A Liu estimator for the beta regression model and its application to chemical data

Peter Karlsson\(^1\) | Kristofer Månsson\(^{1,2}\) | B.M. Golam Kibria\(^3\)

\(^1\)Department of Economics and Statistics, Linneaus University, Jönköping, Sweden
\(^2\)International Business School, Jönköping University, Jönköping, Sweden
\(^3\)Department of Mathematics and Statistics, Florida International University, Miami, Florida, USA

Abstract

Beta regression has become a popular tool for performing regression analysis on chemical, environmental, or biological data in which the dependent variable is restricted to the interval \([0, 1]\). For the first time, in this paper, we propose a Liu estimator for the beta regression model with fixed dispersion parameter that may be used in several realistic situations when the degree of correlation among the regressors differs. First, we show analytically that the new estimator outperforms the maximum likelihood estimator (MLE) using the mean square error (MSE) criteria. Second, using a simulation study, we investigate the properties in finite samples of six different suggested estimators of the shrinkage parameter and compare it with the MLE. The simulation results indicate that in the presence of multicollinearity, the Liu estimator outperforms the MLE uniformly. Finally, using an empirical application on chemical data, we show the benefit of the new approach to applied researchers.

**KEYWORDS**

beta regression, Liu estimator, Monte Carlo methods, multicollinearity, relative MSE

1 | INTRODUCTION

In many empirical contexts in which the dependent variable is restricted to the interval \([0, 1]\), such as rates and proportions, the classical solution has been to transform the dependent variable so that it is mapped from the interval \([0, 1]\) onto the real line (\(\mathbb{R}\)). In this situation (which is common when analyzing chemical, environmental, or biological data), the logistic regression model is considered, where the log-odds is used as the dependent variable in a linear multiple regression model. Another example of a common transformation is the inverse of a suitable cumulative density function, which also leads to the dependent variable \(Y \in [0, 1]\) being mapped onto the real line. A well-known example of the latter is to use the probit model. But these kinds of solutions have several demerits; for example, the model parameters cannot be easily interpreted in terms of the original dependent variable. Another demerit is that proportion measures typically show asymmetry, and hence any inference based on the normality assumption can be deceptive, especially for a small sample.

As a remedy to the aforementioned problems, Ferrari and Cribari-Neto\(^1\) proposed a beta regression model for a continuous dependent variable \((Y)\) with support in the open interval \([0, 1]\). This model is based on the assumption that the dependent variable follows the beta distribution. One of its advantages is that the regression parameters are interpretable in terms of \(E(Y)\), and further, the model can also accommodate asymmetries and
heteroskedasticity. Later developments have expanded the model so that it is also possible to include covariates for modeling dispersion.\(^2\)

While the performance of the maximum likelihood estimator (MLE) for the model parameters in a beta regression model has been investigated under orthogonal regressors, it is of great interest to investigate the performance of MLE under conditions where the regressors are non-orthogonal and even more interestingly to investigate the performance of a Liu estimator by Liu,\(^3\) as an alternative to the MLE, for situations with orthogonal and nonorthogonal covariates. Shrinkage estimators for the linear regression model were first considered by Hoerl and Kennard\(^4\) where the classical ridge estimator was introduced. This method was further developed and applied to chemical data by Vigneau et al,\(^5\) Muniz and Kibria,\(^6\) and Algamal\(^7\) among others. A major development was the introduction of the Liu estimator by Liu,\(^3\) which has an advantage over the ridge estimator since the shrinkage is a linear function of the shrinkage parameter. This method has also been further developed and applied to chemical data by Kaçiranlar et al,\(^8\) Alheety and Kibria,\(^9\) and Lukman et al.\(^10\) The extent to which these types of methods have been studied is due to the importance of the situation when the regressors are non-orthogonal since this is the most common situation when analyzing chemical, environmental, or biological data.

In this paper, for the first time, we propose some Liu estimators for the beta regression model with fixed dispersion parameter. To estimate the shrinkage parameter of the Liu estimator, we use five different methods from previous literature and one novel method based on the derived optimal value. These estimators will be investigated in the case where we have a multicollinearity issue. We evaluate the MSE properties analytically, use a simulation study to investigate the finite sample properties and show the benefit of using this method on two different chemical data sets. In the first application, we use the same dataset analyzed by Ospina et al.\(^11\) In this data set, the dependent variable is defined as the proportion of crude oil that is converted to petrol after fractional distillation. This dependent variable is related to some explanatory variables that are multicollinear. In the second application, we follow Şimsêk,\(^12\) where we study the color of hazelnuts. The color of roasted hazelnuts is measured using a colorimeter which is one of the most widely used methods in chemistry to determine color characteristics. In both examples, the benefit of our new method is clear to applied researchers analyzing chemical data.

The organization of this paper is as follows: The beta regression model and Liu estimators are discussed in Section 2. A simulation study to compare the performance of the estimators are provided in Section 3. To illustrate the findings of the paper, two real-world data are analyzed in Section 4. Concluding remarks are presented in Section 5.

2 | BETA REGRESSION MODEL AND LIU ESTIMATORS

The beta regression model and some Liu estimators are discussed in this section.

2.1 | Beta regression model

In contrast to the multiple linear regression model in which the dependent variable \(Y\) is assumed to be normally distributed, the beta regression model assumes that \(Y\) has a beta distribution with a support on the interval \([0, 1]\). The beta regression model by Ferrari and Cribari-Neto\(^1\) is based on an alternative parameterization of the beta probability density function (pdf) in terms of a variable mean (\(\mu(x)\)), a precision parameter (\(\varphi\)), and \(x\), which is an \((k \times 1)\) vector of predictors. Hence, using the same parameterization for the beta distribution, we have the following pdf \(Y \sim B(\mu(x), \varphi)\):

\[
\begin{align*}
  f_Y(y; \mu, \varphi) &= \frac{\Gamma(\varphi)}{\Gamma(\mu(x)\varphi)\Gamma((1-\mu(x))\varphi)} y^{\mu(x)\varphi-1} (1-y)^{(1-\mu(x))\varphi-1}, \quad y \in [0,1], \\
  &\mu(x) \in [0,1], \text{ and } \varphi \in [0, \infty].
\end{align*}
\]

In particular,

\[
E[Y|X=x] = \mu(x) \quad \text{and} \quad Var(Y|X=x) = \frac{\mu(x)(1-\mu(x))}{1+\varphi},
\]

where
and thus, for a fixed value of $X$, and hence fixed $\mu(x)$, the variance of $Y$ increases/decreases with decreasing/increasing $\phi$, which explains the name dispersion parameter. Further, for fixed $\phi$, the variance of $Y$ is maximized (unimodality) when $\mu(x) = 0.5$ and minimized as we move to the endpoints of the support for $Y$. Hence, the variance is a function of $\mu(x)$ so that a regression model using this parameterization is heteroscedastic. From Equation 1, the log pdf becomes

$$
\ln(f_Y(y; \mu(x), \phi)) = \ln(\Gamma(\phi)) - \ln(\Gamma(\mu(x)\phi)) - \ln(\Gamma((1-\mu(x))\phi)) + (\mu(x)\phi - 1)\ln(y) + ((1-\mu(x))\phi - 1)\ln(1-y).
$$

Now let $\{Y_i\}_{i=1}^n$ be a random sample from the population $Y_i \sim B(\mu(x_i), \phi)$, and let $g(.)$ be a strictly monotonic and twice differentiable link function such that

$$
g(\mu(x_i)) = \eta_i = x_i^T \beta,
$$

where $x_i$ is a $(k \times 1)$ vector of predictors and $\beta' = [\beta_1, ..., \beta_k]$ is a $(k \times 1)$ vector of unknown regression parameters.

