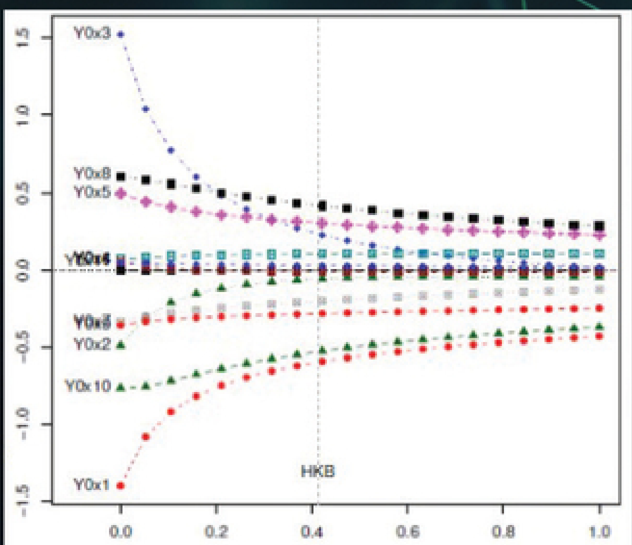
WILEY SERIES IN PROBABILITY AND STATISTICS

Theory of Ridge Regression Estimators with Applications

A.K. Md. Ehsanes Saleh | Mohammad Arashi
B.M. Golam Kibria



WILEY

Theory of Ridge Regression Estimation with Applications

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by *Walter A. Shewhart and Samuel S. Wilks*

The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches. This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

A complete list of titles in this series can be found at
http://www.wiley.com/go/wsps

Theory of Ridge Regression Estimation with Applications

A.K. Md. Ehsanes Saleh
Carleton University
Ottawa, Canada

Mohammad Arashi
Shahrood University of Technology, Shahrood, Iran
University of Pretoria, Pretoria, South Africa

B.M. Golam Kibria
Florida International University
Miami, Florida, USA

WILEY

This edition first published 2019

Registered Office
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data Applied For
ISBN: 9781118644614 (hardback)

Cover design: Wiley
Cover image: © hrui/Shutterstock, Courtesy of B. M. Golam Kibria

Set in 10/12pt WarnockPro by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To our wives
Shahidara Saleh
Reihaneh Arashi
Farhana Kibria
(Orchi)

Contents in Brief

1 Introduction to Ridge Regression *1*

2 Location and Simple Linear Models *15*

3 ANOVA Model *43*

4 Seemingly Unrelated Simple Linear Models *79*

5 Multiple Linear Regression Models *109*

6 Ridge Regression in Theory and Applications *143*

7 Partially Linear Regression Models *171*

8 Logistic Regression Model *197*

9 Regression Models with Autoregressive Errors *221*

10 Rank-Based Shrinkage Estimation *251*

11 High-Dimensional Ridge Regression *285*

12 Applications: Neural Networks and Big Data *303*

Contents

List of Figures *xvii*
List of Tables *xxi*
Preface *xxvii*
Abbreviations and Acronyms *xxxi*
List of Symbols *xxxiii*

1 Introduction to Ridge Regression *1*
1.1 Introduction *1*
1.1.1 Multicollinearity Problem *3*
1.2 Ridge Regression Estimator: Ridge Notion *5*
1.3 LSE vs. RRE *6*
1.4 Estimation of Ridge Parameter *7*
1.5 Preliminary Test and Stein-Type Ridge Estimators *8*
1.6 High-Dimensional Setting *9*
1.7 Notes and References *11*
1.8 Organization of the Book *12*

2 Location and Simple Linear Models *15*
2.1 Introduction *15*
2.2 Location Model *16*
2.2.1 Location Model: Estimation *16*
2.2.2 Shrinkage Estimation of Location *17*
2.2.3 Ridge Regression–Type Estimation of Location Parameter *18*
2.2.4 LASSO for Location Parameter *18*
2.2.5 Bias and MSE Expression for the LASSO of Location Parameter *19*
2.2.6 Preliminary Test Estimator, Bias, and MSE *23*
2.2.7 Stein-Type Estimation of Location Parameter *24*
2.2.8 Comparison of LSE, PTE, Ridge, SE, and LASSO *24*
2.3 Simple Linear Model *26*
2.3.1 Estimation of the Intercept and Slope Parameters *26*

2.3.2 Test for Slope Parameter *27*
2.3.3 PTE of the Intercept and Slope Parameters *27*
2.3.4 Comparison of Bias and MSE Functions *29*
2.3.5 Alternative PTE *31*
2.3.6 Optimum Level of Significance of Preliminary Test *33*
2.3.7 Ridge-Type Estimation of Intercept and Slope *34*
2.3.7.1 Bias and MSE Expressions *35*
2.3.8 LASSO Estimation of Intercept and Slope *36*
2.4 Summary and Concluding Remarks *39*

