2449	125.96	0.1640	139.00	0.0433	132.03
200	0.3189	147.61	0.1794	141.48	0.0417	135.80
500	0.2781	141.35	0.1814	139.23	0.0446	214.53
1000	0.3125	142.34	0.1646	136.53	0.0415	139.32

Table 4.2 The Relative Efficiencies of k_{opt}, k_{HK}, k_{HKB}, k_{SRW1}, k_{SRW2}, k_{GM} and k_{WA} in the Case of Five Explanatory Variables

Sample size	$\rho_{12}=0.90, \rho_{34}=0.90$		$\rho_{12}=0.99, \rho_{34}=0.90$		$\rho_{12}=0.99, \rho_{34}=0.99$	
	k_{opt}	***RE***	k_{opt}	***RE***	k_{opt}	***RE***
100	0.0247	201.61	0.0052	247.35	0.0027	227.47
200	0.0322	211.30	0.0047	240.18	0.0030	234.79
500	0.0333	212.52	0.0059	256.07	0.0031	244.74
1000	0.0317	217.79	0.0053	246.25	0.0033	234.39
	k_{HK}	***RE***	k_{HK}	***RE***	k_{HK}	***RE***
100	0.0059	157.33	0.0016	209.96	0.0007	177.35
200	0.0085	165.41	0.0017	209.63	0.0008	177.29
500	0.0088	169.47	0.0023	227.31	0.0009	187.77
1000	0.0092	172.97	0.0021	217.67	0.0009	182.06
	k_{HKB}	***RE***	k_{HKB}	***RE***	k_{HKB}	***RE***
100	0.0235	198.61	0.0074	225.56	0.0033	224.15
200	0.0332	208.50	0.0082	214.32	0.0035	232.77
500	0.0372	209.39	0.0110	225.47	0.0039	241.69
1000	0.0359	214.97	0.0100	216.89	0.0042	230.73
	k_{SRW1}	***RE***	k_{SRW1}	***RE***	k_{SRW1}	***RE***
100	0.0144	191.45	0.0041	244.74	0.0017	217.92
200	0.0188	199.39	0.0039	237.61	0.0019	223.67
500	0.0183	200.13	0.0049	253.50	0.0020	232.68
1000	0.0194	205.31	0.0045	243.60	0.0020	222.52
	k_{SRW2}	***RE***	k_{SRW2}	***RE***	k_{SRW2}	***RE***
100	0.0578	170.16	0.0186	171.37	0.0078	180.14
200	0.0732	181.84	0.0186	165.37	0.0089	187.79
500	0.0781	182.74	0.0231	178.84	0.0093	197.60
1000	0.0753	187.89	0.0212	170.54	0.0095	188.46
	k_{GM}	***RE***	k_{GM}	***RE***	k_{GM}	***RE***
100	0.1756	106.15	0.1102	97.00	0.0842	95.88
200	0.2147	117.02	0.1390	96.74	0.0655	103.66
500	0.2588	114.72	0.1769	105.54	0.0971	110.20
1000	0.2350	120.92	0.1456	97.61	0.0952	101.06
	k_{WA}	***RE***	k_{WA}	***RE***	k_{WA}	***RE***
100	0.1533	118.13	0.0443	122.14	0.0197	126.26
200	0.1904	129.67	0.0458	120.35	0.0214	132.41
500	0.2013	130.65	0.0545	131.44	0.0227	141.62
1000	0.1953	132.86	0.0513	122.57	0.0233	131.96

Table 4.3 The Mean and Median of k_{ub} in the Case of Three Explanatory Variables

Sample Size	$\rho = 0.90$		$\rho = 0.95$		$\rho = 0.99$	
	mean	***median***	***mean***	***median***	***mean***	***median***
100	6.8624	6.5188	6.9255	6.6553	6.9561	6.6929
200	8.8410	8.4438	8.7374	8.3725	8.6410	8.3298
500	11.7957	11.6883	11.8902	11.7225	11.7974	11.5859
1000	11.5886	11.6508	11.4886	11.4834	11.5503	11.5986

Table 4.4 The Mean and Median of k_{ub} in the Case of Five Explanatory Variables

Sample Size	$\rho_{12} = 0.90, \rho_{34} = 0.90$		$\rho_{12} = 0.99, \rho_{34} = 0.90$		$\rho_{12} = 0.99, \rho_{34} = 0.99$	
	mean	***median***	***mean***	***median***	***mean***	***median***
100	7.0166	6.8764	6.9464	6.7896	7.1101	6.9007
200	9.2462	9.0394	9.2953	9.0181	7.1033	7.0632
500	12.8172	12.6166	12.9257	12.7355	9.1942	9.1594
1000	12.2826	12.2581	12.2817	12.3705	10.5164	10.5190

4.3 A Real-Life Data Example

The Lee cancer remission dataset (Lee, 1974; Marx, 1988) taken from SAS, the SUGI Supplementary Guide (Hastings, 1986), was used to demonstrate the efficacy of k_{opt}. The binary response was 1 if a patient went into complete cancer remission and 0 otherwise. The explanatory variables are cell of the marrow clot section (CELL), smear differential percentage of blasts (SMEAR), percentage of absolute marrow leukemia cell infiltrate (INFIL), percentage labeling index of the bone marrow leukemia cells (LI), and the maximum temperature ahead of treatment (TEMP). There were 27 patients in this study. Before applying the LRR technique, the explanatory variables were centered and scaled by using a unit length scaling method. The correlation matrix between the explanatory variables is reported in Table 4.5. Noticeably, SMEAR and INFIL are highly correlated, while CELL and INFIL are moderately correlated, which implies multicollinearity.

Table 4.5 The Correlation Matrix of the Explanatory Variables in the Lee Cancer Remission Dataset ($n = 27$).

	Correlation Matrix of the Explanatory Variables				
	CELL	**SMEAR**	**INFIL**	**LI**	**TEMP**
CELL	1.0000	0.2918	0.6071	0.1902	0.1082
SMEAR		1.0000	0.9297	0.3175	-0.1125
INFIL			1.0000	0.3211	-0.0445
LI				1.0000	-0.0548
TEMP					1.0000

The model corresponding to these standardized variables is

$$\log it(\hat{\pi}(\mathbf{x}_i)) = \hat{\beta}_0 + \hat{\beta}_1 CELL_i + \hat{\beta}_2 SMEAR_i + \hat{\beta}_3 INFIL_i + \hat{\beta}_4 LI_i + \hat{\beta}_5 TEMP_i, \quad i = 1, 2, ..., n. \tag{4.8}$$

Estimates of the standardized regression coefficients and standard error (in parentheses), ridge parameter, MSE, relative error (RE), and deviance (DEV) by the ML and LRR estimators are reported in Table 4.5, in which CELL, SMEAR, and INFIL are obviously highly correlated. The standard errors are based, in part, on the correlation between the variables in the model, and it is evident that those of the estimated regression coefficients of CELL, SMEAR, and INFIL (corresponding to $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, respectively) were explicitly inflated by the ML estimator for Model (4.8) due to the multicollinearity problem. Next, the standardized regression coefficients in Table 4.6 were converted back into the original units of data, as reported in Table 4.7.

Table 4.6 Estimates of Standardized Regression Coefficients (Standard Error), Ridge Parameter (k), MSE, RE and DEV by the ML and LRR estimators.

Variable	Method							
	ML	k_{opt}	k_{HK}	k_{HKB}	k_{SRW1}	k_{SRW2}	k_{GM}	k_{WA}
Constant	-2.3111	-1.7855	-2.0011	-1.7972	-1.7882	-1.5098	-1.0071	-1.2802
	(1.8001)	(1.0565)	(1.2319)	(1.0621)	(1.0578)	(0.9312)	(0.7163)	(0.8279)
CELL	23.0121	8.5009	13.7348	8.7021	8.5463	5.8148	2.8942	4.3921
	(44.975)	(7.7147)	(19.6764)	(8.0339)	(7.7849)	(5.2077)	(3.2967)	(4.3109)
SMEAR	20.0497	0.7390	7.4003	0.9551	0.7871	-0.7988	-0.4977	-0.7402
	(61.3591)	(7.2466)	(25.854)	(7.8438)	(7.3793)	(2.8289)	(2.0315)	(2.3497)
INFIL	-22.3814	0.1783	-7.5863	-0.0712	0.1229	1.8532	1.2099	1.6580
	(71.7846)	(8.3556)	(30.2193)	(9.0643)	(8.5133)	(2.9498)	(1.876)	(2.2955)
LI	9.5107	8.8752	9.2313	8.9057	8.8824	7.8773	5.6784	6.9103
	(4.536)	(4.264)	(4.3991)	(4.277)	(4.2671)	(3.811)	(2.8411)	(3.3765)
TEMP	-6.5271	-6.0361	-6.3496	-6.0663	-6.0433	-5.0097	-2.8163	-4.0230
	(4.9092)	(4.7173)	(4.8512)	(4.7318)	(4.7208)	(4.199)	(3.0594)	(3.6936)
k	0	0.00074	0.00013	0.00067	0.00072	0.00382	0.01682	0.00814
MSE	10,988.64	1,316.74	2,478.38	1,318.01	1,316.80	1,400.18	1,450.10	1,426.02
RE		834.53	443.38	833.73	834.50	784.80	757.78	770.58
DEV	21.7550	21.8746	21.8002	21.8702	21.8736	22.0482	23.2243	22.3968

In the experiment, the value of the ridge parameter was in the range $0.0001 < k < 0.017$. The ridge method was effective in significantly reducing the MSE of the ML estimator, and the size of the regression coefficients shrank depending on the value of k . The estimated regression coefficients due to the LRR methods were smaller than those of the ML estimator, especially those of the explanatory variables with multicollinearity (i.e. CELL, SMEAR and INFIL corresponding to $\hat{\beta}_{LRR,1}$, $\hat{\beta}_{LRR,2}$ and $\hat{\beta}_{LRR,3}$ respectively), and the regression coefficient sign of INFIL differed from the ML estimator. The proposed estimator, k_{opt} , produced the lowest MSE. Interestingly, k_{opt} and k_{SRW1} realized quite similar MSE and k values. The deviances of the methods were not very different.

Marx and Smith (1990) recommended a formula to convert the standardized regression coefficients to the original unit (uncentered and unscaled regression coefficients):

$$b_j = q_j^{-1}\hat{\beta}_{LRR,j}, \quad j = 1, 2, ..., p, \tag{4.9}$$

and
$$b_0 = \hat{\beta}_{LRR,0} - \sum_j q_j^{-1}\bar{x}_j\hat{\beta}_{LRR,j}, \quad j = 1, 2, ..., p, \tag{4.10}$$

where $q_j = \sqrt{\sum_i (x_{ij} - \bar{x}_j)^2}$.

The standard errors associated with the uncentered and unscaled LRR estimators are defined as

$$SE(b_j) = q_j^{-1}SE(\hat{\beta}_{LRR,j}), \quad j = 1, 2, ..., p, \tag{4.11}$$

and
$$SE(b_0) = \Big\{Var(\hat{\beta}_{LRR,0}) + \sum_{j=1}^{p} (q_j^{-1}\bar{x}_j)^2 Var(\hat{\beta}_{LRR,j}) + 2\sum_{i<j}\sum_{j\neq 0} q_i^{-1}q_j^{-1}\bar{x}_i\bar{x}_j Cov(\hat{\beta}_{LRR,i}, \hat{\beta}_{LRR,j}) - 2\sum_{j=1}^{p} q_j^{-1}\bar{x}_j Cov(\hat{\beta}_{LRR,0}, \hat{\beta}_{LRR,j})\Big\}^{1/2}. \tag{4.12}$$

The estimates of regression coefficients and standard error with the ML and LRR estimators are summarized in Table 4.7.

The estimates of standardized and unstandardized regression coefficients of each method in Table 4.6 and Table 4.7 are the same except for the constant term. When comparing the estimates of regression coefficients with the ML and LRR estimators in Table 4.7, it was found that b_3 with k_{HK} and k_{HKB} was similar to ML but the original negative estimate changed positive with k_{opt} and k_{SRW1} . Similarly, the original positive estimate of b_2 changed to negative with k_{SRW2}, k_{GM} , and k_{WA}. Therefore, in practice, the appropriate methods are k_{opt} and k_{SRW1}, which also provided values of k

and MSE that were identical. In addition, when comparing the prediction percentage with both methods, the results were the same.

Table 4.7 Estimates of Regression Coefficients (Standard Error) by ML and LRR Estimators.

Variable	Method							
	ML	k_{opt}	k_{HK}	k_{HKB}	k_{SRW1}	k_{SRW2}	k_{GM}	k_{WA}
Constant	57.1285	65.5110	64.2034	65.6879	65.5553	55.2955	30.6967	44.3251
	(69.9768)	(57.7647)	(61.0266)	(57.9488)	(57.8079)	(51.9808)	(39.1675)	(46.3809)
CELL	24.1799	8.9323	14.4318	9.1437	8.9800	6.1099	3.0411	4.6149
	(47.2573)	(8.1062)	(20.6749)	(8.4415)	(8.18)	(5.472)	(3.4639)	(4.5297)
SMEAR	18.3697	0.6771	6.7802	0.8751	0.7211	-0.7319	-0.4560	-0.6782
	(56.2177)	(6.6394)	(23.6877)	(7.1865)	(6.761)	(2.5919)	(1.8612)	(2.1528)
INFIL	-18.4763	0.1472	-6.2627	-0.0588	0.1014	1.5298	0.9988	1.3687
	(59.2597)	(6.8977)	(24.9467)	(7.4828)	(7.0279)	(2.4352)	(1.5486)	(1.895)
LI	3.9872	3.7208	3.8701	3.7336	3.7238	3.3024	2.3806	2.8970
	(1.9017)	(1.7876)	(1.8442)	(1.7931)	(1.7889)	(1.5977)	(1.1911)	(1.4156)
TEMP	-86.1371	-79.6578	-83.7940	-80.0551	-79.7520	-66.1124	-37.1658	-53.0908
	(64.7854)	(62.2538)	(64.0202)	(62.4449)	(62.2993)	(55.4136)	(40.3742)	(48.7443)

As illustrated clearly in the real-life example, the LRR approach is a good alternative to the ML estimator when faced with the multicollinearity problem. Choosing an appropriate LRR estimator depends on the purpose of the user and the processing capabilities of the computer used. The optimal LRR method proposed in this study searches for an efficient LRR parameter whereas the others use methods to estimate the unknown LRR parameter for the dataset.

CHAPTER 5

CONCLUSIONS AND FUTURE RESEARCH

This dissertation presents a solution to solve the problem of determining the optimal ridge parameter in logistic regression by using one of the efficient searches for finding the optimal of a non-linear performance measure. A theorem on the upper-bound of the ridge parameter based on the eigenvalues of the explanatory variables is developed to facilitate the numerical search. The following are the conclusions of the study in section 5.1 and recommendations for future work in section 5.2.

5.1 Conclusions

This dissertation demonstrates that it is quite convenient to compute the optimal ridge parameter, instead of the conventional approximations of ridge parameter, due to the ubiquity of powerful computing capability. A theorem on the upper-bound of the optimal ridge parameter estimator, k_{opt}, is developed by following the eigen approach such that k_{opt} which minimizes the MSE of the estimates of the coefficients in the LRR model can be searched effectively in a specified small interval as shown in Table 4.3 and 4.4. A simulation is used to evaluate the relative efficiencies of the proposed k_{opt} and other six well-known ridge parameter estimators, $k_{HK}, k_{HKB}, k_{SW1}, k_{SW2}, k_{GM}$ and k_{WA} with respect to the ML estimator. The simulation results show that the relative efficiency of the proposed k_{opt} is highest among the compared well-known estimators k_{HKB} and k_{SRW1} are good alternatives as shown in Table 4.1 and 4.2. Also, the simulation result suggests that the ridge parameter in some cases may be greater than unity. Additionally, using a real-life data set of small size, comparisons with the same six estimators show that the relative efficiency of the estimator with the optimal ridge parameter is also better than or equal to others.

5.2 Recommendations for Future Work

The effectiveness of direct search could be improved further by decreasing the upper-bound k_{ub} with a better approximation than the first order approximation as in the proof of theorem. Other better direct search than the iterative Nelder-Mead Algorithm should be investigated. Furthermore, the concept of direct search should be extended to solve the ill-conditioned information matrix in other statistical estimation.

BIBLIOGRAPHY

Akay, K. U. 2014. A Graphical Evaluation of Logistic Ridge Estimator in Mixture Experiments. **Journal of Applied Statistics.** 41 (6): 1217 – 1232.

Al Turk, L. I. and Alsomahi, A. A. 2014. On Enhancing the Dorugade and Kashid's Ridge Parameter in Ridge Regression. **Applied Mathematical Sciences.** 8 (152): 7553 – 7565.

Alkhamisi, M., Khalaf, G. and Shukur, G. 2006. Some Modifications for Choosing Ridge Parameters. **Communications in Statistics – Theory and Methods.** 35 (11): 2005 – 2020.

Allison, P. D. 2000. **Logistic Regression Using SAS: Theory and Application.** Cary, NC: SAS Institute.

Antoniadis, A. and Fan, J. 2001. Regularization of Wavelets Approximations. **Journal of the American Statistical Association.** 96 (455): 939 – 967.

Asar, Y. 2017. Some New Methods to Solve Multicollinearity in Logistic Regression. **Communications in Statistics – Simulation and Computation.** 46 (4): 2576 – 2586.

Asar, Y.; Arashi, M. and Wu, J. 2017. Restricted Ridge Estimator in the Logistic Regression Model. **Communications in Statistics – Simulation and Computation.** 46 (8): 1 – 7.

Baeyens, E.; Herreros, A. and Perán, J. R. 2016. A Direct Search Algorithm for Global Optimization. **Algorithms.** 9 (2): 40.

Conniffe, D. and Stone, J. 1973. A Critical View of Ridge Regression. **Journal of the Royal Statistical Society Series D (The Statistician).** 22 (3): 181 – 187.

Cox, D. R. and Hinkley, D. V. 1974. **Theoretical Statistics.** London: Chapman and Hall.

Crotty, M. and Barker, C. 2014. **Penalizing Your Models: An Overview of the Generalized Regression Platform.** Cary, NC: SAS Institute.

De Grange, L.; Fariña, P. and De Dios Ortúzar, J. 2015. Dealing with Collinearity in Travel Time Valuation. **Transportmetrica A: Transport Science.** 11 (4): 317 – 332.

Dorugade, A. V. 2014. New Ridge Parameters for Ridge Regression. **Journal of the Association of Arab Universities for Basic and Applied Sciences.** 15: 94 – 99.

Duffy, D. E. and Santner, T. J. 1989. On the Small Sample Properties of Norm-Restricted ML Estimators for Logistic Regression Models. **Communications in Statistics – Theory and Methods.** 18 (3): 959 – 980.

Gujarati, D. N. and Porter, D. C. 2010. **Essentials of Econometrics.** 4^{th} ed. Singapore: The McGraw – Hill.

Hastings, R. P. 1986. **SUGI Supplemental Library User's Guide.** Cary, NC: SAS Institute.

Hoerl, A. E. and Kennard, R. W. 1970A. Ridge Regression: Applications to Nonorthogonal Problems. **Technometrics.** 12 (1): 69 – 82.

Hoerl, A. E. and Kennard, R. W. 1970B. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics.** 12 (1): 55 – 67.

Hoerl, A. E.; Kannard, R. W. and Baldwin, K. F. 1975. Ridge Regression: Some Simulations. **Communications in Statistics – Theory and Methods.** 4 (2): 105 – 123.

Hosmer, D. and Lemeshow, S. 1989. **Applied Logistic Regression.** New York: John Wiley & Sons.

Hosmer, D. W. and Lemeshow, S. 2000. **Applied Logistic Regression.** 2^{nd} ed. New York: John Wiley & Sons.

Hosmer, D. W.; Taber, S. and Lemeshow, S. 1991. The Importance of Assessing the Fit of Logistic Regression Models: A Case Study. **American Journal of Public Health.** 81 (12): 1630 – 1635.

Hotelling, H. and Pabst, M. R. 1936. Rank Correlation and Tests of Significance Involving No Assumptions of Normality. **Annals of Mathematical Statistics.** 7: 29 – 43.

Jirawan Jitthavech. 2015. **Regression Analysis.** Bangkok: WVO Office of Printing Mill. [In Thai]

Khalaf, G. and Shukur, G. 2005. Choosing Ridge Parameter for Regression Problems. **Communications in Statistics – Theory and Methods.** 34 (5): 1177 – 1182.

Kibria, B. G. 2003. Performance of Some New Ridge Regression Estimators. **Communications in Statistics – Simulation and Computation.** 32 (2): 419 – 435.

Kibria, B. M. G.; Månsson, K. and Shukur, G. 2012. Performance of Some Logistic Ridge Regression Estimators. **Computational Economics.** 40 (4): 401 – 414.

Le Cessie, S. and Van Houwelingen, J. C. 1992. Ridge Estimators in Logistic Regression. **Applied statistics.** 41 (1): 191 – 201.

Lee, A. H. and Silvapulle, M. J. 1988. Ridge Estimation in Logistic Regression. **Communications in Statistics – Simulation and Computation.** 17 (4): 1231 – 1257.

Lee, E. T. 1974. A Computer Program for Linear Logistic Regression Analysis. **Computer Programs in Biomedicine.** 4 (2): 80 – 92.

Makalic, E. and Schmidt, D. F. 2011. Review of Modern Logistic Regression Methods with Application to Small and Medium Sample Size Problems. In **AI 2010: Advances in Artificial Intelligence.** Li J., ed. Berlin: Springer. Pp. 213 – 222.

Mansson, K. and Shukur, G. 2011. On Ridge Parameters in Logistic Regression. **Communications in Statistics – Theory and Methods.** 40 (18): 3366 – 3381.

Marx, B. D. 1988. **Ill-Conditioned Information Matrices and Generalized Linear Model: An Asymptotically Biased Estimation Approach.** Doctoral dissertation, Virginia Polytechnic Institute and State University, USA.

Marx, B. D. and Smith, E. P. 1990. Weighted Multicollinearity in Logistic Regression: Diagnostics and Biased Estimation Techniques with an Example from Lake Acidification. **Canadian Journal of Fisheries and Aquatic Sciences.** 47 (6): 1128 – 1135.

Meier, L.; Van De Geer, S. and Bühlmann, P. 2008. The Group Lasso for Logistic Regression. **Journal of the Royal Statistical Society: Series B (Statistical Methodology).** 70 (1): 53 – 71.

Midi, H.; Sarkar, S. K. and Rana, S. 2010. Collinearity Diagnostics of Binary Logistic Regression Model. **Journal of Interdisciplinary Mathematics.** 13 (3): 253 – 267.

Muniz, G. and Kibria, B. G. 2009. On Some Ridge Regression Estimators: An Empirical Comparison. **Communications in Statistics – Simulation and Computation.** 38 (3): 621 – 630.

Muniz, G. B. M.; Kibria, G.; Mansson, K. and Shukur, G. 2012. On Developing Ridge Regression Parameters: A Graphical Investigation. **Statistics and Operations Research Transactions.** 36 (2): 115 – 138.

Nelder, J. A. and Mead, R. 1965. A Simplex Method for Function Minimization. **The Computer Journal.** 7: 308 – 313.

Okeh, U. M. and Oyeka, I. C. A. 2013. Estimating the Fisher's Scoring Matrix Formula from Logistic Model. **American Journal of Theoretical and Applied Statistics.** 2 (6): 221 – 227.

Özkale, M. R. 2016. Iterative Algorithms of Biased Estimation Methods in Binary Logistic Regression. **Statistical Papers.** 57 (4): 991 – 1016.

Özkale, M. R. and Arıcan, E. 2016. A New Biased Estimator in Logistic Regression Model. **Statistics.** 50 (2): 233 – 253.

Özkale, M. R.; Lemeshow, S. and Sturdivant, R. 2017. Logistic Regression Diagnostics in Ridge Regression. **Computational Statistics.** 1 – 31.

Rashid, M. 2008. **Inference on Logistic Regression Model.** Doctoral dissertation, Bowling Green State University, USA.

Rashid, M. and Shifa, N. 2009. Consistency of the Maximum Likelihood Estimator in Logistic Regression Model: A Different Approach. **Journal of Statistics.** 16 (1): 1 – 11.

Ryan, T. P. 1997. **Modern Regression Methods.** New York: John Wiley & Sons.

Schaefer, R. L. 1979. **Multicollinearity and Logistic Regression.** Doctoral dissertation, the University of Michigan, USA.

Schaefer, R. L. 1986. Alternative Estimators in Logistic Regression When the Data are Collinear. **Journal of Statistical Computation and Simulation.** 25 (1 – 2): 75 – 91.

Schaefer, R. L.; Roi, L. D. and Wolfe, R. A. 1984. A Ridge Logistic Estimator. **Communications in Statistics – Theory and Methods.** 13 (1): 99 – 113.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society. Series B (Methodological).** 58 (1): 267 – 288.

Vágó, E. and Kemény, S. 2006. Logistic Ridge Regression for Clinical Data Analysis (A Case Study). **Applied Ecology and Environmental Research.** 4 (2): 171 – 179.

Wu, J., and Asar, Y. 2016. On Almost Unbiased Ridge Logistic Estimator for the Logistic Regression Model. **Hacettepe Journal of Mathematics and Statistics.** 45 (3): 989 – 998.

APPENDICES

Appendix A

The Results of the Simulation Study

The results of the simulation study are presented in two parts. First, in case of three explanatory variables, the ridge parameter k and the estimators at each correlation level are reported in Appendix A.1. The median of ridge parameter k, the estimated standardized regression coefficients, the squared bias, variance, MSE, and deviance of the estimated coefficients are reported in Tables A.1.1 to Tables A.1.12. Second, in case of five explanatory variables, the ridge parameter k and the estimators at each correlation level are presented in Appendix A.2. The medians of the ridge parameter k, the estimated standard regression coefficients, the absolute bias, MSE, RE and deviance of the estimators are summarized in Tables A.2.1 – A.2.12.

A.1 The Results of the Simulation Study in case of Three Explanatory Variables

The effect of varying the correlation level and sample size on the performance of the ML and LRR estimators are presented.

Table A.1.1 The Results of the ML and LRR Estimator Performances for $\rho = 0.90$ and $n = 100$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2919	0.0000	0.0440	
	w_1	2.9193	0.0000	23.6342	
	w_2	0.7548	0.0000	23.5304	
	w_3	-1.8411	0.0000	4.5733	
	DEV=129.8437	Total	0.0000	51.7819	100.00
KOPT	constant	0.2838	0.0082	0.0436	
k = 0.0474	w_1	2.2681	1.9455	9.9337	
	w_2	0.8174	1.8468	9.6290	
	w_3	-1.4494	0.5226	3.3266	
	DEV=130.2447	Total	4.3231	22.9328	225.80
HK	constant	0.2887	0.0032	0.0439	
k = 0.0150	w_1	2.6274	1.1348	11.7771	
	w_2	0.8470	1.1040	11.6535	
	w_3	-1.6941	0.2068	3.8571	
	DEV=129.9360	Total	2.4489	27.3315	189.46
HKB	constant	0.2850	0.0070	0.0438	
k = 0.0405	w_1	2.2894	2.0525	10.2913	
	w_2	0.9458	1.9803	10.0413	
	w_3	-1.5296	0.4265	3.3464	
	DEV=130.1691	Total	4.4662	23.7227	218.28
SRW1	constant	0.2858	0.0063	0.0438	
k = 0.0358	w_1	2.3793	1.8077	10.0728	
	w_2	0.8912	1.7424	9.8589	
	w_3	-1.5494	0.3949	3.4522	
	DEV=130.1047	Total	3.9512	23.4275	221.03
SRW2	constant	0.2799	0.0123	0.0435	
k = 0.0943	w_1	1.8881	2.8660	12.7305	
	w_2	0.9667	2.7290	12.1257	
	w_3	-1.2790	0.7389	3.0388	
	DEV=130.6210	Total	6.3462	27.9386	185.34
GM	constant	0.2726	0.0197	0.0430	
k = 0.2329	w_1	1.2275	3.8805	22.9882	
	w_2	0.9441	3.6589	21.1044	
	w_3	-0.8625	1.2219	3.4236	
	DEV=131.8539	Total	8.7811	47.5592	108.88
WA	constant	0.2709	0.0214	0.0431	
k = 0.2449	w_1	1.3047	3.7233	19.6195	
	w_2	0.8647	3.4908	18.0261	
	w_3	-0.8762	1.2267	3.4210	
	DEV=131.6744	Total	8.4622	41.1096	125.96

Note: $k_{ub} = 6.5188$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.2 The Results of the ML and LRR Estimator Performances for $\rho = 0.90$ and $n = 200$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2832	0.0000	0.0211	
	w_1	2.4273	0.0000	22.3219	
	w_2	0.7909	0.0000	22.2898	
	w_3	-1.4765	0.0000	4.2925	
	DEV=267.3959	Total	0.0000	48.9253	100.00
KOPT	constant	0.2788	0.0044	0.0209	
k = 0.0684	w_1	1.8019	1.7928	8.2990	
	w_2	0.8494	1.7578	8.1161	
	w_3	-1.1340	0.5005	2.8914	
	DEV=267.8046	Total	4.0555	19.3273	253.14
HK	constant	0.2816	0.0016	0.0210	
k = 0.0196	w_1	2.1311	1.1081	9.8597	
	w_2	0.8898	1.0879	9.7911	
	w_3	-1.3407	0.2058	3.4615	
	DEV=267.4962	Total	2.4035	23.1334	211.49
HKB	constant	0.2797	0.0036	0.0210	
k = 0.0561	w_1	1.8326	1.9255	8.6796	
	w_2	0.9535	1.8885	8.4774	
	w_3	-1.1926	0.4191	2.9189	
	DEV=267.7363	Total	4.2366	20.0969	243.45
SRW1	constant	0.2802	0.0030	0.0210	
k = 0.0438	w_1	1.9092	1.6614	8.4961	
	w_2	0.9347	1.6234	8.3179	
	w_3	-1.2257	0.3678	3.0644	
	DEV=267.6492	Total	3.6556	19.8993	245.86
SRW2	constant	0.2772	0.0061	0.0208	
k = 0.1219	w_1	1.5201	2.5649	10.5613	
	w_2	0.9434	2.4838	10.0160	
	w_3	-0.9977	0.6839	2.6291	
	DEV=268.1331	Total	5.7386	23.2273	210.64
GM	constant	0.2727	0.0106	0.0207	
k = 0.2868	w_1	1.0852	3.4027	17.8811	
	w_2	0.8525	3.2014	16.0686	
	w_3	-0.7047	1.0706	2.8194	
	DEV=269.1092	Total	7.6853	36.7899	132.99
WA	constant	0.2722	0.0111	0.0207	
k = 0.3189	w_1	1.0629	3.2884	15.8358	
	w_2	0.7908	3.1083	14.3530	
	w_3	-0.6696	1.1144	2.9354	
	DEV=269.1005	Total	7.5222	33.1449	147.61

Note: $k_{ub} = 8.4438$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.3 The Results of the ML and LRR Estimator Performances for $\rho = 0.90$ and $n = 500$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2350	0.0000	0.0082	
	w_1	2.1443	0.0000	21.6145	
	w_2	1.0165	0.0000	21.6150	
	w_3	-1.6124	0.0000	4.1397	
	DEV=680.7005	Total	0.0000	47.3775	100.00
KOPT	constant	0.2336	0.0014	0.0082	
k = 0.0570	w_1	1.6513	1.7853	8.3273	
	w_2	0.9965	1.7581	8.2532	
	w_3	-1.2469	0.4727	2.8805	
	DEV=681.1003	Total	4.0175	19.4693	243.34
HK	constant	0.2345	0.0006	0.0082	
k = 0.0195	w_1	1.9020	1.1452	9.6941	
	w_2	1.0786	1.1368	9.6763	
	w_3	-1.4648	0.1958	3.3770	
	DEV=680.8064	Total	2.4783	22.7557	208.20
HKB	constant	0.2338	0.0012	0.0082	
k = 0.0542	w_1	1.6411	2.0081	8.7982	
	w_2	1.1121	1.9832	8.7062	
	w_3	-1.3032	0.4078	2.8370	
	DEV=681.0678	Total	4.4002	20.3495	232.80
SRW1	constant	0.2340	0.0010	0.0082	
k = 0.0409	w_1	1.7605	1.6525	8.5112	
	w_2	1.0743	1.6372	8.4712	
	w_3	-1.3556	0.3350	3.0372	
	DEV=680.9469	Total	3.6257	20.0279	236.56
SRW2	constant	0.2330	0.0020	0.0082	
k = 0.1063	w_1	1.4218	2.5878	10.5677	
	w_2	1.0526	2.5469	10.3879	
	w_3	-1.1169	0.6421	2.5741	
	DEV=681.4344	Total	5.7787	23.5379	201.28
GM	constant	0.2313	0.0037	0.0081	
k = 0.3165	w_1	0.9995	3.4539	17.9786	
	w_2	0.9068	3.3738	17.3252	
	w_3	-0.7904	1.0600	2.5738	
	DEV=682.4427	Total	7.8913	37.8858	125.05
WA	constant	0.2314	0.0036	0.0081	
k = 0.2781	w_1	1.0267	3.3026	15.5631	
	w_2	0.8516	3.2288	15.1251	
	w_3	-0.7615	1.0733	2.8215	
	DEV=682.3935	Total	7.6083	33.5178	141.35

Note: $k_{ub} = 11.6883$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.4 The Results of the ML and LRR Estimator Performances for $\rho = 0.90$ and $n = 1000$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2715	0.0000	0.0041	
	w_1	2.0590	0.0000	21.4991	
	w_2	0.9502	0.0000	21.4967	
	w_3	-1.5177	0.0000	4.1199	
	DEV=1362.7780	Total	0.0000	47.1198	100.00
KOPT	constant	0.2707	0.0008	0.0041	
k = 0.0618	w_1	1.6413	1.7386	8.1334	
	w_2	0.8536	1.7220	8.1226	
	w_3	-1.1680	0.4899	2.8738	
	DEV=1363.1779	Total	3.9514	19.1338	246.26
HK	constant	0.2712	0.0003	0.0041	
k = 0.0200	w_1	1.8672	1.1187	9.4706	
	w_2	0.9483	1.1084	9.4677	
	w_3	-1.3746	0.2102	3.3295	
	DEV=1362.8857	Total	2.4376	22.2719	211.57
HKB	constant	0.2708	0.0007	0.0041	
k = 0.0554	w_1	1.6369	1.9839	8.6315	
	w_2	0.9539	1.9613	8.5877	
	w_3	-1.2170	0.4245	2.8135	
	DEV=1363.1543	Total	4.3704	20.0369	235.17
SRW1	constant	0.2710	0.0006	0.0041	
k = 0.0408	w_1	1.7286	1.6178	8.3315	
	w_2	0.9391	1.6034	8.3161	
	w_3	-1.2690	0.3554	3.0029	
	DEV=1363.0308	Total	3.5771	19.6545	239.74
SRW2	constant	0.2704	0.0011	0.0041	
k = 0.1117	w_1	1.4114	2.5484	10.4009	
	w_2	0.9124	2.5195	10.2873	
	w_3	-1.0402	0.6594	2.5853	
	DEV=1363.5214	Total	5.7284	23.2774	202.43
GM	constant	0.2694	0.0021	0.0041	
k = 0.3136	w_1	0.9836	3.3814	17.6298	
	w_2	0.8510	3.3224	17.0132	
	w_3	-0.7653	1.0371	2.5518	
	DEV=1364.4292	Total	7.7430	37.1990	126.67
WA	constant	0.2694	0.0021	0.0041	
k = 0.3125	w_1	1.0035	3.2318	15.2991	
	w_2	0.7641	3.2041	14.9278	
	w_3	-0.7096	1.0826	2.8728	
	DEV=1364.4444	Total	7.5206	33.1038	142.34

Note: $k_{ub} = 11.6507$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.5 The Results of the ML and LRR Estimator Performances for $\rho = 0.95$ and $n = 100$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2992	0.0000	0.0439	
	w_1	2.3046	0.0000	45.7270	
	w_2	1.2920	0.0000	45.7353	
	w_3	-1.9021	0.0000	4.5569	
	DEV=130.0871	Total	0.0000	96.0632	100.00
KOPT	constant	0.2921	0.0073	0.0435	
k = 0.0240	w_1	1.8282	2.6069	17.3223	
	w_2	1.2980	2.5942	17.2437	
	w_3	-1.5690	0.4083	3.4933	
	DEV=130.4239	Total	5.6168	38.1027	252.12
HK	constant	0.2964	0.0029	0.0438	
k = 0.0096	w_1	2.0721	1.6429	20.1705	
	w_2	1.3682	1.6406	20.1523	
	w_3	-1.7717	0.1637	3.9763	
	DEV=130.1756	Total	3.4500	44.3430	216.64
HKB	constant	0.2930	0.0064	0.0437	
k = 0.0290	w_1	1.8385	2.9378	18.0282	
	w_2	1.3984	2.9256	17.9183	
	w_3	-1.6168	0.3528	3.5057	
	DEV=130.3993	Total	6.2227	39.4957	243.22
SRW1	constant	0.2938	0.0055	0.0437	
k = 0.0217	w_1	1.9097	2.5199	17.4384	
	w_2	1.3689	2.5163	17.3843	
	w_3	-1.6466	0.3144	3.6300	
	DEV=130.3218	Total	5.3562	38.4963	249.54
SRW2	constant	0.2884	0.0110	0.0435	
k = 0.0678	w_1	1.5918	3.9123	22.8760	
	w_2	1.3260	3.8828	22.6310	
	w_3	-1.3906	0.6187	3.1771	
	DEV=130.7836	Total	8.4248	48.7275	197.14
GM	constant	0.2803	0.0194	0.0428	
k = 0.1971	w_1	1.1372	5.1654	39.8641	
	w_2	1.0648	5.0817	39.4491	
	w_3	-0.9360	1.1299	3.3354	
	DEV=131.9489	Total	11.3965	82.6912	116.17
WA	constant	0.2808	0.0189	0.0431	
k = 0.1640	w_1	1.2333	4.8139	33.1328	
	w_2	1.1114	4.7519	32.5870	
	w_3	-1.0281	1.0351	3.3479	
	DEV=131.6260	Total	10.6198	69.1109	139.00

Note: $k_{ub} = 6.6553$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.6 The Results of the ML and LRR Estimator Performances for $\rho = 0.95$ and $n = 200$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2915	0.0000	0.0211	
	w_1	2.2700	0.0000	43.6179	
	w_2	1.1057	0.0000	43.6069	
	w_3	-1.3964	0.0000	4.2942	
	DEV=267.1187	Total	0.0000	91.5402	100.00
KOPT	constant	0.2879	0.0036	0.0210	
k = 0.0254	w_1	1.8163	2.4876	16.0464	
	w_2	1.1177	2.4845	16.0244	
	w_3	-1.1336	0.3948	3.2010	
	DEV=267.4550	Total	5.3705	35.2927	259.37
HK	constant	0.2901	0.0014	0.0211	
k = 0.0111	w_1	2.0586	1.6407	18.2886	
	w_2	1.1653	1.6336	18.2643	
	w_3	-1.2920	0.1662	3.6825	
	DEV=267.2138	Total	3.4419	40.2565	227.39
HKB	constant	0.2884	0.0031	0.0210	
k = 0.0346	w_1	1.7831	2.8961	16.8926	
	w_2	1.2566	2.8859	16.8336	
	w_3	-1.1734	0.3400	3.2211	
	DEV=267.4468	Total	6.1251	36.9683	247.62
SRW1	constant	0.2889	0.0026	0.0210	
k = 0.0235	w_1	1.8916	2.4199	16.1372	
	w_2	1.1879	2.4084	16.0977	
	w_3	-1.2008	0.2989	3.3552	
	DEV=267.3540	Total	5.1297	35.6111	257.06
SRW2	constant	0.2863	0.0052	0.0210	
k = 0.0744	w_1	1.5398	3.7417	21.2421	
	w_2	1.2243	3.7277	21.0957	
	w_3	-1.0144	0.5629	2.8960	
	DEV=267.7978	Total	8.0376	45.2547	202.28
GM	constant	0.2820	0.0095	0.0208	
k = 0.2331	w_1	1.1655	4.9622	38.2977	
	w_2	1.0744	4.8936	37.3076	
	w_3	-0.7368	0.9346	2.5295	
	DEV=268.6905	Total	10.7999	78.1556	117.13
WA	constant	0.2824	0.0091	0.0209	
k = 0.1794	w_1	1.1943	4.6176	31.1223	
	w_2	1.0517	4.5824	30.6715	
	w_3	-0.7473	0.9232	2.8851	
	DEV=268.5816	Total	10.1324	64.6998	141.48

Note: $k_{ub} = 8.3725$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.7 The Results of the ML and LRR Estimator Performances for $\rho = 0.95$ and $n = 500$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2329	0.0000	0.0082	
	w_1	2.0678	0.0000	42.1896	
	w_2	0.9015	0.0000	42.1846	
	w_3	-1.4649	0.0000	4.1399	
	DEV=680.8537	Total	0.0000	88.5223	100.00
KOPT	constant	0.2318	0.0012	0.0082	
k = 0.0270	w_1	1.7286	2.4820	15.7689	
	w_2	0.8737	2.4748	15.7370	
	w_3	-1.2172	0.3718	3.1332	
	DEV=681.1758	Total	5.3298	34.6472	255.50
HK	constant	0.2324	0.0005	0.0082	
k = 0.0123	w_1	1.9086	1.7021	17.5102	
	w_2	0.9225	1.6984	17.4903	
	w_3	-1.3532	0.1671	3.5372	
	DEV=680.9583	Total	3.5681	38.5459	229.65
HKB	constant	0.2319	0.0011	0.0082	
k = 0.0365	w_1	1.6958	2.9731	16.7647	
	w_2	0.9644	2.9659	16.7280	
	w_3	-1.2291	0.3501	3.0879	
	DEV=681.2088	Total	6.2902	36.5888	241.94
SRW1	constant	0.2321	0.0008	0.0082	
k = 0.0252	w_1	1.7890	2.4091	15.8389	
	w_2	0.9325	2.4026	15.8058	
	w_3	-1.2740	0.2854	3.2649	
	DEV=681.0857	Total	5.0979	34.9178	253.52
SRW2	constant	0.2312	0.0017	0.0082	
k = 0.0727	w_1	1.4910	3.7410	20.7899	
	w_2	0.9585	3.7293	20.7097	
	w_3	-1.0893	0.5548	2.8289	
	DEV=681.5247	Total	8.0269	44.3367	199.66
GM	constant	0.2295	0.0035	0.0081	
k = 0.2819	w_1	0.9980	5.0314	38.3214	
	w_2	0.9110	4.9991	37.7953	
	w_3	-0.7434	1.0402	2.5887	
	DEV=682.5349	Total	11.0742	78.7135	112.46
WA	constant	0.2299	0.0030	0.0082	
k = 0.1814	w_1	1.1311	4.6280	30.4658	
	w_2	0.8762	4.6014	30.2426	
	w_3	-0.8234	0.9307	2.8653	
	DEV=682.2807	Total	10.1631	63.5820	139.23

Note: $k_{ub} = 11.7225$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.8 The Results of the ML and LRR Estimator Performances for $\rho = 0.95$ and $n = 1000$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2721	0.0000	0.0041	
	w_1	1.9599	0.0000	42.0866	
	w_2	1.0420	0.0000	42.0860	
	w_3	-1.3999	0.0000	4.1223	
	DEV=1362.3482	Total	0.0000	88.2990	100.00
KOPT	constant	0.2714	0.0007	0.0041	
k = 0.0240	w_1	1.5570	2.4818	15.9734	
	w_2	1.0828	2.4569	15.8706	
	w_3	-1.1586	0.3445	3.1382	
	DEV=1362.6743	Total	5.2838	34.9861	252.38
HK	constant	0.2718	0.0003	0.0041	
k = 0.0110	w_1	1.7390	1.6819	17.8410	
	w_2	1.1326	1.6787	17.8176	
	w_3	-1.2987	0.1463	3.5483	
	DEV=1362.4494	Total	3.5072	39.2110	225.19
HKB	constant	0.2715	0.0006	0.0041	
k = 0.0326	w_1	1.5413	2.9869	17.0178	
	w_2	1.1638	2.9692	16.9245	
	w_3	-1.1838	0.3148	3.0937	
	DEV=1362.6996	Total	6.2715	37.0402	238.39
SRW1	constant	0.2716	0.0005	0.0041	
k = 0.0225	w_1	1.6303	2.4010	16.0598	
	w_2	1.1363	2.3909	16.0111	
	w_3	-1.2227	0.2529	3.2776	
	DEV=1362.5748	Total	5.0453	35.3527	249.77
SRW2	constant	0.2711	0.0010	0.0041	
k = 0.0660	w_1	1.3795	3.7648	21.0940	
	w_2	1.1165	3.7330	20.9254	
	w_3	-1.0491	0.5087	2.8226	
	DEV=1363.0144	Total	8.0075	44.8461	196.89
GM	constant	0.2700	0.0021	0.0041	
k = 0.2388	w_1	0.9897	5.0666	38.7957	
	w_2	0.9863	5.0159	38.3293	
	w_3	-0.7591	0.9596	2.5467	
	DEV=1364.0212	Total	11.0442	79.6758	110.82
WA	constant	0.2703	0.0018	0.0041	
k = 0.1646	w_1	1.0829	4.6849	31.0396	
	w_2	0.9566	4.6284	30.7351	
	w_3	-0.7846	0.8845	2.8969	
	DEV=1363.8115	Total	10.1996	64.6756	136.53

Note: $k_{ub} = 11.4834$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.9 The Results of the ML and LRR Estimator Performances for $\rho = 0.99$ and $n = 100$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3081	0.0000	0.0439	
	w_1	2.6117	0.0000	230.0279	
	w_2	0.7746	0.0000	229.8651	
	w_3	-1.8810	0.0000	4.5715	
	DEV=130.0404	Total	0.0000	464.5084	100.00
KOPT	constant	0.3027	0.0055	0.0437	
k = 0.0051	w_1	2.2206	5.5303	84.0022	
	w_2	0.9006	5.5849	84.1294	
	w_3	-1.6678	0.2716	3.8782	
	DEV=130.2898	Total	11.3924	172.0535	269.98
HK	constant	0.3060	0.0021	0.0439	
k = 0.0022	w_1	2.4420	3.6742	94.5846	
	w_2	0.8579	3.6886	94.5330	
	w_3	-1.7964	0.1145	4.2234	
	DEV=130.1182	Total	7.4794	193.3847	240.20
HKB	constant	0.3034	0.0048	0.0437	
k = 0.0083	w_1	2.1652	6.7060	88.5993	
	w_2	1.0195	6.7358	88.5903	
	w_3	-1.6894	0.2468	3.9217	
	DEV=130.3243	Total	13.6935	181.1550	256.41
SRW1	constant	0.3043	0.0039	0.0438	
k = 0.0050	w_1	2.2635	5.5205	84.0735	
	w_2	0.9479	5.5487	84.0834	
	w_3	-1.7192	0.2090	4.0124	
	DEV=130.2328	Total	11.2821	172.2131	269.73
SRW2	constant	0.3003	0.0079	0.0437	
k = 0.0190	w_1	1.8992	8.7525	116.7375	
	w_2	1.1088	8.8046	116.7391	
	w_3	-1.5461	0.4137	3.6772	
	DEV=130.6162	Total	17.9787	237.1975	195.83
GM	constant	0.2913	0.0171	0.0433	
k = 0.1406	w_1	1.1996	12.1051	241.3894	
	w_2	1.1574	12.1282	239.7083	
	w_3	-1.0500	0.9738	3.1864	
	DEV=131.7032	Total	25.2243	484.3275	95.91
WA	constant	0.2957	0.0127	0.0435	
k = 0.0433	w_1	1.5397	10.6652	174.1364	
	w_2	1.1610	10.7497	174.0410	
	w_3	-1.3240	0.6647	3.5999	
	DEV=131.1708	Total	22.0923	351.8207	132.03

Note: $k_{ub} = 6.6929$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.10 The Results of the ML and LRR Estimator Performances for $\rho = 0.99$ and $n = 200$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2868	0.0000	0.0211	
	w_1	2.1554	0.0000	218.3029	
	w_2	1.1403	0.0000	218.2959	
	w_3	-1.6564	0.0000	4.3001	
	DEV=267.0747	Total	0.0000	440.9201	100.00
KOPT	constant	0.2843	0.0026	0.0210	
k = 0.0049	w_1	2.0320	5.4607	78.7929	
	w_2	1.0341	5.4857	78.8755	
	w_3	-1.5132	0.2300	3.6678	
	DEV=267.3185	Total	11.1790	161.3573	273.26
HK	constant	0.2858	0.0010	0.0211	
k = 0.0023	w_1	2.1797	3.7206	87.5826	
	w_2	1.0422	3.7309	87.6117	
	w_3	-1.5955	0.0957	3.9777	
	DEV=267.1576	Total	7.5483	179.1931	246.06
HKB	constant	0.2845	0.0023	0.0211	
k = 0.0085	w_1	1.9692	6.7624	84.0956	
	w_2	1.1502	6.7822	84.1608	
	w_3	-1.5194	0.2066	3.6900	
	DEV=267.3735	Total	13.7535	171.9675	256.40
SRW1	constant	0.2850	0.0018	0.0211	
k = 0.0049	w_1	2.0745	5.4367	78.8007	
	w_2	1.0777	5.4563	78.8715	
	w_3	-1.5441	0.1699	3.7985	
	DEV=267.2666	Total	11.0647	161.4917	273.03
SRW2	constant	0.2831	0.0038	0.0210	
k = 0.0188	w_1	1.7797	8.6044	108.8197	
	w_2	1.1939	8.6400	108.9367	
	w_3	-1.4173	0.3430	3.4599	
	DEV=267.6407	Total	17.5912	221.2373	199.30
GM	constant	0.2786	0.0082	0.0209	
k = 0.1516	w_1	1.2061	11.8468	220.5848	
	w_2	1.2004	11.8829	220.0011	
	w_3	-1.0137	0.8223	2.6725	
	DEV=268.6075	Total	24.5603	443.2794	99.47
WA	constant	0.2807	0.0062	0.0210	
k = 0.0417	w_1	1.4911	10.4655	160.5676	
	w_2	1.2110	10.5333	160.7669	
	w_3	-1.2401	0.5649	3.3289	
	DEV=268.1691	Total	21.5699	324.6844	135.80

Note: $k_{ub} = 8.3298$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.11 The Results of the ML and LRR Estimator Performances for $\rho = 0.99$ and $n = 500$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2338	0.0000	0.0082	
	w_1	2.1460	0.0000	208.7358	
	w_2	1.0567	0.0000	208.7251	
	w_3	-1.6437	0.0000	4.1409	
	DEV=680.7125	Total	0.0000	421.6100	100.00
KOPT	constant	0.2333	0.0005	0.0082	
k = 0.0053	w_1	1.7353	6.2579	66.4014	
	w_2	1.3925	6.2453	66.3074	
	w_3	-1.5740	0.1032	3.8826	
	DEV=680.9225	Total	12.6069	136.5995	308.65
HK	constant	0.2335	0.0003	0.0082	
k = 0.0026	w_1	1.8500	4.7000	74.0362	
	w_2	1.3202	4.6948	73.9960	
	w_3	-1.6121	0.0520	4.0059	
	DEV=680.8230	Total	9.4471	152.0461	277.29
HKB	constant	0.2332	0.0006	0.0082	
k = 0.0099	w_1	1.6662	7.4605	71.4128	
	w_2	1.4401	7.4463	71.2771	
	w_3	-1.5522	0.1303	3.7601	
	DEV=681.0108	Total	15.0377	146.4582	287.87
SRW1	constant	0.2333	0.0004	0.0082	
k = 0.0053	w_1	1.7419	6.2420	66.4095	
	w_2	1.3966	6.2314	66.3216	
	w_3	-1.5815	0.0931	3.8972	
	DEV=680.9144	Total	12.5670	136.6365	308.56
SRW2	constant	0.2330	0.0008	0.0082	
k = 0.0198	w_1	1.5697	8.4894	83.1276	
	w_2	1.4477	8.4618	82.8461	
	w_3	-1.4718	0.2311	3.5109	
	DEV=681.1348	Total	17.1832	169.4928	248.75
GM	constant	0.2315	0.0023	0.0082	
k = 0.1632	w_1	1.2062	9.7855	108.8929	
	w_2	1.1735	9.6988	107.7686	
	w_3	-1.0472	0.8081	2.3952	
	DEV=681.7209	Total	20.2947	219.0648	192.46
WA	constant	0.2325	0.0013	0.0082	
k = 0.0446	w_1	1.4356	9.2691	96.9684	
	w_2	1.3784	9.2078	96.3313	
	w_3	-1.2979	0.4475	3.2152	
	DEV=681.3534	Total	18.9257	196.5232	214.53

Note: $k_{ub} = 11.5859$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.1.12 The Results of the ML and LRR Estimator Performances for $\rho = 0.99$ and $n = 1000$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2717	0.0000	0.0041	
	w_1	2.0809	0.0000	207.5837	
	w_2	0.8897	0.0000	207.5749	
	w_3	-1.5778	0.0000	4.1206	
	DEV=1362.6041	Total	0.0000	419.2833	100.00
KOPT	constant	0.2712	0.0005	0.0041	
k = 0.0050	w_1	1.8036	5.3266	73.5617	
	w_2	0.9551	5.3209	73.5608	
	w_3	-1.4348	0.2058	3.4656	
	DEV=1362.8503	Total	10.8538	150.5921	278.42
HK	constant	0.2715	0.0002	0.0041	
k = 0.0025	w_1	1.9258	3.8065	80.2250	
	w_2	0.9728	3.8012	80.2039	
	w_3	-1.5157	0.0912	3.7668	
	DEV=1362.6963	Total	7.6990	164.1998	255.35
HKB	constant	0.2713	0.0005	0.0041	
k = 0.0093	w_1	1.7696	6.7742	79.9804	
	w_2	1.0365	6.7631	79.9248	
	w_3	-1.4426	0.1914	3.4780	
	DEV=1362.9211	Total	13.7292	163.3873	256.62
SRW1	constant	0.2714	0.0003	0.0041	
k = 0.0050	w_1	1.8439	5.3074	73.5756	
	w_2	0.9969	5.2982	73.5455	
	w_3	-1.4707	0.1539	3.6060	
	DEV=1362.7987	Total	10.7598	150.7313	278.17
SRW2	constant	0.2710	0.0007	0.0041	
k = 0.0183	w_1	1.6230	8.3358	101.1638	
	w_2	1.0675	8.3198	101.0625	
	w_3	-1.3562	0.3059	3.2766	
	DEV=1363.1549	Total	16.9621	205.5069	204.02
GM	constant	0.2701	0.0016	0.0041	
k = 0.1635	w_1	1.1345	11.4379	204.3483	
	w_2	1.0816	11.3816	202.8520	
	w_3	-1.0020	0.7872	2.4299	
	DEV=1364.0300	Total	23.6083	409.6343	102.36
WA	constant	0.2706	0.0012	0.0041	
k = 0.0415	w_1	1.3660	10.1202	149.0197	
	w_2	1.0935	10.1003	148.7959	
	w_3	-1.1937	0.5130	3.1266	
	DEV=1363.6378	Total	20.7346	300.9463	139.32

Note: $k_{ub} = 11.5986$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

A.2 The Results of the Simulation Study in case of Five Explanatory Variables

Table A.2.1 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.90,\ \rho_{34} = 0.90$ and $n = 100$.

