Ridge Regression and Generalized Maximum Entropy: an improved version of the Ridge-GME parameter estimator

Pedro Macedo

CIDMA – Center for Research and Development in Mathematics and Applications, Department of Mathematics, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

pmacedo@ua.pt

In this paper, the Ridge-GME parameter estimator, which combines Ridge Regression and Generalized Maximum Entropy, is improved in order to eliminate the subjectivity in the analysis of the ridge trace. A serious concern with the visual inspection of the ridge trace to define the supports for the parameters in the Ridge-GME parameter estimator is the misinterpretation of some ridge traces, in particular where some of them are very close to the axes. A simulation study and two empirical applications are used to illustrate the performance of the improved estimator. A MATLAB code is provided as supplementary material.

Keywords: generalized maximum entropy, ridge regression, shrinkage estimation

MSC 2010: 62J07, 94A17

---

Final version of this paper is published in Communications in Statistics – Simulation and Computation, 46(5), 3527-3539, 2017 (http://dx.doi.org/10.1080/03610918.2015.1096378).
1 Introduction

It is well-known that under fairly regularity conditions, linear regression becomes a powerful statistical tool. In practice, however, some of these conditions are usually not satisfied and regression models become ill-posed, implying that the application of traditional estimation methods, such as the Ordinary Least Squares (OLS) or the Maximum Likelihood (ML) estimators, may lead to non-unique or highly unstable solutions. One of the major concerns in multiple linear regression is the collinearity\(^2\) problem, which is responsible for inflating the variance associated with the regression coefficients estimates\(^3\) and, in general, to affect the signs of the estimates, as well as statistical inference. Common examples arise from many fields including astronomy, chemistry, medicine and economics, where it is necessary to make the best possible predictions with limited and noisy information.

Various methods, including Ridge Regression, Principal Component Regression, Partial Least Squares Regression, Continuum Regression, Lasso, Elastic Net, Least Angle Regression and Generalized Maximum Entropy are well suited to cope with collinearity problems; see, among others, Brown (1994), Ciavolino and Al-Nasser (2009), Efron et al. (2004), Golan (2006), Hoerl and Kennard (1970b), Satici and Demirhan (2012), Stone and Brooks (1990), Tibshirani (1996) and Zou and Hastie (2005) for reviews. However, it is interesting to note that, despite some more recent approaches, Ridge Regression continues to play a key role in regression models affected by collinearity and outperforms other competitors in many cases; e.g., Hastie et al. (2009), Maronna (2011), McDonald (2009), Zou and Hastie (2005). The importance of Ridge Regression is discussed by McDonald (2009) that analyzed the number of publications in the Technometrics, the Journal of the American Statistical Association, the Communications in Statistics – Theory and Methods, and the Communications in Statistics – Simulation and Computation. Approximately 320 articles related to Ridge Regression has been published in these four scientific journals since the seventies.\(^4\)

\(^2\)Naturally, this notion is not used here in a literal sense; e.g., Belsley et al. (2004, pp. 85–98).

\(^3\)Small sample size (micronumerosity) is also responsible for the inflation of variance.

\(^4\)This number of publications results from a recent update based on the study of McDonald (2009).
The Ridge Regression introduced by Hoerl and Kennard (1970b) is an estimation procedure to handle collinearity without removing variables from the regression model. By adding a small non-negative constant (tuning, ridge or shrinkage parameter) to the diagonal of the correlation matrix of the explanatory variables, it is possible to reduce the variance of the OLS estimator through the introduction of some bias. Although the resulting estimators are biased, the biases are small enough for these estimators to be substantially more precise than the unbiased estimators. Regarding this usual concern between the use of biased and unbiased estimators, it is important to note that OLS and ML may also be biased; see, for example, Ryan (2009, p. 466) and Greene (2008, p. 487). Furthermore, and considering other criticisms on Ridge Regression, it is worth to mention that ML estimator may not be unique and it is attractive mainly due to the large-sample properties. Moreover, considering \( \beta \) as the vector of unknown parameters in a regression model, OLS considers that \( \beta' \beta \) is unbounded. Naturally, \( \beta' \beta \) is unknown but should be finite, which means that assuming an upper bound on \( \beta' \beta \) and applying OLS subject to this restriction, the general form of the Ridge Regression is obtained (Ryan, 2009, p. 470). Thus, it is difficult to argue against Ridge Regression and the usual criticisms in choosing biased estimators may not be always totally fair.

