- Home
- Local Control
- ICE Inference
- Shrinkage
- eBook
- Author Info

softRX freeware

Personal Computing Treatments for
Data Analysis Infirmities ...since 1983.

We champion modern statistical approaches that make realistic assumptions, use robust methods and present results graphically.

We provide statistical concept tutorials and freeware computing algorithms in several different languages that introduce and implement these exciting new approaches.

Risk Benefit Statistics LLC

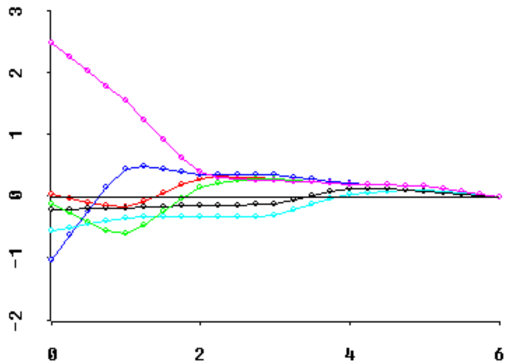
Bob Obenchain, PhD, FASA
13212 Griffin Run, Carmel, IN, 46033-8835
(317) 580-0144, wizbob at att dot net

Download: 1999 ASA Council of Chapters "Statistical Thinking" Software [1010K]

Shrinkage Regression Downloads

Intro to Regression Shrinkage Concepts

The TRACE display below shows how fitted regression coefficients for the infamous Longley(1967) dataset change due to shrinkage along a **Q-shape** = -1.5 path through 6-dimensional coefficient likelihood space.



Shrinkage methods can drastically reduce variability, but shrinkage also results in biased estimates. Since mean-squared-error risk is made up of variance plus squared-bias, shrinkage reduces risk whenever the unknown squared-bias introduced is less than the known reduction in (relative) variance.

To apply shrinkage methodology, the two key questions that a regression practitioner must answer are:

- **Which shrinkage path should I try?**
- **Where along that path should I stop shrinking?**

softRX freeware is proud to provide computer algorithms for R, XLisp-Stat, Stata, Gauss and SAS-IMSL to guide you along your shrinkage regression "journey" by providing powerful, maximum likelihood statistical inferences and dynamic graphical insights along the "route" of your choice! Related materials are...

- **MSE risk of Shrinkage**
- **Influential Observations and Shrinkage**
- **Shrinkage Equations and MCAL**
- **Literature References**

Nav: **Back**

softRX freeware

by Bob Obenchain, Ph.D.

Downloads for XLisp-Stat

Downloads for R, Stata, Gauss or SAS/IML

Download Stand-Alone Applications

Nav: **Back**

Nav: **Back**

Path Shapes: The traditional 2-Parameter Goldstein-Smith **(Q-shape, M-extent)** Paths based upon X-coordinate eigenvalues (Lambda spreads) are featured.
However, three "p-parameter" Shrinkage Path Shapes based upon X-coordinate principal correlations can also be specified: generalized GARROTE, generalized LASSO, and CROSR=Constant-Ratio-Of-Shrinkage-Ratios. The horizontal axis for TRACE displays remains **M-extent** of shrinkage.

XLisp-Stat RXridge Ver.2008 for Microsoft Windows [50k]

Download this zip archive and unpack it on a Windows machine to a sub-directory (SRX, say) of the directory containing WXLS32.EXE to install 1996 *.LSP code and numerical examples plus 2008 PDF documentation.

XLisp-Stat RXridge Ver.9605 for Unix [34k]

Download this rxridge.tar.gz archive to your UNIX machine, and enter the commands "gzip -df rxridge.tar.gz" then "tar -xvf rxridge.tar". This creates UNIX text files of *.LSP code, documentation, and numerical examples.

XLisp-Stat RXridge Ver.9605 for Macintosh [77k]

Download this .HQX file to your Mac and decode it using BinHex to create a self-extracting Stuffit archive named RXridge.sea. Then double click on this icon to unpack Mac text files containing *.LSP code, documentation, and numerical examples.

Extra RXridge Numerical Example Files [20k]

Download this zip archive and unpack it on a Windows machine for 9 additional ill-conditioned datasets formatted for RXridge in XlispStat.

Nav: **Back**

Nav: **Back**

RXshrink_1.0-2 "Zip" Package (Oct 2005) for Windows R 2.7+ [282k]

RXshrink_1.0-2 "Tar.Gz" Package (Oct 2005) for Linux R 2.7+ [167k]

Download a **R Package** installation archive for **maximum likelihood shrinkage** via "generalized (2-parameter) ridge" or "least angle" regression.

RXridge Ver.2008 for Stata (tm) 11.0+ [233K]

Download a Win-Zip archive of *.ADO functions for ridge/shrinkage calculations in Stata version 11.0+ ...with numerical example *.DO files, *.PDF documentation with embedded fonts, and *.TXT installation instructions.

RXridge Ver.9411 for SAS/IML (tm) [36k]

Download a self-extracting, MS-DOS archive of source code for shrinkage/ridge calculations in SAS/IML ...plus ASCII text documentation and numerical examples.

RXridge Ver.2008 for GAUSS (tm) [36k]

Download a self-extracting, MS-DOS archive of source code for shrinkage/ridge calculations in GAUSS ...plus ASCII text documentation and numerical examples.

