

1 Diagnosing and Correcting the Effects of Multicollinearity: Bayesian 2 Implications of Ridge Regression 3 4

5 A. George Assaf (PhD)
6 Associate Professor
7 Isenberg School of Management
8 University of Massachusetts-Amherst
9 Email: assaf@isenberg.umass.edu

10
11 Mike Tsionas (PhD)
12 Professor of Econometrics
13 Lancaster University Management School
14 Email: m.tsionas@lancaster.ac.uk
15

16
17 Anastasios Tasiopoulos (PhD)
18 Researcher
19 Hellenic Parliamentary Budget Office (HPBO)
20 Email: atassiopoulos@gmail.com
21
22
23
24
25
26
27

28 **Abstract**

29
30 When faced with the problem of multicollinearity most tourism researchers recommend mean-
31 centering the variables. This procedure however does not work. It is actually one of the biggest
32 misconceptions we have in the field. We propose instead using Bayesian ridge regression and treat
33 the biasing constant as a parameter about which inferences are to be made. It is well known that
34 many estimates of the biasing constant have been proposed in the literature. When the coefficients
35 in ridge regression have a conjugate prior distribution, formal selection can be based on the marginal
36 likelihood. In the non-conjugate case, we propose a conditionally conjugate prior for the biasing
37 constant, and show that Gibbs sampling can be employed to make inferences about ridge regression
38 parameters as well as the biasing constant itself. We examine posterior sensitivity and apply the
39 techniques to a tourism data set.

40
41
42 *Keywords:* Multicollinearity; Bayesian analysis; ridge regression; Gibbs sampling.
43

44

45 Introduction

46

47 The problem of multicollinearity is highly common in tourism research. One particular example is
48 the regression model with moderators. Such model is usually highly prone to having collinearity
49 problems because the interaction term is created by multiplying two exogenous variables to create
50 another exogenous variable. To “alleviate” the potential problems of collinearity, tourism
51 researchers routinely mean center the variables by subtracting the item value from the mean value of
52 the item. This simply does not fix the problem. Mean centering does not really help or harm
53 (Echambadi and Hess, 2007; and Dalal and Zickar, 2011). While the mean-centered coefficients
54 have different interpretations than the original coefficients, we rarely see them being compared
55 against each other in the tourism literature. In fact, anytime an interaction is included in the model,
56 the original coefficients should not be used directly to assess the impact of X on Y. Instead, one
57 needs to use the marginal effect which is actually what we obtain when we mean center the variables.

58

59 Assuming that data for the dependent variable are arranged in the $n \times 1$ vector \mathbf{y} and the data for
60 the explanatory variables are in the $n \times p$ matrix \mathbf{X} , so that we have n observations and p
61 regressors, it is well established that the least squares (LS) estimator $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$, under the
62 stated assumptions about the error term is the best linear unbiased estimator (BLUE). However,
63 multicollinearity can result in ill conditioning of the matrix $\mathbf{X}'\mathbf{X}$ rendering the LS estimator
64 undesirable. For example when this matrix is nearly non-invertible, the covariance matrix will have
65 large elements in the diagonal, implying that standard errors of LS estimators will be quite large.
66 Effectively, in specific samples, it is quite likely that we may end up with LS coefficients having the
67 wrong sign, being non-significant, *etc.*

68

69 A regularization method that has been proposed is the use of the ridge regression estimator (Hoerl
70 and Kennard, 1970), with a biasing constant k , usually small. Effectively, “the procedure can be
71 used to portray the sensitivity of the estimates to the particular set of data being used, and it can be
72 used to obtain a point estimate with a smaller mean square error” (Hoerl and Kennard, 1970, p.55).
73 As a matter of fact, Hoerl and Kennard (1970) discussed the Bayesian foundation of their approach
74 (p.64) and also proposed a more general ridge regression.

75

76 A main challenge in the literature has been finding the appropriate value of k , as different
77 procedures (Dorugade and Kashid, 2010; Uslu, Egrioglu and Bas, 2014) have been used for that
78 purpose. Hoerl and Kennard (1970) suggested using the ridge trace to find the appropriate value of
79 k , for which the regression coefficients have been stabilized. Hoerl and Kennard (1976) proposed
80 an iterative approach for selecting k . However, their procedure does not necessarily converge. As
81 there is no consensus on what is a reasonable procedure to select the value of k , we propose here a
82 Bayesian approach to address this issue. Our aim is to provide tourism researchers with more
83 flexibility in estimating ridge regressions. The Bayesian approach is appealing because it treats k as a
84 parameter which is to be selected in light of the data. In fact, we do not select a single value of k ,
85 but we produce the whole marginal posterior of this parameter given the data. This, in turn, is one
86 attractive way to address the uncertainty about k .

87
88 The push for Bayesian estimation is taking place across several disciplines such as management
89 (Zyphur & Oswald, 2015; Cabantous and Gond, 2015; McKee and Miller, 2015), marketing (Rossi
90 and Allenby, 2003; Rossi et al. 2012), psychology (Van De Schoot, et al., 2017) and tourism (Assaf
91 and Tsionas, 2018 a, b). Over the last decade, we have seen a strong increase in the use of the
92 Bayesian methodology in tourism and other related fields (Wong et al. 2006; Wang et al. 2011; Assaf,
93 2012; Barros, 2014; Assaf et al. 2017; Assaf et al., 2018a). A recent special issue in the Journal of
94 Management is a clear indication on the growing popularity of this method (Zyphur & Oswald,
95 2015). Across several research areas in tourism, recent studies have demonstrated the effectiveness
96 of the Bayesian approach. For instance, Wong et al. (2006) have shown that a Bayesian vector
97 autoregressive model resulted in better forecasting accuracy than traditional non-Bayesian models.
98 Assaf et al. (2018b) have also recently shown that the Bayesian global vector autoregressive
99 (BGVAR) consistently outperforms other non-Bayesian models. Moreover, in related areas, such as
100 tourism performance, recent studies have also demonstrated how the Bayesian approach can handle
101 more complicated models than traditional estimation techniques (Assaf and Tsionas, 2018 a, b).

102
103 Recent papers has provided comprehensive introductions on the advantages of the Bayesian
104 approach (Muthen, 2010, Zyphur and Oswald, 2015). The Bayesian approach is not simply about
105 fitting more advanced models with MCMC (Markov chain Monte Carlo) but is a completely
106 different paradigm and philosophy in statistics. It offers several advantages in the estimation of
107 regression models including “ rich diagnostic information about parameters and models; controlling
108 for multiple comparisons as a function of the data; handling low-frequency, unbalanced, missing
109 data; and exploration of prior assumptions about model parameters” (Zyphur and Oswald, 2013,
110 p.7). Probably, on the most known advantages of the method is its ability to incorporate prior
111 information about a parameter and form a prior distribution. For instance, the Bayes’ theorem can
112 be expressed as: $p(\theta | y) \propto p(y | \theta)p(\theta)$, where \propto is the proportionality symbol. Here, $p(\theta | y)$ is
113 the posterior distribution which is used to carry out all inferences, and is proportional to the product
114 of the prior $p(\theta)$ and the likelihood function $p(y | \theta)$ ¹. Different choices of priors can be used
115 such as conjugate vs. non-conjugate priors. The prior is said to be conjugate if it belongs to the
116 family of distribution as the posterior distribution (i.e. the posterior has posterior has the same
117 distributional form as the prior distribution). For example, in the context when the likelihood
118 function is binomial $y \sim Bin(n, \theta)$, a conjugate prior in the form of a beta distribution on θ will
119 also lead to a posterior distribution that follows a beta distribution. A prior distribution which is not
120 conjugate is called a non-conjugate prior.