Method	Variable	^a^Coef.	|Bias|	MSE	RE
ML	constant	0.4027	0.0000	0.0462	
	w_1	2.8366	0.0000	25.1127	
	w_2	0.7594	0.0000	25.0471	
	w_3	-1.8464	0.0000	25.4739	
	w_4	3.1873	0.0000	25.5151	
	w_5	-1.5365	0.0000	4.8319	
	DEV=126.0612	Total	0.0000	106.0269	100.00
KOPT	constant	0.3886	0.0142	0.0460	
k = 0.0247	w_1	2.2178	1.9133	12.0266	
	w_2	0.9021	1.8625	11.7683	
	w_3	-1.0284	2.0226	12.4461	
	w_4	2.1912	2.0563	12.6080	
	w_5	-1.2702	0.4123	3.6956	
	DEV=126.6439	Total	8.2812	52.5905	201.61
HK	constant	0.3982	0.0045	0.0461	
k = 0.0059	w_1	2.6294	0.8197	15.6367	
	w_2	0.8389	0.8075	15.5527	
	w_3	-1.5258	0.8636	15.8671	
	w_4	2.8195	0.8729	15.9170	
	w_5	-1.4568	0.1276	4.3710	
	DEV=126.1527	Total	3.4959	67.3906	157.33
HKB	constant	0.3903	0.0125	0.0460	
k = 0.0235	w_1	2.2534	1.8968	12.0640	
	w_2	0.9844	1.8576	11.8435	
	w_3	-0.9478	2.0479	12.7429	
	w_4	2.1542	2.0783	12.9097	
	w_5	-1.3168	0.3405	3.7794	
	DEV=126.5593	Total	8.2335	53.3854	198.61
SRW1	constant	0.3934	0.0094	0.0460	
k = 0.0144	w_1	2.4282	1.4521	12.6869	
	w_2	0.8936	1.4240	12.5447	
	w_3	-1.2182	1.5417	13.0013	
	w_4	2.4570	1.5622	13.0947	
	w_5	-1.3689	0.2627	4.0086	
	DEV=126.3557	Total	6.2520	55.3821	191.45
SRW2	constant	0.3811	0.0219	0.0461	
k = 0.0578	w_1	1.8689	2.7638	13.9794	
	w_2	1.0320	2.6826	13.4140	
	w_3	-0.4817	2.9788	15.5549	
	w_4	1.5604	3.0408	15.9891	
	w_5	-1.1249	0.6168	3.3272	
	DEV=127.2638	Total	12.1047	62.3107	170.16

Table A.2.1 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.3680	0.0352	0.0464	
k = 0.1756	w_1	1.2850	3.7058	22.0825	
	w_2	1.0458	3.5337	20.0272	
	w_3	0.1048	3.9957	26.3218	
	w_4	0.7179	4.1633	28.3147	
	w_5	-0.8210	1.0077	3.0907	
	DEV=128.8027	Total	16.4414	99.8833	106.15
WA	constant	0.3675	0.0357	0.0464	
k = 0.1533	w_1	1.3496	3.6081	20.0828	
	w_2	0.9665	3.4519	18.6442	
	w_3	-0.0443	3.8521	23.3037	
	w_4	0.9008	3.9746	24.4224	
	w_5	-0.8218	1.0358	3.2573	
	DEV=128.6212	Total	15.9582	89.7568	118.13

Note: $k_{ub} = 6.8765$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.2 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.90$, $\rho_{34} = 0.90$ and $n = 200$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3341	0.0000	0.0215	
	w_1	1.9298	0.0000	23.0035	
	w_2	1.0494	0.0000	22.9942	
	w_3	-1.1465	0.0000	23.0413	
	w_4	2.2041	0.0000	23.0598	
	w_5	-1.1753	0.0000	4.3984	
	DEV=263.9890	Total	0.0000	96.5187	100.00
KOPT	constant	0.3285	0.0056	0.0214	
k = 0.0322	w_1	1.5726	1.9270	10.6715	
	w_2	1.0549	1.9253	10.6423	
	w_3	-0.5932	1.8667	10.4741	
	w_4	1.5289	1.8906	10.5482	
	w_5	-0.9476	0.3700	3.3220	
	DEV=264.5599	Total	7.9852	45.6795	211.30
HK	constant	0.3322	0.0019	0.0215	
k = 0.0085	w_1	1.8110	0.8992	13.6383	
	w_2	1.0712	0.8948	13.6311	
	w_3	-0.9258	0.8546	13.5412	
	w_4	1.9511	0.8620	13.5692	
	w_5	-1.1083	0.1169	3.9514	
	DEV=264.0962	Total	3.6295	58.3527	165.41
HKB	constant	0.3290	0.0051	0.0214	
k = 0.0332	w_1	1.5990	1.9789	10.8090	
	w_2	1.1000	1.9686	10.7503	
	w_3	-0.5108	1.9249	10.6494	
	w_4	1.4733	1.9434	10.7107	
	w_5	-0.9866	0.3160	3.3501	
	DEV=264.5185	Total	8.1369	46.2909	208.50
SRW1	constant	0.3304	0.0036	0.0214	
k = 0.0188	w_1	1.7023	1.4763	11.2836	
	w_2	1.0764	1.4724	11.2714	
	w_3	-0.7462	1.4153	11.0819	
	w_4	1.7359	1.4299	11.1272	
	w_5	-1.0417	0.2269	3.6218	
	DEV=264.2860	Total	6.0245	48.4074	199.39
SRW2	constant	0.3256	0.0085	0.0214	
k = 0.0732	w_1	1.3932	2.7030	12.6282	
	w_2	1.0567	2.6888	12.4916	
	w_3	-0.2390	2.6367	12.4348	
	w_4	1.1143	2.6719	12.5821	
	w_5	-0.8435	0.5466	2.9208	
	DEV=265.1143	Total	11.2555	53.0789	181.84

Table A.2.2 (Continued)

Method	**Variable**	**[a]Coef.**	**|Bias|**	**MSE**	**RE**
GM	constant	0.3197	0.0144	0.0214	
k = 0.2147	w_1	1.0429	3.5541	19.7841	
	w_2	0.9225	3.5297	19.5741	
	w_3	0.1699	3.4835	19.8101	
	w_4	0.5298	3.5567	20.5222	
	w_5	-0.5846	0.9403	2.7682	
	DEV=266.5182	Total	15.0787	82.4802	117.02
WA	constant	0.3201	0.0140	0.0214	
k = 0.1904	w_1	1.0634	3.4449	18.0254	
	w_2	0.9080	3.4171	17.6877	
	w_3	0.0433	3.3558	17.7315	
	w_4	0.6625	3.4235	18.1956	
	w_5	-0.6036	0.9222	2.7717	
	DEV=266.3062	Total	14.5774	74.4333	129.67

Note: $k_{ub} = 9.0394$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.3 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.90$, $\rho_{34} = 0.90$ and $n = 500$.

Method	Variable	^a^Coef.	|Bias|	MSE	RE
ML	constant	0.2651	0.0000	0.0083	
	w_1	1.6139	0.0000	21.8392	
	w_2	0.8327	0.0000	21.8409	
	w_3	-1.2352	0.0000	21.9014	
	w_4	2.0367	0.0000	21.9134	
	w_5	-1.0238	0.0000	4.1698	
	DEV=677.5409	Total	0.0000	91.6729	100.00
KOPT	constant	0.2633	0.0018	0.0083	
k = 0.0333	w_1	1.3209	1.7834	9.7136	
	w_2	0.8473	1.7683	9.6587	
	w_3	-0.5934	1.9526	10.3072	
	w_4	1.3158	1.9524	10.3513	
	w_5	-0.8573	0.3268	3.0999	
	DEV=678.1039	Total	7.7853	43.1390	212.51
HK	constant	0.2644	0.0006	0.0083	
k = 0.0088	w_1	1.5126	0.8686	12.5264	
	w_2	0.8544	0.8617	12.5133	
	w_3	-0.9397	0.9423	12.6615	
	w_4	1.7197	0.9426	12.6720	
	w_5	-0.9672	0.1085	3.7134	
	DEV=677.6580	Total	3.7244	54.0949	169.47
HKB	constant	0.2634	0.0017	0.0083	
k = 0.0372	w_1	1.3253	1.8929	9.7546	
	w_2	0.8941	1.8811	9.7130	
	w_3	-0.5208	2.0922	10.5879	
	w_4	1.2567	2.0930	10.6193	
	w_5	-0.8761	0.2909	3.0973	
	DEV=678.1085	Total	8.2517	43.7803	209.39
SRW1	constant	0.2640	0.0011	0.0083	
k = 0.0183	w_1	1.4361	1.3519	10.4389	
	w_2	0.8575	1.3416	10.4097	
	w_3	-0.7579	1.4718	10.7517	
	w_4	1.5169	1.4731	10.7724	
	w_5	-0.9191	0.1991	3.4260	
	DEV=677.8306	Total	5.8387	45.8069	200.13
SRW2	constant	0.2624	0.0027	0.0083	
k = 0.0781	w_1	1.1671	2.4981	11.1230	
	w_2	0.8695	2.4778	11.0432	
	w_3	-0.2781	2.7558	12.5934	
	w_4	0.9576	2.7608	12.6921	
	w_5	-0.7710	0.4867	2.7052	
	DEV=678.6442	Total	10.9818	50.1652	182.74

Table A.2.3 (Continued)

Method	**Variable**	**[a]Coef.**	**|Bias|**	**MSE**	**RE**
GM	constant	0.2602	0.0048	0.0082	
k = 0.2588	w_1	0.8291	3.3694	17.9653	
	w_2	0.7738	3.3250	17.6306	
	w_3	0.1063	3.7074	20.7325	
	w_4	0.4214	3.7232	21.2305	
	w_5	-0.5500	0.9061	2.3398	
	DEV=680.0756	Total	15.0359	79.9069	114.72
WA	constant	0.2607	0.0044	0.0082	
k = 0.2013	w_1	0.8896	3.1907	15.7318	
	w_2	0.7612	3.1601	15.5594	
	w_3	-0.0104	3.4954	17.9968	
	w_4	0.5638	3.5232	18.3726	
	w_5	-0.5671	0.8464	2.4982	
	DEV=679.7693	Total	14.2201	70.1671	130.65

Note: $k_{ub} = 12.6166$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.4 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.90$, $\rho_{34} = 0.90$ and $n = 1,000$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3126	0.0000	0.0041	
	w_1	1.8427	0.0000	21.7771	
	w_2	0.9368	0.0000	21.7734	
	w_3	-1.5380	0.0000	21.8117	
	w_4	2.6285	0.0000	21.8157	
	w_5	-1.0408	0.0000	4.1493	
	DEV=1355.0308	Total	0.0000	91.3313	100.00
KOPT	constant	0.3115	0.0011	0.0041	
k = 0.0317	w_1	1.5166	1.7593	9.5514	
	w_2	0.9202	1.7622	9.5151	
	w_3	-0.8244	1.8415	9.8517	
	w_4	1.7673	1.8691	9.9511	
	w_5	-0.8470	0.3407	3.0623	
	DEV=1355.6090	Total	7.5738	41.9358	217.79
HK	constant	0.3122	0.0004	0.0041	
k = 0.0092	w_1	1.7350	0.8607	12.2472	
	w_2	0.9494	0.8618	12.2399	
	w_3	-1.2263	0.8887	12.3058	
	w_4	2.2757	0.8985	12.3249	
	w_5	-0.9797	0.1111	3.6807	
	DEV=1355.1459	Total	3.6211	52.8026	172.97
HKB	constant	0.3115	0.0010	0.0041	
k = 0.0359	w_1	1.5424	1.8435	9.6026	
	w_2	0.9688	1.8461	9.5721	
	w_3	-0.7246	1.9489	10.0592	
	w_4	1.6974	1.9755	10.1667	
	w_5	-0.8770	0.2942	3.0814	
	DEV=1355.5846	Total	7.9092	42.4862	214.97
SRW1	constant	0.3119	0.0007	0.0041	
k = 0.0194	w_1	1.6427	1.3420	10.1769	
	w_2	0.9502	1.3444	10.1606	
	w_3	-1.0038	1.4020	10.3571	
	w_4	2.0154	1.4204	10.4078	
	w_5	-0.9262	0.2064	3.3779	
	DEV=1355.3249	Total	5.7159	44.4844	205.31
SRW2	constant	0.3109	0.0016	0.0041	
k = 0.0753	w_1	1.3579	2.4412	10.8814	
	w_2	0.9349	2.4400	10.7885	
	w_3	-0.4019	2.6009	11.9862	
	w_4	1.2837	2.6502	12.2610	
	w_5	-0.7641	0.4949	2.6890	
	DEV=1356.1418	Total	10.6289	48.6102	187.89

Table A.2.4 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.3098	0.0028	0.0041	
k = 0.2350	w_1	1.0165	3.2365	17.0472	
	w_2	0.8522	3.2131	16.6309	
	w_3	0.0534	3.4530	19.2832	
	w_4	0.6442	3.5649	20.3364	
	w_5	-0.5643	0.8394	2.2272	
	DEV=1357.4274	Total	14.3098	75.5289	120.92
WA	constant	0.3098	0.0027	0.0041	
k = 0.1953	w_1	1.0265	3.1321	15.4266	
	w_2	0.8193	3.1131	15.0906	
	w_3	-0.0415	3.3364	17.4634	
	w_4	0.7431	3.4330	18.2913	
	w_5	-0.5593	0.8522	2.4668	
	DEV=1357.3286	Total	13.8696	68.7428	132.86

Note: $k_{ub} = 12.2580$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.5 The Results of the ML and LRR Estimator Performances for $\rho_{12}=0.99$, $\rho_{34}=0.90$ and $n=100$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3956	0.0000	0.0463	
	w_1	2.4569	0.0000	250.3804	
	w_2	1.3547	0.0000	250.2962	
	w_3	-2.0037	0.0000	25.4287	
	w_4	3.1694	0.0000	25.4930	
	w_5	-1.7626	0.0000	4.8530	
	DEV=125.8812	Total	0.0000	556.4976	100.00
KOPT	constant	0.3867	0.0091	0.0461	
k = 0.0052	w_1	2.2450	6.1029	93.1235	
	w_2	1.3052	6.1000	93.0933	
	w_3	-1.4877	1.1854	17.1834	
	w_4	2.5703	1.1892	17.2675	
	w_5	-1.5720	0.2720	4.2737	
	DEV=126.2472	Total	14.8585	224.9876	247.35
HK	constant	0.3923	0.0034	0.0462	
k = 0.0016	w_1	2.3734	3.7274	109.9871	
	w_2	1.3725	3.7263	109.9396	
	w_3	-1.7899	0.5179	20.1907	
	w_4	2.9334	0.5189	20.2616	
	w_5	-1.7080	0.0905	4.6232	
	DEV=125.9689	Total	8.8544	265.0484	209.96
HKB	constant	0.3865	0.0092	0.0462	
k = 0.0074	w_1	2.2387	7.7903	104.9629	
	w_2	1.3780	7.7851	104.8807	
	w_3	-1.4088	1.3035	16.2154	
	w_4	2.5120	1.3144	16.3089	
	w_5	-1.6100	0.2371	4.3066	
	DEV=126.3175	Total	18.4396	246.7207	225.56
SRW1	constant	0.3891	0.0066	0.0462	
k = 0.0041	w_1	2.2948	5.8572	93.6427	
	w_2	1.3733	5.8528	93.5624	
	w_3	-1.5850	0.9590	17.8044	
	w_4	2.7035	0.9654	17.8844	
	w_5	-1.6466	0.1786	4.4394	
	DEV=126.1247	Total	13.8196	227.3795	244.74
SRW2	constant	0.3802	0.0156	0.0462	
k = 0.0186	w_1	2.0094	9.9451	144.7054	
	w_2	1.4130	9.9303	144.5000	
	w_3	-1.0079	2.0754	15.6295	
	w_4	2.0494	2.0957	15.8335	
	w_5	-1.4708	0.4210	4.0250	
	DEV=126.7958	Total	24.4830	324.7396	171.37

Table A.2.5 (Continued)

Method	**Variable**	**[a]Coef.**	**|Bias|**	**MSE**	**RE**
GM	constant	0.3658	0.0301	0.0466	
k = 0.1102	w_1	1.4182	12.6553	263.4700	
	w_2	1.3914	12.5859	261.2868	
	w_3	-0.0676	3.6272	22.1830	
	w_4	0.8935	3.7509	23.4573	
	w_5	-1.0844	0.8737	3.2672	
	DEV=128.2829	Total	33.5232	573.7109	97.00
WA	constant	0.3714	0.0244	0.0465	
k = 0.0443	w_1	1.6774	11.6432	207.3713	
	w_2	1.4124	11.6225	206.9010	
	w_3	-0.5195	2.9551	18.4481	
	w_4	1.4538	2.9958	18.9664	
	w_5	-1.2505	0.6958	3.8880	
	DEV=127.6332	Total	29.9369	455.6211	122.14

Note: $k_{ub} = 6.7896$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.6 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.99$, $\rho_{34} = 0.90$ and $n = 200$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3304	0.0000	0.0215	
	w_1	1.7310	0.0000	224.5251	
	w_2	1.0618	0.0000	224.5003	
	w_3	-1.4547	0.0000	23.1174	
	w_4	2.5626	0.0000	23.1423	
	w_5	-1.2865	0.0000	4.3952	
	DEV=264.1144	Total	0.0000	499.7017	100.00
KOPT	constant	0.3270	0.0034	0.0214	
k = 0.0047	w_1	1.5962	5.8673	86.1817	
	w_2	1.0595	5.8785	86.2325	
	w_3	-1.0485	0.9942	15.8317	
	w_4	2.0865	1.0203	15.9228	
	w_5	-1.1810	0.1958	3.8604	
	DEV=264.4507	Total	13.9595	208.0505	240.18
HK	constant	0.3291	0.0014	0.0215	
k = 0.0017	w_1	1.7035	3.7260	98.6046	
	w_2	1.0537	3.7277	98.6062	
	w_3	-1.2932	0.4557	18.4551	
	w_4	2.3850	0.4597	18.4807	
	w_5	-1.2549	0.0667	4.2063	
	DEV=264.2082	Total	8.4371	238.3743	209.63
HKB	constant	0.3267	0.0037	0.0214	
k = 0.0082	w_1	1.4799	7.8596	100.0370	
	w_2	1.1994	7.8701	100.0810	
	w_3	-0.9502	1.1751	14.5276	
	w_4	2.0053	1.1912	14.6038	
	w_5	-1.1948	0.1756	3.8848	
	DEV=264.5865	Total	18.2754	233.1556	214.32
SRW1	constant	0.3280	0.0025	0.0214	
k = 0.0039	w_1	1.6104	5.6374	86.6257	
	w_2	1.1084	5.6407	86.6353	
	w_3	-1.1443	0.7995	16.4633	
	w_4	2.2207	0.8079	16.5037	
	w_5	-1.2249	0.1209	4.0576	
	DEV=264.3480	Total	13.0088	210.3072	237.61
SRW2	constant	0.3246	0.0059	0.0214	
k = 0.0186	w_1	1.3508	9.7971	135.5886	
	w_2	1.2216	9.8147	135.6581	
	w_3	-0.6648	1.7769	13.5632	
	w_4	1.6748	1.8098	13.7446	
	w_5	-1.1173	0.2952	3.5921	
	DEV=264.9969	Total	23.4995	302.1681	165.37

Table A.2.6 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.3178	0.0126	0.0214	
k = 0.1390	w_1	1.0455	12.4808	237.2729	
	w_2	1.0360	12.5314	237.8926	
	w_3	0.0675	3.3988	18.9067	
	w_4	0.7236	3.4916	19.8215	
	w_5	-0.8157	0.7691	2.6322	
	DEV=266.4642	Total	32.6844	516.5474	96.74
WA	constant	0.3214	0.0090	0.0214	
k = 0.0458	w_1	1.2114	11.4731	190.7445	
	w_2	1.1552	11.4971	190.8717	
	w_3	-0.2885	2.5671	14.8952	
	w_4	1.2100	2.6229	15.3857	
	w_5	-0.9776	0.5038	3.2731	
	DEV=265.7009	Total	28.6730	415.1917	120.35

Note: $k_{ub} = 9.0181$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.7 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.99,\ \rho_{34} = 0.90$ and $n = 500$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.2615	0.0000	0.0083	
	w_1	1.6732	0.0000	211.3365	
	w_2	0.8831	0.0000	211.3355	
	w_3	-1.2852	0.0000	21.8253	
	w_4	2.2428	0.0000	21.8337	
	w_5	-0.9525	0.0000	4.1692	
	DEV=677.8282	Total	0.0000	470.5084	100.00
KOPT	constant	0.2604	0.0011	0.0083	
k = 0.0059	w_1	1.7159	5.5622	75.6525	
	w_2	0.6962	5.5745	75.7392	
	w_3	-0.9296	0.9663	14.3284	
	w_4	1.8288	0.9846	14.3828	
	w_5	-0.8636	0.1936	3.6276	
	DEV=678.1681	Total	13.2823	183.7387	256.07
HK	constant	0.2610	0.0004	0.0083	
k = 0.0023	w_1	1.7660	3.7274	84.6491	
	w_2	0.7566	3.7290	84.6418	
	w_3	-1.1258	0.4748	16.8616	
	w_4	2.0705	0.4774	16.8696	
	w_5	-0.9295	0.0598	3.9620	
	DEV=677.9307	Total	8.4688	206.9924	227.31
HKB	constant	0.2603	0.0012	0.0083	
k = 0.0110	w_1	1.6279	7.5134	89.4944	
	w_2	0.8148	7.5205	89.5277	
	w_3	-0.8186	1.1903	12.9997	
	w_4	1.7343	1.2017	13.0362	
	w_5	-0.8755	0.1725	3.6107	
	DEV=678.3155	Total	17.5998	208.6769	225.47
SRW1	constant	0.2607	0.0007	0.0083	
k = 0.0049	w_1	1.7410	5.3347	75.8115	
	w_2	0.7500	5.3379	75.8113	
	w_3	-0.9980	0.7782	15.0686	
	w_4	1.9298	0.7845	15.0844	
	w_5	-0.9096	0.1066	3.8196	
	DEV=678.0576	Total	12.3425	185.6036	253.50
SRW2	constant	0.2597	0.0018	0.0083	
k = 0.0231	w_1	1.4706	9.0672	117.6608	
	w_2	0.8780	9.0787	117.7376	
	w_3	-0.5861	1.7040	12.1191	
	w_4	1.4650	1.7240	12.2137	
	w_5	-0.8194	0.2861	3.3469	
	DEV=678.6742	Total	21.8618	263.0863	178.84

Table A.2.7 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.2574	0.0041	0.0082	
k = 0.1769	w_1	0.9503	11.5095	205.6181	
	w_2	0.9148	11.5362	205.2973	
	w_3	0.0643	3.1357	15.9533	
	w_4	0.6455	3.2102	16.5493	
	w_5	-0.5588	0.7761	2.4006	
	DEV=680.0883	Total	30.1719	445.8268	105.54
				RE	
WA	constant	0.2587	0.0028	0.0082	
k = 0.0545	w_1	1.2030	10.5714	164.3453	
	w_2	0.9487	10.5875	164.4823	
	w_3	-0.2600	2.3956	12.8822	
	w_4	1.0623	2.4365	13.1711	
	w_5	-0.7092	0.4961	3.0808	
	DEV=679.3362	Total	26.4899	357.9700	131.44

Note: $k_{ub} = 12.7354$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.8 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.99,\ \rho_{34} = 0.90$ and $n = 1{,}000$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3118	0.0000	0.0041	
	w_1	1.8676	0.0000	209.8126	
	w_2	0.9309	0.0000	209.8143	
	w_3	-1.2086	0.0000	21.7579	
	w_4	2.2570	0.0000	21.7643	
	w_5	-1.0885	0.0000	4.1482	
	DEV=1355.0240	Total	0.0000	467.3013	100.00
KOPT	constant	0.3112	0.0007	0.0041	
k = 0.0053	w_1	1.6949	5.6674	78.3538	
	w_2	0.9384	5.6773	78.3839	
	w_3	-0.8896	0.9644	14.6942	
	w_4	1.8686	0.9780	14.7358	
	w_5	-1.0076	0.1481	3.5921	
	DEV=1355.3653	Total	13.4359	189.7640	246.25
HK	constant	0.3116	0.0003	0.0041	
k = 0.0021	w_1	1.7999	3.7464	88.2333	
	w_2	0.9619	3.7499	88.2376	
	w_3	-1.0705	0.4544	17.1208	
	w_4	2.1033	0.4559	17.1257	
	w_5	-1.0686	0.0432	3.9574	
	DEV=1355.1230	Total	8.4500	214.6789	217.67
HKB	constant	0.3111	0.0007	0.0041	
k = 0.0100	w_1	1.6212	7.6834	92.5318	
	w_2	1.0602	7.6897	92.5189	
	w_3	-0.7943	1.1619	13.3758	
	w_4	1.7929	1.1712	13.4017	
	w_5	-1.0207	0.1310	3.6183	
	DEV=1355.5066	Total	17.8380	215.4506	216.89
SRW1	constant	0.3114	0.0005	0.0041	
k = 0.0045	w_1	1.7360	5.4377	78.6712	
	w_2	0.9891	5.4436	78.6752	
	w_3	-0.9602	0.7650	15.3297	
	w_4	1.9778	0.7692	15.3374	
	w_5	-1.0490	0.0804	3.8104	
	DEV=1355.2523	Total	12.4964	191.8281	243.60
SRW2	constant	0.3107	0.0011	0.0041	
k = 0.0212	w_1	1.4846	9.3601	122.7785	
	w_2	1.0938	9.3682	122.7329	
	w_3	-0.5693	1.7065	12.5476	
	w_4	1.5266	1.7243	12.6297	
	w_5	-0.9622	0.2315	3.3172	
	DEV=1355.8843	Total	22.3917	274.0101	170.54

Table A.2.8 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.3094	0.0024	0.0041	
k = 0.1456	w_1	1.0593	12.0344	222.2062	
	w_2	1.0474	12.0121	220.5507	
	w_3	0.0680	3.1732	16.5923	
	w_4	0.7180	3.2232	17.1171	
	w_5	-0.7235	0.6653	2.2621	
	DEV=1357.2997	Total	31.1106	478.7326	97.61
WA	constant	0.3101	0.0018	0.0041	
k = 0.0513	w_1	1.2698	11.0106	175.3745	
	w_2	1.0963	11.0185	175.2076	
	w_3	-0.2439	2.4493	13.7207	
	w_4	1.1193	2.4875	14.0103	
	w_5	-0.8470	0.4234	2.9312	
	DEV=1356.5943	Total	27.3911	381.2484	122.57

Note: $k_{ub} = 12.3705$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.9 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.99$, $\rho_{34} = 0.99$ and $n = 100$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3899	0.0000	0.0460	
	w_1	2.6157	0.0000	248.1356	
	w_2	1.0282	0.0000	248.0858	
	w_3	-2.8003	0.0000	250.7020	
	w_4	4.0050	0.0000	250.6873	
	w_5	-1.4002	0.0000	4.8008	
	DEV=126.4370	Total	0.0000	1002.4576	100.00
KOPT	constant	0.3815	0.0085	0.0459	
k = 0.0027	w_1	2.2133	6.1726	109.2940	
	w_2	1.2915	6.1705	109.2714	
	w_3	-1.3722	6.0464	108.7649	
	w_4	2.5275	6.0673	108.8804	
	w_5	-1.3223	0.1756	4.4520	
	DEV=126.8505	Total	24.6408	440.7086	227.46
HK	constant	0.3867	0.0032	0.0459	
k = 0.0007	w_1	2.4681	3.0075	140.2593	
	w_2	1.1354	3.0066	140.2485	
	w_3	-2.1138	2.8891	140.0022	
	w_4	3.3047	2.8920	140.0284	
	w_5	-1.3802	0.0673	4.6640	
	DEV=126.5304	Total	11.8657	565.2483	177.35
HKB	constant	0.3812	0.0088	0.0459	
k = 0.0033	w_1	2.1772	6.7595	110.3104	
	w_2	1.3429	6.7541	110.2411	
	w_3	-1.1969	6.6936	111.0590	
	w_4	2.3609	6.7027	111.1020	
	w_5	-1.3296	0.1730	4.4610	
	DEV=126.9185	Total	27.0917	447.2193	224.15
SRW1	constant	0.3837	0.0063	0.0459	
k = 0.0017	w_1	2.3509	5.0312	114.1448	
	w_2	1.2078	5.0288	114.1186	
	w_3	-1.6340	4.8994	113.5512	
	w_4	2.8113	4.9066	113.5874	
	w_5	-1.3538	0.1249	4.5591	
	DEV=126.6969	Total	19.9972	460.0071	217.92
SRW2	constant	0.3761	0.0140	0.0460	
k = 0.0078	w_1	1.9837	9.1541	135.5813	
	w_2	1.4233	9.1462	135.3849	
	w_3	-0.5781	9.1884	140.5547	
	w_4	1.7076	9.2090	140.6861	
	w_5	-1.2586	0.2831	4.2462	
	DEV=127.4020	Total	36.9947	556.4992	180.14

Table A.2.9 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.3627	0.0276	0.0464	
k = 0.0842	w_1	1.4236	12.6132	252.5444	
	w_2	1.4128	12.5666	250.6083	
	w_3	0.3683	12.8529	268.9274	
	w_4	0.5313	12.8988	270.1859	
	w_5	-0.9262	0.7184	3.1939	
	DEV=128.8905	Total	51.6775	1045.5064	95.88
WA	constant	0.3700	0.0202	0.0462	
k = 0.0197	w_1	1.7289	11.2054	191.6185	
	w_2	1.4877	11.1939	191.1592	
	w_3	-0.0126	11.3562	203.4055	
	w_4	1.0843	11.3972	203.7772	
	w_5	-1.1420	0.4442	3.9871	
	DEV=128.1025	Total	45.6171	793.9939	126.26

Note: $k_{ub} = 6.9007$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.10 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.99$, $\rho_{34} = 0.99$ and $n = 200$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.4977	0.0000	0.0225	
	w_1	3.6077	0.0000	232.3441	
	w_2	0.4026	0.0000	232.2606	
	w_3	-1.0104	0.0000	233.9651	
	w_4	2.3918	0.0000	233.9892	
	w_5	-1.7509	0.0000	4.5809	
	DEV=256.4165	Total	0.0000	937.1624	100.00
KOPT	constant	0.4922	0.0054	0.0225	
k = 0.0030	w_1	2.7925	5.8656	97.7974	
	w_2	1.0768	5.8580	97.6846	
	w_3	-0.2788	6.0187	99.7477	
	w_4	1.6080	6.0069	99.6445	
	w_5	-1.6748	0.1479	4.2555	
	DEV=256.8360	Total	23.9025	399.1522	234.79
HK	constant	0.4957	0.0020	0.0225	
k = 0.0008	w_1	3.2818	2.6867	130.4903	
	w_2	0.6938	2.6847	130.4293	
	w_3	-0.7022	2.7704	131.6019	
	w_4	2.0700	2.7681	131.5952	
	w_5	-1.7308	0.0490	4.4742	
	DEV=256.5005	Total	10.9608	528.6134	177.29
HKB	constant	0.4922	0.0055	0.0225	
k = 0.0035	w_1	2.7482	6.2829	98.2481	
	w_2	1.1517	6.2759	98.1468	
	w_3	-0.2002	6.4676	100.9968	
	w_4	1.5412	6.4596	100.9163	
	w_5	-1.6830	0.1362	4.2898	
	DEV=256.8754	Total	25.6276	402.6202	232.77
SRW1	constant	0.4938	0.0039	0.0225	
k = 0.0019	w_1	3.0011	4.6856	102.9654	
	w_2	0.9304	4.6836	102.9029	
	w_3	-0.4568	4.8094	104.3895	
	w_4	1.8090	4.8036	104.3455	
	w_5	-1.7036	0.0991	4.3699	
	DEV=256.6704	Total	19.0852	418.9956	223.67
SRW2	constant	0.4888	0.0089	0.0225	
k = 0.0089	w_1	2.3598	8.7131	122.7683	
	w_2	1.4202	8.7011	122.5455	
	w_3	0.1171	8.8815	124.9063	
	w_4	1.1828	8.8666	124.7436	
	w_5	-1.6052	0.2466	4.0577	
	DEV=257.3627	Total	35.4179	499.0439	187.79

Table A.2.10 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.4805	0.0172	0.0227	
k = 0.0655	w_1	1.7398	11.9432	222.8291	
	w_2	1.5771	11.9068	221.7649	
	w_3	0.4711	12.1284	228.3726	
	w_4	0.6442	12.0948	227.9301	
	w_5	-1.3229	0.5937	3.1175	
	DEV=258.6292	Total	48.6840	904.0369	103.66
WA	constant	0.4846	0.0131	0.0226	
k = 0.0214	w_1	2.0060	10.7349	174.8088	
	w_2	1.5674	10.7086	174.2855	
	w_3	0.3778	10.8948	177.5425	
	w_4	0.8527	10.8693	177.3059	
	w_5	-1.4738	0.4128	3.7920	
	DEV=258.0510	Total	43.6335	707.7573	132.41

Note: $k_{ub} = 7.0632$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.11 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.99$, $\rho_{34} = 0.99$ and $n = 500$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.4247	0.0000	0.0085	
	w_1	0.8543	0.0000	217.2867	
	w_2	1.7656	0.0000	217.3002	
	w_3	-2.0822	0.0000	216.3521	
	w_4	3.0030	0.0000	216.3596	
	w_5	-1.2144	0.0000	4.2826	
	DEV=664.7201	Total	0.0000	871.5897	100.00
KOPT	constant	0.4229	0.0018	0.0085	
k = 0.0031	w_1	1.0079	5.6345	88.6548	
	w_2	1.5141	5.6273	88.6920	
	w_3	-0.9494	5.5414	87.4119	
	w_4	1.8356	5.5430	87.4253	
	w_5	-1.1578	0.1311	3.9343	
	DEV=665.1476	Total	22.4791	356.1267	244.74
HK	constant	0.4240	0.0007	0.0085	
k = 0.0009	w_1	0.9432	2.7574	115.2821	
	w_2	1.6543	2.7546	115.2860	
	w_3	-1.5962	2.7376	114.7081	
	w_4	2.5102	2.7393	114.7182	
	w_5	-1.1980	0.0412	4.1699	
	DEV=664.8154	Total	11.0308	464.1728	187.77
HKB	constant	0.4228	0.0018	0.0085	
k = 0.0039	w_1	1.0572	6.2081	90.2963	
	w_2	1.4921	6.2056	90.3397	
	w_3	-0.8078	6.0721	87.9802	
	w_4	1.7054	6.0777	88.0281	
	w_5	-1.1646	0.1119	3.9633	
	DEV=665.2060	Total	24.6772	360.6162	241.69
SRW1	constant	0.4234	0.0013	0.0085	
k = 0.0020	w_1	0.9800	4.5208	92.9779	
	w_2	1.5915	4.5158	92.9889	
	w_3	-1.2080	4.4592	92.2568	
	w_4	2.1139	4.4631	92.2812	
	w_5	-1.1797	0.0791	4.0695	
	DEV=664.9769	Total	18.0393	374.5827	232.68
SRW2	constant	0.4219	0.0028	0.0085	
k = 0.0093	w_1	1.0727	8.2475	112.2964	
	w_2	1.4042	8.2449	112.4102	
	w_3	-0.3334	8.0064	106.2543	
	w_4	1.2075	8.0166	106.3829	
	w_5	-1.1156	0.1962	3.7357	
	DEV=665.6366	Total	32.7143	441.0881	197.60

Table A.2.11 (Continued)

Method	Variable	[a]Coef.	|Bias|	MSE	RE
GM	constant	0.4193	0.0054	0.0085	
k = 0.0971	w_1	1.0729	11.2583	205.0956	
	w_2	1.0827	11.2888	205.5814	
	w_3	0.2933	10.8373	188.4532	
	w_4	0.4499	10.8638	189.2019	
	w_5	-0.8978	0.5475	2.5993	
	DEV=666.7977	Total	44.8012	790.9398	110.20
WA	constant	0.4207	0.0040	0.0085	
k = 0.0227	w_1	1.1016	10.0776	158.9490	
	w_2	1.2475	10.0820	159.2014	
	w_3	0.0632	9.7186	146.7712	
	w_4	0.7693	9.7351	147.0870	
	w_5	-1.0301	0.3300	3.4371	
	DEV=666.2448	Total	39.9473	615.4542	141.62

Note: $k_{ub} = 9.1594$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Table A.2.12 The Results of the ML and LRR Estimator Performances for $\rho_{12} = 0.99,\ \rho_{34} = 0.99$ and $n = 1{,}000$.

