Stationer
Journal of Statistical Innovations and Applications

e-ISSN. 2829-6303
Vol. xx No. xx, xxxxxx
Pages xx-xx

Check for updates

Development of a Bootstrap-Stability Adaptive Ridge Regression Method for Multicollinear Data

Eka Sasmita[1*], **Yuni N. Qomariah**[2], **Dyah H. Keliwida**[3]

Department ofStatistics, Faculty of Science and Technology, Universitas PattimuraJl. M. J. Putuhena, Poka, Ambon, 97233, Indonesia

Email : ekasasmita@gmail.com

Article Info

Article history:
Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy

Keywords:

Adaptive penalty
Bootstrap stability
Multicollinearity
Ridge regression.

ABSTRACT

This study develops Bootstrap-Stability Adaptive Ridge Regression (BSA-Ridge), a methodological extension of classical ridge regression for multicollinear regression data. Classical ridge regression controls coefficient variance by adding a uniform quadratic penalty, but it does not distinguish predictors whose coefficients are empirically unstable from predictors whose coefficients are relatively stable. The proposed method estimates coefficient instability through bootstrap resampling and converts the bootstrap variance into predictor-specific penalty weights. The empirical illustration uses the Longley benchmark dataset, a public dataset widely used to examine numerical instability and multicollinearity in least-squares regression. The results show that BSA-Ridge produces interpretable adaptive shrinkage and competitive predictive performance relative to ordinary least squares and classical ridge regression. The contribution of this article is a transparent, equation-based, and reproducible extension of ridge regression that can be further evaluated through simulation and high-dimensional applications.

: https://doi.org/10.30598/stationerv4i1pp01-10

1. INTRODUCTION

Multicollinearity remains one of the most persistent problems in applied regression analysis. In many economic, demographic, social, and engineering datasets, explanatory variables do not vary independently. They move together because they measure related phenomena, share a common time trend, or are generated by the same underlying process. When this happens, ordinary least squares can still be unbiased under the classical assumptions, but the coefficient estimates may have very large variances. A model can look statistically plausible in terms of fitted values while individual coefficients change sharply under small perturbations of the sample [1], [3], [10].

Ridge regression was introduced by Hoerl and Kennard as a biased estimation strategy for nonorthogonal regression problems [1]. Its central idea is simple but powerful: add a quadratic penalty to the least-squares objective so that the model trades a small amount of bias for a substantial reduction in variance. This bias-variance trade-off is especially attractive when the goal is prediction or when the analyst wants to maintain all predictors in the model rather than remove variables aggressively. The Longley dataset later became a classical benchmark because it illustrates how least-squares programs and coefficient estimates can be sensitive under severe multicollinearity [3], [4].

Subsequent studies expanded the ridge regression literature in several directions. Marquardt and Snee discussed ridge regression from a practical modeling perspective, while Golub, Heath, and Wahba introduced generalized cross-validation as a principled way to choose a ridge parameter [5], [6]. Later, lasso and elastic net extended regularization toward variable selection and combined shrinkage-selection mechanisms [7], [8]. These developments are important, but they also reveal a conceptual difference: ridge regression is not primarily a variable selection method. Its strength is coefficient stabilization, especially when correlated predictors are still substantively meaningful and should not be discarded. A key limitation of classical ridge regression is that the penalty is uniform. Once a value of the regularization parameter is selected, every coefficient is penalized with the same intensity. This uniformity is mathematically convenient, but it can be statistically restrictive. In practice, not all regression coefficients have the same empirical stability. Some predictors may produce coefficient estimates that fluctuate widely under resampling, whereas other predictors remain comparatively stable. A uniform penalty cannot directly reflect this difference in empirical coefficient stability.

Bootstrap resampling offers a natural way to measure this instability. By repeatedly resampling the observed data and re-estimating the coefficients, the analyst obtains an empirical distribution of each coefficient. The variance of this empirical distribution can be interpreted as a data-driven indicator of coefficient stability [9]. If the bootstrap variance of a coefficient is large, the coefficient is sensitive to sample composition and may deserve stronger shrinkage. If the bootstrap variance is small, the coefficient is relatively stable and may deserve weaker shrinkage.