The challenge in Ridge Regression remains on the selection of the ridge parameter. One straightforward approach is based on simply plotting the coefficients against several possible values for the ridge parameter and inspecting the resulting traces; e.g., Hoerl and Kennard (1970b,a), Zhang and McDonald (2005). In addition to this graphical procedure, several formal methods have been proposed to estimate the ridge parameter. The literature on methods for choosing the ridge parameter is massive and includes, among many others, Alkhamisi and Shukur (2007), Gibbons (1981), Golub et al. (1979), Hoerl and Kennard (1970b,a), Hoerl et al. (1975), Khalaf and Shukur (2005), Kibria (2003), McDonald and Galarneau (1975) and Muniz and Kibria (2009).

Recently, based on the Generalized Maximum Entropy (GME) estimator, the Ridge-GME parameter estimator (Macedo et al., 2010) appears in the literature as one interesting
approach. The use of the GME estimator tries to solve the problem of the subjective sele-
ction of the tuning parameter based on the visual inspection of the ridge trace, and the
Ridge Regression approach provides information for the selection of the supports for the
parameters in the regression model, which is the main drawback of the GME estimator.
Thus, by circumventing criticisms and difficulties of traditional Ridge Regression and GME
estimation, the Ridge-GME parameter estimator seems to be a perfect combination between
the estimators that highlights the best of both.

However, the Ridge-GME parameter estimator requires the use of subjective information
from the analysis of the ridge trace, which represents the main drawback of the estimator.
In this paper, the Ridge-GME parameter estimator is developed so that no subjective in-
formation from the visual inspection of the ridge trace is needed. Furthermore, a MATLAB
code with the improved Ridge-GME parameter estimator is provided.

The remainder of the paper is laid out as follows: in Section 2, a brief overview on Ridge
Regression and GME estimation is presented. In Section 3, the improvement on the Ridge-
GME parameter estimator is discussed. A simulation study is provided in Section 4. In
Section 5, two empirical applications are accomplished. Finally, some concluding remarks
are given in Section 6.

2 Ridge Regression and Generalized Maximum Ent-
tropy

Consider the multiple linear regression model defined by

$$y = X\beta + u,$$

where $y$ denotes a $(N \times 1)$ vector of noisy observations, $\beta$ is a $(K \times 1)$ vector of unknown
parameters, $X$ is a known $(N \times K)$ matrix of explanatory variables and $u$ is a $(N \times 1)$ vector
of random disturbances (errors), usually assumed to have a conditional expected value of
zero and representing spherical disturbances, i.e, $E[u|X] = 0$ and $E[uu'|X] = \sigma^2 I$, where
\(I\) is a \((N \times N)\) identity matrix and \(\sigma^2\) is the error variance.

The OLS and the Ridge Regression estimators of \(\beta\) are well-known.

**Definition 2.1.** The Ordinary Least Squares (OLS) estimator of \(\beta\) in model (1) is given by

\[
\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'y. \tag{2}
\]

**Definition 2.2.** The Ridge Regression estimator of \(\beta\) in model (1) takes the form

\[
\hat{\beta}_{\text{ridge}} = (X'X + \eta I)^{-1}X'y, \tag{3}
\]

where \(\eta \geq 0\) denotes the ridge parameter and \(I\) is a \((K \times K)\) identity matrix.

Definition 2.2 represents the biased estimator proposed by Hoerl and Kennard (1970b). Note that when \(\eta \to 0\), the Ridge Regression estimator approaches the OLS estimator whereas the Ridge Regression estimator approaches the zero vector when \(\eta \to \infty\). Thus, a trade-off between variance and bias is needed. Hoerl and Kennard (1970b, pp. 62–63) proved that the Ridge Regression estimator is superior to the OLS estimator (in a Mean Squared Error sense) for a range of values of \(\eta\), say \(0 < \eta < \sigma^2/\alpha^2_{\text{max}}\), where \(\alpha^2_{\text{max}}\) is the largest squared value from a vector \(\alpha\) that depends on \(\beta\) (considering model (1) in the canonical form). Thus, is the problem of choosing \(\eta\) solved? Unfortunately the answer is no, because \(\sigma^2\) and \(\beta\) are unknown.