Moreover, we assume that the link function $g(.) : [0,1] \to \mathbb{R}$ and there exist several different link functions that map the linear predictor onto the space $[0, 1]$, such as

$$
\text{logit} \quad g(\mu(x_i)) = \ln\left(\frac{\mu(x_i)}{1-\mu(x_i)}\right) = \eta_i,
$$

$$
\text{probit} \quad g(\mu(x_i)) = \Phi^{-1}(\mu(x_i)) = \eta_i,
$$

where $\ln(.)$ is the natural logarithm and is the standard normal cumulative distribution function. Hence, the beta regression model assumes that the mean of the dependent variable can be represented in the following form:

$$
g(\mu(x_i)) = \eta_i = x^T \beta.
$$

When using the logit link function, the beta regression model is obtained by assuming that the conditional mean of $Y_i$ can be written as

$$
\mu(x_i) = \frac{e^{x^T \beta}}{1 + e^{x^T \beta}}.
$$

Based on the random sample, the log-likelihood function for this family of beta regression models is then given by

$$
\ell(\mu(x_i), \phi; y, x_i) = \sum_{i=1}^n \{\ln(\Gamma(\phi)) - \ln(\Gamma(\mu(x_i)\phi)) - \ln(\Gamma((1-\mu(x_i))\phi)) + (\mu(x_i)\phi - 1)\ln(y_i) + ((1-\mu(x_i))\phi - 1)\ln(1-y_i)\},
$$

where $\mu(x_i) = g^{-1}(\eta_i)$.

MLEs for $\beta$, $\phi$ are usually found by numerically maximizing Equation 9. Ospina and Ferrari13 show that MLEs can be obtained by using a re-weighted least square algorithm (Greene14). These estimators are not unbiased, especially in small samples (Cribari-Neto and Vasconcellos15), although Simas et al2 have obtained analytical bias corrections for the MLEs. We do not perform any bias corrections for the MLEs because we use the betareg package that only offers the possibility to estimate beta regressions without any bias corrections in the R system for statistical computing (Cribari-Neto and Zeileis16). The motivation for this paper is twofold: first, to examine the effect of multicollinearity on MLEs and, second, to use inbuilt packages for estimating a beta regression model (as is common among practitioners).
2.2 | Liu estimators for a beta regression model

Different types of shrinkage estimators are available for estimating the regression parameters when the regressors are ill-conditioned in the model. The ordinary ridge regression estimator was first developed for the linear regression model by Hoerl and Kennard. Later developments led to different new types of ridge regression estimators. Liu proposed an alternative but similar regression estimator with a different biasing parameter. In a multiple regression context, the regression parameter vector \( \beta \) for the Liu estimator is specified as follows:

\[
\hat{\beta}(d) = (X'X + I)^{-1}(X'X + dI)\hat{\beta}_{ML}
\]

(10)

where \( d \) is a shrinkage parameter defined in the interval \([0, 1]\) and \( \hat{\beta}_{ML} \) is the MLE of \( \beta \).

In a beta regression context, the parameter vector \( \beta \) for the Liu estimator becomes

\[
\hat{\beta}(d) = (X'WX + I)^{-1}(X'WX + dI)\hat{\beta}_{ML},
\]

(11)

where \( d \) is a shrinkage parameter defined in the interval \([0, 1]\) and \( \hat{\beta}_{ML} \) is the MLE of \( \beta \) obtained by a re-weighted least squares algorithm.

2.3 | A comparison of the MSE properties of Liu and ML

Let \( \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_p) \) be a diagonal matrix with the eigenvalues of \( X'WX \) and let \( \Gamma = [\gamma_1, \cdots, \gamma_p] \) be a matrix with corresponding orthonormal eigenvectors. Then, based on the risk function in 27, the ML estimator is given by

\[
\text{MSE} \left( \hat{\beta}_{ML} \right) = E \left[ \left( \hat{\beta}_{ML} - \beta \right)' \left( \hat{\beta}_{ML} - \beta \right) \right] = tr \left\{ E \left[ \left( \hat{\beta}_{ML} - \beta \right)' \left( \hat{\beta}_{ML} - \beta \right) \right] \right\} = tr \left\{ \left( \phi X'WX \right)^{-1} \right\} = \frac{1}{\varphi} \sum_{j=1}^{p} \frac{1}{\lambda_j}
\]

(12)

The Liu estimator can be written as follows:

\[
\hat{\beta}_d = (X'WX + I)^{-1}(X'WX + dI)\hat{\beta}_{ML} = M\hat{\beta}_{ML}.
\]

(13)

To derive the MSE of the Liu estimator 13, some simplifying lemmas are listed below.

**Lemma 1:** \( M'M = \Gamma \text{diag} \left( \frac{(\lambda + d)^2}{\lambda + 1} \right) \Gamma' \).

**Proof**: The proof of Lemma 1 can be seen in Appendix A.

**Lemma 2:** \( (M - I)'(M - I) = \Gamma \text{diag} \left( \frac{(d-1)^2}{\lambda + 1} \right) \Gamma' \).

**Proof**: The proof of Lemma 2 can be seen in Appendix B.

Based on the risk function in (3.4) and by 13 we have for the Liu estimator:

\[
\text{MSE} \left( \hat{\beta}(d) \right) = E \left[ \left( M\hat{\beta}_{ML} - \beta \right)' \left( M\hat{\beta}_{ML} - \beta \right) \right] = E \left[ \left( M\hat{\beta}_{ML} - M\beta \right)' \left( M\hat{\beta}_{ML} - M\beta \right) \right] + 2E \left[ \left( M\hat{\beta}_{ML} - M\beta \right)'(M\beta - \beta) \right] + E \left[ (M\beta - \beta)'(M\beta - \beta) \right] =
\]

\[
E \left[ tr \left( \left( \hat{\beta}_{ML} - \beta \right)' \left( \hat{\beta}_{ML} - \beta \right) \right) M'M \right] + \beta'(M - I)'(M - I)\beta.
\]