3 ANOVA Model *43*
3.1 Introduction *43*
3.2 Model, Estimation, and Tests *44*
3.2.1 Estimation of Treatment Effects *45*
3.2.2 Test of Significance *45*
3.2.3 Penalty Estimators *46*
3.2.4 Preliminary Test and Stein-Type Estimators *47*
3.3 Bias and Weighted L_2 Risks of Estimators *48*
3.3.1 Hard Threshold Estimator (Subset Selection Rule) *48*
3.3.2 LASSO Estimator *49*
3.3.3 Ridge Regression Estimator *51*
3.4 Comparison of Estimators *52*
3.4.1 Comparison of LSE with RLSE *52*
3.4.2 Comparison of LSE with PTE *52*
3.4.3 Comparison of LSE with SE and PRSE *53*
3.4.4 Comparison of LSE and RLSE with RRE *54*
3.4.5 Comparison of RRE with PTE, SE, and PRSE *56*
3.4.5.1 Comparison Between $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}(\alpha)$ *56*
3.4.5.2 Comparison Between $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ *56*
3.4.5.3 Comparison of $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ with $\hat{\boldsymbol{\theta}}_n^{\mathrm{S+}}$ *57*
3.4.6 Comparison of LASSO with LSE and RLSE *58*
3.4.7 Comparison of LASSO with PTE, SE, and PRSE *59*
3.4.8 Comparison of LASSO with RRE *60*
3.5 Application *60*
3.6 Efficiency in Terms of Unweighted L_2 Risk *63*
3.7 Summary and Concluding Remarks *72*
3A. Appendix *74*

4 Seemingly Unrelated Simple Linear Models *79*
4.1 Model, Estimation, and Test of Hypothesis *79*
4.1.1 LSE of θ and β *80*
4.1.2 Penalty Estimation of β and θ *80*

4.1.3 PTE and Stein-Type Estimators of β and θ *81*
4.2 Bias and MSE Expressions of the Estimators *82*
4.3 Comparison of Estimators *86*
4.3.1 Comparison of LSE with RLSE *86*
4.3.2 Comparison of LSE with PTE *86*
4.3.3 Comparison of LSE with SE and PRSE *87*
4.3.4 Comparison of LSE and RLSE with RRE *87*
4.3.5 Comparison of RRE with PTE, SE, and PRSE *89*
4.3.5.1 Comparison Between $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}$ *89*
4.3.5.2 Comparison Between $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ *89*
4.3.5.3 Comparison of $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ with $\hat{\boldsymbol{\theta}}_n^{\mathrm{S+}}$ *90*
4.3.6 Comparison of LASSO with RRE *90*
4.3.7 Comparison of LASSO with LSE and RLSE *92*
4.3.8 Comparison of LASSO with PTE, SE, and PRSE *92*
4.4 Efficiency in Terms of Unweighted L_2 Risk *93*
4.4.1 Efficiency for $\boldsymbol{\beta}$ *94*
4.4.2 Efficiency for $\boldsymbol{\theta}$ *95*
4.5 Summary and Concluding Remarks *96*

5 Multiple Linear Regression Models *109*
5.1 Introduction *109*
5.2 Linear Model and the Estimators *110*
5.2.1 Penalty Estimators *111*
5.2.2 Shrinkage Estimators *113*
5.3 Bias and Weighted L_2 Risks of Estimators *114*
5.3.1 Hard Threshold Estimator *114*
5.3.2 Modified LASSO *116*
5.3.3 Multivariate Normal Decision Theory and Oracles for Diagonal Linear Projection *117*
5.3.4 Ridge Regression Estimator *119*
5.3.5 Shrinkage Estimators *119*
5.4 Comparison of Estimators *120*
5.4.1 Comparison of LSE with RLSE *120*
5.4.2 Comparison of LSE with PTE *121*
5.4.3 Comparison of LSE with SE and PRSE *121*
5.4.4 Comparison of LSE and RLSE with RRE *122*
5.4.5 Comparison of RRE with PTE, SE, and PRSE *123*
5.4.5.1 Comparison Between $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{PT}}(\alpha)$ *123*
5.4.5.2 Comparison Between $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\theta}}_n^{\mathrm{S}}$ *124*
5.4.5.3 Comparison of $\hat{\boldsymbol{\theta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$ with $\hat{\boldsymbol{\theta}}_n^{\mathrm{S+}}$ *124*
5.4.6 Comparison of MLASSO with LSE and RLSE *125*