Nav: **Back**

Nav: Back

PC Stand-Alone Systems

RXridge for Windows: Ver.9803 [899k]

Download a WinZip self-installing archive (srx_9803.exe) containing a stand-alone windows application for shrinkage/ridge calculations and high-resolution graphics. Somewhat less functionality (no VRR or Inference Intervals) than Version 9605 for MS-DOS, but this is a true 32-bit Microsoft Windows application. Includes royalty-free run-time Graphics Server® library modules.

RXridge MS-DOS Archive: Ver.9605 [162k]

Download a self-extracting archive containing the stand-alone MS-DOS RXridge.EXE system for shrinkage/ridge calculations and interactive CGA graphics ...plus ASCII text documentation and numerical examples.

RXtrace Ver.9603 MS-DOS Archive [70k]

Download a self-extracting archive containing the stand-alone MS-DOS RXtrace.EXE system for interactive re-display of TRACE results from RXridge.EXE ...plus ASCII text documentation and numerical examples.

PathProj Ver.9603 MS-DOS Archive [76k]

Download a self-extracting archive containing the stand-alone MS-DOS PathProj.EXE system for interactive display of the PROJECTION of the shrinkage path onto any 2-dimensional subspace ...plus ASCII text documentation and numerical examples.

RXmsesim Ver.9511 MS-DOS Archive [79k]

Download a self-extracting archive containing the stand-alone MS-DOS RXmsesim.exe system for simulation of MSE risk profiles associated with shrinkage rules ...plus ASCII text documentation and numerical examples.

ARXfiles MS-DOS Archive [22k]

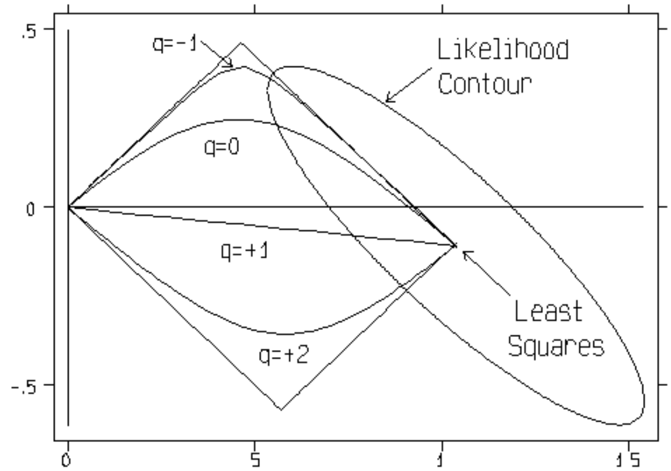
Download a self-extracting archive of batch-input (*.ARX) files for well known numerical examples of ill-conditioned regression models.

Nav: Back

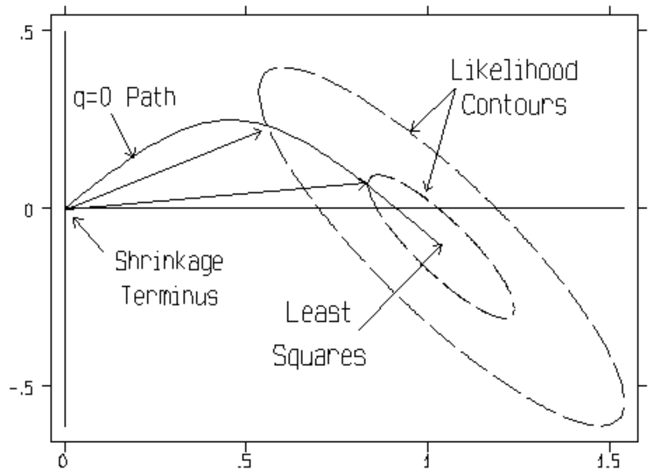
Nav: **Back | FAQ | Refs**

Shrinkage Regression Path Q-shapes

The display below shows a variety of shrinkage path **Q-shapes** for the rank(X) = p = 2 case.



The best known special case of a Q-shaped path is probably Q = 0 for Hoerl-Kennard(1970) **"ordinary" ridge regression**. This path has a dual "characteristic property," illustrated in the figure below. Namely, the Q = 0 path contains not only the shortest beta estimate vector of any given likelihood but also the most likely beta estimate of any given length.



Another well known special case of a Q-shaped path is Q = +1 for **uniform** shrinkage. The coefficient TRACE and shrinkage factor TRACE for this path are both rather "dull," but the estimated risk and inferior direction TRACES can still be quite interesting even when Q = +1.

An extremely important limiting case is Q = minus infinity for **principal components regression**. [Marquardt(1970) calls this limit "assigned rank" regression.] My experience is that the Q = -5 path is frequently quite close, numerically, to this limiting case. Note in the top figure on this page that the path with Q = -1 shape is already near the limit in the p = 2 dimensional case being depicted here.

As a general rule-of-thumb, paths with Q-shapes in the [-1,+2] range generally tend to be fairly **smooth** ...i.e. have "rounded" corners. Paths with Q-shapes greater that +2 or less than -1 can display quite "sharp" corners. In fact, the paths with limiting shapes of +/- infinity are actually linear splines with join points at integer MCAL values!

My computing algorithms provide strong, objective guidance on the choice of the Q-shape that is best for your data. For example, they implement the methods of Obenchain(1975b, 1981, 1997a) to identify the path Q-shape (and the MCAL-extent of shrinkage along that path) which have **maximum likelihood** (under a classical, fixed coefficient, normal-theory model) of achieving overall minimum MSE risk in estimation of regression coefficients.

