

# Variations on Ridge Traces in Regression

DONALD R. JENSEN AND DONALD E. RAMIREZ

Department of Mathematics, University of Virginia, Charlottesville,  
Virginia, USA

*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. In contrast to ridge traces and their derivatives, neither of the latter depends on observed responses, and both support deterministic assessments.*

**Keywords** Derivative traces; Deterministic traces; Ill-conditioned models; Ridge regression; Stability of ridge traces.

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

## 1. Introduction

### 1.1. Overview

In a full-rank model  $\{Y = X\beta + \epsilon\}$  having zero-mean, uncorrelated, and homoscedastic errors, the Ordinary Least Squares (OLS) estimators  $\hat{\beta}_L$  solve the  $p$  equations  $\{X'X\beta = X'Y\}$ . These solutions are unbiased with dispersion matrix  $V(\hat{\beta}_L) = \sigma^2(X'X)^{-1}$ . Near-dependency among the columns of  $X$ , as *ill-conditioning*, engenders “crucial elements of  $X'X$  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  $X$ ,” (Belsley, 1986). Among continuing palliatives are the ridge system  $\{(X'X + kI_p)\beta = X'Y; k \geq 0\}$  (Hoerl and Kennard, 1970a) (hereafter H&K, 1970a) with solutions  $\{\hat{\beta}_R(k); k \geq 0\}$ . 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  $\hat{\beta}_R(k) = [\hat{\beta}_R^1(k), \dots, \hat{\beta}_R^p(k)]'$  and taking the *ridge trace* as graphs of  $\{k \rightarrow \hat{\beta}_R^i(k); 1 \leq i \leq p\}$ , this provides “a two-dimensional graphical procedure for portraying the complex relationships in multifactor data” (H&K, 1970b).

Received August 31, 2010; Accepted May 2, 2011

Address correspondence to Donald E. Ramirez, Department of Mathematics, University of Virginia, Charlottesville, VA 22904-4137, USA; E-mail: der@virginia.edu

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  $\{\hat{\beta}_R^i(k); 1 \leq i \leq p\}$  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 and McDonald (1984) showed 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 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 and 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  $Y$ , 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  $X$ , 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 of 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. 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  $\{[\hat{\beta}_R^i(k_u) - \hat{\beta}_R^i(k_v)]/(k_u - k_v); 1 \leq i \leq 5\}$ , as changes in  $\hat{\beta}_R^i(k)$  per unit change in  $k$ , accurately portray the local variation in  $\{\hat{\beta}_R^i(k); 1 \leq i \leq 5\}$  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.

**Table 1**

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

$k$	$\hat{\beta}_R^1(k)$	$\hat{\beta}_R^2(k)$	$\hat{\beta}_R^3(k)$	$\hat{\beta}_R^4(k)$	$\hat{\beta}_R^5(k)$
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 $\{[\hat{\beta}_R^i(k_u) - \hat{\beta}_R^i(k_v)]/(k_u - k_v); 1 \leq i \leq 5\}$					
$(k_1, k_2)$	11988	0	-394	-150	2000
$(k_2, k_3)$	9300	2	-310	-22	4600
Derivatives $\{d\hat{\beta}_R^i/dk; 1 \leq i \leq 5\}$					
$k_1 = 0.00030$	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

## 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  $A$  are  $A'$ ,  $A^{-1}$ , and  $\text{tr}(A)$ . Special arrays include the identity  $I_n$ , the unit vector  $\mathbf{1}_n = [1, 1, \dots, 1]'$   $\in \mathbb{R}^n$ , the diagonal matrix  $D(a_i) = \text{Diag}(a_1, \dots, a_p)$ , and  $B_n = (I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n')$ . Let  $Z(n \times p)$  have rank  $p < n$ ; its *singular decomposition* is  $Z = PD_\xi Q'$ , with  $D_\xi = \text{Diag}(\xi_1, \dots, \xi_p)$  as its *ordered singular values*  $\{\xi_1 \geq \xi_2 \geq \dots \geq \xi_p > 0\}$ , and with the columns of  $P = [p_1, \dots, p_p]$  and of  $Q = [q_1, \dots, q_p]$  comprising the *left-* and *right-singular vectors* of  $Z$ , such that  $P'P = I_p$  and  $Q$  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  $\{Y = \beta_0\mathbf{1}_n + X\beta + \epsilon\}$  with intercept, take  $Y_0 = B_n Y$  with  $B_n = (I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n')$ ; observe that  $B_n\mathbf{1}_n = \mathbf{0}$  and  $B_n^2 = B_n$ ; and let  $S_c = \text{Diag}(S_1, \dots, S_p)$  comprise reciprocals of square roots of the diagonals of  $X'B_n X$ . This serves to center and scale elements of  $X$ , so that  $Z'Z = S_c X' B_n X S_c$  is in the conventional “correlation form.” These changes support the transitions