5.4.7 Comparison of MLASSO with PTE, SE, and PRSE *126*
5.4.8 Comparison of MLASSO with RRE *127*
5.5 Efficiency in Terms of Unweighted L_2 Risk *127*
5.6 Summary and Concluding Remarks *129*

6 Ridge Regression in Theory and Applications *143*
6.1 Multiple Linear Model Specification *143*
6.1.1 Estimation of Regression Parameters *143*
6.1.2 Test of Hypothesis for the Coefficients Vector *145*
6.2 Ridge Regression Estimators (RREs) *146*
6.3 Bias, MSE, and L_2 Risk of Ridge Regression Estimator *147*
6.4 Determination of the Tuning Parameters *151*
6.5 Ridge Trace *151*
6.6 Degrees of Freedom of RRE *154*
6.7 Generalized Ridge Regression Estimators *155*
6.8 LASSO and Adaptive Ridge Regression Estimators *156*
6.9 Optimization Algorithm *158*
6.9.1 Prostate Cancer Data *160*
6.10 Estimation of Regression Parameters for Low-Dimensional Models *161*
6.10.1 BLUE and Ridge Regression Estimators *161*
6.10.2 Bias and L_2-risk Expressions of Estimators *162*
6.10.3 Comparison of the Estimators *165*
6.10.4 Asymptotic Results of RRE *166*
6.11 Summary and Concluding Remarks *168*

7 Partially Linear Regression Models *171*
7.1 Introduction *171*
7.2 Partial Linear Model and Estimation *172*
7.3 Ridge Estimators of Regression Parameter *174*
7.4 Biases and L_2 Risks of Shrinkage Estimators *177*
7.5 Numerical Analysis *178*
7.5.1 Example: Housing Prices Data *182*
7.6 High-Dimensional PLM *188*
7.6.1 Example: Riboflavin Data *192*
7.7 Summary and Concluding Remarks *193*

8 Logistic Regression Model *197*
8.1 Introduction *197*
8.1.1 Penalty Estimators *199*
8.1.2 Shrinkage Estimators *200*
8.1.3 Results on MLASSO *201*
8.1.4 Results on PTE and Stein-Type Estimators *202*

8.1.5 Results on Penalty Estimators *204*
8.2 Asymptotic Distributional L_2 Risk Efficiency Expressions of the Estimators *204*
8.2.1 MLASSO vs. MLE *205*
8.2.2 MLASSO vs. RMLE *206*
8.2.3 Comparison of MLASSO vs. PTE *206*
8.2.4 PT and MLE *207*
8.2.5 Comparison of MLASSO vs. SE *208*
8.2.6 Comparison of MLASSO vs. PRSE *208*
8.2.7 RRE vs. MLE *209*
8.2.7.1 RRE vs. RMLE *209*
8.2.8 Comparison of RRE vs. PTE *211*
8.2.9 Comparison of RRE vs. SE *211*
8.2.10 Comparison of RRE vs. PRSE *212*
8.2.11 PTE vs. SE and PRSE *212*
8.2.12 Numerical Comparison Among the Estimators *213*
8.3 Summary and Concluding Remarks *213*

9 Regression Models with Autoregressive Errors *221*
9.1 Introduction *221*
9.1.1 Penalty Estimators *223*
9.1.2 Shrinkage Estimators *224*
9.1.2.1 Preliminary Test Estimator *224*
9.1.2.2 Stein-Type and Positive-Rule Stein-Type Estimators *225*
9.1.3 Results on Penalty Estimators *225*
9.1.4 Results on PTE and Stein-Type Estimators *226*
9.1.5 Results on Penalty Estimators *229*
9.2 Asymptotic Distributional L_2-risk Efficiency Comparison *230*
9.2.1 Comparison of GLSE with RGLSE *230*
9.2.2 Comparison of GLSE with PTE *231*
9.2.3 Comparison of LSE with SE and PRSE *231*
9.2.4 Comparison of LSE and RLSE with RRE *232*
9.2.5 Comparison of RRE with PTE, SE and PRSE *233*
9.2.5.1 Comparison Between $\hat{\boldsymbol{\beta}}_n^{\mathrm{GRR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\beta}}_n^{\mathrm{G(PT)}}$ *233*
9.2.5.2 Comparison Between $\hat{\boldsymbol{\beta}}_n^{\mathrm{GRR}}(k_{\mathrm{opt}})$ and $\hat{\boldsymbol{\beta}}_n^{\mathrm{G(S)}}$ *234*
9.2.5.3 Comparison of $\hat{\boldsymbol{\beta}}_n^{\mathrm{GRR}}(k_{\mathrm{opt}})$ with $\hat{\boldsymbol{\beta}}_n^{\mathrm{G(S+)}}$ *234*
9.2.6 Comparison of MLASSO with GLSE and RGLSE *235*
9.2.7 Comparison of MLASSO with PTE, SE, and PRSE *236*
9.2.8 Comparison of MLASSO with RRE *236*
9.3 Example: Sea Level Rise at Key West, Florida *237*
9.3.1 Estimation of the Model Parameters *237*
9.3.1.1 Testing for Multicollinearity *237*