Nav: **Back | FAQ | Refs**

Nav: **Back** | **FAQ** | **Refs**

"Stopping Rules" for Shrinkage

Once I start shrinking, where do I stop???

Yes, this may well be **the** question!!! This is probably why shrinkage methods in regression have been considered highly controversial for much of the last 40 years. A brief review of the history of **ridge regression** may reveal some potential "root causes" for this controversy...

Relatively widespread interest in ridge regression was initially sparked by Hoerl and Kennard(1970) when they suggested plotting the elements of their shrinkage estimator of regression beta-coefficients in a graphical display called the **RIDGE TRACE.** They observed that the relative magnitudes of the fitted coefficients tend to stabilize as shrinkage occurs. And they over-optimistically implied that it is "easy" to pick an extent of shrinkage, via visual examination of that coefficient trace, that achieves lower MSE risk than least squares.

Today, we **know** it just ain't easy. Even the relatively "tiny" amount of shrinkage that results using James-Stein-like estimators [Strawderman(1978), Casella(1981,1985)] when R-squared is large (greater than .8, say) is already "too much" to allow them to dominate least squares in any matrix-valued MSE risk sense. In other words, least squares is **admissible** in all meaningful (multivariate) senses [Brown(1975), Bunke(1975).]

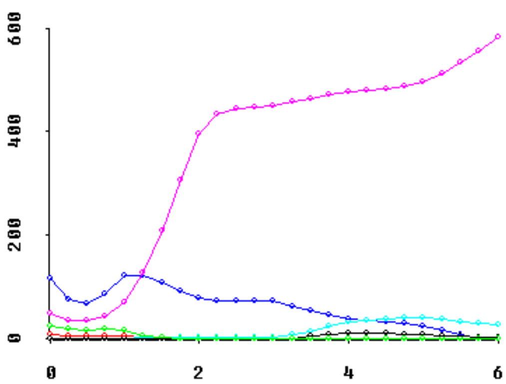
Ok, so Hoerl-Kennard were wrong about **guaranteed** risk improvements. But, in the eyes of their major critics, this probably wasn't their **BIG** mistake! No, they were unabashedly telling regression practitioners to... **LOOK AT THEIR DATA** (via that trace display) ...before subjectively "picking" a solution. And all "purists" certainly consider any tactic like this a real **NO! NO!**

In all fairness, almost **everybody** does this sort of thing in one way or another. Regression practitioners are constantly being encouraged to actively explore many different potential models for their data. Some of these alternatives change the functional form of the model, drop relatively uninteresting variables, or set-aside relatively influential observations. But who tells those practitioners that it can be misleading to simply report the least squares estimates and confidence intervals for that final model as if it were the **only model** they ever even considered?

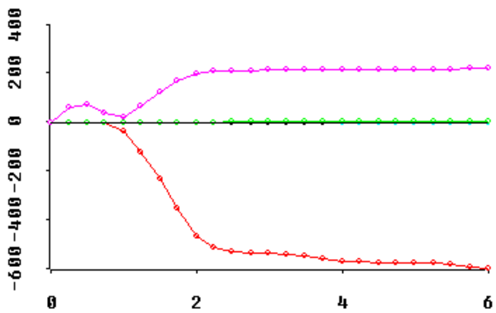
In other words, shrinkage/ridge methods have served as a convenient "whipping-boy" for all sorts of statistical practices that are questionable simply because they are based mainly upon heuristics.

My implementations of shrinkage/ridge regression algorithms skirt the above somewhat delicate issues by providing theoretically sound and **objective** (rather than subjective) criteria for deciding which path to follow and where to stop shrinking along that path. My algorithms use a normal-theory, **maximum likelihood** formulation to quantify the effects of shrinkage. Simulation studies suggest that my 2-parameter (Q-shape & M-extent) approach can work very, very well in practice [Gibbons(1981).]

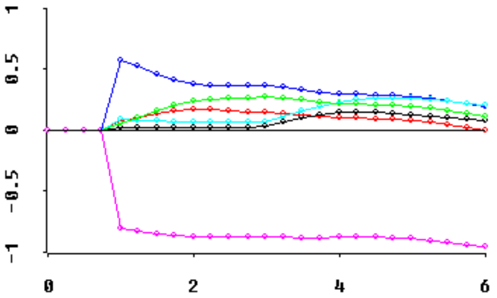
My shrinkage algorithms also display a **wide spectrum** of "trace" visualizations. For example, they display traces of scaled (relative) MSE risk estimates for individual coefficients...



traces of "excess eigen-value" estimates (with at most one negative estimate)...



traces of "inferior direction cosine" estimates (for the one negative eigen-value above)...



and even traces of multiplicative shrinkage factors (not shown) as well as the more traditional traces of shrunken coefficients.

Note that the horizontal axis for all of these traces is the **Multicollinearity Allowance** parameter, 0 <= MCAL <= p. This MCAL can usually be interpreted as the approximate rank deficiency in the predictor variables X-matrix. Displays of fitted coefficients in least angle regression (LAR) use a horizontal axis scaling equivalent to MCAL!

Why should these trace displays be of interest to YOU? Because... **OH!, WHAT A DATA ANALYTIC STORY THEY CAN TELL!** In Obenchain(1980), I called this the **"see power" of shrinkage in regression**.

For explanations of (and interpretations for) these sorts of traces, see Obenchain(1984, 1995); for the underlying maximum-likelihood estimation theory, see Obenchain(1978).