121
122 We illustrate below the flexibility of the Bayesian approach and prior information within the context
123 of ridge regression. In particular, we introduce a Bayesian ridge estimator for both conjugate and
124 non-conjugate priors though we rely more on the non-conjugate prior as the conjugate priors are
125 restrictive and have certain problems, for example they have the same tails with the likelihood and

¹ The likelihood function summarizes the information from the data.

126 they are rarely used in practice. A singular advantage of the Bayesian approach is that ridge
 127 regression can be interpreted as Bayes posterior mean when the prior on the regression parameters
 128 is multivariate normal with zero mean and covariance matrix a diagonal matrix whose diagonal
 129 elements have the same variance / precision. Moreover, the significance of the Bayesian approach to
 130 regression is that the celebrated James-Stein estimator has a direct empirical Bayes estimator. The
 131 James-Stein estimator is well-known to improve on maximum likelihood / OLS estimator in terms
 132 of risk and MSE across all values of the parameter space.

133
 134 In this paper we proceed as follows: In section 2 we provide an introduction to ridge regression.
 135 Sections 3 and 4 present the Bayesian ridge regression approach with conjugate and non-conjugate
 136 setting in comparison with the diffuse prior assumptions. We conduct a Monte Carlo study in
 137 section 5 to illustrate the issue diagnosing and correcting the effect multicollinearity. We then
 138 present illustration on the Bayesian ridge regression using a tourism application.
 139

140
 141 **2. How to Proceed?**

142 So, if mean centering does not work, how to proceed from here? One of the most common
 143 approaches is to use ridge regression to analyze regression data that is subject to multicollinearity. As
 144 mentioned, with OLS the regression parameters can be estimated using the following formula:
 145

146 $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$

147
 148 The ridge regression differentiates by adding a biased constant $k > 0$ to the diagonal elements of the
 149 correlations matrix:
 150

151 $\mathbf{b}_k = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}'\mathbf{y},$

152
 153 This is where the term “ridge regression” comes from as the diagonal of one in the correlation
 154 matrix are thought of as a ridge). What we know from Hoerl and Kennard (1970) is that there is
 155 always a $k \in (0, \bar{k})$ for which ridge regression dominates OLS in terms of mean squared error
 156 (MSE), and $\bar{k} = \frac{\sigma^2}{\alpha_{\max}^2}$, where $\mathbf{X}'\mathbf{X} = \mathbf{P}'\mathbf{\Lambda}\mathbf{P}$, and $\boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta}$. Here, \mathbf{P} is the orthonormal matrix of
 157 eigenvectors of $\mathbf{X}'\mathbf{X}$, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$, where $\lambda_1, \dots, \lambda_p$ represent the distinct eigenvalues
 158 of $\mathbf{X}'\mathbf{X}$. Another result of Hoerl and Kennard (1970) was that the total MSE of the ridge estimator
 159 is²:
 160

161
$$MSE(\mathbf{b}_k) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \boldsymbol{\beta}' (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-2} \boldsymbol{\beta}$$

² The notation A^{-2} for a matrix A , means $A^{-2} = A^{-1}A^{-1}$.

162
163
164
165 Minimizing the MSE, unfortunately, depends on the ratio of β/σ . Depending on this result
166 several settings for the parameter k have been proposed. See for example Khalaf and Shukur
167 (2005), Lawless and Wang (1976), Nomura (1988) and Maruyama and Strawderman (2005). A similar
168 idea is the Bayesian lasso regression (Park and Casella, 2008, Hans, 2009).

169
170 The goal of this paper is to propose a more flexible way to estimate k using the Bayesian approach.
171 As mentioned, one of the advantages is that with the Bayesian approach we do not (necessarily)
172 select a single value of k but we produce the whole marginal posterior of this parameter given the
173 data. We aim here to diagnose and correct the effects of multicollinearity through a full non-
174 conjugate Bayesian approach to Ridge Regression. In particular we take up Bayesian inference in
175 conjugate and non-conjugate ridge regression models by using the fact that a prior can be placed on
176 the ridge parameter(s) k and proceed with posterior analysis on all parameters using MCMC
177 techniques. We run different simulations to illustrate the performance of the method. We also
178 provided evidence based on a real dataset from the hotel industry. Our goal is to show that
179 collinearity can be simultaneously diagnosed and corrected using priors on all parameters. Our
180 techniques detect and correct the adverse effects of collinearity in a transparent way.

181
182 Specifically, given the general regression model, we consider first ridge regression from the Bayesian
183 point of view treating the biasing constant (k) as a parameter about which inferences are to be
184 made to avoid selecting a particular value of k . For the conjugate case we have derived the marginal
185 likelihoods and showed how selection of the k parameter can be performed to choose the
186 appropriate value. It is important to notice that the original ridge regression estimators depend
187 crucially on a conjugacy assumption, namely that the regression coefficients, $\beta | \sigma, k \sim N_p\left(0, \frac{\sigma^2}{k} \mathbf{I}_p\right)$.
188 Conjugate priors have certain problems, for example they have the same tails with the likelihood and
189 they are rarely used in practice.

190
191
192 The reader can refer to Leamer (1969) and Judge et al (1985) regarding this point. As they mention,
193 despite the fact that the natural conjugate setting is a convenient approach (since it provides an
194 analytical solution to the integrations involved), it has been criticized because it employs the prior
195 information as a previous imaginary sample from the same process: When we set the degrees of
196 freedom and the precision matrix equal to zero to obtain the limiting distribution of the normal-
197 (inverse) gamma prior, the resulting ignorant prior is different to the usual diffuse prior and the
198 posterior distribution has different degrees of freedom. Therefore, a non - conjugate prior can be
199 adopted instead, and numerical posterior inference can rely on the Gibbs sampler. In the Gibbs
200 sampler k is treated as a parameter and, therefore, formal statistical inferences can be made about
201 this parameter thus solving a long-standing problem in the literature. Moreover, a formal test for
202 collinearity can be developed if we compare the marginal posterior of k with its value at $k=0$
203 (corresponding to OLS or Bayes with diffuse prior). An equivalent test is to compare the marginal
204 likelihood at the optimal k with its value when $k=0$.

205

206 We discuss below a Bayesian ridge estimator for both conjugate and non-conjugate priors though we
 207 rely more on the non-conjugate prior as the conjugate priors are restrictive and have certain
 208 problems, for example they have the same tails with the likelihood and they are rarely used in
 209 practice.