**Definition 2.3.** The Mean Squared Error (MSE) of the Ridge Regression estimator of \(\beta\) in model (1) is given by

\[
\text{MSE}(\hat{\beta}_{\text{ridge}}) = \sigma^2 \sum_{k=1}^{K} \frac{\lambda_k}{(\lambda_k + \eta)^2} + \eta^2 \beta'(X'X + \eta I)^{-2} \beta, \tag{4}
\]

where \(\lambda_k\) are the eigenvalues of the \(X'X\) matrix in correlation form.

It is interesting to note that the variance of \(\hat{\beta}_{\text{ridge}}\) decreases as \(\eta\) increases, whereas the bias increases with \(\eta\). Due to its biasedness, the Ridge Regression estimator is superior to the OLS estimator (in a MSE sense) if the reduction in the variance is greater than the increase...
of the squared bias. Since the range of values for which the Ridge Regression estimator is
superior to the OLS estimator depends on the unknown parameters $\beta$ and $\sigma^2$, the challenge
is to select an estimate of $\eta$ such that the Ridge Regression estimator has a smaller MSE
than the OLS estimator.

To better understand the drawbacks of the OLS estimator under collinearity, consider
the expected squared distance between $\hat{\beta}_{\text{OLS}}$ and $\beta$, defined as

$$E \left[ (\hat{\beta}_{\text{OLS}} - \beta)'(\hat{\beta}_{\text{OLS}} - \beta) \right]. \quad (5)$$

Considering $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K > 0$ as the ordered eigenvalues of the $X'X$ matrix in
correlation form, it follows that

$$E \left[ (\hat{\beta}_{\text{OLS}} - \beta)'(\hat{\beta}_{\text{OLS}} - \beta) \right] = \sigma^2 \sum_{k=1}^{K} \frac{1}{\lambda_k} \quad (6)$$

and

$$E \left[ \hat{\beta}_{\text{OLS}}'\hat{\beta}_{\text{OLS}} \right] = \beta'\beta + \sigma^2 \sum_{k=1}^{K} \frac{1}{\lambda_k}. \quad (7)$$

Thus, as $\lambda_K$ becomes smaller (leading to an increase in collinearity), the vector $\hat{\beta}_{\text{OLS}}$ can be
expected to be farther from the vector $\beta$; e.g., Hoerl and Kennard (1970a).

As noted by Golan et al. (1996), statistical data are frequently limited and affected by
collinearity implying that the associated statistical models may be ill-posed, unless simplifying
assumptions/procedures are imposed to generate seemingly well-posed statistical models
that can be estimated with traditional statistical tools. Giving heed to this problem, Golan
et al. (1996) generalized the Maximum Entropy formalism (Jaynes, 2003) to linear inverse
problems with noise, expressed in model (1). The idea is to treat each $\beta_k$ as a discrete
random variable with a compact support and $2 \leq M < \infty$ possible outcomes, and each $u_n$
as a finite and discrete random variable with $2 \leq J < \infty$ possible outcomes. Assuming that
both the unknown parameters and the unknown error terms may be bounded a priori, the
linear model (1) can be presented as

$$y = XZp + Vw, \quad (8)$$
where

$$\beta = Zp = \begin{bmatrix} z_1' & 0 & \cdots & 0 \\ 0 & z_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_K' \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix},$$  \tag{9}

with \(Z\) a \((K \times KM)\) matrix of support values and \(p\) a \((KM \times 1)\) vector of unknown probabilities, and

$$u = Vw = \begin{bmatrix} v_1' & 0 & \cdots & 0 \\ 0 & v_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_N' \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix},$$  \tag{10}

with \(V\) a \((N \times NJ)\) matrix of support values and \(w\) a \((NJ \times 1)\) vector of unknown probabilities.

**Definition 2.4.** For the linear regression model specified in (1), the Generalized Maximum Entropy (GME) estimator is given by

$$\arg\max_{p, w} \{ -p' \ln p - w' \ln w \},$$  \tag{11}

subject to the model constraint

$$y = XZp + Vw,$$  \tag{12}

and the additivity constraints for \(p\) and \(w\), respectively,

$$1_K = (I_K \otimes 1_M)p,$$
$$1_N = (I_N \otimes 1_J)w,$$  \tag{13}

where \(\otimes\) represents the Kronecker product, \(1\) is a column vector of ones with a specific dimension, \(I\) is an identity matrix with a specific dimension and, as defined in (9) and (10), \(Z\) and \(V\) are the matrices of supports, and \(p > 0\) and \(w > 0\) are probability vectors to be estimated.

