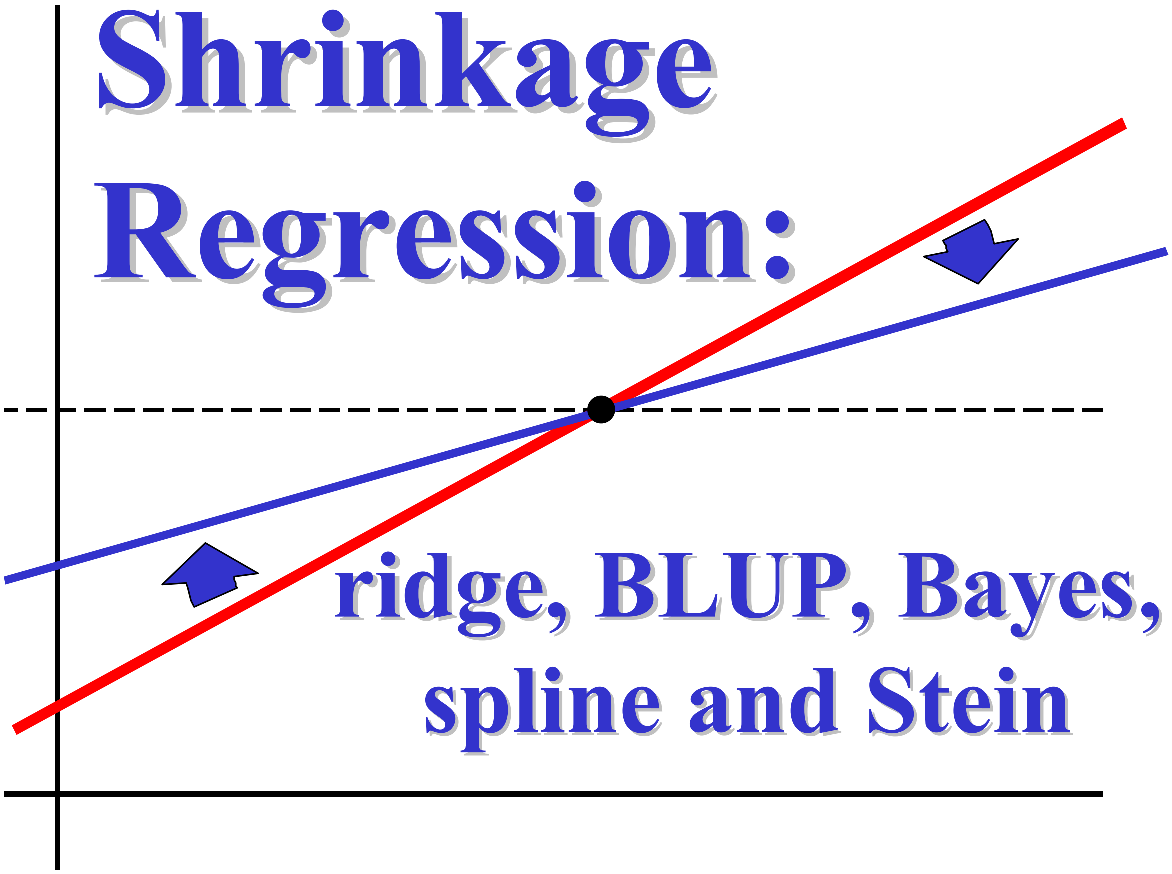
Shrinkage Regression:

ridge, BLUP, Bayes, spline and Stein



Chapter 00: Preface and Table of Contents

Bob Obenchain, Ph.D.
softR$_X$ freeware
13212 Griffin Run
Carmel, Indiana 46033-8835

To the great ladies who shaped and enriched my life...

Lynne, Tiffany, Anne, Lottie and mistress mathematics.

Preface

This book introduces, motivates, and explores a variety of methodologies for “shrinking" the regression coefficients that result when fitting models to possibly ill-conditioned and/or imprecise (errors-in-variables) data. The four main, broad categories of shrinkage methodology we consider here are ridge, BLUP, Bayes and Stein estimation. In ridge estimation, the underlying model usually views the unknown, true regression coefficients as being **fixed** constants. And the corresponding objective for shrinkage is to exploit variance-bias trade-offs to reduce mean-squared-error in estimation. In this same context, Stein methods are highly specialized forms of **uniform** coefficient shrinkage that focus interest on admissibility and/or minimax properties relative to specific, **univariate** (scalar valued) measures of risk. In BLUP and Bayes estimation, true regression coefficients are viewed as being unknown realizations of **random variables**, but estimation procedures suggested by these approaches again correspond to various forms of coefficient shrinkage.