210
 211 **3. Bayesian ridge regression**

212
 213 For the Bayesian interpretation of the ridge regression estimator, the model is given by:

214
 215
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \mathbf{u} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

216 The prior on the unknown parameters is

217
$$\boldsymbol{\beta} | \sigma \sim N_p\left(\mathbf{0}, \frac{\sigma^2}{k} \mathbf{I}_p\right), p(\sigma) \propto \sigma^{-1},$$

218 where $k > 0$ is prior precision *relative* to the error variance, σ^2 . It is not difficult to show that,
 219 under these conditions, the posterior mean is given by: $E(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}) = \mathbf{b}_k$, that is the ridge regression
 220 estimator. As mentioned, since the choice of parameter k has been an active area of research for
 221 many years and, many choices have been proposed, it is natural to investigate the implications of a
 222 fully Bayesian approach to the problem. To this effect, we consider both a conjugate and non-
 223 conjugate prior on the regression parameters, $\boldsymbol{\beta}$

224
 225 The conjugate prior provides explicitly in analytical form the ridge regression estimator so there is
 226 much in favor of it. However, the non-conjugate case is also interesting and can be considered as an
 227 alternative.

228
 229 **3.1 Optimal biasing parameter through conjugacy**

230
 231 Suppose $\boldsymbol{\beta} | \sigma \sim N_p\left(\mathbf{0}, \frac{\sigma^2}{k} \mathbf{I}_p\right)$, and $\frac{\underline{\nu} \underline{s}^2}{\sigma^2} \sim \chi^2(\underline{\nu})$, where $\underline{\nu}, \underline{s}^2$ are prior hyperparameters. This prior
 232 is conjugate because it depends on σ and matches exactly the likelihood to provide as posterior mean
 233 \mathbf{b}_k below.

234
 235 The marginal likelihood (“or evidence”), for a given value of k , can be derived analytically in this
 236 case³:

237
 238
$$p_k(\mathbf{y}) \propto \left(\frac{|\mathbf{V}|}{|\underline{\mathbf{V}}|}\right)^{1/2} (\bar{\nu} \bar{s}^2)^{-\bar{\nu}/2}, \text{ where}$$

$$\mathbf{V} = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}, \underline{\mathbf{V}} = k^{-1}\mathbf{I}_p,$$

$$\bar{\nu} \bar{s}^2 = \underline{\nu} \underline{s}^2 + \mathbf{y}' \left(\mathbf{I}_p - \mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}'\right) \mathbf{y}$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \text{ and } \bar{\nu} = \underline{\nu} + n.$$

³See Zellner (1971), p.309.

239 Moreover, $\frac{|V|}{|Y|} = \frac{k^p}{\prod_{i=1}^p (k + \lambda_i)}$, where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of $\mathbf{X}'\mathbf{X}$, and

240 $\bar{v}s^2 = \underline{v}s^2 + RSS + \mathbf{b}'(\mathbf{X}'\mathbf{X})(\mathbf{b} - \mathbf{b}_k)$, $RSS = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$. By (2.3) in Hoerl and Kennard

241 (1970), we have $\left[\mathbf{I}_p + k(\mathbf{X}'\mathbf{X})^{-1}\right]^{-1} \mathbf{b} = \mathbf{b}_k$, where $\mathbf{b}_k = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}'\mathbf{y}$ is the ridge estimate.

242 So the log marginal likelihood simplifies to the expression:

243

244
$$\log p_k(\mathbf{y}) = 0.5p \log k - 0.5 \sum_{i=1}^p \log(\lambda_i + k) - 0.5\bar{v} \log(\underline{v}s^2 + RSS + \mathbf{b}'(\mathbf{X}'\mathbf{X})(\mathbf{b} - \mathbf{b}_k)) \quad (1)$$

245

246 This expression involves only the eigenvalues of $\mathbf{X}'\mathbf{X}$, standard LS quantities and the ridge
247 estimates.

248

249

250 3.2. Bayesian ridge regression in the non-conjugate case

251

252 In the Bayesian context it is reasonable to treat k as unknown parameter whose prior is $p(k)$
253 *independently* of $\boldsymbol{\beta}$ and σ . Therefore, it is useful to depart from the conjugate case which involves the
254 unpleasant feature that the tails of the posterior and the prior are the same. Then we have:

255

256
$$\boldsymbol{\beta} | \sigma, k \sim N_p\left(\mathbf{0}, \frac{1}{k}\mathbf{I}_p\right), p(\sigma | k) \propto \sigma^{-1},$$

257 and the prior of k is proportional to $p(k)$. It is not necessary for this prior to be proper. The
258 joint posterior is as follows:

259

260
$$p(\boldsymbol{\beta}, \sigma, k | \mathbf{y}, \mathbf{X}) \propto \sigma^{-(n+1)} k^{p/2} p(k) \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + k\sigma^2 \boldsymbol{\beta}'\boldsymbol{\beta}}{2\sigma^2}\right] \quad (2)$$

261 Completing the square $Q = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + k\sigma^2 \boldsymbol{\beta}'\boldsymbol{\beta}$, we obtain the expression:

262
$$Q = (\boldsymbol{\beta} - \mathbf{b}_k)'(\mathbf{X}'\mathbf{X} + k\sigma^2\mathbf{I}_p)(\boldsymbol{\beta} - \mathbf{b}_k) + \mathbf{y}'\mathbf{M}_k\mathbf{y},$$

263

264 where $\mathbf{M}_k = \mathbf{I}_p - \mathbf{X}\mathbf{V}_k\mathbf{X}'$ and $\mathbf{V}_k = (\mathbf{X}'\mathbf{X} + k\sigma^2\mathbf{I}_p)^{-1}$. Therefore, the posterior distribution is:

265

266
$$p(\boldsymbol{\beta}, \sigma, k | \mathbf{y}, \mathbf{X}) \propto \sigma^{-(n+1)} k^{p/2} p(k) \exp\left[-\frac{(\boldsymbol{\beta} - \mathbf{b}_k)'(\mathbf{X}'\mathbf{X} + k\sigma^2\mathbf{I}_p)(\boldsymbol{\beta} - \mathbf{b}_k) + \mathbf{y}'\mathbf{M}_k\mathbf{y}}{2\sigma^2}\right] \quad (3)$$

267 In this expression, $\mathbf{b}_k = (\mathbf{X}'\mathbf{X} + k\sigma^2\mathbf{I}_p)^{-1} \mathbf{X}'\mathbf{y}$ is the ridge regression estimate. From the expression
268 in (3), we can extract the following posterior conditional distributions:

269

270
$$\boldsymbol{\beta} | \sigma, k, \mathbf{y}, \mathbf{X} \sim N_p \left(\mathbf{b}_k, \sigma^2 (\mathbf{X}'\mathbf{X} + k\sigma^2 \mathbf{I}_p)^{-1} \right),$$

271
$$\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\sigma^2} \Big| \boldsymbol{\beta}, k, \mathbf{y}, \mathbf{X} \sim \chi^2(n),$$