The GME estimator generates the optimal probability vectors \(\hat{p}\) and \(\hat{w}\) that can be used to form point estimates of the unknown parameters and the unknown random errors through
the reparameterizations (9) and (10), respectively. As noted by Golan et al. (1996), since the objective function (11) is strictly concave in the interior of the additivity constraint set, a unique solution for the GME estimator is guaranteed if the intersection of the model and the additivity constraint sets is non-empty.

The supports in matrices $Z$ and $V$ are defined as being closed and bounded intervals within which each parameter or error is restricted to lie, implying that researchers need to provide exogenous information (which, unfortunately, it is not always available). This is considered the main weakness of the GME estimator; see, for example, Caputo and Paris (2008) for further details. Golan et al. (1996) discuss these issues in the case of minimal prior information: for the unknown parameters, the authors recommend the use of wide bounds (this is naturally subjective) for the supports in $Z$, without extreme risk consequences; for the unknown errors, the authors suggest the use of the three-sigma rule with a sample scale parameter. The number of points, $M$ and $J$, in the supports is less controversial and are usually used in the literature between 3 and 7 points, since there is likely no significant improvement in the estimation with more points in the supports. The three-sigma rule, considering the standard deviation of the noisy observations and $J = 3$ points in the supports, is adopted in the improved Ridge-GME parameter estimator discussed in Section 3.

### 3 Improved Ridge-GME parameter estimator

The Ridge-GME parameter estimator was introduced by Macedo et al. (2010) and was adapted for a jackknife procedure by Erdugan and Akdeniz (2012). The basic idea underlying the Ridge-GME parameter estimator is to combine the ridge trace and the GME estimator. The two key issues are: how can the information provided by the ridge trace be used without making a subjective selection of an estimate of $\eta^2$, and how can the ridge trace and the GME estimator be efficiently combined? In answering to these questions, Macedo et al. (2010) illustrate how the Ridge-GME parameter estimator can be implemented. Moreover, based on the results from the simulation studies conducted by Macedo et al. (2010) and Erdugan
and Akdeniz (2012), it seems reasonable to state that, in the case of regression models with small samples sizes affected by collinearity, the Ridge-GME parameter estimator has a good performance and it may be recommended to practitioners.

However, the original Ridge-GME parameter estimator has two drawbacks: it requires subjective information from visual inspection of the ridge trace, namely the definition of the ridge interval (a set of possible values for the ridge parameter) and it requires the supports for the parameters of the regression model. These drawbacks are overcome with the improved version of the Ridge-GME parameter estimator discussed next.

First, considering the definition of the ridge interval, since it is widely accepted in the literature that an estimate of the ridge parameter must be selected within the interval \([0, 1]\), the improved Ridge-GME parameter estimator considers this interval by default, with 1000 possible values for the ridge parameter. Naturally, the user has always the option to define another ridge interval as a subset of the default interval \([0, 1]\).\(^5\)

Second, considering the definition of the supports for the parameters, Figure 1 illustrates the improved Ridge-GME parameter estimator, where the subjective choice of the supports for the parameters based on the visual inspection of the ridge trace is eliminated. In the original Ridge-GME parameter estimator, the ridge trace provides “guidelines” for the selection of the supports for the model regression parameters, which leads to subjective choices for the supports.\(^6\) A serious concern with the visual inspection of the ridge trace to define the supports for the parameters is the misinterpretation of some ridge traces, in particular where some of them overlap (or are very close to) the axes. For example, this difficulty occurs in both empirical applications discussed in Section 5: in Figure 2 one trace exceeds the value \(-2\) and in Figure 3 one trace exceeds the value 62. In both cases, the limits of these traces are very difficult to obtain only by visual inspection of the ridge traces.

\(^5\)In this case, the ridGME code (the MATLAB code provided as supplementary material) presents the ridge trace to help the user in the selection of the ridge interval.

\(^6\)For example, based on the same ridge trace for the Portland cement model, different supports for the parameters are considered by Macedo et al. (2010) and Erdugan and Akdeniz (2012).
Definition 3.1. Considering the vector $\hat{p}$ obtained through the GME estimator, where the matrix $Z$ is a $(K \times 5K)$ matrix of support values with limits given by

$$z_k = \left[\min \left\{ \hat{\beta}_{k,\text{ridge}}(\eta) \right\}, \max \left\{ \hat{\beta}_{k,\text{ridge}}(\eta) \right\} \right],$$

for all the 1000 possible values of $\eta$ in the ridge interval, $[\eta_1, \eta_{1000}]$, which is $[0, 1]$ by default, the improved Ridge-GME parameter estimator is given by

$$\hat{\eta} = \arg\min_{\eta} \left\| Z\hat{p} - (X'X + \eta I)^{-1}X'y \right\|_\infty$$

subject to the specified ridge interval.

