Title: Variations on Ridge Traces in Regression
Running Title: Ridge Traces
Authors: Donald R. Jensen, (email: djensen@vt.edu),
Donald E. Ramirez (email: der@virginia.edu)
Authors' Affiliation: Department of Mathematics, University of Virginia
Corresponding Author: Donald E. Ramirez, Department of Mathematics,
University of Virginia, Charlottesville, VA 22904–4137 email: der@virginia.edu

VARIATIONS ON RIDGE TRACES IN REGRESSION

D. R. JENSEN AND D. E. RAMIREZ

Abstract. Ridge regression, perturbing the design moment matrix via a parameter k, persists in the study of ill-conditioned systems. Ridge traces, exhibiting solutions as functions of k, are intended to reflect stability as k evolves, in contrast to transient instabilities in ordinary least squares. This study examines derivative traces as analytic tools regarding stability, and develops rational representations for them. Two further gauges of stability are derivatives of variances of the ridge solutions, and the variances of the derivative traces, both tending to zero as k increases. In contrast to ridge traces and their derivatives, neither of the latter depends on observed responses, and both support deterministic assessments.

1. Introduction

1.1. **Overview.** In a full–rank model $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ having zero–mean, uncorrelated, and homoscedastic errors, the Ordinary Least Squares (*OLS*) estimators $\widehat{\boldsymbol{\beta}}_L$ solve the p equations $\{\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{Y}\}$. These solutions are unbiased with dispersion matrix $\mathrm{V}(\widehat{\boldsymbol{\beta}}_L) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Near–dependency among the columns of $\boldsymbol{X}$, as *ill–conditioning,* engenders "crucial elements of $\boldsymbol{X}'\boldsymbol{X}$ to be large and unstable," "creating inflated variances," the *OLS* solutions often inflated in size and of questionable signs, and "very sensitive to small changes in $\boldsymbol{X}$;" (Belsley, 1986). Among continuing palliatives are the *ridge system* $\{(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p)\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{Y}; k \geq 0\}$ (Hoerl and Kennard, 1970a) (hereafter H&K, 1970a) with solutions $\{\widehat{\boldsymbol{\beta}}_R(k); k \geq 0\}$. The *OLS* solution at $k = 0$ is known to be unstable: "A slight movement away from this point can give completely different estimates of the coefficients" (H&K, 1970b). Letting $\widehat{\boldsymbol{\beta}}_R(k) = [\widehat{\beta}^1_R(k), \ldots, \widehat{\beta}^p_R(k)]'$ and taking the *ridge trace* as graphs of $\{k \to \widehat{\beta}^i_R(k);\ 1 \leq i \leq p\}$, this provides "a two–dimensional graphical procedure for portraying the complex relationships in multifactor data" (H&K, 1970b). Of particular concern is the *stability* of solutions as k evolves, and the use of evidential stability in choosing k. As a staple in the analyst's toolbox, it thus is germane to reexamine the use of ridge traces for gauging stability, and the role of diverse other choices for k as advocated in the literature.

In particular, since $\{\widehat{\beta}^i_R(k); 1 \leq i \leq p\}$ not only are continuous but are differentiable in k, their derivatives hold promise in regard to stability and other such features as local maxima and minima. Gibbons & McDonald (1984) show that ridge estimators may be expressed as rational functions of the parameter k, and critical properties emerge on examining polynomials in those representations. Specifically, sign changes, crossings, and rates–of–change of ridge coefficients, as functions of k, emerge on examining derivatives and identifying zeros of polynomials, where locations of those changes coincide with the positive roots. "These

1991 *Mathematics Subject Classification.* Primary: 62J07; Secondary: 62J20.

Key words and phrases. Ill–conditioned models, ridge regression, stability of ridge traces, derivative traces, deterministic traces.

Research of the first author supported in part by the Department of Mathematics, University of Virginia.

characterizations, in the aggregate, serve to 'quantify' the rationale (such as 'stability' and 'reasonable' coefficient signs) for selecting a specific ridge estimator in a specific application;" see Zhang & McDonald (2005) and, for a recent survey, McDonald (2009). Parallel developments are given subsequently for derivative traces.

Moreover, ridge traces and their derivatives, depending on $\boldsymbol{Y}$, are stochastic and thus subject to the vagaries of experimental variation. It remains to ask whether evidence towards stability might reside exclusively in the matrix $\boldsymbol{X}$, signaling stability "in distribution" of a ridge trace or its derivative. This question is answered in the affirmative, giving a deterministic assessment of the attenuation of those attributes as k evolves. Connections are drawn to other aspects of ridge regression from the literature. Unfortunately, ridge traces on occasion have been misconstrued, as noted in the following.

1.2. **Case Study: A First Look.** The Hospital Manpower Data (Myers, 1990), detailed subsequently, exhibits $p = 5$ highly ill–conditioned regressors. In support of an algorithm for choosing k, Table 8.12 (Myers, 1990) gives sections of the ridge traces in the original coordinates, extracted here in part as the first three rows of Table 1. [**Table 1 here.**] Based on these, Myers asserts that the ridge coefficients have stabilized at $k = 0.0004$, taking this to be a viable choice for k based on ridge traces.

Unfortunately, these conclusions are flawed: Nearness of ridge coefficients for nearby k in the first three rows of Table 1 reflects continuity of the ridge traces, not stability; even better agreement would accompany smaller increments in k.

On the other hand, the divergences $\{[\widehat{\beta}^i_R(k_u) - \widehat{\beta}^i_R(k_v)]/(k_u - k_v); 1 \le i \le 5\}$, as changes in $\widehat{\beta}^i_R(k)$ per unit change in k, accurately portray the local variation in $\{\widehat{\beta}^i_R(k); 1 \le i \le 5\}$ as k ranges from k_v to k_u. Values for these divergences are listed in the midsection of Table 1, categorically rejecting any prospects that stability might have been achieved across these values for k. These facts in part motivate ensuing developments in which instantaneous rates of change, *i.e.* derivatives as limits of divergences, are germane in assessing stability of ridge traces as k evolves. Looking ahead, these derivatives are listed in the bottom portion of Table 1.

2. Preliminaries

2.1. **Notation.** Designate $\mathbb{R}^n$ and $\mathbb{R}^n_+$ as Euclidean n–space and its positive orthant. Matrices and vectors are set in bold type; the transpose, inverse, and trace of $\boldsymbol{A}$ are $\boldsymbol{A}'$, $\boldsymbol{A}^{-1}$, and $\mathrm{tr}(\boldsymbol{A})$. Special arrays include the identity $\boldsymbol{I}_n$, the unit vector $\boldsymbol{1}_n = [1, 1, \ldots, 1]' \in \mathbb{R}^n$, the diagonal matrix $\boldsymbol{D}(a_i) = \mathrm{Diag}(a_1, \ldots, a_p)$, and $\boldsymbol{B}_n = (\boldsymbol{I}_n - n^{-1}\boldsymbol{1}_n\boldsymbol{1}_n')$. Let $\boldsymbol{Z}(n \times p)$ have rank $p < n$; its *singular decomposition* is $\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{Q}'$, with $\boldsymbol{D}_\xi = \mathrm{Diag}(\xi_1, \ldots, \xi_p)$ as its ordered *singular values* $\{\xi_1 \ge \xi_2 \ge \ldots \ge \xi_p > 0\}$, and with the columns of $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_p]$ and of $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_p]$ comprising the *left*– and *right–singular vectors* of $\boldsymbol{Z}$, such that $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_p$ and $\boldsymbol{Q}$ is orthogonal.