(14)
By Lemmas 1 and 2, the MSE for the Liu estimator (14) becomes

$$\text{MSE}(\hat{\beta}(d)) = tr\left\{ E\left[ (\hat{\beta}_{\text{ML}} - \beta) (\hat{\beta}_{\text{ML}} - \beta)^\top \right] \Gamma \text{diag}\left( \frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2} \right) \right\} + \beta^T \text{diag}\left( \frac{(d-1)^2}{(\lambda_j + 1)^2} \right) \Gamma \beta =$$

$$tr\left\{ \frac{1}{\varphi} (X'WX)^{-1} \Gamma \text{diag}\left( \frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2} \right) \Gamma' \right\} + \beta^T \text{diag}\left( \frac{(d-1)^2}{(\lambda_j + 1)^2} \right) \Gamma \beta =$$

$$tr\left\{ \frac{1}{\varphi} \Gamma \text{diag}\left( \frac{1}{\lambda_j} \right) \Gamma \text{diag}\left( \frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2} \right) \Gamma' \right\} + \beta^T \text{diag}\left( \frac{(d-1)^2}{(\lambda_j + 1)^2} \right) \Gamma \beta =$$

$$tr\left\{ \Gamma \text{diag}\left( \frac{(\lambda_j + d)^2}{\varphi \lambda_j (\lambda_j + 1)} \right) \Gamma' \right\} + \beta^T \text{diag}\left( \frac{(d-1)^2}{(\lambda_j + 1)^2} \right) \Gamma \beta.$$

Substituting $\alpha = \Gamma \beta$ into 15 and using the fact that the trace of a square matrix is the sum of its eigenvalues, we obtain

$$\text{MSE}(\hat{\beta}(d)) = \frac{1}{\varphi} \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)} + \alpha^T \text{diag}\left( \frac{(d-1)^2}{(\lambda_j + 1)^2} \right) \alpha = \frac{1}{\varphi} \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)} + (d-1)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2}. \quad (16)$$

Finally, we want to prove that the Liu estimator is efficient compared with the ML estimator in terms of MSE. If the Liu estimator is efficient, then there must exist a value of $d \in [0,1]$ such that the MSE of the ML estimator is larger than the MSE of the Liu estimator.

**Theorem 1:** There exists a value of $d \in [0,1]$ such that

$$\Delta \text{MSE}(\hat{\beta}_{\text{ML}}, \hat{\beta}(d)) = \text{MSE}(\hat{\beta}_{\text{ML}}) - \text{MSE}(\hat{\beta}(d)) > 0.$$

**Proof:** The proof of Theorem 1 can be seen in Appendix C.

To find an optimal $\hat{\beta}(d)$, we need to determine the optimal value of $d$ such that the MSE of $\hat{\beta}(d)$ is minimized in the closed interval from zero to one. From the proof of Theorem 1, we have that

$$\frac{\partial}{\partial d} \text{MSE}(\hat{\beta}(d)) = \frac{1}{\varphi} \sum_{j=1}^{p} \frac{2(\lambda_j + d)}{\lambda_j (\lambda_j + 1)} + 2(d-1) \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2}. \quad (17)$$

Setting 17 equal to zero yields

$$0 = \frac{1}{\varphi} \sum_{j=1}^{p} \frac{2(\lambda_j + d)}{\lambda_j (\lambda_j + 1)} + 2(d-1) \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2} \Rightarrow$$

$$0 = d \varphi^{-1} \sum_{j=1}^{p} \frac{1}{\lambda_j (\lambda_j + 1)} + d \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2} + \varphi^{-1} \sum_{j=1}^{p} \frac{1}{(\lambda_j + 1)^2} - \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2} \Rightarrow$$

$$d = \frac{\varphi^{-1} \sum_{j=1}^{p} \frac{1}{\lambda_j (\lambda_j + 1)} - \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2}}{\sum_{j=1}^{p} \frac{\varphi \alpha_j^2}{(\lambda_j + 1)^2} + \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2}}. \quad (18)$$

Hence, this is a regular local extreme value problem, where the solution is limited to the domain $d \in [0,1]$. Given that the critical point given by 18 lies within the interval 0 to 1, the choice of $d$ is based on a comparison of the MSE for the points $d = 0$ and the calculated $d$ by 18. Endpoint $d = 1$ is excluded due to the result in Theorem 1.
2.4 | Estimation of the shrinkage parameter

To obtain such a shrinkage estimator, several estimators of the shrinkage parameter \( d \) have been suggested. We have selected a number of these that have shown good results in other applications. These estimators were mainly developed in a multiple regression context for the case where \( d \) is restricted to a scalar, hence yielding the same shrinkage for all parameters in \( \beta \).\(^{3,17,19} \)

These estimators are easily extendable to a beta regression context because the estimator of the unknown \( \beta \) parameter vector can be expressed in the form of a weighted least squares estimator, that is, the form of the Liu estimator.\(^{11} \)

To present the investigated Liu estimators of the shrinkage parameter \( d \) in the paper, we first define \( \hat{\lambda}_j \) to be the \( j \)th eigenvalue of the matrix \( X^\top W X \), where \( X \) is a \((n \times p)\) matrix containing the regressors in the beta regression model and \( W \) is a diagonal weight matrix with the elements \( \left\{ \frac{1}{\hat{\sigma}_j^2}, \cdots, \frac{1}{\hat{\sigma}_n^2} \right\} \) in its diagonal. Furthermore, \( \hat{\lambda}_{\text{max}} = \max(\hat{\lambda}_1, \cdots, \hat{\lambda}_p) \) and \( \hat{\beta}_{\text{max}}^2 \) is the largest element in \( \hat{\beta} \odot \hat{\beta} \) where \( \odot \) is the Hadamard product.

The first proposed estimator is based on the work of Hoerl and Kennard\(^{4} \) and is defined as follows:

\[
d_1 = \max \left( 0, \frac{\hat{\beta}_{\text{max}}^2 - 1}{\hat{\lambda}_{\text{max}} - 1} + \hat{\beta}_{\text{max}}^2 \right). \tag{19}
\]

The following two estimators are based on the work of Kibria,\(^{17} \) which were developed for a linear multiple regression context:

\[
d_2 = \max \left( 0, \text{Median} \left( \left\{ \frac{\hat{\beta}_j^2 - 1}{\hat{\lambda}_j - 1} + \hat{\beta}_j^2 \right\}_{j=1}^p \right) \right). \tag{20}
\]

\[
d_3 = \max \left( 0, \sum_{j=1}^p \frac{\hat{\beta}_j^2 - 1}{\hat{\lambda}_j - 1} + \hat{\beta}_j^2 \right). \tag{21}
\]

The remaining two estimators are based on the work of Khalaf and Shukur,\(^{19} \) which were also developed for a linear multiple regression context:

\[
d_4 = \max \left( 0, \max \left( \left\{ \frac{\hat{\beta}_j^2 - 1}{\hat{\lambda}_j - 1} + \hat{\beta}_j^2 \right\}_{j=1}^p \right) \right). \tag{22}
\]

\[
d_5 = \max \left( 0, \min \left( \left\{ \frac{\hat{\beta}_j^2 - 1}{\hat{\lambda}_j - 1} + \hat{\beta}_j^2 \right\}_{j=1}^p \right) \right). \tag{23}
\]