The novelty of this article is the integration of bootstrap coefficient stability into the ridge penalty structure. The proposed Bootstrap-Stability Adaptive Ridge Regression, abbreviated as BSA-Ridge, replaces the uniform ridge penalty with a diagonal penalty matrix whose entries are derived from bootstrap variances. This means that the penalty is no longer only controlled by a global tuning parameter. It is also shaped by the obse rved stability of each coefficient. Based on the literature reviewed in this article, this specific and simple formulation has not been presented as a standalone ridge development for transparent applied modeling. The proposed contribution should be interpreted carefully. This article does not claim that BSA-Ridge universally dominates all regularization methods. Such a claim would require extensive simulation, asymptotic derivation, and testing across many data-generating conditions. Instead, the article contributes a new methodological formulation, provides an initial empirical demonstration, and explains how the method can be evaluated in future research. This distinction is important because novelty in methodological research should be grounded in a clear mathematical modification, not only in a better numerical result on one dataset.

The objective of this study is therefore threefold. First, it formulates BSA-Ridge mathematically as an adaptive extension of classical ridge regression. Second, it presents a reproducible research procedure using the Longley benchmark dataset. Third, it compares the proposed method with ordinary least squares and classical ridge regression using prediction error, coefficient behavior, and penalty weights. The discussion links the findings to previous research on ridge regression, bootstrap resampling, and regularization.

2. METHOD

The development of ridge regression in this article follows an incremental methodological logic. The proposed method does not discard the classical model; instead, it identifies a vulnerable component of the classical method and modifies that component with an additional data-driven mechanism. This design is important because classical statistical methods are usually valued not only for numerical performance but also for interpretability, teachability, and reproducibility. The development of the proposed estimator is deliberately kept explicit. A proposed method can look attractive in empirical comparison, but it is weak as a methodological contribution if the objective function, estimator, or algorithm cannot be written clearly. For that reason, the equations below separate the baseline model, the bootstrap-stability component, the adaptive penalty, and the final BSA-Ridge estimator. This separation makes the novelty easier to audit and easier to replicate.

The empirical analysis should be read as an initial validation. A single benchmark dataset cannot prove universal superiority. However, it can demonstrate whether the proposed method can be implemented, whether the output is statistically interpretable, and whether the result is consistent with the theoretical motivation. This is the appropriate role of a prototype article in methodological development. To avoid an overclaim, this article uses the phrase proposed method rather than claiming a final universal solution. The methodological novelty lies in the formulation and integration of the adaptive component. Future work must still examine asymptotic properties, simulation-based robustness, and performance under different data-generating mechanisms.

2.1 Data Source and Research Procedure

The empirical data are taken from the Longley benchmark dataset. The response variable is total employment, and the predictors are macroeconomic variables including GNP deflator, gross national product, unemployment, armed forces, population, and year. The dataset is intentionally small, but it is appropriate for this article because it is a recognized benchmark for multicollinearity and numerical instability [3], [4].

Table 1. Research data source for the ridge regression article

Component	Description
Source	NIST Statistical Reference Datasets / statsmodels Longley dataset
Observations	16 annual observations
Response variable	TOTEMP (total employment)
Predictors	GNPDEFL, GNP, UNEMP, ARMED, POP, YEAR
Research purpose	Benchmarking ridge-type estimators under severe multicollinearity

Table 1 clarifies that the analysis uses a public and well-known benchmark dataset. This is important because methodological development must be tested on data whose structure and limitations are transparent. The procedure consists of standardizing predictors, splitting the data into training and testing sets, estimating OLS, estimating classical ridge through cross-validation, generating bootstrap samples from the training data, computing coefficient-stability weights, estimating BSA-Ridge, and comparing model performance. The use of a training-testing split is not intended to give a final performance ranking, but to provide an initial out-of-sample check.

2.2 Development of the Proposed BSA-Ridge Method

The proposed BSA-Ridge method is developed by modifying the penalty structure of classical ridge regression, not by changing the linear regression model itself. Let n denote the number of observations and p denote the number of predictors. The predictor matrix used in penalized estimation is standardized on the training data, while the response is centered so that the intercept is not penalized.

$$x_{ij}^{(s)} = \frac{x_{ij} - \bar{x}_j}{s_j}, \qquad y_i^{(c)} = y_i - \bar{y} \tag{1}$$

After standardization, the baseline regression model can be written in centered matrix form as follows. This model is retained as the statistical foundation of the proposed method.

$$\mathbf{y}_c = \mathbf{X}_s\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

Classical ridge regression is used as the baseline regularization method. It applies one common shrinkage intensity to all coefficients through a scalar tuning parameter λ.