2.2. **A Canonical Form.** Successive transformations yield a basic canonical form; inverse mappings then recover the original coordinates. Beginning with the standard model $\{\boldsymbol{Y} = \beta_0\boldsymbol{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ with intercept, take $\boldsymbol{Y}_0 = \boldsymbol{B}_n\boldsymbol{Y}$ with $\boldsymbol{B}_n = (\boldsymbol{I}_n - n^{-1}\boldsymbol{1}_n\boldsymbol{1}_n')$; observe that $\boldsymbol{B}_n\boldsymbol{1}_n = \boldsymbol{0}$ and $\boldsymbol{B}_n^2 = \boldsymbol{B}_n$; and let $\boldsymbol{S}_c = \mathrm{Diag}(S_1, \ldots, S_p)$ comprise reciprocals of square roots of the diagonals of $\boldsymbol{X}'\boldsymbol{B}_n\boldsymbol{X}$. This serves to center and scale elements of $\boldsymbol{X}$, so that $\boldsymbol{Z}'\boldsymbol{Z} = \boldsymbol{S}_c\boldsymbol{X}'\boldsymbol{B}_n\boldsymbol{X}\boldsymbol{S}_c$ is in the conventional "correlation form." These changes support the transitions

$$\{\boldsymbol{Y} = \beta_0\boldsymbol{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\} \to \{\boldsymbol{B}_n\boldsymbol{Y} = \beta_0\boldsymbol{B}_n\boldsymbol{1}_n + \boldsymbol{B}_n\boldsymbol{X}\boldsymbol{S}_c\boldsymbol{S}_c^{-1}\boldsymbol{\beta} + \boldsymbol{\eta}\} \tag{1}$$

$$\to \{\boldsymbol{Y}_0 = \boldsymbol{Z}\boldsymbol{\omega} + \boldsymbol{\eta}\} \to \{\boldsymbol{Y}_0 = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{\theta} + \boldsymbol{\eta}\}. \tag{2}$$

The singular decomposition of $\boldsymbol{Z} = \boldsymbol{B}_n\boldsymbol{X}\boldsymbol{S}_c$ is $\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{Q}'$; $\boldsymbol{\omega} = \boldsymbol{S}_c^{-1}\boldsymbol{\beta}$; $\boldsymbol{\eta} = \boldsymbol{B}_n\boldsymbol{\epsilon}$; and $\boldsymbol{\theta} = \boldsymbol{Q}'\boldsymbol{\omega} = \boldsymbol{Q}'\boldsymbol{S}_c^{-1}\boldsymbol{\beta}$ is linear in the original $\boldsymbol{\beta}$. From this our ultimate canonical form is

$$\{\boldsymbol{Y}_0 = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{\theta} + \boldsymbol{\eta}\} \rightarrow \{\boldsymbol{P}'\boldsymbol{Y}_0 = \boldsymbol{P}'\boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{\theta} + \boldsymbol{P}'\boldsymbol{\eta}\} \tag{3}$$

$$\rightarrow \{\boldsymbol{W} = \boldsymbol{D}_\xi\boldsymbol{\theta} + \boldsymbol{\tau}\}, \tag{4}$$

where $\boldsymbol{Y}_0 \rightarrow \boldsymbol{W} = \boldsymbol{P}'\boldsymbol{Y}_0$, $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_p$, and $\boldsymbol{\tau} = \boldsymbol{P}'\boldsymbol{\eta}$.

Ridge regression traditionally is carried out as $\{(\boldsymbol{Z}'\boldsymbol{Z} + k\boldsymbol{I}_p)\boldsymbol{\omega} = \boldsymbol{Z}'\boldsymbol{Y}_0\}$, or equivalently, $\{(\boldsymbol{D}_\xi^2 + k\boldsymbol{I}_p)\boldsymbol{\theta} = \boldsymbol{D}_\xi\boldsymbol{W}\}$, giving $\widehat{\boldsymbol{\theta}}_R(k) = \boldsymbol{D}(\xi_i/(\xi_i^2 + k))\boldsymbol{W}$ and $\widehat{\boldsymbol{\omega}}_R(k) = \boldsymbol{Q}\widehat{\boldsymbol{\theta}}_R(k)$ as the ridge solutions in correlation form, where

$$\boldsymbol{D}(\xi_i/(\xi_i^2 + k)) = \mathrm{Diag}(\xi_1/(\xi_1^2 + k), \ldots, \xi_p/(\xi_p^2 + k)). \tag{5}$$

These in turn map back to the original coordinates through $\widehat{\boldsymbol{\beta}}_R(k) = \boldsymbol{A}\widehat{\boldsymbol{\theta}}_R(k)$, with $\boldsymbol{A} = \boldsymbol{S}_c\boldsymbol{Q}$, *i.e.*, $\{\widehat{\beta}_R^i(k) = \boldsymbol{a}_i'\widehat{\boldsymbol{\theta}}_R(k);\ 1 \le i \le p\}$ with $\boldsymbol{A}' = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p]$. For an orthogonal system with $\boldsymbol{X}'\boldsymbol{X}$ diagonal, the orthogonal matrices are $\boldsymbol{P} = \boldsymbol{I}_n$ and $\boldsymbol{Q} = \boldsymbol{I}_p$, with the pull-back matrix $\boldsymbol{A} = \boldsymbol{I}_p$.

For subsequent reference observe that

$$\mathrm{V}(\widehat{\boldsymbol{\theta}}_R(k)) = \sigma^2\boldsymbol{D}(\xi_i^2/(\xi_i^2 + k)^2) \Longrightarrow \mathrm{V}(\widehat{\boldsymbol{\beta}}_R(k)) = \sigma^2\boldsymbol{A}\boldsymbol{D}(\xi_i^2/(\xi_i^2 + k)^2)\boldsymbol{A}'. \tag{6}$$

Moreover, since $\{d(\hat{\theta}_R^i(k))/dk = -(\xi_i/(\xi_i^2 + k)^2)W_i;\ 1 \le i \le p\}$, the canonical traces $\{k \rightarrow \hat{\theta}_R^i(k);\ 1 \le i \le p\}$ are monotone in k. On the original scale, however, $\{k \rightarrow \widehat{\beta}_R^i(k) = \boldsymbol{a}_i'\widehat{\boldsymbol{\theta}}_R(k)\}$ need not be monotone, as seen in Section 4.

3. The Principal Findings

3.1. **Ridge Traces.** H&K(1970a,b) promulgate ridge traces as fundamental in identifying values of k yielding stable solutions. These are stochastic displays; evidence of stability thus is obscured in part by random disturbances in $\boldsymbol{Y}$. On the other hand, evidence of tendencies to stabilize *in distribution* is provided by their evolving point–wise variances: Diminishing variances would point to increasing stability of ridge traces about their evolving means. This concept may be quantified through Chebychev's inequalities, for example.

Accordingly, were "variance traces" $\{k \rightarrow \mathrm{Var}(\widehat{\beta}_R^i(k));\ 1 \le i \le p\}$ to be monitored as adjuncts to ridge traces, this would serve to exhibit stabilizing trends in their variances, hence their concentrations in probability. Moreover, the latter entities are deterministic, depending only on $\boldsymbol{X}$, free of random disturbances in $\boldsymbol{Y}$. On the other hand, since instabilities often may be discerned more readily through derivatives, we further seek to examine derivative variances of ridge traces through $\{k \rightarrow d(\mathrm{Var}(\widehat{\beta}_R^i(k))/dk;\ 1 \le i \le p\}$. Some properties may be listed as follows.