With the improved Ridge-GME parameter estimator, given the ridge interval, $[\eta_1, \eta_{1000}]$, the supports for the parameters are no longer subjectively defined.

Figure 1: Ridge interval and supports in the improved Ridge-GME parameter estimator.
4 Simulation study

A similar simulation study to the one conducted by Macedo et al. (2010) is provided in this section to compare the original Ridge-GME parameter estimator with the improved version proposed in Definition 3.1. Results from the OLS estimator are also provided for comparison.

The five explanatory variables are generated through

\[ x_{ij} = (1 - \alpha^2)^{\frac{1}{2}} z_{ij} + \alpha z_{i6}, \]

where \( z_{ij} \) are independent standard normal pseudo-random numbers and \( \alpha \) is specified so that the correlation between any two explanatory variables is given by \( \alpha^2 \); e.g., Gibbons (1981), Kibria (2003) and McDonald and Galarneau (1975). Four different combinations of sample sizes are considered, namely \( n = 10, 20, 50, 100 \). By choosing the true coefficient vector \( \beta \) as the normalized eigenvector corresponding to the largest eigenvalue of the \( X'X \) matrix, the \( n \) observations for the dependent variable are obtained by

\[ y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + u_i, \]

where \( u_i \) are independent normal pseudo-random numbers with zero mean and variance \( \sigma^2 \). Three different values for \( \sigma \) and five different values for \( \alpha \) are tested, namely \( \sigma = 0.5, 1.0, 1.5 \) and \( \alpha = 0.750, 0.900, 0.950, 0.975, 0.999 \).

The ridge interval for the original Ridge-GME parameter estimator is defined as \([0, 1]\) and the GME estimator is performed using different supports, namely \([-5, 5]\) and \([-10, 10]\), for the five parameters of the model, illustrating the possible subjectivity in the selection of the supports. The 3\( \sigma \) rule is used to define the supports for the error component, where \( \sigma \) represents the empirical standard deviation of the noisy observations. The number of support points used for the unknown parameters and the error component is 5 and 3, respectively. The improved version of the Ridge-GME parameter estimator is implemented according to Definition 3.1, where the ridge interval is considered as \([0, 1]\) by default.

Since the variables were standardized, the estimated standardized coefficients are transformed back to the original model and for the 1000 trials performed, the mean squared
error loss (MSEL), with \( \text{SEL} \left( \hat{\beta} \right) = \left( \| \hat{\beta} - \beta \|_2 \right)^2 \), is the measure used to evaluate the performance of the three estimators.

Table 1: MSEL for OLS and Ridge Regression \((n = 10 \text{ and } n = 20)\).

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(\sigma = 0.5)</th>
<th>(\sigma = 1.0)</th>
<th>(\sigma = 1.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 10)</td>
<td>(n = 20)</td>
<td>(n = 10)</td>
</tr>
<tr>
<td>OLS</td>
<td>Ridge-GME</td>
<td>Ridge-GME</td>
<td>OLS</td>
</tr>
<tr>
<td>0.750</td>
<td>0.4560</td>
<td>0.2601</td>
<td>0.0892</td>
</tr>
<tr>
<td>0.900</td>
<td>1.8412</td>
<td>0.3757</td>
<td>0.0717</td>
</tr>
<tr>
<td>0.950</td>
<td>2.0110</td>
<td>0.2408</td>
<td>0.0439</td>
</tr>
<tr>
<td>0.975</td>
<td>4.1578</td>
<td>0.3749</td>
<td>0.0718</td>
</tr>
<tr>
<td>0.999</td>
<td>69.7136</td>
<td>0.1243</td>
<td>0.9997</td>
</tr>
</tbody>
</table>

Table 1 and Table 2 presents the results for OLS and Ridge Regression with the original Ridge-GME parameter estimator, considering the support \([-10, 10]\) for the five parameters.
Table 2: MSEL for OLS and Ridge Regression \((n = 50 \text{ and } n = 100)\).