Finally, the last estimator is based on optimal \( \hat{\beta}(d) \) where \( d \) is determined by plugging in MLE’s of \( \phi, \{ \hat{\lambda}_j \}_{j=1}^p, \beta \) in 18; hence, the final estimator is

\[
d_6 = \text{Max}(0, \text{Min}(d^*, 1)),
\]

where \( d^* = \sum_{j=1}^p \frac{\hat{\sigma}_j^{-1} - \hat{\alpha}_j^2}{(\hat{\lambda}_j + 1)^2} \sum_{j=1}^p \frac{\hat{\phi}\hat{\lambda}_j}{(\hat{\lambda}_j + 1)^2} + \hat{\alpha}_j^2 \), \( \hat{\alpha} = \Gamma^\top \hat{\beta} \) and \( \text{X}^\top \text{WX} = \Gamma^\top \Gamma^* \).
3 | THE MONTE CARLO EXPERIMENT

To evaluate the different point estimators of the regression parameters $\mathbf{\beta}' = [\beta_1 \cdot \beta_p]$ in the context of a beta regression model, we need to perform a Monte Carlo simulation study under some different contexts. To make the evaluation similar to that in empirical applications, the data generating process is specified with both stochastic error terms and stochastic regressors. This is achieved by generating the data in a two-step procedure. Hence, in each replication of the Monte Carlo experiments, the first step involves generating $n$ independent observations of the regressors which are drawn from a multivariate normal distribution by a pre-specified correlation structure among the regressors. The corresponding mean is then calculated by using the inverse link function $8$ which yields $n$ different mean values. These mean values are then used as parameters in the second step. In the second step, $n$ independent observations of the dependent variable are obtained by conducting $n$ random draw from a beta distribution with fixed dispersion and varying mean based on the mean values generated in step 1. Hence, for the $i$th observation of regressors, the inverse link function $8$ produces the $i$th mean value, which is then used as a parameter in the second step along with a fixed dispersion parameter.

The inverse link function $8$ usually needs to be adjusted marginally to avoid ones and zeros in the observations of the dependent variable ($Y$). We follow the general recommendation (Smithson and Verkuilen20) for solving this problem by rescaling the values of the dependent variable ($Y_i$) by the following:

$$\tilde{Y}_i = Y_i \left( \frac{n-1}{n} \right) + \frac{0.5}{n}.$$  \hspace{1cm} (24)

3.1 | The design of the Monte Carlo experiment

The main property of the data generating process (DGP) is that the dependent variable follows a beta distribution with a fixed dispersion parameter and variable mean.

3.1.1 | The DGP used in the Monte Carlo experiments

The dependent variable is generated according to the following formula:

$$Y_i \sim B(\mu(X_i), \phi),$$  \hspace{1cm} (25)

where $\mu(X_i) = \exp(X_i') \beta (1 + \exp(X_i' \beta))^{-1}$ and $X_i' = [X_{i,1}, \ldots, X_{i,p}]$ is a vector with dimension $p \times 1$ in which we have $p$ different stochastic regressors; more information regarding the specification of the regressors is found below. The variability of the dependent variable is affected by two parts (Equation 2), where $\phi$ is a fixed scalar that partly controls variability; more details about the values used for $\phi$ are available in next subsection.

3.1.2 | The correlation structure among the stochastic regressors

In each Monte Carlo experiment, we generated a $(n \times p)$ matrix ($Z$), where $p$ is the number of regressors obtained by random draw from a multivariate normal distribution with uncorrelated variables and uncorrelated observations; that is, $Z \sim N_p(0, I_p, I_n)$. Hence, the regressors are initially uncorrelated which will be changed in the Monte Carlo experiments. Our choice of the regressors’ correlation matrix is a Toeplitz matrix that is determined by one parameter, $\rho$, as shown below:

$$P_x = \text{Toeplitz}(\rho) = \begin{bmatrix} \rho^0 & \rho^1 & \cdots & \rho^{p-1} \\ \rho^1 & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \rho^1 \\ \rho^{p-1} & \rho^1 & \cdots & \rho^0 \end{bmatrix},$$  \hspace{1cm} (26)
where \( \mathbf{P} \) is the population correlation matrix. By using this type of correlation structure, we can generalize the same correlation pattern to any dimension on the matrix of regressors. Setting \( \rho \) to zero yields the uncorrelated regressors, and for cases in which \( \rho \) is larger than zero, an individual regressor is uncorrelated at the same level as the rest of the regressors (which is more realistic in some empirical applications). The regressors (\( \mathbf{X} \)) are generated by post-multiplying \( \mathbf{Z} \) by the transpose of the matrix root from the Cholesky decomposition of \( \mathbf{P} \) so that they follow the correlation structure in question in the DGP.

### 3.1.3 Motivation for the parameters used in the Monte Carlo experiments

We have chosen three main settings in which the number of regressors in the DGP is equal to 3, 6, and 12. We chose these numbers based on empirical studies analyzing chemical data with the use of more than 12 continuous regressors is unusual. In this study, we used three values for the degrees of freedom \( (df = n - p - 1) \) covering large and small sample sizes: 20, 50, and 100.

We have chosen a relatively large resolution for the \( \rho \) parameter, which controls the correlation structure among the regressors. This is because the main focus of this paper is to investigate how correlated regressors affect the precision of the estimators of \( \mathbf{\beta} \) in the beta regression model and the potential for a Liu estimator under different settings. We use \( \rho = 0.7, 0.8, 0.9, 0.95, 0.99 \).

The beta coefficients were chosen by first setting all coefficients to be one, except for the intercept which was set to zero and then allocating a +/- sign to the coefficient by randomly drawing from a Bernoulli distribution with probability for success equal to 0.5. Finally, the coefficients were adjusted so that the length of the beta vector is always one.

The final parameter in the Monte Carlo experiments is the specification of the stochastic terms in the DGP. It is well known that the performance of ridge and Liu estimators are affected by the signal-to-noise ratio in the DGP. The dispersion parameter is therefore set to 1, 10, and 100, where 1 corresponds to a low signal-to-noise ratio.

We investigated a much richer variety of simulation parameter values, but to keep the results presentable in this paper; we have chosen to present the above values.

### 3.2 Performance criteria for evaluating point estimators of the unknown regression parameters in a beta regression model

To assess how good an estimator is, we first need to determine a loss function that measures how close the estimator is to the unknown parameter. Then, by taking the expectation of the loss function, we obtain the corresponding risk function. We choose the estimator with the smallest risk, based on the risk function. In this paper, we have chosen \( L_2 \) norm as the loss function, which yields the mean squared error (MSE) as the risk function for evaluating the performance of the competing estimators of unknown regression parameters \( \mathbf{\beta}' = [\beta_1, \ldots, \beta_k] \) in the beta regression model. Hence, we will treat \( \rho \) as a nuisance parameter and focus solely on the evaluation of the estimators of \( \mathbf{\beta} \). We choose the \( L_2 \) norm as the loss function (and hence the MSE) because it is one of the most common risk functions and allows us to directly compare our results with other results reported in the literature. Hence, we have following risk function:

\[
\mathcal{R}_2 = \text{MSE}(\hat{\beta}_j) = E_{\hat{\beta}} \left[ \left( \hat{\beta}_j - \beta \right)' \left( \hat{\beta}_j - \beta \right) \right], \text{where} j = 1, 2, 3, 4. \tag{27}
\]

The risk function in 27 is estimated by the corresponding empirical risk function below:

\[
\hat{\mathcal{R}}_2 = R^{-1} \sum_{r=1}^{R} \left( \hat{\beta}_{j,r} - \beta \right)' \left( \hat{\beta}_{j,r} - \beta \right), \text{where} j = 1, 2, 3, 4. \tag{28}
\]

To simplify the comparison of the different Liu estimators to the MLE, we calculate the relative MSE defined as follows:
3.3 Discussion of simulation results

The simulation results for different degrees of freedom, correlation, and dispersion parameters are provided in Table 1a for \( p = 3 \), Table 1b for \( p = 6 \), and Table 1c for \( p = 12 \). All tables show the relative MSE where a value above one indicates that the Liu estimator outperforms the MLE. Increasing the degree of correlation increases the benefit of using the Liu estimator. The estimator \( d_5 \) is nearly always in the optimal choice with the largest decrease in MSE followed by \( d_6 \) that sometimes can yield a lower MSE when the degree of multicollinearity is severe.