Part One: Shrinkage Regression Theory and Methodology

We place special emphasis in Part One of our exposition upon normal-distribution-theory, maximum likelihood formulations that unify and contrast the ridge, BLUP, and empirical Bayes approaches to shrinkage. But we also explore a spectrum of alternative motivations for classical (fixed coefficient), random coefficient, and Bayes (added information) methods. We employ a uniform collection of terminology and notation throughout, and we exploit this formalism to give insights into interrelationships among diverse methods.

When I am voicing my own personal opinions about the relative strengths and weaknesses in a methodology, I will attempt to consistently use first person, singular pronouns (I, my).

People who know me well say that they “can almost hear me talking" when they read what I write. I suspect this is partly due to my use of somewhat non-standard typography. For example, I tend to place words in **bold face** or ALL CAPS for emphasis. And I tend to place quotation marks (“") around words or phrases that I am using in a somewhat ambiguous way...as in an analogy that is much less than complete. If I have trouble choosing between a pair of words in a particular phrase, I tend to write down both words with only a slash (/) between them. I know that referees and editors have found my typography disconcerting...but, then, I don't usually give the above “explanation" when I submit a paper for publication! Anyway, if you don't find my usage of **bolds**, CAPS, “", and /s helpful, ...please try to ignore them.

Part Two: Shrinkage Regression Applications and Implementations

In the sense that practical applications of shrinkage regression methodology **can** be highly graphical, then they **should** be highly graphical ...RIGHT? (Let's hope we can all agree on at least THAT!) While the technical materials of Part One provide, hopefully, a firm foundation for shrinkage theory and methodology, fundamental questions about general regression strategies and tactics are not addressed in Part One. Rather, questions like “Which displays should I examine?" and “How should I interpret them?" and overview materials on the **psychology of graphical perception** are primary topics for Part Two of our exposition. For example, our discussion of shrinkage TRACE displays in Chapter 11 discusses numerous advantages associated with using Multicollinearity Allowance (MCAL) scaling along the horizontal (shrinkage extent) axis on ridge TRACE displays.

Because shrinkage regression techniques tend to be computationally and graphically intensive, effective study/application of the ideas outlined in this book **requires** access to state-of-the-art computer software and hardware. I have developed prototype systems for IBM-compatible (MS-DOS and Windows) personal computers for this express purpose, and I market them as freeware to maximize their availability to potential practitioners of shrinkage regression. But I also illustrate (i) computational procedures for LISP-STAT, GAUSS, SAS/IML and S as well as (ii) commercial software systems such as SAS proc MIXED, BMDP (especially 4R and 5V), and GLMM.

I hope users will find that my personal computer software systems provide an interface that is sufficiently intuitive and self-explanatory that the full depth of detailed understanding provided by this book is not a prerequisite for effective shrinkage regression applications. But, by pulling together diverse materials - ranging through introductory remarks, common misconceptions, historical commentaries, simulation results, and theoretical fine points - we provide here not only an unquestionably strong foundation but also a full set of practically useful road-maps for the theory and application of shrinkage regression.

Shrinkage Regression: ridge, BLUP, Bayes, spline & Stein

Table of Contents

CHAPTER

1. INTRODUCTION

This introductory chapter provides an short overview of our topic. We first consider the original sense in which “regression" implies “shrinkage" by tracing standard least-squares fitting terminology back to the work of Francis Galton in the late 1800's. Then we introduce our PRIMARY THEME, the **multivariate analysis point-of-view on shrinkage regression**. Next we provide a very brief thumb-nail-sketch of how modern shrinkage methods might be applied to a simple numerical example. At the end of chapter one, we detail three key motivations for exploiting the **principal axes rotation** of regressor coordinates as a basic **canonical form** for possibly ill-conditioned regressor problems.

Part One: Shrinkage Regression Theory and Methodology

2. BASIC LINEAR MODEL CONCEPTS

This chapter reviews many aspects of the theory of **general linear models** in the possibly less-than-full-rank case.

3. SHRINKAGE REGRESSION FUNDAMENTALS

This chapter introduces generalized shrinkage (ridge regression) estimators and points out several special cases which played major roles in the early history of shrinkage regression.

4. THE RISK OF SHRINKAGE

How much shrinkage is "best"? We start out by showing that this question is difficult to answer - even in theory when we pretend that true values of regression parameters are **known!** Then we introduce the concepts of "optimal", "good", and "ultimate" choices for shrinkage factors. We also develop an exact parallel between the fixed coefficient and the random coefficient formulations of minimum mean-squared-error shrinkage of a single parameter.