Nav: **Back** | **FAQ** | **Refs**

***Nav:* [Back]**

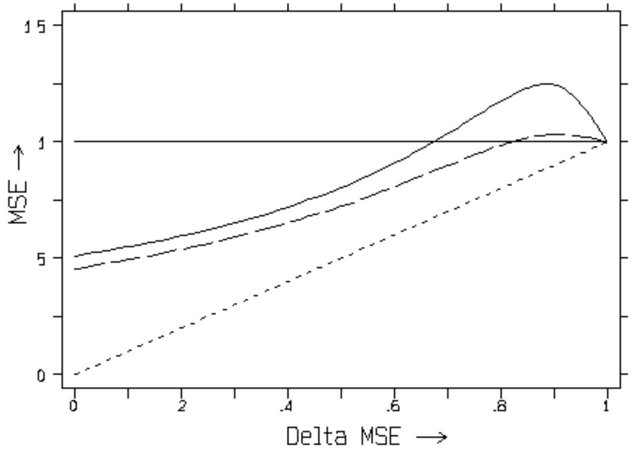
The MSE Risk of Shrinkage in Regression

What does a **Risk Profile** for Maximum Likelihood Shrinkage typically look like?

The figure below is a plot of simulated Mean-Squared-Error RISK profiles for two maximum likelihood estimators over the full range of potential, optimal extents of shrinkage. In other words, the plot below depicts situations where the unknown, true **MSE optimal shrinkage factor** for a single regression coefficient ranges all of the way from 0 to 1.

Shrinking a coefficient all of the way to zero is optimal (Delta-MSE=0) when either the unknown, true value of that coefficient is zero or else in the limit as the error variance approaches +infinity. This is the LEFT-hand extreme in the plot below.

No shrinking at all is optimal (Delta-MSE=1) when either the error variance is zero or else in the limit as the unknown, true value of that coefficient approaches +infinity or -infinity. This is the RIGHT-hand extreme in the plot below.



The horizontal line in the above plot (constant risk normalized to equal one) represents the MSE of ordinary least squares (OLS). There is no real need to simulate a risk profile for OLS; OLS is well known to be the essentially unique, minimax rule.

The diagonal, dotted line in the above plot (from Risk=0 at Delta-MSE=0 to Risk=1 at Delta-MSE=1) represents the theoretical lower bound on MSE that would result if it were possible to always apply the exactly correct extent of shrinkage.

Of the two curves in the above plot, the top, solid curve is for the unrestricted maximum likelihood shrinkage of a single coefficient under normal distribution theory when the degrees-of-freedom for error equal 5. Note that MSE may decrease here by as much as 50% but never increases by more than 25%.

The lower, dashed curve in the above plot is for maximum likelihood, equal shrinkage of a pair of coefficients in the "null" case where their unknown, true values actually are equal. This second curve is also simulated under normal distribution theory when the degrees-of- freedom-for-error equal 5. Here MSE may decrease by more than 50% and yet never increases by more than only 6%.

See Obenchain(1997a) for more information about the above plot. Highly portable C-language source code and an MS-DOS executable for shrinkage risk simulation are also available for download from this web site.

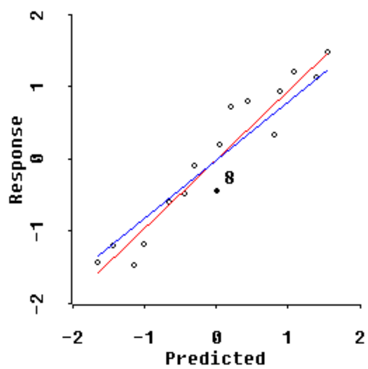
***Nav:* [Back]**

Nav: **Back | FAQ | Refs |**

Influential Observations in Shrinkage/Ridge Regression

The RXridge algorithms for XLISP-Stat display two very different types of plots that display the potential effects of INFLUENTIAL OBSERVATIONS on model fits. Specifically, an observation can be influential because it has an **outlying response value** or because it represents a **high leverage regressor combination** ...or even for both reasons!

The first type of influence plot shows the observed response values, Y, vertically against their "standardized" predicted values along the horizontal axis. Since predictions are always linear combinations of the given regressor coordinates, the horizontal axis is best viewed as giving coordinates for a single, standardized **composite regressor** variable, x-star, that depends only upon the ORIENTATION of the shrinkage regression beta-star vector in p-dimensional space. The LENGTH of the shrinkage beta-star vector determines only the slope of the line on the Y versus x-star plot that represents the shrinkage fit.



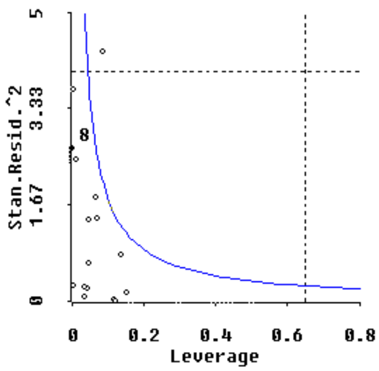
The plot above corresponds to rather extreme shrinkage of the Longley data (p=6) to MCAL=5 along the Q= -1.5 path. The BLUE line represents this shrinkage fit while the RED line shows the "Visual Re-Regression" of Y onto the standardized x-star coordinates. Since the RED line is a clearly better fit here than the BLUE line, we see that this MCAL=5 extent of shrinkage is excessive.

