

Theory of Ridge Regression Estimation with Applications

A guide to the systematic analytical results for ridge, LASSO, preliminary test, and Stein-type estimators with applications

Theory of Ridge Regression Estimation with Applications offers a comprehensive guide to the theory and methods of estimation. Ridge regression and LASSO are at the center of all penalty estimators in a range of standard models that are used in many applied statistical analyses. Written by noted experts in the field, the book contains a thorough introduction to penalty and shrinkage estimation and explores the role that ridge, LASSO, and logistic regression play in the computer intensive area of neural network and big data analysis.

Designed to be accessible, the book presents detailed coverage of the basic terminology related to various models such as the location and simple linear models, normal and rank theory-based ridge, LASSO, preliminary test and Stein-type estimators. The authors also include problem sets to enhance learning. This book is a volume in the Wiley Series in Probability and Statistics series that provides essential and invaluable reading for all statisticians. This important resource:

- Offers theoretical coverage and computer-intensive applications of the procedures presented
- Contains solutions and alternate methods for prediction accuracy and selecting model procedures
- Presents the first book to focus on ridge regression and unifies past research with current methodology

Written for graduate students, practitioners, and researchers in various fields of science, *Theory of Ridge Regression Estimation with Applications* is an authoritative guide to the theory and methodology of statistical estimation.

A. K. Md. EHSANES SALEH, PhD, is a Professor Emeritus and Distinguished Research Professor in the school of Mathematics and Statistics, Carleton University, Ottawa, Canada.


MOHAMMAD ARASHI, PhD, is an Associate Professor at Shahrood University of Technology, Iran and Extraordinary Professor and C2 rated researcher at University of Pretoria, Pretoria, South Africa.

B. M. GOLAM KIBRIA, PhD, is a Professor in the Department of Mathematics and Statistics at Florida International University, Miami, FL.

Cover Design: Wiley
Cover Images: © hrui/Shutterstock,
Courtesy of B. M. Golam Kibria

www.wiley.com

WILEY

 Also available
as an e-book

ISBN 978-1-118-64461-4



9 781118 644614

List of Figures	<i>xvii</i>
List of Tables	<i>xxi</i>
Preface	<i>xxvii</i>
Abbreviations and Acronyms	<i>xxxix</i>
List of Symbols	<i>xxxiii</i>

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 $\hat{\theta}_n^{RR}(k_{opt})$ and $\hat{\theta}_n^{PT}(\alpha)$	56
3.4.5.2	Comparison Between $\hat{\theta}_n^{RR}(k_{opt})$ and $\hat{\theta}_n^S$	56
3.4.5.3	Comparison of $\hat{\theta}_n^{RR}(k_{opt})$ with $\hat{\theta}_n^{S+}$	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 $\hat{\theta}_n^{RR}(k_{opt})$ and $\hat{\theta}_n^{PT}$	89
4.3.5.2	Comparison Between $\hat{\theta}_n^{RR}(k_{opt})$ and $\hat{\theta}_n^S$	89
4.3.5.3	Comparison of $\hat{\theta}_n^{RR}(k_{opt})$ with $\hat{\theta}_n^{S+}$	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 β	94
4.4.2	Efficiency for θ	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 $\hat{\theta}_n^{RR}(k_{opt})$ and $\hat{\theta}_n^{PT}(\alpha)$	123
5.4.5.2	Comparison Between $\hat{\theta}_n^{RR}(k_{opt})$ and $\hat{\theta}_n^S$	124
5.4.5.3	Comparison of $\hat{\theta}_n^{RR}(k_{opt})$ with $\hat{\theta}_n^{S+}$	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 $\hat{\beta}_n^{GRR}(k_{opt})$ and $\hat{\beta}_n^{G(P.T)}$	233
9.2.5.2	Comparison Between $\hat{\beta}_n^{GRR}(k_{opt})$ and $\hat{\beta}_n^{G(S)}$	234
9.2.5.3	Comparison of $\hat{\beta}_n^{GRR}(k_{opt})$ with $\hat{\beta}_n^{G(S+)}$	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
--------------	------------