9.3.1.2 Testing for Autoregressive Process *238*
9.3.1.3 Estimation of Ridge Parameter k *239*
9.3.2 Relative Efficiency *240*
9.3.2.1 Relative Efficiency (REff) *240*
9.3.2.2 Effect of Autocorrelation Coefficient ϕ *243*
9.4 Summary and Concluding Remarks *245*

10 Rank-Based Shrinkage Estimation *251*
10.1 Introduction *251*
10.2 Linear Model and Rank Estimation *252*
10.2.1 Penalty R-Estimators *256*
10.2.2 PTREs and Stein-type R-Estimators *258*
10.3 Asymptotic Distributional Bias and L_2 Risk of the R-Estimators *259*
10.3.1 Hard Threshold Estimators (Subset Selection) *259*
10.3.2 Rank-based LASSO *260*
10.3.3 Multivariate Normal Decision Theory and Oracles for Diagonal Linear Projection *261*
10.4 Comparison of Estimators *262*
10.4.1 Comparison of RE with Restricted RE *262*
10.4.2 Comparison of RE with PTRE *263*
10.4.3 Comparison of RE with SRE and PRSRE *263*
10.4.4 Comparison of RE and Restricted RE with RRRE *265*
10.4.5 Comparison of RRRE with PTRE, SRE, and PRSRE *266*
10.4.6 Comparison of RLASSO with RE and Restricted RE *267*
10.4.7 Comparison of RLASSO with PTRE, SRE, and PRSRE *267*
10.4.8 Comparison of Modified RLASSO with RRRE *268*
10.5 Summary and Concluding Remarks *268*

11 High-Dimensional Ridge Regression *285*
11.1 High-Dimensional RRE *286*
11.2 High-Dimensional Stein-Type RRE *288*
11.2.1 Numerical Results *291*
11.2.1.1 Example: Riboflavin Data *291*
11.2.1.2 Monte Carlo Simulation *291*
11.3 Post Selection Shrinkage *293*
11.3.1 Notation and Assumptions *296*
11.3.2 Estimation Strategy *297*
11.3.3 Asymptotic Distributional L_2-Risks *299*
11.4 Summary and Concluding Remarks *300*

12 Applications: Neural Networks and Big Data *303*
12.1 Introduction *304*
12.2 A Simple Two-Layer Neural Network *307*

12.2.1 Logistic Regression Revisited *307*
12.2.2 Logistic Regression Loss Function with Penalty *310*
12.2.3 Two-Layer Logistic Regression *311*
12.3 Deep Neural Networks *313*
12.4 Application: Image Recognition *315*
12.4.1 Background *315*
12.4.2 Binary Classification *316*
12.4.3 Image Preparation *318*
12.4.4 Experimental Results *320*
12.5 Summary and Concluding Remarks *323*
References *325*

Index *333*

List of Figures

Figure 2.1 Relative efficiencies of the estimators. *26*

Figure 2.2 Graph of $\text{REff}(\hat{\theta}_n; \tilde{\theta}_n)$ and $\text{REff}(\hat{\theta}_n^{\text{PT}}(\alpha); \tilde{\theta}_n)$ for $n = 8$ and $\bar{x}^2/Q = 0.5$. *32*

Figure 2.3 Relative efficiency of the estimators for θ. *38*

Figure 2.4 Relative efficiency of the estimators for β. *39*

Figure 3.1 RWRE for the restricted estimator. *54*

Figure 3.2 RWRE for the preliminary test estimator. *55*

Figure 3.3 RWRE for the Stein-type and its positive-rule estimator. *55*

Figure 3.4 Weighted L_2 risk for the ridge, preliminary test, and Stein-type and its positive-rule estimators for $p_1 = 5,\ p_2 = 15$, and $\alpha = 0.20$. *58*

Figure 3.5 RWRE for the ridge, preliminary test, and Stein-type and its positive-rule estimators for $p_1 = 7,\ p_2 = 33$, and $\alpha = 0.20$. *59*

Figure 3.6 RWRE for LASSO, ridge, restricted, preliminary test, and Stein-type and its positive-rule estimators. *61*

Figure 3.7 RWRE of estimates of a function of Δ^2 for $p_1 = 5$ and different p_2. *73*

Figure 4.1 RWRE for the restricted, preliminary test, and Stein-type and its positive-rule estimators. *88*