$$\{Y = \beta_0\mathbf{1}_n + X\beta + \epsilon\} \rightarrow \{B_n Y = \beta_0 B_n \mathbf{1}_n + B_n X S_c S_c^{-1} \beta + \eta\} \tag{1}$$

$$\rightarrow \{Y_0 = Z\omega + \eta\} \rightarrow \{Y_0 = PD_\xi \theta + \eta\}. \tag{2}$$

The singular decomposition of  $Z = B_n X S_c$  is  $Z = PD_\xi Q'$ ;  $\omega = S_c^{-1}\beta$ ;  $\eta = B_n \epsilon$ ; and  $\theta = Q'\omega = Q'S_c^{-1}\beta$  is linear in the original  $\beta$ . From this, our ultimate

canonical form is

$$\{Y_0 = PD_\xi\theta + \eta\} \rightarrow \{P'Y_0 = P'PD_\xi\theta + P'\eta\} \quad (3)$$

$$\rightarrow \{W = D_\xi\theta + \tau\}, \quad (4)$$

where  $Y_0 \rightarrow W = P'Y_0$ ,  $P'P = I_p$ , and  $\tau = P'\eta$ .

Ridge regression traditionally is carried out as  $\{(Z'Z + kI_p)\omega = Z'Y_0\}$ , or equivalently,  $\{(D_\xi^2 + kI_p)\theta = D_\xi W\}$ , giving  $\hat{\theta}_R(k) = D(\xi_i/(\xi_i^2 + k))W$  and  $\hat{\omega}_R(k) = Q\hat{\theta}_R(k)$  as the ridge solutions in correlation form, where

$$D(\xi_i/(\xi_i^2 + k)) = \text{Diag}(\xi_1/(\xi_1^2 + k), \dots, \xi_p/(\xi_p^2 + k)). \quad (5)$$

These in turn map back to the original coordinates through  $\hat{\beta}_R(k) = A\hat{\theta}_R(k)$ , with  $A = S_c Q$ , i.e.,  $\{\hat{\beta}_R^i(k) = a_i'\hat{\theta}_R(k); 1 \leq i \leq p\}$  with  $A' = [a_1, \dots, a_p]$ . For an orthogonal system with  $X'X$  diagonal, the orthogonal matrices are  $P = I_n$  and  $Q = I_p$ , with the pull-back matrix  $A = I_p$ .

For subsequent reference observe that

$$V(\hat{\theta}_R(k)) = \sigma^2 D(\xi_i^2/(\xi_i^2 + k)^2) \implies V(\hat{\beta}_R(k)) = \sigma^2 A D(\xi_i^2/(\xi_i^2 + k)^2) A'. \quad (6)$$

Moreover, since  $\{d(\hat{\theta}_R^i(k))/dk = -(\xi_i/(\xi_i^2 + k)^2)W_i; 1 \leq i \leq p\}$ , the canonical traces  $\{k \rightarrow \hat{\theta}(k); 1 \leq i \leq p\}$  are monotone in  $k$ . On the original scale, however,  $\{k \rightarrow \hat{\beta}_R^i(k) = a_i'\hat{\theta}_R(k)\}$  need not be monotone, as seen in Sec. 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  $Y$ . 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"  $\{k \rightarrow \text{Var}(\hat{\beta}_R^i(k)); 1 \leq i \leq p\}$  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  $X$ , free of random disturbances in  $Y$ . 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  $\{k \rightarrow d(\text{Var}(\hat{\beta}_R^i(k)))/dk; 1 \leq i \leq p\}$ . Some properties may be listed as follows.