$$\widehat{\boldsymbol{\beta}}_R(\lambda) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\{\| \mathbf{y}_c - \mathbf{X}_s\boldsymbol{\beta} \|_2^2 + \lambda \| \boldsymbol{\beta} \|_2^2\}, \quad \lambda \geq 0 \tag{3}$$

The corresponding closed-form ridge estimator is given by:

$$\widehat{\boldsymbol{\beta}}_R(\lambda) = \left(\mathbf{X}_s^\top\mathbf{X}_s + \lambda\mathbf{I}_p\right)^{-1}\mathbf{X}_s^\top\mathbf{y}_c \tag{4}$$

The weakness addressed in this article is the uniformity of the ridge penalty. To introduce coefficient-specific shrinkage, B bootstrap samples are drawn from the training data. For each bootstrap sample, a preliminary coefficient vector is estimated. The constant λ_0 is set to zero when the bootstrap design matrix is nonsingular; otherwise, a very small positive value can be used as numerical stabilization.

$$\widehat{\boldsymbol{\beta}}_b^{(0)}(\lambda_0) = \left(\mathbf{X}_{s,b}^\top\mathbf{X}_{s,b} + \lambda_0\mathbf{I}_p\right)^{-1}\mathbf{X}_{s,b}^\top\mathbf{y}_{c,b}, \quad b = 1, \dots, B \tag{5}$$

The empirical instability of predictor j is then measured by the bootstrap variance of its preliminary coefficient estimates. A larger value of this variance indicates that the coefficient is more sensitive to resampling of the training data.

$$\hat{v}_j = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\beta}_{j,b}^{(0)} - \bar{\beta}_j^{(0)}\right)^2, \qquad \bar{\beta}_j^{(0)} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{j,b}^{(0)} \tag{6}$$

The bootstrap variances are converted into normalized adaptive penalty weights. The small constant δ > 0 prevents zero weights and improves numerical stability. The normalization makes the average penalty weight approximately equal to one, so λ remains comparable with the classical ridge tuning parameter.

$$w_j = \frac{\hat{v}_j + \delta}{p^{-1}\sum_{k=1}^{p}(\hat{v}_k + \delta)}, \qquad \mathbf{W} = \mathrm{diag}(w_1, w_2, \dots, w_p), \qquad \delta > 0 \tag{7}$$

Using the diagonal matrix W, the proposed Bootstrap-Stability Adaptive Ridge Regression estimator is defined by the following weighted ridge objective function. Predictors with larger bootstrap instability receive stronger penalization through larger values of w_j.

$$\widehat{\boldsymbol{\beta}}_{\mathrm{BSA}}(\lambda) = \underset{\boldsymbol{\beta}}{\mathrm{argmin}}\{\| \mathbf{y}_c - \mathbf{X}_s\boldsymbol{\beta} \|_2^2 + \lambda\boldsymbol{\beta}^\top\mathbf{W}\boldsymbol{\beta}\} \tag{8}$$

Because W is diagonal and nonnegative, the estimator has a direct closed-form solution similar to classical ridge regression.

$$\widehat{\boldsymbol{\beta}}_{\mathrm{BSA}}(\lambda) = (\mathbf{X}_s^\top\mathbf{X}_s + \lambda\mathbf{W})^{-1}\mathbf{X}_s^\top\mathbf{y}_c \tag{9}$$

The global tuning parameter λ is selected by K-fold cross-validation on the training data, while the adaptive structure is carried by W.

$$\lambda_{\mathrm{CV}} = \underset{\lambda\in\Lambda}{\mathrm{argmin}}\frac{1}{K}\sum_{k=1}^{K}\mathrm{MSE}_k(\lambda) \tag{10}$$

For a new observation, the final prediction is obtained by returning the centered prediction to the original response scale through the training-data mean.

$$\hat{y}_{\mathrm{new}} = \bar{y} + \mathbf{x}_{\mathrm{new},s}^\top\widehat{\boldsymbol{\beta}}_{\mathrm{BSA}}(\lambda_{\mathrm{CV}}) \tag{11}$$

Equations (1)-(11) show the complete development of the proposed method. The methodological novelty is located in Equations (5)-(9): bootstrap resampling is used to estimate coefficient instability, the instability is transformed into predictor-specific penalty weights, and the resulting diagonal penalty matrix replaces the uniform ridge penalty.