Figure 4.2 Weighted L_2 risk for the ridge, preliminary test, and Stein-type and its positive-rule estimators for $p_1 = 5,\ p_2 = 15$, and $\alpha = 0.20$. *91*

Figure 4.3 Weighted L_2 risk for the ridge, preliminary test, and Stein-type and its positive-rule estimators for $p_1 = 7,\ p_2 = 33$, and $\alpha = 0.20$. *91*

Figure 4.4 RWRE for the LASSO, ridge, restricted, preliminary test, and Stein-type and its positive-rule estimators. *93*

Figure 4.5 RWRE of estimates of a function of Δ^2 for $p_1 = 5$ and different p_2. *106*

Figure 5.1 Weighted L_2 risk for the ridge, preliminary test, and Stein-type and its positive-rule estimators for $p_1 = 5, p_2 = 15$, and $\alpha = 0.20$. *125*

Figure 5.2 Weighted L_2 risk for the ridge, preliminary test, and Stein-type and its positive-rule estimators for $p_1 = 7,\ p_2 = 33$, and $\alpha = 0.20$. *126*

Figure 6.1 Graph of L_2 risk of RRE in case of $p = 10$. *149*

Figure 6.2 Ridge trace for gasoline mileage data. *153*

Figure 7.1 Plots of Δ and risks vs. k for different values of γ. *183*

Figure 7.2 Estimation of the mixtures of normal p.d.fs by the kernel approach. Solid lines are the estimates and dotted lines are the true functions. *184*

Figure 7.3 Plot of CV vs. h_n. *185*

Figure 7.4 Plots of individual explanatory variables vs. dependent variable, linear fit (dash line), and local polynomial fit (solid line). *186*

Figure 7.5 Estimation of nonlinear effect of ANI on dependent variable by kernel fit. *187*

Figure 7.6 The diagrams of Δ and risk vs. k for housing prices data. *188*

Figure 7.7 Estimation of f(ANI) by kernel regression after removing the linear part by the proposed estimators in housing prices data. *189*

Figure 7.8 Added-variable plot of explanatory variables *DNAJ_at* vs. dependent variable, linear fit (solid line) and kernel fit (dashed line). *192*

Figure 8.1 Relative efficiency for $p = 10$ and $p_1 = 3$. *217*

Figure 8.2 Relative efficiency for $p = 10$ and $p_2 = 5$. *217*

Figure 8.3 Relative efficiency for $p = 10$ and $p_2 = 7$. *218*

Figure 9.1 Relative efficiency of the estimators for $p_1 = 2, p_2 = 7$, $\phi = -0.34$, and $k = 0.10$. *242*

Figure 9.2 Relative efficiency of the estimators for $p_1 = 2, p_2 = 7$, $\phi = 0.35$, and $k = 0.10$. *245*

Figure 9.3 Relative efficiency of the estimators for $p_1 = 2, p_2 = 7$, $\phi = 0.55$, and $k = 0.10$. *245*

Figure 9.4 Relative efficiency of the estimators for $p_1 = 2, p_2 = 7$, $\phi = 0.75$, and $k = 0.10$. *246*

Figure 9.5 Relative efficiency of the estimators for $p_1 = 2, p_2 = 7$, $\phi = 0.85$, and $k = 0.10$. *247*

Figure 9.6 Relative efficiency of the estimators for $p_1 = 2, p_2 = 7$, $\phi = -0.85$, and $k = 0.10$. *247*

Figure 10.1 RWRE for the restricted, preliminary test, Stein-type, and its positive-rule R-estimators. *264*

Figure 10.2 RWRE for the modified RLASSO (MRLASSO), ridge, restricted, preliminary test and the Stein-type and its positive rule estimators. *280*

Figure 10.3 RWRE of R-estimates of a function of Δ^2 for $p_1 = 3$, $\rho^2 = 0.9$, and different p_2. *281*

Figure 11.1 Estimation of RSS based on the ridge parameter for riboflavin data example. *292*

Figure 11.2 Estimated risks for the estimators of model (11.3), for $\gamma = 0.60$ and different values of p. *293*

Figure 11.3 Estimated risks for the estimators of model (11.3), for $\gamma = 0.75$ and different values of p. *294*

Figure 11.4 Estimated risks for the estimators of model (11.3), for $\gamma = 0.90$ and different values of p. *295*

Figure 12.1 Data set aspect ratios and suitable methods. (a) Very small p, very small n = logistic regression. (b) Large p, small n = shallow neural network. (c) Smaller p, very large n = deep neural network. *305*