The user of RXridge.LSP can use the MCAL **slider** control to reduce the shrinkage extent back to the MCAL=1.0 to 1.33 range to verify that the BLUE Q-shape= -1.5 fit is virtually identical to the RED VRR fit in this range.

Outliers show in this plot as large **residuals** ...i.e. these response Y values represent relatively large deviations from the fitted BLUE shrinkage line.

And the points with highest leverage along the 1-dimensional, composite x-star axis are the points toward the extreme left-hand and right-hand ends of the plot. Unfortunately, considerable information can be lost in attempting to display p-dimensional leverage information in one dimension. Anyway, these x-star axis leverages can be somewhat misleading. So, **linked** to the first plot, RXridge.LSP also displays a second plot of standardized residuals and p-dimensional leverages!

This second plot shows squared, standardized residuals (i.e. corrected for any differences in variance), vertically, against its p-dimensional regressor leverage ratio (prediction variance divided by residual variance) along the horizontal axis.



The Cook(1977) measure of overall influence for each observation is proportion to the **product** of its squared, standardized residual times its leverage ratio. Each contour of constant overall influence thus display as a **hyperbola** on our second type of plot. And this hyperbola can be **moved** up and down using the overall "influence" slider control.

Nav: **Back | FAQ | Refs |**

Nav: **Back** | **FAQ** | **Refs**

Formulas for Generalized Shrinkage Estimators

The following formulas describe the FORM and EXTENT of shrinkage yielding 2-parameter generalized ridge regression estimators.

$$\beta^* = [X'X + k \cdot (X'X)^Q]^{-1} X'y$$

Our first formula, above, represents the 2-parameter family using notation like that of Goldstein and Smith(1974). Here we have assumed that the response vector, y, and all p columns of the (nonconstant) regressors matrix, X, have been "centered" by subtracting off observed mean value from each of the n observations. Thus Rank(X) = r can exceed neither p nor (n-1).

Insight into the form of the shrinkage path that results as k increases (from zero to infinity) for a fixed value of Q is provided by the "singular value decomposition" of the regressor X matrix and the corresponding "eigenvalue decomposition" of X'X.

$$X = H\Lambda^{+1/2}G'$$

$$(X'X)^Q = G\Lambda^Q G'$$

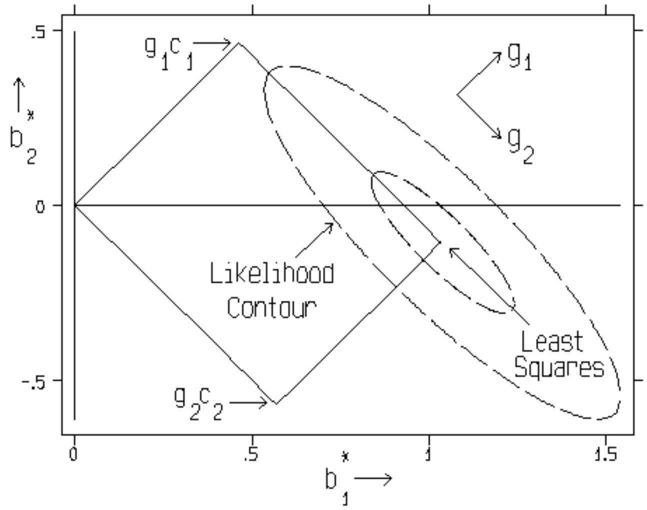
The H matrix above of "principal axis regressor coordinates" is (n by r) and semi-orthogonal (H'H = I.) And the G matrix of "principal axis direction cosines" is (p by r) and semi-orthogonal (G'G = I.) In the full-column-rank case (r = p), G is orthogonal; i.e. GG' is then also an identity matrix.

The (r by r) diagonal "Lambda" matrix above contains the **ordered** and **strictly positive** eigenvalues of X'X; Lambda(1) >= ... >= Lambda(r) > 0. Thus our operational rule for determining the Q-th power of X'X (where Q may not be an integer) will simply be to raise all of the positive eigenvalues of X'X to the Q-th power, pre-multiply by G, and post-multiply by G'.

Taken together, these decompositions allow us to recognize the above 2-parameter (k and Q) family of shrinkage estimators, beta-star, as being a special case of r-dimensional generalized ridge regression...

$$\beta^* = G\Delta\Lambda^{-1/2}H'y = G\Delta c$$

where the (r by r) diagonal "Delta" matrix contains the multiplicative **shrinkage factors** along the r principal axes of X. Each of these Delta(i) factors range from 0 to 1 (i = 1, 2, ..., r.) And the (r by 1) column vector, c, contains the **uncorrelated components** of the ordinary least squares estimate, beta-hat = G c, of the true regression coefficient beta vector. The variance matrix of c is the diagonal Lambda-inverse matrix times the scalar value of the error sigma-square.



In fact, we now see that the 2-parameter family of shrinkage estimators from our first equation, above, is the special case of the last equation in which...

$$\delta_i = \frac{\lambda_i}{(\lambda_i + k \cdot \lambda_i^Q)} = \frac{1}{(1 + k \cdot \lambda_i^{Q-1})}$$

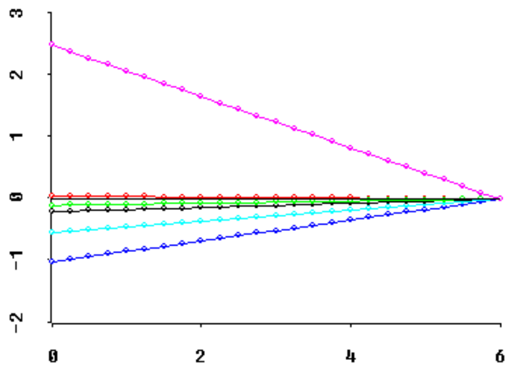
Actually, the "k" parameter is not a very good measure of the **extent** of shrinkage in the sense that the sizes of all r shrinkage factors, Delta, can depend more on one's choice of Q than on one's choice of k. Specifically, the k-values corresponding to two different choices of Q are usually **not** comparable.