**Theorem 3.1.** Consider the ridge traces  $\{\hat{\beta}_R^i(k); k \geq 0\}$ , their variances  $\{\text{Var}(\hat{\beta}_R^i(k)); k \geq 0\}$ , and the derivatives  $\{d\text{Var}(\hat{\beta}_R^i(k))/dk; k \geq 0\}$ , for  $1 \leq i \leq p$ . Then:

- (i) variances are given by  $\{\text{Var}(\hat{\beta}_R^i(k)) = \sigma^2 a_i' D(\xi_i^2/(\xi_i^2 + k)^2) a_i; 1 \leq i \leq p\}$ , each monotone decreasing with increasing  $k$ ;

(ii) rates of change in  $\{\text{Var}(\hat{\beta}_R^i(k)); 1 \leq i \leq p\}$  are given by

$$\{d\text{Var}(\hat{\beta}_R^i(k))/dk = -2\sigma^2 \mathbf{a}'_i \mathbf{D}(\xi_i^2/(\xi_i^2 + k)^3) \mathbf{a}_i < 0; 1 \leq i \leq p\}$$

independently of  $\mathbf{Y}$ ;

(iii) the negative functions  $\{d\text{Var}(\hat{\beta}_R^i(k))/dk; 1 \leq i \leq p\}$  are monotone increasing as  $k \uparrow$  for  $k \geq 0$ , 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  $\{d^2\text{Var}(\hat{\beta}_R^i(k))/dk^2 = 6\sigma^2 \mathbf{a}'_i \mathbf{D}(\xi_i^2/(\xi_i^2 + k)^4) \mathbf{a}_i > 0; 1 \leq i \leq p\}$  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.  $\square$

### 3.2. Derivative Traces

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

**Theorem 3.2.** Consider the ridge trace  $\{k \rightarrow \hat{\beta}_R(k)\}$  and its derivative as the vector gradient  $\frac{d\hat{\beta}_R(k)}{dk} = \left[ \frac{d\hat{\beta}_R^1(k)}{dk}, \frac{d\hat{\beta}_R^2(k)}{dk}, \dots, \frac{d\hat{\beta}_R^p(k)}{dk} \right]'$ . Then the following properties hold.

- (i)  $d\hat{\beta}_R(k)/dk = \mathbf{A}d\hat{\theta}_R(k)/dk = -\mathbf{A}\mathbf{D}(\xi_i/(\xi_i^2 + k)^2)\mathbf{W}$ .
- (ii) For each  $k \geq 0$ , the dispersion matrix is  $\text{V}(d\hat{\beta}_R(k)/dk) = \sigma^2 \mathbf{A}\mathbf{D}(\xi_i^2/(\xi_i^2 + k)^4)\mathbf{A}'$  independently of  $\mathbf{Y}$ .
- (iii) The ratio of variances of  $\hat{\theta}_R^i(k)$  to  $d\hat{\theta}_R^i(k)/dk$  is  $\text{Var}(\hat{\theta}_R^i(k))/\text{Var}(d\hat{\theta}_R^i(k)/dk) = (\xi_i^2 + k)^2$ , for  $1 \leq i \leq p$  and each  $k > 0$ .

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

Ridge traces, and now their derivatives, are subject to chance disturbances intrinsic to  $\mathbf{Y}$  as noted and, from conclusion (iii), each may be more or less variable than the other, depending on the weights  $\{\mathbf{a}_i; 1 \leq i \leq p\}$  and signs of  $\{[(\xi_i^2 + k) - 1]; 1 \leq i \leq p\}$ . We next seek versions of traces devoid of disturbances in  $\mathbf{Y}$ , depending only on  $\mathbf{X}$ . 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,  $\{\text{Var}(d\hat{\beta}_R^i(k)/dk); 1 \leq i \leq p\}$ , and derivatives of the latter as  $\{d[\text{Var}(d\hat{\beta}_R^i(k)/dk)]/dk; 1 \leq i \leq p\}$ , in seeking evidence for evolving stochastic stability of the ridge solutions. Details may be collected as follows.