272 where $\chi^2(n)$ denotes the chi-square distribution with n degrees of freedom. Finally, the posterior
273 conditional distribution of the biasing parameter can be derived from (2) as:

274
$$p(k | \boldsymbol{\beta}, \sigma, \mathbf{y}, \mathbf{X}) \propto k^{p/2} p(k) \exp\left(-\frac{k\boldsymbol{\beta}'\boldsymbol{\beta}}{2}\right) \quad (4)$$

275 The conditionally conjugate prior for the biasing parameter, k , is clearly a *Gamma* $(\bar{A}/2, \bar{B}/2)$ ⁴
276 distribution whose density is of the following form: $p(k) = \frac{(\bar{B}/2)^{\bar{A}/2}}{\Gamma(\bar{A}/2)} k^{\bar{A}/2-1} \exp(-\frac{\bar{B}}{2}k)$, where $\bar{A} \geq 0$
277 and $\bar{B} \geq 0$ are hyperparameters. Then we obtain:

278
$$p(k | \boldsymbol{\beta}, \sigma, \mathbf{y}, \mathbf{X}) \propto k^{(p+\bar{A})/2-1} \exp\left(-\frac{\bar{B} + \boldsymbol{\beta}'\boldsymbol{\beta}}{2}k\right)$$

279 Therefore, the posterior conditional distribution of the biasing parameter is:
280

281
$$k | \boldsymbol{\beta}, \sigma, \mathbf{y}, \mathbf{X} \sim \text{Gamma}\left(\frac{n + \bar{A}}{2}, \frac{\bar{B} + \boldsymbol{\beta}'\boldsymbol{\beta}}{2}\right) \quad (5)$$

282
283 Prior elicitation of the hyperparameters \bar{A} and \bar{B} is facilitated by the fact that, in the prior,
284 $E(k) = \frac{\bar{A}}{\bar{B}}$ and $Var(k) = 2\frac{\bar{A}}{\bar{B}^2}$. If we believe that $E(k) = 0.1$ and the standard deviation of the
285 biasing parameter is σ_k , then $\bar{B} = \frac{0.2}{\sigma_k^2}$. If σ_k^2 is 0.1, 0.5, 1 or 5 then we obtain respectively that \bar{B} is
286 2.0, 0.40, 0.20 or 0.04. Therefore \bar{A} must be, respectively, 0.2, 0.004, 0.02 or 0.004. In what follows
287 we adopt the reference prior $p(\sigma) \propto \sigma^{-1}$.
288

289
290

291 4. Sampling properties of diagnosing and correcting multicollinearity

292

293

294 4.1. Diagnosing Collinearity

295

296 Two reasonable questions: how we diagnose for collinearity using the Bayesian approach, and how
297 the Bayesian ridge model in section 3.2⁵ behaves compared to ordinary least square (OLS)⁶.

⁴Notice that *Gamma* $(\bar{A}/2, \bar{B}/2)$ reduce to an exponential prior for k , by setting $\bar{A} = 2$.

⁵ As mentioned, we recommend relying on the non-conjugate prior for the reasons mentioned in Section 4.2.

298
 299 Specifically, the question is whether the Bayesian approach can be useful in both diagnosing and
 300 correcting the possibly harmful effects of multicollinearity in circumstances that are encountered in
 301 practice. To illustrate this, we choose a design with $n=500$ observations and $p=10$ regressors.
 302 The first regressor, say X_{i1} , is generated from a standard normal distribution. The remaining
 303 regressors are $X_{ij} = \alpha X_{i,j-1} + \omega Z_{ij}$, $Z_{ij} \sim iidN(0,1)$, $j=2, \dots, p$, and ω is set to $1/500$ with $\alpha=1$
 304 for collinear data, and $\alpha=0$, $\omega=1$ for independent data. The data generating process is
 305 $y_i = \sum_{j=1}^p \beta_j x_{ij} + u_i$, where all regression coefficients are $\beta_j=1$ ($j=1, \dots, p$), and $u_i \sim iidN(0, \sigma^2)$,
 306 with $\sigma=1$ ⁷. Setting all coefficients equal to one is done only for simplicity and the results in no way
 307 depend on the exact true values of the coefficients.
 308

309 As our prior on the biasing constant, k , we choose an exponential with parameter $\bar{B}=10^3$ implying
 310 a prior average value of k equal to $E(k)=10^{-3}$ which seems reasonable in view of experience with
 311 collinear data. Holding the matrix of regressors, X , fixed we generate $D=10,000$ different data
 312 sets. For each data set, the Gibbs sampling technique presented in section 4.2 is applied using
 313 11,000 draws, omitting the first 1,000 and taking only every other tenth draw (for a total for 1,000
 314 draws). From the 1,000 available, approximately independent, draws we compute the posterior
 315 means, $\bar{\beta}^{(d)} = E(\beta | \mathbf{y})$, $d=1, \dots, D$. Our objective is to compare the sampling distribution resulting
 316 from $\bar{\beta}^{(d)}$ s (as an approximation to the actual sampling distribution) with the sampling distribution
 317 of posterior means resulting from a diffuse prior, which is based on OLS quantities that are readily
 318 available for each different data set.
 319

320 With a diffuse prior (i.e. OLS), the sampling distribution of posterior means should be more
 321 dispersed compared to the sampling distribution of $E(\beta | \mathbf{y})$ under the stated prior on the biasing
 322 constant, k . The sampling results are presented in the three panels of Figure 1. For orthogonal data
 323 (i.e. no collinearity) ridge and diffuse posteriors are extremely close (Figure 1a), but this is not the
 324 case when we have collinear data (Figures 1b and 1c). In other words, for collinear data, one would
 325 observe significant difference between the diffuse posterior (i.e. OLS) and the ridge regression
 326 results.
 327

328 This, in turn, would provide a useful way to detect whether there is harmful multicollinearity and, at
 329 the same time, correct it based on the Bayesian ridge estimator. The test can be made more formal
 330 by using a Kolmogorov-Smirnov test for testing the equality of the two distributions. We believe,
 331 however, that visual presentation is much more informative. We can also use a formal test for
 332 collinearity in Bayesian analysis. In OLS settings such tests are not possible. For example, variance
 333 inflation factors (VIF) commonly used are diagnostics of collinearity, not statistical tests.
 334

⁶ This is similar to Bayesian analysis using diffuse priors of the form $p(\beta, \sigma) \propto \sigma^{-1}$, from the sampling-theory point of view

⁷ All computations were performed using the WinGauss software. The codes can be provided by the authors upon request.

335 In our case, one can formally test for collinearity using the Bayes factor. Given the marginal
336 posterior $p(k|\mathbf{y})$ the Bayes factor in favor of ridge regression and against OLS can be
337 approximated using

$$338 \quad BF \simeq \frac{p(\hat{k}|\mathbf{y})}{p(k=0|\mathbf{y})} \quad (6)$$

339 where \hat{k} is the modal value of the marginal posterior. In our case the denominator is practically
340 zero, so the BF diverges to a very large value, indicating that the ridge regression model fits the data
341 best. For this approach see Berger (1980, p. 156).