Thus my algorithms use the m = MCAL = "multicollinearity allowance" parameter of Obenchain and Vinod(1974) to index the **M-extent of Shrinkage** along paths. This parameter is defined as follows...

$$MCAL = p - \delta_1 - \delta_2 - \ldots - \delta_p = rank(X) - trace(\Delta)$$

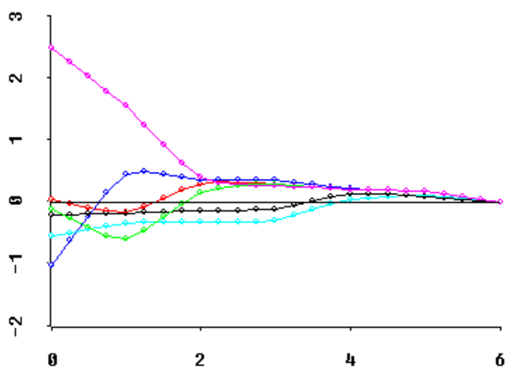
Note that the range of MCAL is finite; MCAL ranges from 0 to r=Rank(X), inclusive. Whatever may be your choice of Q-shape, the OLS solution always occurs at the beginning of the shrinkage path at MCAL=0 (k=0 and D=I) and the terminus of the shrinkage path, where the fitted regression hyperplane becomes "horizontal" (slope=0 in all p-directions of X space) and y-hat = y-bar, always occurs at MCAL = r (k = +infinity and D = 0). RXridge.LSP uses Newtonian descent methods to compute the numerical value of k corresponding to given values of MCAL and Q-shape.

In addition to having finite (rather than infinite) range, MCAL has a large number of other advantages over k when used as the scaling for the horizontal axis of ridge TRACE displays. For example, shrunken regression coefficients with stable relative magnitudes form **straight lines** when plotted versus MCAL.



Similarly, the average value of all r shrinkage factors is (r-MCAL)/r, which is the Theil(1963) proportion of Bayesian posterior precision due to sample information (rather than to prior information.) And this proportion decreases linearly as MCAL increases.

Perhaps most importantly, MCAL can frequently be interpreted as the **approximate deficiency in the rank of X.** For example, if a regressor X'X matrix has only two relatively small eigenvalues, then the coefficient ridge trace of best Q-shape typically "stabilizes" at about MCAL = 2. I.E., the coefficient trace then consists primarily of fairly straight lines between MCAL = 2 and MCAL = r = 6 in the graphic below.



Nav: **Back** | **FAQ** | **Refs**

Nav: **Back**

Shrinkage Regression REFERENCES

Brown, L. (1975). "Estimation with incompletely specified loss functions (the case of several location parameters.)" **Journal American Statistical Association** 70, 417-427.

Breiman, L. (1995). "Better subset regression using the non-negative garrote." **Technometrics** 37, 373-384.

Bunke, 0. (1975). "Least squares estimators as robust and minimax estimators." **Math. Operations forsch u. Statist.** 6, 687-688.

Burr, TL and Fry, HA. (2005). "Biased regression: the case for cautious application." **Technometrics** 47, 284-296. **[**Includes references to material on this website.]

Casella, G. (1980). "Minimax ridge regression estimation." **Annals of Statistics** 8, 1036-1056.

Casella, G. (1985). "Condition numbers and minimax ridge-regression estimators." **Journal American Statistical Association** 80, 753-758.

Chatterjee, S and Hadi, A. S. (1988). **Sensitivity Analysis in Regression.** New York: John Wiley.

Cook, R. D. and Weisberg, S. (1994). **Introduction to Regression Graphics.** New York: John Wiley.

Cook, R. D. (1977). "Detection of influential observations in linear regression." **Technometrics** 19, 15-18.

Croghan, T.W., Obenchain. R.L. and Crown, W.E. (1998). "What does treatment of depression really cost?" **Health Affairs** 17(4); 198-208. **View/Download Paper**

Computation of marginal effects from an ill-conditioned model of log(cost.) **View/Download Appendix** Technical details of shrinkage and smearing methods.

Efron B. and Morris, C. N. (1976). "Discussion" (of Dempster, Schatzoff and Wermuth.) **Journal American Statistical Association** 72, 91-93.

Efron B, Hastie T, Johnstone I, Tibshirani R. (2004). "Least Angle Regression." **Annals of Statistics** 32: 407-499 (with discussion.)

Frank, I. E. and Freidman, J. H. (1993). "A statistical view of some chemometrics regression tools." (with discussion) **Technometrics** 35, 109-148.

Fu, W. J. (1997). "Penalized Regressions: the Bridge versus the Lasso." One of 4 winners in the 1997 student paper competition sponsored by the **ASA Statistical Computing Section**.

Wenjiang's thesis under Rob Tibshani at the University of Toronto is entitled "A Statistical Shrinkage Model and Its Applications."