Theorem 1. *Consider the ridge traces* $\{\widehat{\beta}_R^i(k);\ k \ge 0\}$, *their variances* $\{\mathrm{Var}(\widehat{\beta}_R^i(k));\ k \ge 0\}$, *and the derivatives* $\{d\mathrm{Var}(\widehat{\beta}_R^i(k))/dk;\ k \ge 0\}$, *for* $1 \le i \le p$. *Then*

(i) Variances are given by $\{\mathrm{Var}(\widehat{\beta}_R^i(k)) = \sigma^2\boldsymbol{a}_i'\boldsymbol{D}(\xi_i^2/(\xi_i^2 + k)^2)\boldsymbol{a}_i;\ 1 \le i \le p\}$, *each monotone decreasing with increasing* k.

(ii) Rates of change in $\{\mathrm{Var}(\widehat{\beta}_R^i(k));\ 1 \le i \le p\}$ *are given by*

$$\{d\mathrm{Var}(\widehat{\beta}_R^i(k))/dk = -2\sigma^2\boldsymbol{a}_i'\boldsymbol{D}(\xi_i^2/(\xi_i^2 + k)^3)\boldsymbol{a}_i < 0;\ 1 \le i \le p\}$$

independently of $\boldsymbol{Y}$.

(iii) The negative functions $\{d\mathrm{Var}(\widehat{\beta}_R^i(k))/dk; 1 \le i \le p\}$ *are monotone increasing as* $k \uparrow$ *for* $k \ge 0$, *their values progressing from large to small in magnitude.*

Proof. Conclusion (i) extracts diagonal elements from (6); differentiation yields conclusion (ii); and monotonicity in (i) follows from negative slopes and that $\{d^2\mathrm{Var}(\widehat{\beta}_R^i(k))/dk^2 = 6\sigma^2\boldsymbol{a}_i'\boldsymbol{D}(\xi_i^2/(\xi_i^2+k)^4)\boldsymbol{a}_i > 0;\ 1\le i\le p\}$ are positive as noted in H&K (1970a, p.60). Conclusion (iii) follows on checking signs of their first and second derivatives, to complete our proof. □

3.2. **Derivative Traces.** Pursuing the notion that instabilities are revealed through derivatives, we proceed to examine the derivative traces $\{k \to d\widehat{\beta}_R^i(k)/dk;\ 1\le i\le p\}$ and their properties, *in lieu of* ridge traces. Principal findings follow.

Theorem 2. *Consider the ridge trace* $\{k \to \widehat{\boldsymbol{\beta}}_R(k)\}$ *and its derivative as the vector gradient* $\frac{d\widehat{\boldsymbol{\beta}}_R(k)}{dk} = \left[\frac{d\widehat{\beta}_R^1(k)}{dk}, \frac{d\widehat{\beta}_R^2(k)}{dk}, \ldots, \frac{d\widehat{\beta}_R^p(k)}{dk}\right]'$. *Then the following properties hold.*

(i) $d\widehat{\boldsymbol{\beta}}_R(k)/dk = \boldsymbol{A}\, d\widehat{\boldsymbol{\theta}}_R(k)/dk = -\boldsymbol{A}\boldsymbol{D}(\xi_i/(\xi_i^2+k)^2)\boldsymbol{W}$.

(ii) For each $k \ge 0$, *the dispersion matrix is* $\mathrm{V}(d\widehat{\boldsymbol{\beta}}_R(k)/dk) = \sigma^2\boldsymbol{A}\boldsymbol{D}(\xi_i^2/(\xi_i^2+k)^4)\boldsymbol{A}'$ *independently of* $\boldsymbol{Y}$.

(iii) The ratio of variances of $\hat{\theta}_R^i(k)$ *to* $d\hat{\theta}_R^i(k)/dk$ *is* $\mathrm{Var}(\hat{\theta}_R^i(k))/\mathrm{Var}(d\hat{\theta}_R^i(k)/dk) = (\xi_i^2 + k)^2$, *for* $1\le i\le p$ *and each* $k > 0$.

Proof. From $\widehat{\boldsymbol{\beta}}_R(k) = \boldsymbol{A}\widehat{\boldsymbol{\theta}}_R(k)$ and $\{\widehat{\beta}_R^i(k) = a_{i1}\hat{\theta}_R^1(k) + \ldots + a_{ip}\hat{\theta}_R^p(k); 1 \le i \le p\}$, determine that $\{d\widehat{\beta}_R^i(k)/dk = a_{i1}\, d\hat{\theta}_R^1(k)/dk + \ldots + a_{ip}\, d\hat{\theta}_R^p(k)/dk; 1\le i\le p\}$. It follows that $d\widehat{\boldsymbol{\beta}}_R(k)/dk = \boldsymbol{A}\, d\widehat{\boldsymbol{\theta}}_R(k)/dk$, since $d\widehat{\boldsymbol{\theta}}_R(k)/dk = d\,\boldsymbol{D}(\xi_i/(\xi_i^2+k))\boldsymbol{W}/dk = -\boldsymbol{D}(\xi_i/(\xi_i^2+k)^2)\boldsymbol{W}$, to give conclusion (i). To see conclusion (ii), observe that $\mathrm{V}(\boldsymbol{\eta}) = \sigma^2\boldsymbol{B}_n$ at equation (2) since $\boldsymbol{B}_n$ is idempotent. Accordingly, $\mathrm{V}(\boldsymbol{W}) = \sigma^2\boldsymbol{P}'\boldsymbol{B}_n\boldsymbol{P} = \sigma^2\boldsymbol{P}'\boldsymbol{P} = \sigma^2\boldsymbol{I}_p$, so that $\mathrm{V}(d\,\widehat{\boldsymbol{\theta}}_R(k)/dk) = \mathrm{V}(d\,\boldsymbol{D}(\xi_i/(\xi_i^2+k))\boldsymbol{W}/dk) = \sigma^2\boldsymbol{D}(\xi_i^2/(\xi_i^2+k)^4)$ and $\mathrm{V}(d\widehat{\boldsymbol{\beta}}_R(k)/dk) = \sigma^2\boldsymbol{A}\boldsymbol{D}(\xi_i^2/(\xi_i^2+k)^4)\boldsymbol{A}'$. Conclusion (iii) follows from (6) and the proof for (ii), to complete our proof. □

Ridge traces, and now their derivatives, are subject to chance disturbances intrinsic to $\boldsymbol{Y}$ as noted and, from conclusion (iii), each may be more or less variable than the other, depending on the weights $\{\boldsymbol{a}_i; 1\le i\le p\}$ and signs of $\{[(\xi_i^2+k)-1]; 1\le i\le p\}$. We next seek versions of traces devoid of disturbances in $\boldsymbol{Y}$, depending only on $\boldsymbol{X}$. As noted, random traces may be thought to "stabilize in distribution" as their variances diminish. Accordingly, we focus next on stability of derivative traces in terms of their point–wise variability, namely, $\{\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk); 1\le i\le p\}$, and derivatives of the latter as $\{d[\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk)]/dk; 1\le i\le p\}$, in seeking evidence for evolving stochastic stability of the ridge solutions. Details may be collected as follows.