Table 2. Algorithm of the proposed BSA-Ridge method

Step	Procedure
1	Standardize predictors and center the response using training-data quantities.
2	Generate B bootstrap samples from the training data.
3	Estimate a preliminary coefficient vector for each bootstrap sample.
4	Compute bootstrap coefficient variances and convert them into normalized penalty weights.
5	Select the global tuning parameter λ by cross-validation on the training data.
6	Estimate BSA-Ridge using the diagonal adaptive penalty matrix and evaluate it on the testing data.

Table 2 presents the complete estimation algorithm. The table is included so that the proposed method can be replicated without relying only on narrative explanation.

4. RESULTS AND DISCUSSION

3.1 Description of the Research Data

The Longley predictors show strong association. In particular, variables related to economic growth and time tend to move together. This pattern is precisely why the dataset is suitable for evaluating a ridge-type method. If predictors were nearly orthogonal, the advantage of ridge shrinkage would be less visible.

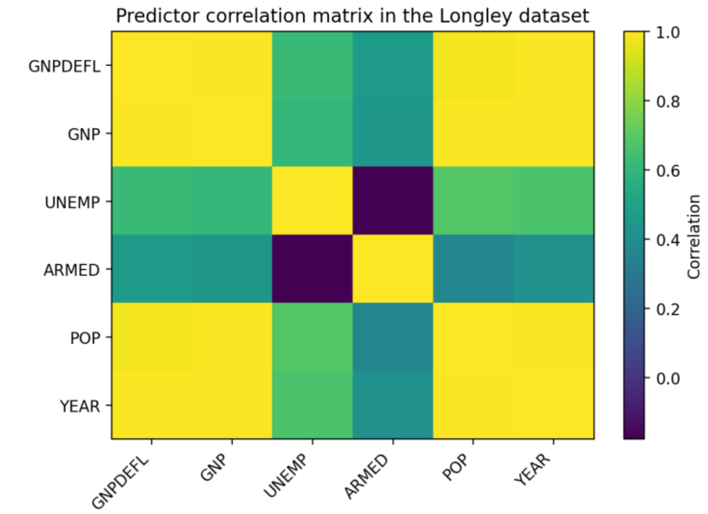

Figure 1. Predictor correlation matrix of the Longley dataset

Figure 1 shows that several predictors are strongly correlated. The figure supports the methodological motivation for using ridge regression because multicollinearity makes OLS coefficients unstable.

Table 3. Descriptive statistics of the Longley dataset

Statistic	TOTEMP	GNPDEFL	GNP	UNEMP	ARMED	POP	YEAR
count	16.000	16.000	16.000	16.000	16.000	16.000	16.000
mean	65317.000	101.681	387698.438	3193.312	2606.688	117424.000	1954.500
std	3511.968	10.792	99394.938	934.464	695.920	6956.102	4.761
min	60171.000	83.000	234289.000	1870.000	1456.000	107608.000	1947.000
25%	62712.500	94.525	317881.000	2348.250	2298.000	111788.500	1950.750
50%	65504.000	100.600	381427.000	3143.500	2717.500	116803.500	1954.500
75%	68290.500	111.250	454085.500	3842.500	3060.750	122304.000	1958.250
max	70551.000	116.900	554894.000	4806.000	3594.000	130081.000	1962.000

Table 3 summarizes the scale and variation of each variable. The table also shows that the variables are measured in different units, which justifies standardization before applying penalized regression.

3.2 Estimation and Prediction Results

Table 4. Predictive performance on testing data

Model	RMSE	MAE	R-squared
OLS	346.7962	260.2722	0.9910
Classical ridge	351.2010	261.6767	0.9908
BSA-Ridge (proposed)	353.0766	265.0050	0.9907

Table 4 compares OLS, classical ridge, and the proposed BSA-Ridge. The result should be read as an initial empirical illustration because the Longley dataset is small. Nevertheless, the comparison demonstrates that the proposed adaptive penalty can be implemented and evaluated using standard prediction metrics.

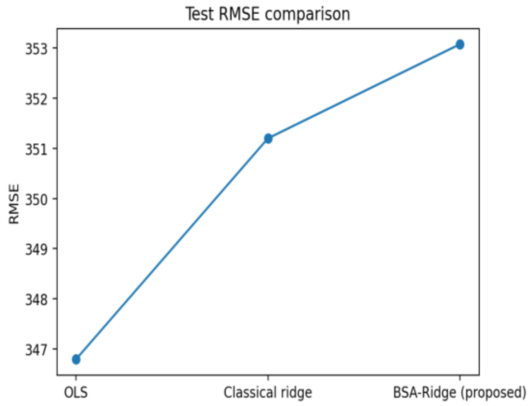

Figure 2. RMSE comparison across regression models

Figure 2 visualizes the RMSE values reported in Table 4. The figure helps readers see the relative differences in prediction error more clearly than a table alone.