Gibbons, D. G. (1981). "A simulation study of some ridge estimators." **Journal of the American Statistical Association** 76, 131-139.

Goldstein M. and Smith, A. F. M (1974). "Ridge-type estimators for regression analysis." **Journal of the Royal Statistical Society B** 36, 284-291.

Golub, G. H., Heath, M., and Wahba, G. (1979). "Generalized cross- validation as a method for choosing a good ridge parameter." **Technometrics** 21, 215-223.

Gruber, M. H. J. (1998). **Improving Efficiency by Shrinkage:** The James-Stein and Ridge Regression Estimators. New York: Marcel Dekker.

Hoerl, A. E. and Kennard, R. W. (1970a). "Ridge regression: biased estimation for nonorthogonal problems." **Technometrics** 12, 55-67.

Hoerl, A. E. and Kennard, R. W. (1970b). "Ridge regression: applications to nonorthogonal problems." **Technometrics** 12, 69-82.

LeBlanc, M. and Tibshirani, R. (1998). "Monotone shrinkage of trees." **Journal of Computational and Graphical Statistics** 7, 417-433.

Longley, J. W. (1967). "An appraisal of least squares programs for the electronic computer from the point of view of the user." **Journal of the American Statistical Association** 62, 819-841.

Mallows, C. L. (1973). "Some comments on Cp." **Technometrics** 15, 661-677.

Mallows, C. L. (1995). "More comments on Cp." **Technometrics** 37, 362-372.

Marquardt, D. W. (1970). "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation." **Technometrics** 12, 591-612.

Marquardt, D. W. (1980). "Comment: You should standardize the predictor variables in your regression models." **Journal of the American Statistical Association** 75, 87-91.

Massy, W. F. (1965). "Principal components regression in exploratory statistical research." **Journal American Statistical Association** 60, 234-256.

Nash, J. C. (1992). "Statistical shareware: illustrations from regression techniques." **The American Statistician** 46, 312-318.

Nash reviewed PC software that included softRX freeware stand-alone (DOS) applications.

Obenchain, R. L. and Vinod, H. (1974). "Estimates of partial derivatives from ridge regression on ill-conditioned data." NBER-NSF Seminar on Bayesian

Inference in Econometrics, Ann-Arbor, Michigan.

Obenchain, R. L. (1975a). "Residual optimality: ordinary vs. weighted vs. biased least squares." **Journal of the American Statistical Association** 70, 375-379.

> Ordinary least squares and diagonally-weighted least squares always produce residuals that are optimal estimators of lack-of-fit ...even when the expectation model and/or dispersion model you have specified are/is wrong!

Obenchain, R. L. (1975b). "Ridge analysis following a preliminary test of the shrunken hypothesis" (with discussion.) **Technometrics** 17, 431-445.

> Classical, normal-theory monitoring of the likelihood-of-MSE-optimality along ANY shrinkage path.

Obenchain, R. L. (1977). "Classical F-tests and confidence regions for ridge regression." **Technometrics** 19, 429-439.

> See this paper and/or my Frequently-Asked-Questions page for information on confidence intervals in shrinkage regression.

Obenchain, R. L. (1978). "Good and optimal ridge estimators." **Annals of Statistics** 6, 1111-1121.

> The "ridge-function" theorem; maximum likelihood estimation of the "inferior direction" and of scaled MSE risk along any direction in p-dimensional regression coefficient space; MSE risk optimality of 0 or X'y along directions orthogonal to or parallel to beta, respectively.

Obenchain, R. L. (1980). Comment on "A critique of some ridge regression methods." **Journal of the American Statistical Association** 75, 95-96.

Obenchain, R. L. (1981). "Maximum likelihood ridge regression and the shrinkage pattern alternatives." **I.M.S. Bulletin** 10, 37; Absract 81t-23. (67 page review article.)

Obenchain, R. L. (1984). "Maximum likelihood ridge displays." **Communications in Statistics A** 13, 227-240.

> Proceedings of the Fordham Ridge Symposium, ed. H. D. Vinod; illustrations of usage of RXridge SAS/IML freeware.

Obenchain, R. L. (1995). "Maximum likelihood ridge regression." **Stata Technical Bulletin** 28, 22-36.

> Introduction to basic shrinkage/ridge concepts; illustrations of usage of RXridge .ADO freeware.

Obenchain, R. L. (1997a). "Maximum likelihood shrinkage in regression."

> Closed form expressions for classical, fixed coefficient, maximum likelihood estimation within the 2-parameter shrinkage family, and some simulated MSE risk profiles.

Obenchain, R. L. (1997b). "Influential observations in ridge regression."

> Exposition of basic "Visual Re-Regression" concepts.

Obenchain, R. L. (1997c). **Shrinkage Regression: ridge, BLUP, Bayes and Stein.** Preliminary draft (200+ pages.)

Obenchain, R. L. (2005). **RXshrink:** an R package for maximum likelihood shrinkage in generalized (2-parameter) ridge regression or least angle regression (LAR). **www.r-project.org**

Shumway, R. H. (1982). "Maximum likelihood estimation of the ridge parameter in linear regression." Technical Report, Department of Statistics, University of California at Davis.

Stein, C. M. (1956). "Inadmissibility of the usual estimator of the mean of a multivariate normal distribution." **Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Pobability** 1, 197-206. University of California Press.

Strawderman, W. E. (1978). "Minimax adaptive generalized ridge regression estimators." **Journal American Statistical Association** 73, 623-627.