**Theorem 3.3.** Consider the derivative traces  $\{k \rightarrow d\hat{\beta}_R^i(k)/dk; 1 \leq i \leq p\}$  and their variances  $\{\text{Var}(d\hat{\beta}_R^i(k)/dk); 1 \leq i \leq p\}$ . Then:

- (i) for each  $k > 0$ ,  $\{\text{Var}(d\hat{\beta}_R^i(k)/dk) = \sigma^2 \mathbf{a}'_i \mathbf{D}(\xi_i^2/(\xi_i^2 + k)^4) \mathbf{a}_i; 1 \leq i \leq p\}$ . Moreover, each decreases monotonically with increasing  $k > 0$  for  $1 \leq i \leq p$ ;
- (ii) rates of change in  $\{\text{Var}(d\hat{\beta}_R^i(k)/dk); 1 \leq i \leq p\}$  are given by  $\{d[\text{Var}(d\hat{\beta}_R^i(k)/dk)]/dk = -4\sigma^2 \mathbf{a}'_i \mathbf{D}(\xi_i^2/(\xi_i^2 + k)^5) \mathbf{a}_i < 0; 1 \leq i \leq p\}$  independently of  $\mathbf{Y}$ .
- (iii) the negative functions  $\{d[\text{Var}(d\hat{\beta}_R^i(k)/dk)]/dk; 1 \leq i \leq p\}$  are monotone increasing as  $k \uparrow$  for  $k \geq 0$ , their values progressing from large to small in magnitude.

*Proof.* Conclusion (i) extracts  $\{\text{Var}(d\hat{\beta}_R^i(k)/dk); 1 \leq i \leq p\}$  as diagonal elements from Theorem 3.2(ii); differentiation yields conclusion (ii); and monotonicity in (i) follows from negative slopes and that  $\{d^2 \text{Var}(d\hat{\beta}_R^i(k)/dk)/dk^2 = 20\sigma^2 \mathbf{a}'_i \mathbf{D}(\xi_i^2/(\xi_i^2 + k)^6) \mathbf{a}_i > 0\}$  are positive. Conclusion (iii) follows on checking signs of their first and second derivatives as in Theorem 3.1, to complete our proof.  $\square$

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  $\mathbf{X}$  alone, whereas ridge traces, and now their derivatives, must await the empirical outcome of  $\mathbf{Y}$  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  $\div 1000$  ( $X_4$ ); and average length of patients' stay in days ( $X_5$ ). The applicable model is

$$\{\mathbf{Y}_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon_i \leq i \leq 17\}. \quad (7)$$

Our computations utilize *Proc IML* of the SAS Programming System and the symbolic program *Maple*. The *OLS* estimates  $\hat{\beta}_L = [\hat{\beta}_L^1, \hat{\beta}_L^2, \hat{\beta}_L^3, \hat{\beta}_L^4, \hat{\beta}_L^5]'$  and their variances, in the original coordinates, are listed in Table 2. Following convention (for example the *Ridge* option of *Proc Reg* in the SAS system), ridge regression proceeds on first centering and scaling, taking  $\mathbf{X}'\mathbf{X} \rightarrow \mathbf{Z}'\mathbf{Z}$  in correlation form; solving the ridge equations; then mapping back onto the natural coordinates as in Sec. 2.2. The data are remarkably ill conditioned: singular values of  $\mathbf{Z}$  are  $\mathbf{D}_\xi = \text{Diag}(2.048687, 0.816997, 0.307625, 0.201771, 0.007347)$ , and the condition number is  $c_1(\mathbf{Z}'\mathbf{Z}) = 77, 754.86$ . Throughout  $24.30 \text{ E}02$  designates  $24.30 \times 10^2$ , for example, in scientific notation.

**Table 2**

*OLS* estimates  $\{\hat{\beta}_L^i; 1 \leq i \leq 5\}$  and their variances for the Hospital Manpower Data

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

**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, together with their values for the Hospital Manpower Data. These encompass  $DF_k = \text{tr}(\mathbf{H}_k) = \sum_{i=1}^p \xi_i^2 / (\xi_i^2 + k)$  with  $\mathbf{H}_k = [\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p)^{-1}\mathbf{Z}']$ ;  $PRESS_k = \sum_{i=1}^n e_{(i,\lambda)}^2$  as the cross-validation statistic (Allen, 1974);  $GCV_k = SS_{Res,k} / [n - (1 + \text{tr}(\mathbf{H}_k))]^2$ , a rotation-invariant analog called *Generalized Cross Validation* (Golub et al., 1979);  $C_k = [(SS_{Res,k} / \hat{\sigma}^2) - n + 2 + 2\text{tr}(\mathbf{H}_k)]$  to achieve a variance-bias trade-off (Mallows, 1973); and  $HKB_k = \hat{\sigma}^2 / \hat{\beta}'_L \hat{\beta}_L$  as in (Hoerl et al., 1975) from simulation studies. Here,  $SS_{Res,k}$  and  $\hat{\sigma}^2$ , 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 Kibra (2009), among others.