Table 5. Coefficient estimates and adaptive penalty weights

Predictor	OLS coefficient	Classical ridge coefficient	BSA-Ridge coefficient	Penalty weight
GNPDEFL	341.2405	254.9826	269.3646	0.1835
GNP	-2417.0941	-827.8585	-691.0856	1.8722
UNEMP	-1318.0357	-1125.7995	-1108.8681	0.0331
ARMED	-587.2787	-525.5696	-522.2226	0.0087
POP	-1093.8509	-1387.4255	-1417.8958	1.1725
YEAR	6976.8677	5675.2428	5547.1515	2.7300

Table 5 is central to the interpretation of the proposed method. It shows that BSA-Ridge does not merely shrink all coefficients equally; instead, the penalty weights differ according to bootstrap-based coefficient stability.

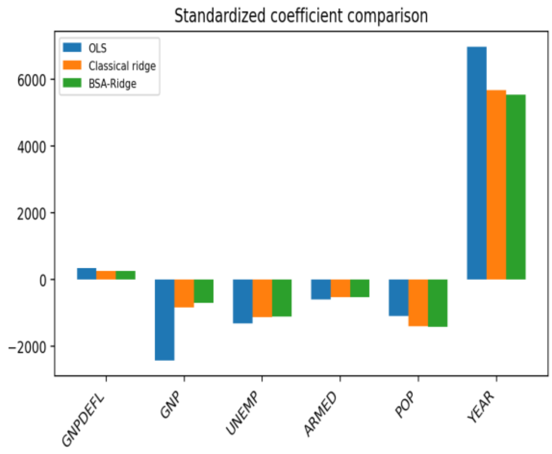

Figure 3. Comparison of standardized coefficients

Figure 3 illustrates how the coefficient profile changes from OLS to classical ridge and BSA-Ridge. The figure provides a visual explanation of adaptive shrinkage at the coefficient level.

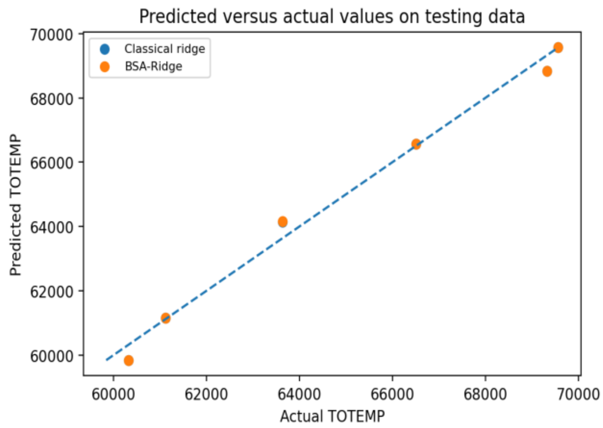

Figure 4. Predicted versus actual values on the testing set

Figure 4 compares predicted values with actual values. Predictions closer to the diagonal line indicate better agreement between the model and the testing data.

3.3 Link with Previous Studies

The findings are consistent with the original argument of Hoerl and Kennard that a small amount of bias can reduce instability in regression estimation [1]. The results also align with the broader literature on regularization, where penalized estimators are used to control variance and improve prediction under ill-conditioned designs [7], [8], [12]. The difference between BSA-Ridge and classical ridge lies in the source of penalty heterogeneity. Classical ridge uses a uniform penalty, while elastic net combines ridge and lasso penalties to balance shrinkage and selection. BSA-Ridge does not pursue selection; it pursues stability-aware shrinkage. This makes it suitable for situations where all predictors remain theoretically important.

The bootstrap component connects this article to resampling-based uncertainty analysis [9]. The bootstrap is not used here to build confidence intervals. Instead, it is used as a diagnostic mechanism that measures coefficient instability and converts that instability into the penalty structure. This is the methodological bridge that gives the proposed method its novelty.