Figure 12.2 Computational flow graph for logistic regression. *308*

Figure 12.3 Logistic function used in logistic regression. *308*

Figure 12.4 Simplified flow graph for logistic regression. *311*

Figure 12.5 Two-layer neural network. *312*

Figure 12.6 Detailed equations for a two-layer neural network. *313*

Figure 12.7 A four-layer neural network. *314*

Figure 12.8 The relu activation function used in deep neural networks. *314*

Figure 12.9 Qualitative assessment of neural networks. *315*

Figure 12.10 Typical setup for supervised learning methods. *316*

Figure 12.11 Preparing the image for model building. *319*

Figure 12.12 Over-fitting vs. regularized training data. (a) Over-fitting. (b) Effect of L_2 penalty. *319*

List of Tables

Table 1.1 Model fit indices for Portland cement data. *5*

Table 1.2 Coefficient estimates for Portland cement data. *5*

Table 2.1 Table of relative efficiency. *25*

Table 2.2 Maximum and minimum guaranteed relative efficiency. *34*

Table 2.3 Relative efficiency of the estimators for θ. *37*

Table 2.4 Relative efficiency of the estimators for β. *38*

Table 3.1 One-way ANOVA table. *62*

Table 3.2 Estimated values of different estimators. *62*

Table 3.3 RWRE for the estimators. *64*

Table 3.4 RWRE of the estimators for $p = 10$ and different Δ^2 values for varying p_1. *65*

Table 3.5 RWRE of the estimators for $p = 20$ and different Δ^2 values for varying p_1. *66*

Table 3.6 RWRE of the estimators for $p = 40$ and different Δ^2 values for varying p_1. *67*

Table 3.7 RWRE of the estimators for $p = 60$ and different Δ^2 values for varying p_1. *68*

Table 3.8 RWRE values of estimators for $p_1 = 5$ and different values of p_2 and Δ^2. *69*

Table 3.9 RWRE values of estimators for $p_1 = 7$ and different values of p_2 and Δ^2. *70*

Table 3.10 RWRE values of estimators for $p_2 = 5$ and different values of p_1 and Δ^2. *71*

Table 3.11 RWRE values of estimators for $p_2 = 7$ and different values of p_1 and Δ^2. *75*

Table 3A.1 Sample efficiency table of estimators under Hansen's method. *76*

Table 4.1 RWRE for the estimators. *97*

Table 4.2 RWRE of the estimators for $p = 10$ and different Δ^2-value for varying p_1. *98*

Table 4.3 RWRE of the estimators for $p = 20$ and different Δ^2 values for varying p_1. *99*

Table 4.4 RWRE of the estimators for $p = 40$ and different Δ^2 values for varying p_1. *100*

Table 4.5 RWRE of the estimators for $p = 60$ and different Δ^2 values for varying p_1. *101*

Table 4.6 RWRE values of estimators for $p_1 = 5$ and different values of p_2 and Δ^2. *102*

Table 4.7 RWRE values of estimators for $p_1 = 7$ and different values of p_2 and Δ^2. *103*

Table 4.8 RWRE values of estimators for $p_2 = 5$ and different values of p_1 and Δ^2. *104*

Table 4.9 RWRE values of estimators for $p_2 = 7$ and different values of p_1 and Δ^2. *105*

Table 5.1 Relative weighted L_2-risk efficiency for the estimators for $p_1 = 5, p_2 = 15$. *130*

Table 5.2 Relative weighted L_2-risk efficiency for the estimators for $p_1 = 7, p_2 = 33$. *131*

Table 5.3 Relative weighted L_2-risk efficiency of the estimators for $p = 10$ and different Δ^2 values for varying p_1. *132*

Table 5.4 Relative weighted L_2-risk efficiency of the estimators for $p = 20$ and different Δ^2 values for varying p_1. *133*

Table 5.5 Relative weighted L_2-risk efficiency of the estimators for $p = 40, 60$ and different Δ^2 values for varying p_1. *134*

Table 5.6 Relative weighted L_2-risk efficiency of the estimators for $p = 60$ and different Δ^2 values for varying p_1. *135*

Table 5.7 Relative weighted L_2-risk efficiency values of estimators for $p_1 = 5$ and different values of p_2 and Δ^2. *136*

Table 5.8 Relative weighted L_2-risk efficiency values of estimators for $p_1 = 7$ and different values of p_2 and Δ^2. *137*

Table 5.9 Relative weighted L_2-risk efficiency values of estimators for $p_2 = 5$ and different values of p_1 and Δ^2. *138*

Table 5.10 Relative weighted L_2-risk efficiency values of estimators for $p_2 = 7$ and different values of p_1 and Δ^2. *139*

Table 6.1 Correlation coefficients for gasoline mileage data. *152*

Table 6.2 Estimated coefficients (standard errors) for prostate cancer data using LS, LASSO, and ARR estimators. *160*