342
343

344 4.2. Correcting for Collinearity

345

346 To correct for collinearity we propose the Bayesian ridge estimator proposed in Section 4.2. While
347 we saw in Figure 1.c how the Bayesian diffuse prior (i.e. OLS) can seriously affect the regression
348 results in the presence of collinear data, the important issue that remains is how reliable our Bayesian
349 estimator is in the presence of collinearity.

350 We run a similar experiment to the above where we compared the performance of the Bayesian
351 diffuse prior against the Bayesian ridge regression. We also included in the comparison a traditional
352 ridge regression model (non-Bayesian) with a pre-specified value for $k = 0.001^8$. For the diffuse
353 prior context, we followed a common practice in the literature and tried to drop the collinear
354 variables from the model.

355 We use 10,000 Monte Carlo replications to compare the above models. We tried three versions of
356 the diffuse prior model by dropping one, two and three variables at a time. The true model for the
357 simulation is

358 $y_i = x_{i1} + x_{i2} + x_{i3} + x_{i4} + x_{i5} + 0.1u_i$ where $u_i \sim N(0,1)$ and the regressors are generated as follows.

359 $x_{i1} \sim N(0,1)$

360 $x_{ij} = x_{i1} + 0.1v_{ij}, v_{ij} \sim N(0,1), j = 2, \dots, 5.$

361 For testing purposes we set all β_j ($j = 1, \dots, p$) as equal to 1. Again, the results do not depend on
362 the exact values of these coefficients.

363 The results are presented in Figure 2. We can see that the Bayesian ridge regression based on the
364 optimal prior seems to performs best and is the one most centered around the true value of β .
365 Contrary to common belief, the practice of dropping variables from the models, on the other hand,

⁸ This is a standard value for k used in the literature. Of course, our proposed approach focuses on optimizing the value of k and not on pre-specifying the value of k , as discussed previously.

366 does not seem to be a good choice for correcting the results of the regression model. The closest to
367 our model is the traditional ridge regression, but this has the problem of pre-specifying the value of
368 k in advance. Further illustration with real data is presented next.

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

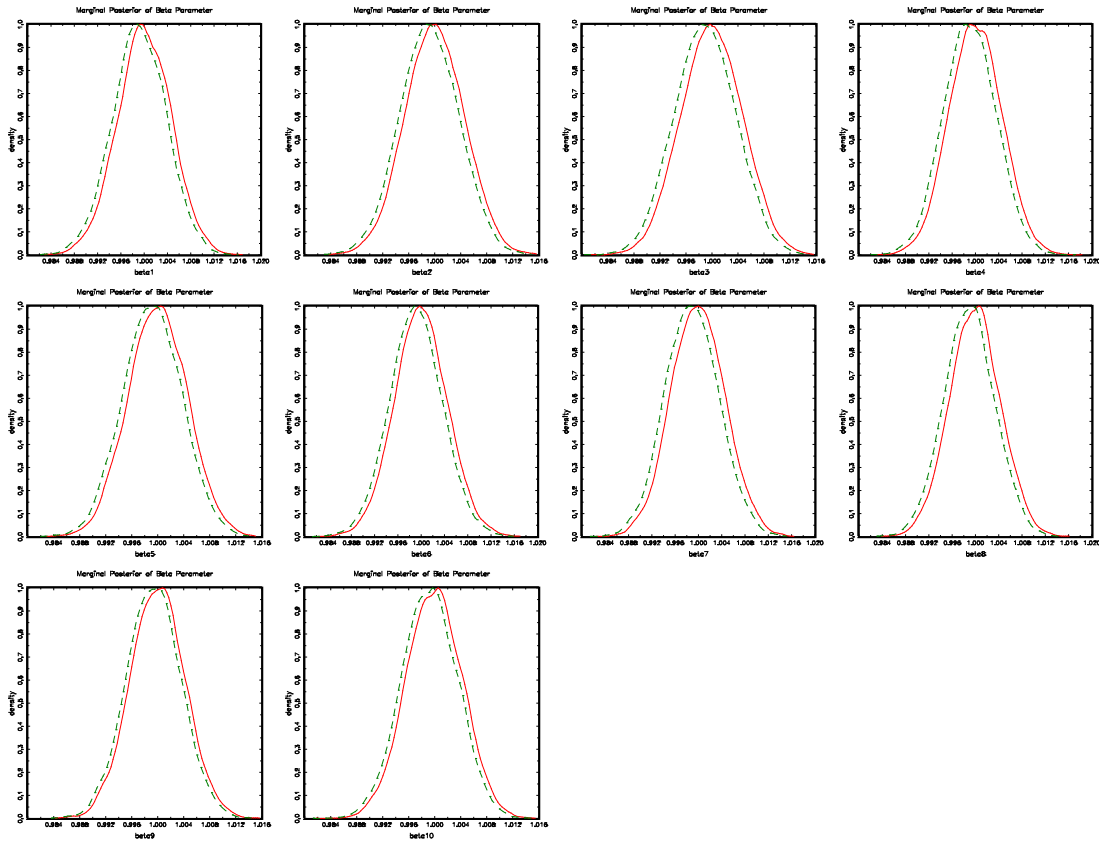
394

395 Figure 1. Sampling distributions of different estimators of β_2 .

396

397 **Figure 1a. Sampling Distributions of Posterior Mean from Bayesian diffuse analysis and**
398 **Bayesian ridge analysis ($\bar{B} = 1$). Orthogonal data.**