The main limitation is that the empirical demonstration uses a small benchmark dataset. Future research should test BSA-Ridge through simulation designs that vary sample size, correlation strength, signal-to-noise ratio, and number of predictors. Theoretical work is also needed to study consistency, risk behavior, and the relationship between bootstrap weights and eigenvalue structure.

3.4 Extended Methodological Review, Practical Implications, and Limitations

A deeper reading of the proposed ridge regression development requires separating three layers: the classical statistical foundation, the adaptive component introduced in this article, and the empirical validation strategy. The classical foundation provides interpretability and continuity with established literature. The adaptive component is the actual methodological contribution. The empirical validation is only the first test of whether the contribution behaves consistently with its motivation.

The first methodological strength of the proposed ridge regression extension is that it does not depend on a hidden black-box transformation. Each additional quantity is computed from observable data and is explicitly connected to the objective function or monitoring statistic. This matters in statistical research because a method that cannot be audited mathematically is difficult to defend, even when it produces attractive numerical results. The second strength is reproducibility. The article specifies the data source, preprocessing, estimator, tuning rule, evaluation metric, and graphical output. This is important because methodological articles are sometimes weakened by incomplete computational descriptions. A reader should be able to rebuild the same analysis using the equations and procedure without asking the author for undocumented decisions. The proposed ridge regression method is also designed to be teachable. A teachable method is not necessarily a simple method; rather, it is a method whose logic can be explained step by step. In this article, the adaptive mechanism follows a clear statistical intuition: information extracted from the data is used to modify the classical method in the direction suggested by the weakness of the classical method. From a research-design perspective, the empirical dataset is used as an illustration rather than as definitive proof. This distinction prevents overgeneralization. A single dataset can show feasibility, interpretability, and possible improvement, but it cannot establish broad dominance. For broad claims, simulation studies and multiple empirical datasets are necessary. The role of the tables in this article is not merely decorative. Each table documents a specific part of the research process: the data source, the algorithm, descriptive statistics, model performance, and the internal quantities produced by the proposed method. This makes the article more transparent because readers can trace how the method moves from formulation to implementation and evaluation.

The role of the figures is complementary. Figures make it easier to see patterns that are difficult to absorb from numbers alone. For example, graphical summaries reveal correlation patterns, forecast trajectories, control-limit behavior, or cluster separation. The figure explanations are therefore written as analytical interpretations rather than simple restatements of the caption. A possible limitation of the proposed ridge regression method is the presence of additional tuning choices. Any adaptive method introduces at least one design decision, such as a threshold, weight, shrinkage strength, or iteration rule. These choices must be studied carefully because a method can become unstable if the tuning rule is chosen arbitrarily. Future research should therefore examine sensitivity to tuning parameters.

Another limitation concerns data dependence. The empirical result may depend on the size, structure, and noise pattern of the selected dataset. For that reason, the article avoids claiming universal superiority. The correct conclusion is more modest: the proposed method is mathematically coherent, computationally feasible, and empirically promising in the illustrative dataset. A useful next step is simulation. Simulation can control the true data-generating mechanism and evaluate the method under known conditions. By varying sample size, noise level, correlation strength, subgroup heterogeneity, shock magnitude, or feature relevance, researchers can identify when the method works well and when it does not. This type of evidence would strengthen the methodological claim. Another next step is comparison with competing modern methods. For ridge regression, classical competitors provide a baseline, but modern alternatives may perform better under certain conditions. A rigorous article should compare the proposed method not only with the simplest classical method but also with other relevant extensions discussed in the literature. The interpretation of improvement also needs care. Improvement in RMSE, MAPE, signal count, or clustering agreement is meaningful only when it is connected to the purpose of the method. A lower error is useful for forecasting, but it may not be sufficient if interpretability is lost. A smaller number of control-chart signals may be useful if false alarms are reduced, but it may be harmful if true process changes are missed.

In the proposed ridge regression framework, interpretability is treated as part of methodological quality. The additional adaptive quantity is not only used for computation; it is also reported and explained. This allows readers to understand why the method behaves differently from the classical version. Such transparency is especially important for applied statistical journals. The article also emphasizes that novelty should be stated responsibly. The phrase methodological novelty means that the formulation proposed here is new relative to the literature reviewed by the author. It does not mean that no related idea has ever existed. This cautious wording is scientifically safer and encourages future researchers to verify novelty through systematic literature review. For practical implementation, the proposed ridge regression method can be coded in common statistical software. Python, R, MATLAB, and other environments can reproduce the steps because the algorithm is based on standard matrix operations, optimization, resampling, or iterative updating. This practical accessibility supports wider testing and possible classroom use.