Table 7.1 Evaluation of the Stein-type generalized RRE at different k values in model (7.24) with $\gamma = 0.80$. *180*

Table 7.2 Evaluation of PRSGRE at different k values in model (7.24) with $\gamma = 0.80$. *180*

Table 7.3 Evaluation of SGRE at different k values in model (7.24) with $\gamma = 0.90$. *181*

Table 7.4 Evaluation of PRSGRE at different k values in model (7.24) with $\gamma = 0.90$. *181*

Table 7.5 Evaluation of SGRE at different k values in model (7.24) with $\gamma = 0.99$. *181*

Table 7.6 Evaluation of PRSGRE at different k values in model (7.24) with $\gamma = 0.99$. *182*

Table 7.7 Correlation matrix. *185*

Table 7.8 Fitting of parametric and semi-parametric models to housing prices data. *186*

Table 7.9 Evaluation of SGRRE at different k values for housing prices data. *187*

Table 7.10 Evaluation of PRSGRRE at different k values for housing prices data. *188*

Table 7.11 Evaluation of proposed estimators for real data set. *193*

Table 8.1 Relative efficiency table for different values of p and p_1. *205*

Table 8.2 Relative efficiency table for different values of p_1. *206*

Table 8.3 Relative efficiency for $\alpha = 0.05$. *207*

Table 8.4 Relative efficiency for $n = 20$, $p = 10$, and $p_1 = 3$. *214*

Table 8.5 Relative efficiency for $n = 20$, $p = 10$, and $p_1 = 5$. *215*

Table 8.6 Relative efficiency for $n = 20$, $p = 10$, and $p_1 = 7$. *216*

Table 9.1 Correlation coefficients among the variables. *238*

Table 9.2 VIF values related to sea level rise at Key West, Florida data set. *238*

Table 9.3 Estimation of parameter using different methods ($p_1 = 2$, $p_2 = 7$, $\widehat{\phi} = -0.34$, and $k = 0.10$). *239*

Table 9.4 Estimation of parameter using LSE ($p_1 = 2$, $p_2 = 7$, $\widehat{\phi} = -0.34$, and $k = 0.10$). *240*

Table 9.5 Relative efficiency of the proposed estimators ($p_1 = 2$, $p_2 = 7$, $\widehat{\phi} = -0.34$, and $k = 0.10$). *242*

Table 9.6 The relative efficiency of the proposed estimators ($p_1 = 6$, $p_2 = 3$, $\widehat{\phi} = -0.34$, and $k = 0.10$). *243*

Table 9.7 The relative efficiency of the proposed estimators ($p_1 = 2$, $p_2 = 7$, $\phi = 0.35$, and $k = 0.10$). *244*

Table 9.8 The relative efficiency of the proposed estimators ($p_1 = 2$, $p_2 = 7$, $\phi = 0.55$, and $k = 0.10$). *244*

Table 9.9 The relative efficiency of the proposed estimators ($p_1 = 2$, $p_2 = 7$, $\phi = 0.75$, and $k = 0.10$). *246*

Table 9.10 The relative efficiency of the proposed estimators ($p_1 = 2$, $p_2 = 7$, $\phi = 0.85$, and $k = 0.10$). *248*

Table 9.11 The relative efficiency of the proposed estimators ($p_1 = 2$, $p_2 = 7$, $\phi = -0.85$, and $k = 0.10$). *249*

Table 10.1 RWRE for the estimators for $p_1 = 3$ and $p_2 = 7$. *269*

Table 10.2 RWRE for the estimators for $p_1 = 7$ and $p_2 = 13$. *270*

Table 10.3 RWRE of the R-estimators for $p = 10$ and different Δ^2-values for varying p_1. *271*

Table 10.4 RWRE of the R-estimators for $p = 15$ and different Δ^2-values for varying p_1. *272*

Table 10.5 RWRE of the R-estimators for $p = 20$ and different Δ^2-values for varying p_1. *273*

Table 10.6 RWRE of the R-estimators for $p = 30$ and different Δ^2-values for varying p_1. *274*

Table 10.7 RWRE values of estimators for $p_1 = 5$ and different values of p_2 and Δ^2. *275*

Table 10.8 RWRE values of estimators for $p_1 = 7$ and different values of p_2 and Δ^2. *276*

Table 10.9 RWRE values of estimators for $p_2 = 3$ and different values of p_1 and Δ^2. *277*

Table 10.10 RWRE values of estimators for $p_2 = 7$ and different values of p_1 and Δ^2. *278*

Table 11.1 Model fit characteristics for the proposed high-dimensional estimators: riboflavin data example. *291*

Table 11.2 REff values of the proposed high-dimensional estimators relative to high-dimensional RRE. *295*