Theil, H. (1963). "On the use of incomplete prior information in regression analysis." **Journal of the American Statistical Association** 58, 401-414.

Thisted, R. (1976). "Ridge regression, minimax estimation, and empirical bayes methods." Technical Report No. 28, Division of Biostatistics, Stanford University.

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso." **Journal of the Royal Statistical Society B** 58, 267-288.

Tierney, Luke. (1990). **LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics.** New York: John Wiley and Sons.

Tukey, J. W. (1975). "Instead of Gauss-Markov Least Squares; What?" **Applied Statistics,** ed. R. P. Gupta. Amsterdam-New York: North Holland Publishing Company.

Vinod, H. D. and Ullah, A. (1981). **Recent Advances in Regression Methods.** New York: Marcel Dekker.

Vinod, H. D. (1995). "Double bootstrap for shrinkage estimators." **Journal of Econometrics** 68, 287-302.

Walter, Bernhard (1994). "XLisp-Stat code for shrinkage/ridge regression." Techniche Universitad Munich.

***Nav:* Back**

Frequently Asked "Pointed" Questions (FAQ)

"Isn't shrinkage in regression a dead topic? I haven't seen any new papers in years!!!"

I don't know of any recent papers critical of shrinkage methods. **Technometrics** continues to publish important articles on shrinkage in regression, such as Burr and Fry (2005). Of course, there's also the exciting “Least Angle Regression” work of Efron, Hastie, Johnstone and Tibshirani (2004) in **Annals of Statistics,** at least when their LAR beta vector ultimately becomes shorter than the vector of least squares estimates.

Frank and Friedman(1993) and Breiman(1995) express great confidence in "cross validation" methods in shrinkage estimation. Although Mallows(1995) observes that minimizing C-sub-p to pick a regressor subset can be misleading in situations that aren't "clear-cut," he apparently still recommends calculating C-sub-p while shrinking along "smooth" paths. See also: Tibshirani(1996), Fu(1997) and LeBlanc and Tibshirani (1998).

"Aren't some of the early shrinkage/ridge methods still considered rather controversial?"

In a word: Yes! But the great, subjective "passions" (both for and against ridge methods) of the 1970's are now muted if not forgotten. In my opinion, the keys to avoiding controversy are (1) to use statistical inference to decide how much shrinkage of what type to perform and (2) to be rather conservative as stressed by Burr and Fry (2005). For example, in my computing algorithms, maximum likelihood methods under normal distribution theory are stressed. See my "Shrinkage" pages for more details on this.

"How can I form confidence intervals for shrinkage estimates?"

A reasonable (and simple!) approach is to simply use classical confidence intervals, centered at least-squares estimates, computed using your favorite statistics package; see Obenchain(1977). In other words, even though point-estimates of effects change as shrinkage is imposed, there really is no basis in "classical" statistical theory for either shifting the location or changing the width of interval-estimates. In fact, a shrunken estimate can look quite different, numerically, from the least-squares solution without being significantly different, statistically. (Obviously, you don't want to shrink so much that your point estimate ends up OUTSIDE your reported interval!)

If you feel you ABSOLUTELY MUST have an interval centered near or at your shrunken estimate, you are going to have to use either bootstrap resampling, Vinod(1995), or Bayesian methods. Highest Posterior Density (HPD) intervals incorporate "added information" from your prior (centered at zero) to that from your sample. This characterizes your shrunken estimates as "unbiased" compromises between prior and sample information.

"How can a so-called OPTIMAL shrinkage estimator be inferior to a so-called GOOD shrinkage estimator?"

"Optimal" shrinkage estimators attempt to minimize a single (scalar valued) measure of overall MSE risk. "Good" shrinkage estimators are simply those that are better than Ordinary-Least-Squares (OLS) ...but they have to dominate OLS in EVERY (matrix valued) MSE sense. So good shrinkage estimators generally do much less shrinkage (are much closer to OLS, numerically) than optimal shrinkage estimators. In fact, a useful guideline is provided by the "2/p-ths rule-of-thumb," Obenchain(1978), where p=Rank(X). Namely, in terms of the MCAL measure of extent-of-shrinkage, the upper-limit on good shrinkage extents is only 2/p-ths of the extent of shrinkage most likely to the MSE optimal. For example, p=6 for the Longley data, and MCAL = 4 along the Q-shape = -1.5 path is most likely to be MSE optimal; thus good shrinkage estimates tend to be limited to MCAL of no more than 2*4/6 = 1.33 ...which is confirmed by the corresponding excess eigenvalue and inferior direction TRACES for the Q-shape = -1.5 path.

"Why not simply use either Stein-like or minimum-estimated-risk rules?

Minimax rules can tend to do so little shrinkage that they are almost indistinguishable, for all practical purposes, from least squares. Minimum estimated risk rules, like those of Mallows(1973), can shrink quite aggressively. This can lead to a big reduction in MSE risk in "favorable" cases, but aggressive shrinkage can also lead to even bigger MSE risk penalties in "unfavorable" cases. Maximum likelihood approaches represent some sort of "middle ground" between these "extremes." They reduce risk by only about 50% even in the most favorable cases ...where the risk could be reduced 100% by shrinkage all of the way to ZERO. But they also tend to increase MSE risk by at most 25% when truly unfavorable cases are encountered (i.e. when shrinkage factors in the .8 to .9 range are MSE optimal.) See Gibbons(1981) and Obenchain(1996) for more on this.