Tourism Management 71 (2019) 1–8

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression

A. George Assaf[a,*,1], Mike Tsionas[b], Anastasios Tasiopoulos[c]

[a] Isenberg School of Management, University of Massachusetts-Amherst, 90 Campus Center Way, 209A Flint Lab, Amherst, MA, 01003, USA
[b] Lancaster University Management School, UK
[c] Hellenic Parliamentary Budget Office (HPBO), Greece

ARTICLE INFO

Keywords:
Multicollinearity
Bayesian analysis
Ridge regression
Gibbs sampling

ABSTRACT

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

1. Introduction

The problem of multicollinearity is highly common in tourism research. One particular example is the regression model with moderators. Such model is usually highly prone to having collinearity problems because the interaction term is created by multiplying two exogenous variables to create another exogenous variable. To "alleviate" the potential problems of collinearity, tourism researchers routinely mean center the variables by subtracting the item value from the mean value of the item. This simply does not fix the problem. Mean centering does not really help or harm (Echambadi & Hess, 2007; and Dalal & Zickar, 2012). While the mean-centered coefficients have different interpretations than the original coefficients, we rarely see them being compared against each other in the tourism literature.

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

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

A main challenge in the literature has been finding the appropriate value of k, as different procedures (Dorugade & Kashid, 2010; Uslu, Egrioglu, & Bas, 2014) have been used for that purpose. Hoerl and Kennard (1970) suggested using the ridge trace to find the appropriate value of k, for which the regression coefficients have been stabilized. Hoerl and Kennard (1976) proposed an iterative approach for selecting k. However, their procedure does not necessarily converge. As there is no consensus on what is a reasonable procedure to select the value of k, we propose here a Bayesian approach to address this issue. Our aim is to provide tourism researchers with more flexibility in estimating ridge regressions. The Bayesian approach is appealing because it treats k as a parameter which is to be selected in light of the data. In fact, we do not

* Corresponding author.
E-mail addresses: assaf@isenberg.umass.edu (A.G. Assaf), m.tsionas@lancaster.ac.uk (M. Tsionas), a.tasiopoulos@parliament.gr (A. Tasiopoulos).
[1] Authors are listed in this order based on their contribution to the paper.

https://doi.org/10.1016/j.tourman.2018.09.008
Received 20 April 2018; Received in revised form 17 September 2018; Accepted 19 September 2018
0261-5177/

select a single value of k, but we produce the whole marginal posterior of this parameter given the data. This, in turn, is one attractive way to address the uncertainty about k.

The push for Bayesian estimation is taking place across several disciplines such as management (Cabantous & Gond, 2015; McKee & Miller, 2015; Zyphur & Oswald, 2015), marketing (Rossi & Allenby, 2003; Rossi, Allenby, & McCulloch, 2012), psychology (Van De Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017) and tourism (Assaf and Tsionas, 2018a,b). Over the last decade, we have seen a strong increase in the use of the Bayesian methodology in tourism and other related fields (Assaf, 2012; Assaf et al. 2017, 2018a; Barros, 2014; Wang, Zeng, & Tang, 2011; Wong, Song, & Chon, 2006). A recent special issue in the Journal of Management is a clear indication on the growing popularity of this method (Zyphur & Oswald, 2015). Across several research areas in tourism, recent studies have demonstrated the effectiveness of the Bayesian approach. For instance, Wong et al. (2006) showed that a Bayesian vector autoregressive model resulted in better forecasting accuracy than traditional non-Bayesian models. Assaf, Li, Song, and Tsionas (2018b) showed that the Bayesian global vector autoregressive (BGVAR) consistently outperforms other non-Bayesian models. Moreover, in related areas, such as tourism performance, recent studies have demonstrated how the Bayesian approach can handle more complicated models than traditional estimation techniques (Assaf and Tsionas, 2018a,b).