This is especially clear when the value of the degrees of freedom is low. The benefit of the Liu estimator decreases with increasing sample size. Finally, the relative gain of applying the Liu estimator increases with the number of regressors; for example, for case of 12 regressors, the gain of applying the Liu method is substantial in the presence of high multicollinearity. Finally increasing the variability increases the relative benefit of applying a Liu estimator, and this pattern is prevailing regardless of the number of regressors used in the simulation study.

In summary, we always recommend the use of a shrinkage estimator in the presence of moderate to strong multicollinearity. The estimators \( d_5 \) and \( d_6 \) perform especially well. These estimators may also be recommended in some situations where there is a low degree of correlation and even sometimes when orthogonal regressors are present. Simulation results for orthogonal regressors are available on request.

4 APPLICATIONS TO CHEMICAL DATA

For illustration purposes, we will consider two real-life applications in this section.

4.1 Gasoline yield data

In this example, we will analyze the classical data set from Prater,\(^{21}\) which was also analyzed by Ospina et al.\(^{11}\) In this data set, the dependent variable is defined as the proportion of crude oil that is converted to petrol after fractional distillation. We use all four regressors originally considered in Prater.\(^{21}\) The first regressor (gravity) is defined as crude oil gravity and measured as the American Petroleum Institute (API) gravity. It is a measure of how heavy or light a petroleum liquid is compared with water. The second regressor is defined as the vapor pressure of crude oil (denoted vapor). We use the temperature (measured in degrees Fahrenheit) at which 10% of the crude oil has vaporized (denoted tempten). Finally, we use the temperature at which all petrol in the amount of crude oil vaporizes (temp). The condition index of the explanatory variables equals 68.1, which indicates a multicollinearity problem. The estimated coefficients for the different estimation methods along with the MSE are shown in Table 2. In some cases, the variable changes sign when applying a Liu-type estimator. However, the effect for that variable is very close to zero, and the variable seems to have no effect on the proportion of crude oil that is converted to petrol after fractional distillation. The MSE decreases when applying a Liu-type estimator, especially when using \( d_2, d_3, d_5, \) and \( d_6 \) estimators. In general, the result confirms the simulation study results, and the \( d_5 \) estimator always minimizes the MSE.

4.2 Color of hazelnuts

In this example, we follow Şimşek,\(^{12}\) where we study the color of hazelnuts. The color of roasted hazelnuts is measured using a colorimeter, which is one of the most widely used methods in chemistry to determine color characteristics. The
| $\varphi$ = 1 | $\varphi$ = 10 | $\varphi$ = 100 |
|---|---|---|
| $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ |
| 0.70 | 1.15 | 1.21 | 1.23 | 1.16 | 1.23 | 1.23 | 1.07 | 1.08 | 1.08 | 1.07 | 1.08 | 1.08 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.80 | 1.24 | 1.37 | 1.42 | 1.25 | 1.43 | 1.42 | 1.11 | 1.13 | 1.13 | 1.11 | 1.13 | 1.14 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.90 | 1.37 | 1.70 | 1.91 | 1.43 | 2.03 | 2.03 | 1.20 | 1.27 | 1.30 | 1.20 | 1.30 | 1.31 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 |
| 0.95 | 1.39 | 1.96 | 2.49 | 1.52 | 3.11 | 3.31 | 1.30 | 1.50 | 1.62 | 1.32 | 1.64 | 1.64 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 |
| 0.99 | 1.17 | 1.91 | 2.73 | 1.39 | 6.16 | 14.31 | 1.35 | 2.00 | 2.71 | 1.50 | 3.82 | 4.31 | 1.23 | 1.32 | 1.36 | 1.23 | 1.36 | 1.36 |

### $\text{Relative MSE, } df = 50$

| $\rho$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.70 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.06 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.80 | 1.09 | 1.10 | 1.11 | 1.09 | 1.11 | 1.12 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.90 | 1.24 | 1.33 | 1.36 | 1.25 | 1.36 | 1.34 | 1.11 | 1.13 | 1.13 | 1.11 | 1.13 | 1.13 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.95 | 1.40 | 1.70 | 1.86 | 1.45 | 1.90 | 1.88 | 1.21 | 1.27 | 1.29 | 1.21 | 1.29 | 1.29 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| 0.99 | 1.37 | 2.26 | 3.14 | 1.64 | 5.10 | 6.11 | 1.42 | 1.91 | 2.29 | 1.52 | 2.51 | 2.54 | 1.14 | 1.15 | 1.16 | 1.14 | 1.16 | 1.16 |

### $\text{Relative MSE, } df = 100$

| $\rho$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.70 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 1.02 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.80 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.04 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.90 | 1.12 | 1.14 | 1.14 | 1.12 | 1.14 | 1.14 | 1.06 | 1.07 | 1.07 | 1.06 | 1.07 | 1.07 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.95 | 1.28 | 1.39 | 1.43 | 1.29 | 1.43 | 1.39 | 1.13 | 1.15 | 1.15 | 1.13 | 1.15 | 1.15 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 |
| 0.99 | 1.48 | 2.23 | 2.82 | 1.68 | 3.53 | 3.62 | 1.37 | 1.62 | 1.77 | 1.40 | 1.79 | 1.79 | 1.08 | 1.08 | 1.08 | 1.08 | 1.08 | 1.08 |
| Table 1b | Estimated relative MSE of proposed estimator compared to MLE when \( p = 6 \) |

**Relative MSE, \( df = 20 \)**

| \( \rho \) | \( \phi = 1 \) | \( \phi = 10 \) | \( \phi = 100 \) |
|---|---|---|---|
| 0.70 | 1.26 1.40 1.40 1.27 1.40 1.36 | 1.10 1.11 1.11 1.10 1.11 1.11 | 1.00 1.00 1.00 1.00 1.00 1.00 |
| 0.80 | 1.35 1.66 1.68 1.37 1.68 1.63 | 1.16 1.17 1.17 1.16 1.17 1.17 | 1.01 1.01 1.01 1.01 1.01 1.01 |
| 0.90 | 1.39 2.31 2.48 1.46 2.54 2.50 | 1.27 1.36 1.36 1.27 1.36 1.36 | 1.03 1.03 1.03 1.03 1.03 1.03 |
| 0.95 | 1.31 2.94 3.58 1.44 4.32 4.37 | 1.37 1.72 1.74 1.39 1.74 1.74 | 1.07 1.07 1.07 1.07 1.07 1.07 |
| 0.99 | 1.10 2.95 4.02 1.28 15.34 21.54 | 1.28 3.10 3.86 1.42 4.83 4.92 | 1.28 1.39 1.39 1.29 1.39 1.39 |