Method	Variable	[a]Coef.	|Bias|	MSE	RE
ML	constant	0.3791	0.0000	0.0042	
	w_1	1.5135	0.0000	211.7635	
	w_2	1.0008	0.0000	211.7590	
	w_3	-1.2978	0.0000	212.2370	
	w_4	2.0746	0.0000	212.2491	
	w_5	-1.1677	0.0000	4.1950	
	DEV=1344.3251	Total	0.0000	852.2077	100.00
KOPT	constant	0.3783	0.0008	0.0042	
k = 0.0033	w_1	1.4114	5.3679	87.3130	
	w_2	1.0380	5.3612	87.2925	
	w_3	-0.7419	5.8300	92.5622	
	w_4	1.4952	5.8244	92.5255	
	w_5	-1.1236	0.1019	3.8866	
	DEV=1344.7482	Total	22.4862	363.5840	234.39
HK	constant	0.3788	0.0003	0.0042	
k = 0.0009	w_1	1.4806	2.6411	115.2019	
	w_2	1.0211	2.6393	115.1949	
	w_3	-1.0726	2.8706	116.7888	
	w_4	1.8447	2.8702	116.7906	
	w_5	-1.1544	0.0302	4.1063	
	DEV=1344.4220	Total	11.0517	468.0867	182.06
HKB	constant	0.3782	0.0008	0.0042	
k = 0.0042	w_1	1.4285	6.0012	88.1483	
	w_2	1.0410	5.9981	88.1501	
	w_3	-0.6372	6.5088	94.5757	
	w_4	1.3975	6.5081	94.5617	
	w_5	-1.1248	0.0882	3.9195	
	DEV=1344.8291	Total	25.1052	369.3596	230.73
SRW1	constant	0.3785	0.0006	0.0042	
k = 0.0020	w_1	1.4582	4.2618	93.0725	
	w_2	1.0286	4.2585	93.0669	
	w_3	-0.8807	4.6279	96.4067	
	w_4	1.6474	4.6273	96.4000	
	w_5	-1.1404	0.0577	4.0244	
	DEV=1344.5761	Total	17.8338	382.9746	222.52
SRW2	constant	0.3778	0.0013	0.0042	
k = 0.0095	w_1	1.3716	7.9399	107.2112	
	w_2	1.0481	7.9349	107.2264	
	w_3	-0.3251	8.5576	117.0363	
	w_4	1.0679	8.5560	117.0169	
	w_5	-1.0846	0.1568	3.7057	
	DEV=1345.2559	Total	33.1465	452.2007	188.46

Table A.2.12 (Continued)

Method	**Variable**	**[a]Coef.**	**|Bias|**	**MSE**	**RE**
GM	constant	0.3765	0.0026	0.0042	
k = 0.0952	w_1	1.0545	11.1549	207.6856	
	w_2	1.0286	11.1618	207.4891	
	w_3	0.2496	11.7897	212.4173	
	w_4	0.3893	11.7993	213.1414	
	w_5	-0.8487	0.5573	2.5514	
	DEV=1346.5847	Total	46.4655	843.2889	101.06
WA	constant	0.3773	0.0018	0.0042	
k = 0.0233	w_1	1.2385	9.8771	156.4721	
	w_2	1.0804	9.8738	156.5031	
	w_3	0.0259	10.5065	164.7063	
	w_4	0.6826	10.5052	164.7292	
	w_5	-1.0065	0.2813	3.3890	
	DEV=1345.9021	Total	41.0457	645.8039	131.96

Note: $k_{ub} = 10.5190$; results of ridge parameter reported as medians; [a]estimated standardized regression coefficients

Appendix B

Distribution of Ridge Parameter in Case of Having Three Explanatory Variables

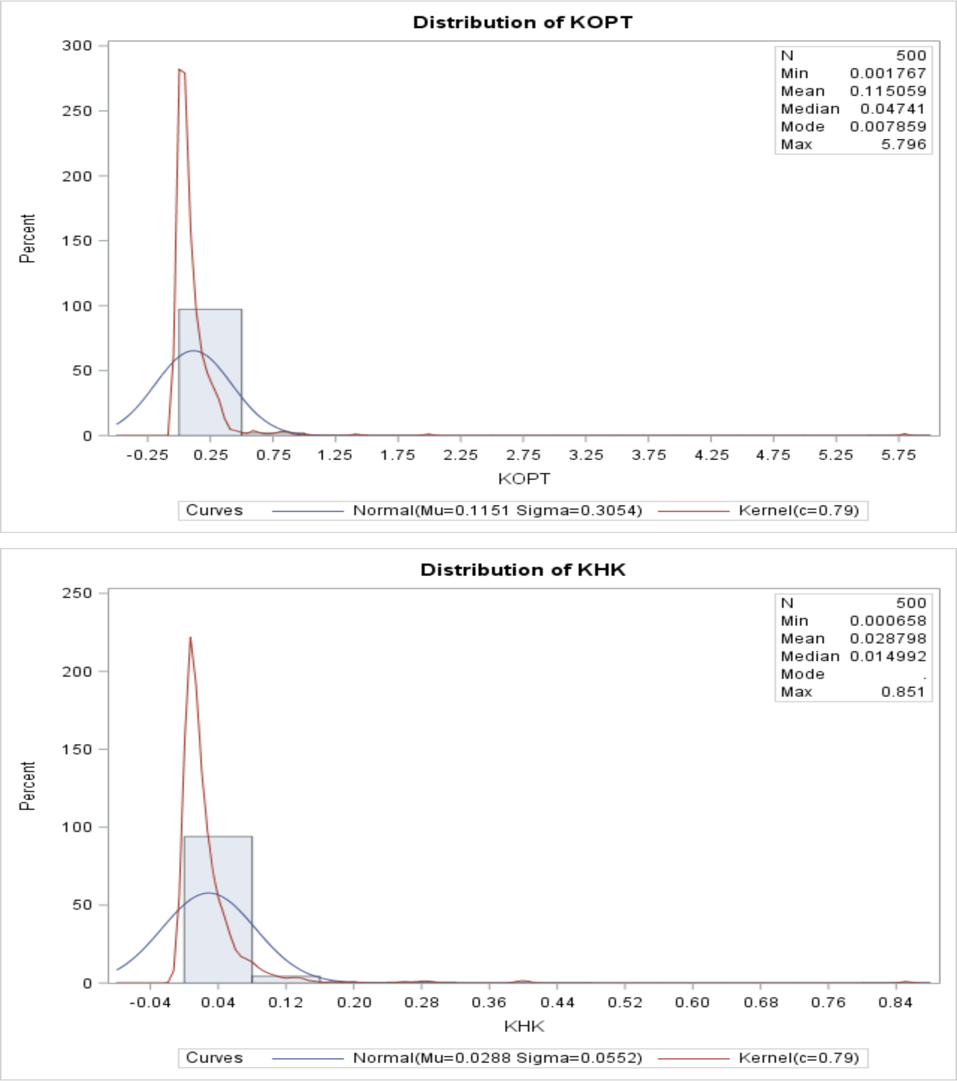
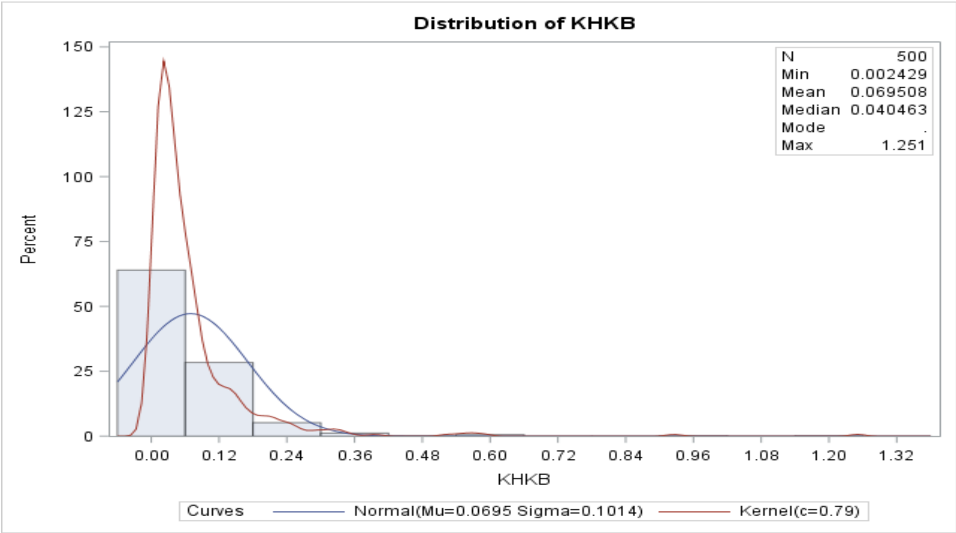
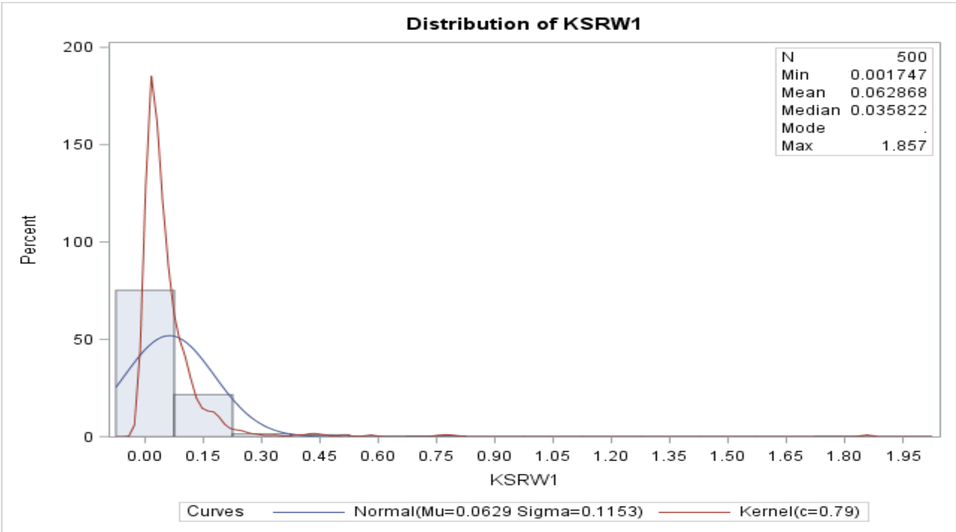


Figure B.1 Distribution of Ridge Parameter for $\rho = 0.90,\ n = 100$.





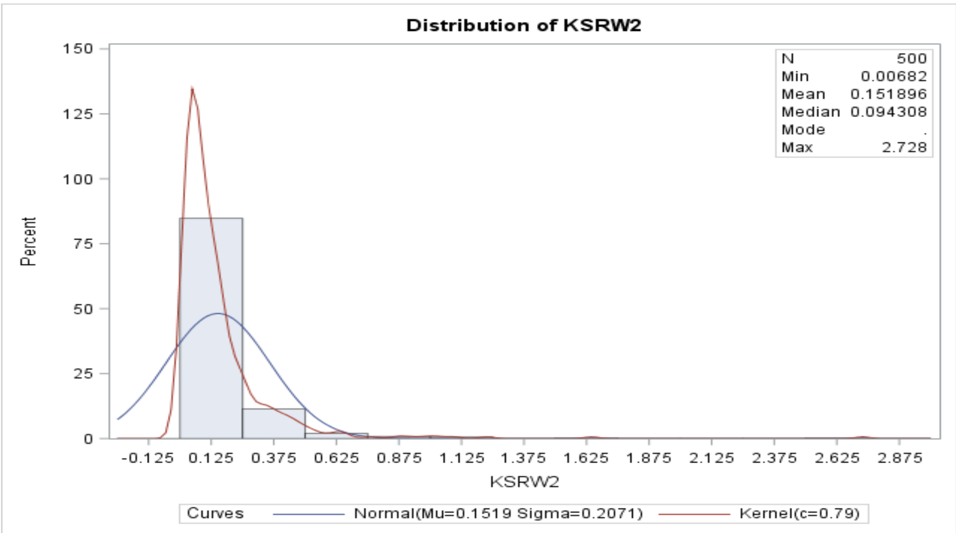


Figure B.1 (Continued)

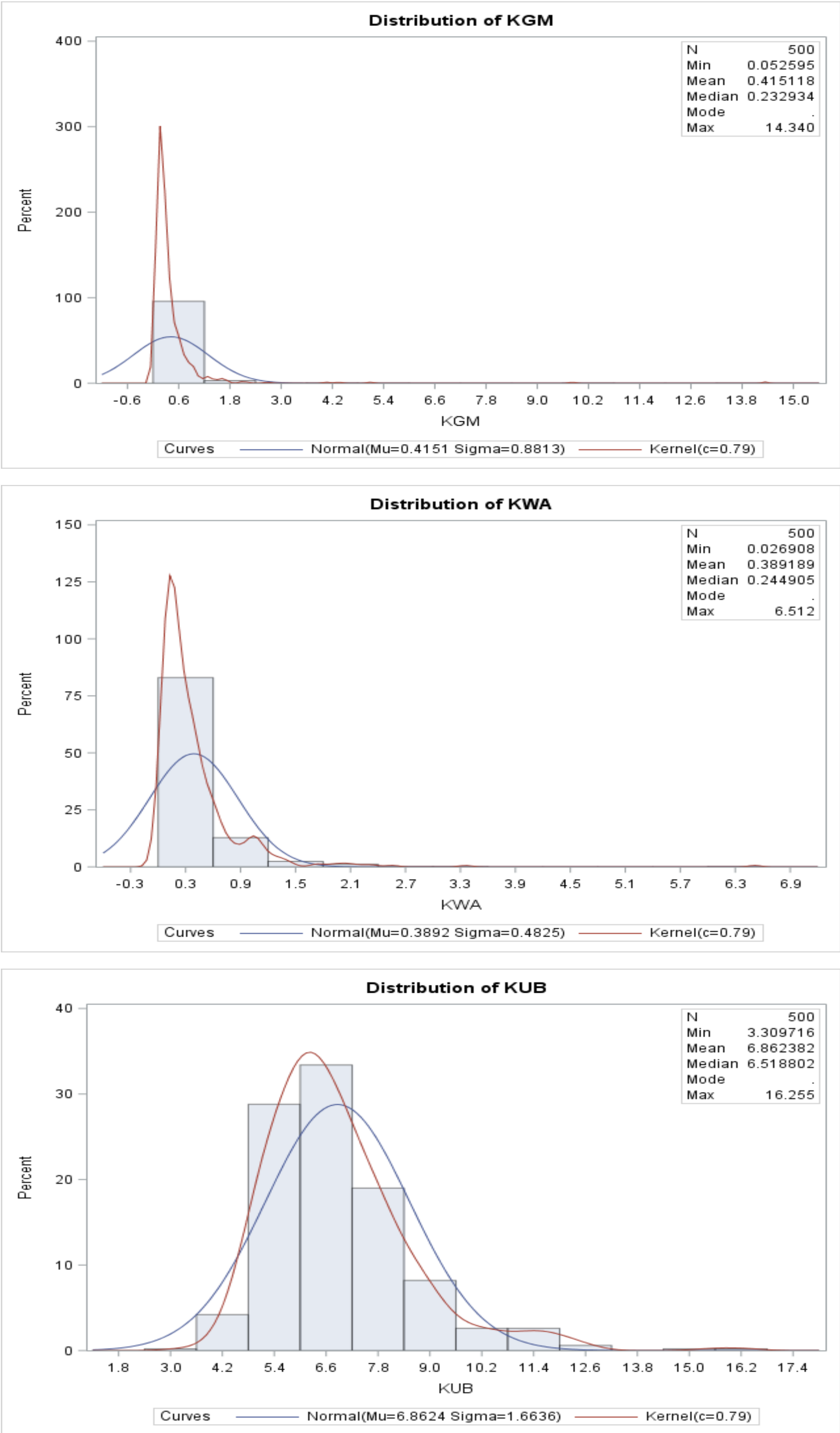


Figure B.1 (Continued)

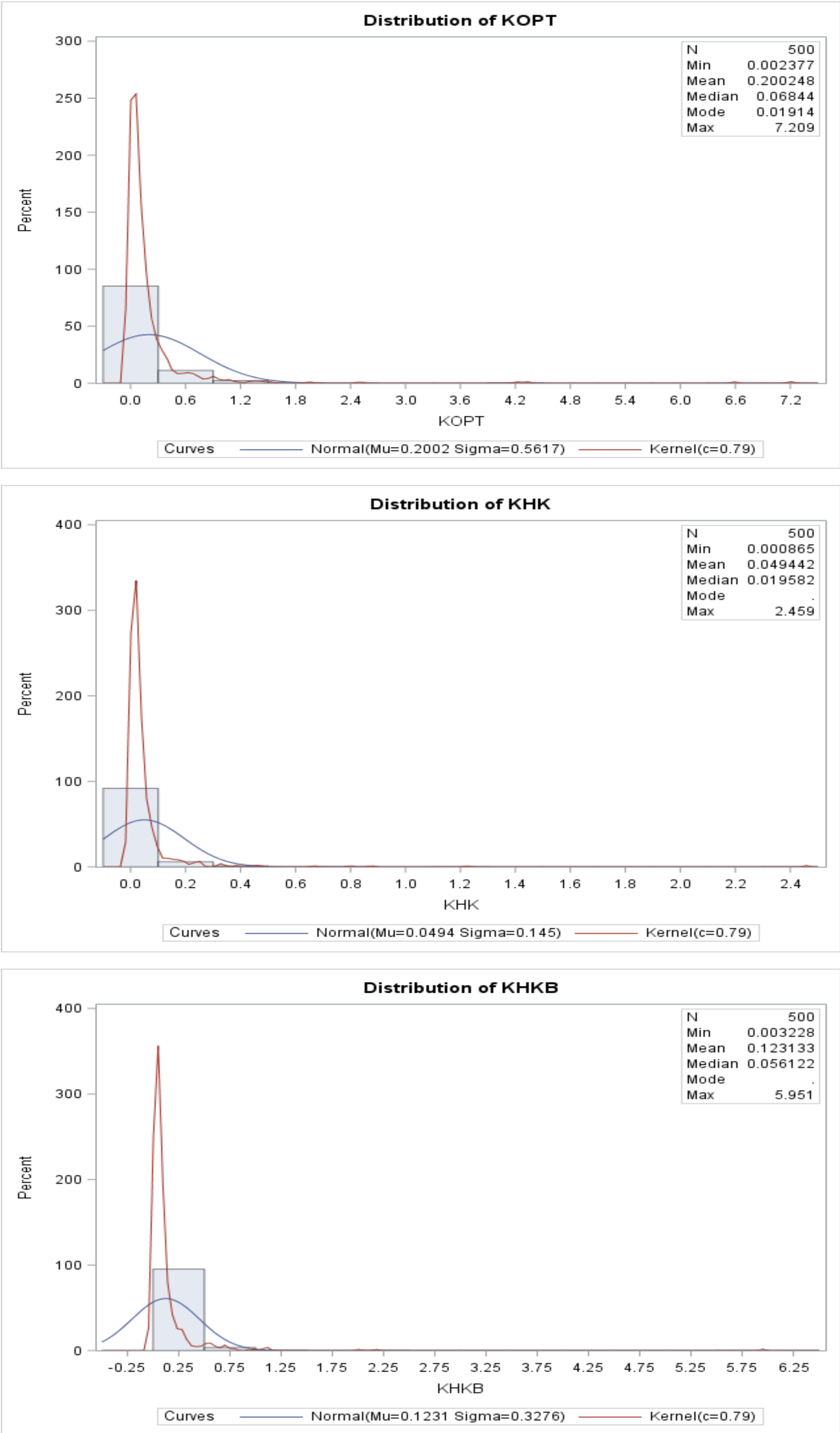
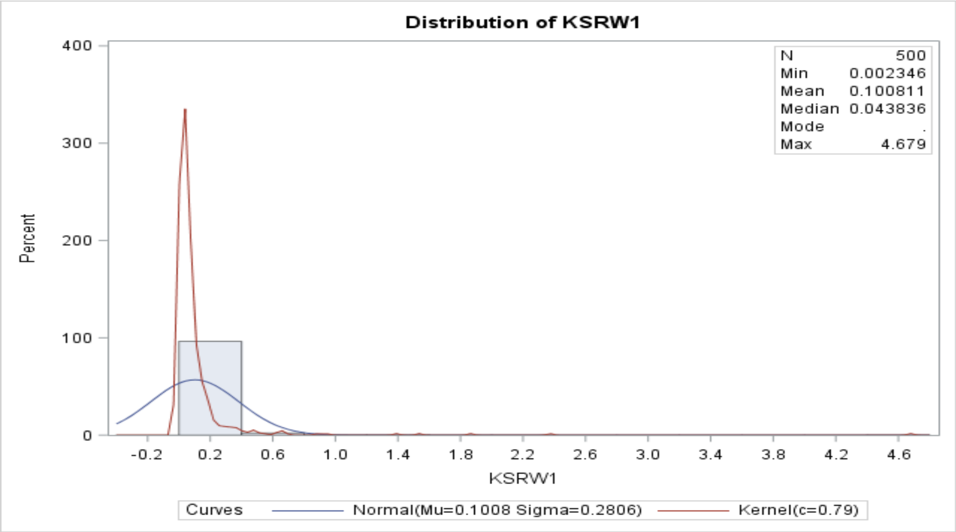
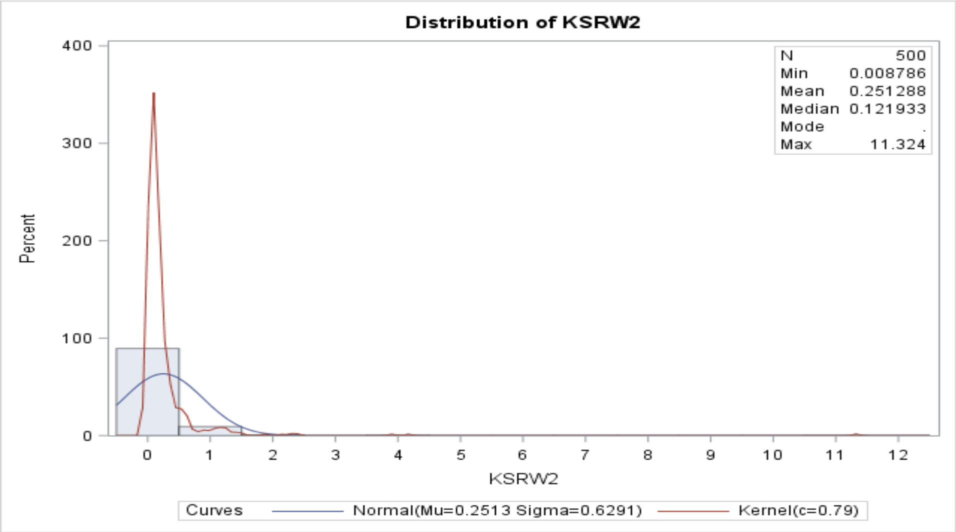


Figure B.2 Distribution of Ridge Parameter for $\rho = 0.90,\ n = 200$.





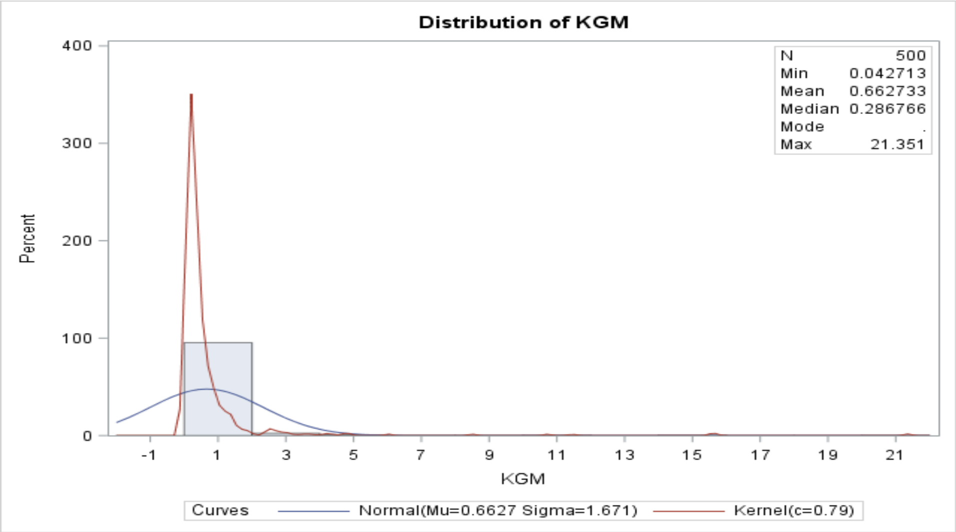
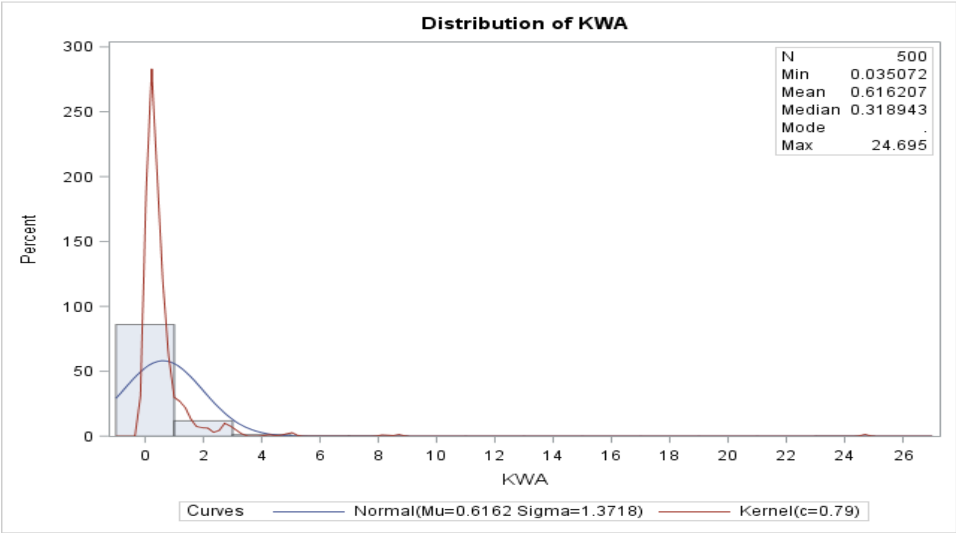


Figure B.2 (Continued)



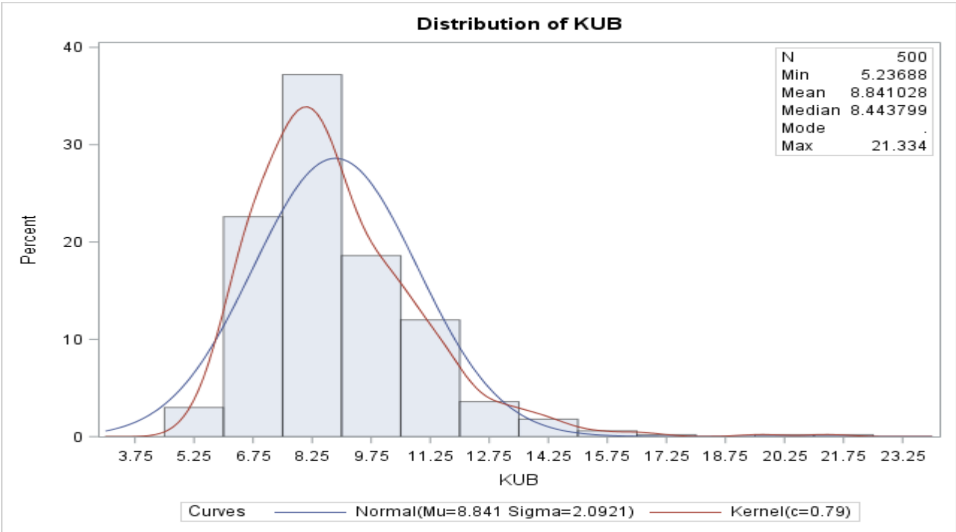
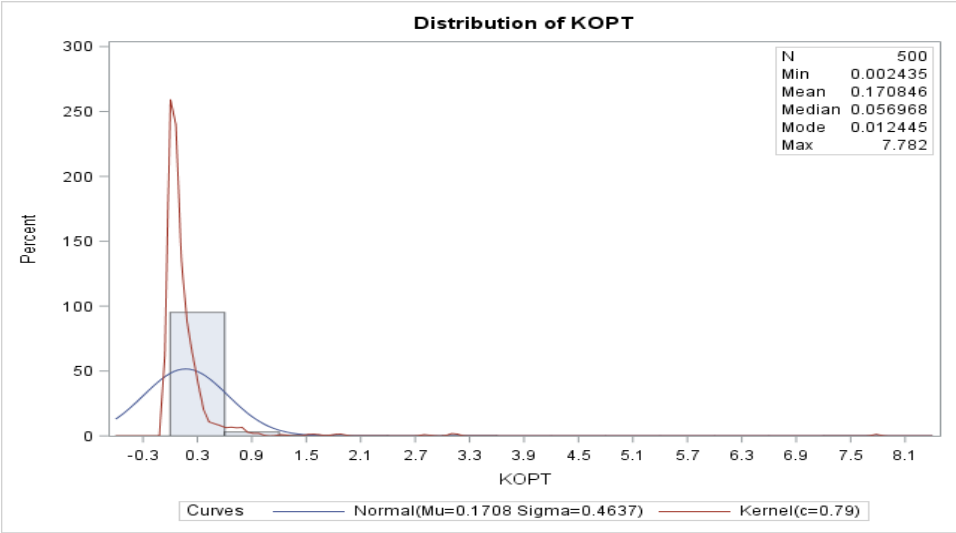
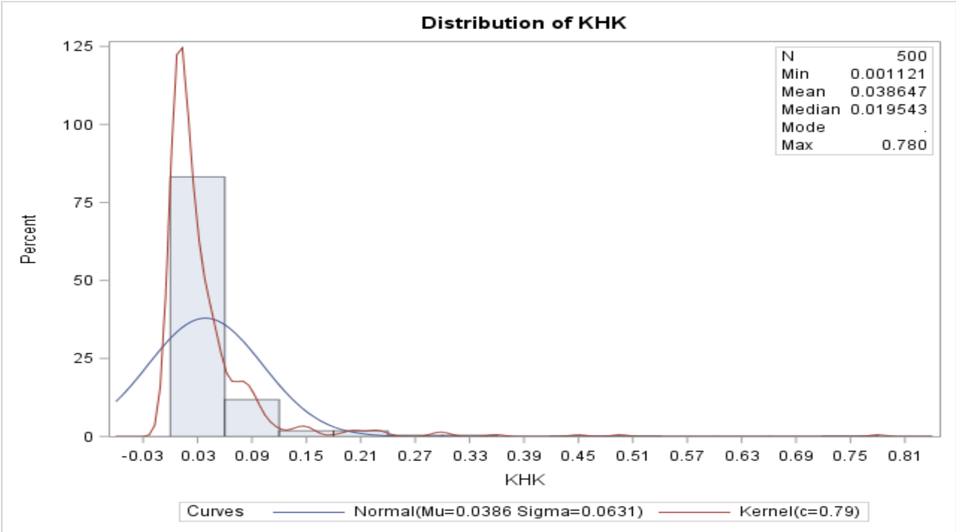


Figure B.2 (Continued)





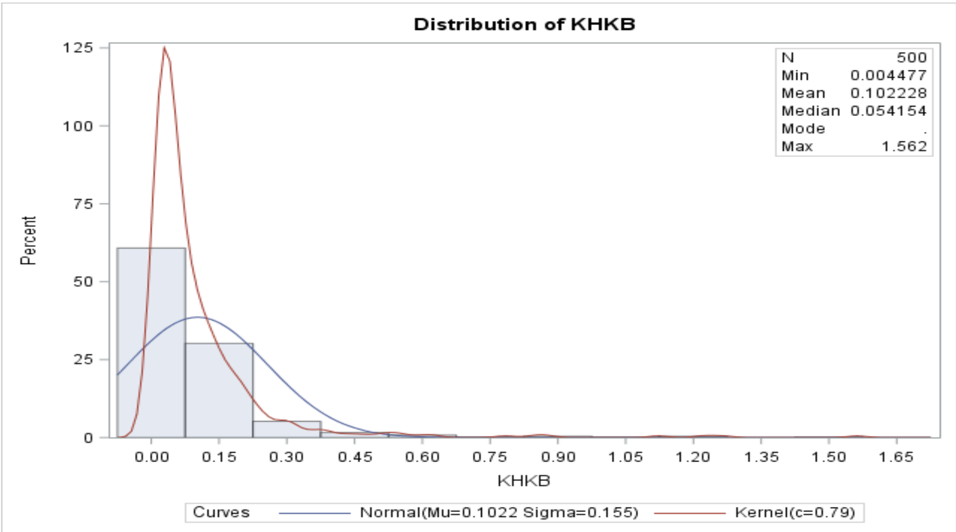
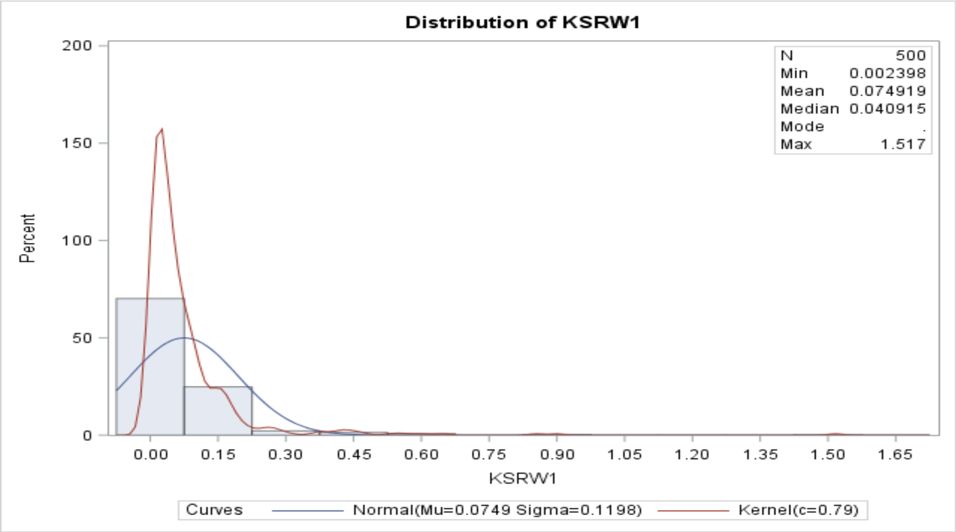
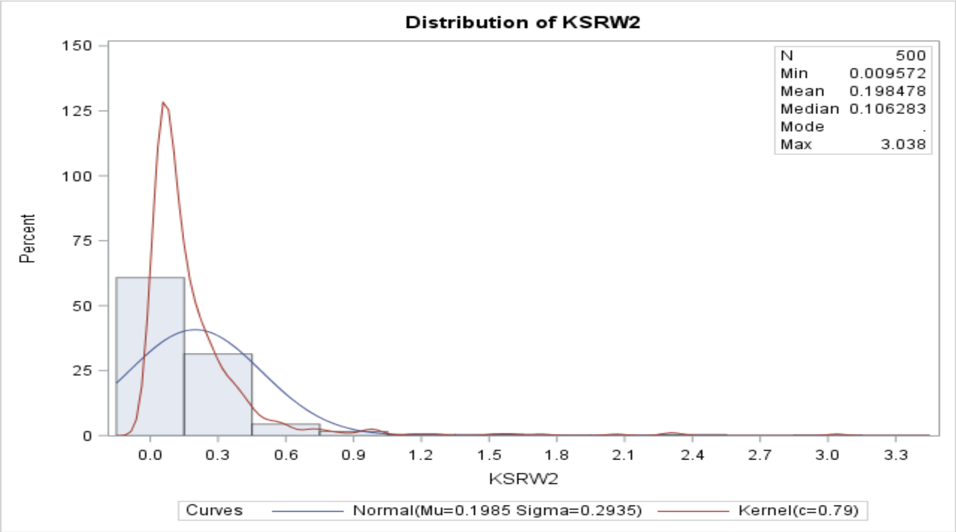


Figure B.3 Distribution of Ridge Parameter for $\rho = 0.90$, $n = 500$.





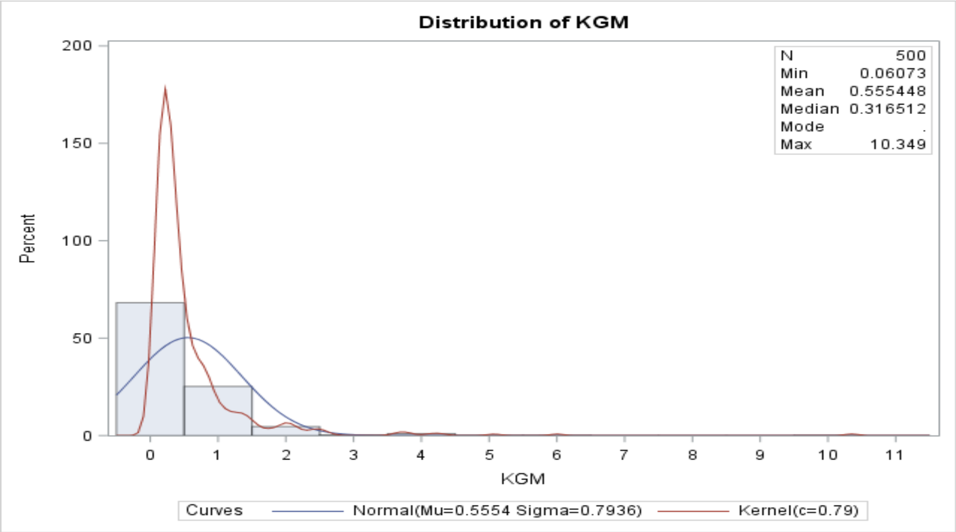
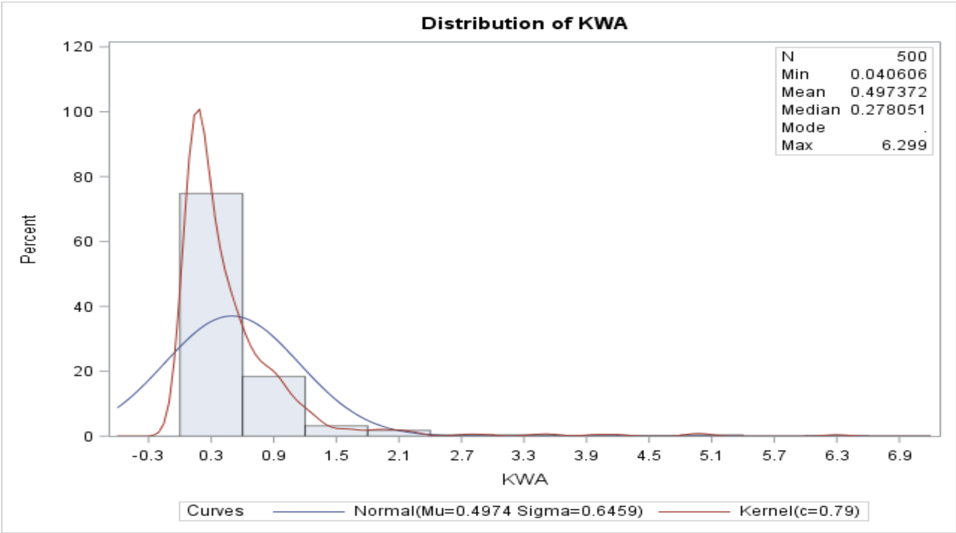


Figure B.3 (Continued)



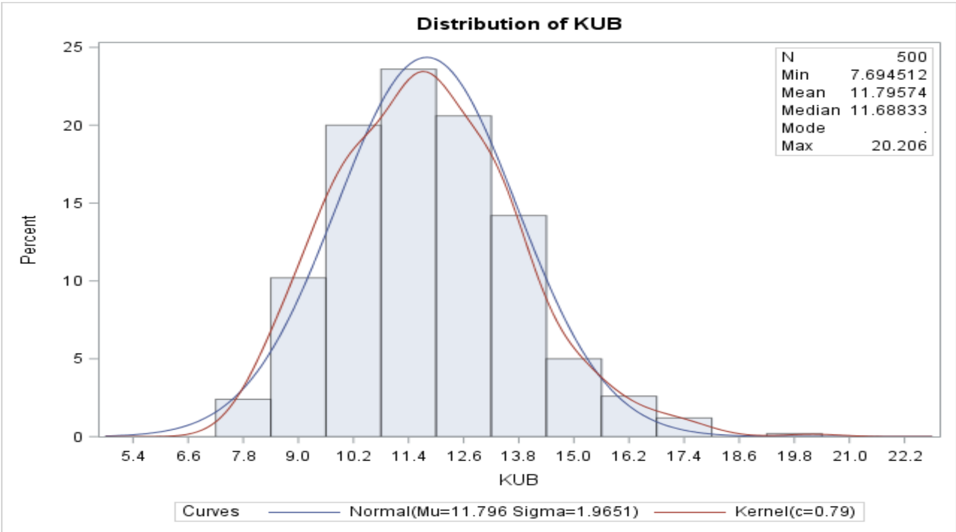
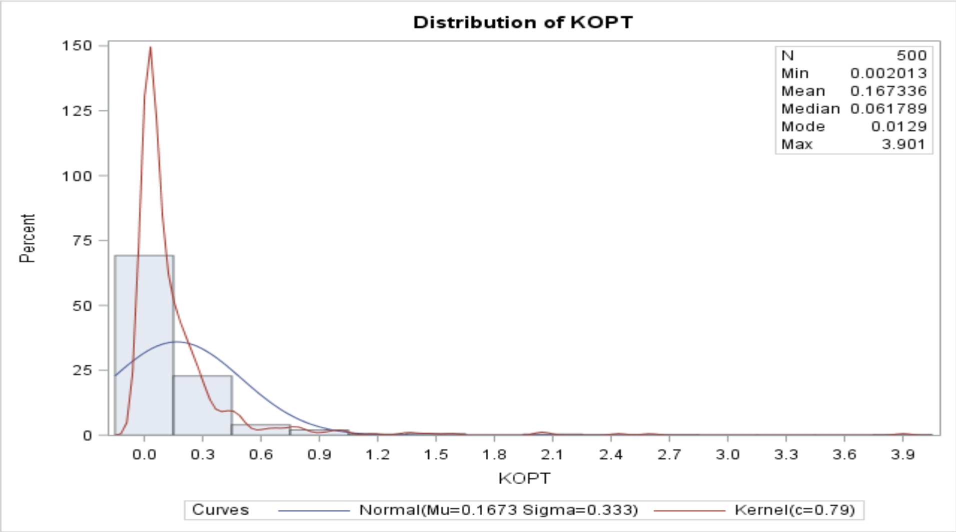
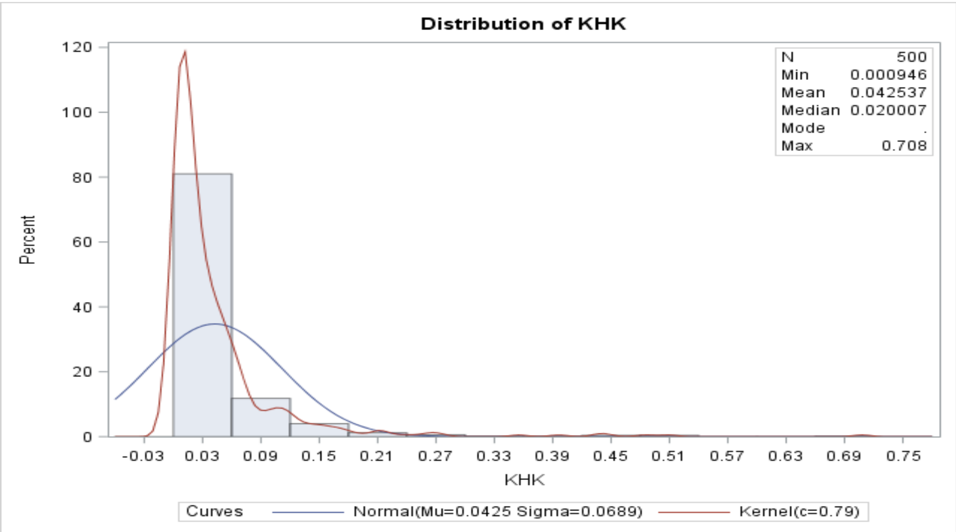


Figure B.3 (Continued)





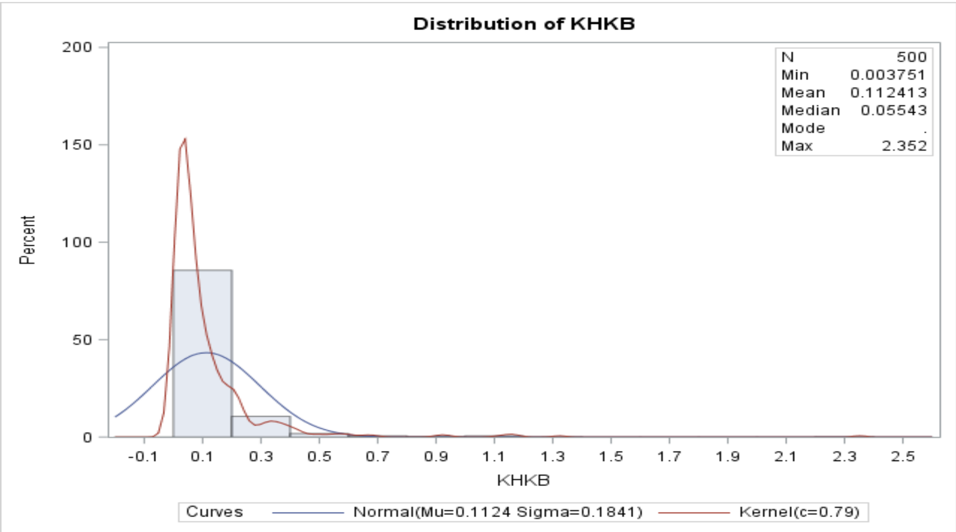
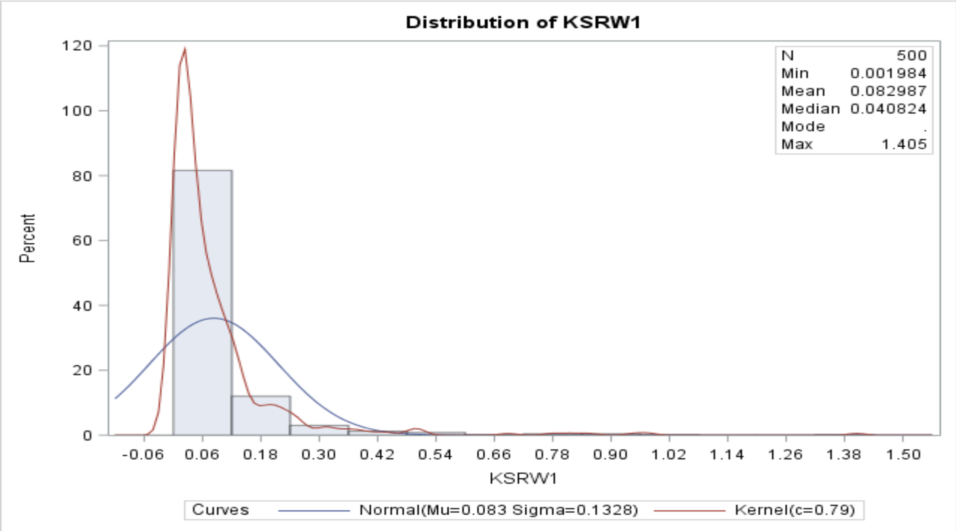
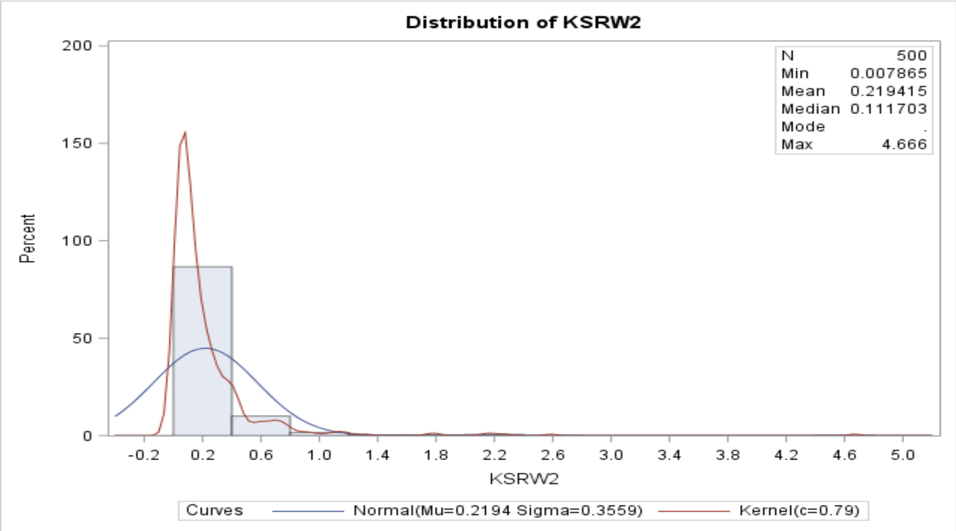


Figure B.4 Distribution of Ridge Parameter for $\rho = 0.90,\ n = 1000$.