Recent papers have provided comprehensive introductions on the advantages of the Bayesian approach (Muthén, 2010; Zyphur & Oswald, 2015). The Bayesian approach offers several advantages in the estimation of regression models including " rich diagnostic information about parameters and models; controlling for multiple comparisons as a function of the data; handling low-frequency, unbalanced, missing data; and exploration of prior assumptions about model parameters" (Zyphur & Oswald, 2013, p. 7). Probably, one of the most known advantages of the method is its ability to incorporate prior information about a parameter and form a prior distribution. The Bayes' theorem can be expressed as: $p(\theta|y) \propto p(y|\theta)p(\theta)$, where $\propto$ is the proportionality symbol. Here, $p(\theta|y)$ is the posterior distribution which is used to carry out all inferences, and is proportional to the product of the prior $p(\theta)$and the likelihood function$p(y|\theta)$.[2] Different choices of priors can be used such as conjugate vs. non-conjugate priors. The prior is said to be conjugate if it belongs to the family of distribution as the posterior distribution. For example, in the context when the likelihood function is binomial $y \sim Bin(n,\ \theta)$, a conjugate prior in the form of a beta distribution on θ will also lead to a posterior distribution that follows a beta distribution. A prior distribution which is not conjugate is called a non-conjugate prior.

We illustrate below the flexibility of the Bayesian approach and prior information within the context of ridge regression. In particular, we introduce a Bayesian ridge estimator for both conjugate and non-conjugate priors. We rely more on the non-conjugate prior as conjugate priors are restrictive and have certain problems, for example they have the same tails with the likelihood and they are rarely used in practice. A singular advantage of the Bayesian approach is that ridge regression can be interpreted as Bayes posterior mean when the prior on the regression parameters is multivariate normal with zero mean and diagonal covariance matrix whose diagonal elements have the same variance/precision. Moreover, the significance of the Bayesian approach to regression is that the celebrated James-Stein estimator has a direct empirical Bayes estimator. The James-Stein estimator is well-known to improve on maximum likelihood/OLS estimator in terms of risk and MSE across all values of the parameter space.

In this paper, we proceed as follows: In section 2 we provide an introduction to ridge regression. Sections 3 and 4 present the Bayesian ridge regression approach with conjugate and non-conjugate setting in comparison with the diffuse prior assumptions. We conduct a Monte Carlo study in section 5. We then present an illustration of the Bayesian ridge regression using a tourism application.

2. How to proceed?

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

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

The ridge regression differentiates by adding a biasing constant $k > 0$ to the diagonal elements of the correlation matrix:

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

This is where the term "ridge regression" comes from as the diagonal of one in the correlation matrix are thought of as a ridge. What we know from Hoerl and Kennard (1970) is that there is always a $k \in (0,\ \bar{k})$ for which ridge regression dominates OLS in terms of mean squared error (MSE), and $\bar{k} = \frac{\sigma^2}{\alpha_{\max}^2}$, where $\mathbf{X}'\mathbf{X} = \mathbf{P}'\boldsymbol{\Lambda}\mathbf{P}$, and $\boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta}$. Here, $\mathbf{P}$ is the orthonormal matrix of eigenvectors of $\mathbf{X}'\mathbf{X}$, and $\Lambda = diag(\lambda_1, .., \lambda_p)$, where $\lambda_1, .., \lambda_p$ represent the distinct eigenvalues of $\mathbf{X}'\mathbf{X}$. Another result of Hoerl and Kennard (1970) was that the total MSE of the ridge estimator is[3]:

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

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

The goal of this paper is to propose a more flexible way to estimate k using the Bayesian approach. As mentioned, one of the advantages is that with the Bayesian approach we do not (necessarily) select a single value of k but we produce the whole marginal posterior of this parameter given the data. We aim here to diagnose and correct the effects of multicollineatity through a full non-conjugate Bayesian approach to ridge regression. In particular, we take up Bayesian inference in conjugate and non-conjugate ridge regression models by using the fact that a prior can be placed on the ridge parameter(s) k and proceed with posterior analysis on all parameters using MCMC techniques. We run different simulations to illustrate the performance of the method. We also provide evidence based on a dataset from the hotel industry.

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

The reader can refer to Leamer (1978) and Judge, Griffith, Hill, Lee, and Lutkepohl (1985) regarding this point. As they mention, despite the fact that the natural conjugate setting is a convenient approach (since it provides an analytical solution to the integrations involved), it has been

[2] The likelihood function summarizes the information from the data.

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