**Relative MSE, \( df = 50 \)**

| \( \rho \) | \( \phi = 1 \) | \( \phi = 10 \) | \( \phi = 100 \) |
|---|---|---|---|
| 0.70 | 1.11 1.11 1.11 1.11 1.09 | 1.05 1.05 1.05 1.05 1.05 | 1.00 1.00 1.00 1.00 1.00 |
| 0.80 | 1.20 1.23 1.23 1.20 1.23 1.18 | 1.08 1.08 1.08 1.08 1.08 | 1.00 1.00 1.00 1.00 1.00 |
| 0.90 | 1.38 1.57 1.58 1.39 1.58 1.48 | 1.15 1.16 1.16 1.15 1.16 | 1.01 1.01 1.01 1.01 1.01 |
| 0.95 | 1.47 2.22 2.28 1.54 2.30 2.21 | 1.26 1.33 1.33 1.26 1.33 | 1.03 1.03 1.03 1.03 1.03 |
| 0.99 | 1.24 3.89 4.96 1.49 8.47 8.82 | 1.44 2.51 2.65 1.54 2.70 | 1.16 1.17 1.17 1.16 1.17 |

**Relative MSE, \( df = 100 \)**

| \( \rho \) | \( \phi = 1 \) | \( \phi = 10 \) | \( \phi = 100 \) |
|---|---|---|---|
| 0.70 | 1.03 1.03 1.03 1.03 1.03 1.02 | 1.02 1.02 1.02 1.02 1.02 | 1.00 1.00 1.00 1.00 1.00 |
| 0.80 | 1.08 1.09 1.09 1.08 1.09 1.06 | 1.04 1.04 1.04 1.04 1.04 | 1.00 1.00 1.00 1.00 1.00 |
| 0.90 | 1.23 1.27 1.27 1.23 1.27 1.19 | 1.08 1.08 1.08 1.08 1.08 | 1.00 1.00 1.00 1.00 1.00 |
| 0.95 | 1.40 1.63 1.63 1.42 1.63 1.51 | 1.16 1.17 1.17 1.16 1.17 | 1.01 1.01 1.01 1.01 1.01 |
| 0.99 | 1.39 3.62 4.17 1.62 4.81 4.78 | 1.44 1.86 1.88 1.48 1.88 | 1.09 1.09 1.09 1.09 1.09 |
| \( \rho \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.70 | 1.32 | 1.56 | 1.56 | 1.33 | 1.56 | 1.50 | 1.13 | 1.14 | 1.14 | 1.13 | 1.14 | 1.14 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 |
| 0.80 | 1.34 | 1.89 | 1.90 | 1.36 | 1.90 | 1.86 | 1.20 | 1.21 | 1.21 | 1.20 | 1.21 | 1.21 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| 0.90 | 1.28 | 2.72 | 2.97 | 1.35 | 2.97 | 2.96 | 1.31 | 1.44 | 1.44 | 1.31 | 1.44 | 1.44 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
| 0.95 | 1.20 | 3.45 | 4.66 | 1.29 | 5.21 | 5.20 | 1.36 | 1.88 | 1.89 | 1.38 | 1.89 | 1.89 | 1.09 | 1.09 | 1.09 | 1.09 | 1.09 |
| 0.99 | 1.06 | 2.95 | 4.71 | 1.15 | 26.59 | 28.42 | 1.19 | 3.64 | 5.01 | 1.29 | 5.65 | 5.65 | 1.32 | 1.46 | 1.46 | 1.32 | 1.46 |

| \( \rho \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.70 | 1.19 | 1.20 | 1.20 | 1.19 | 1.20 | 1.08 | 1.06 | 1.06 | 1.06 | 1.06 | 1.06 | 1.06 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.80 | 1.30 | 1.34 | 1.34 | 1.30 | 1.34 | 1.18 | 1.09 | 1.09 | 1.09 | 1.09 | 1.09 | 1.09 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 |
| 0.90 | 1.45 | 1.77 | 1.77 | 1.48 | 1.77 | 1.61 | 1.19 | 1.19 | 1.19 | 1.19 | 1.19 | 1.19 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| 0.95 | 1.44 | 2.64 | 2.66 | 1.53 | 2.66 | 2.57 | 1.33 | 1.40 | 1.40 | 1.33 | 1.40 | 1.40 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
| 0.99 | 1.16 | 5.17 | 7.39 | 1.35 | 11.43 | 11.44 | 1.37 | 3.04 | 3.13 | 1.47 | 3.14 | 3.14 | 1.20 | 1.21 | 1.21 | 1.20 | 1.21 |

| \( \rho \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( d_4 \) | \( d_5 \) | \( d_6 \) |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.70 | 1.08 | 1.08 | 1.08 | 1.08 | 1.08 | 1.00 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.80 | 1.15 | 1.15 | 1.15 | 1.15 | 1.15 | 1.02 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.90 | 1.32 | 1.36 | 1.36 | 1.33 | 1.36 | 1.14 | 1.10 | 1.10 | 1.10 | 1.10 | 1.10 | 1.10 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 |
| 0.95 | 1.50 | 1.80 | 1.80 | 1.52 | 1.80 | 1.54 | 1.20 | 1.20 | 1.20 | 1.20 | 1.20 | 1.20 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| 0.99 | 1.31 | 4.97 | 5.66 | 1.51 | 5.92 | 5.89 | 1.47 | 2.09 | 2.09 | 1.52 | 2.09 | 2.09 | 1.11 | 1.11 | 1.11 | 1.11 | 1.11 |
measurement of colorimeter can differ, but in this example, it ranges from 0% to 100%. To determine the color characteristics of hazelnuts is a common research question in food chemistry, and it has been studied by, for example, Chiou and Tsai, Demir et al, Özdemir and Devres, Pattee et al, Robbins and Fryer, Sanders et al, and Şimsėk, among others. The changes in color of hazelnuts during roasting are explained by non-enzymatic browning that includes the Maillard reaction, caramelization, pigment destruction, and ascorbic acid browning. In Şimsėk, it is explained by the temperature in the oven time \((x_1)\) and the time \((x_2)\). They used a paraboloid model to investigate the effect of these factors on the color of hazelnuts. The condition index of this model is 665.8, which indicates a severe multicollinearity. From the results presented in Table 3, we can see the same paraboloid effect as was shown in Simsek (1997). However, the fit of the regression model increases substantially when using the new Liu estimator. The MSE of the optimal choice of ridge parameter \(d_5\) and \(d_6\) decreases the MSE almost with 10 times. Hence, it is clear that when modeling these type of relationships, the beta regression is a better choice than the linear regression model employed by Simsek (1997). Furthermore, our new suggested method leads to a substantial improvement of the estimated parameters and the MSE.