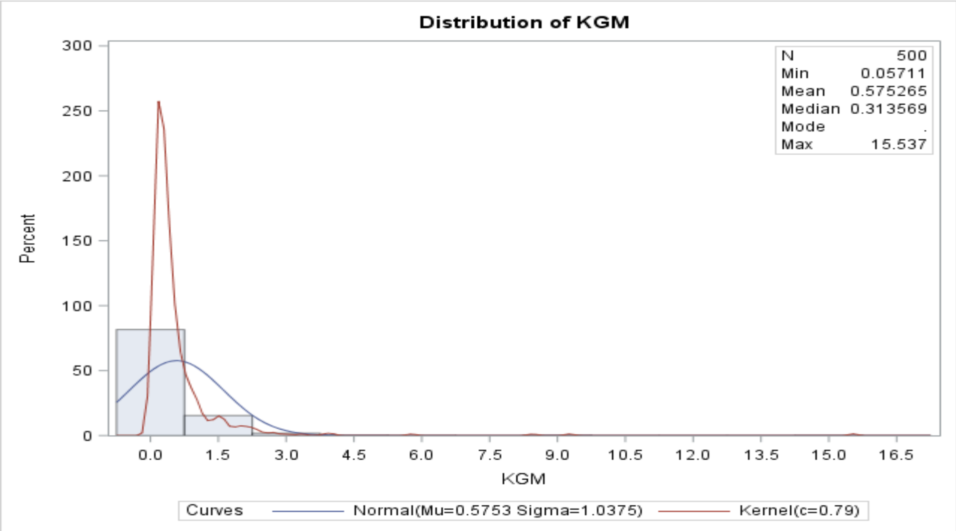
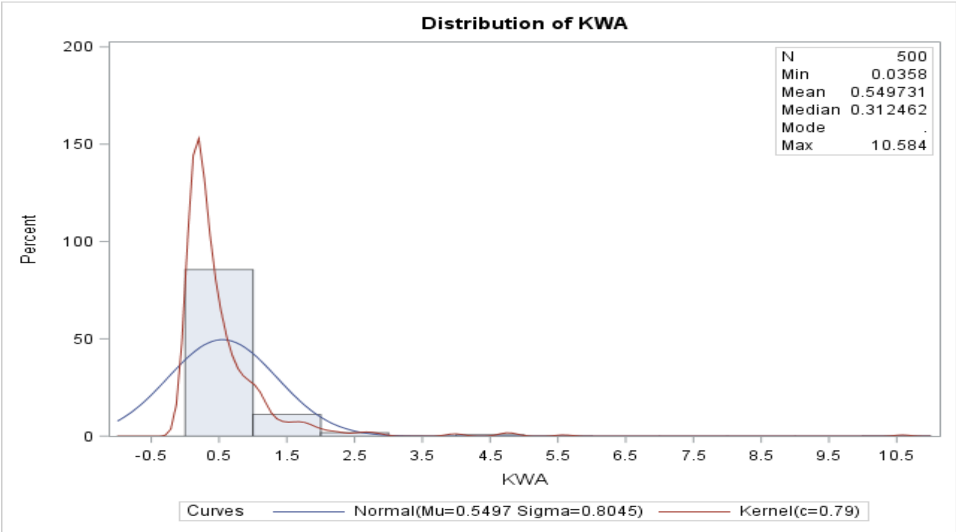


Figure B.4 (Continued)



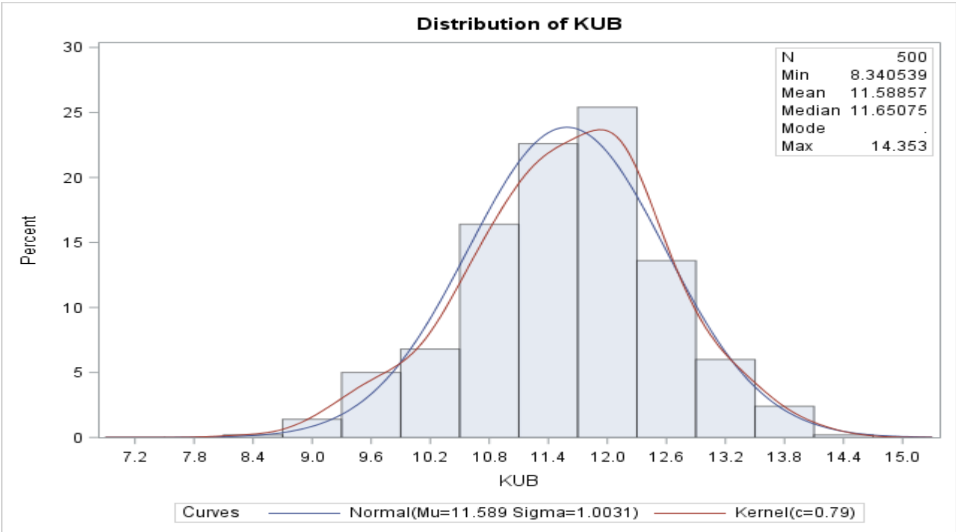


Figure B.4 (Continued)

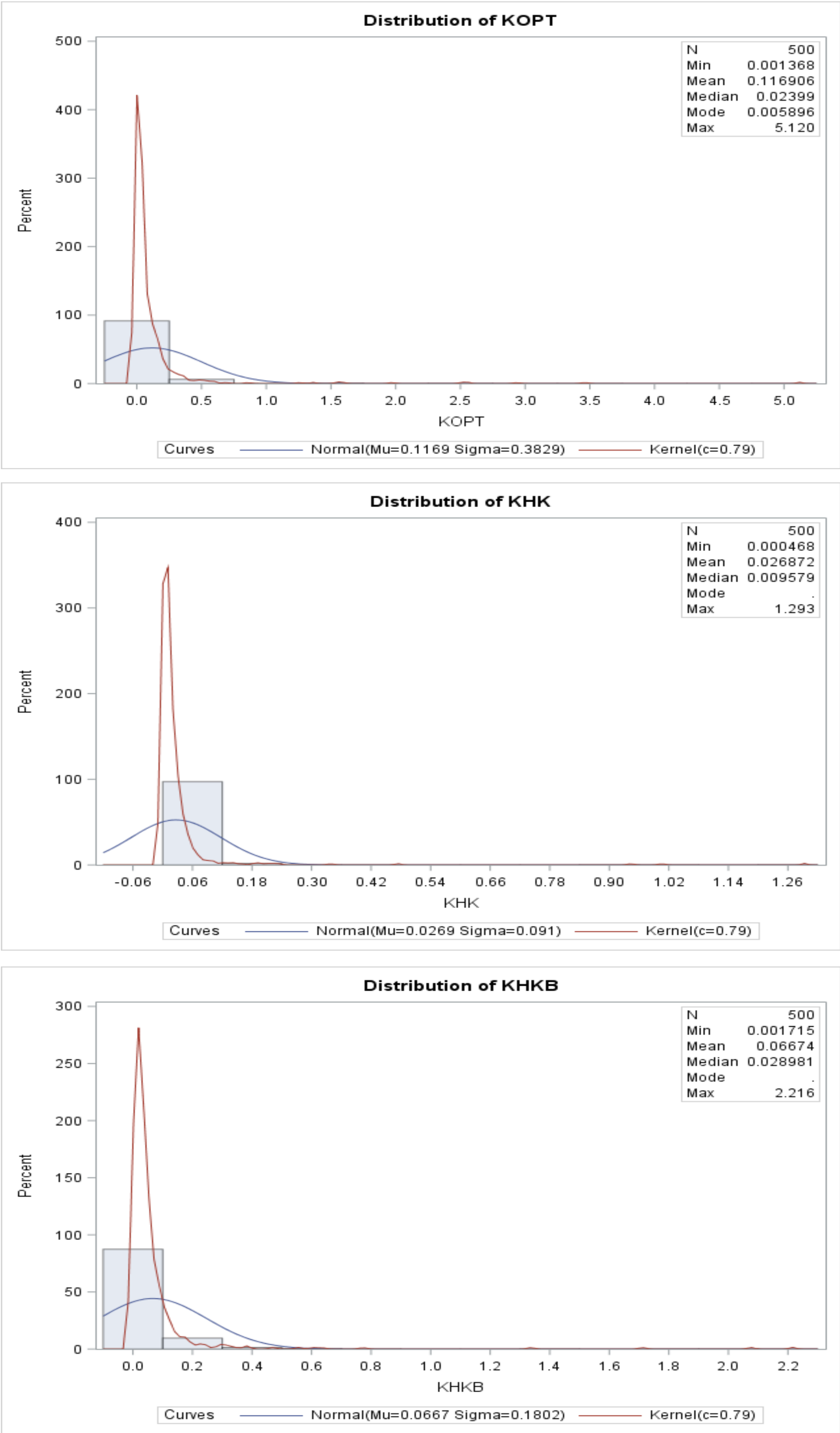
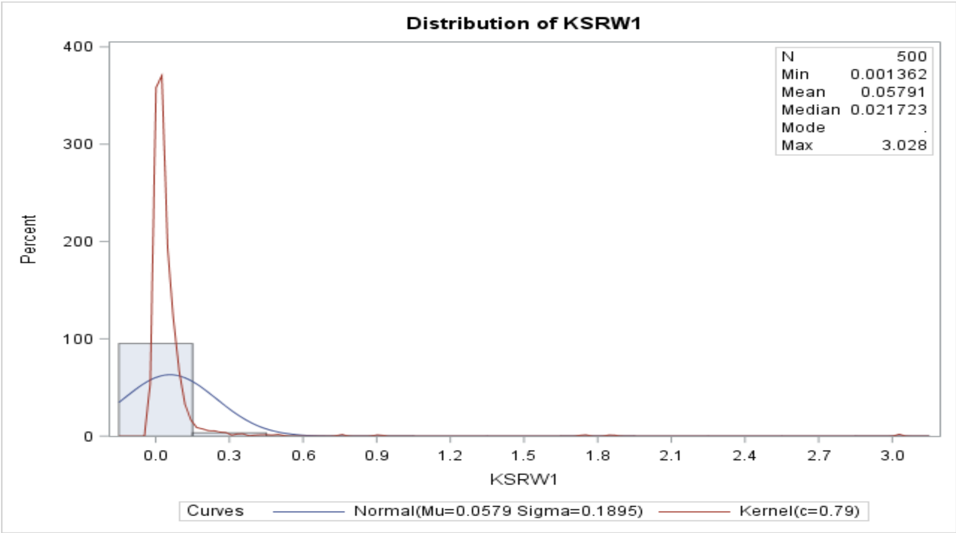
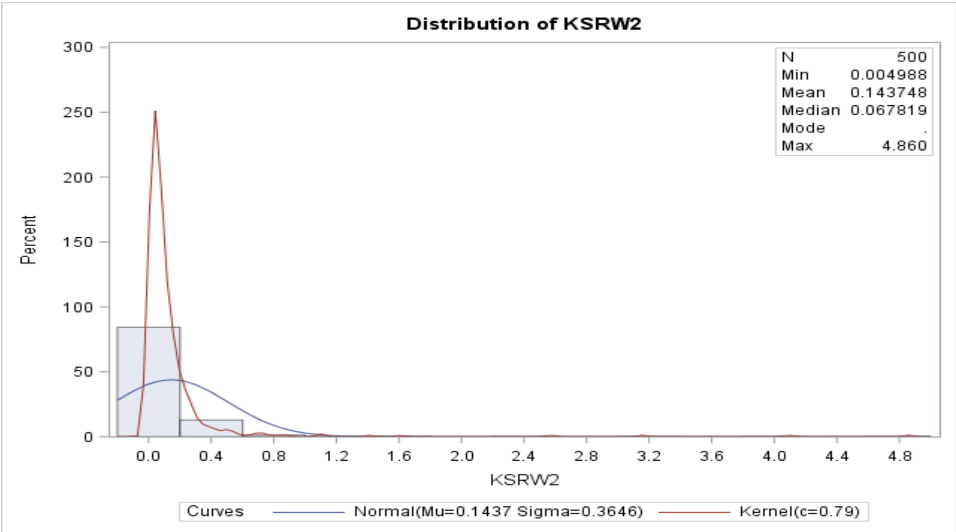


Figure B.5 Distribution of Ridge Parameter for $\rho = 0.95$, $n = 100$.





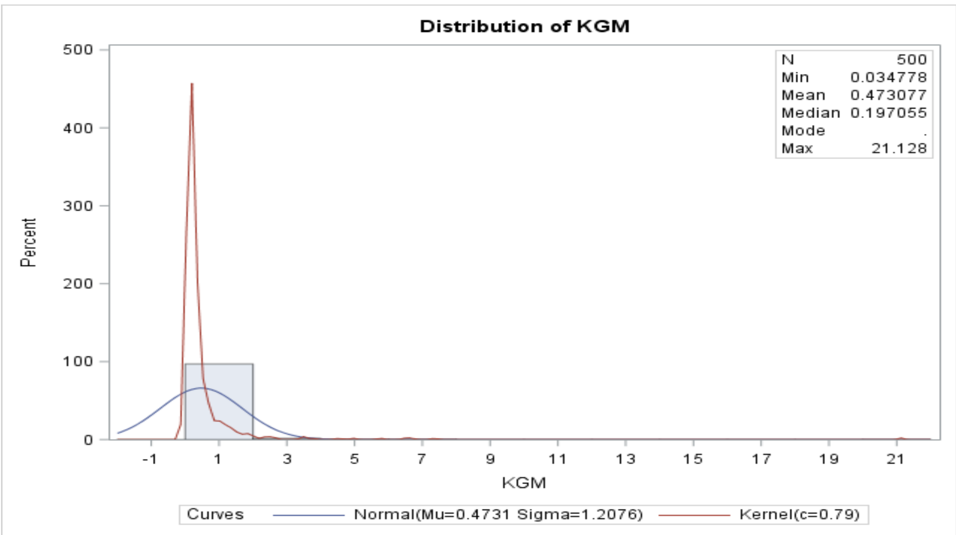
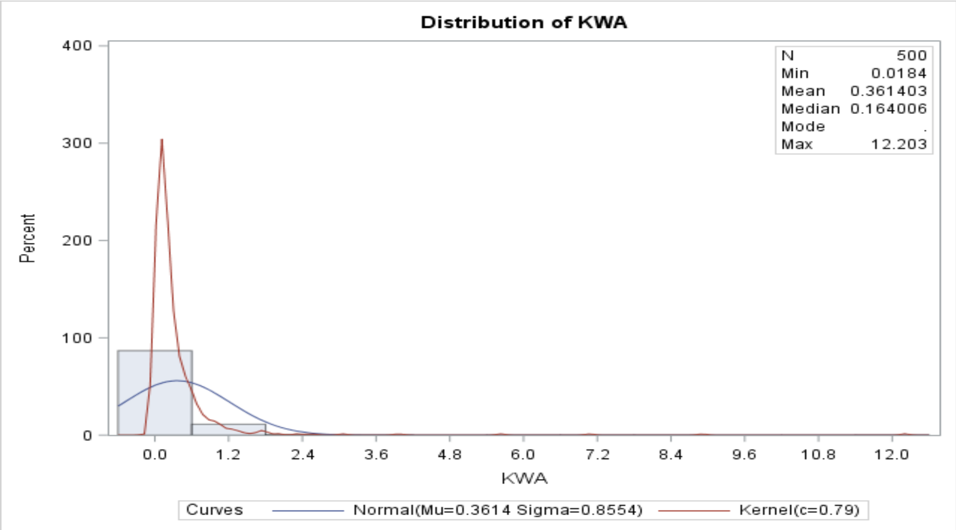


Figure B.5 (Continued)



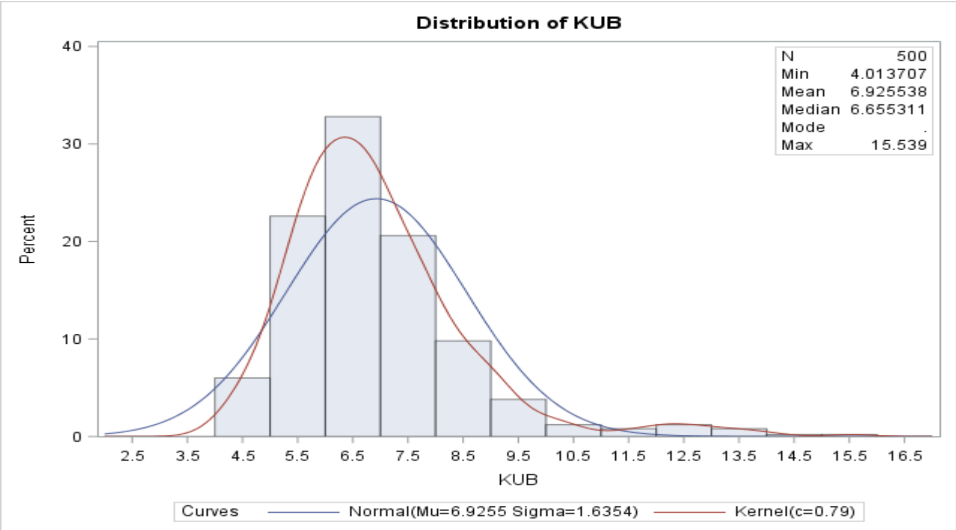


Figure B.5 (Continued)

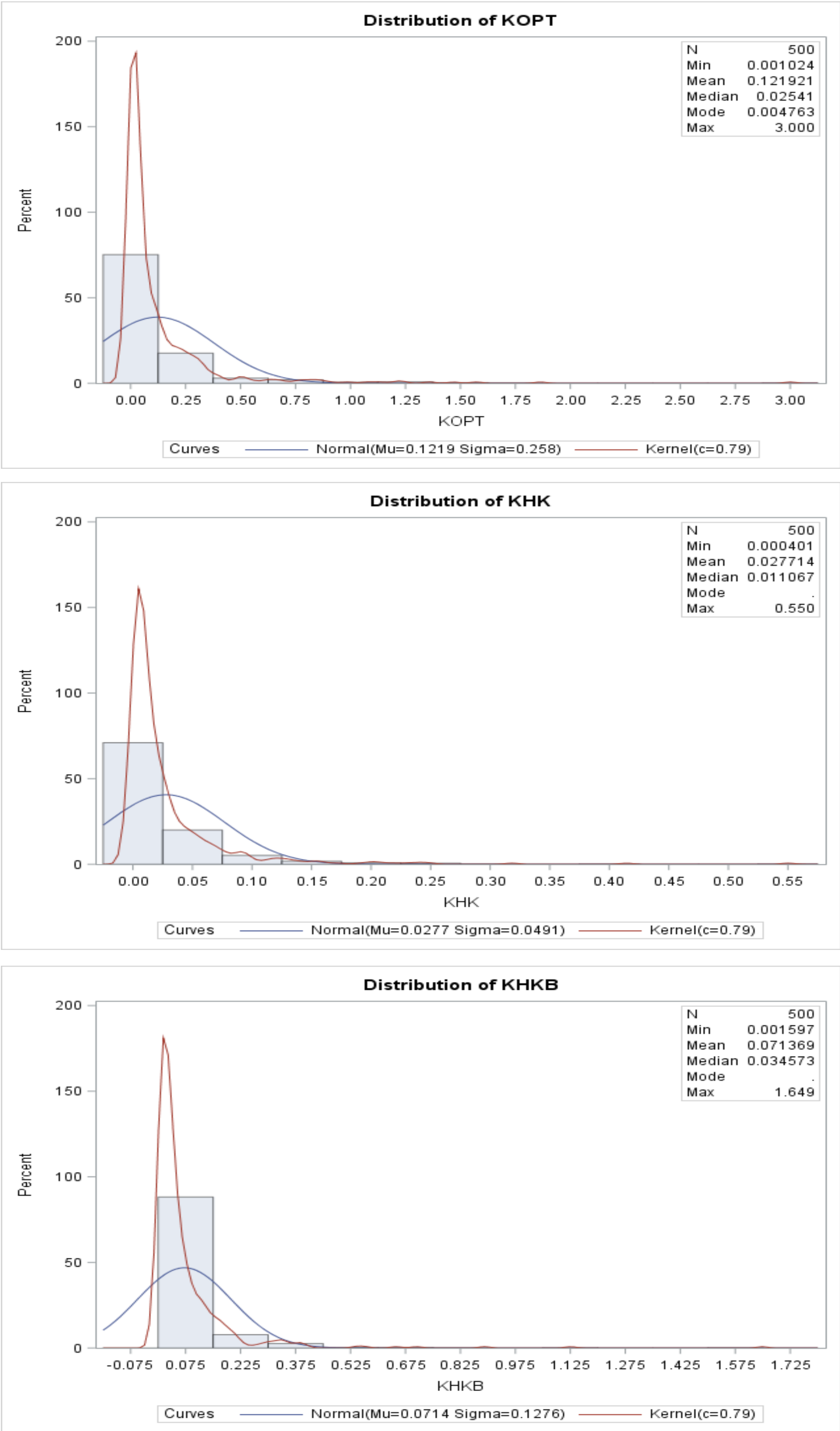
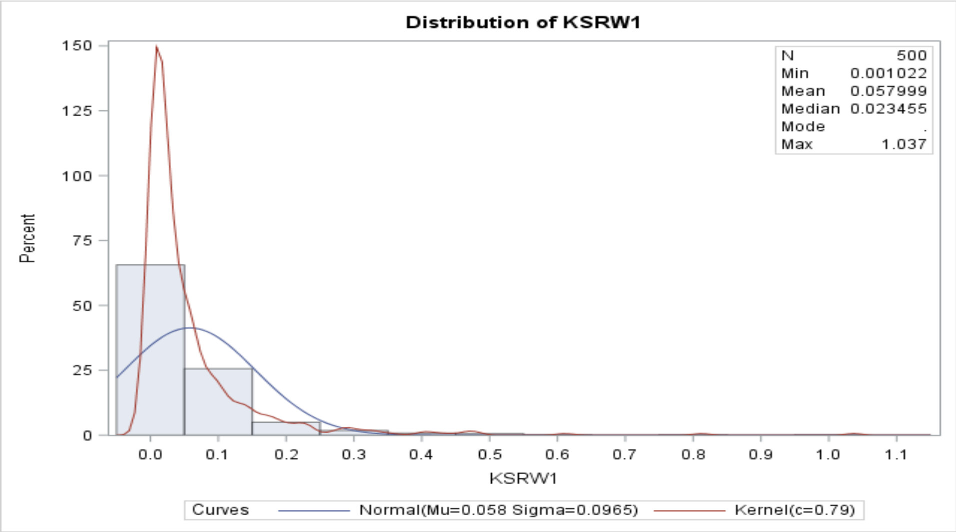
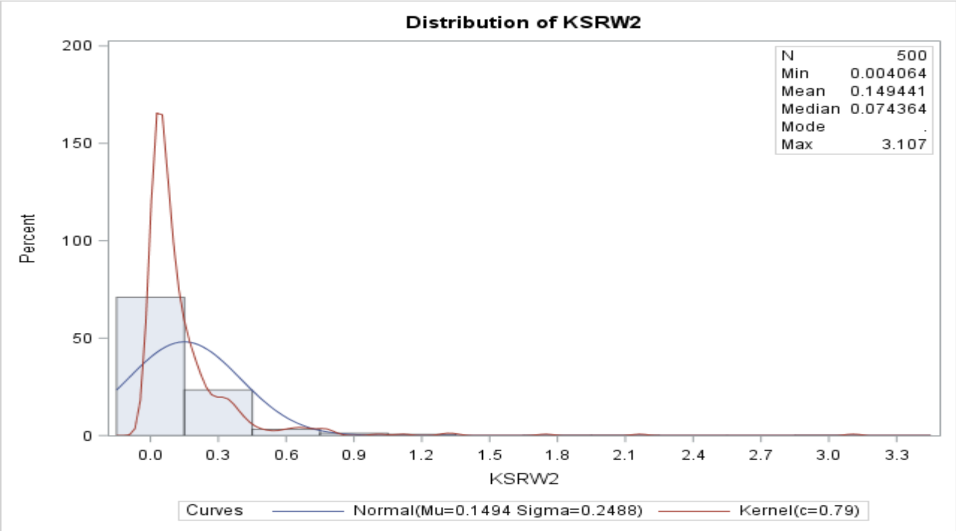


Figure B.6 Distribution of Ridge Parameter for $\rho = 0.95,\ n = 200$.





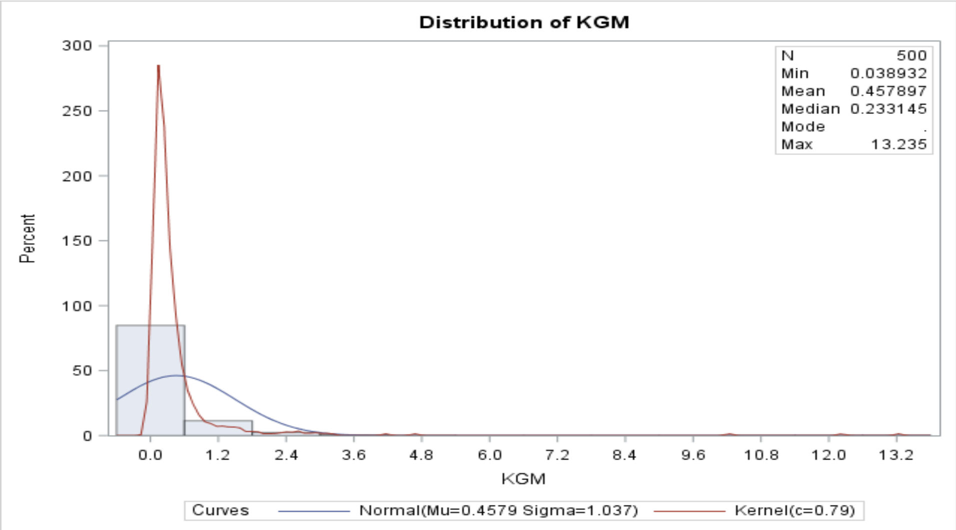
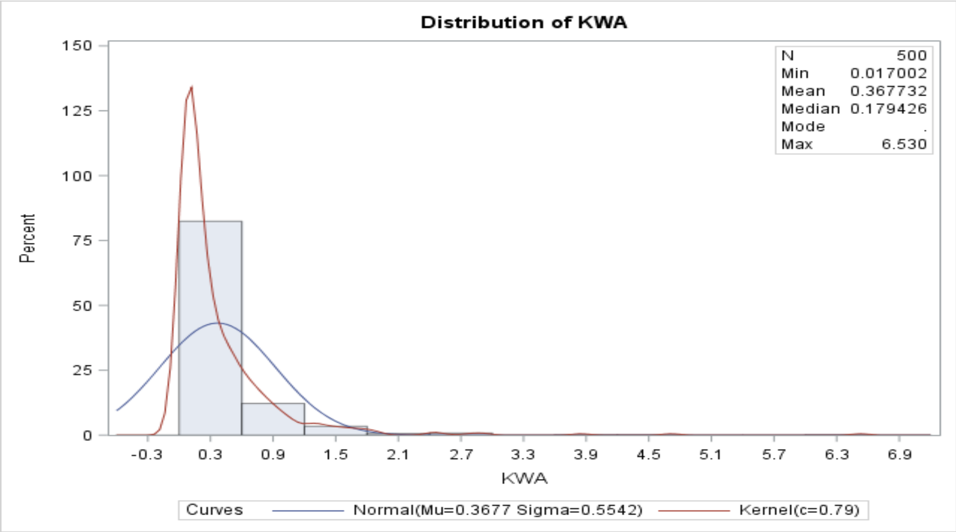


Figure B.6 (Continued)



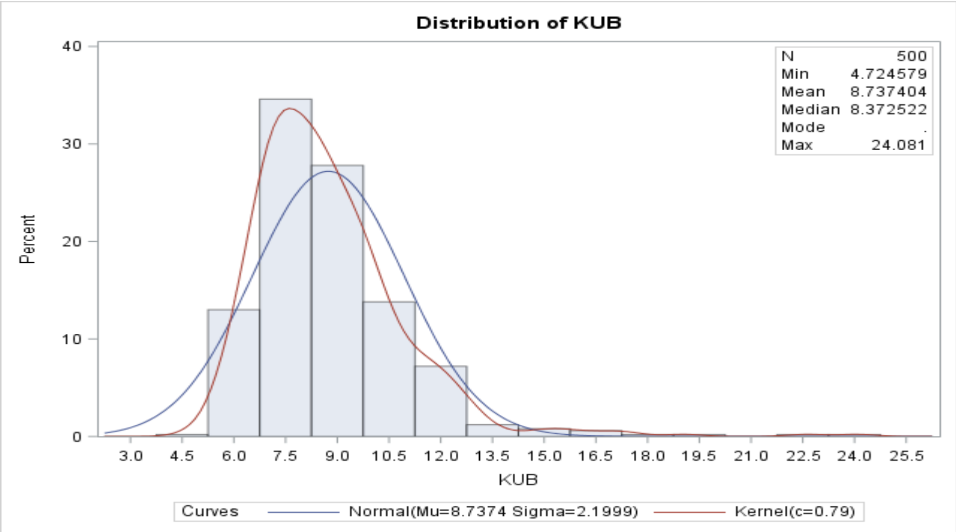


Figure B.6 (Continued)

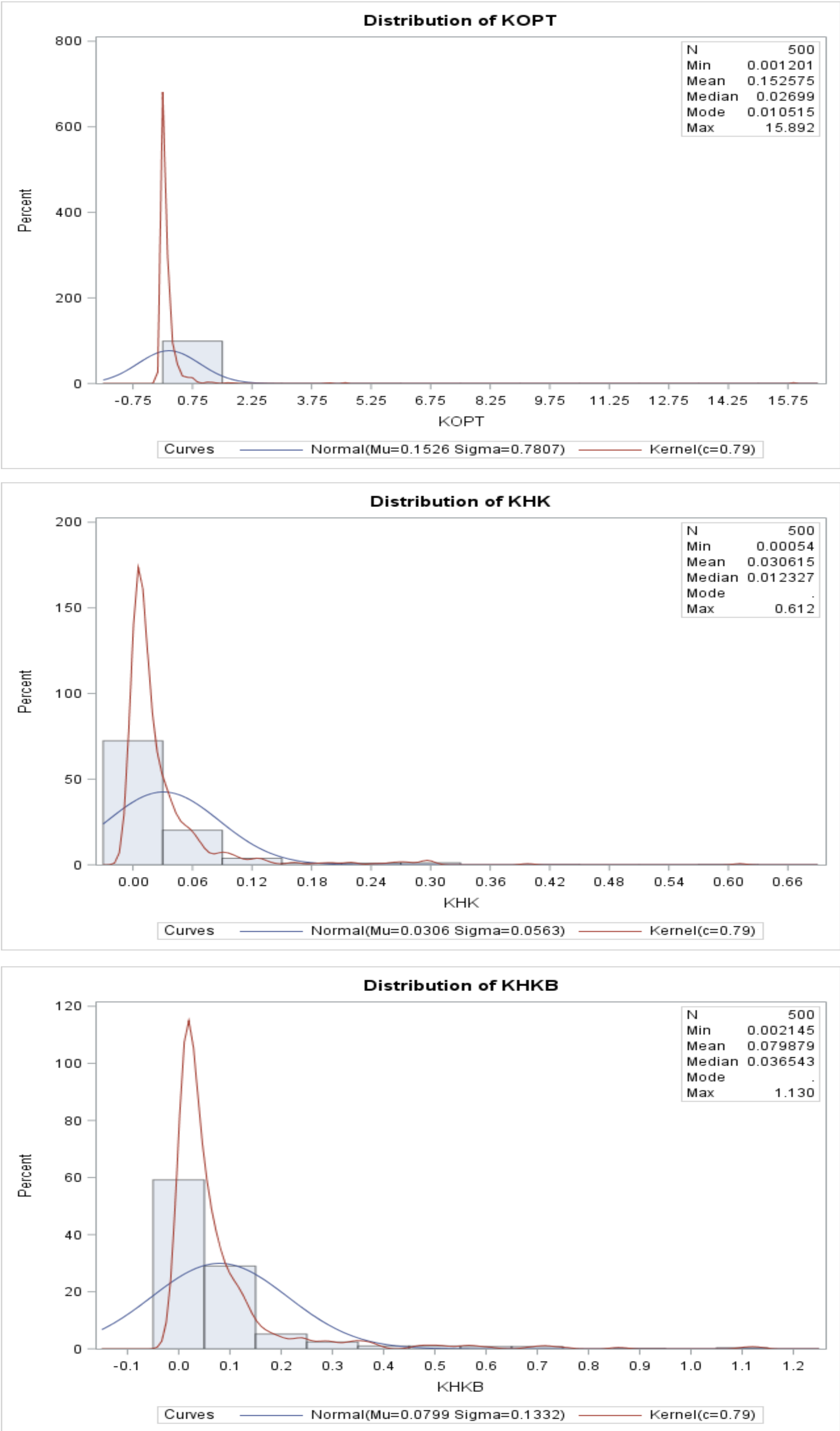
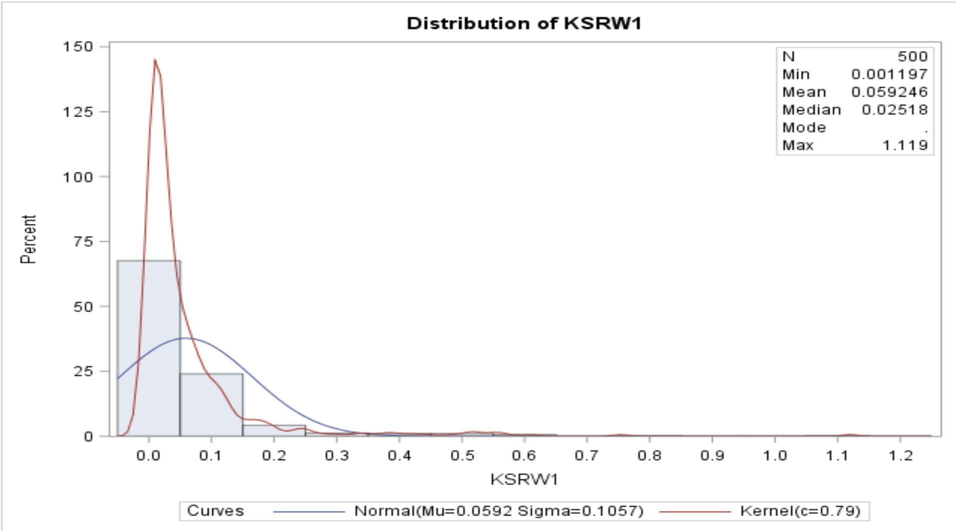
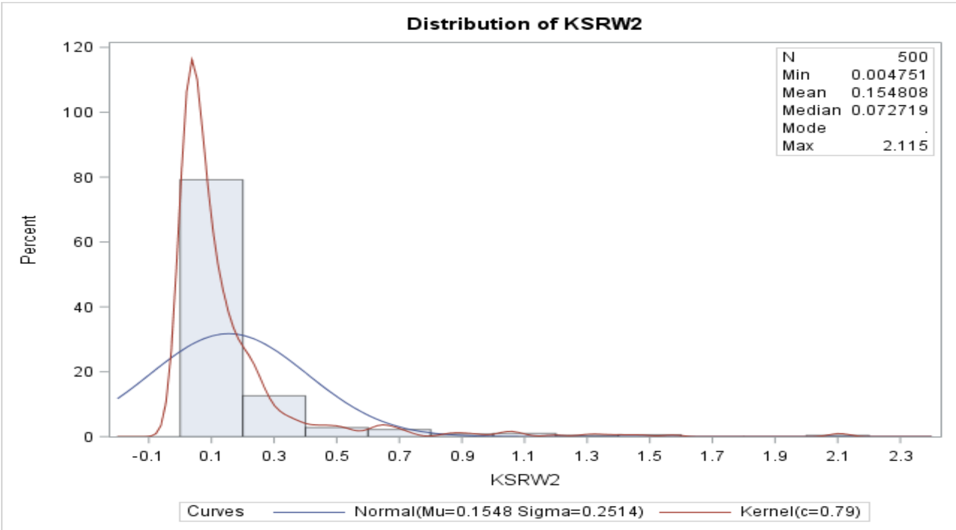


Figure B.7 Distribution of Ridge Parameter for $\rho = 0.95,\ n = 500$.





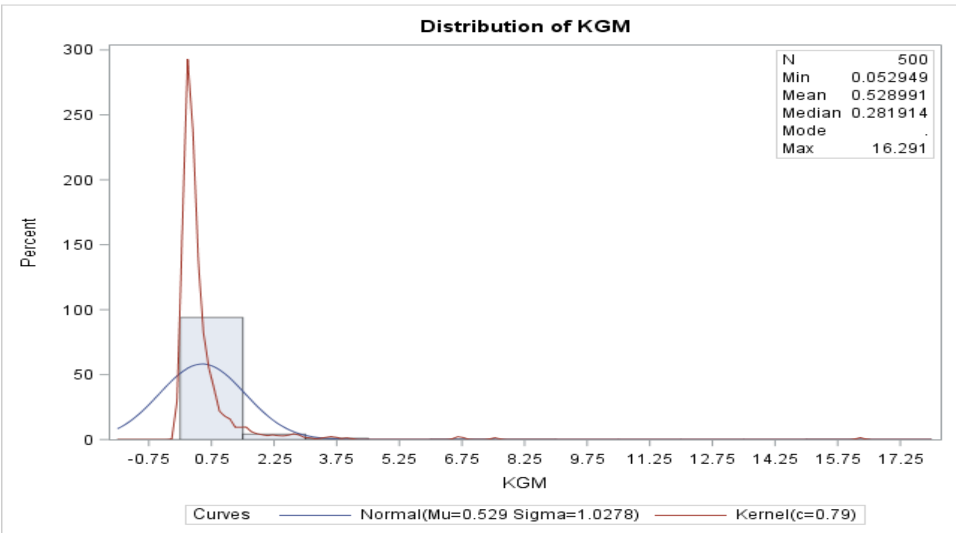
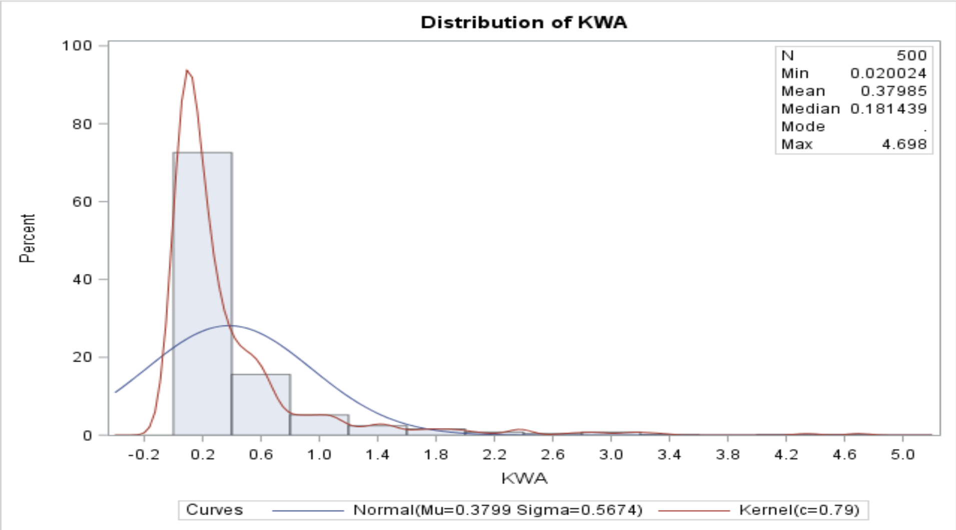


Figure B.7 (Continued)



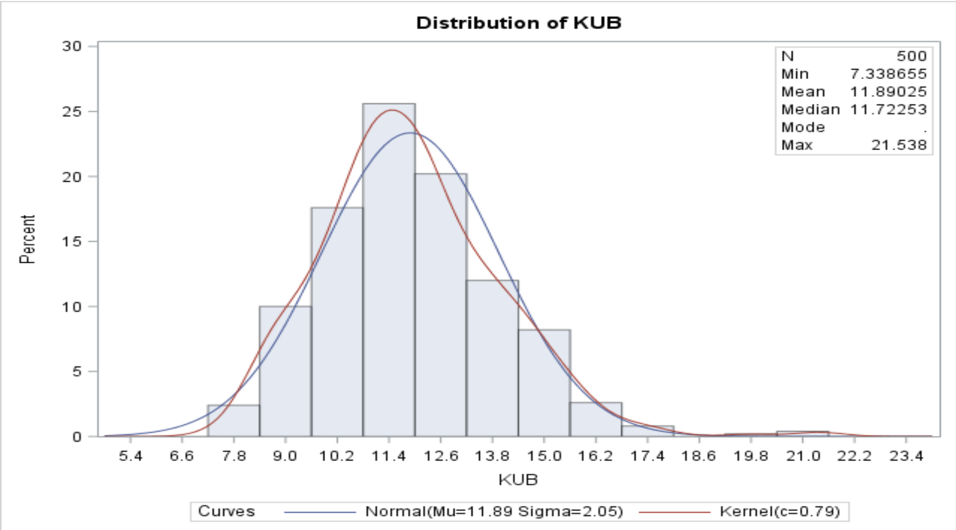
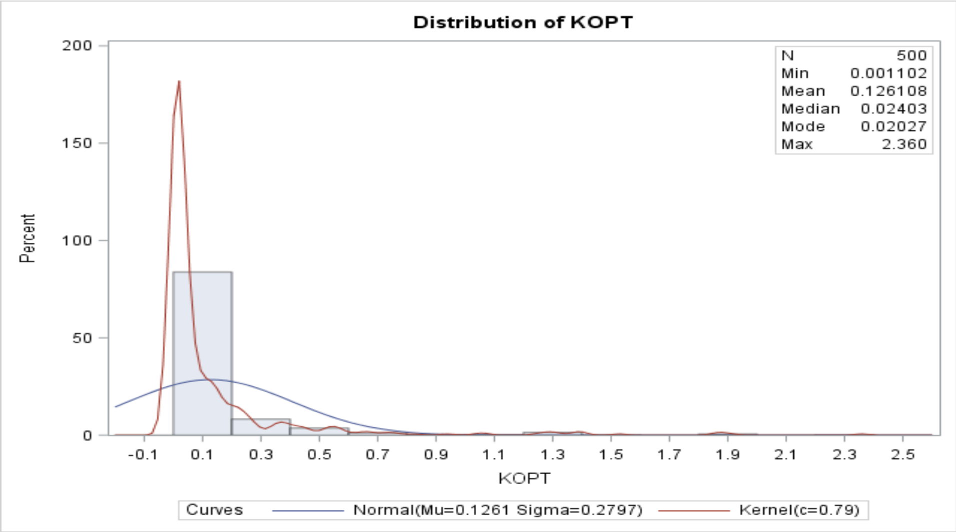
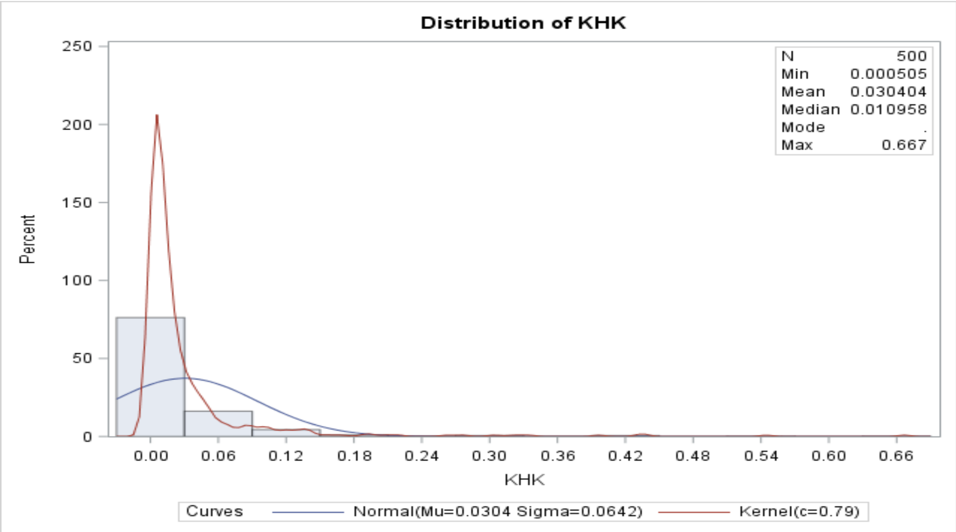


Figure B.7 (Continued)





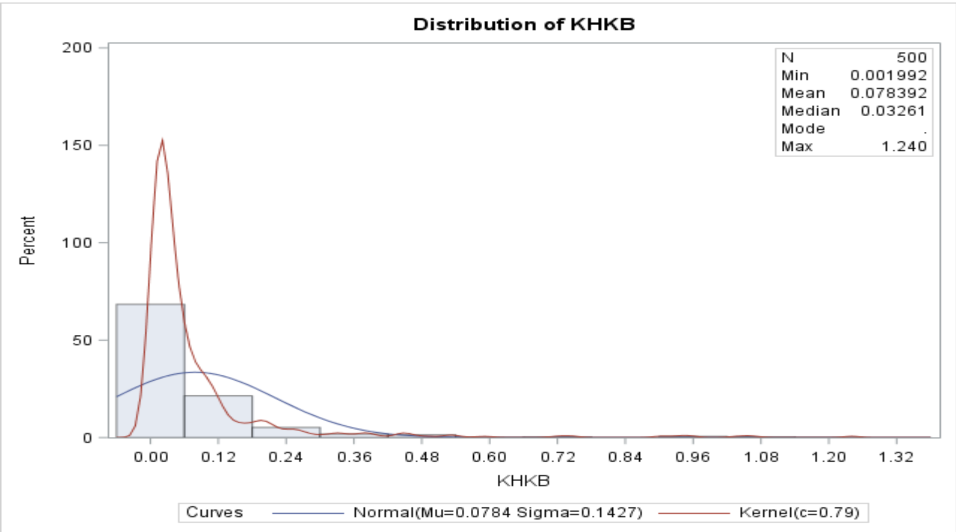
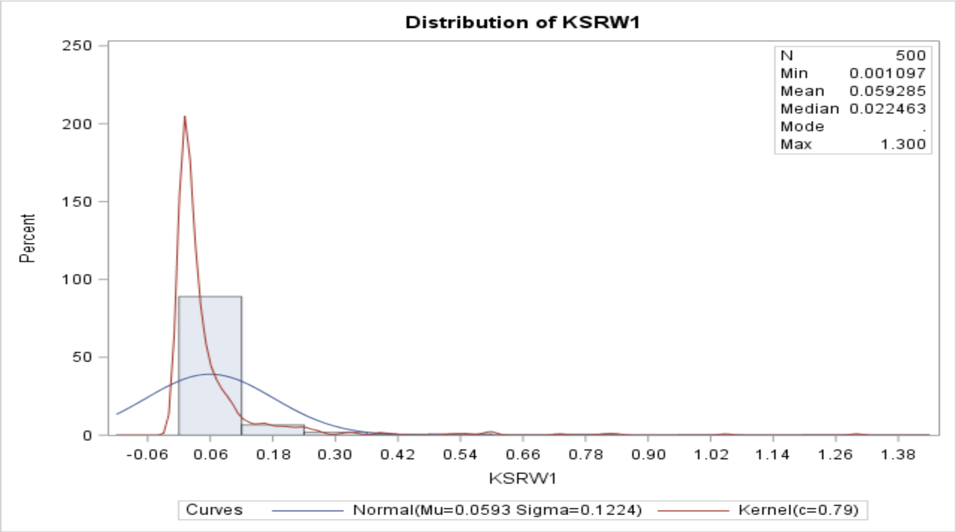
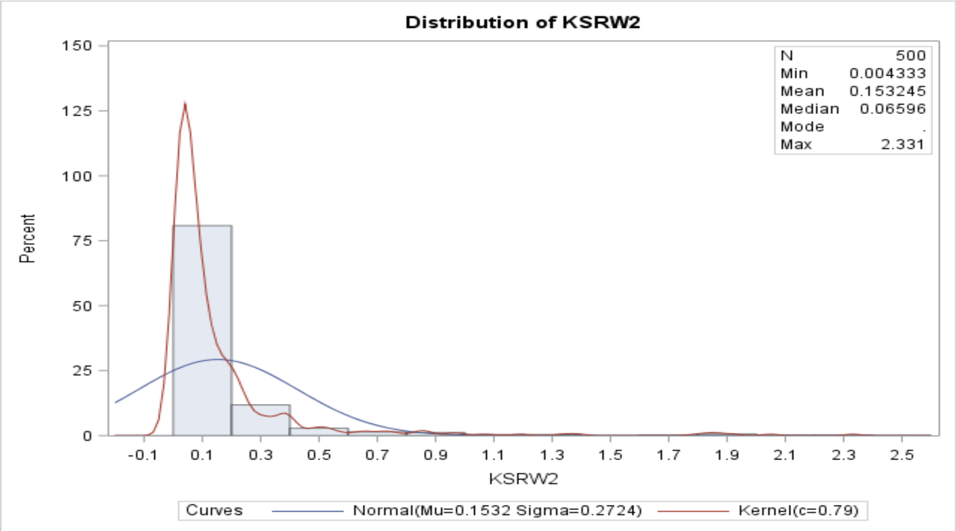


Figure B.8 Distribution of Ridge Parameter for $\rho = 0.95$, $n = 1000$.