<table>
<thead>
<tr>
<th>(\alpha )</th>
<th>OLS</th>
<th>Ridge-GME</th>
<th>Ridge-GME</th>
<th>OLS</th>
<th>Ridge-GME</th>
<th>Ridge-GME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma = 0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.750</td>
<td>0.0622</td>
<td>0.0428</td>
<td>0.0572</td>
<td>0.0399</td>
<td>0.0259</td>
<td>0.0396</td>
</tr>
<tr>
<td>0.900</td>
<td>0.1217</td>
<td>0.0616</td>
<td>0.0395</td>
<td>0.0549</td>
<td>0.0277</td>
<td>0.0371</td>
</tr>
<tr>
<td>0.950</td>
<td>0.2433</td>
<td>0.0811</td>
<td>0.0348</td>
<td>0.1111</td>
<td>0.0369</td>
<td>0.0331</td>
</tr>
<tr>
<td>0.975</td>
<td>0.4971</td>
<td>0.0577</td>
<td>0.0299</td>
<td>0.1949</td>
<td>0.0347</td>
<td>0.0307</td>
</tr>
<tr>
<td>0.999</td>
<td>14.2421</td>
<td>0.0183</td>
<td>0.2319</td>
<td>5.2173</td>
<td>0.0064</td>
<td>0.0967</td>
</tr>
<tr>
<td>(\sigma = 1.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.750</td>
<td>0.2586</td>
<td>0.1899</td>
<td>0.0868</td>
<td>0.0950</td>
<td>0.0734</td>
<td>0.0710</td>
</tr>
<tr>
<td>0.900</td>
<td>0.5576</td>
<td>0.3140</td>
<td>0.0594</td>
<td>0.2272</td>
<td>0.1289</td>
<td>0.0466</td>
</tr>
<tr>
<td>0.950</td>
<td>1.0015</td>
<td>0.3124</td>
<td>0.0410</td>
<td>0.4943</td>
<td>0.1576</td>
<td>0.0365</td>
</tr>
<tr>
<td>0.975</td>
<td>2.2038</td>
<td>0.3022</td>
<td>0.0416</td>
<td>0.8207</td>
<td>0.1535</td>
<td>0.0310</td>
</tr>
<tr>
<td>0.999</td>
<td>48.4125</td>
<td>0.0373</td>
<td>0.8342</td>
<td>21.9658</td>
<td>0.0204</td>
<td>0.3562</td>
</tr>
<tr>
<td>(\sigma = 1.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.750</td>
<td>0.4922</td>
<td>0.3411</td>
<td>0.0861</td>
<td>0.2495</td>
<td>0.1783</td>
<td>0.0699</td>
</tr>
<tr>
<td>0.900</td>
<td>1.2959</td>
<td>0.6825</td>
<td>0.0764</td>
<td>0.5530</td>
<td>0.2898</td>
<td>0.0549</td>
</tr>
<tr>
<td>0.950</td>
<td>2.3680</td>
<td>0.7513</td>
<td>0.0632</td>
<td>0.9839</td>
<td>0.2881</td>
<td>0.0372</td>
</tr>
<tr>
<td>0.975</td>
<td>3.7130</td>
<td>0.8399</td>
<td>0.0677</td>
<td>1.9774</td>
<td>0.3450</td>
<td>0.0406</td>
</tr>
<tr>
<td>0.999</td>
<td>78.4112</td>
<td>0.1075</td>
<td>1.1941</td>
<td>46.1627</td>
<td>0.0520</td>
<td>0.7002</td>
</tr>
</tbody>
</table>

As expected, in terms of MSEL, both Ridge Regression procedures outperform the OLS estimator in all the scenarios. Concerning the two Ridge Regression procedures, in the 60 simulations performed, the Ridge Regression estimator with the improved version of
the Ridge-GME parameter estimator outperforms the Ridge Regression estimator with the original version in 45 scenarios, in terms of MSEL. In general, the worst results with the improved version of the Ridge-GME parameter estimator are obtained only for $\alpha = 0.999$, regardless the value of $n$ and $\sigma$, which represents the cases of extreme severe collinearity.

This simulation study reveals that the performance of the original Ridge-GME parameter estimator depends on the supports for the parameters. For example, considering some supports of higher amplitude than with $[-10, 10]$, the Ridge Regression estimator with the improved version of the Ridge-GME parameter estimator outperforms the Ridge Regression estimator with the original version in all the 60 simulations. However, for example, considering the support $[-5, 5]$ for the five parameters of the model, the Ridge Regression estimator with the improved version of the Ridge-GME parameter estimator outperforms the Ridge Regression estimator with the original version only in 38 of the 60 simulations.\(^7\)

These results highlight the concerns with the visual inspection of the ridge trace to define the supports for the parameters in the original Ridge-GME parameter estimator. With the improved version proposed in Definition 3.1, the supports for the parameters are no longer arbitrarily defined and seem to be appropriately specified in most of the cases considered in this simulation study.