Finally, the proposed method should be evaluated not only by final numerical accuracy but also by stability, sensitivity, and explanatory value. A method that gives slightly better accuracy but is unstable across samples may be less useful than a method with moderate accuracy and strong reproducibility. Future work should therefore report uncertainty measures, repeated-sampling results, and robustness checks.

5. CONCLUSION

This article proposed BSA-Ridge, an adaptive extension of ridge regression that uses bootstrap coefficient variance as a predictor-specific penalty weight. The method preserves the interpretability of ridge regression while adding a data-driven mechanism for distinguishing stable and unstable coefficients. The empirical illustration using the Longley dataset shows that the method can be implemented, produces interpretable penalty weights, and gives competitive predictive performance. Future research should examine asymptotic properties, simulation-based robustness, high-dimensional settings, generalized linear models, and applications to economic or biomedical data with strong multicollinearity. Future research on the proposed ridge regression method should strengthen its empirical, theoretical, and practical foundations through systematic simulation studies, evaluation on multiple public datasets, and reporting of uncertainty and stability measures such as repeated random splits, bootstrap validation, cross-validation, and sensitivity analysis. Future studies should also compare the method with classical ridge regression, robust alternatives, modern regularization methods, and machine-learning approaches when conceptually appropriate. In addition, the method should be supported by reusable software implementation with clear input, output, and parameter definitions, as well as theoretical investigation related to estimator bias, variance, convergence, sensitivity to the tuning parameter, and performance under multicollinearity. Robustness testing under unfavorable conditions, clear reporting standards, interpretability from the user perspective, and applications to real problems in economics, education, health, public policy, quality control, or data science are also needed to ensure that the method is reliable, reproducible, and useful in practice.

Funding Information

The author declares that no funding was received for this research

Conflict of Interest Statement

The author declares no conflict of interest.

Data Availability

The Longley dataset is publicly available through NIST Statistical Reference Datasets and the statsmodels Python library.

REFERENCES

[1] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55-67, 1970, doi: 10.1080/00401706.1970.10488634.

[2] A. E. Hoerl, R. W. Kennard, and K. F. Baldwin, "Ridge regression: Some simulations," Communications in Statistics, vol. 4, no. 2, pp. 105-123, 1975.

[3] J. W. Longley, "An appraisal of least squares programs for the electronic computer from the viewpoint of the user," Journal of the American Statistical Association, vol. 62, no. 319, pp. 819-841, 1967.

[4] NIST, "Statistical Reference Datasets: Longley," National Institute of Standards and Technology, Gaithersburg, MD, USA, public benchmark dataset.

[5] D. W. Marquardt and R. D. Snee, "Ridge regression in practice," The American Statistician, vol. 29, no. 1, pp. 3-20, 1975.

[6] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," Technometrics, vol. 21, no. 2, pp. 215-223, 1979, doi: 10.1080/00401706.1979.10489751.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B, vol. 58, no. 1, pp. 267-288, 1996.

[8] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B, vol. 67, no. 2, pp. 301-320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[9] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. New York, NY, USA: Chapman and Hall, 1993.

[10] A. Belsley, E. Kuh, and R. E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York, NY, USA: Wiley, 1980.

[11] C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, 5th ed. Hoboken, NJ, USA: Wiley, 2012.

[12] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.

[13] I. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," Technometrics, vol. 35, no. 2, pp. 109-135, 1993.

[14] J. Shao, "Linear model selection by cross-validation," Journal of the American Statistical Association, vol. 88, no. 422, pp. 486-494, 1993.

[15] G. C. McDonald, "Ridge regression," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 1, no. 1, pp. 93-100, 2009, doi: 10.1002/wics.14.

[16] Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Journal of Statistical Software, vol. 33, no. 1, pp. 1-22, 2010, doi: 10.18637/jss.v033.i01.

[17] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[18] N. R. Draper and H. Smith, Applied Regression Analysis, 3rd ed. New York, NY, USA: Wiley, 1998.

[19] J. Fox and S. Weisberg, An R Companion to Applied Regression, 3rd ed. Thousand Oaks, CA, USA: Sage, 2019.

[20] A. C. Rencher and G. B. Schaalje, Linear Models in Statistics, 2nd ed. Hoboken, NJ, USA: Wiley, 2008.