5. NORMAL-THEORY MAXIMUM LIKELIHOOD: BLUEs and BLUPs

This chapter includes a review of the basic theory of BLUEs and BLUPs. For shrinkage in fixed coefficient models, we develop not only general methods for identifying maximum likelihood estimators within arbitrary shrinkage families but also **closed form** expressions for optimal shrinkage estimators within the 2-parameter family. For shrinkage in random coefficient models, we discuss why maximum likelihood estimates are rarely "linear" or "unbiased". And we review the maximum likelihood methods of Golub, Heath, and Wahba(1979) and of Shumway(1982) for the special case of completely random models with a single variance component.

6. RISK (MEAN SQUARED ERROR) ESTIMATION and SIMULATION

In this chapter, we display normal-theory maximum likelihood estimates of scaled (or relative) mean-squared-error (MSE) risk. In addition to estimates of risk in individual coefficients, we consider estimates for arbitrary linear combinations. And we explore corrections for bias and "range." Then we examine Monte-Carlo simulation results showing that reduced MSE risk is easier to actually achieve when coefficients are random than when they are fixed.

7. RANDOM COEFFICIENT FORMULATIONS

Here we review iterative methods (Newton-Raphson, Fisher Scoring, EM, REML) for solving Henderson's “mixed model equations." And we detail specific applications in which random-coefficient models are more realistic (and provide “better" estimates) than fixed coefficient models.

8. BAYESIAN FORMULATIONS

We review the normal-theory, hierarchical models of Lindley and Smith(1972) as well as the empirical Bayes formulation of Efron and Morris(1977). We also discuss exactly why and how the Bayesian variance of a shrinkage estimate exceeds its classical variance. We derive both Theil's measure of the proportion of posterior precision due to sample information and also Shannon's measure of information gain. Finally, we comment on proposals for nonconjugate Bayes formulations.

9. COMPUTATIONALLY INTENSE METHODS

We start with two sections on “errors-in-variables" models under which least-squares estimates are biased and inconsistent; we first show how random data un-rounding can suggest shrinkage to improve stability of estimates, but we also explore maximum likelihood methods for multivariate normal “structural" models that suggest expansions. Next, we review iterative methods which, although traditionally applied to regressor variable subsetting or robust fitting, can also be used in shrinkage regression estimation. Specifically, we discuss methods of cross-validation for choice and assessment of shrinkage and methods that down-weight certain types of otherwise “influential" observations.

10. TOPICS of HISTORICAL INTEREST, HEURISTIC ARGUMENTS, and COMMON MISCONCEPTIONS

We start with a review and critique of the many contributions of Art Hoerl and Robert Kennard to the theory and practice of **ridge regression**. But a wide spectrum of alternative approaches are also described.

Part Two: Shrinkage Regression Applications and Implementations

11. TRACE DISPLAYS: THE PSYCHOLOGY OF PERCEPTION

We start with a discussion of the many advantages of adopting the Multicollinearity Allowance (MCAL) choice for scaling along the horizontal (shrinkage extent) axis of a TRACE display.

Then we explain not only how to “read" the 5 major types of ridge TRACE displays but also how to focus on specific details using **Path Projection** onto two-dimensional linear subspaces.

> “The greatest thing a human soul ever does in this world is to SEE something, and tell what it SAW in a plain way. ... To see clearly is poetry, prophecy and religion - all in one."
>
> John Ruskin
> Modern Painters, 1888

12. USAGE of RXridge

RXridge performs normal-theory maximum likelihood computations for 2-parameter generalized shrinkage regression. It features user-friendly windows, menus, graphics, and visual review of ridge files that have been written to disk. RXridge handles **exact** (numerical) singularities as well as nearly **multicollinear** (statistically ill-conditioned) examples.

13. USAGE of RXtraces

RXtraces creates **interactive** CGA graphics displays of all 5 types of ridge TRACES ...including estimates of (i) regression coefficients, (ii) scaled mean squared errors, (iii) excess (ordinary-least-squares minus ridge) eigenvalues, (iv) inferior direction cosines, and (v) the shrinkage factor pattern. RXtraces is ideal for ridge **training** applications (workshops and/or self-paced learning) in which students simply review RXridge or RelaxR outputs across a spectrum of previously-computed numerical examples.

14. USAGE of PathProj

PathProj creates **interactive** CGA graphics displays of the projection of ridge shrinkage path statistics (regression coefficients, inferior direction, regressor coordinates, etc.) onto any 2-dimensional linear subspace (specified via orthogonal direction cosine vectors.) Literally **WATCH** as an inferior direction first “appears" and then changes its orientation as you step along a ridge shrinkage path.

15. USAGE of RXrisk, RXshrink, and RXmsesim

16. CASE STUDY 1: PORTLAND CEMENT (MIXTURE) DATA

17. CASE STUDY 2: GASOLINE MILEAGE DATA

18. CASE STUDY 3: PHARMACEUTICAL SHELF LIFE ESTIMATION