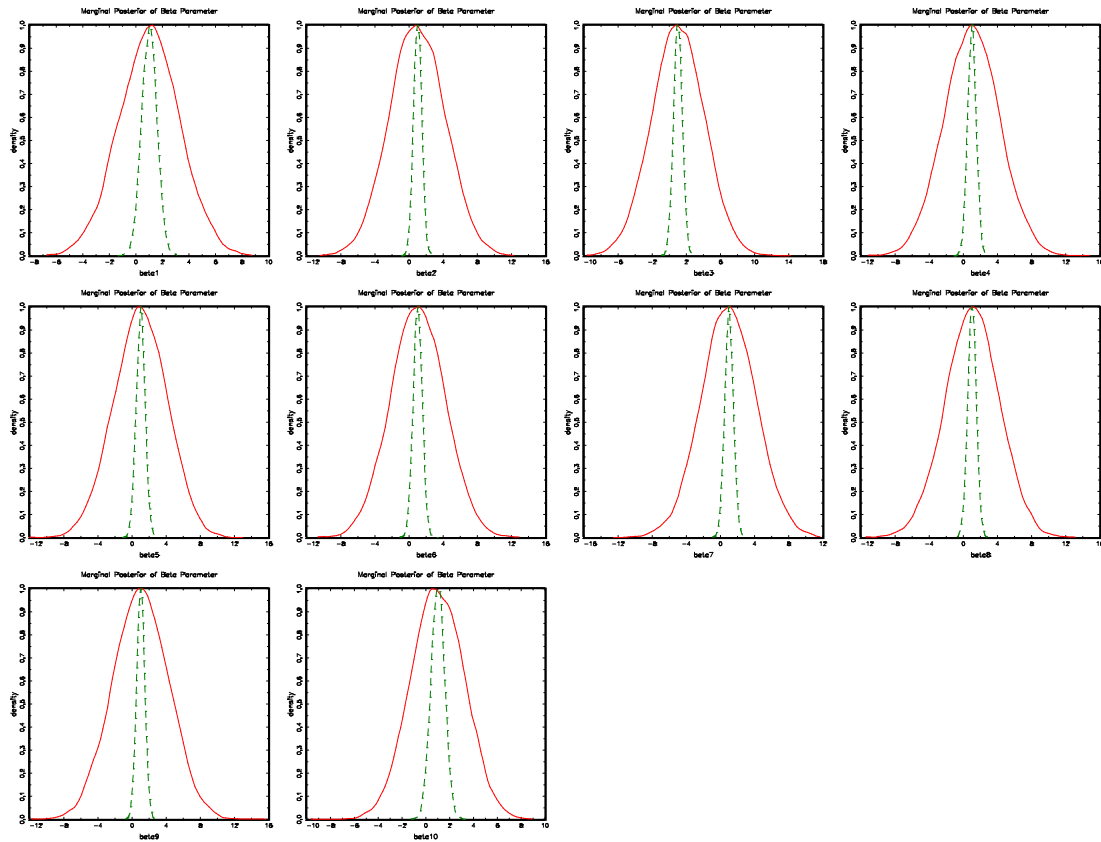
399



400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416

417 **Figure 1b. Sampling Distributions of Posterior Mean from Bayesian diffuse analysis and**
418 **Bayesian ridge analysis ($\bar{B} = 10^3$). Collinear data.**

419
420



421
422

423

424

425

426

427

428

429

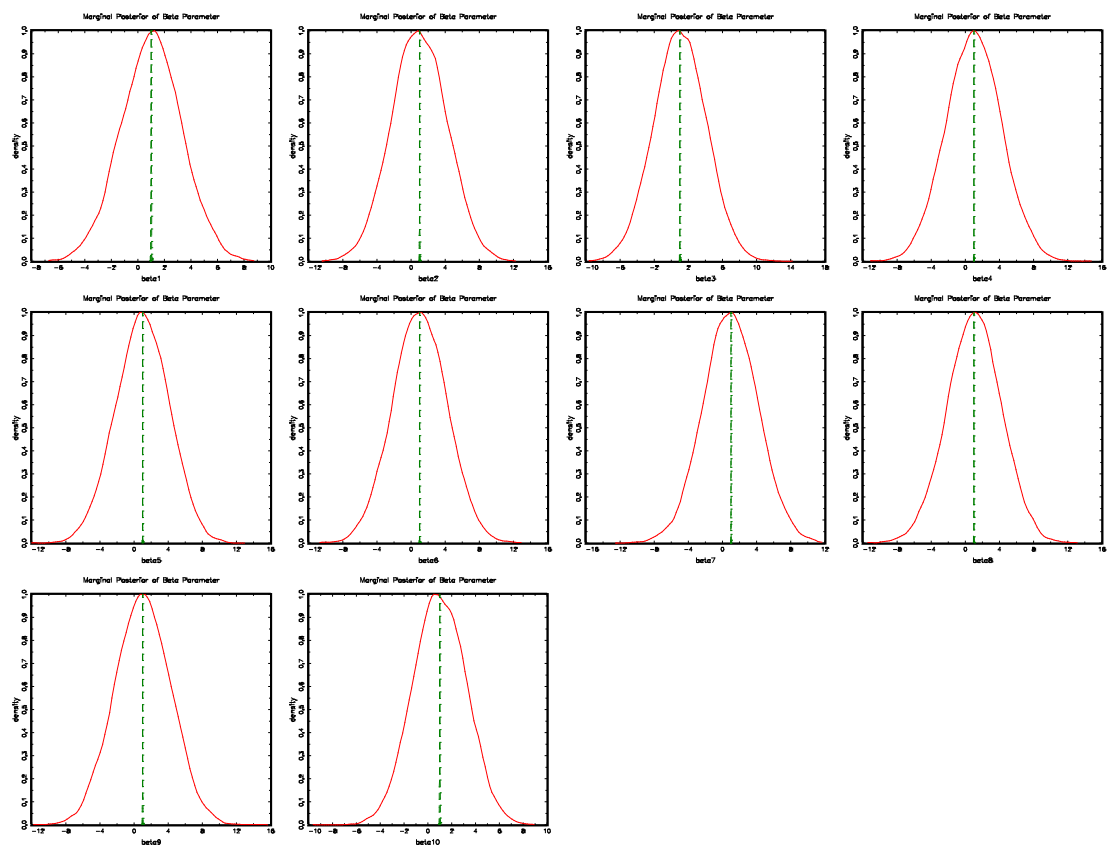
430

431

432

433

434 **Figure 1c. Sampling Distributions of Posterior Mean from Bayesian diffuse analysis and**
435 **Bayesian ridge analysis ($\bar{B} = 1$). Collinear data.**
436

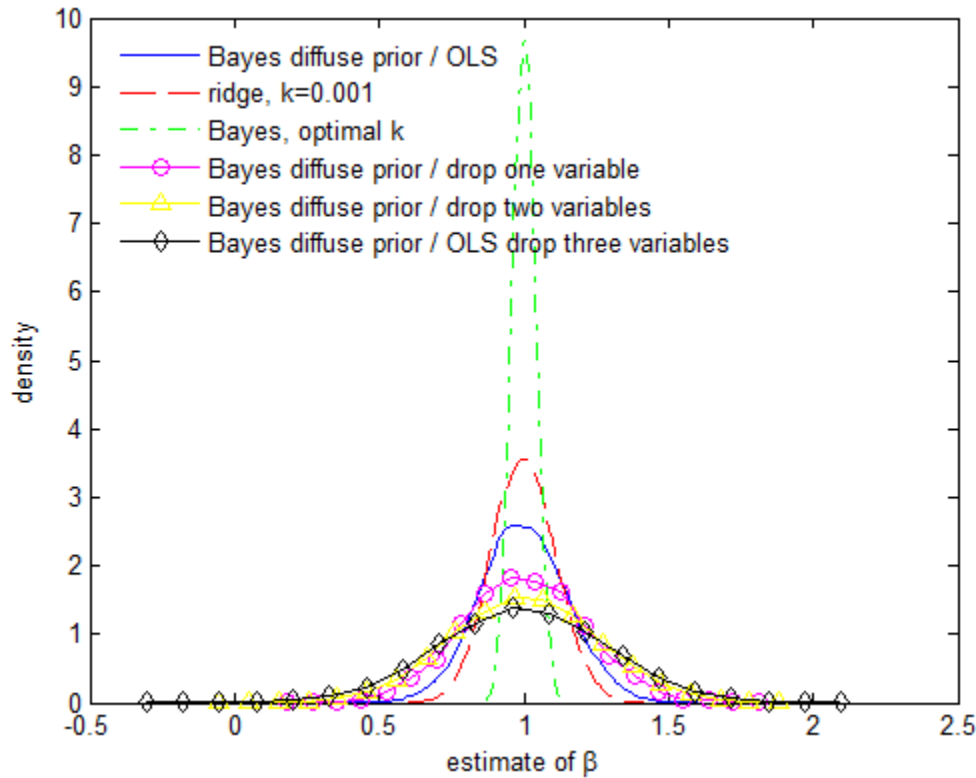


437
438
439
440
441
442
443
444
445
446
447
448
449

450

451 **Figure 2. Performance of the Bayesian ridge regression against other alternatives**

452



453

454

455

456 5. Illustration using Real Data

457 We also test our Bayesian ridge regression using a real application the hotel industry. The model
458 focuses on the relationship between room revenue and the following covariates: room expenses,
459 food and beverage (F&B) expenses, utility expenses, marketing expenses, property and maintenance
460 (POM) expenses and number of rooms. All these variables are expected to be positively correlated
461 with room revenue as higher revenue usually results in higher expenses in these categories.