Marquardt (1970) noted that the variation inflation factors  $VIF_i(k)$  for the ridge estimators are the diagonal elements of  $(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p)^{-1}$ , and gives the rule of thumb that  $\{VIF_i(k) < 10; 1 \leq i \leq p\}$  in choosing *k*. Values for  $\max[VIF_i(k); 1 \leq i \leq p]$  are shown in Table 3 for *k* as listed there. We note that all  $\{VIF_i(k) < 10; 1 \leq i \leq p\}$  for  $k > 0.00706$ . Moreover, Jensen and Ramirez (2010) gave the cross-over value for the inequality  $MSE(\hat{\beta}_R(k)) < MSE(\hat{\beta}_L)$ ; 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 3.1–3.3. We then undertake a synthesis of these findings with reference to the criteria of Table 3.

**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  $\max[VIF(\hat{\beta}_i); 1 \leq i \leq 5]$

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

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

**4.3. Derivative Traces**

Turning to derivative traces  $\{d\hat{\beta}_R^i(k)/dk; 1 \leq i \leq 5\}$  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  $\{|d\hat{\beta}_R^i(k)/dk|/|\hat{\beta}_L^i|; 1 \leq i \leq 5\}$ , taking absolute values in the denominators so as to preserve signs of the derivatives. These appear in Table 5. Suppose instead that these had been scaled by  $\{\hat{\beta}_R^i(k); 1 \leq i \leq p\}$  or their moduli. Then of the series  $\{|[d\hat{\beta}_R^i(k)/dk]/\hat{\beta}_R^i(k)|; 1 \leq i \leq 5\}$ , two would exhibit singularities since their ridge traces change signs at  $\hat{\beta}_R^4(0.0188) = 0 = \hat{\beta}_R^5(0.0983)$ . 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  $\sigma^2$ . Nonetheless, Table 5 entries are random, subject to experimental variation in  $\mathbf{Y}$ , if not their variance.

Pursuant to Theorem 3.1, we next examine the derivative variances as  $k$  evolves, but taking  $\{|d\text{Var}(\hat{\beta}_R^i(k))/dk|/\text{Var}(\hat{\beta}_L^i)\}$  to adjust for the widely disparate *OLS* variances in Table 2. These appear in Table 6. 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  $\mathbf{Y}$ . Monotonicity as in Theorem 3.1(iii), is clearly evident.

Values  $\{\text{Var}(d\hat{\beta}_R^i(k)/dk); 1 \leq i \leq 5\}$  and  $\{d[\text{Var}(d\hat{\beta}_R^i(k)/dk)]/dk; 1 \leq i \leq 5\}$ , companion to these, provide the deterministic traces of Theorem 3.3, as parameters of the derivative trace distributions, in concert with their stability as  $k$  evolves.

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

$k$	$\frac{\hat{\beta}_R^1(k)}{ \hat{\beta}_L^1 }$	$\frac{\hat{\beta}_R^2(k)}{ \hat{\beta}_L^2 }$	$\frac{\hat{\beta}_R^3(k)}{ \hat{\beta}_L^3 }$	$\frac{\hat{\beta}_R^4(k)}{ \hat{\beta}_L^4 }$	$\frac{\hat{\beta}_R^5(k)}{ \hat{\beta}_L^5 }$
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

Downloaded by [Donald Jensen] at 07:09 11 October 2011



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

$k$	$\frac{[d\hat{\beta}_R^1(k)/dk]}{ \hat{\beta}_L^1 }$	$\frac{[d\hat{\beta}_R^2(k)/dk]}{ \hat{\beta}_L^2 }$	$\frac{[d\hat{\beta}_R^3(k)/dk]}{ \hat{\beta}_L^3 }$	$\frac{[d\hat{\beta}_R^4(k)/dk]}{ \hat{\beta}_L^4 }$	$\frac{[d\hat{\beta}_R^5(k)/dk]}{ \hat{\beta}_L^5 }$
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