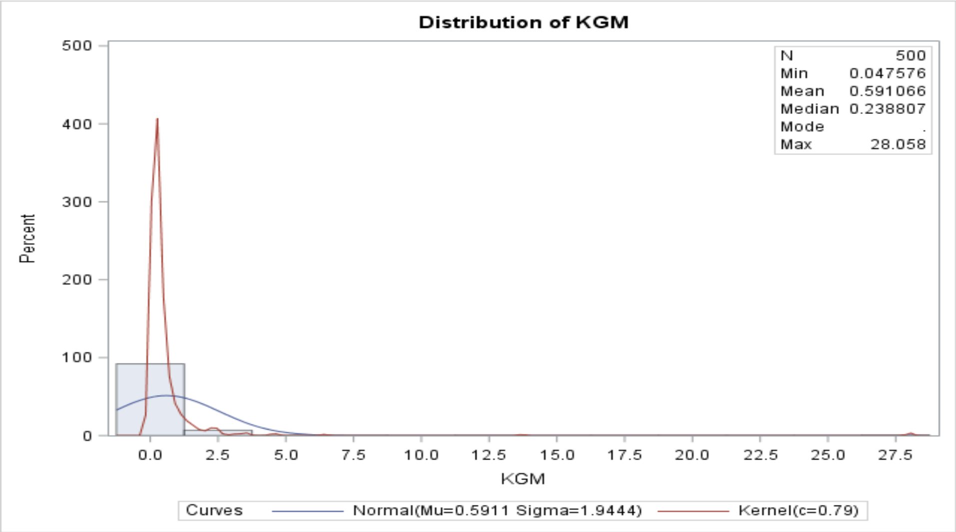
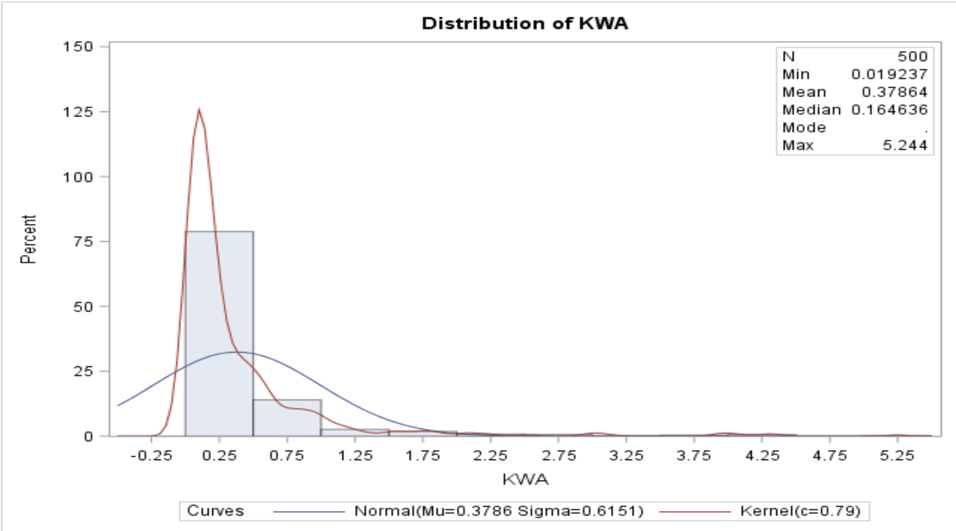


Figure B.8 (Continued)



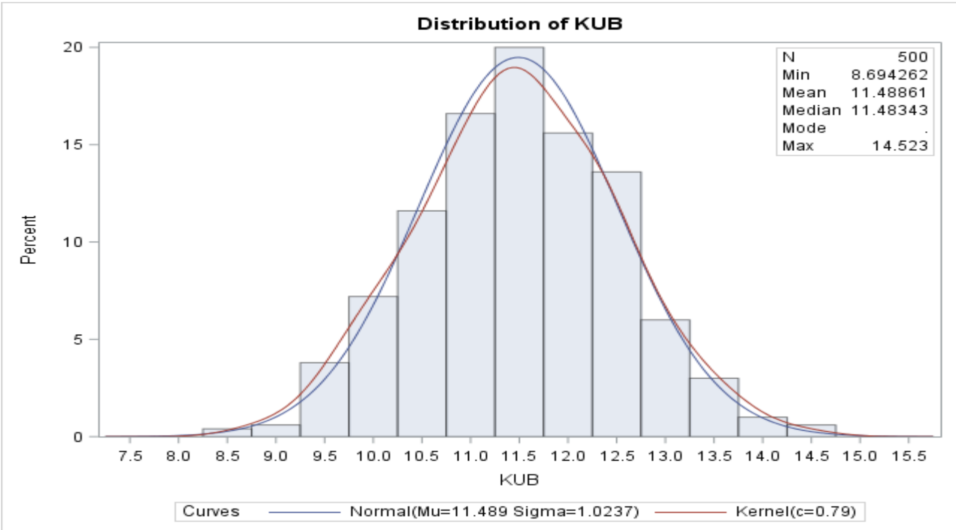


Figure B.8 (Continued)

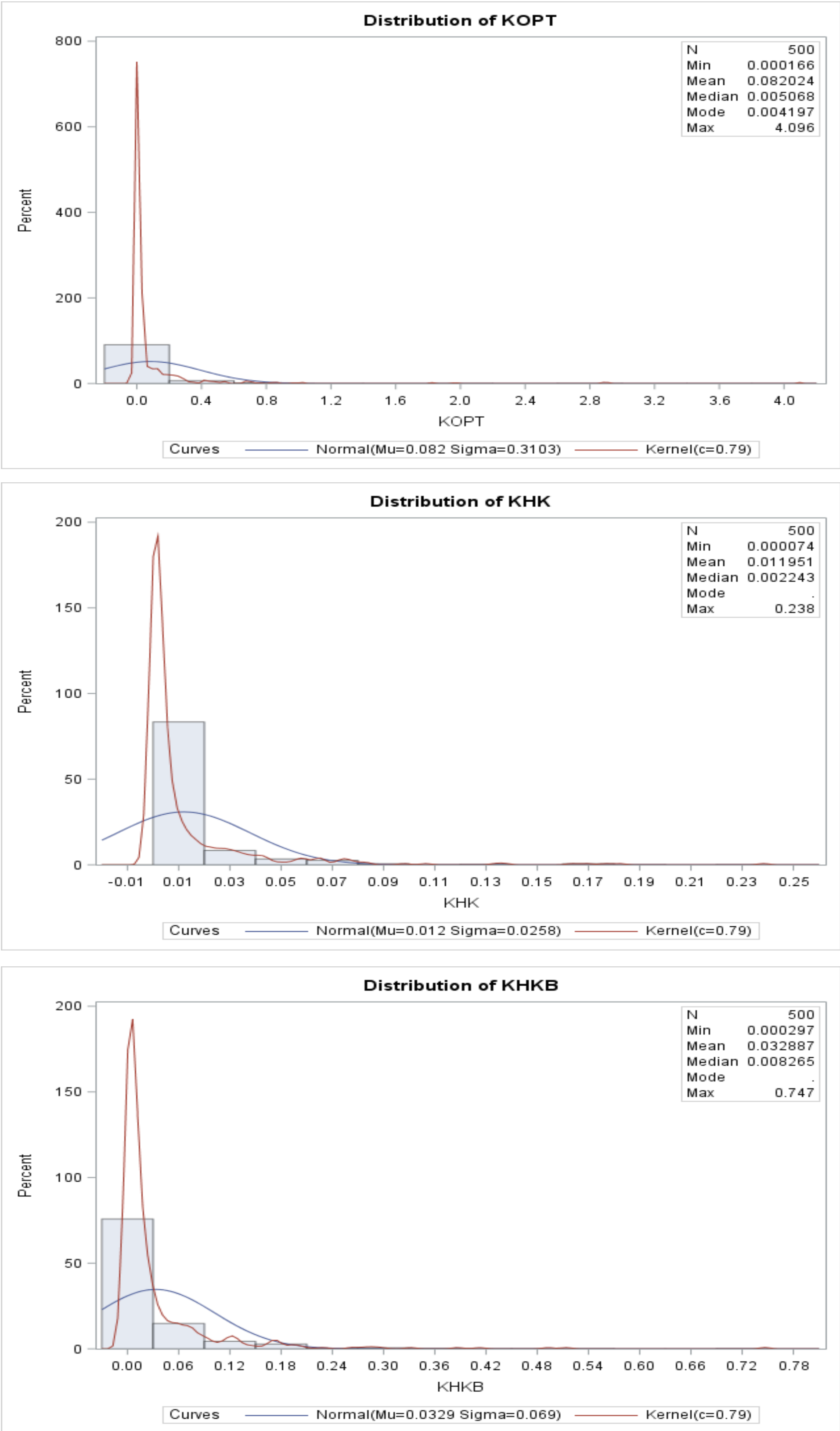
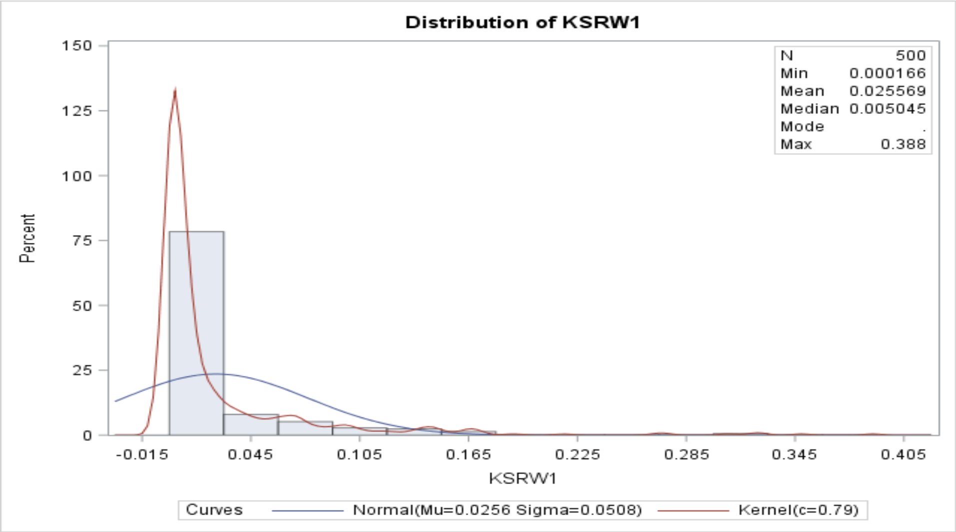
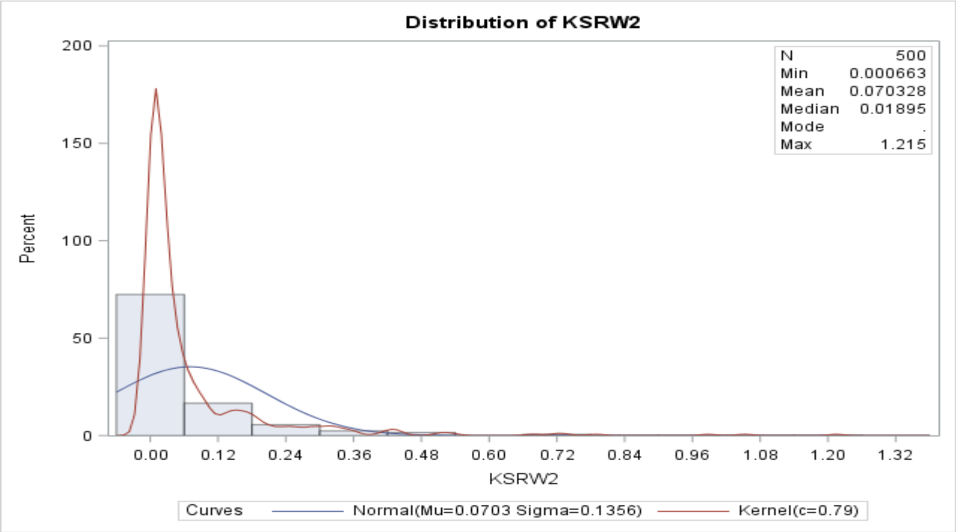


Figure B.9 Distribution of Ridge Parameter for $\rho = 0.99$, $n = 100$.





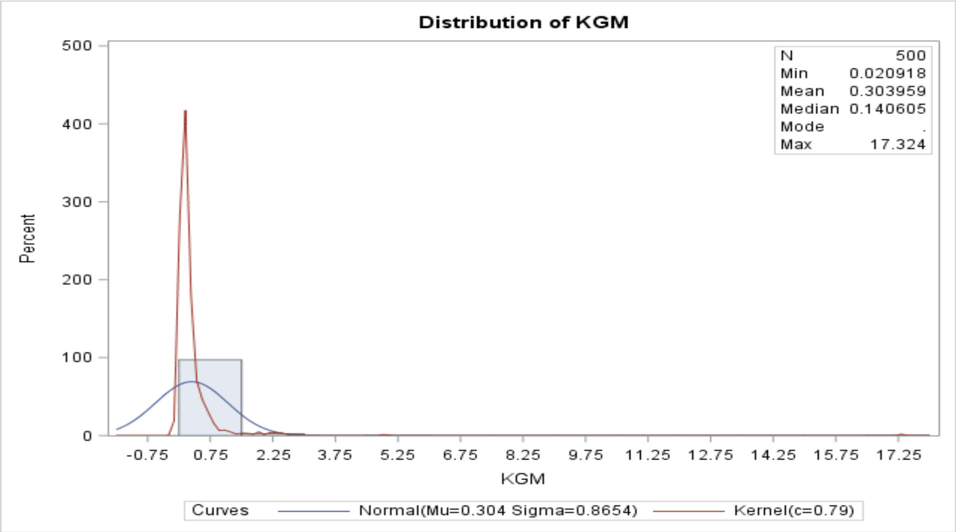
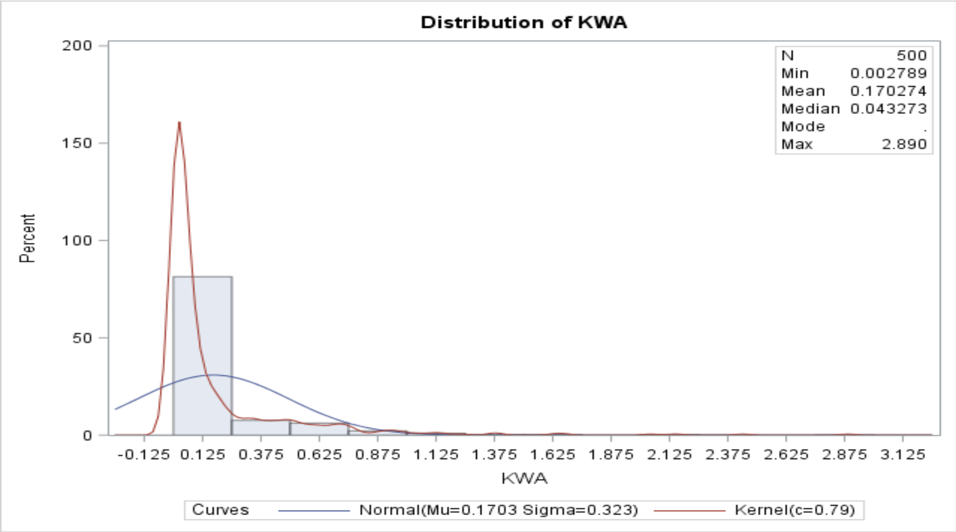


Figure B.9 (Continued)



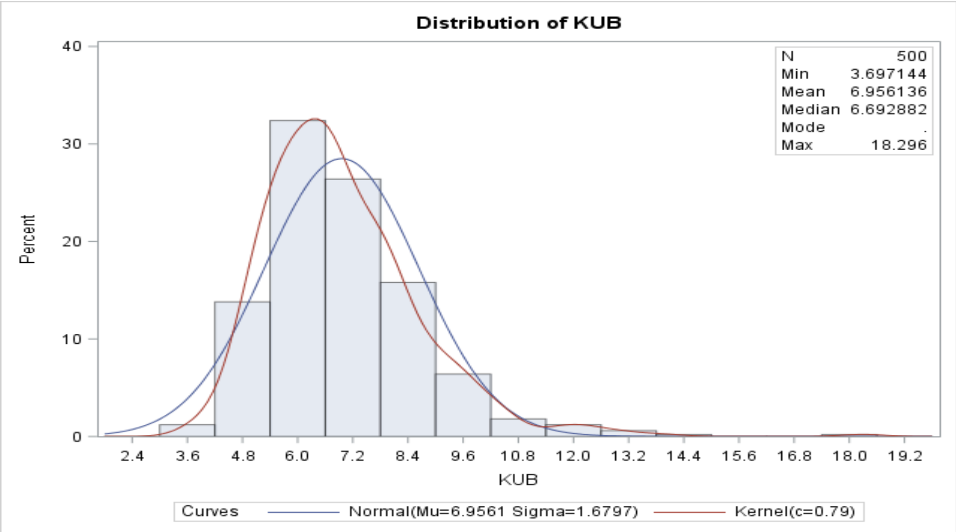


Figure B.9 (Continued)

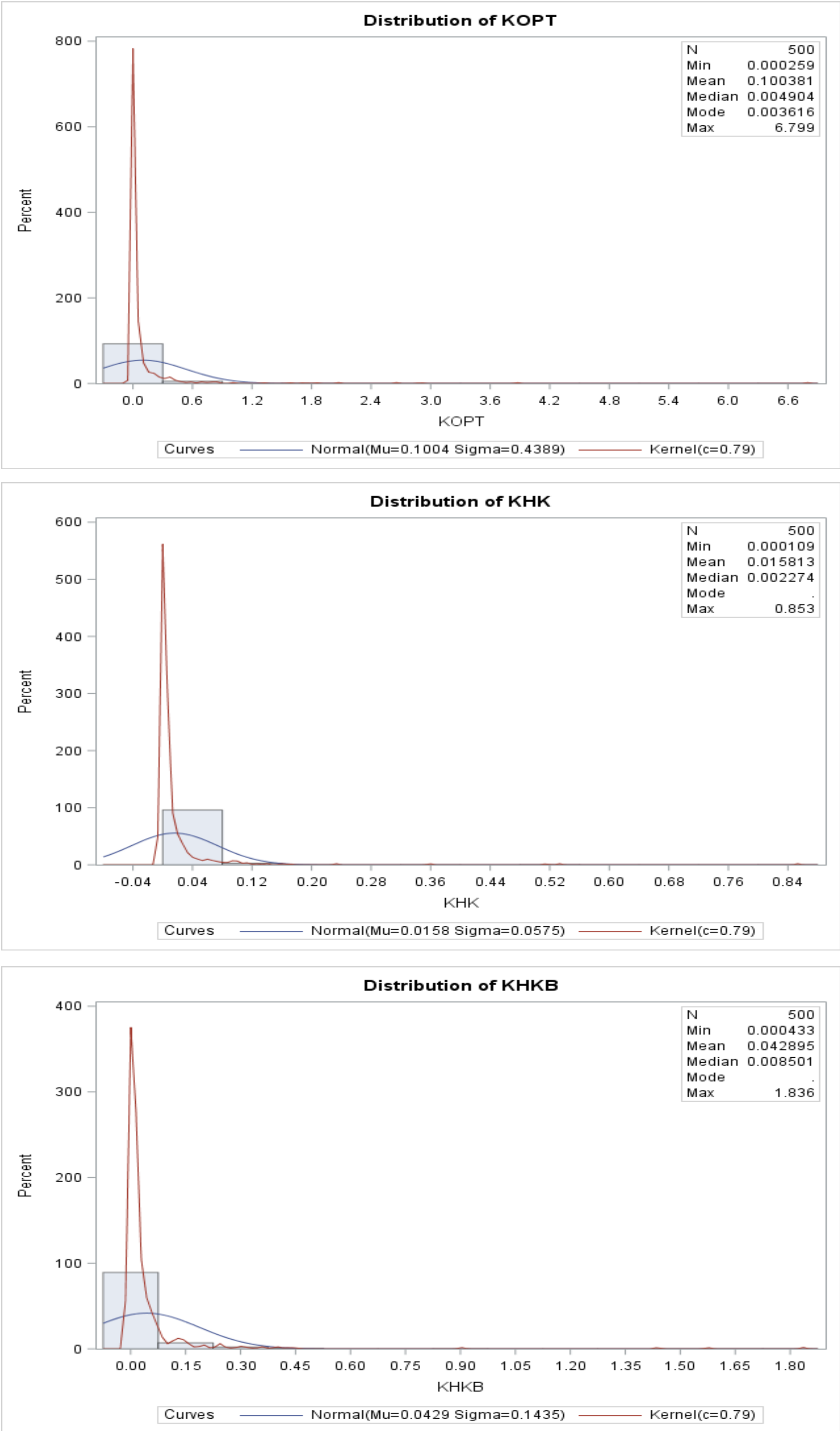
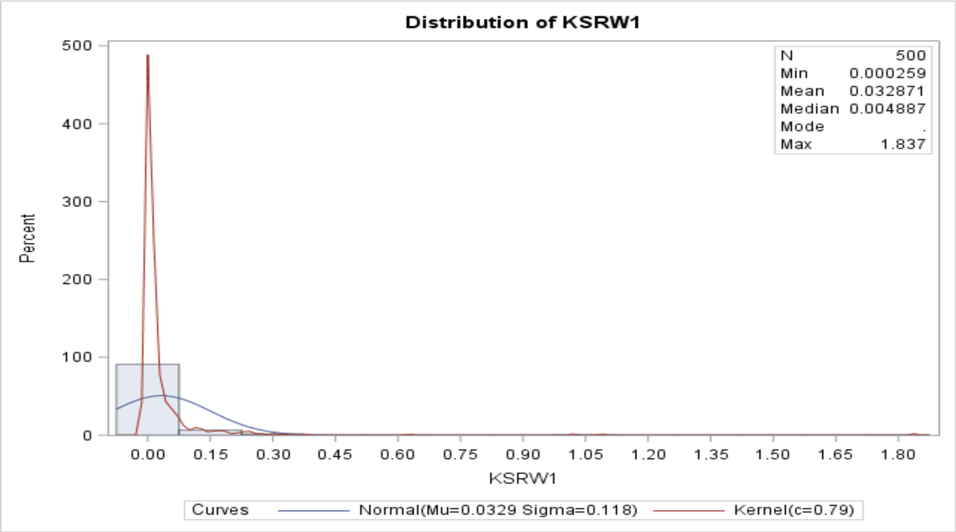
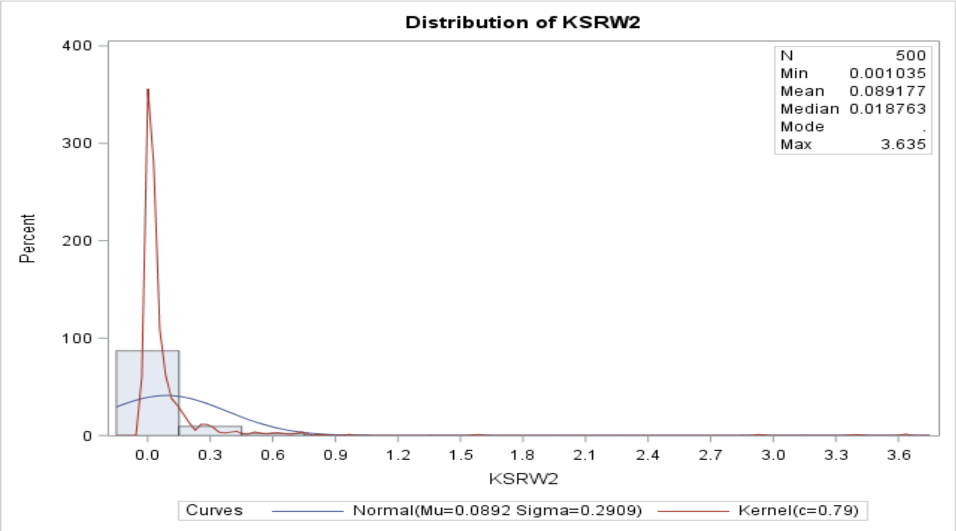


Figure B.10 Distribution of Ridge Parameter for $\rho = 0.99,\ n = 200$.





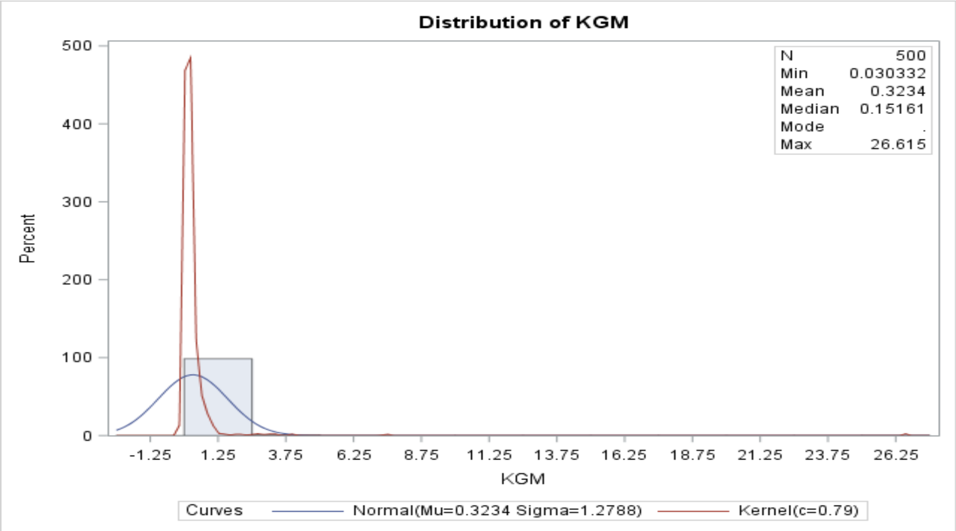
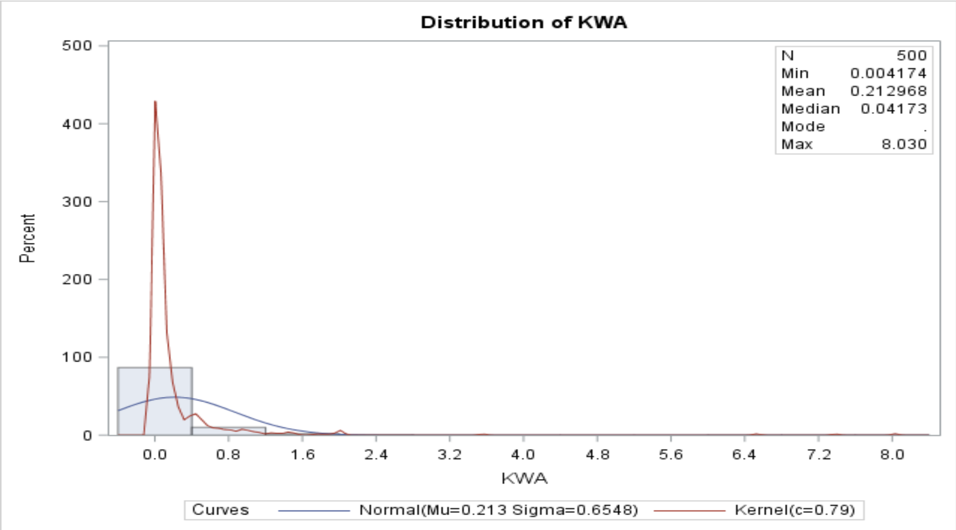


Figure B.10 (Continued)



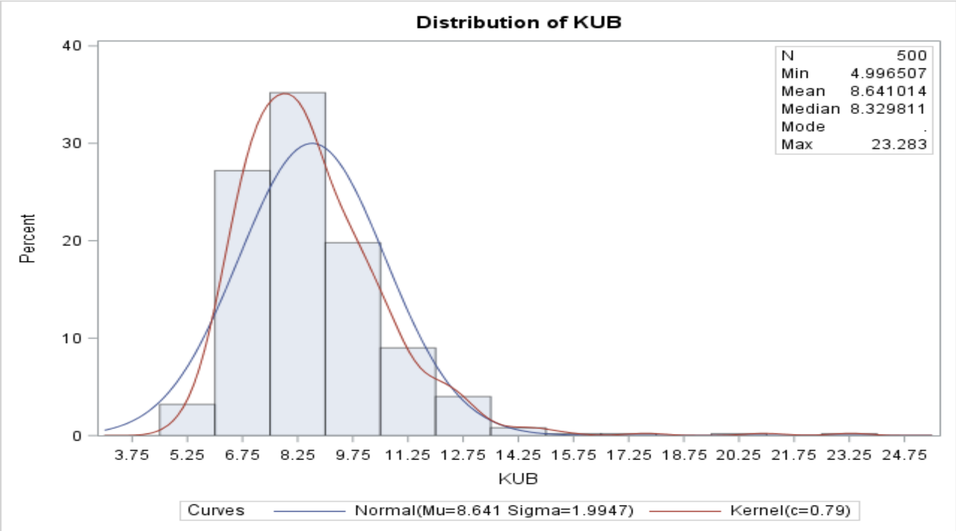


Figure B.10 (Continued)

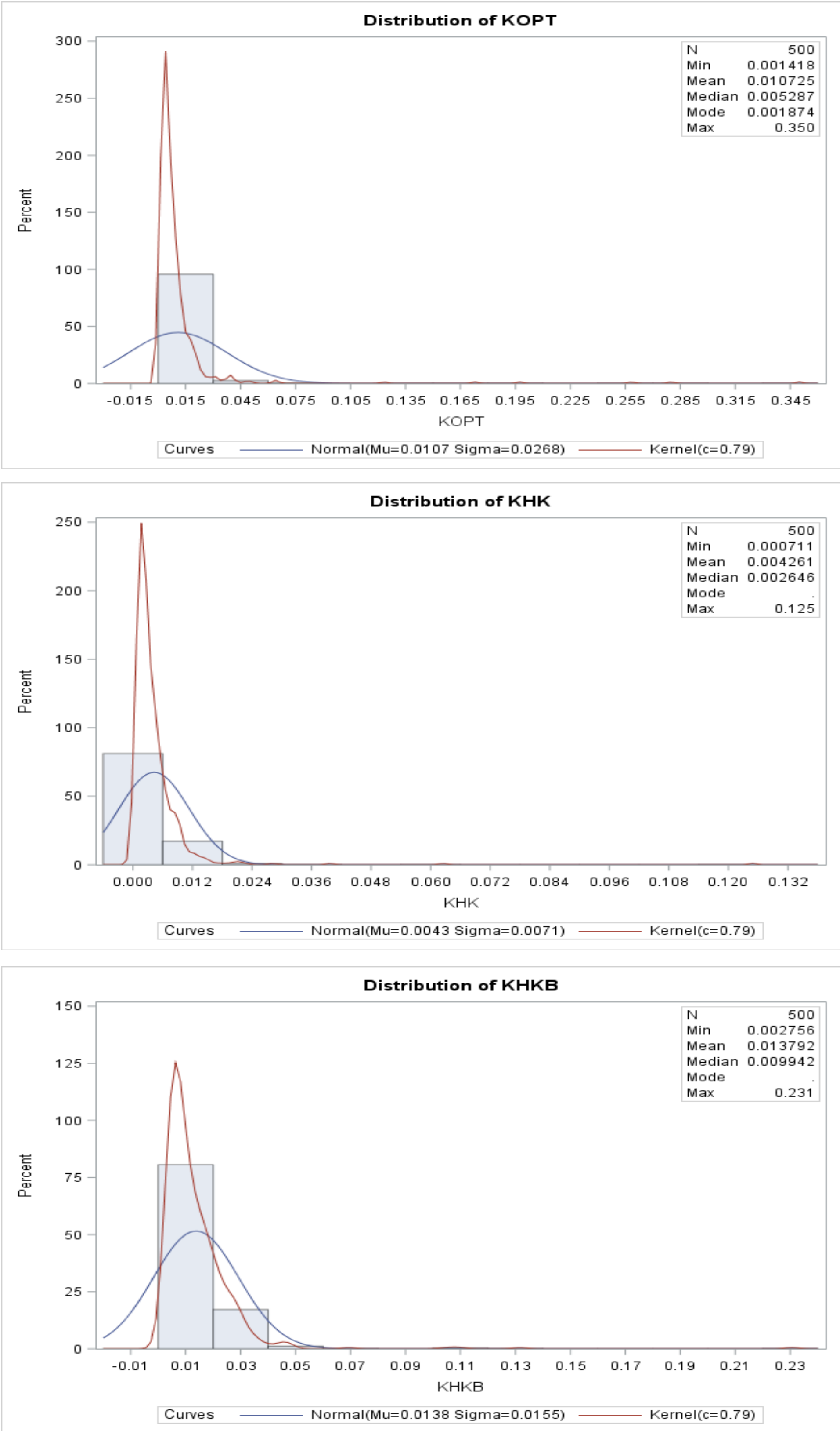
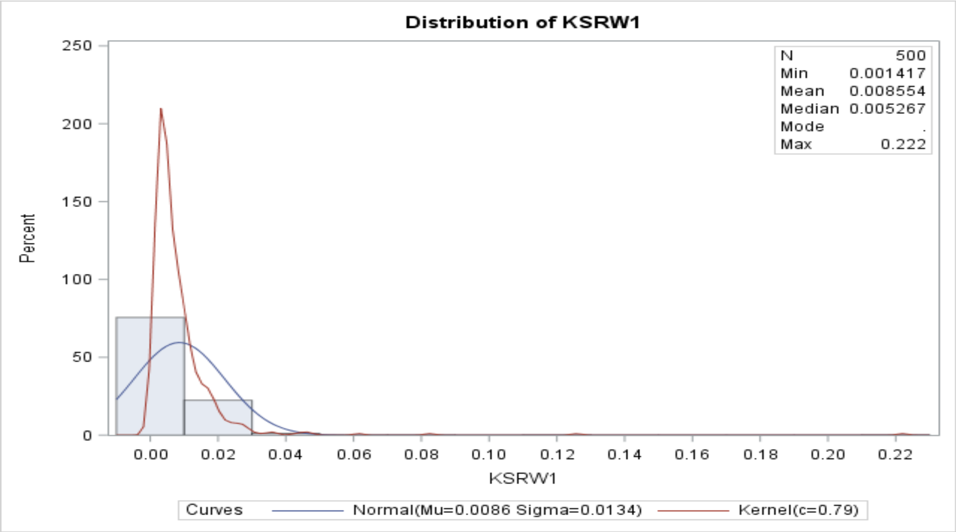
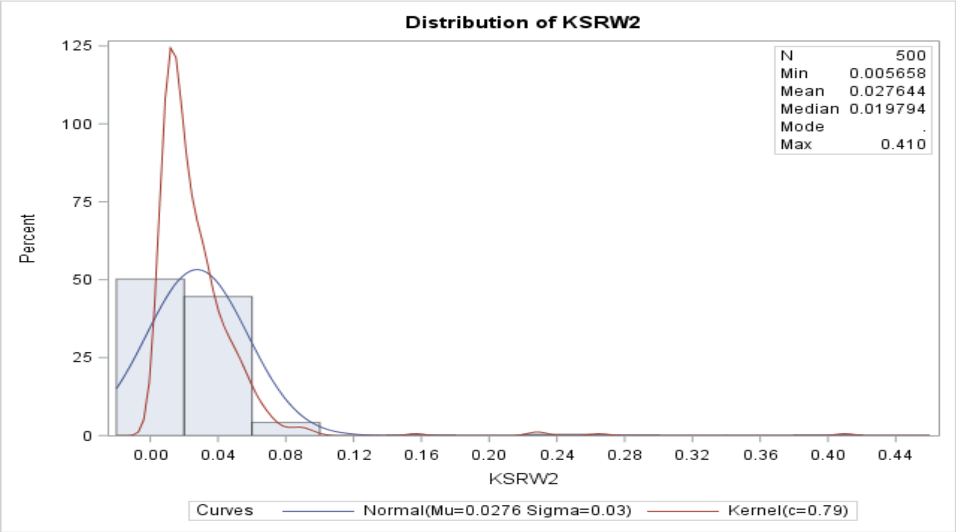


Figure B.11 Distribution of Ridge Parameter for $\rho = 0.99,\ n = 500$.





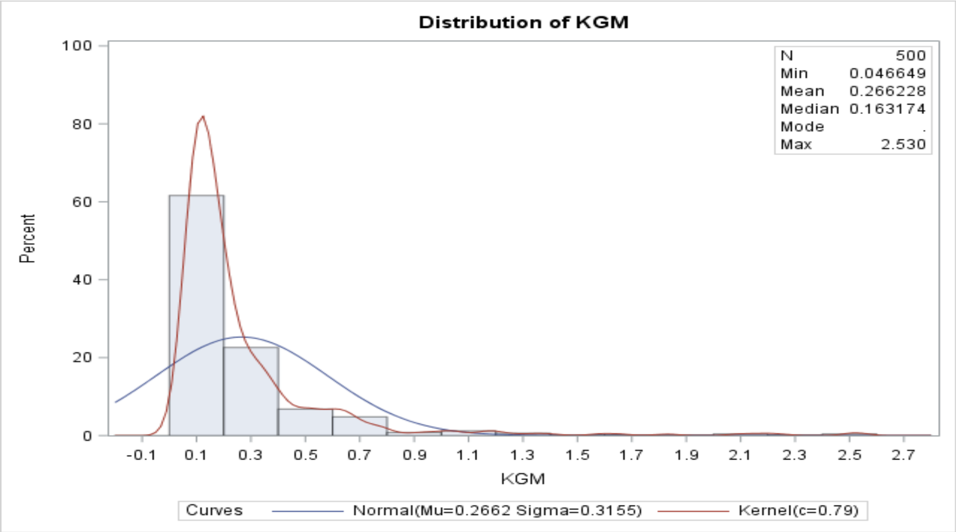
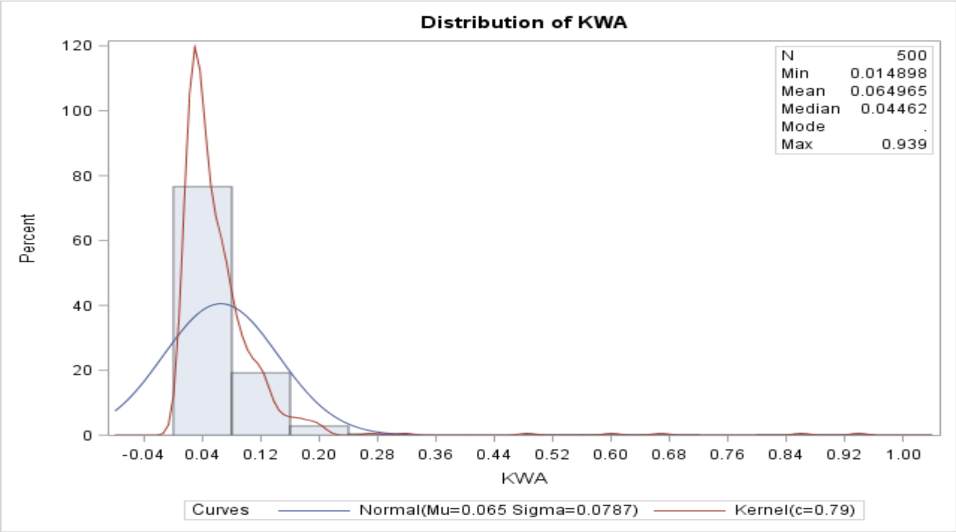


Figure B.11 (Continued)



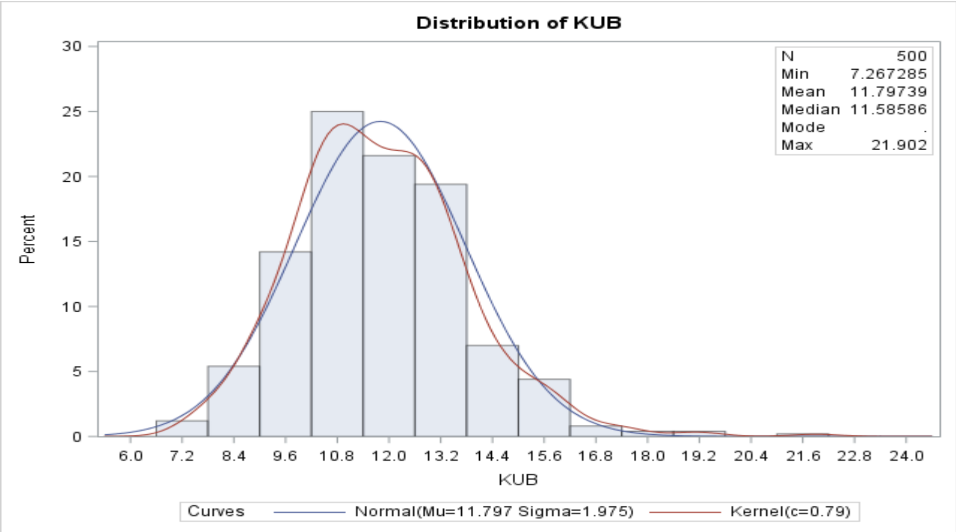


Figure B.11 (Continued)

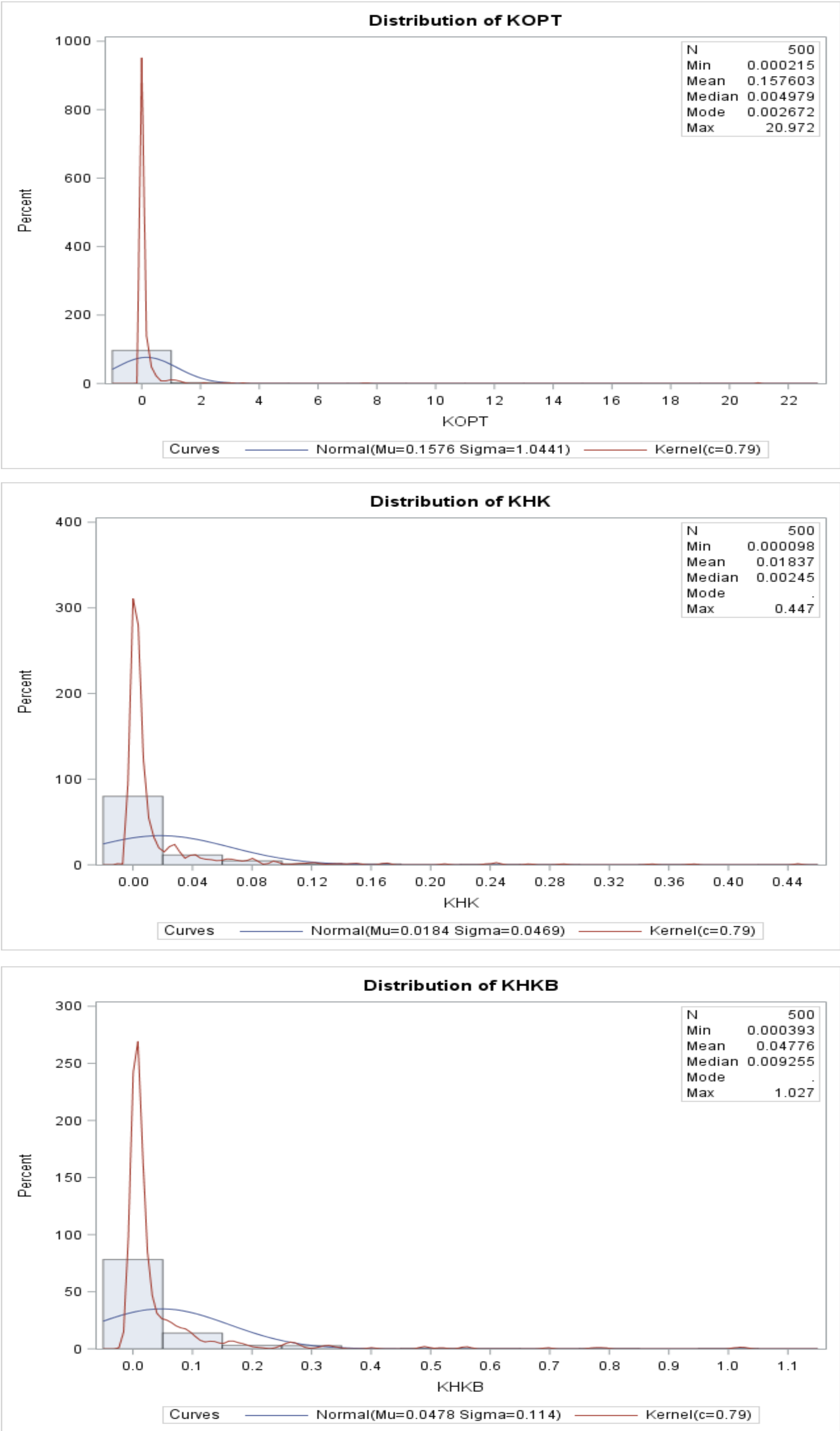
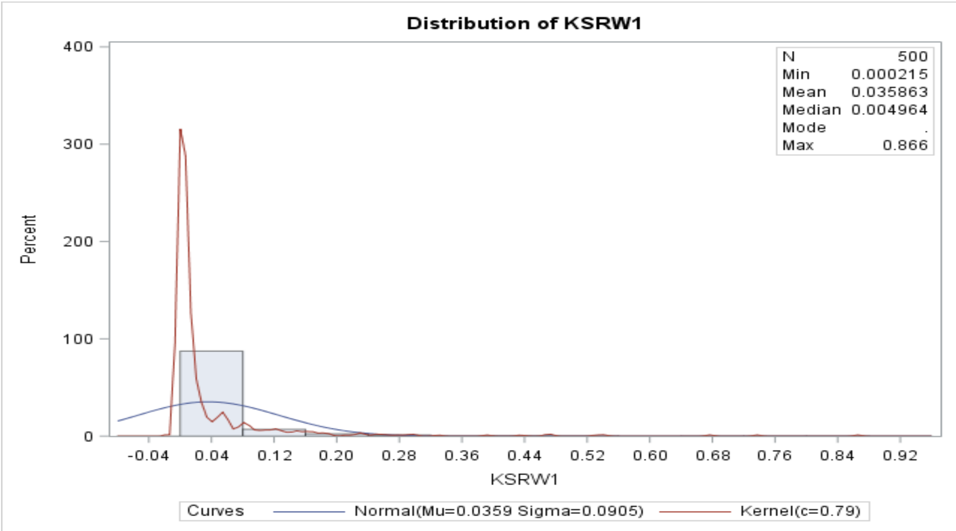
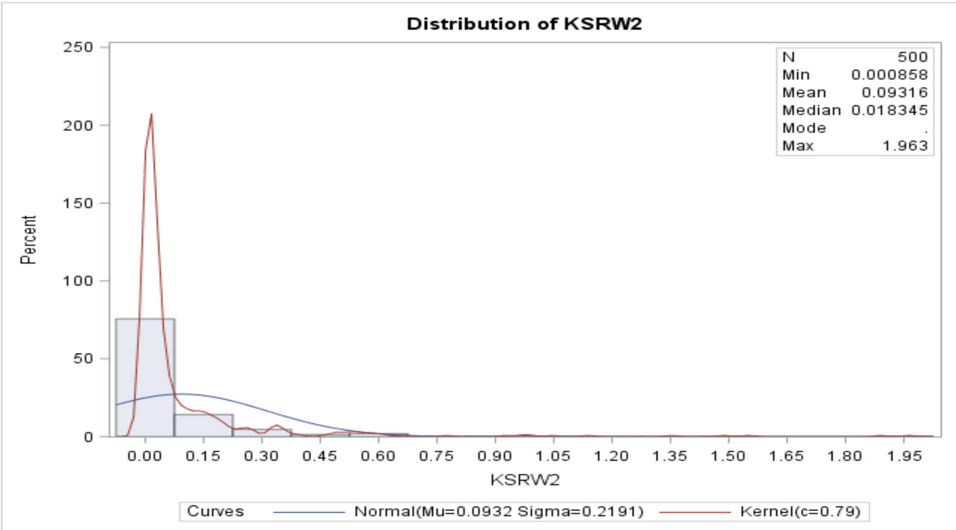


Figure B.12 Distribution of Ridge Parameter for $\rho = 0.99$, $n = 1000$.