Theorem 3. *Consider the derivative traces* $\{k \to d\widehat{\beta}_R^i(k)/dk;\ 1\le i\le p\}$ *and their variances* $\{\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk);\ 1\le i\le p\}$. *Then*

(i) For each $k > 0$, $\{\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk) = \sigma^2\boldsymbol{a}_i'\boldsymbol{D}(\xi_i^2/(\xi_i^2+k)^4)\boldsymbol{a}_i;\ 1\le i\le p\}$. *Moreover, each decreases monotonically with increasing* $k > 0$ *for* $1\le i\le p$.

(ii) Rates of change in $\{\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk);\ 1\le i\le p\}$ *are given by* $\{d[\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk)]/dk = -4\sigma^2\boldsymbol{a}_i'\boldsymbol{D}(\xi_i^2/(\xi_i^2+k)^5)\boldsymbol{a}_i < 0;\ 1\le i\le p\}$ *independently of* $\boldsymbol{Y}$.

(iii) The negative functions $\{d[\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk)]/dk;\ 1\le i\le p\}$ *are monotone increasing as* $k\uparrow$ *for* $k \ge 0$, *their values progressing from large to small in magnitude.*

Proof. Conclusion (i) extracts $\{\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk);\ 1\le i\le p\}$ as diagonal elements from Theorem 2(ii); differentiation yields conclusion (ii); and monotonicity in (i) follows from negative slopes and that $\{d^2\text{Var}(d\widehat{\beta}_R^i(k)/dk)/dk^2 = 20\sigma^2\boldsymbol{a}_i'\boldsymbol{D}(\xi_i^2/(\xi_i^2+k)^6)\boldsymbol{a}_i > 0\}$ are positive. Conclusion (iii) follows on checking signs of their first and second derivatives as in Theorem 1, to complete our proof. □

The following deserves emphasis. Whereas the ridge and derivative traces are random, convergence of their distributions towards stability may be gauged *deterministically* through variance and derivative variance traces. Such features may be assessed beforehand, based on $\boldsymbol{X}$ alone, whereas ridge traces, and now their derivatives, must await the empirical outcome of $\boldsymbol{Y}$ at the conclusion of an experiment.

4. Case Studies

4.1. **The Setting.** The Hospital Manpower Data, as reported in Table 3.8 (Myers, 1990), consist of records at $n = 17$ U.S. Naval Hospitals, to include: Monthly man–hours (Y); Average daily patient load (X_1); Monthly X–ray exposures (X_2); Monthly occupied bed days (X_3); Eligible population in the area ÷ 1000 (X_4); and Average length of patients' stay in days (X_5). The applicable model is

$$\{Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon_i;\ 1 \le i \le 17\}. \tag{7}$$

Our computations utilize *Proc IML* of the SAS Programming System and the symbolic program *Maple.* The *OLS* estimates $\widehat{\boldsymbol{\beta}}_L = [\widehat{\beta}_L^1, \widehat{\beta}_L^2, \widehat{\beta}_L^3, \widehat{\beta}_L^4, \widehat{\beta}_L^5]'$ and their variances, in the original coordinates, are listed in Table 2. [**Table 2 here.**] Following convention (for example the *Ridge* option of *Proc Reg* in the SAS system), ridge regression proceeds on first centering and scaling, taking $\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{Z}'\boldsymbol{Z}$ in correlation form; solving the ridge equations; then mapping back onto the natural coordinates as in Section 2.2. The data are remarkably ill–conditioned: Singular values of $\boldsymbol{Z}$ are $\boldsymbol{D}_\xi = \text{Diag}(2.048687, 0.816997, 0.307625, 0.201771, 0.007347)$, and the condition number is $c_1(\boldsymbol{Z}'\boldsymbol{Z}) = 77{,}754.86$. Throughout 24.30E02 designates 24.30×10^2, for example, in scientific notation.

4.2. **Ridge Regression.** A striking diversity persists in choices for k as advocated in the literature, underscoring the problem as less than well–posed, often with profound differences among solutions. We invoke findings reported here as a common lens through which those choices may be viewed, with special reference to stability of solutions in k.

Choices in wide usage are identified in Table 3, [**Table 3 here**] together with their values for the Hospital Manpower Data. These encompass $DF_k = \text{tr}(\boldsymbol{H}_k) = \sum_{i=1}^p \xi_i^2/(\xi_i^2+k)$ with $\boldsymbol{H}_k = [\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z}+k\boldsymbol{I}_p)^{-1}\boldsymbol{Z}']$; $PRESS_k = \sum_{i=1}^n e_{(i,\lambda)}^2$ as the cross–validation statistic (Allen, 1974); $GCV_k = SS_{Res,k}/[n-(1+\text{tr}(\boldsymbol{H}_k))]^2$, a rotation–invariant analog called *Generalized Cross Validation* (Golub et al., 1979); $C_k = [(SS_{Res,k}/\widehat{\sigma}^2) - n + 2 + 2\text{tr}(\boldsymbol{H}_k)]$ to achieve a variance–bias trade–off (Mallows, 1973); and $HKB_k = \widehat{\sigma}^2/\widehat{\boldsymbol{\beta}}_L'\widehat{\boldsymbol{\beta}}_L$ as in (Hoerl et al., 1975) from simulation studies. Here $SS_{Res,k}$ and $\widehat{\sigma}^2$ respectively are the residual sum of squares using ridge and the *OLS* residual mean square; and $\{e_{(i,\lambda)}^2\}$ are the *PRESS* residuals from ridge regression. See (Myers, 1990), including numerical values for DF_k, C_k, and $PRESS_k$ as reported here. These values are marked with asteristics in the tables to follow. For further details on estimating k, see (Kibria, 2003) and (Muniz and Kibria, 2009), among others.

Marquardt (1970) notes that the variation inflation factors $VIF_i(k)$ for the ridge estimators are the diagonal elements of $(\boldsymbol{Z}'\boldsymbol{Z}+k\boldsymbol{I}_p)^{-1}(\boldsymbol{Z}'\boldsymbol{Z})(\boldsymbol{Z}'\boldsymbol{Z}+k\boldsymbol{I}_p)^{-1}$, and gives the rule of thumb that $\{VIF_i(k) < 10; 1 \le i \le p\}$ in choosing k. Values for $\max[VIF_i(k); 1 \le i \le p]$ are shown in Table 3 for k as listed there. We note that all $\{VIF_i(k) < 10; 1 \le i \le p\}$ for $k > 0.00706$. Moreover, Jensen & Ramirez (2010) have given the cross-over value for the inequality $MSE(\widehat{\boldsymbol{\beta}}_R(k)) < MSE(\widehat{\boldsymbol{\beta}}_L)$; and, for the Hospital Manpower Data, they estimate the set of admissible values for k to be (0, 0.007263).

In the paragraphs to follow, we first present numerical evidence from the Hospital Manpower Data as it pertains to Theorems 1–3. We then undertake a synthesis of these findings with reference to the criteria of Table 3.

Ridge Traces. Ridge traces for the Hospital Manpower Data are listed in Table 8.9 (Myers, 1990) for $k \in [0.00, 0.24]$ by increments of 0.01. Companion to those [**Table 4 here**] are the scaled traces $\{\widehat{\beta}_R^i(k)/ \mid \widehat{\beta}_L^i \mid ; 1 \le i \le 5\}$ at selected values for k, as listed in Table 4. These are intended to adjust for scale. Additional quantities of interest are $\{\mathrm{Var}(\widehat{\beta}_R^i(k)); 1 \le i \le 5\}$, and their derivatives $\{d\mathrm{Var}(\widehat{\beta}_R^i(k))/dk; 1 \le i \le 5\}$, as *parameters* of their distributions, thus deterministic.