462 The dataset for this study was obtained from Smith Travel Research, an independent company that
463 tracks lodging supply and demand data for most major hotels in the US and internationally. The
464 STR's data are highly comprehensive, reliable and mostly commonly by hotels to track their
465 performance⁹.

466 We use here a unique panel sample of 78 US hotels (for the years 2012-2016). So, in total we have
467 390 observations. The correlation matrix for all variables included in the model (Table 4) clearly

⁹ At least in the United States.

468 illustrates the high collinearity problem. Further evidence on the collinearity problem in this dataset
469 is illustrated in Table 5 where we can see that the variance inflation factors (VIFs) for five of the six
470 covariates are >10 . Our Bayes factor (equation 6) also diverges to a very large value, indicating that
471 the ridge regression model fits the data best.

472 We report in Table 6 the results from Bayesian ridge regression and linear regression (i.e. OLS). For
473 the Bayesian estimation, we used the non-conjugate prior described in Section 4.2. As mentioned,
474 one of the advantages is that with the Bayesian approach, we do not pre-set or select a single value
475 of k but we produce the whole marginal posterior of this parameter given the data. For example, we
476 report in Figure 3 the overall posterior density of k .

477 The posterior mean of k is also included in Table 6. We can clearly the differences between the
478 results obtained from the Bayesian ridge regression vs. OLS. For instance, despite the high positive
479 correlation between the various covariates and the dependent variable, several coefficients from
480 OLS have a negative sign and only three of them are significant. The Bayesian ridge regression
481 however, indicates that all coefficients are positive and significant. This confirms our earlier results
482 from the simulation that when collinearity exists, ridge and least square results can be very different.

483 Of course, we are not implying that collinearity is of less concern than is often implied in the
484 literature. While our method seems to be more tolerant to collinearity, the results should not
485 encourage tourism researchers to throw any variable into the model and expect the results to come
486 out perfectly. The selection of variables should still be based on an educated theoretical approach.
487 In contexts when collinearity cannot be avoided, the practices of mean centering, or dropping
488 variables do not seem to be good choices for correcting the results of the regression model. Rather,
489 the regression estimation should be conducted using more robust approaches such as the one we
490 propose in this study.

491
492
493
494
495
496
497
498
499
500
501
502

503

504

505 Table 4. Correlation Matrix

	1	2	3	4	5	6	7
Room revenue (1)	1						
Room Expenses (2)	0.9867	1					
F&B expenses (3)	0.9640	0.9744	1				
Utility Expenses (4)	0.9815	0.9741	0.9543	1			
Marketing Expenses(5)	0.9463	0.9279	0.9379	0.9523	1		
POM Expenses (6)	0.9615	0.9717	0.9741	0.9602	0.9514	1	
Number of Rooms (7)	0.8773	0.8839	0.8685	0.9075	0.9242	0.8967	1

506

507

508 Table 5. Multicollinearity Diagnostic Criteria

	Eigen Values	VIF	1/VIF
Room Expenses	5.7033	34.3711	0.0291
F&B expenses	0.157	25.0432	0.0399
Utility Expenses	0.0566	28.6689	0.0349
Marketing Expenses	0.0423	19.6543	0.0509
POM Expenses	0.0242	33.7577	0.0296
Number of Rooms	0.0165	8.3613	0.1196

509

510

511 Table 6. Bayesian Ridge Regression vs. OLS

Variable	Bayesian Ridge		OLS	
	Posterior Mean	Posterior t-stat	Estimate	t-stat
Room Expenses	5.871	11.068	10.442	16.202
F&B expenses	0.803	3.902	-0.504	-1.127
Utility Expenses	1.414	11.542	0.957	1.844
Marketing Expenses	3.913	38.391	4.628	12.172
POM Expenses	0.960	3.357	-1.378	-2.018
Number of Rooms	1.101	4.871	-0.911	-1.953
sigma	0.211	0.010		
k	0.129	0.075		

512

513

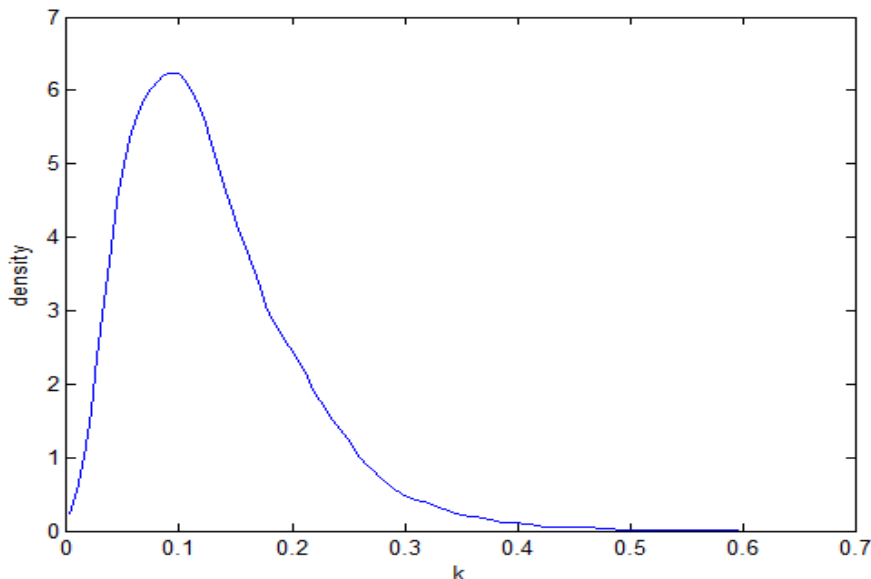
514

515

516

517

518 **Figure 3. Posterior Distribution of the Bayesian ridge parameter (k)**



519

520

521

522

523

524 **6. Concluding Remarks**

525

526 In this paper, we have taken up Bayesian inference in conjugate and non-conjugate ridge regression
527 models by using the fact that a prior can be placed on the ridge parameter(s) k proceed with
528 posterior analysis on all parameters using standard MCMC techniques. For the conjugate case we
529 have derived the marginal likelihoods and showed how selection of the k or g parameter can be
530 based in an empirical Bayes context to choose the appropriate value. It is important to notice that
531 the original ridge regression estimators depend crucially on a conjugacy assumption, namely that the
532 regression coefficients, $\boldsymbol{\beta} | \sigma, k \sim N_p\left(0, \frac{\sigma^2}{k} \mathbf{I}_p\right)$. In the absence of σ the prior of $\boldsymbol{\beta}$ is no longer in
533 the normal-gamma prior form which is necessary for ordinary ridge regression to emerge. A non-
534 conjugate prior of the form $\boldsymbol{\beta} | k \sim N_p\left(0, \frac{1}{k} \mathbf{I}_p\right)$ can be adopted instead, and numerical posterior
535 inference can rely on the Gibbs sampler. Conjugate priors have certain problems, for example they
536 have the same tails with the likelihood and they are rarely used in practice.