### 5 | CONCLUSIONS

Some Liu shrinkage estimators for the beta regression model are proposed in this paper. Since a theoretical comparison among the estimators is not possible, a simulation study was conducted to compare the performance of the estimators. In the design of the experiment, we change several factors, including the number of degrees of freedom, the number of explanatory variables, the dispersion parameter, and the degree of correlation. We find that the new Liu estimators outperform the MLE most of the time, especially when there is high to strong correlation among the explanatory variables. Furthermore, among the proposed shrinkage estimators, \(d_5\) and \(d_6\) performed better than the rest. Therefore, these estimators may be recommended for practitioners since it moderately shrinks the vector of parameters. We analyze real data to illustrate the findings of the paper and find some support for the simulation study results.

**ORCID**

Peter Karlsson [https://orcid.org/0000-0002-3623-5034](https://orcid.org/0000-0002-3623-5034)
Kristofer Månsson [https://orcid.org/0000-0002-4535-3630](https://orcid.org/0000-0002-4535-3630)
B.M. Golam Kibria [https://orcid.org/0000-0002-6073-1978](https://orcid.org/0000-0002-6073-1978)

---

**TABLE 2** Estimated regression coefficients and MSE

| Variable name | ML   | \(d_1\) | \(d_2\) | \(d_3\) | \(d_4\) | \(d_5\) | \(d_6\) |
|---------------|------|---------|---------|---------|---------|---------|---------|
| Gravity       | 0.0045 | 0.0037 | −0.0013 | −0.0013 | 0.0037 | −0.0013 | −0.0019 |
| Vapor         | 0.0304 | 0.0268 | 0.0039  | 0.0039  | 0.0268 | 0.0039  | 0.0038  |
| Tempten       | −0.0110 | −0.0114 | −0.0137 | −0.0137 | −0.0114 | −0.0137 | −0.0133 |
| Temp          | 0.0106 | 0.0106  | 0.0106  | 0.0106  | 0.0106  | 0.0106  | 0.0103  |
| MSE           | 0.0023 | 0.0021  | 0.0009  | 0.0009  | 0.0021  | 0.0009  | 0.0008  |

**TABLE 3** Estimated regression coefficients and MSE

| Variable name | ML   | \(d_1\) | \(d_2\) | \(d_3\) | \(d_4\) | \(d_5\) | \(d_6\) |
|---------------|------|---------|---------|---------|---------|---------|---------|
| \(x_1\)       | −0.0115 | −0.0077 | −0.0122 | −0.0122 | −0.0077 | −0.0122 | −0.0122 |
| \(x_2\)       | −0.0206 | −0.0207 | −0.0212 | −0.0212 | −0.0207 | −0.0212 | −0.0211 |
| \(x_1^2\)     | 0.0000 | 0.0001  | 0.0001  | 0.0001  | 0.0000  | 0.0001  | 0.0001  |
| \(x_2^2\)     | 0.0002 | 0.0002  | 0.0002  | 0.0002  | 0.0002  | 0.0002  | 0.0002  |
| MSE           | 0.0015 | 0.0012  | 0.0002  | 0.0002  | 0.0012  | 0.0002  | 0.0002  |
REFERENCES

1. Ferrari SLP, Cribari-Neto F. Beta regression for modeling rates and proportions. *J Appl Stat*. 2004;31(7):799-815.
2. Simas AB, Barreto-Souza W, Rocha AV. Improved estimators for a general class of beta regression models. *Comput stat data an*. 2010;54(2):348-366.
3. Liu K. A new class of biased estimate in linear regression. *Commun Stat-Theor M*. 1993;22:393-402.
4. Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. *Dent Tech*. 1970;12:55-67.
5. Vigneau E, Devaux MF, Qannari EM, Robert P. Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration. *J Chemometr*. 1998;11:239-249.
6. Muniz G, Kibria BMG. On some ridge regression estimators: an empirical comparisons. *Commun Stat-Simul C*. 2009;38(3):621-630.
7. Algamal ZY. Developing a ridge estimator for the gamma regression model. *J Chemometr*. 2018;32:e3054. https://doi.org/10.1002/cem.3054
8. Kaçiranlar S, Sakallioglu S, Akdeniz F, Styan GPH, Werner HJ. A new biased estimator in linear regression and detailed analysis of the widely-analysed dataset on Portland cement. *Sankhya B*. 1999;61:443-459.
9. Alheety MI, Kibria BMG. Modified Liu-type estimator based on \((r - k)\) class estimator. *Commun Stat-Theor M*. 2013;42(2):304-319.
10. Lukman AF, Ayinde K, Binuomote S, Clement OA. Modified ridge-type estimator to combat multicollinearity: application to chemical data. *J Chemometr*. 2019;33(5):e3125. https://doi.org/10.1002/cem.3125
11. Ospina R, Cribari-Neto F, Vasconcellos K. Improved point and interval estimation for a beta regression model. *Comput Stat Data An*. 2006;51(2):960-981.
12. Şimşek A. The use of 3D-nonlinear regression analysis in mathematics modeling of colour change in roasted hazelnuts. *J Food Eng*. 2007;78(4):1361–1370.
13. Ospina R, Ferrari SLP. A general class of zero-or-one inflated beta regression models. *Comput Stat Data An*. 2012;56(6):1609-1623.
14. Greene WH. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J R Stat Soc B*. 1984;46:149-192.
15. Cribari-Neto F, Vasconcellos KLP. Nearly unbiased maximum-likelihood estimation for the beta distribution. *J Stat Comput Simul*. 2002;72(2):107-118.
16. Cribari-Neto F, Zeilei A. Beta regression in R. *J Stat Softw*. 2010;34:1-24.
17. Kibria BMG. Performance of some new ridge regression estimators. *Commun Stat-Theor M*. 2003;32:419-435.
18. Kibria BMG, Banik S. Some ridge regression estimators and their performances. *J Mod Appl Stat Methods*. 2016;15(1):206-238.
19. Khalaf G, Shukur G. Choosing ridge parameter for regression problems. *Commun Stat-Theor M*. 2005;34(5):1177-1182.
20. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol Methods*. 2006;11(1):54-71.
21. Prater NH. Estimate gasoline yields from crude. *PetroLium Refiner*. 1956;35:236-238.
22. Chiou RY, Tsai TT. Characterization of peanut proteins during roasting as affected by initial moisture content. *J Agric Food Chem*. 1989;37(5):1377–1381.
23. Demir AD, Celayeta JMF, Cronin K, Abodayeh K. Modelling of the kinetics of colour change in hazelnuts during air roasting. *J Food Eng*. 2002;55(4):283–292.
24. Özdemir M, Devres YO. The thin layer drying characteristics of hazelnuts during roasting. *J Food Eng*. 1999;42(4):225–233.
25. Pattee HE, Pearson JL, Young CT, Giesbrecht FG. Changes in roasted peanut flavor and other quality factors with seed size and storage time. *J Food Sci*. 1982;47(2):455–456.
26. Robbins PT, Fryer PJ. The spouted-bed roasting of barley: development of a predictive model for moisture and temperature. *J Food Eng*. 2003;59(2-3):199–208.
27. Sanders TH, Vercellotti JR, Crippen KL, Civille GV. Effect of maturity on roast color and descriptive flavor of peanuts. *J Food Sci*. 1989;54(2):475–477.