4.3. **Derivative Traces.** Turning to derivative traces $\{d\widehat{\beta}_R^i(k)/dk;\ 1 \le i \le 5\}$ as schemata for gauging instabilities, we first note from Table 2 that the ratio of the largest to smallest OLS values in magnitude is 7054. To adjust for these disparities, we scale the derivative traces as $\{[d\widehat{\beta}_R^i(k)/dk]/ \mid \widehat{\beta}_L^i \mid ; 1 \le i \le 5\}$, taking absolute values in the denominators so as to preserve signs of the derivatives. These appear in Table 5. [**Table 5 here.**] Suppose instead that these had been scaled by $\{\widehat{\beta}_R^i(k);\ 1 \le i \le p\}$ or their moduli. Then of the series $\{[d\widehat{\beta}_R^i(k)/dk]/ \mid \widehat{\beta}_R^i(k) \mid ; 1 \le i \le 5\}$, two would exhibit singularities since their ridge traces change signs at $\widehat{\beta}_R^4(0.0188) = 0 = \widehat{\beta}_R^5(0.0983)$. Having obviated those discontinuities, we see that entries in Table 5, as in Table 4, serve to equilibrate the unscaled values; moreover, they are scale–invariant, thus dimensionless, and hence free of σ^2. Nonetheless, Table 5 entries are random, subject to experimental variation in $\boldsymbol{Y}$, if not their variance.

Pursuant to Theorem 1, we next examine the derivative variances as k evolves, but taking $\{[d\mathrm{Var}(\widehat{\beta}_R^i(k))/dk]/\ \mathrm{Var}(\widehat{\beta}_L^i)\}$ to adjust for the widely disparate *OLS* variances in Table 2. These appear in Table 6. [**Table 6 here.**] This scaling serves to equilibrate entries in the table; more importantly, these ratios are scale–invariant and thus free of experimental variation in the observed $\boldsymbol{Y}$. Monotonicity as in Theorem 1(iii) is clearly evident.

Values $\{\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk);\ 1 \le i \le 5\}$ and $\{d[\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk)]/dk;\ 1 \le i \le 5\}$, companion to these, provide the deterministic traces of Theorem 3, as parameters of the derivative trace distributions, in concert with their stability as k evolves. Standard deviations are reported in Table 7, [**Table 7 here**] scaled again by corresponding *OLS* values, once more equilibrating the unscaled values, and assuring their scale–invariance and freedom from σ^2. Further, as in Theorem 3(i), these are seen to decrease monotonically with increasing k. In addition, derivatives $\{d[\mathrm{Var}(d\widehat{\beta}_R^i(k)/dk)]/dk;\ 1 \le i \le 5\}$, scaled by *OLS* values but not tabulated here, stand in further support of trends reported here. Moreover, these increase with increasing k as in Theorem 3(iii), often becoming vanishingly small.

A Synthesis. The variations on ridge traces of this study, identified as Tables 5, 6, and 7, are themselves somewhat disparate yet related. Let $\{TE_i(k);\ 1 \le i \le 5\}$ designate tabular entries in the columns of a typical table. Then values of $\{|TE_i(k)|;\ 1 \le i \le 5\}$, diminishing as k increases, characterize enhanced stability by each criterion. There is a plethora of entries; nonetheless, some concensus might emerge on seeking $k^{\dagger}(\delta)$ such that all entries across columns are dominated by a given threshold value $\delta > 0$. Specifically, take

$$k^{\dagger}(\delta) = \arg\left(\min_k \{|TE_1(k)|,\ |TE_2(k)|,\ |TE_3(k)|,\ |TE_4(k)|,\ |TE_5(k)|\} < \delta\right).$$

This undertaking was carried out using *Maple* software and three threshold values, namely, $\delta \in \{1, 10, 100\}$. Results are compiled in Table 8, [**Table 8 here**] showing remarkable consistency across the three tables, despite their diverse but related origins.

4.4. **Polynomial Representations.** As noted by Gibbons & McDonald (1984), the ridge traces are rational functions in k of degree $(p-1, p)$. From the canonical form, we use Maple to calculate the rational functions $\{k \to \boldsymbol{A}(\boldsymbol{Z'Z} + k\boldsymbol{I}_p)^{-1}\boldsymbol{Z'Y}_0; 1 \leq i \leq p\}$ which computes the ridge estimators in correlation units and transforms them back into natural units by $\boldsymbol{A}$. The rational representation of $\widehat{\beta}_R^i(k)$ as $R_i(k) = P_i(k)/Q(k)$ has degrees $(p-1, p)$, all having the common denominator $Q(k)$ of degree p, such that $R_i(k) \to 0$ as $k \to \infty$. To assist in displaying the coefficients, we again scale the ridge traces as $\widehat{\beta}_R^i(k)/|\widehat{\beta}_L^i| = R_i(k)/|\widehat{\beta}_L^i| = [P_i(k)/|P_i(0)|]/[Q(k)/|Q(0)|] \equiv P_i^0(k)/Q^0(k)$. The coefficients for $P_i^0(k)$ and $Q^0(k)$, to be denoted as $\{c_0, ..., c_4\}$ and $\{c_0, ..., c_5\}$, are shown in Table 9 and Table 10, respectively. Since all the coefficients of $Q^0(k)$ are positive, $Q^0(k) > 0$ for the ridge values $k > 0$. [**Table 9 here**.] Applying Descartes *Rule of Signs,* we find the maximum number of positive zeroes for $\{P_i^0(k); 1 \leq i \leq p\}$ to be $\{1, 0, 0, 1, 1\}$ respectively. Thus $\{\widehat{\beta}_R^2(k), \widehat{\beta}_R^3(k)\}$ have no sign changes, while $\{\widehat{\beta}_R^1(k), \widehat{\beta}_R^4(k), \widehat{\beta}_R^5(k)\}$ each have one sign change. [**Table 10 here**.]

The foregoing representations support explicit rational functions for the derivative traces. For the scaled ridge traces we set $\{\frac{d}{dk}(\widehat{\beta}_R^i(k)/|\widehat{\beta}_L^i|) = T_i^0(k)/[Q^0(k)]^2; 1 \leq i \leq p\}$ with $T_i^0(k)$ of degree $2p$ from the quotient rule for derivatives, and the common denominator $[Q^0(k)]^2$ as above of degree $2p$. The coefficients for the polynomial $T_i^0(k)$, to be denoted as $\{d_0, ..., d_8\}$, are shown in Table 11 for $\{T_i^0; 1 \leq i \leq p\}$. From Descartes *Rule of Signs,* the number of possible positive zeroes, $\#(zeroes)$, for $T_i^0(k)$, and their values as *zeroes*, are shown in the last two rows of Table 11. The denominator $[Q^0(k)]^2 > 0$, so the zeroes of $T_i^0(k)$ are the roots for the derivative traces, equivalently, the critical values for the ridge traces. As the ridge traces $\widehat{\beta}_R^i(k) \to 0$ as $k \to \infty$, the number of critical values of $\widehat{\beta}_R^i(k)$ provide an upper bound for the number of sign changes for $\widehat{\beta}_R^i(k)$. [**Table 11 here**]

For example, $T_3^0(k)$ has no zeroes and so $\widehat{\beta}_R^3(k)$ must tend monotonically to zero; whereas, $T_1^0(k)$ has one zero (equivalently, one critical value) permitting one sign change for $\widehat{\beta}_R^1(k)$, in agreement with the previously noted unique sign change for $\widehat{\beta}_R^1(k)$. These developments for derivatives parallel those for traces in Gibbons & McDonald (1984), Zhang & McDonald (2005), and McDonald (2009).