Standard deviations are reported in Table 7, scaled again by corresponding *OLS* values, once more equilibrating the unscaled values, and assuring their scale-invariance and freedom from  $\sigma^2$ . Further, as in Theorem 3.3(i), these are seen to decrease monotonically with increasing  $k$ . In addition, derivatives  $\{d[\text{Var}(d\hat{\beta}_R^i(k)/dk)]/dk; 1 \leq i \leq 5\}$ , 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 3.3(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  $\{TE_i(k); 1 \leq i \leq 5\}$  designate tabular entries in the columns of a typical table. Then values of

**Table 6**  
 Scaled derivatives  $\{[d\text{Var}(\hat{\beta}_R^i(k))/dk]/\text{Var}(\hat{\beta}_L^i); 1 \leq i \leq 5\}$  of variances of ridge traces  $\hat{\beta}_R(k) = [\hat{\beta}_R^1(k), \dots, \hat{\beta}_R^5(k)]'$ , independently of  $Y$ , in the natural variables at selected  $k$

$k$	$\frac{d\text{Var}(\hat{\beta}_R^1(k))/dk}{\text{Var}(\hat{\beta}_L^1)}$	$\frac{d\text{Var}(\hat{\beta}_R^2(k))/dk}{\text{Var}(\hat{\beta}_L^2)}$	$\frac{d\text{Var}(\hat{\beta}_R^3(k))/dk}{\text{Var}(\hat{\beta}_L^3)}$	$\frac{d\text{Var}(\hat{\beta}_R^4(k))/dk}{\text{Var}(\hat{\beta}_L^4)}$	$\frac{d\text{Var}(\hat{\beta}_R^5(k))/dk}{\text{Var}(\hat{\beta}_L^5)}$
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 ( $d\hat{\beta}_R^i/dk$ ) relative to  $\hat{\beta}_L^i$ , namely,  $\{\phi(\hat{\beta}_i) = [\text{Var}(d\hat{\beta}_R^i/dk)/\text{Var}(\hat{\beta}_L^i)]^{1/2}; 1 \leq i \leq 5\}$ , independent of  $\mathbf{Y}$ , for solutions in the natural variables at selected  $k$

$k$	$\phi(\hat{\beta}_1)$	$\phi(\hat{\beta}_2)$	$\phi(\hat{\beta}_3)$	$\phi(\hat{\beta}_4)$	$\phi(\hat{\beta}_5)$
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

$\{|TE_i(k)|; 1 \leq i \leq 5\}$ , diminishing as  $k$  increases, characterize enhanced stability by each criterion. There is a plethora of entries; nonetheless, some consensus might emerge on seeking  $k^\dagger(\delta)$  such that all entries across columns are dominated by a given threshold value  $\delta > 0$ . Specifically, take

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

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

**4.4. Polynomial Representations**

As noted by Gibbons and 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  $\{k \rightarrow \mathbf{A}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p)^{-1}\mathbf{Z}'\mathbf{Y}_0; 1 \leq i \leq p\}$  which

**Table 8**

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

Threshold values	Table 5	Table 6	Table 7
$\{ TE_i(k)  < \delta; 1 \leq i \leq 5\}$	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  $\{\hat{\beta}_R^i(k)/|\hat{\beta}_L^i| = P_i^0(k)/Q_0(k); 1 \leq i \leq 5\}$

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

computes the ridge estimators in correlation units and transforms them back into natural units by  $A$ . The rational representation of  $\hat{\beta}_R^i(k)$  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  $R_i(k) \rightarrow 0$  as  $k \rightarrow \infty$ . To assist in displaying the coefficients, we again scale the ridge traces as  $\hat{\beta}_R^i(k)/|\hat{\beta}_L^i| = R_i(k)/|\hat{\beta}_L^i| = [P_i(k)/|P_i(0)|]/[Q(k)/|Q(0)|] \equiv P_i^0(k)/Q_0(k)$ . The coefficients for  $P_i^0(k)$  and  $Q_0(k)$ , to be denoted as  $\{c_0, \dots, c_4\}$  and  $\{c_0, \dots, c_5\}$ , are shown in Tables 9 and 10, respectively. Since all the coefficients of  $Q_0(k)$  are positive,  $Q_0(k) > 0$  for the ridge values  $k > 0$ . 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,  $\{\hat{\beta}_R^2(k), \hat{\beta}_R^3(k)\}$  have no sign changes, while  $\{\hat{\beta}_R^1(k), \hat{\beta}_R^4(k), \hat{\beta}_R^5(k)\}$  each have one sign change.