537

538 We have applied these ideas to show that collinearity can be simultaneously diagnosed and corrected

539 using priors on all parameters. We also illustrated that the Bayesian ridge regression performs better
540 than a Bayesian regression with diffuse prior (i.e. OLS). Contrary to common belief, the practice of
541 dropping variables from the models, does not also seem to be a good choice for correcting the
542 results of the regression model. Our method can handle collinearity more robustly when variables
543 are collinear and should be theoretically left in the model. We showed through simulation and
544 through marginal likelihood maximization that the approach we are proposing provide more robust
545 findings. This should provide tourism researchers with more flexibility as they handle such type of
546 data.

547
548 One limitation of the paper is that we focus on a single k . Different k s can be used for each
549 regressor easily, although at the cost of computing several values of the marginal likelihood
550 depending on the value of such k coefficients. Our methods illustrate that biased estimators yielding
551 lower mean squared error (MSE) are clearly desirable and, thus, future research could focus more on
552 generalizations of our procedure. Another limitation is, of course, the assumption of normality of
553 errors which, however, can be relaxed to consider more general models including other elliptical
554 distributions, distributions with fat tails and / or asymmetry, etc.

555

556

557

558

559 **References**

560

561 Assaf, A. G. (2012). Benchmarking the Asia Pacific tourism industry: A Bayesian combination of
562 DEA and stochastic frontier. *Tourism Management*, 33(5), 1122-1127.

563

564 Assaf, A. G., & Tsionas, M. (2018a). Bayes factors vs. P-values. *Tourism Management*, 67, 17-31.

565

566 Assaf, A. G., & Tsionas, M. (2018b). The estimation and decomposition of tourism productivity.
567 *Tourism Management*, 65, 131-142.

568

569 Assaf, A. G., Tsionas, M., & Oh, H. (2018a). The time has come: Toward Bayesian SEM estimation
570 in tourism research. *Tourism Management*, 64, 98-109.

571 Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2018b). Modeling and Forecasting Regional
572 Tourism Demand Using the Bayesian Global Vector Autoregressive (BGVAR) Model. *Journal of*
573 *Travel Research*, in press.

574 Assaf, A. G., Oh, H., & Tsionas, M. (2017). Bayesian Approach for the Measurement of Tourism
575 Performance: A Case of Stochastic Frontier Models. *Journal of Travel Research*, 56(2), 172-186.

576 Barros, C. P. (2014). Airports and tourism in Mozambique. *Tourism Management*, 41, 76-82.

577

- 578 Berger, J. O. (1980). Statistical decision theory and Bayesian analysis, Springer, Berlin.
- 579 Cabantous, L., & Gond, J. P. (2015). The resistible rise of Bayesian thinking in management:
580 Historical lessons from Decision Analysis. *Journal of Management*, 41(2), 441-470.
581
- 582 Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in
583 moderated multiple regression and polynomial regression. *Organizational Research Methods*, 15(3), 339-
584 362.
- 585 Dorugade, A. V., & Kashid, D. N. (2010). Alternative method for choosing ridge parameter for
586 regression. *Applied Mathematical Sciences*, 4(9), 447-456.
- 587 Echambadi, R., & Hess, J. D. (2007). Mean-centering does not alleviate collinearity problems in
588 moderated multiple regression models. *Marketing Science*, 26(3), 438-445.
- 589 Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4), 835-845.
- 590 Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal
591 problems. *Technometrics*, 12(1), 55-67.
- 592 Judge, O. G., W. E. Griffith, R. C. Hill, C. H. Lee, and H. Lutkepohl.(1985). The theory
593 and practice of econometrics, second edition, Wiley, New York.
- 594 Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communication
595 in Statistics: Theory and Methods*, 34, 1177-1182.
596
- 597 Lawless, J. F., & Wang, P. (1976). Simulation study of ridge and other regression estimators.
598 *Communication in Statistics: Theory and Methods*, 4, 307-323.
599
- 600 Leamer, E. E., (1978) Specification Searches, Wiley, New York.
- 601 Maruyama, Y., & Strawderman, W. E. (2005). A new class of generalized Bayes minimax ridge
602 regression estimators. *The Annals of Statistics*, 33(4), 1753-1770.
603
- 604 McKee, R. A., & Miller, C. C. (2015). Institutionalizing Bayesianism within the organizational
605 sciences: A practical guide featuring comments from eminent scholars. *Journal of Management*, 41(2),
606 471-490.
607
- 608 Muthén, B. (2010). Bayesian analysis in Mplus: A brief introduction. *Unpublished manuscript. www.
609 statmodel.com/download/IntroBayesVersion, 203.*
610
- 611 Nomura, M. (1988). On the almost unbiased ridge regression estimator. *Communications in Statistics-
612 Simulation and Computation*, 17(3), 729-743.
613
- 614 Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*,
615 103(482), 681-686.
616
- 617 Rossi, P. E., & Allenby, G. (2003). Bayesian statistics and marketing. *Marketing Science* 22(3), 304-328.

618
619 Rossi, P. E., Allenby, G, and McCulloch, R. (2012). *Bayesian statistics and marketing*. John Wiley &
620 Sons.
621

622 Uslu, V. R., Egrioglu, E., & Bas, E. (2014). Finding optimal value for the shrinkage parameter in
623 ridge regression via particle swarm optimization. *American Journal of Intelligent Systems*, 4(4), 142-147.

624 Van De Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijenburg, M., & Depaoli, S. (2017). A
625 systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2),
626 217.
627

628 Wang, W., Zeng, G., & Tang, D. (2011). Bayesian intelligent semantic mashup for tourism.
629 *Concurrency and Computation: Practice and Experience*, 23(8), 850-862.
630

631 Wong, K. K., Song, H., & Chon, K. S. (2006). Bayesian models for tourism demand forecasting.
632 *Tourism Management*, 27(5), 773-780.
633

634 Zellner, A. (1971). *An introduction to Bayesian inference in econometrics* (No. 519.54 Z4).
635

636 Zyphur, M. J., & Oswald, F. L. (2013). Bayesian probability and statistics in management research: A
637 new horizon. *Journal of Management*, 39(1), 5-13.

638 Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of*
639 *Management*, 41(2), 390-420.
640

641

642