4.5. **Connections to Other Criteria.** As noted, criteria other than stability have driven the wide diversity of choices for k, including those of Table 3. Nonetheless, on the premise that stability of solutions has been a staple of (H&K, 1970a,b) from the beginning, our findings offer a lens through which those other criteria may be gauged. Note that $\{DF_k, GCV_k, C_k\}$ may be grouped as smallest in Table 3. Moreover, $DF_k = \text{tr}(\boldsymbol{H}_k)$ is identified (Myers, 1990) as the "perhaps more appealing" *df–trace* criterion of Tripp (1983), "namely, the effective regression degrees of freedom." Moreover, DF_k is a factor in C_k and GCV_k. Having plotted DF_k *vs* k as his Figure 8.4, Myers concludes: "Certainly $\text{tr}(\boldsymbol{H}_k)$ has stabilized before k=0.0004. Now, in order to illustrate how this reflects stability in coefficients, consider the information in Table 8.12, reflecting the coefficients values (in the natural variables) for k in the interval [0, 0.0004]. The $\text{tr}(\boldsymbol{H}_k)$ appears to be a reasonable composite criterion for reflecting stability in the regression coefficients;" (Myers, 1990). To the contrary, these claims appear *not* to be supported by the evidence. To wit, compare the finite Divergences in Table 1 with rescaled versions of the instantaneous derivatives at $k = 0.0004$ in Table 5, namely, [8254.82, 0.572197, −275.52, −28.6143, 5541.69].

Lesson learned: The similarities of ridge values for nearby k in Table 8.12 reflect continuity of ridge traces, not their stability, as noted previously. Similarly, from Tables 5–7 we see that overall stability arguably may be found in the neighborhood of $k = 0.10$, so that none of the values for $\{DF_k, GCV_k, C_k\}$ would portend stability in the sense of H&K(1970a,b). See also Table 8. Clearly these criteria work at crossed purposes to stability in the context of these data.

Such discrepancies may be found elsewhere. For the Tobacco Data of Table 8.13 (Myers, 1990), his Table 8.15 gives values for ridge traces of $[\widehat{\beta}_R^1(k), \widehat{\beta}_R^2(k), \widehat{\beta}_R^3(k), \widehat{\beta}_R^4(k)]$ for $k \in [0.000, 0.010]$ in increments of 0.001. Based on separate plots of the $\{DF_k, PRESS_k, C_k\}$ criteria and the ridge traces in his Table 8.15, Myers concludes that "if one were to use ridge regression in this data set, a value of k from 0.002 to 0.004 would be appropriate." Again we respectfully disagree. For if, in keeping with our Table 1, we recompute values for $\{[\widehat{\beta}_R^i(k_u) - \widehat{\beta}_R^i(k_v)]/(k_u - k_v); 1 \leq i \leq 4\}$, but now taking $k_u = 0.004$ and $k_v = 0.003$ together with Table 8.15 (Myers, 1990), we get the Divergences = $[-4195.0, 5066.1, 2501.5, -876.5]$ as approximate derivatives. In consequence, ridge traces are seen to diverge wildly for $k \in [0.002, 0.004]$, grossly devoid of evidence towards stability of ridge traces in the tobacco data. Again, the nearness of ridge values for nearby k in Table 8.15 reflects continuity, not stability. On the other hand, a careful examination of derivative traces, *in lieu of* ridge traces, could serve to obviate such misleading assertions.

In summary, users are reminded that the stability of solutions serves as the frontispiece of (H&K, 1970ab) from the outset. Nonetheless, in that numerous and disparate other criteria have been advocated in choosing k, our studies caution that choices for k based on other desiderata need not exhibit the requisite stability.

5. Summary and Conclusions

The widespread use of ridge regression continues apace, prominently now for calibration in chemical engineering and allied fields; see (Frank and Friedman, 1993; Geladi, 2002; Kalivas, 2005; and Sundberg, 1999), for example. The present study complements those basics through a consolidated approach for tracking the stability of prospective solutions in particular applications.

Rates of change of ridge traces are studied in assessing the stability of ridge solutions. Both ridge traces, and their derivatives, are subject to random disturbances in the observed $\boldsymbol{Y}$. On the other hand, since ill–conditioning resides exclusively in the matrix $\boldsymbol{X}$, it is natural to conjecture that critical properties, such as stability, might trace back to $\boldsymbol{X}$ alone, independently of $\boldsymbol{Y}$. An affirmative answer rests on two further metrics, namely, the derivatives of the variances of the ridge estimators, and the variances of the derivative traces. Both tend to zero as the ridge parameter increases, and both reflect the stabilizing of those distributions in a deterministic manner. Case studies in the highly ill–conditioned Hospital Manpower Data serve to illustrate the essential findings. Quantities in these studies are standardized so as to free the diagnostics from dependence on the observational variance σ^2, typically unknown. Users are cautioned that choices for k based on other desiderata need not exhibit the stability taken as the frontispiece of (H&K, 1970ab). It is noted further that, although ridge traces have been misconstrued on occasion as documented, a careful examination of derivative and allied traces could serve to circumvent any false and misleading assertions based on ridge traces.

References

[1] Allen, D.M.(1974). The relationship between variable selection and data augmentation and a method for prediction. Technometrics 16:125–127.

[2] Belsley, D.A.(1986). Centering, the constant, first–differencing, and assessing conditioning. In: Belsley, D.A. Kuh, E. ed., *Model Reliability.* MIT Press.

[3] Frank, I.E. Friedman, J.H.(1993). A statistical view of some chemometrics regression tools. Technometrics 35:109–135.

[4] Geladi, P.(2002). Some recent trends in the calibration literature. Chemometr. Intell. Lab. 60:211–224.

[5] Gibbons, D.I., McDonald, G.C.(1984). A rational interpretation of the ridge trace. Technometrics 26:339-346.

[6] Golub, G.H. Heath, C.G. Wahba, G.(1979). Generalized cross validation as a method for choosing a good ridge parameter. Technometrics 21:215–223.

[7] Hoerl, A.E. Kennard, R.W.(1970a). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55–67.

[8] Hoerl, A.E. Kennard, R.W.(1970b). Ridge regression: Application for nonorthogonal problems. Technometrics 12:69–82.

[9] Hoerl, A.E. Kennard, R.W. Baldwin, K.F.(1975). Ridge regression: Some simulations. Commun. Stat. 4:105–123.

[10] Jensen, D.R., Ramirez, D.E.(2010). Tracking MSE efficiencies in ridge regression. Advances and Applications in Statistical Sciences 1:381-398.

[11] Kalivas, J.H.(2005). Multivariate calibration, an overview. Anal. Lett. 38:2259–2279.

[12] Kibria, G.K.(2003). Performance of some new ridge regression estimators. Commun. Stat.–Simulat. 32:419–435.

[13] Mallows, C.L.(1973). Some comments on Cp. Technometrics 15:661–675.