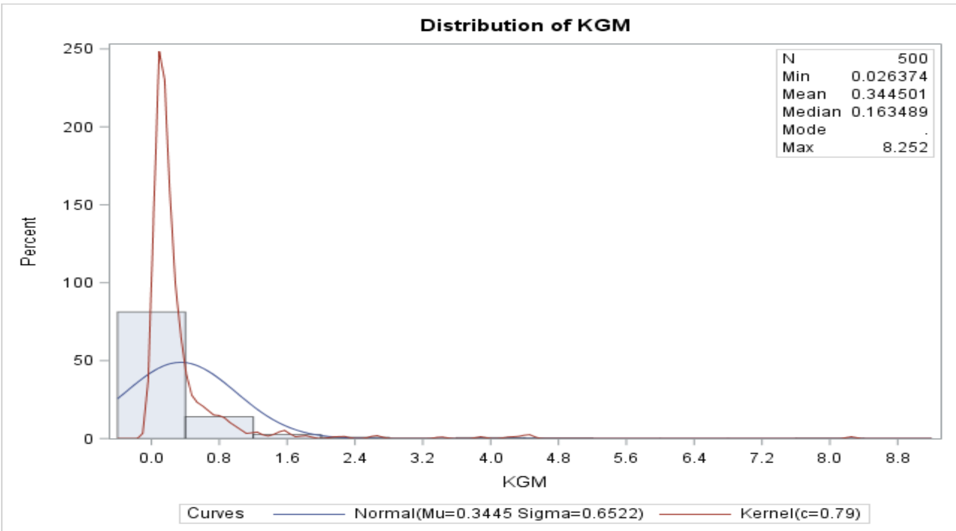
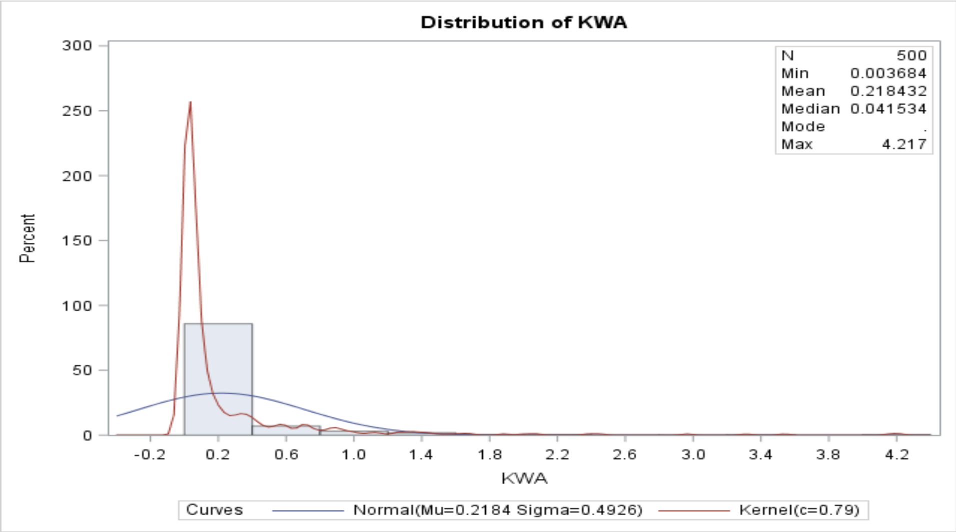


Figure B.12 (Continued)



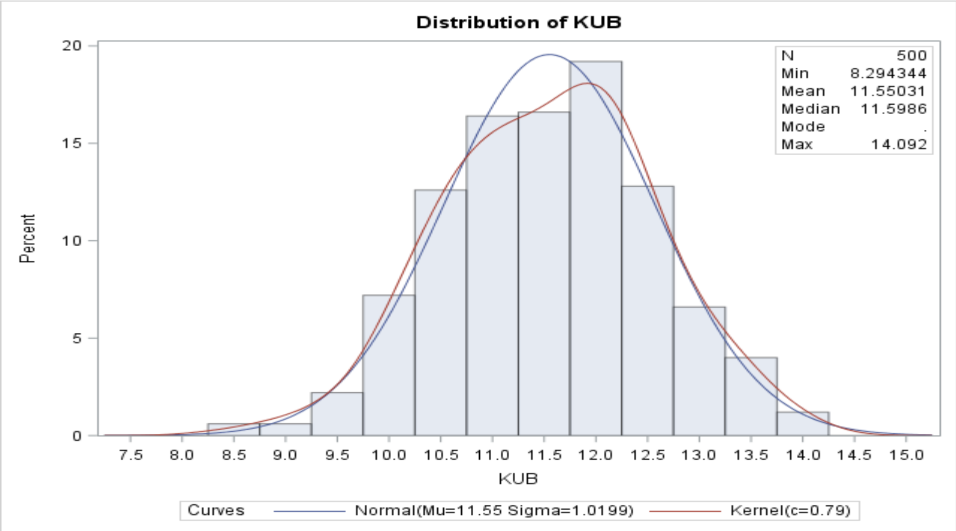


Figure B.12 (Continued)

Appendix C

Distribution of Ridge Parameter in Case of Having Five Explanatory Variables

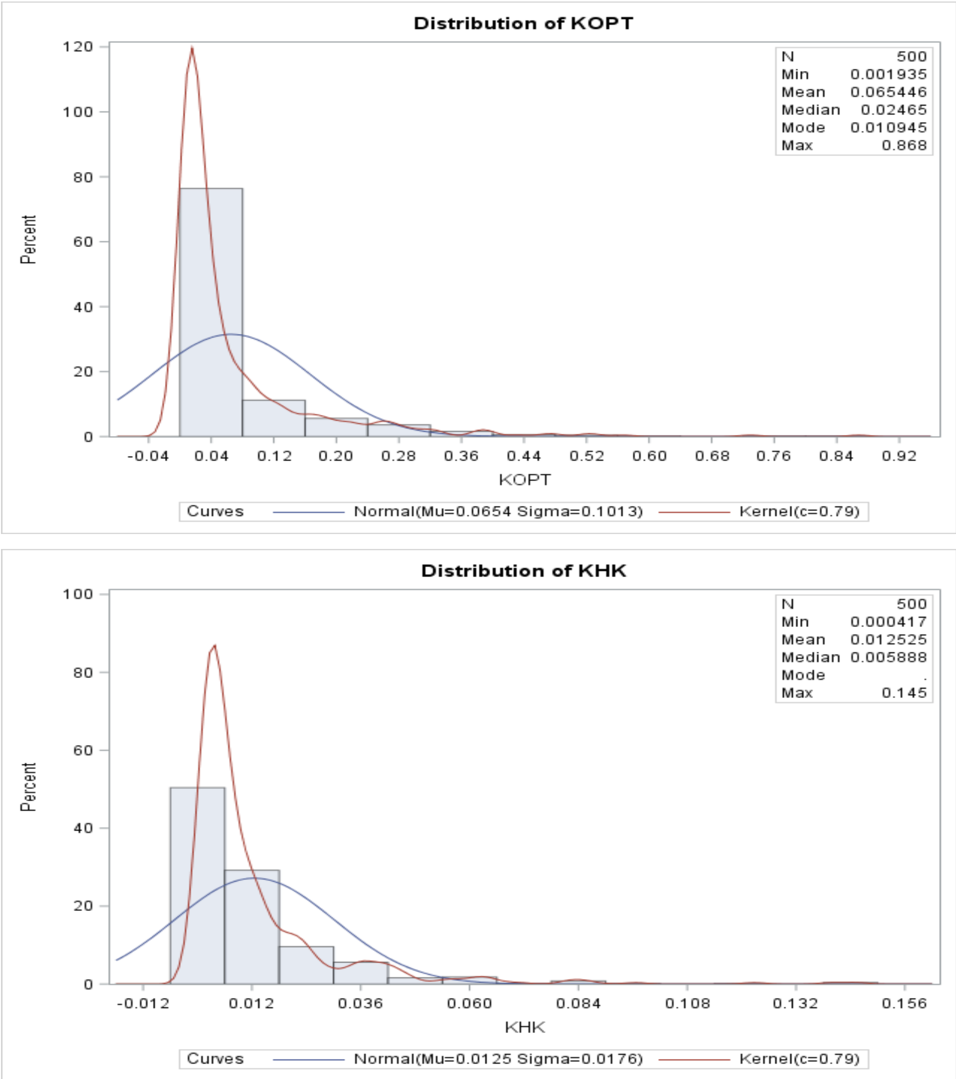


Figure C.1 Distribution of Ridge Parameter for $\rho_{12} = 0.90,\ \rho_{34} = 0.90,\ n = 100$.

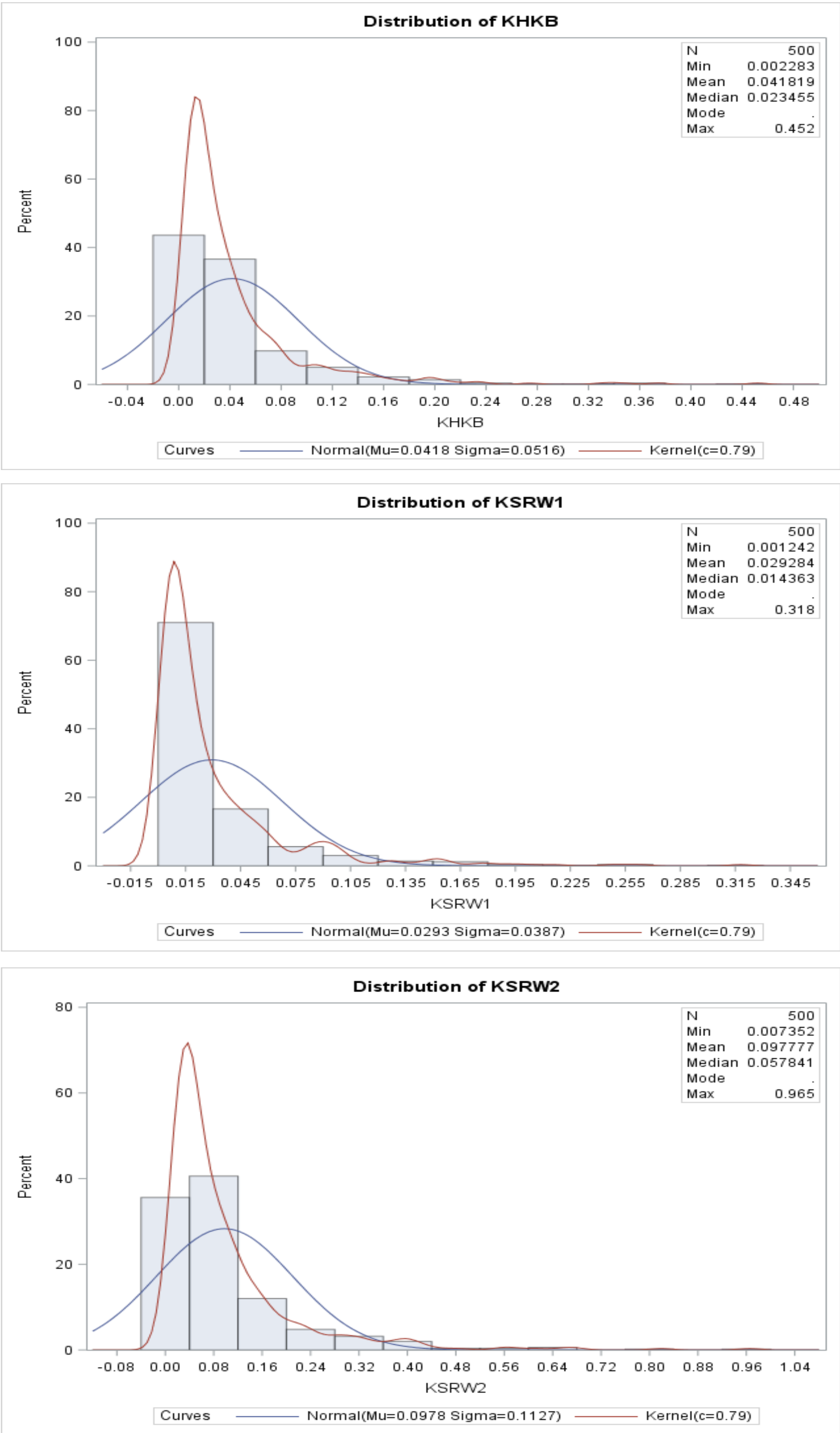
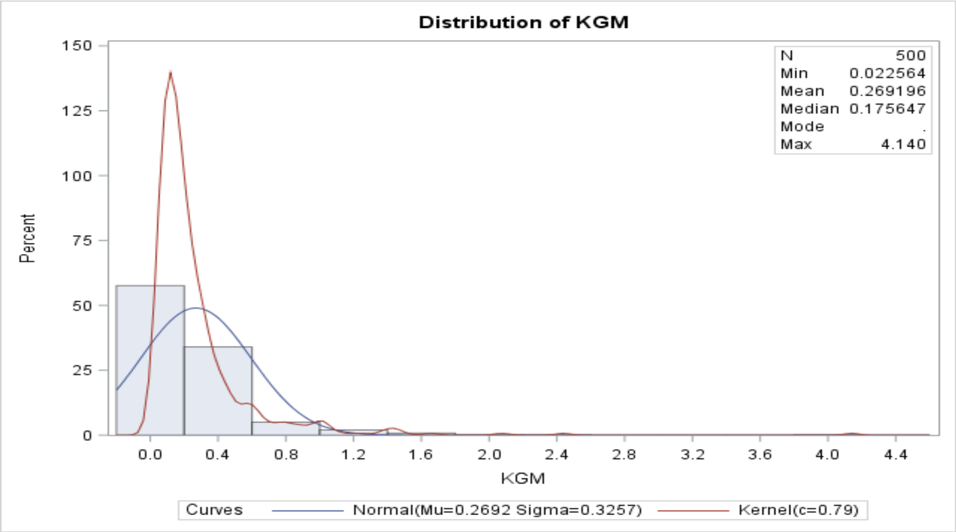
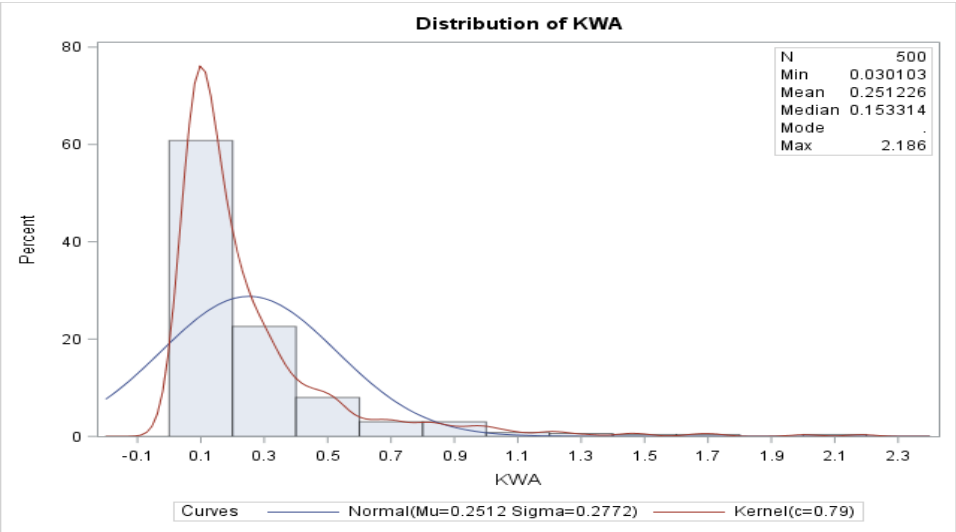


Figure C.1 (Continued)





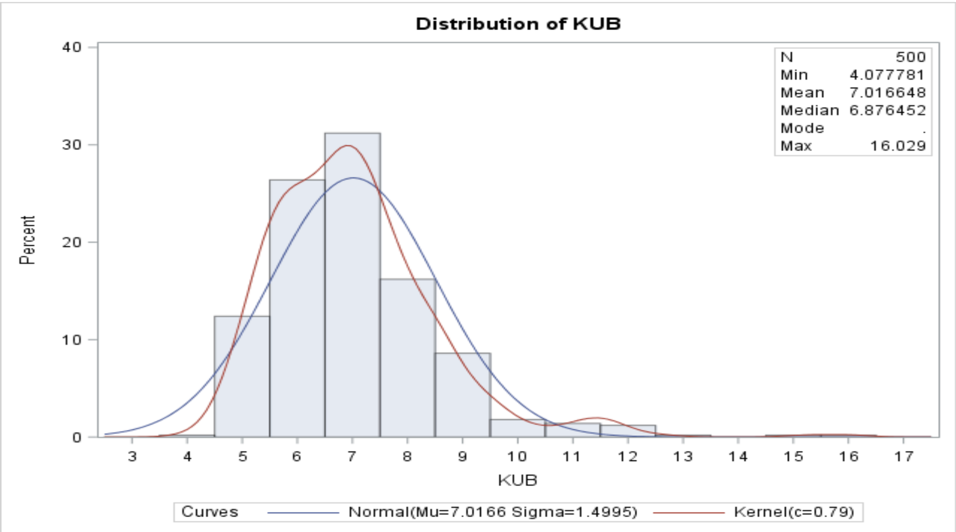


Figure C.1 (Continued)

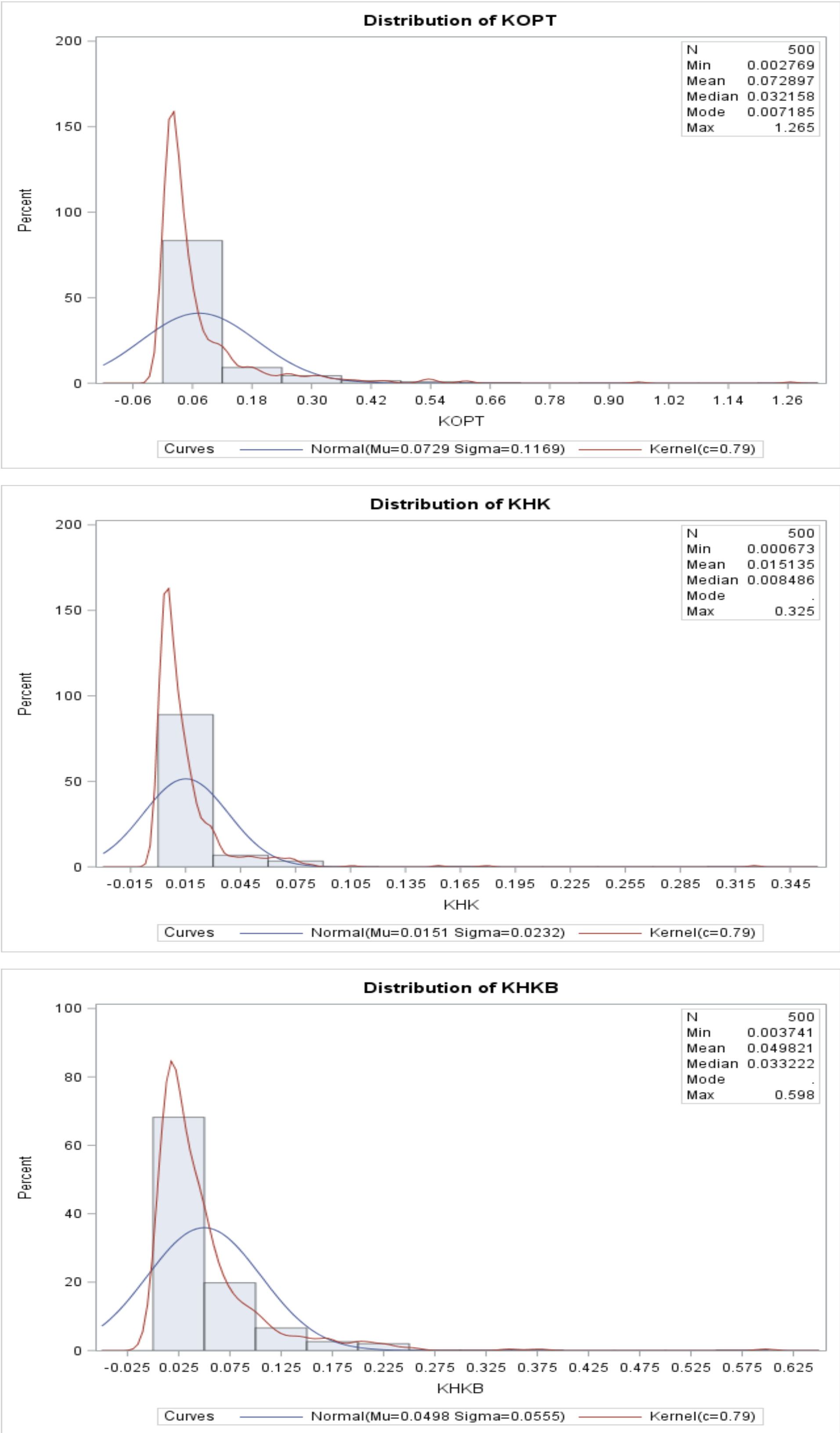


Figure C.2 Distribution of Ridge Parameter for $\rho_{12} = 0.90,\ \rho_{34} = 0.90,\ n = 200$.

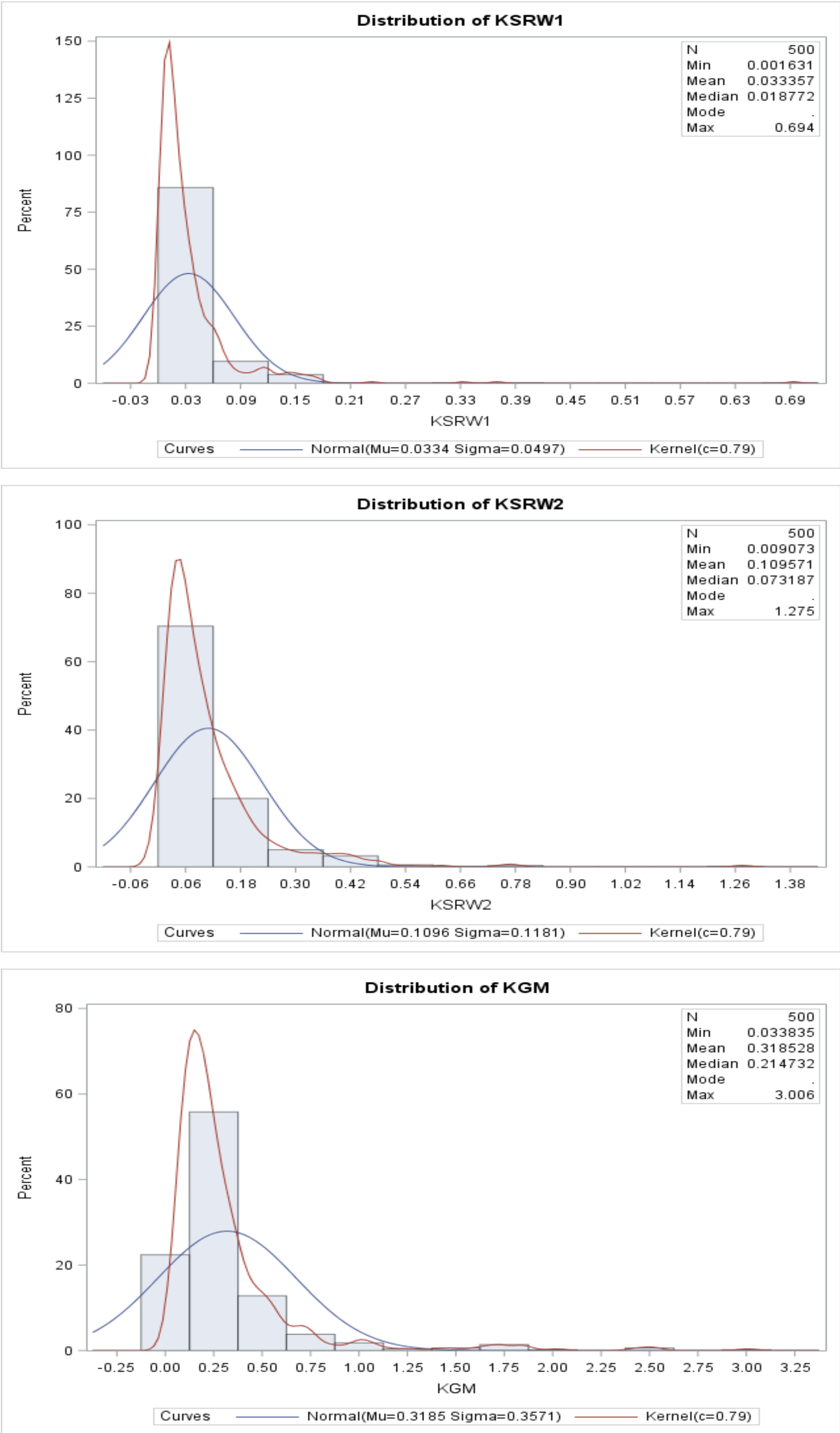
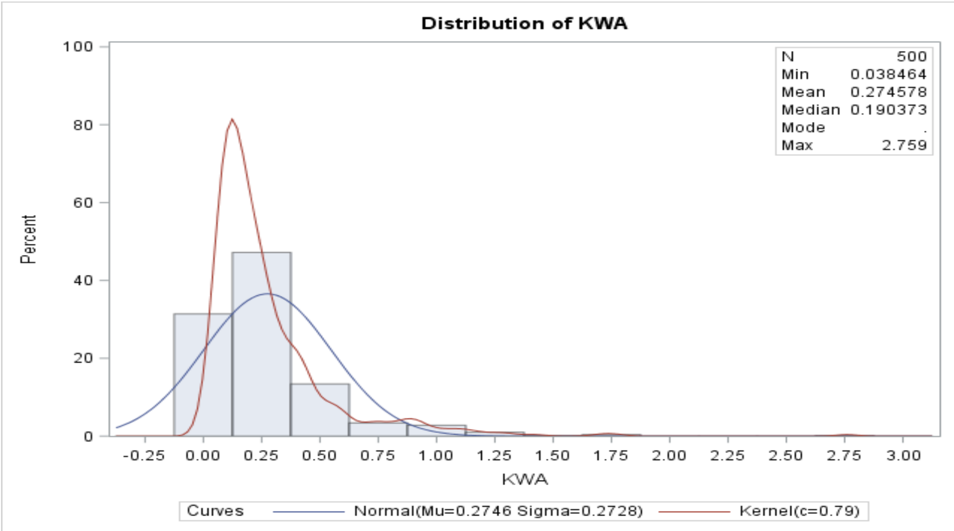


Figure C.2 (Continued)



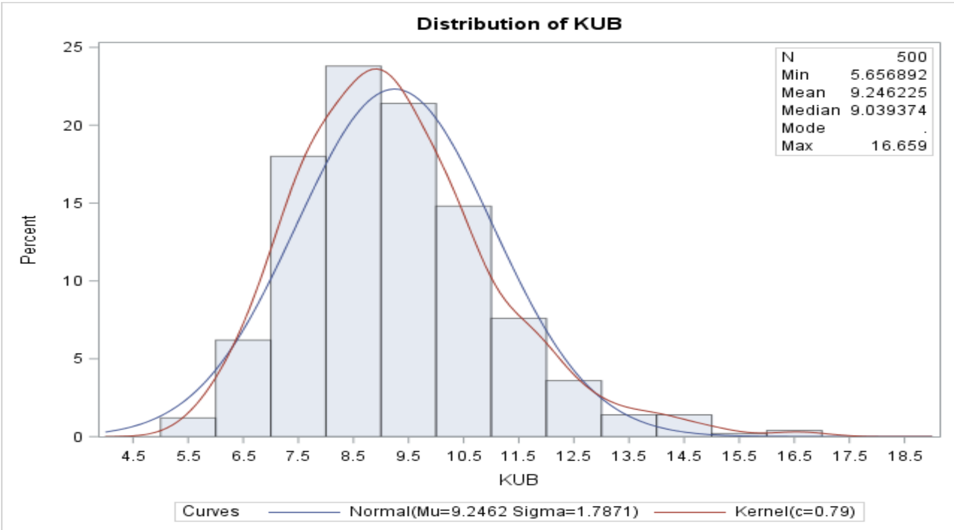


Figure C.2 (Continued)

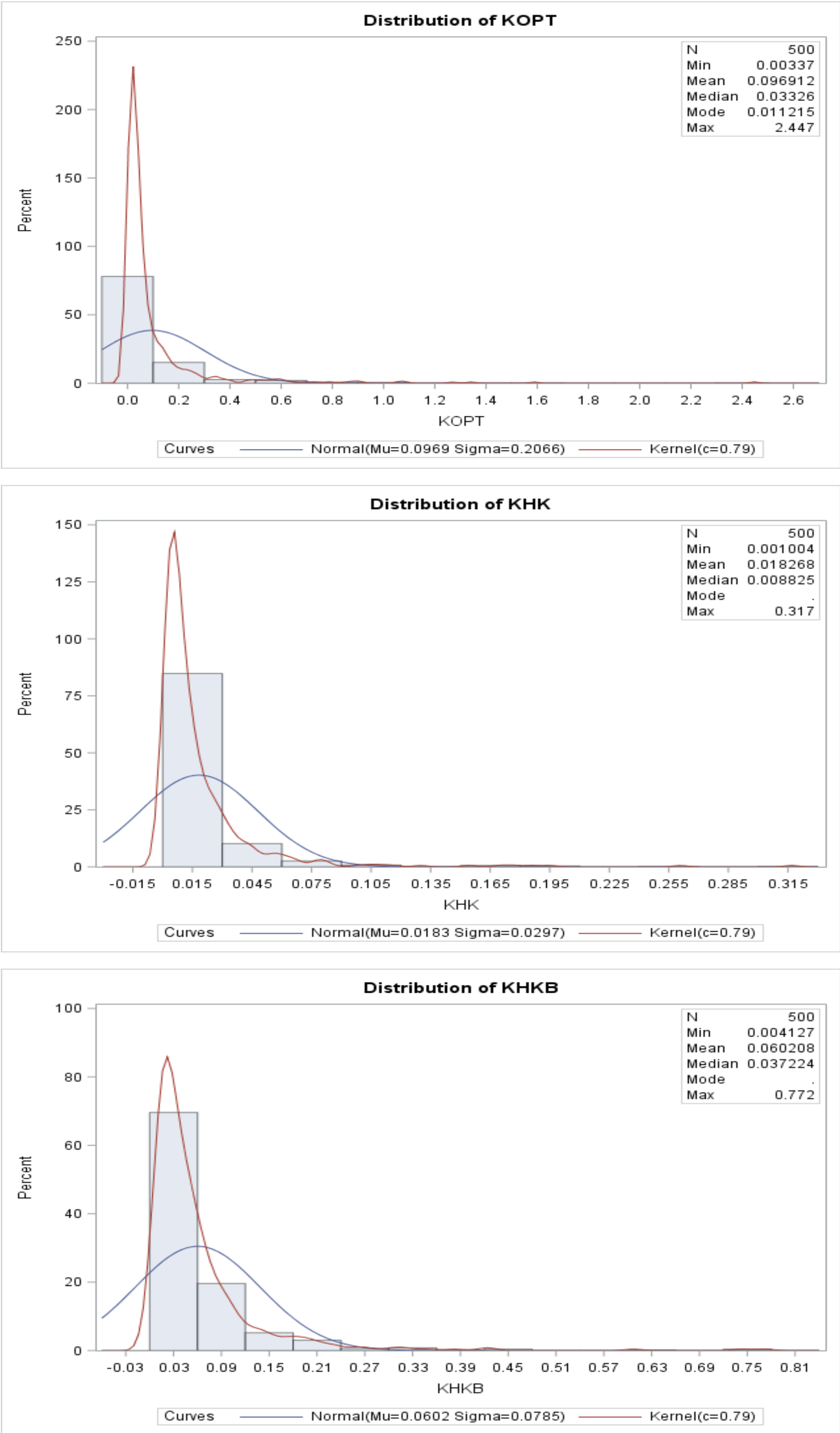
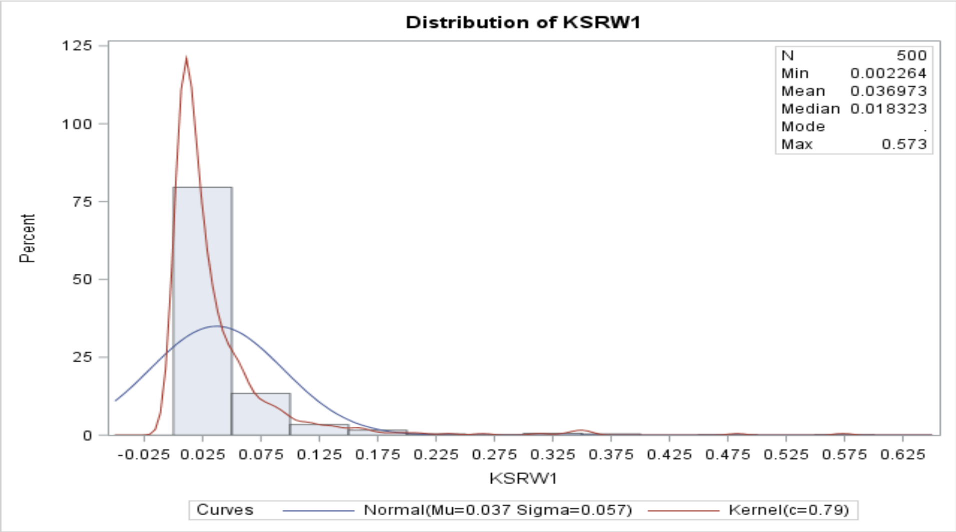
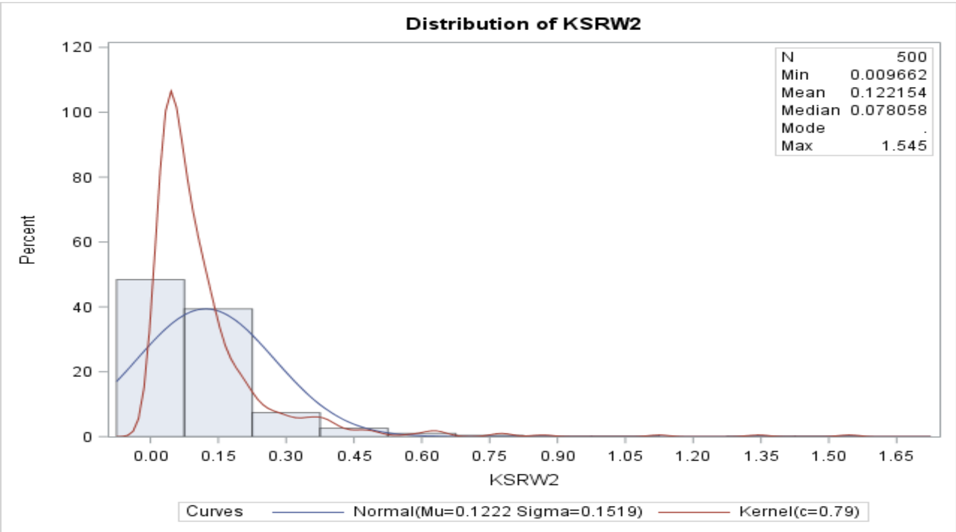


Figure C.3 Distribution of Ridge Parameter for $\rho_{12} = 0.90,\ \rho_{34} = 0.90,\ n = 500$.





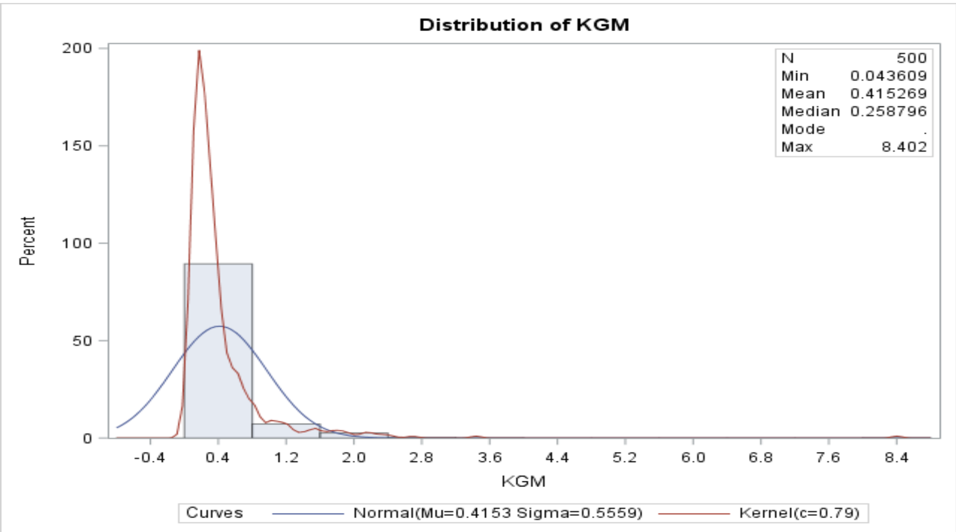
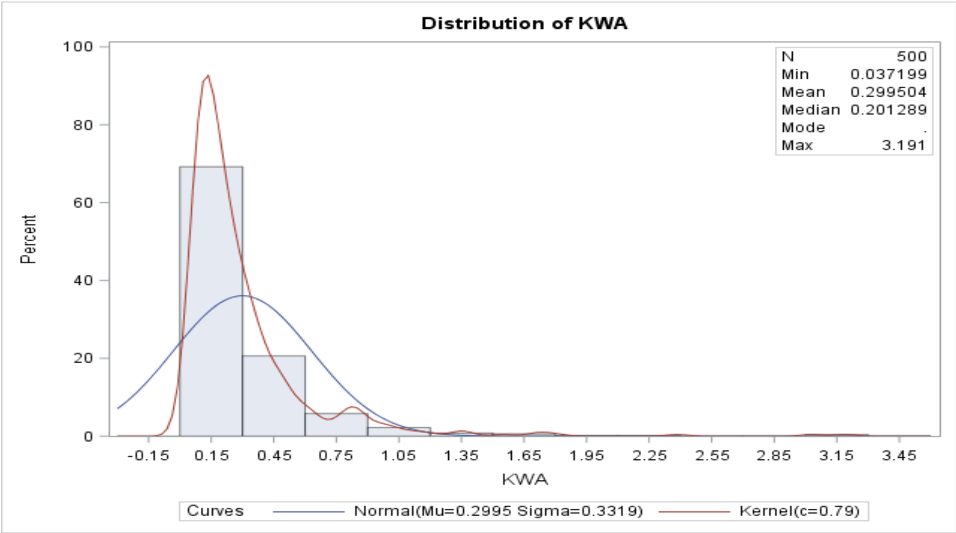


Figure C.3 (Continued)



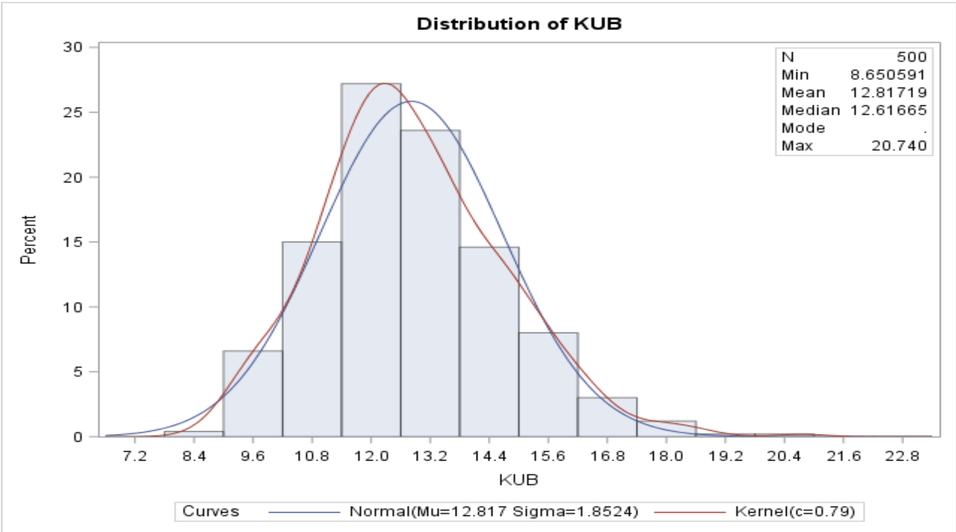


Figure C.3 (Continued)

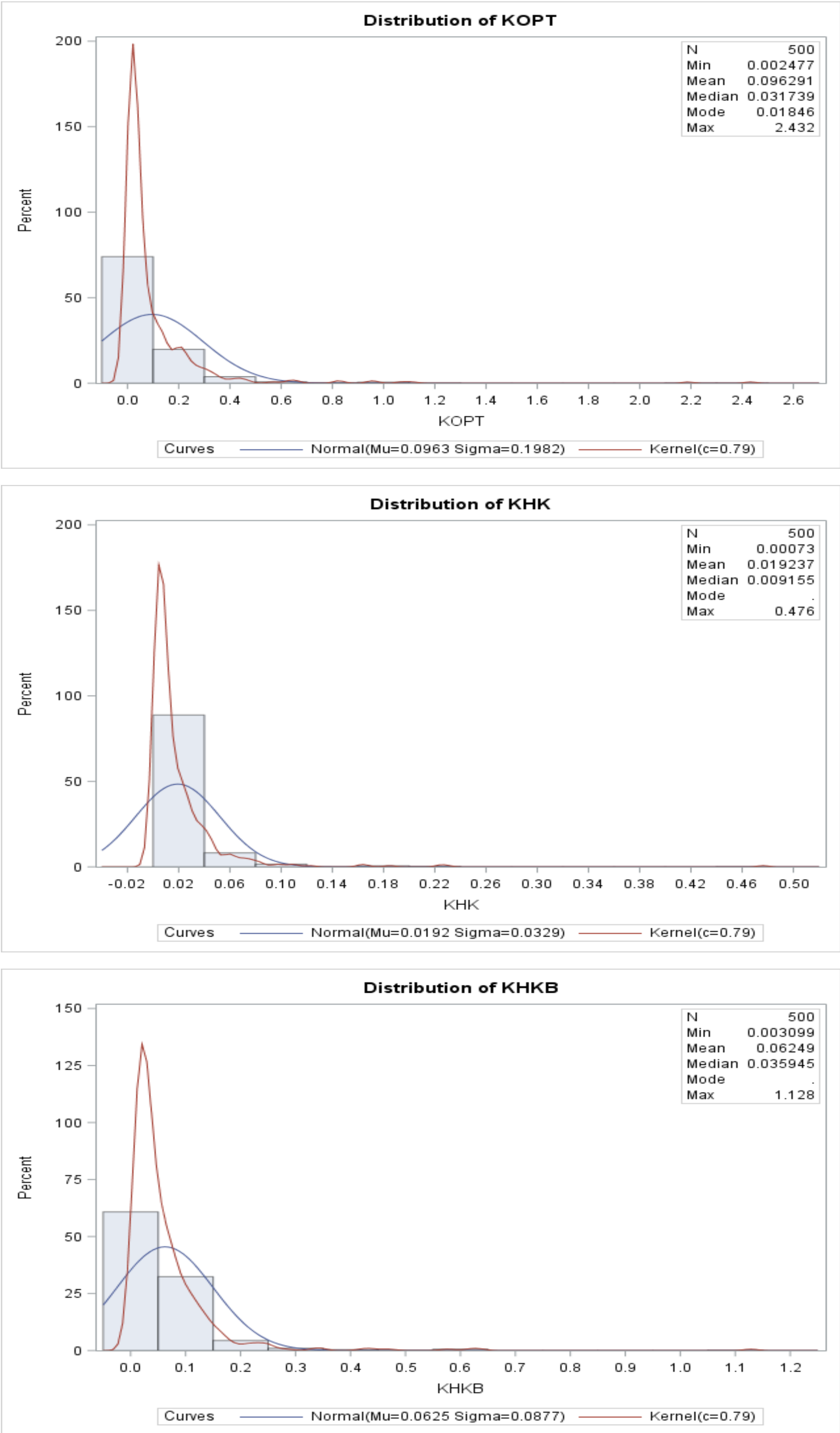
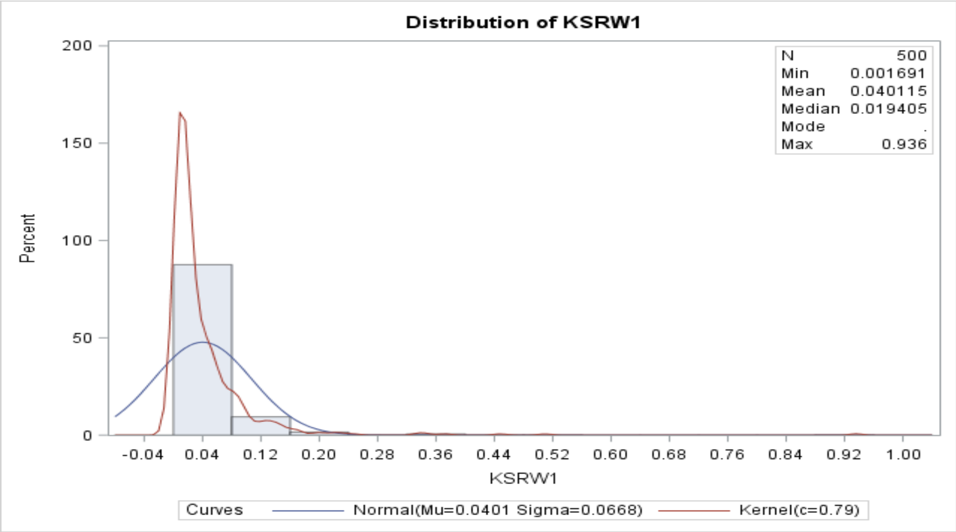
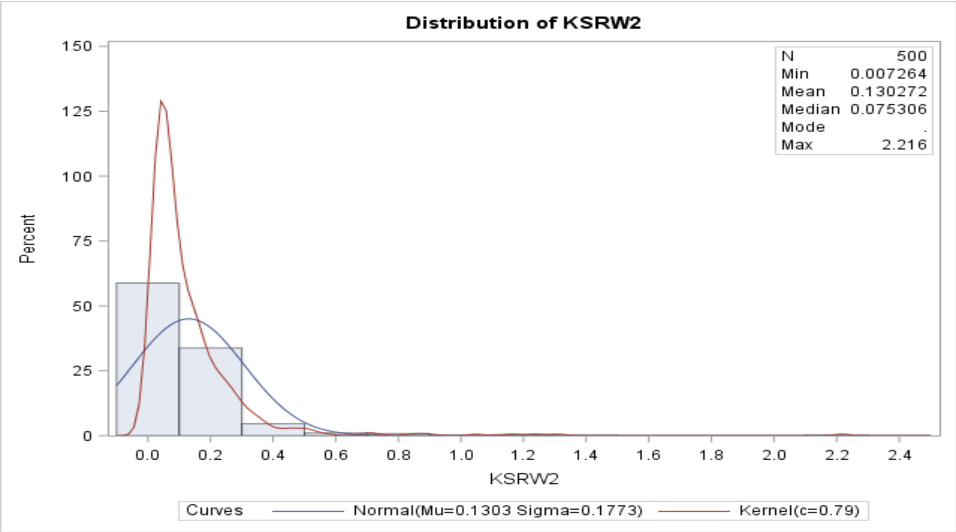


Figure C.4 Distribution of Ridge Parameter for $\rho_{12} = 0.90,\ \rho_{34} = 0.90,\ n = 1000$.





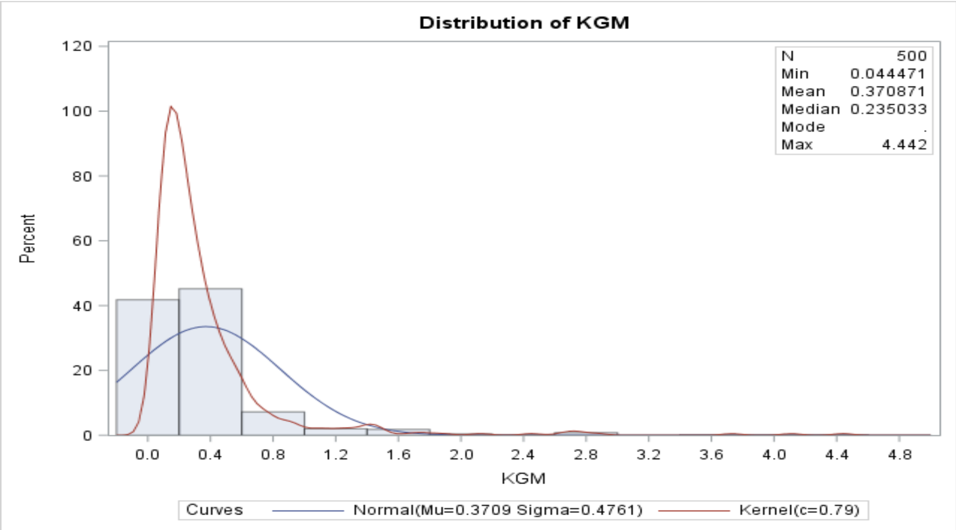
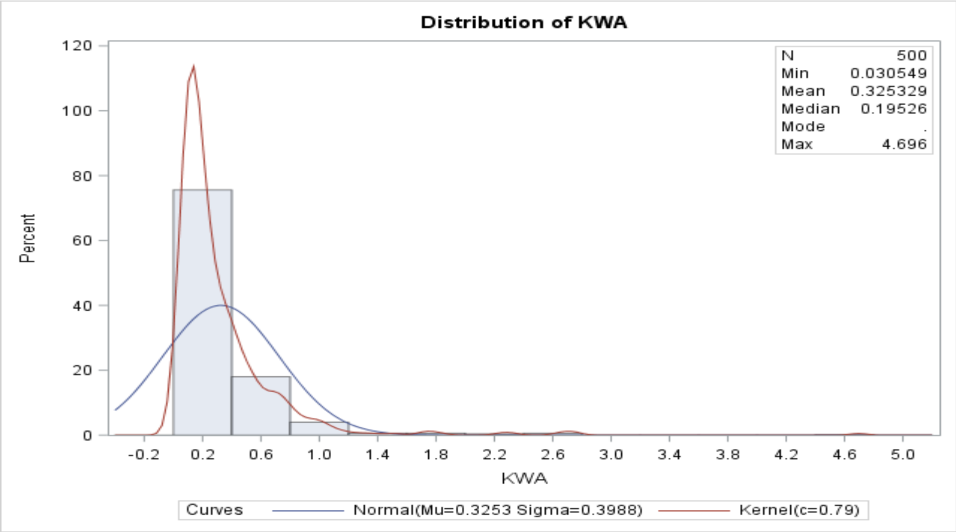


Figure C.4 (Continued)



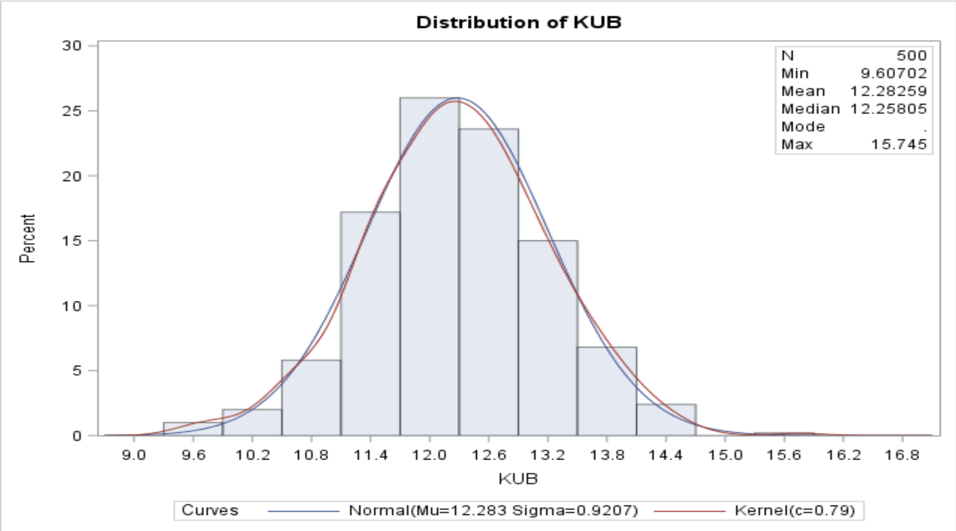


Figure C.4 (Continued)

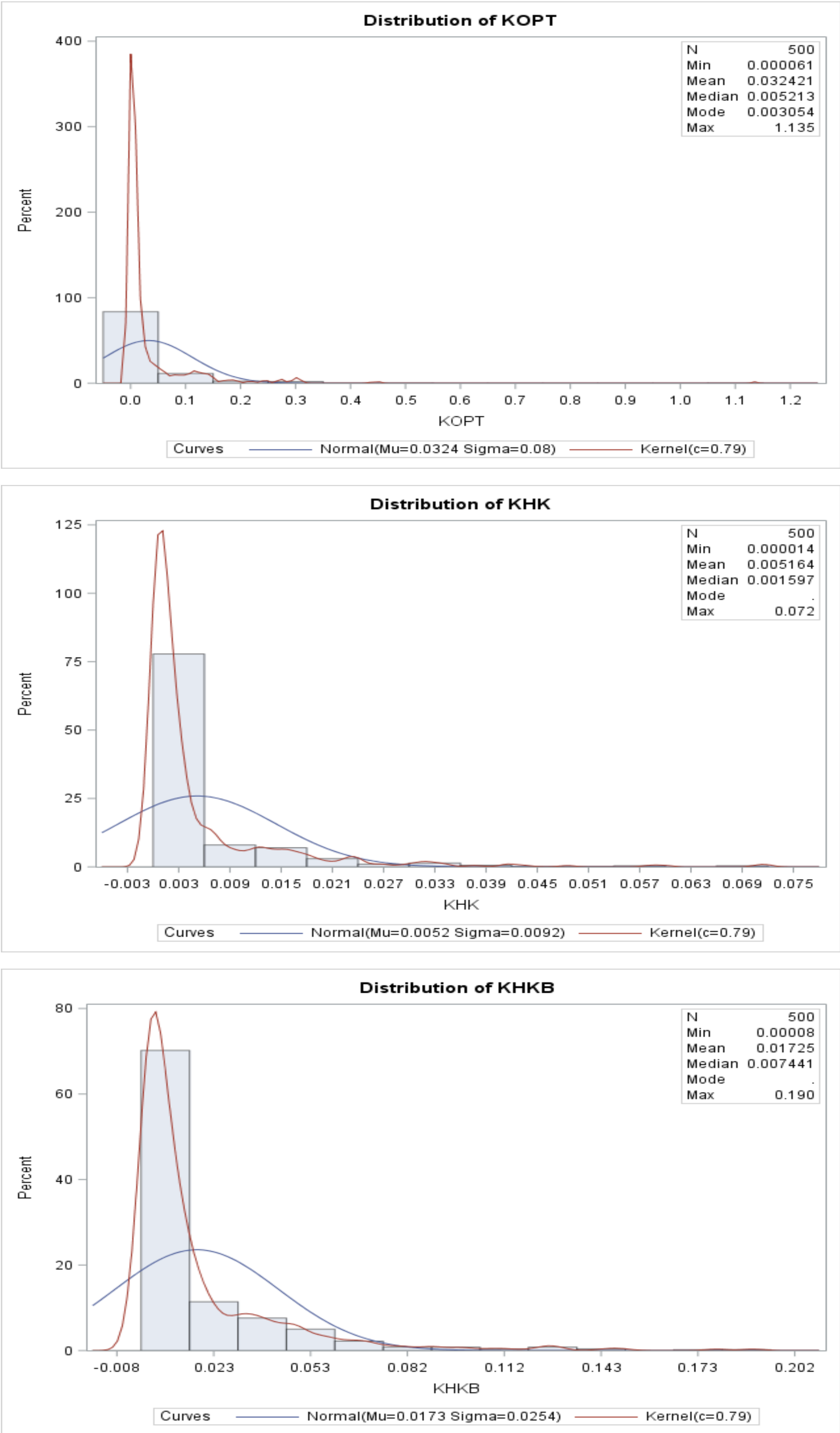


Figure C.5 Distribution of Ridge Parameter for $\rho_{12} = 0.99$, $\rho_{34} = 0.90$, $n = 100$.