Table 12.1 Test data input, output, and predicted values from a binary classification model. *317*

Table 12.2 Interpretation of test set results. *318*

Table 12.3 Results for L_2-penalty (ridge) using LR. *321*

Table 12.4 Results for L_1-penalty (LASSO) using LR. *321*

Table 12.5 Results for L_2-penalty (ridge) using two-layer NN. *321*

Table 12.6 Results for L_2-penalty (ridge) using three-layer NN. *322*

Preface

Regression analysis is the most useful statistical technique for analyzing multifaceted data in numerous fields of science, engineering, and social sciences. The estimation of regression parameters is a major concern for researchers and practitioners alike. It is well known that the least-squares estimators (LSEs) are popular for linear models because they are unbiased with minimum variance characteristics. But data analysts point out some deficiencies of the LSE with respect to prediction accuracy and interpretation. Further, the LSE may not exist if the design matrix is singular. Hoerl and Kennard (1970) introduced "ridge regression," which opened the door for "penalty estimators" based on the Tikhonov (1963) regularization. This methodology is the minimization of the least squares subject to an L_2 penalty. This methodology now impacts the development of data analysis for low- and high-dimensional cases, as well as applications of neural networks and big data analytics. However, this procedure does not produce a sparse solution. Toward this end, Tibshirani (1996) proposed the least absolute shrinkage and selection operator (LASSO) to overcome the deficiencies of LSE such as prediction and interpretation of the reduced model. LASSO is applicable in high- and low-dimensional cases as well as in big data analysis. LASSO simultaneously estimates and selects the parameters of a given model. This methodology minimizes the least squares criteria subject to an L_1 penalty, retaining the good properties of "subset selection" and "ridge regression."

There are many other shrinkage estimators in the literature such as the preliminary test and Stein-type estimators originally developed by Bancroft (1944), and Stein (1956), and James and Stein (1961), respectively. They do not select coefficients but only shrink them toward a predecided target value. There is extensive literature on a parametric approach to preliminary test and Stein-type estimators. The topic has been expanded toward robust rank-based, M-based, and quantile-based preliminary test and Stein-type estimation of regression coefficients by Saleh and Sen (1978–1985) and Sen and Saleh (1979, 1985, 1987). There is extensive literature focused only on this topic, and most recently documented by Saleh (2006). Due to the immense impact of Stein's approach on point estimation, scores of technical papers appeared in various areas of application.

The objective of this book is to provide a clear and balanced introduction of the theory of ridge regression, LASSO, preliminary test, and Stein-type estimators for graduate students and research-oriented statisticians, postdoctoral, and researchers. We start with the simplest models like the location model, simple linear model, and analysis of variance (ANOVA). Then we introduce the seemingly unrelated simple linear models. Next, we consider multiple regression, logistic regression, robust ridge regression, and high dimensional models. And, finally, as applications, we consider neural networks and big data to demonstrate the importance of ridge and logistic regression in these applications.

This book has 12 chapters, according to the given description of materials covered. Chapter 1 presents an introduction to ridge regression and different aspects of it, stressing the multicollinearity problem and its application to high-dimensional problems. Chapter 2 considers the simple linear model and location model, and provides theoretical developments of it. Chapters 3 and 4 deal with the ANOVA model and the seemingly unrelated simple linear models, respectively. Chapter 5 considers ridge regression and LASSO for multiple regression together with preliminary test and Stein-type estimators and a comparison thereof when the design matrix is nonorthogonal. Chapter 6 considers the ridge regression estimator and its relation with LASSO. Further, we study the properties of the preliminary test and Stein-type estimators with low dimension in detail. In Chapter 7, we cover the partially linear model and the properties of LASSO, ridge, preliminary test, and the Stein-type estimators. Chapter 8 contains the discussion of the logistic regression model and the related estimators of diverse kinds as described before in other chapters. Chapter 9 discusses the multiple regression model with autoregressive errors. In Chapter 10, we provide a comparative study of LASSO, ridge, preliminary test, and Stein-type estimators using rank-based theory. In Chapter 11, we discuss the estimation of parameters of a regression model with high dimensions. Finally, we conclude the book with Chapter 12 to illustrate recent applications of ridge, LASSO, and logistic regression to neural networks and big data analysis.

We appreciate the immense work done by Dr Mina Norouzirad with respect to the expert typing, editing, numerical computations, and technical management. Without her help, this book could not have been completed. Furthermore, Chapters 5 and 10 are the results of her joint work with Professor Saleh while she visited Carleton University, Ottawa during 2017. The authors thank Professor Resve Saleh for his assistance in preparing Chapter 12, along with reviewing the chapters and editing for English. We express our sincere thanks to Professor Mahdi Roozbeh (Semnan University, Iran) and Professor