5 Empirical applications

In this section, the well-known acetylene (Marquardt and Snee, 1975) and the Portland cement (Woods et al., 1932) models are used to illustrate the performance of the improved Ridge-GME parameter estimator. These data sets have received considerable attention in the literature; e.g., Montgomery et al. (2006), Hald (1952), Kaçiranlar et al. (1999), Liu (2003), Muniz and Kibria (2009), Sakallıoğlu and Kaçiranlar (2008).\(^7\)

\(^7\)Results not reported here due to space limitations are provided upon request to the author.
5.1 Acetylene model

The response variable in the acetylene model is the percentage of conversion of \( n \)-heptane to acetylene (\( y \)) and the three explanatory variables are the reactor temperature (\( x_1 \)), the ratio of \( H_2 \) to \( n \)-heptane (\( x_2 \)) and the contact time (\( x_3 \)). The data set is composed by 16 observations. Following Montgomery et al. (2006, p. 329), the full quadratic model usually considered for the acetylene data is

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1^2 + \beta_8 x_2^2 + \beta_9 x_3^2 + u. \tag{18}
\]

The model is affected by severe collinearity (\( \text{cond}_2 X \approx 208 \))\(^8\) which means that OLS estimator should not be used. The ridge trace in Figure 2 illustrates the lack of stability of the OLS estimates and the presence of severe collinearity, i.e., large changes in the coefficients for small values of the ridge parameter.

![Ridge Trace](image)

**Figure 2:** Ridge trace for the acetylene model.

From visual inspection of the ridge trace in Figure 2, Montgomery et al. (2006, p. 347) consider that reasonable coefficient stability is achieved in the region established by \( 0.008 < \)

\(^8\)Ratio of the largest singular value of the design matrix, \( X \), with the smallest singular value (condition number). The regressors and the response are centered and scaled.
\( \eta < 0.064 \) and, without a rule to choose the ridge parameter, the authors select \( \eta = 0.032 \) for the Ridge Regression model, obtaining a residual mean square value of, approximately, \( 9.4 \times 10^{-4} \).

Considering the original Ridge-GME parameter estimator (Macedo et al., 2010), from visual inspection of the ridge trace in Figure 2, the ridge interval can be subjectively defined, for example, by \( \eta \in [0, 0.2] \), and different supports can be subjectively specified for the parameters of the model, corresponding to different estimates for the ridge parameter (different combinations of supports were tested in this model, considering equal supports for all the parameters and also considering specific supports for each parameter). The residual mean square values obtained varies between \( 6.1 \times 10^{-4} \) and \( 4.0 \times 10^{-3} \), approximately.

Finally, considering the improved Ridge-GME parameter estimator without any subjective information (considering option 1 in the ridGME code, i.e., considering \( \eta \in [0, 1] \) by default), the estimated ridge parameter is \( \eta = 0.001 \) and the corresponding residual mean square value is, approximately, \( 4.7 \times 10^{-4} \). It is interesting to note that considering, for example, the specific ridge interval defined by Montgomery et al. (2006, p. 347), the estimated ridge parameter is \( \eta = 0.0181 \) (considering option 2 in the ridGME code) and the corresponding residual mean square value is, approximately, \( 7.7 \times 10^{-4} \). In this case, the main advantage of the improved Ridge-GME parameter estimator is that no kind of judgment is required to select an estimate of \( \eta \) in the interval \( [0.008, 0.064] \).

Thus, with lower residual mean square values and, most importantly, with no need for any judgement in the interpretation of the ridge trace, the improved Ridge-GME parameter estimator seems to be an attractive approach.