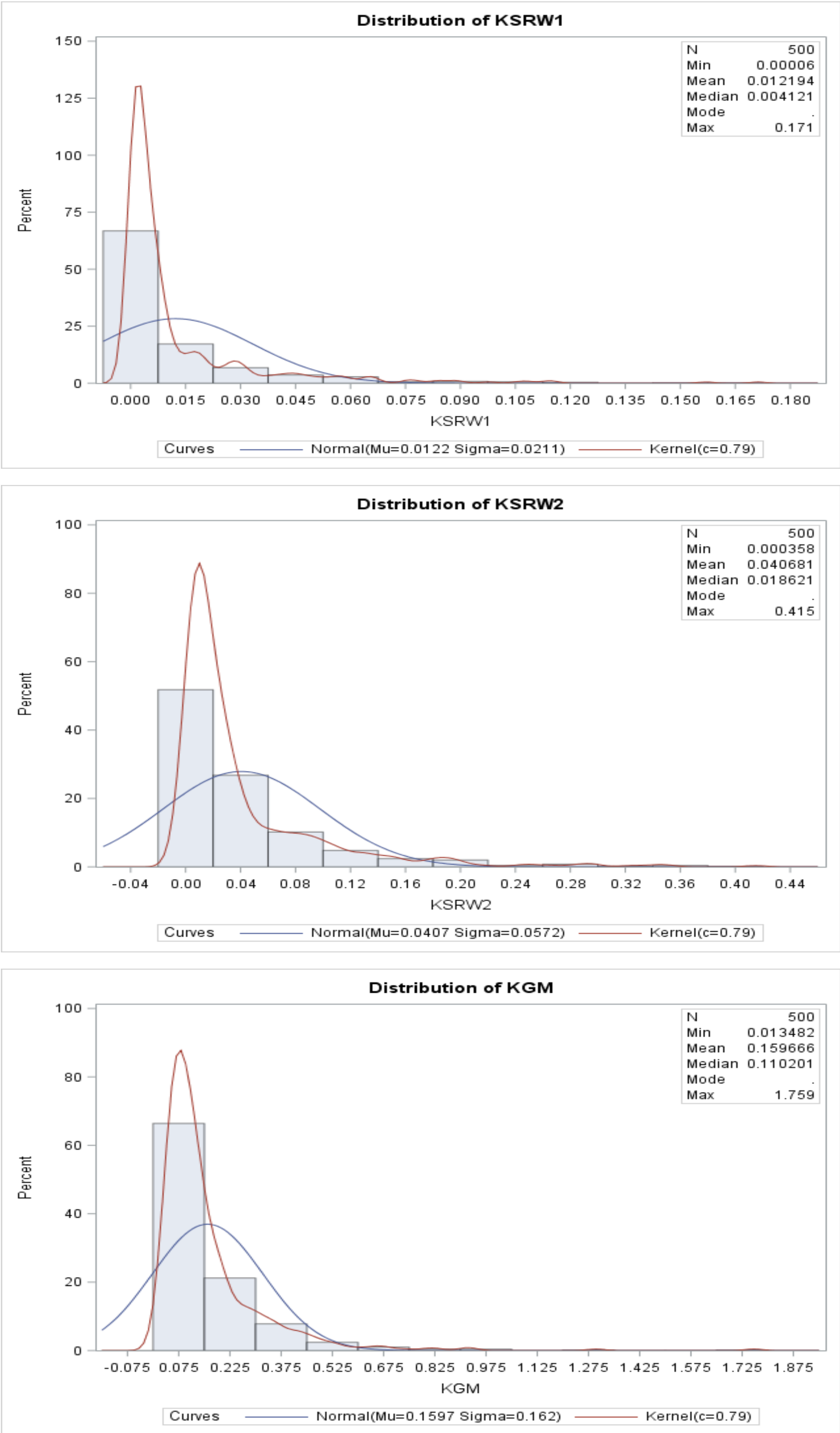
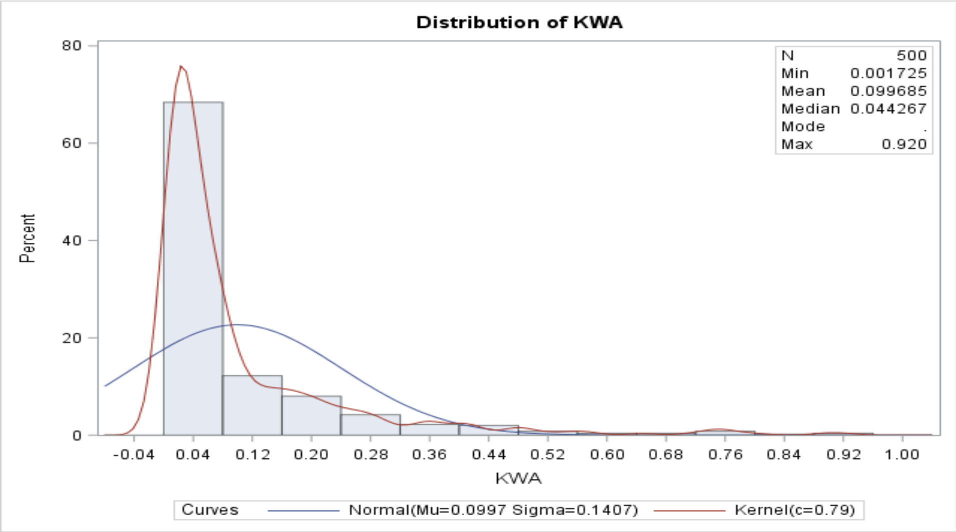


Figure C.5 (Continued)



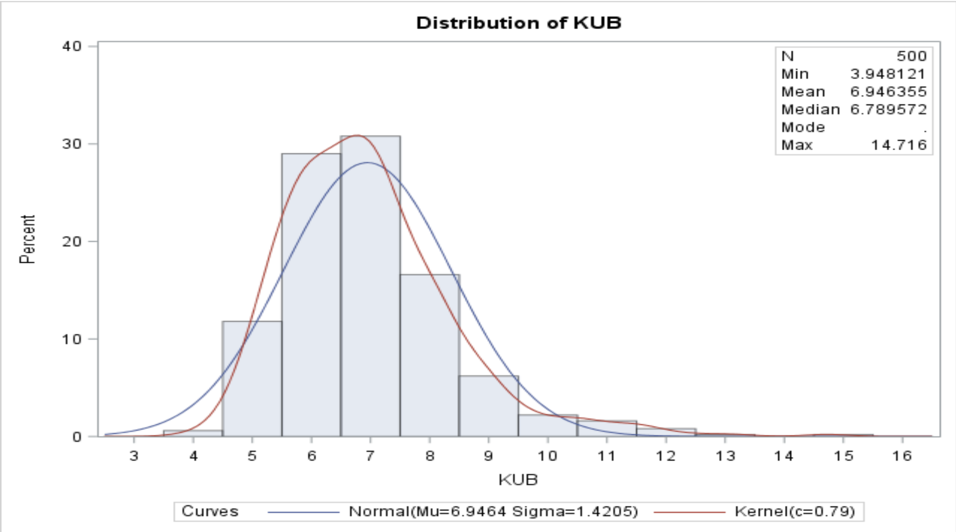
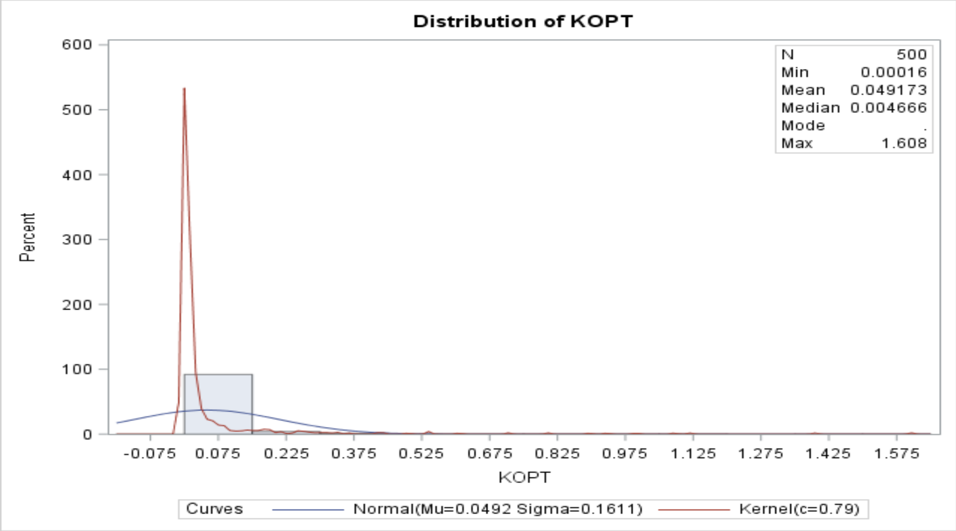
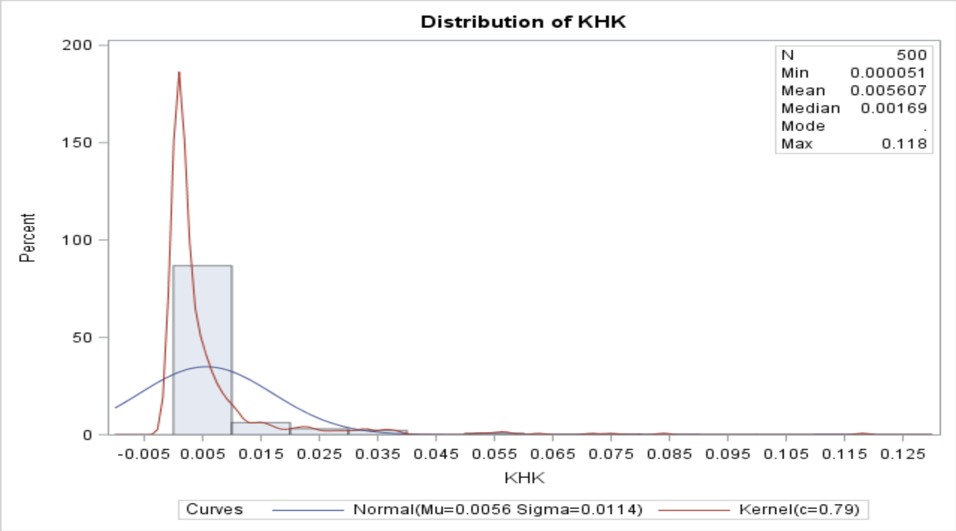


Figure C.5 (Continued)





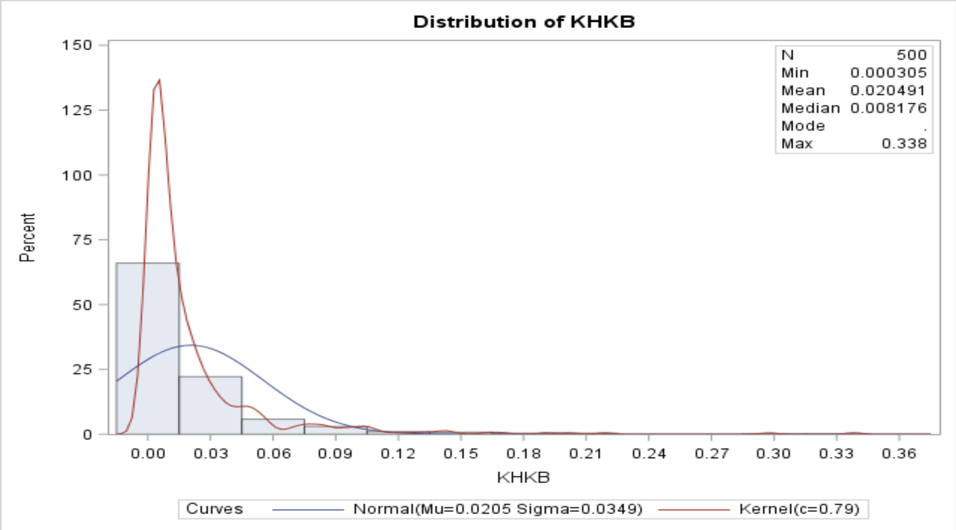


Figure C.6 Distribution of Ridge Parameter for $\rho_{12} = 0.99,\ \rho_{34} = 0.90,\ n = 200$.

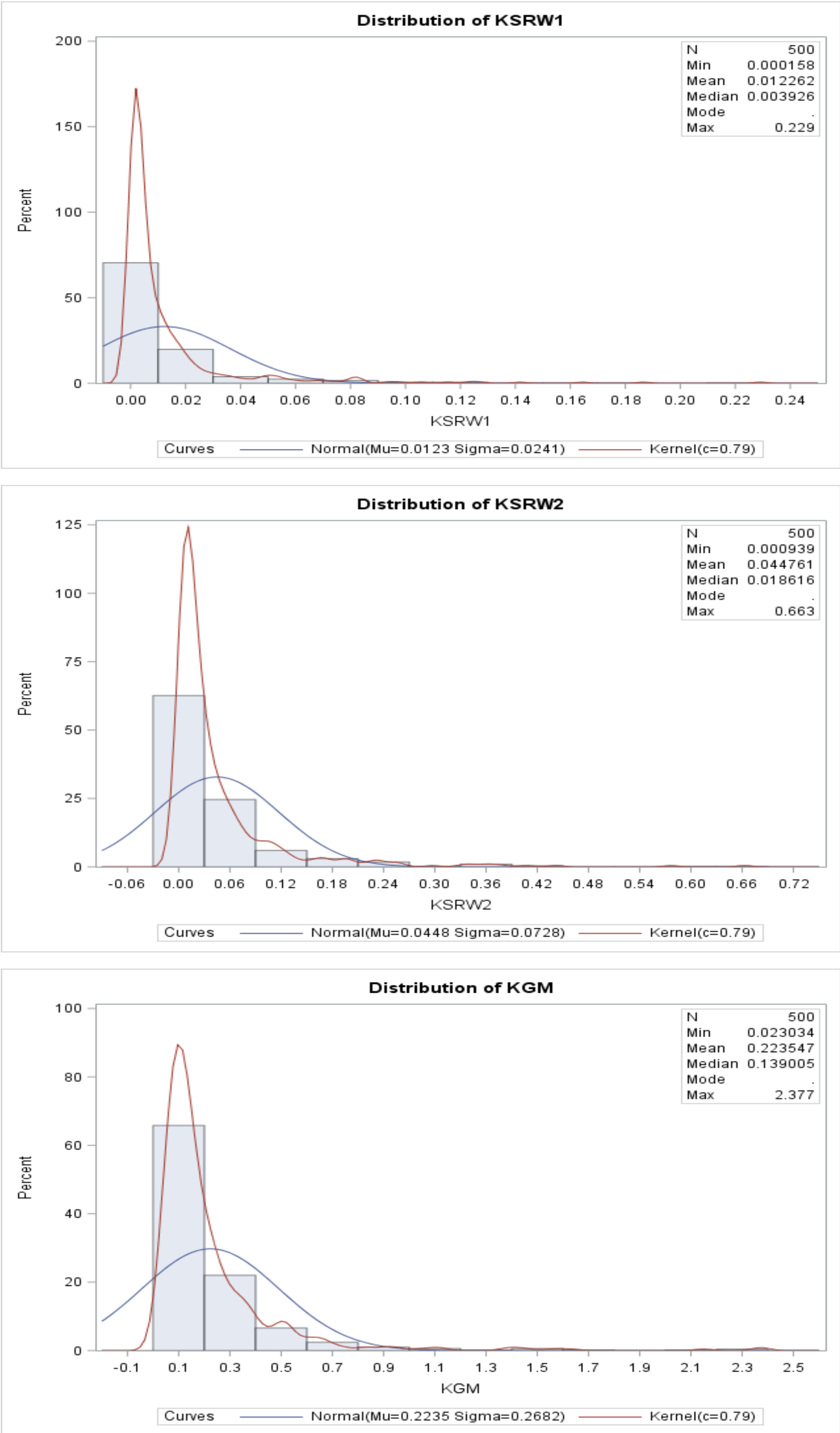
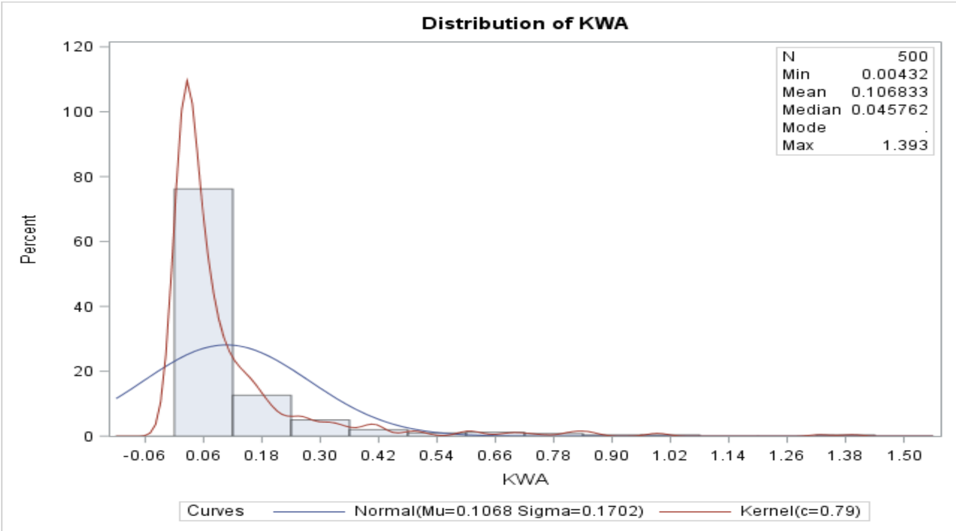


Figure C.6 (Continued)



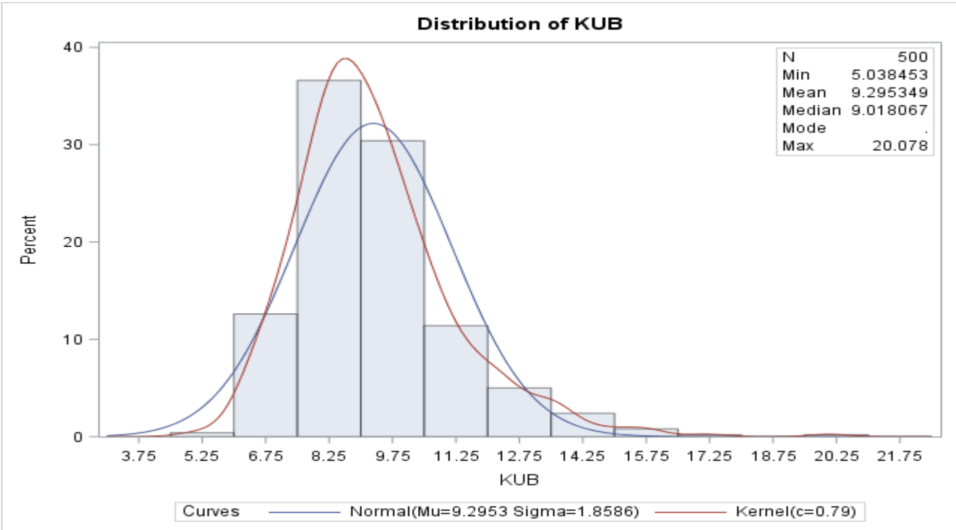


Figure C.6 (Continued)

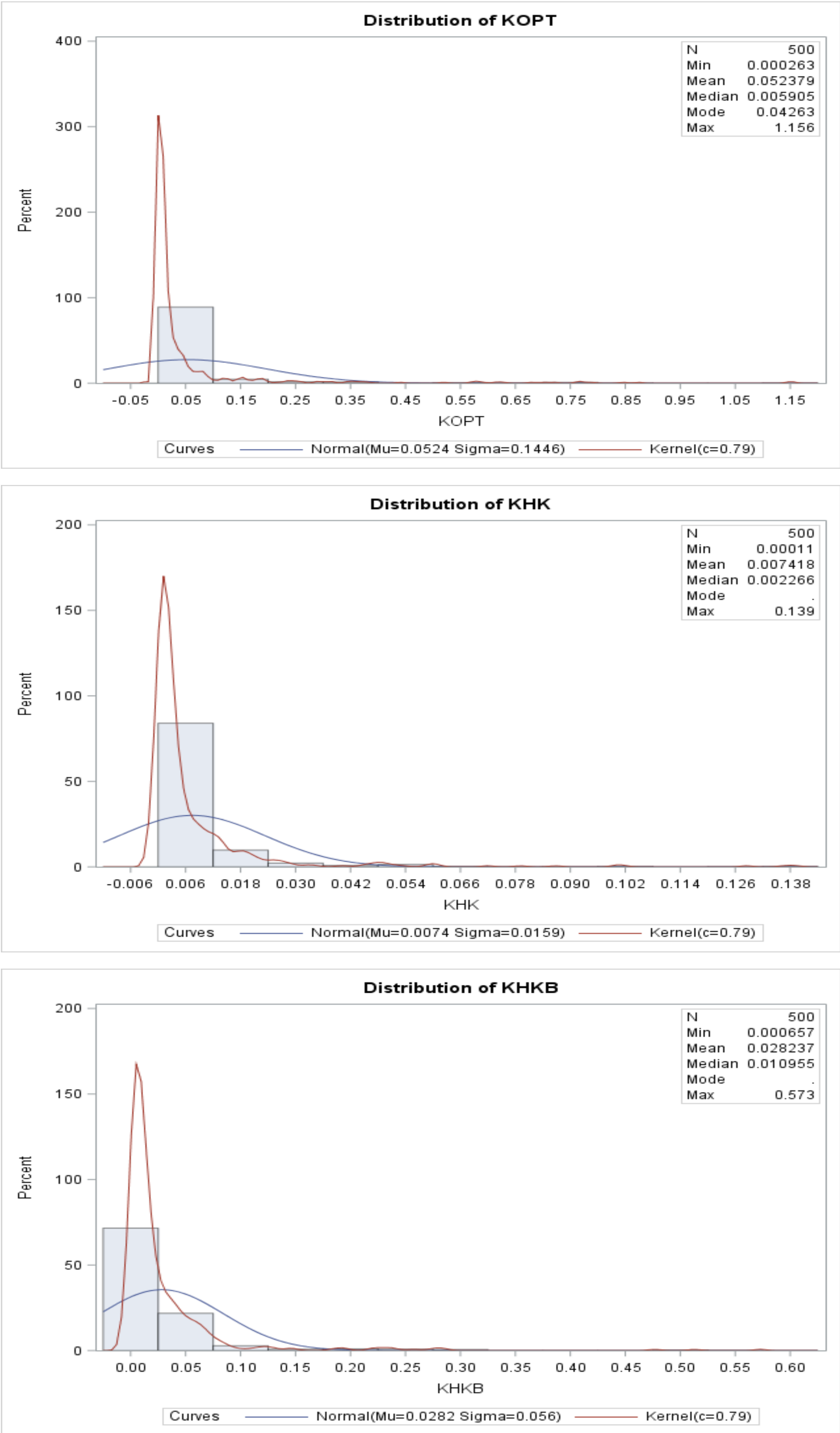
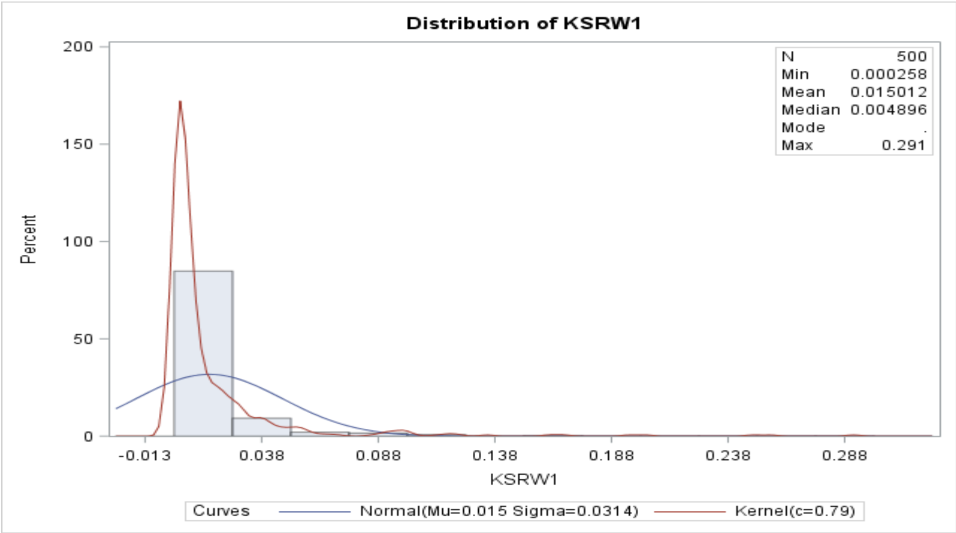
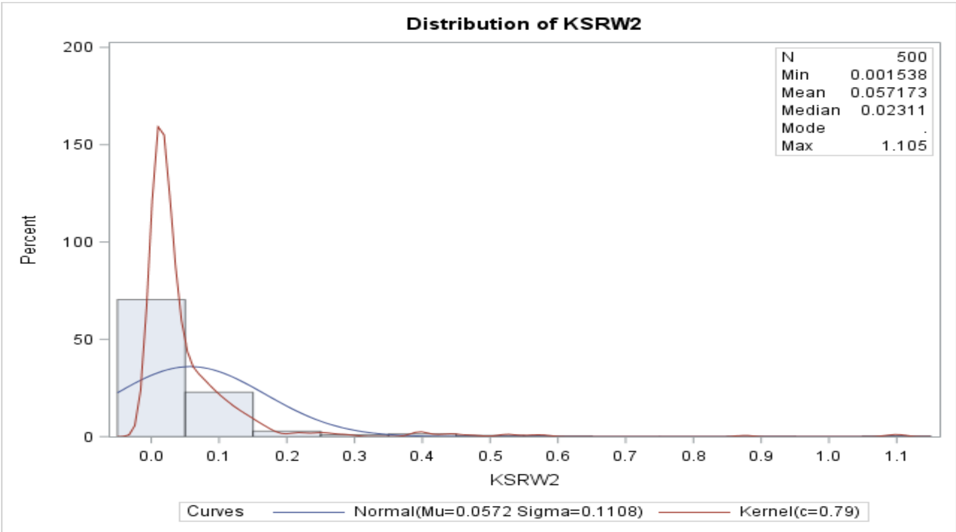


Figure C.7 Distribution of Ridge Parameter for $\rho_{12} = 0.99$, $\rho_{34} = 0.90$, $n = 500$.





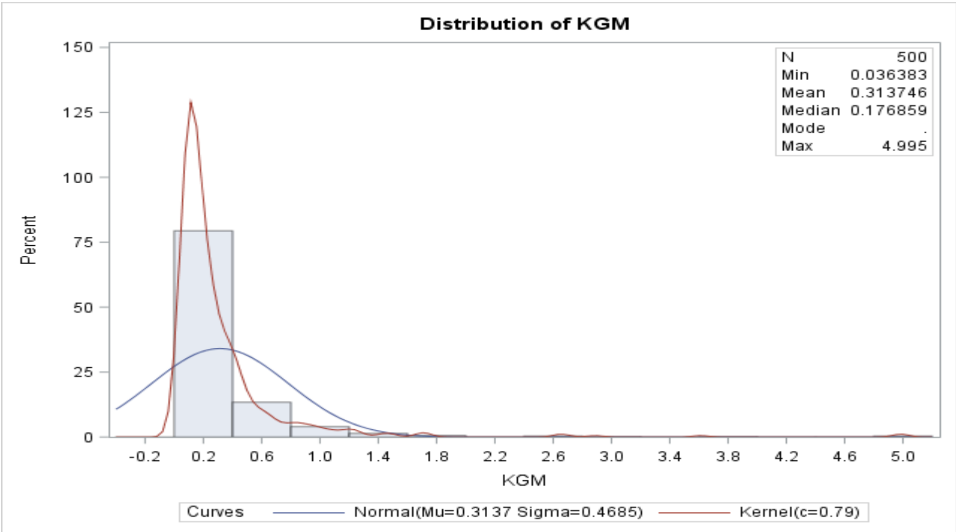
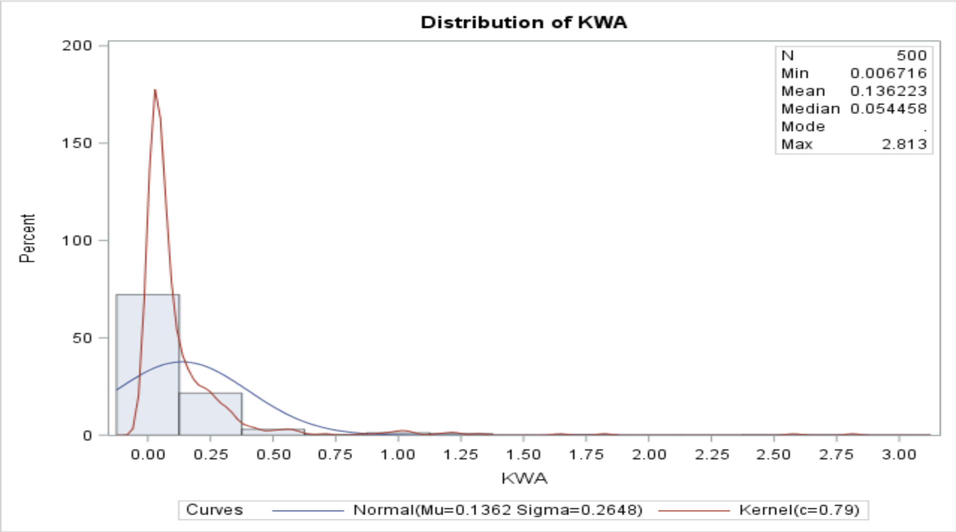


Figure C.7 (Continued)



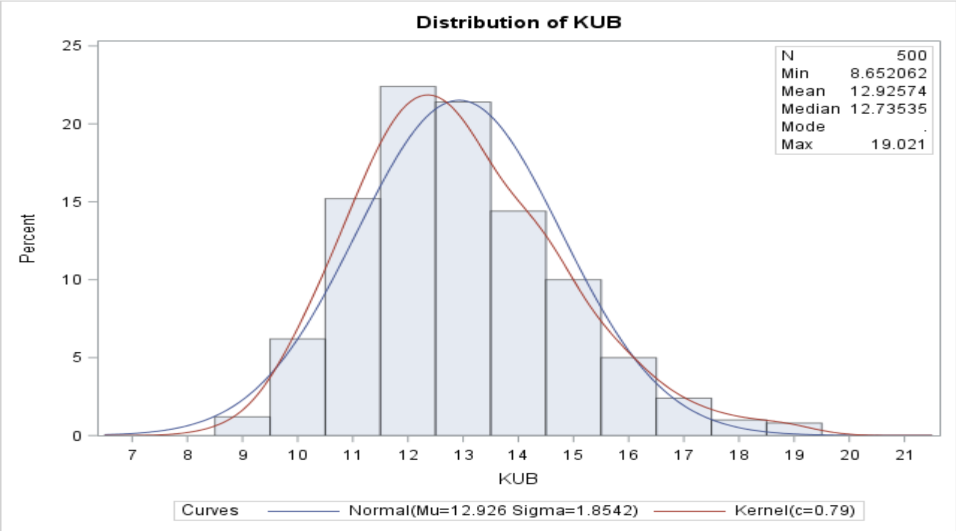


Figure C.7 (Continued)

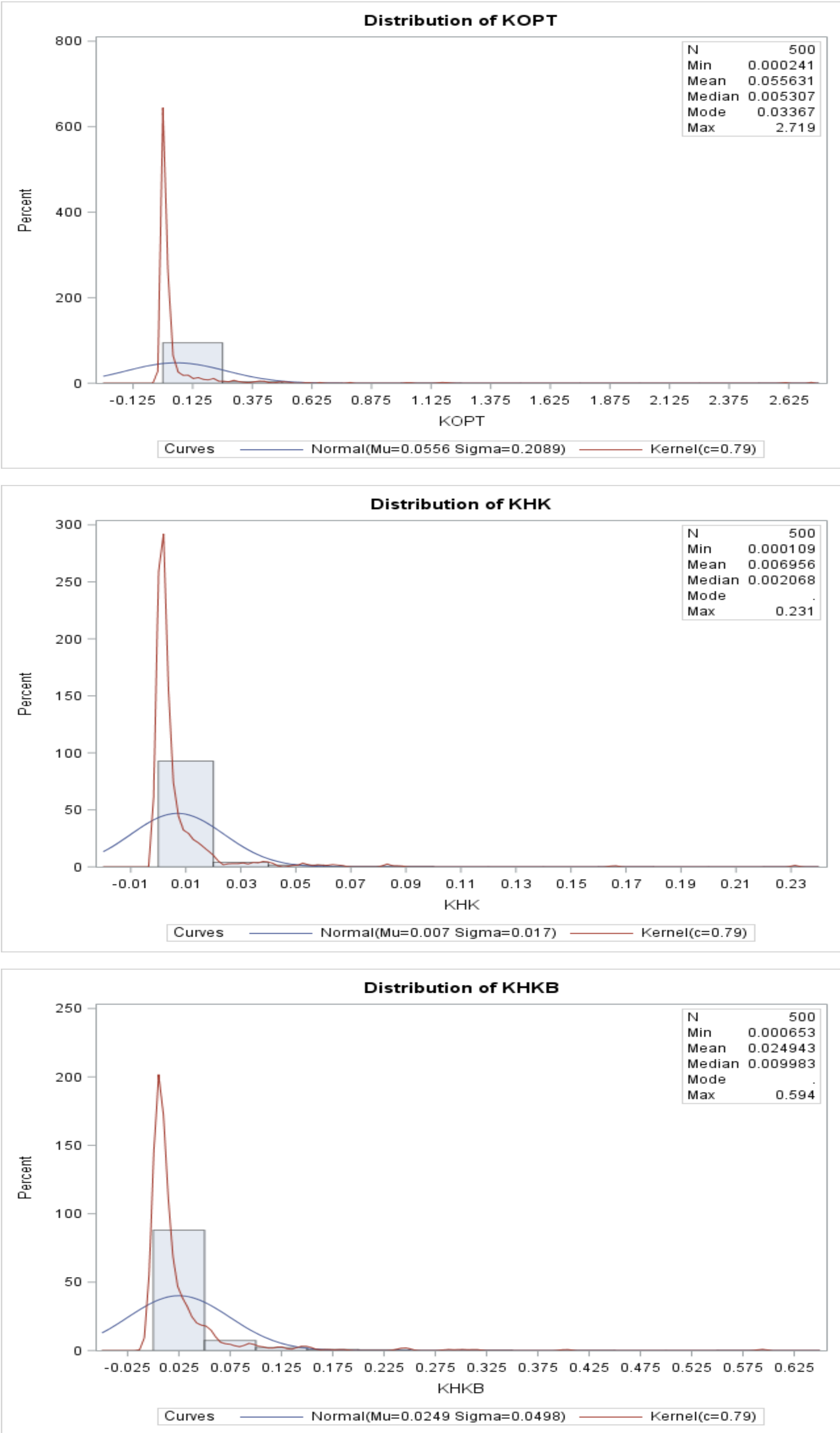
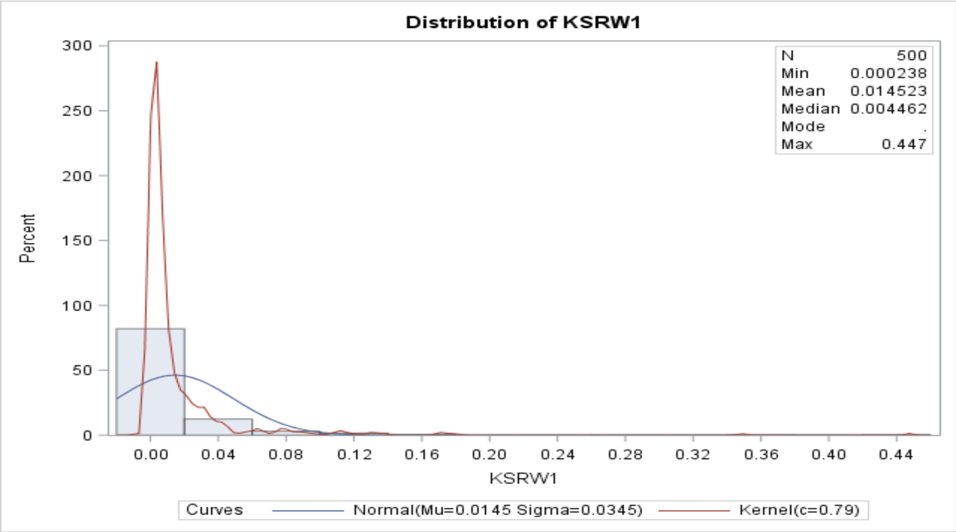
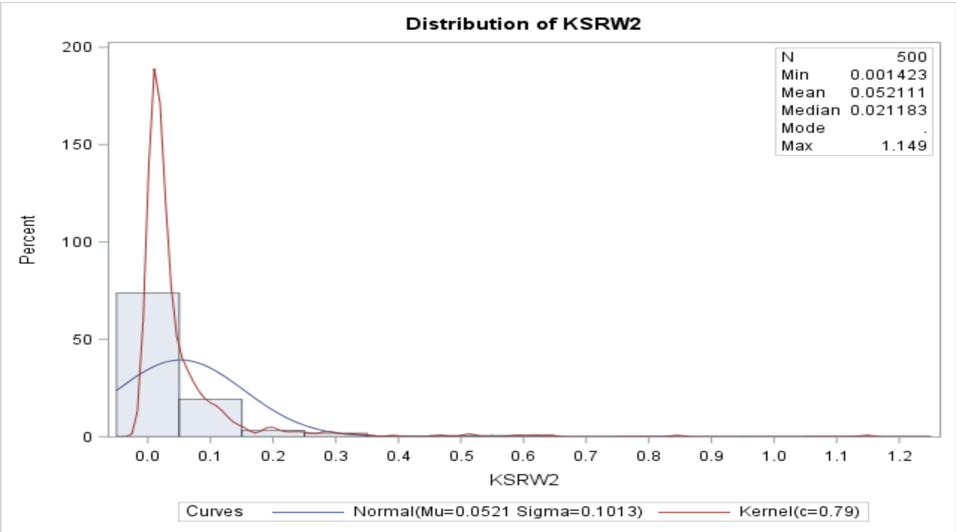


Figure C.8 Distribution of Ridge Parameter for $\rho_{12} = 0.99,\ \rho_{34} = 0.90,\ n = 1000$.





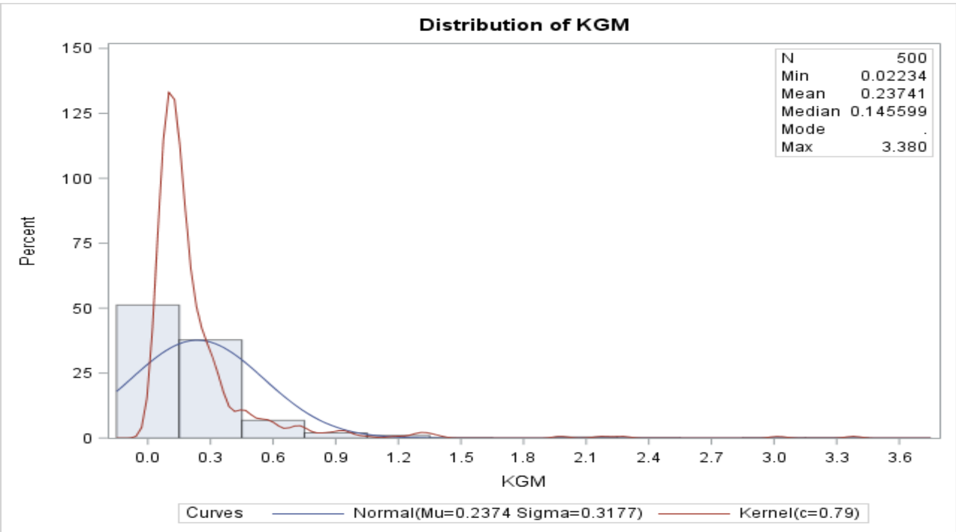
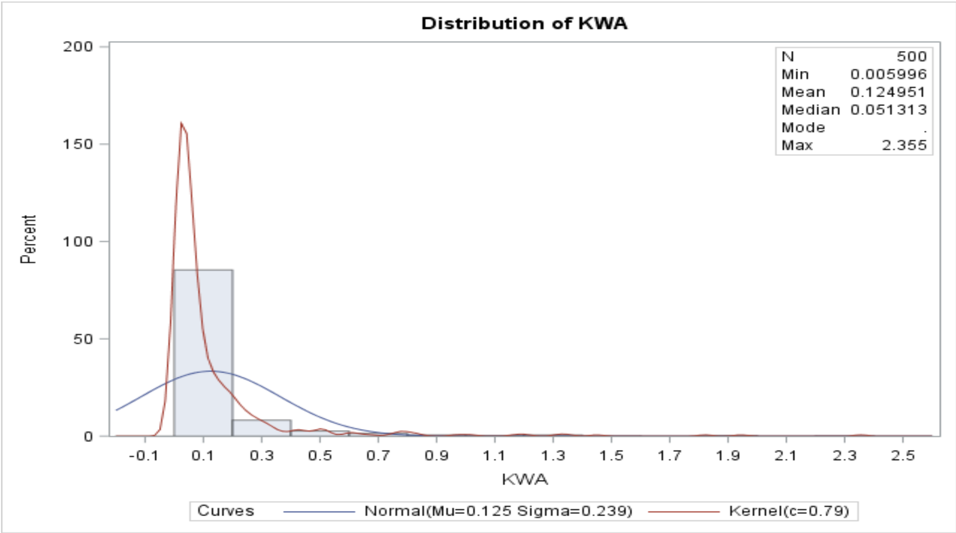


Figure C.8 (Continued)



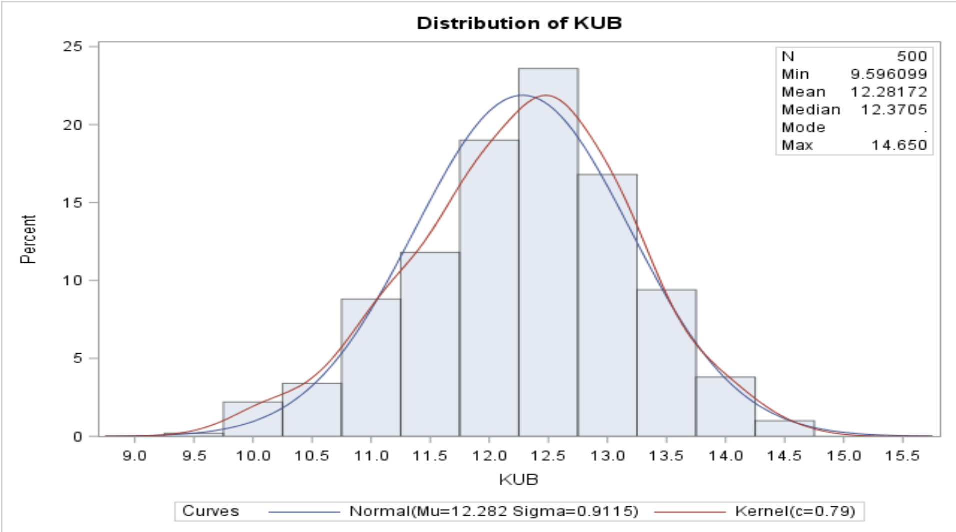


Figure C.8 (Continued)

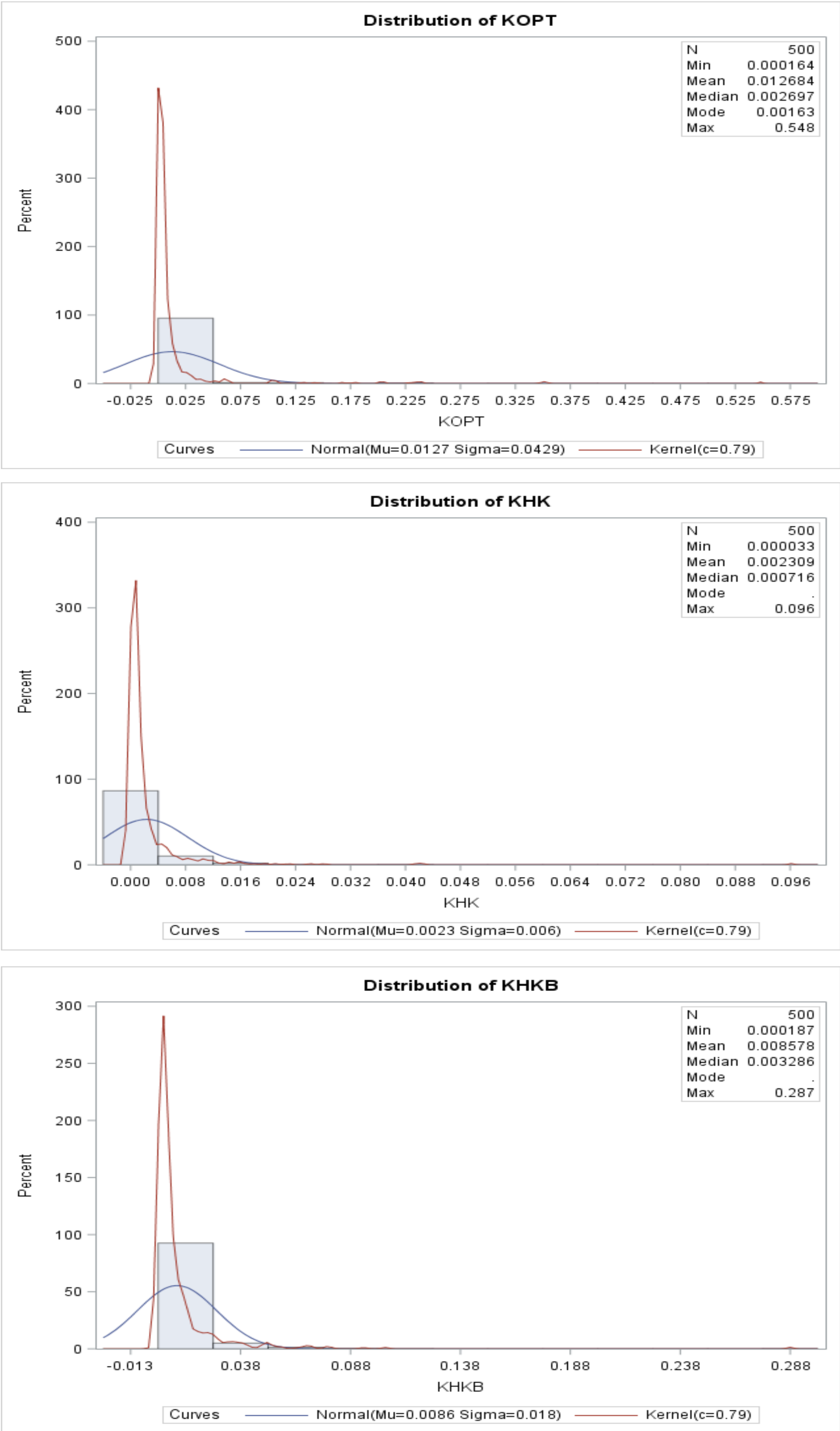
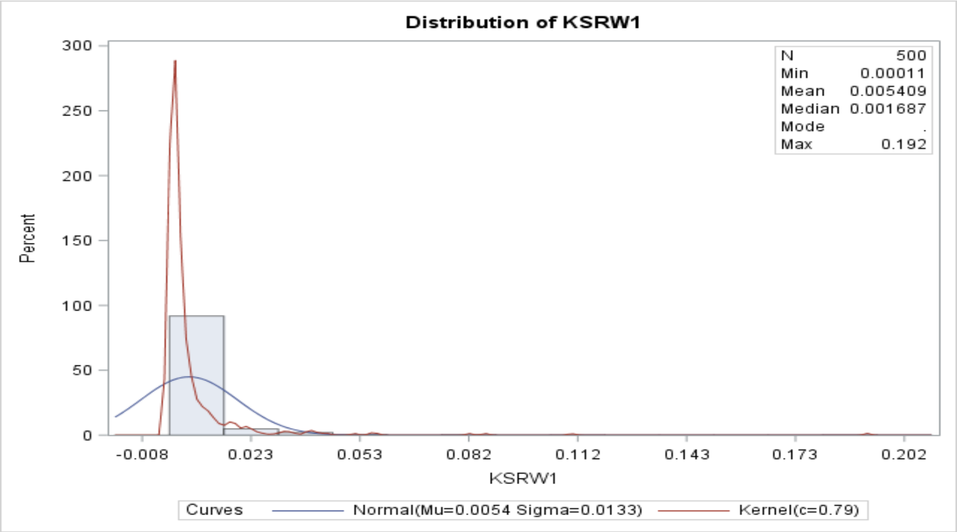
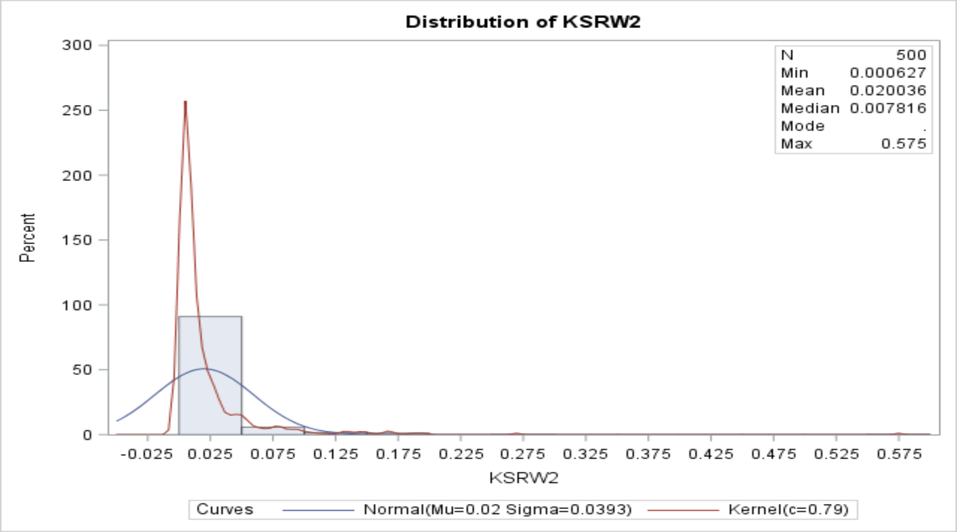


Figure C.9 Distribution of Ridge Parameter for $\rho_{12} = 0.99,\ \rho_{34} = 0.99,\ n = 100$.