How to cite this article: Karlsson P, Månsson K, Kibria BMG. A Liu estimator for the beta regression model and its application to chemical data. *Journal of Chemometrics*. 2020;34:e3300. https://doi.org/10.1002/cem.3300
APPENDIX A: PROOF OF LEMMA 1

Let $\mathbf{X}'\mathbf{WX}$ be a symmetric positive definite matrix and let $\Lambda$ be a diagonal matrix with the $p$ eigenvalues of $\mathbf{X}'\mathbf{WX}$ and let $\Gamma$ be a matrix with corresponding orthonormal eigenvectors.

Then the matrix $\mathbf{X}'\mathbf{WX}$ can be written as $\mathbf{X}'\mathbf{WX} = \Gamma \Lambda \Gamma'$, and based on this, we have

$$
\mathbf{M}'\mathbf{M} = (\mathbf{X}'\mathbf{WX} + d\mathbf{I})(\mathbf{X}'\mathbf{WX} + d\mathbf{I})^{-2}(\mathbf{X}'\mathbf{WX} + d\mathbf{I}) = \Gamma (\Lambda + d\mathbf{I}) \Gamma' (\Lambda + d\mathbf{I})^{-2} \Gamma' \Gamma (\Lambda + d\mathbf{I}) \Gamma' = 
$$

$$
\Gamma (\Lambda + d\mathbf{I})(\Lambda + d\mathbf{I})^{-2}(\Lambda + d\mathbf{I}) \Gamma'.
$$

Since $(\Lambda + d\mathbf{I})(\Lambda + d\mathbf{I})^{-2}(\Lambda + d\mathbf{I}) = \text{diag}\left(\frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2}\right)$, the proof is complete.

APPENDIX B: PROOF OF LEMMA 2

Let $\mathbf{X}'\mathbf{WX}$ be a symmetric positive definite matrix and let $\Lambda$ be a diagonal matrix with the $p$ eigenvalues of $\mathbf{X}'\mathbf{WX}$ and let $\Gamma$ be a matrix with corresponding orthonormal eigenvectors.

Then the matrix $\mathbf{X}'\mathbf{WX}$ can be written as $\mathbf{X}'\mathbf{WX} = \Gamma \Lambda \Gamma'$, and based on this, we have

$$
(\mathbf{M} - \mathbf{I})'(\mathbf{M} - \mathbf{I}) = (\Gamma (\Lambda + d\mathbf{I})^{-1} \Gamma' - \Gamma \Gamma')(\Gamma (\Lambda + d\mathbf{I})^{-1} \Gamma' - \Gamma \Gamma') = 
$$

$$
\Gamma (\Lambda + d\mathbf{I})^{-1}(\Lambda + d\mathbf{I} - \mathbf{I})' (\Lambda + d\mathbf{I})^{-1}(\Lambda + d\mathbf{I} - \mathbf{I}) \Gamma' = \Gamma \text{diag}\left(\frac{\lambda_j + d}{\lambda_j + 1}\right) \text{diag}\left(\frac{\lambda_j + d}{\lambda_j + 1}\right) \Gamma'.
$$

Since $\text{diag}\left(\frac{\lambda_j + d}{\lambda_j + 1}\right) \text{diag}\left(\frac{\lambda_j + d}{\lambda_j + 1}\right) = \text{diag}\left(\frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2}\right)$, the proof is complete.

APPENDIX C: PROOF OF THEOREM 1

In order to prove that there exist a value of $d \in [0,1]$ such that $\Delta \text{MSE}\left(\hat{\beta}_{\text{ML}}, \hat{\beta}(d)\right) = \text{MSE}\left(\hat{\beta}_{\text{ML}}\right) - \text{MSE}\left(\hat{\beta}(d)\right) > 0$, we first note that when $d = 1$, then

$$
\hat{\beta}_{d=1} = (\mathbf{X}'\mathbf{WX} + \mathbf{I})^{-1}(X'\mathbf{WX} + d\mathbf{I}) \hat{\beta}_{\text{ML}}\bigg|_{d=1} = \hat{\beta}_{\text{ML}}.
$$

so that the two estimators coincide causing

$$
\Delta \text{MSE}\left(\hat{\beta}_{\text{ML}}, \hat{\beta}(d = 1)\right) = \text{MSE}\left(\hat{\beta}_{\text{ML}}\right) - \text{MSE}\left(\hat{\beta}(d = 1)\right) = 0.
$$

From 12 and 15 we have that

$$
\Delta \text{MSE}\left(\hat{\beta}_{\text{ML}}, \hat{\beta}(d)\right) = \text{MSE}\left(\hat{\beta}_{\text{ML}}\right) - \text{MSE}\left(\hat{\beta}(d)\right) = \left[\frac{1}{p} \sum_{j=1}^{p} \frac{1}{\lambda_j} \right] - 
$$

$$
\left[\frac{1}{p} \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)^2} + (d - 1)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{\lambda_j (\lambda_j + 1)^2}\right].
$$

Taking the first derivative of C1 with respect to $d$ in the point $d = 1$ gives the following:
\[ \frac{\partial}{\partial d} \Delta \text{MSE}(\hat{\beta}_{\text{ML}}, \hat{\beta}(d)) \bigg|_{d=1} = \frac{\partial}{\partial d} \left( \frac{1}{\varphi} \sum_{j=1}^{p} \frac{1}{\lambda_j} - \frac{1}{\varphi} \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)} - (d-1)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{\lambda_j (\lambda_j + 1)^2} \right) \bigg|_{d=1} = \]

\[ -\left( \frac{1}{\varphi} \sum_{j=1}^{p} \frac{2(\lambda_j + d)}{\lambda_j (\lambda_j + 1)} + 2(d-1) \sum_{j=1}^{p} \frac{\alpha_j^2}{\lambda_j (\lambda_j + 1)^2} \right) \bigg|_{d=1} = -2 \sum_{j=1}^{p} \frac{\lambda_j + 1}{\lambda_j (\lambda_j + 1)^2} = -2 \sum_{j=1}^{p} \frac{1 + \lambda_j^{-1}}{\lambda_j (\lambda_j + 1)^2}. \]

(C2)

Since \( \phi > 0 \) and \( \lambda_j > 0 \) \( \forall \ j = 1, \cdots, p \), we have that each term in the summation in C2 is larger than zero, so that

\[ \frac{\partial}{\partial d} \Delta \text{MSE}(\hat{\beta}_{\text{ML}}, \hat{\beta}(d)) \bigg|_{d=1} < 0. \]

(C3)

Finally, by C3, we have that as \( d \) decrease from \( d = 1 \), \( \Delta \text{MSE}(\hat{\beta}_{\text{ML}}, \hat{\beta}(d)) \) increases and the proof is complete.