5.2 Portland cement model

The response variable in this model is the heat evolved per gram of cement \( (y) \) and the four explanatory variables are the amounts of tricalcium aluminate \( (x_1) \), tricalcium silicate \( (x_2) \), tetracalcium aluminoferrite \( (x_3) \) and \( \beta \)-dicalcium silicate \( (x_4) \). The linear model without intercept presented by Woods et al. (1932) does not suffer from collinearity because
cond_2 X \approx 21$, where $X$ is the matrix of the explanatory variables. However, the linear model with intercept defined as

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \beta_4 x_{n4} + u_n,$$

(19)

$n = 1, 2, \ldots, 13$, is affected by severe collinearity since $\text{cond}_2 X \approx 6056$, with $X$ representing the matrix of the explanatory variables with the first column of ones. As noted by Liu (2003), this dramatic change in $\text{cond}_2$ is explained by the fact that the sum of each row in the original $X$ matrix is approximately equal to 100 (the explanatory variables are presented in rounded percentages) and, thus, the model (19) is affected by severe collinearity.

![Ridge trace for the Portland cement model.](image)

**Figure 3:** Ridge trace for the Portland cement model.

The ridge trace in Figure 3 illustrates the lack of stability of the OLS estimates and the presence of severe collinearity. From visual inspection of the ridge trace in Figure 3, the ridge interval was defined as $\eta \in [0, 0.3]$ by Macedo et al. (2010). Table 3 shows the MSE for the OLS and the Ridge Regression estimator based on the improved Ridge-GME parameter estimator without any subjective information (option 1 in the ridGME code) and using the same ridge interval defined by Macedo et al. (2010) (option 2 in the ridGME code), the generalized cross-validation (GCV) estimator by Golub et al. (1979), the HK estimator
proposed by Hoerl and Kennard (1970b,a), the HKB estimator by Hoerl et al. (1975), the KS estimator by Khalaf and Shukur (2005), and the KM4, KM5 and KM6 estimators by Muniz and Kibria (2009).

**Table 3:** MSE for different estimators in the Portland cement model.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>4912</td>
</tr>
<tr>
<td>Ridge with KM5</td>
<td>3882</td>
</tr>
<tr>
<td>Ridge with KM4</td>
<td>3879</td>
</tr>
<tr>
<td>Ridge with KM6</td>
<td>3876</td>
</tr>
<tr>
<td>Ridge with GCV</td>
<td>3859</td>
</tr>
<tr>
<td>Ridge with HKB</td>
<td>2990</td>
</tr>
<tr>
<td>Ridge with Ridge-GME</td>
<td>2704</td>
</tr>
<tr>
<td>Ridge with improved Ridge-GME (option 1)</td>
<td>2272</td>
</tr>
<tr>
<td>Ridge with improved Ridge-GME (option 2)</td>
<td>2205</td>
</tr>
<tr>
<td>Ridge with KS</td>
<td>2180</td>
</tr>
<tr>
<td>Ridge with HK</td>
<td>2171</td>
</tr>
</tbody>
</table>

The MSE for the Ridge Regression estimator using the improved Ridge-GME parameter estimator is slightly greater than the MSE for the Ridge Regression estimator based on the HK and KS estimates, although it is lower than the MSE for the OLS and the other Ridge Regression estimators presented in Table 3, including the Ridge Regression estimator using the original Ridge-GME parameter estimator. The Ridge Regression estimator using the improved Ridge-GME parameter estimator also provides a MSE less than the MSE of other Ridge Regression estimators considered by Muniz and Kibria (2009, pp. 628–629) and of some estimators considered by Sakallıoğlu and Kaçıranlar (2008, pp. 683–687) in the Portland cement model.
6 Conclusions

The Ridge-GME parameter estimator, by avoiding criticisms and difficulties of the traditional Ridge Regression literature and the GME estimation, seems to be a perfect combination that highlights the best of both methodologies. The Ridge-GME parameter estimator tries to solve the problem of the subjective selection of the tuning parameter by visual inspection of ridge trace in the traditional Ridge Regression procedure (by using the GME estimator) and the subjective selection of the supports for the parameters in the GME estimator (by using the ridge trace).

However, the Ridge-GME parameter estimator also requires the use of subjective information from the analysis of the ridge trace, which represents the main drawback of the estimator. The improvement accomplished in this paper overcomes this weakness of the original Ridge-GME parameter estimator. The simulation study and the empirical applications discussed in this paper, as well as the results obtained by Macedo et al. (2010) and Erdugan and Akdeniz (2012), reveal a good performance of this Ridge Regression procedure in the case of regression models with small samples sizes affected by collinearity.

Acknowledgements

I would like to express my gratitude to the referees. They offered extremely valuable suggestions for improvements. This work was supported by Portuguese funds through the CIDMA – Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (“FCT - Fundação para a Ciência e a Tecnologia”), within project UID/MAT/04106/2013.

References