The foregoing representations support explicit rational functions for the derivative traces. For the scaled ridge traces we set  $\{\frac{d}{dk}(\hat{\beta}_R^i(k)/|\hat{\beta}_L^i|) = T_i^0(k)/[Q_0(k)]^2; 1 \leq i \leq p\}$  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  $\{d_0, \dots, d_8\}$ , 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,  $\#(\text{zeroes})$ , 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  $\hat{\beta}_R^i(k) \rightarrow 0$  as  $k \rightarrow \infty$ , the number of critical values of  $\hat{\beta}_R^i(k)$  provide an upper bound for the number of sign changes for  $\hat{\beta}_R^i(k)$ .

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

**Table 10**  
Coefficients for the normalized functions  $\{P_i^0(k); 1 \leq i \leq 5\}$  for the scaled ridge traces  $\{\hat{\beta}_R^i(k)/|\hat{\beta}_L^i| = P_i^0(k)/Q_0(k); 1 \leq i \leq 5\}$

$c$	$P_1^0(k)$	$P_2^0(k)$	$P_3^0(k)$	$P_4^0(k)$	$P_5^0(k)$
$c_0$	-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  $\{\frac{d}{dk}\hat{\beta}_R^i(k)/|\hat{\beta}_L^i| = T_i^0(k)/[Q_0(k)]^2; 1 \leq i \leq 5\}$

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}

#### 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,  $DF_k = \text{tr}(\mathbf{H}_k)$  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 Fig. 8.4, Myers concludes: “Certainly  $\text{tr}(\mathbf{H}_k)$  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  $\text{tr}(\mathbf{H}_k)$  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  $[\hat{\beta}_R^1(k), \hat{\beta}_R^2(k), \hat{\beta}_R^3(k), \hat{\beta}_R^4(k)]$  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  $\{[\hat{\beta}_R^i(k_u) - \hat{\beta}_R^i(k_v)]/(k_u - k_v); 1 \leq i \leq 4\}$ , 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, 1970a,b) 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  $Y$ . On the other hand, since ill conditioning resides exclusively in the matrix  $X$ , it is natural to conjecture that critical properties, such as stability, might trace back to  $X$  alone, independently of  $Y$ . 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  $\sigma^2$ , 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, 1970a,b). 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.

## Acknowledgments

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

## References

- Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* 16:125–127.
- Belsley, D. A. (1986). Centering, the constant, first-differencing, and assessing conditioning. In: Belsley, D. A. Kuh, E. ed., *Model Reliability*. Cambridge: MIT Press.

- Frank, I. E., Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* 35:109–135.
- Geladi, P. (2002). Some recent trends in the calibration literature. *Chemometry Intelligence Laboratory* 60:211–224.
- Gibbons, D. I., McDonald, G. C. (1984). A rational interpretation of the ridge trace. *Technometrics* 26:339–346.
- 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.
- Hoerl, A. E., Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55–67.
- Hoerl, A. E., Kennard, R. W. (1970b). Ridge regression: Application for nonorthogonal problems. *Technometrics* 12:69–82.
- Hoerl, A. E., Kennard, R. W., Baldwin, K. F. (1975). *Ridge regression: Some simulations. Communications in Statistics* 4:105–123.
- Jensen, D. R., Ramirez, D. E. (2010). Tracking MSE efficiencies in ridge regression. *Advances and Applications in Statistical Sciences* 1:381–398.
- Kalivas, J. H. (2005). Multivariate calibration, an overview. *Analysis Letters* 38:2259–2279.
- Kibria, G. K. (2003). Performance of some new ridge regression estimators. *Communications in Statistics Simulation* 32:419–435.
- Mallows, C. L. (1973). Some comments on Cp. *Technometrics* 15:661–675.
- Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 12:591–612.
- McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics* 1:93–100.
- Myers, R. H. (1990). *Classical and Modern Regression with Applications*. 2nd ed. Boston: PWS-KENT.
- Muniz, G., Kibra, B. M. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics Simulation* 38:621–630.
- Sundberg, R. (1999). Multivariate calibration – direct and indirect regression methodology. *Scandinavian Journal of Statistics* 26:161–207.
- 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, Blacksburg, VA.
- Zhang, R., McDonald, G. C. (2005). Characterization of ridge trace behavior. *Communications in Statistics–Theory Methods* 34:1487–1501.