[14] Marquardt, D.W.(1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12:591-612.

[15] McDonald, G.C.(1984). Discussion of: Ridge analysis following a preliminary test of the shrunken hypothesis. Technometrics 17:443-445.

[16] McDonald, G.C.(2009). Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics 1:93–100.

[17] Myers, R.H.(1990). *Classical and Modern Regression with Applications, Second ed.* Boston: PWS–KENT.

[18] Muniz, G. Kibra, B.M.G.(2009). On some ridge regression estimators: An empirical comparisons. Commun. Stat.–Simulat. 38:621–630.

[19] Sundberg, R.(1999). Multivariate calibration – direct and indirect regression methodology. Scand. J. Stat. 26:161–207.

[20] Tripp, R.E.(1983). *Non–Stochastic Ridge Regression and Effective Rank of the Regressors Matrix.* PhD thesis, Department of Statistics, Virginia Polytechnic Institute and State University.

[21] Zhang, R. McDonald, G.C.(2005). Characterization of ridge trace behavior. Commun. Stat.–Theor. M. 34:1487–1501.

TABLE 1. Ridge solutions $\widehat{\boldsymbol{\beta}}_R(k) = [\widehat{\beta}_R^1(k), \ldots, \widehat{\beta}_R^5(k)]'$ from Table 8.12 of Myers (1990); divergences $\{[\widehat{\beta}_R^i(k_u) - \widehat{\beta}_R^i(k_v)]/(k_u - k_v)$, with $k_u > k_v$; and derivatives $\{d\widehat{\beta}_R^i/dk\}$; for $\{1 \le i \le 5\}$ and coefficient values in the natural variables at selected k.

k	$\widehat{\beta}_R^1(k)$	$\widehat{\beta}_R^2(k)$	$\widehat{\beta}_R^3(k)$	$\widehat{\beta}_R^4(k)$	$\widehat{\beta}_R^5(k)$
	Ridge Solutions				
$k_1 = 0.00030$	11.4767	0.0564	0.7201	-5.4163	-416.42
$k_2 = 0.00035$	12.0761	0.0564	0.7004	-5.4238	-416.32
$k_3 = 0.00040$	12.5411	0.0565	0.6849	-5.4249	-416.09
	Divergences $\{[\widehat{\beta}_R^i(k_u) - \widehat{\beta}_R^i(k_v)]/(k_u - k_v); 1 \le i \le 5\}$				
(k_1, k_2)	11988	0	-394	-150	2000
(k_2, k_3)	9300	2	-310	-22	4600
	Derivatives $\{d\widehat{\beta}_R^i/dk; 1 \le i \le 5\}$				
$k_1 = 0.00030$	13710.4027	0.6414	-448.3150	-234.5197	376.4642
$k_2 = 0.00035$	10478.0469	0.6007	-345.9388	-78.4742	3441.5818
$k_3 = 0.00040$	8253.7726	0.5722	-275.4801	28.5845	5540.9055

TABLE 2. *OLS* estimates $\{\widehat{\beta}_L^i; 1 \le i \le 5\}$ and their variances for the Hospital Manpower Data.

i	1	2	3	4	5
$\widehat{\beta}_L^i$	-15.8508	0.0559	1.5896	-4.2196	-394.3280
$\text{Var}(\widehat{\beta}_L^i)/\sigma^2$	0.023126	1.096E-9	0.000023	0.000125	0.106589

TABLE 3. Choices for k in the Hospital Manpower Data for conventional criteria, DF_k, GCV_k, C_k, $PRESS_k$, and HKB_k; and corresponding values for $\max[VIF(\widehat{\beta}_i); 1 \le i \le 5]$.

Name	DF_k	GCV_k	C_k	$PRESS_k$	HKB_k
Value for k	0.000400	0.004787	0.005000	0.230000	0.616960
$\max[VIF(\widehat{\beta}_i); 1 \le i \le 5]$	1146.4314	112.6210	108.0925	3.9338	2.0374

TABLE 4. Scaled traces $\{\widehat{\beta}_R^i(k)/\mid\widehat{\beta}_L^i\mid; 1\leq i\leq 5\}$ from ridge and *OLS* solutions in the natural variables at selected values for k.

k	$\frac{\widehat{\beta}_R^1(k)}{\lvert\widehat{\beta}_L^1\rvert}$	$\frac{\widehat{\beta}_R^2(k)}{\lvert\widehat{\beta}_L^2\rvert}$	$\frac{\widehat{\beta}_R^3(k)}{\lvert\widehat{\beta}_L^3\rvert}$	$\frac{\widehat{\beta}_R^4(k)}{\lvert\widehat{\beta}_L^4\rvert}$	$\frac{\widehat{\beta}_R^5(k)}{\lvert\widehat{\beta}_L^5\rvert}$
0.0003	0.72400	1.00852	0.45298	-1.28389	-1.05605
0.0004*	0.79115	1.00960	0.43089	-1.28594	-1.05522
0.004748*	0.95833	1.04279	0.34281	-0.91230	-0.92607
0.0050*	0.95700	1.04417	0.34170	-0.89439	-0.92008
0.0100	0.91956	1.07231	0.32106	-0.51513	-0.79304
0.1000	0.64357	1.16936	0.21600	1.69797	0.00635
0.2300*	0.55670	1.13101	0.18497	2.21637	0.29450
0.4000	0.50949	1.07193	0.16833	2.34183	0.45709
0.61696*	0.47311	1.00649	0.15586	2.32467	0.56011
0.8000	0.45008	0.95911	0.14808	2.27055	0.60702
1.0000	0.42898	0.91359	0.14101	2.20025	0.63612

TABLE 5. Scaled derivative traces $\{[d\widehat{\beta}_R^i(k)/dk]/\mid\widehat{\beta}_L^i\mid; 1\leq i\leq 5\}$ from ridge and *OLS* solutions in the natural variables at selected values for k.

k	$\frac{[d\widehat{\beta}_R^1(k)/dk]}{\lvert\widehat{\beta}_L^1\rvert}$	$\frac{[d\widehat{\beta}_R^2(k)/dk]}{\lvert\widehat{\beta}_L^2\rvert}$	$\frac{[d\widehat{\beta}_R^3(k)/dk]}{\lvert\widehat{\beta}_L^3\rvert}$	$\frac{[d\widehat{\beta}_R^4(k)/dk]}{\lvert\widehat{\beta}_L^4\rvert}$	$\frac{[d\widehat{\beta}_R^5(k)/dk]}{\lvert\widehat{\beta}_L^5\rvert}$
0.0003	865.0632	11.46946	-282.0610	-55.56690	0.957827
0.0004*	520.7822	10.23025	-173.3220	6.781230	14.05348
0.004787*	-6.112760	6.518157	-5.339030	84.42942	28.25171
0.0050*	-6.400430	6.439555	-5.180520	83.71277	28.01037
0.0100	-7.614580	4.906354	-3.448770	68.57293	23.02669
0.1000	-1.187950	-0.108010	-0.429810	8.448551	3.721705
0.2300*	-0.394430	-0.357500	-0.139640	1.673329	1.379640
0.4000	-0.207680	-0.328640	-0.071870	0.181898	0.666474
0.61696*	-0.139630	-0.276540	-0.047420	-0.239400	0.331730
0.8000	-0.114410	-0.242590	-0.038480	-0.333940	0.193578
1.0000	-0.097800	-0.213810	-0.032680	-0.361430	0.104965