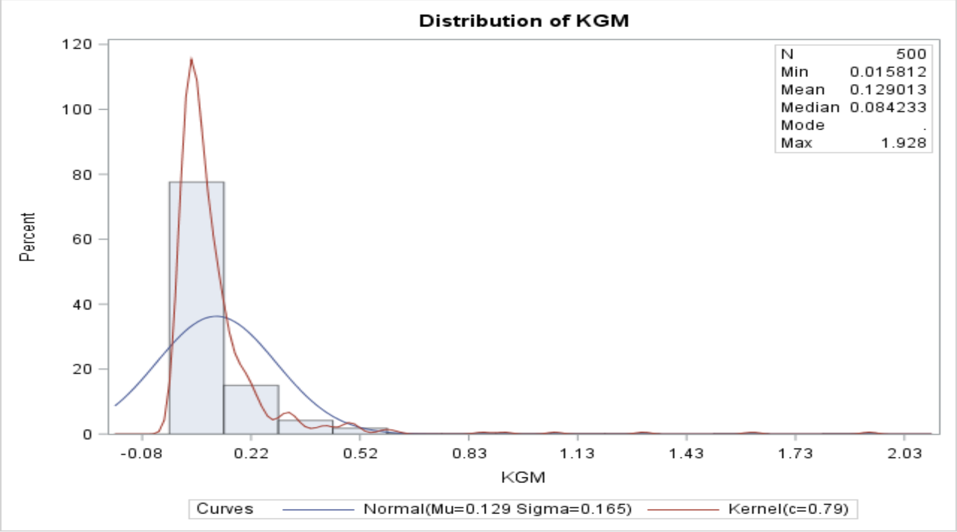
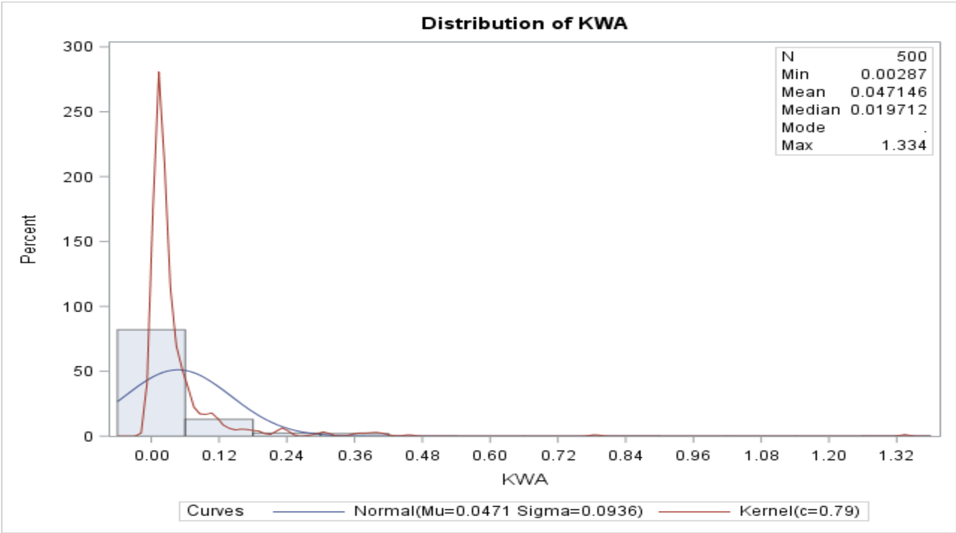


Figure C.9 (Continued)



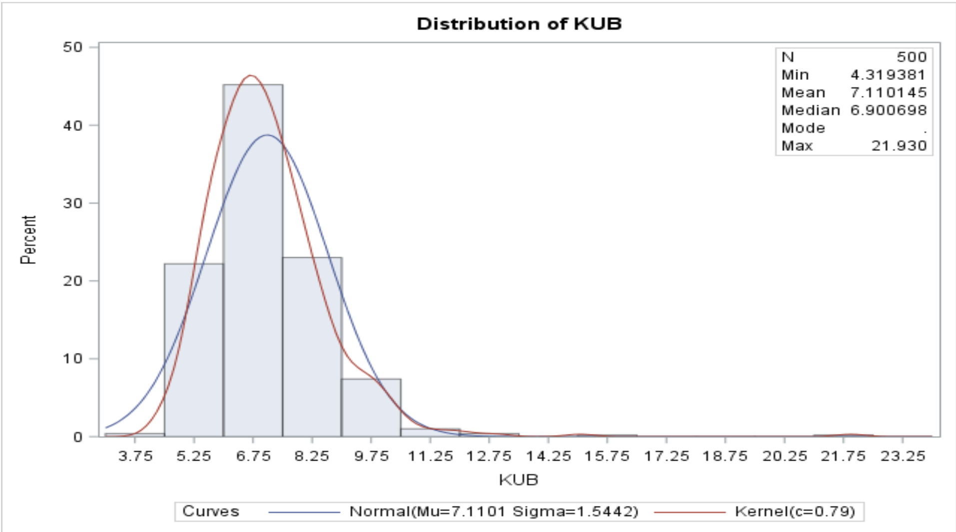


Figure C.9 (Continued)

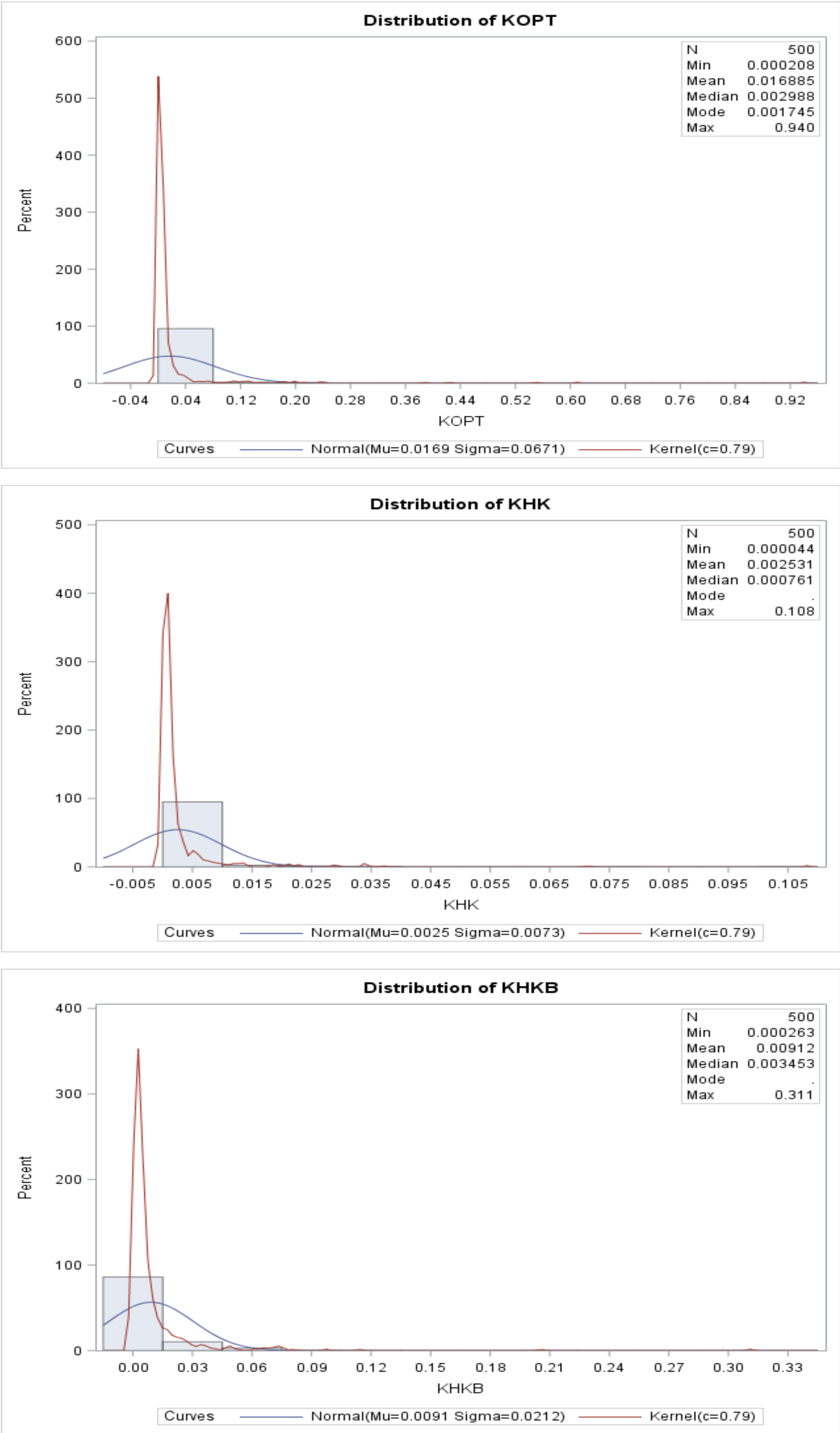
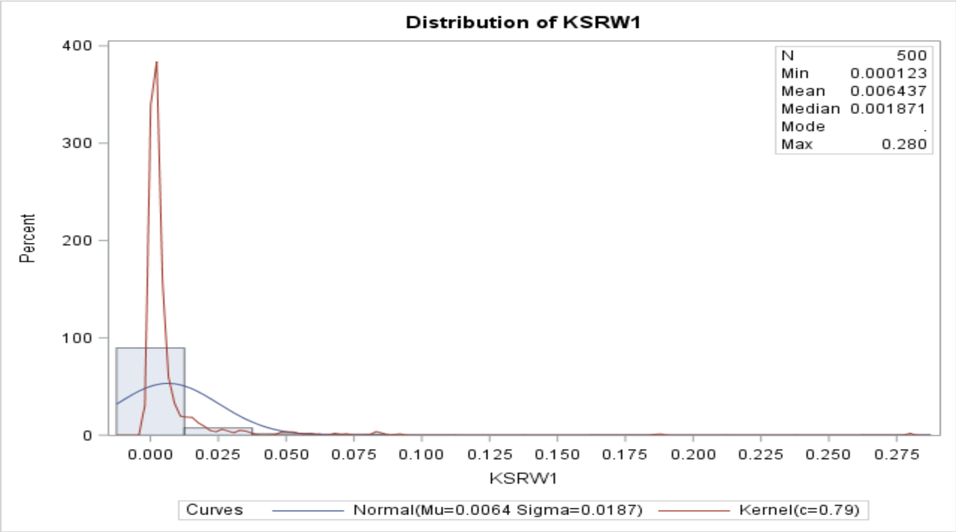
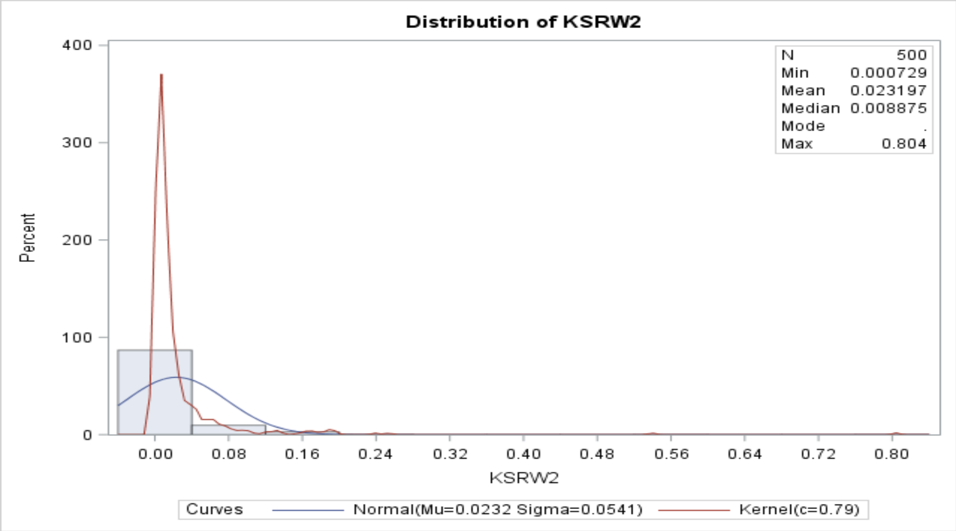


Figure C.10 Distribution of Ridge Parameter for $\rho_{12} = 0.99,\ \rho_{34} = 0.99,\ n = 200$.





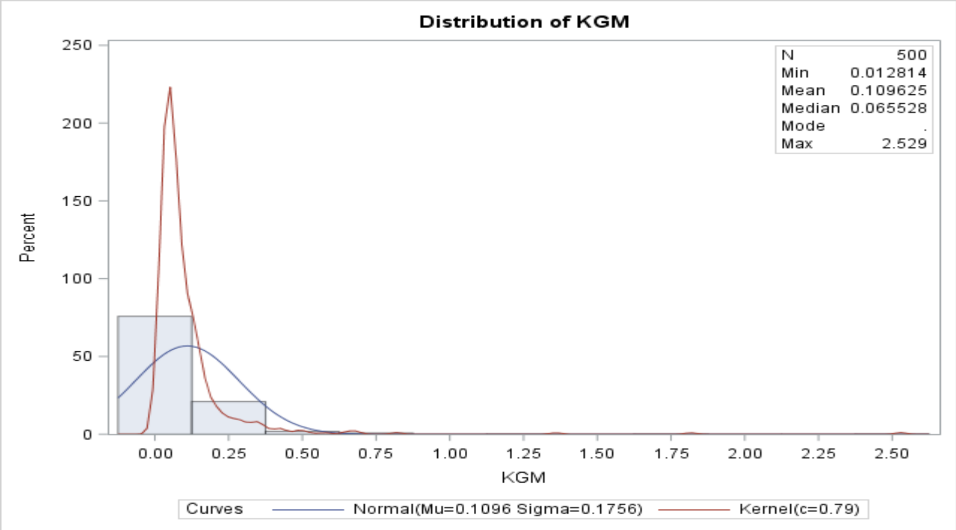
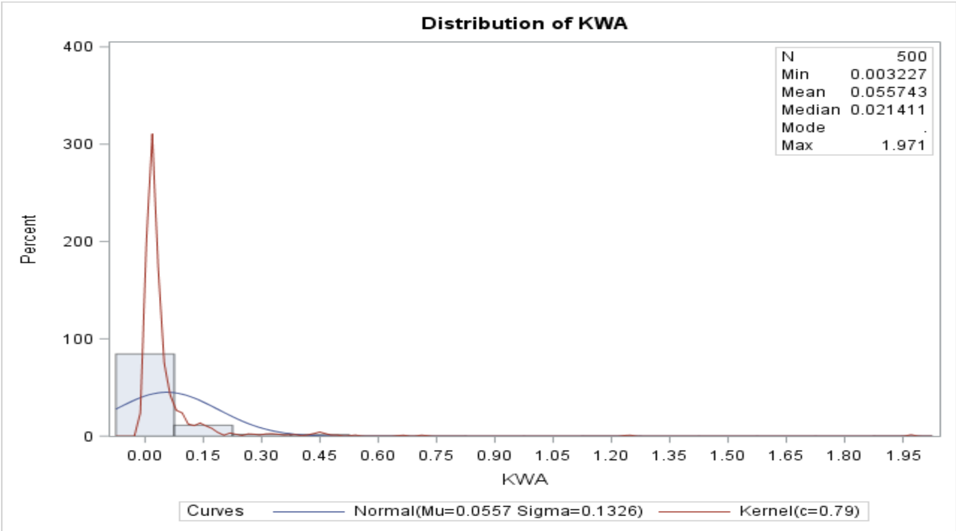


Figure C.10 (Continued)



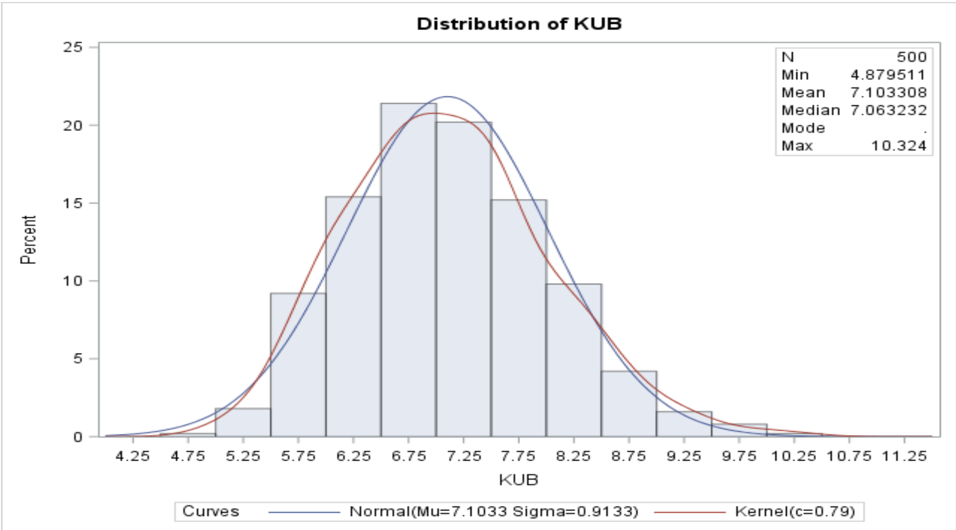


Figure C.10 (Continued)

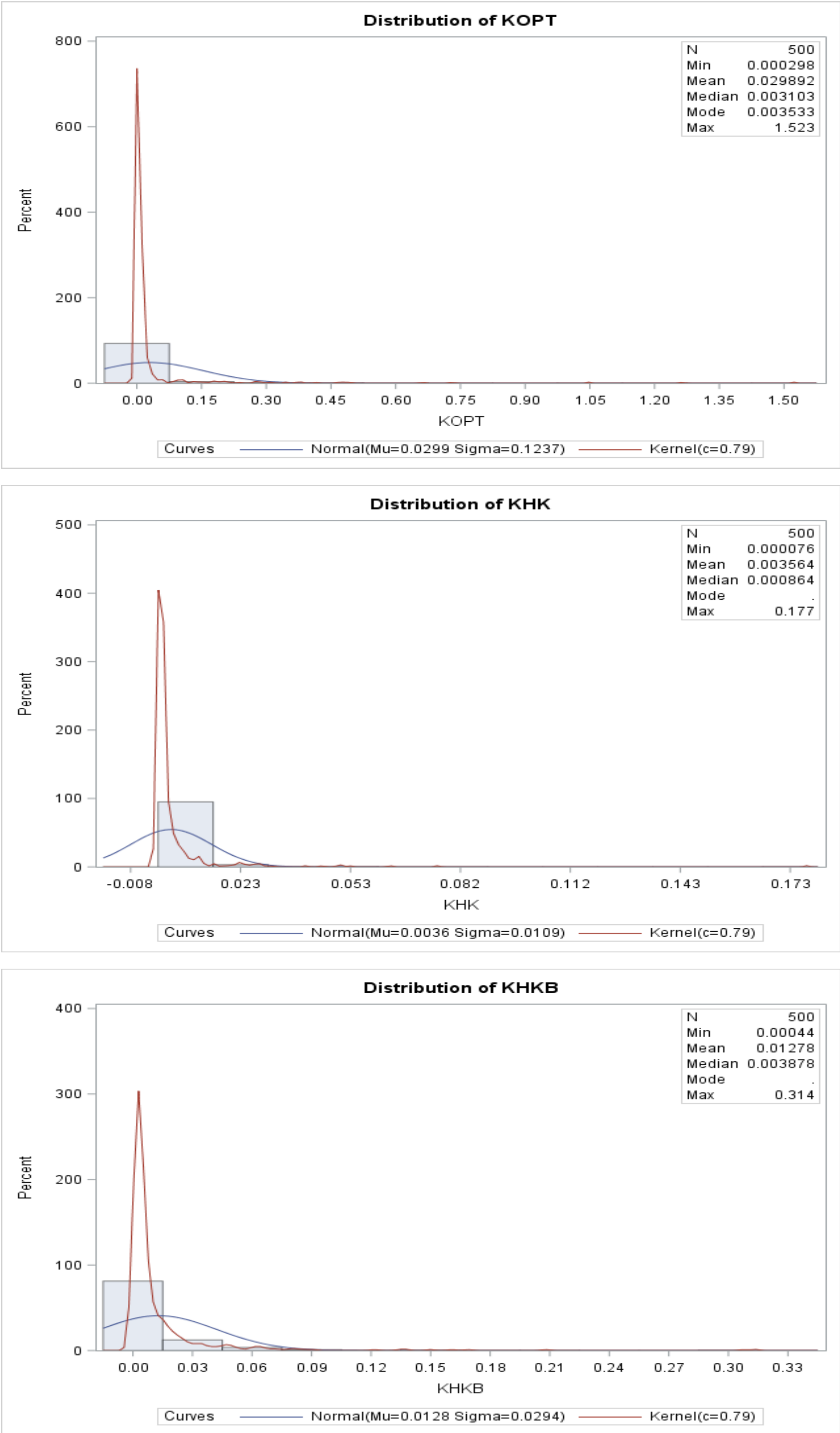


Figure C.11 Distribution of Ridge Parameter for $\rho_{12} = 0.99,\ \rho_{34} = 0.99,\ n = 500$.

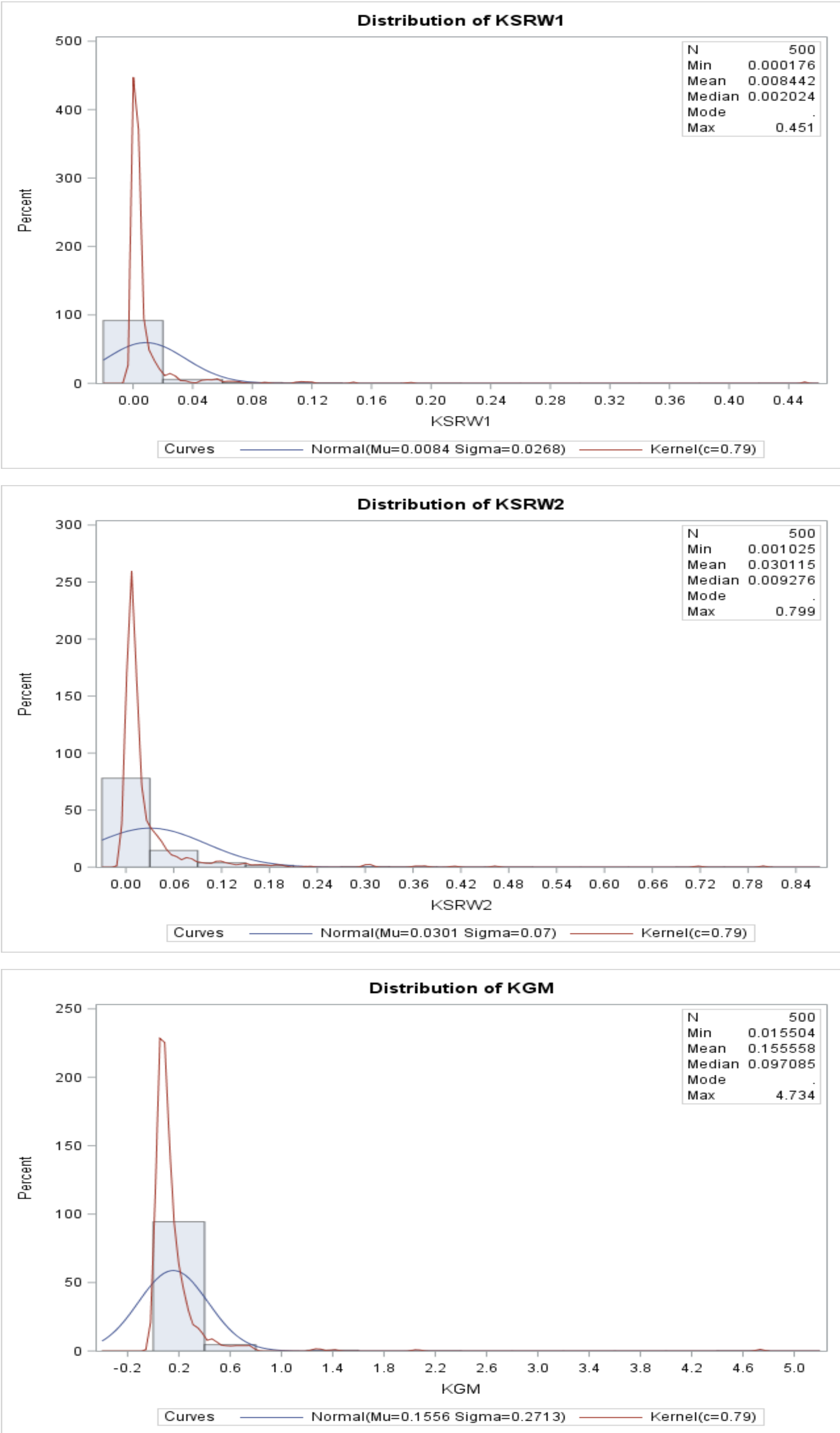
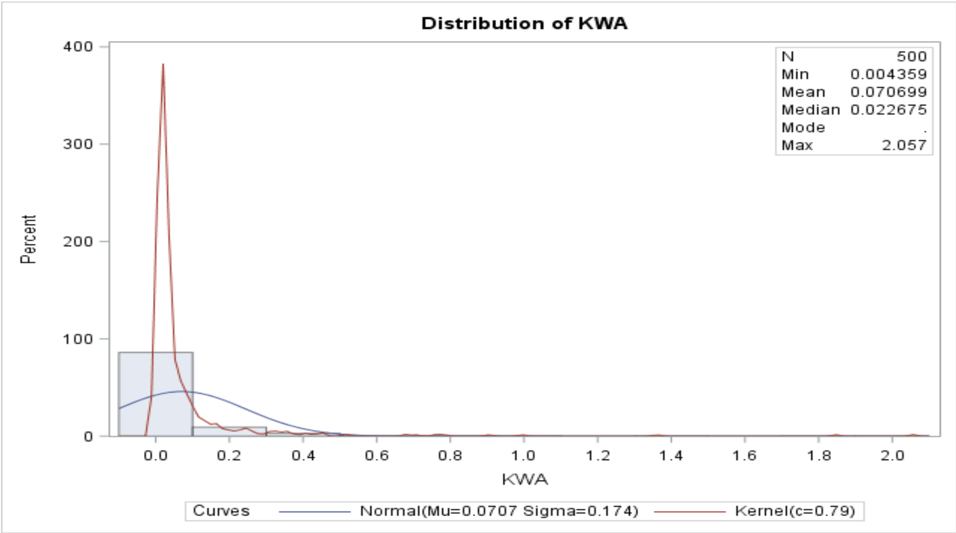


Figure C.11 (Continued)



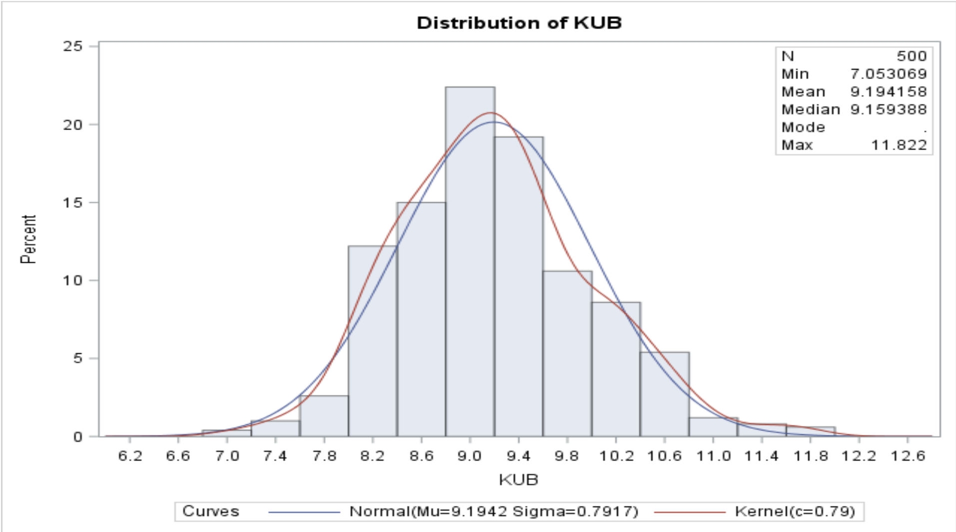


Figure C.11 (Continued)

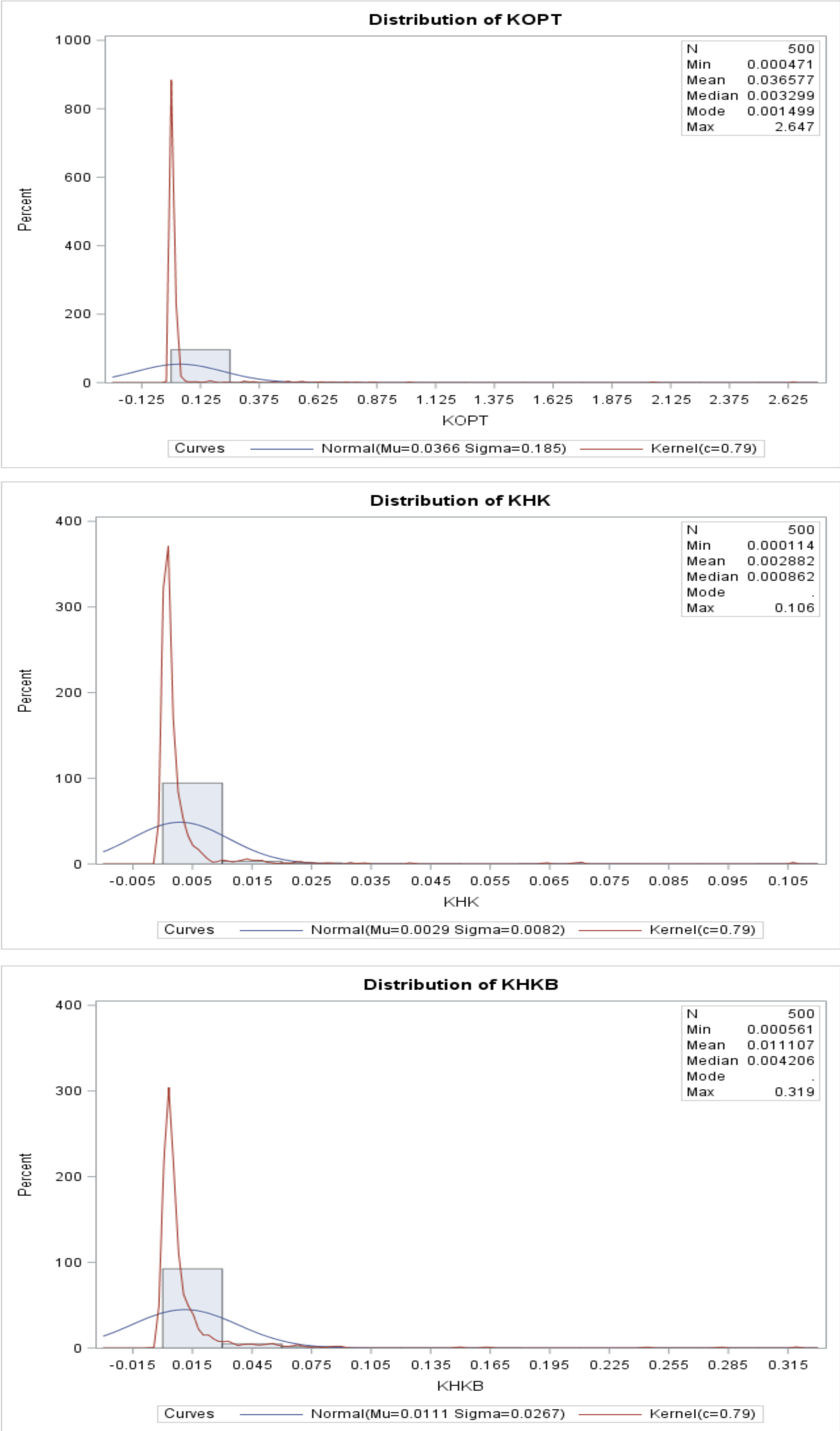
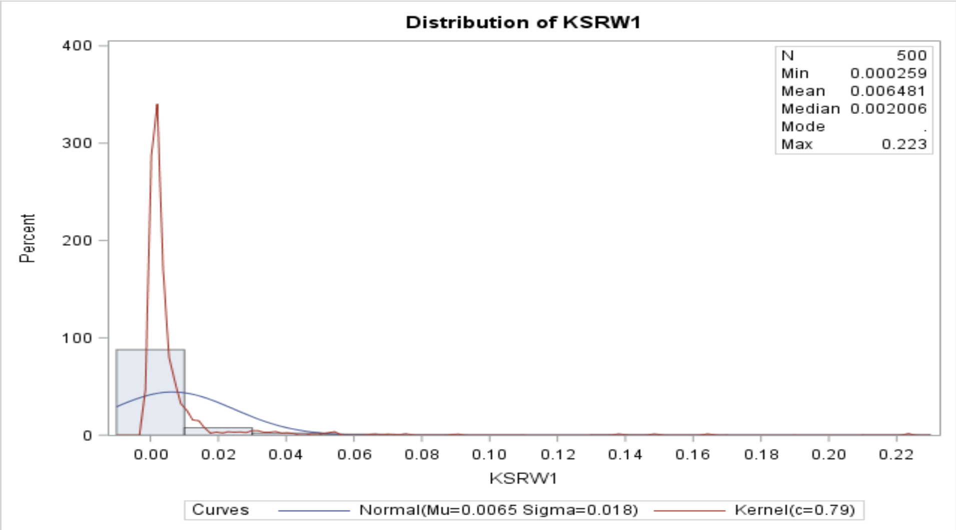
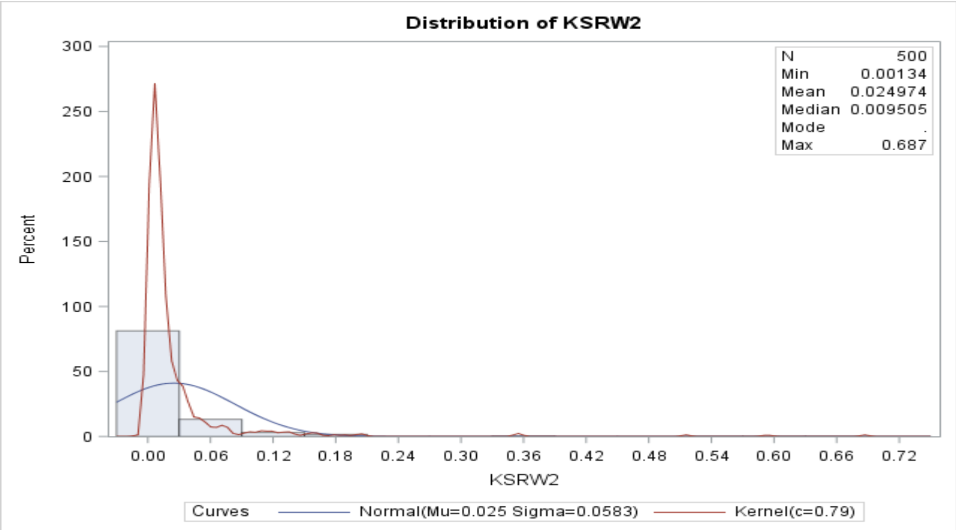


Figure C.12 Distribution of Ridge Parameter for $\rho_{12} = 0.99,\ \rho_{34} = 0.99,\ n = 1000$.





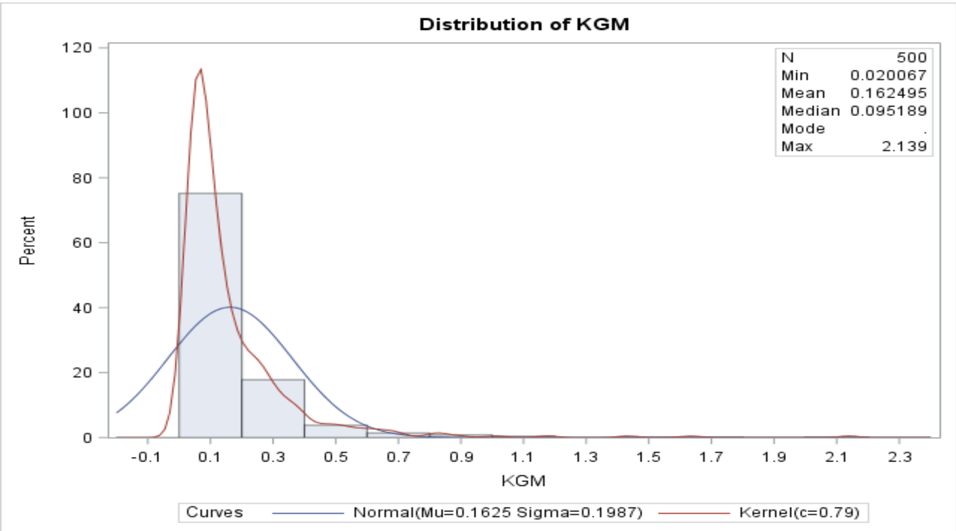
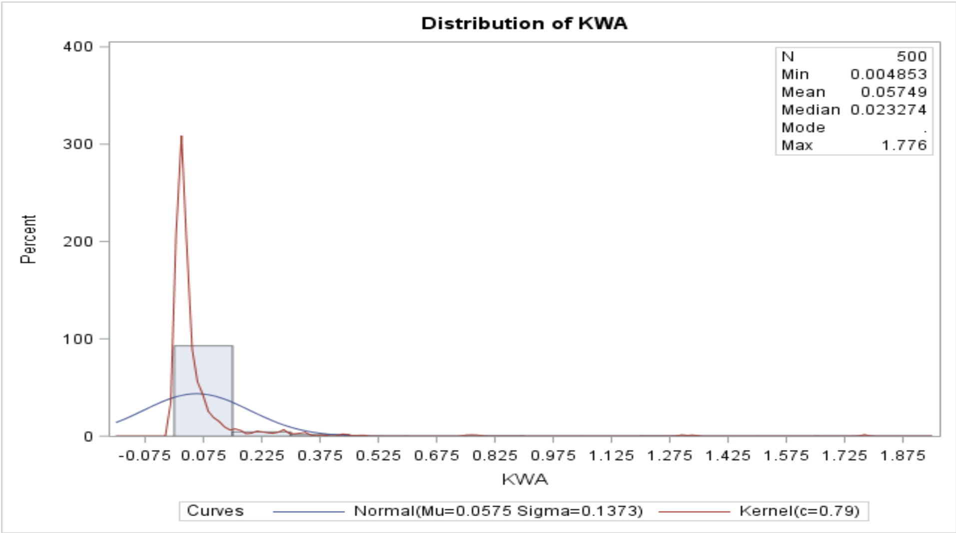


Figure C.12 (Continued)



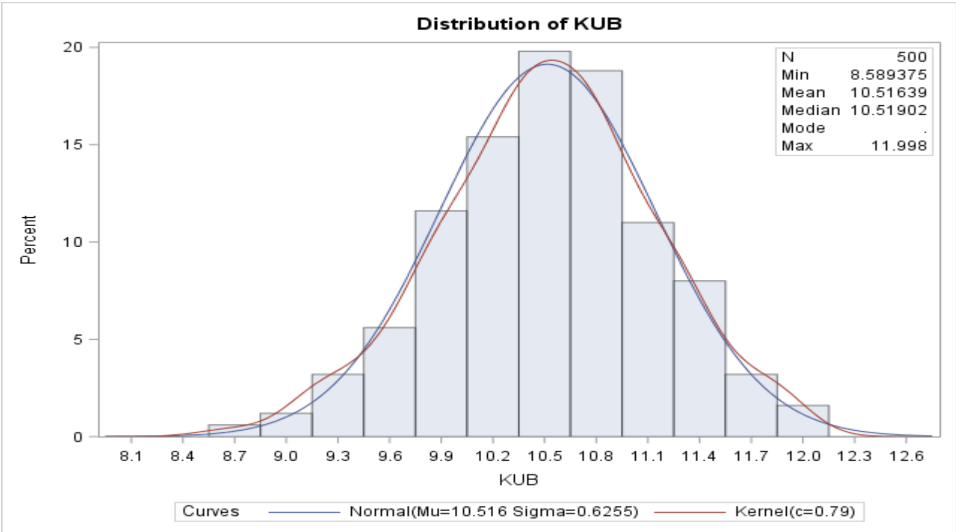


Figure C.12 (Continued)

Appendix D

The Lee Cancer Remission Dataset

The data (Lee, 1974) consist of patient characteristics and feature of cancer remission. The binary response is the cancer remission indicator variable with a value of 1 if the patient received a complete cancer remission and a value of 0 otherwise. The other variables are the risk factors which thought to involve to cancer remission: cellularity of the morrow clot section (CELL), Smear differential percentage of blasts (SMEAR), percentage of absolute morrow leukemia cell infiltrate (INFIL), percentage labeling index of the bone marrow leukemia cells (LI), and the highest temperature prior to star of treatment (TEMP).

Table D.1 The Lee Cancer remission dataset

Obs	Y	CELL	SMEAR	INFIL	LI	TEMP
1	1	0.8	0.83	0.66	1.9	0.996
2	1	0.9	0.36	0.32	1.4	0.992
3	0	0.8	0.88	0.7	0.8	0.982
4	0	1	0.87	0.87	0.7	0.986
5	1	0.9	0.75	0.68	1.3	0.98
6	0	1	0.65	0.65	0.6	0.982
7	1	0.95	0.97	0.92	1	0.992
8	0	0.95	0.87	0.83	1.9	1.02
9	0	1	0.45	0.45	0.8	0.999
10	0	0.95	0.36	0.34	0.5	1.038
11	0	0.85	0.39	0.33	0.7	0.988
12	0	0.7	0.76	0.53	1.2	0.982
13	0	0.8	0.46	0.37	0.4	1.006
14	0	0.2	0.39	0.08	0.8	0.99
15	0	1	0.9	0.9	1.1	0.99
16	1	1	0.84	0.84	1.9	1.02
17	0	0.65	0.42	0.27	0.5	1.014
18	0	1	0.75	0.75	1	1.004
19	0	0.5	0.44	0.22	0.6	0.99
20	1	1	0.63	0.63	1.1	0.986
21	0	1	0.33	0.33	0.4	1.01
22	0	0.9	0.93	0.84	0.6	1.02
23	1	1	0.58	0.58	1	1.002

Table D.2 (Continued)

Obs	Y	CELL	SMEAR	INFIL	LI	TEMP
24	0	0.95	0.32	0.3	1.6	0.988
25	1	1	0.6	0.6	1.7	0.99
26	1	1	0.69	0.69	0.9	0.986
27	0	1	0.73	0.73	0.7	0.986

BIOGRAPHY

NAME	Miss Piyada Phrueksawatnon
ACADEMIC BACKGROUND	Bachelor’s Degree with a major in Mathematics from Naresuan University, Thailand in 2003 and Master’s Degree Applied Statistics from Chiang Mai University, Thailand in 2005.
PRESENT POSITION	Lecturer, School of Science and Technology, University of Phayao, Phayao Province, Thailand