TABLE 6. Scaled derivatives $\{[d\mathrm{Var}(\widehat{\beta}_R^i(k))/dk]/\mathrm{Var}(\widehat{\beta}_L^i); 1 \le i \le 5\}$ of variances of ridge traces $\widehat{\boldsymbol{\beta}}_R(k) = [\widehat{\beta}_R^1(k), \ldots, \widehat{\beta}_R^5(k)]'$, independently of $\boldsymbol{Y}$, in the natural variables at selected k.

k	$\frac{d\mathrm{Var}(\widehat{\beta}_R^1(k))/dk}{\mathrm{Var}(\widehat{\beta}_L^1)}$	$\frac{d\mathrm{Var}(\widehat{\beta}_R^2(k))/dk}{\mathrm{Var}(\widehat{\beta}_L^2)}$	$\frac{d\mathrm{Var}(\widehat{\beta}_R^3(k))/dk}{\mathrm{Var}(\widehat{\beta}_L^3)}$	$\frac{d\mathrm{Var}(\widehat{\beta}_R^4(k))/dk}{\mathrm{Var}(\widehat{\beta}_L^4)}$	$\frac{d\mathrm{Var}(\widehat{\beta}_R^5(k))/dk}{\mathrm{Var}(\widehat{\beta}_L^5)}$
0.0003	-131.3340	-23.7211	-131.3210	-80.5279	-50.9787
0.0004*	-62.2733	-23.4101	-62.2712	-50.1802	-37.0551
0.004787*	-0.07166	-19.5799	-0.07696	-17.3429	-18.4984
0.0050*	-0.06514	-19.4258	-0.07037	-17.1169	-18.2545
0.0100	-0.02047	-16.2852	-0.02430	-12.8401	-13.6422
0.1000	-0.00075	-2.26490	-0.00093	-0.83550	-1.20534
0.2300*	-0.00011	-0.49748	-0.00014	-0.14790	-0.44199
0.4000	-0.00003	-0.14940	-0.00003	-0.04113	-0.22868
0.61696*	-9.92E-6	-0.05493	-0.00001	-0.01475	-0.12565
0.8000	-5.55E-6	-0.02987	-6.46E-6	-0.00801	-0.08334
1.0000	-3.54E-6	-0.01770	-4.06E-6	-0.00477	-0.05664

TABLE 7. Standard deviations of derivative traces $(d\widehat{\beta}_R^i/dk)$ relative to $\widehat{\beta}_L^i$, namely, $\{\phi(\widehat{\beta}_i) = [\mathrm{Var}(d\widehat{\beta}_R^i/dk)/\mathrm{Var}(\widehat{\beta}_L^i)]^{1/2}; 1 \le i \le 5\}$, independent of $\boldsymbol{Y}$, for solutions in the natural variables at selected k.

k	$\phi(\widehat{\beta}_1)$	$\phi(\widehat{\beta}_2)$	$\phi(\widehat{\beta}_3)$	$\phi(\widehat{\beta}_4)$	$\phi(\widehat{\beta}_5)$
0.0003	430.6650	27.18683	430.6328	285.2097	192.9238
0.0004*	261.8325	19.29229	261.8130	173.8692	118.0531
0.004787*	2.350305	11.00050	2.362574	13.35063	13.86791
0.0050*	2.163111	10.93410	2.176214	13.21852	13.73480
0.0100	0.654137	9.570927	0.682494	10.77608	11.16617
0.1000	0.050908	2.420475	0.056744	1.617668	1.616061
0.2300*	0.014107	0.861670	0.015622	0.490641	0.605443
0.4000	0.005436	0.375612	0.005991	0.201063	0.348502
0.61696*	0.002499	0.186296	0.002743	0.097036	0.225747
0.8000	0.001568	0.120870	0.001715	0.062409	0.169821
1.0000	0.001061	0.083042	0.001156	0.042733	0.130359

TABLE 8. Minimal values $k^\dagger(\delta)$ for k required to achieve $\{| TE_i(k) | < \delta;\ 1 \le i \le p\}$, where $\{TE_i(k);\ 1 \le i \le p\}$ designate typical entries in the columns of Tables 5, 6 and 7, successively.

Threshold Values	Table 5	Table 6	Table 7
$\{\lvert TE_i(k)\rvert < \delta;\ 1 \le i \le 5\}$	Minimal $k^\dagger(\delta)$		
$\delta = 100$	0.0009333	0.0003336	0.0006806
$\delta = 10$	0.08952	0.02620	0.01292
$\delta = 1$	0.29670	0.16120	0.20670

TABLE 9. Coefficients for the normalized function $Q^0(k) = Q(k)/|Q(0)|$ for the scaled ridge traces $\{\widehat{\beta}_R^i(k)/|\widehat{\beta}_L^i| = P_i^0(k)/Q^0(k); 1 \le i \le 5\}$.

c_0	c_1	c_2	c_3	c_4	c_5
1	18,566.2	683,431	5,946,835	8,583,693	1,716,739

TABLE 10. Coefficients for the normalized functions $\{P_i^0(k); 1 \le i \le 5\}$ for the scaled ridge traces $\{\widehat{\beta}_R^i(k)/|\widehat{\beta}_L^i| = P_i^0(k)/Q^0(k); 1 \le i \le 5\}$.

c	$P_1^0(k)$	$P_2^0(k)$	$P_3^0(k)$	$P_4^0(k)$	$P_5^0(k)$
c_0	-1	1	1	-1	-1
c_1	19,199.5	18,705.1	6,633.4	-25,358.6	-19,986.1
c_2	459,002.0	847,501.8	154,394.7	1,033,466.4	-82,833.9
c_3	3,106,914.5	7,037,063.0	1,022,197.4	16,573,943.7	2,042,253.5
c_4	3,685,419.0	7,581,429.8	1,206,745.1	19,710,531.2	8,842,320.5

TABLE 11. Coefficients for the rational functions $T_i^0(k)$ for derivatives of the scaled ridge traces $\{\frac{d}{dk}\widehat{\beta}_R^i(k)/|\widehat{\beta}_L^i| = T_i^0(k)/[Q^0(k)]^2; 1 \le i \le 5\}$.

c	$T_1^0(k)$	$T_2^0(k)$	$T_3^0(k)$	$T_4^0(k)$	$T_5^0(k)$
d_0	37.766E03	138.967	-11.933E03	-67.924E02	-14.199E02
d_1	22.849E05	32.814E04	-10.581E05	34.338E05	12.012E05
d_2	-45.725E08	29.545E08	-16.817E08	36.586E09	12.145E09
d_3	-11.294E10	38.825E09	-40.968E09	91.715E10	31.361E10
d_4	-89.535E10	-29.002E10	-32.317E10	69.322E11	28.955E05
d_5	-29.737E11	-43.151E10	-10.467E11	93.738E11	13.646E12
d_6	-71.137E11	-19.683E12	-23.931E11	-30.373E12	35.480E12
d_7	-10.668E12	-24.162E12	-35.097E11	-56.906E12	-70.120E11
d_8	-6.328E12	-13.015E12	-20.717E11	-33.838E12	-15.180E12
$\#(zeroes)$	1	1	0	2	2
zeroes	0.00301	0.08727	*none*	{0.38484E–3, 0.45904}	{0.29507E–3, 1.52119}

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF VIRGINIA, CHARLOTTESVILLE, VA 22904–4137, USA