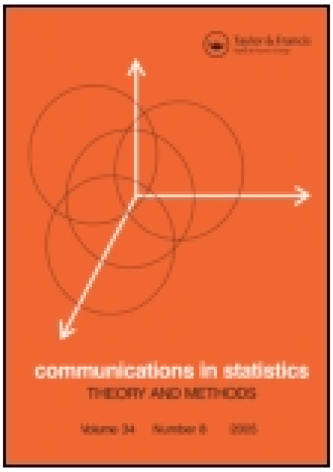

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

Ridge estimator in singulah oesiun with application to age-period-cohort analysis of disease rates

Wenjiang J. Fu PhD [a]
[a] Department of Epidemiology , Michigan State University , Michigan, 48823 E-mail:
Published online: 05 Jul 2007.

COMMUN. STATIST.—THEORY METH., 29(2), 263–278 (2000)

RIDGE ESTIMATOR IN SINGULAR DESIGN WITH APPLICATION TO AGE-PERIOD-COHORT ANALYSIS OF DISEASE RATES

Wenjiang J. Fu, PhD

Department of Epidemiology, Michigan State University
4660 South Hagadorn Road, Suite 600, East Lansing, Michigan 48823
fuw@msu.edu

Key Words: collinearity; identifiability; intrinsic estimator; shrinkage.

ABSTRACT

Ridge estimator of a singular design is considered for linear and generalized linear models. Ridge penalty helps determine a unique estimator in singular design. The tuning parameter of the penalty is selected via generalized cross-validation (GCV) method. It is proven that the ridge estimator lies in a special sub-parameter space and converges to the intrinsic estimator, an estimable function in singular design, as the shrinkage penalty diminishes. The expansion of the ridge estimator and its variance are also obtained. This

method is demonstrated through an application to age-period-cohort (APC) analysis of the incidence rates of cervical cancer in Ontario women 1960-1994.

1. INTRODUCTION

In linear regression

$$Y = X\beta + \varepsilon, \tag{1}$$

ridge estimator is defined as

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y \tag{2}$$

where I is an identity matrix, to minimize the objective function

$$(y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta$$

for $\lambda \geq 0$. It has wide range of applications to deal with non-orthogonal regression problems in applied sciences and engineering (Horle and Kennard (1970a,1970b), Frank and Friedman (1993)), where collinearity of the covariates frequently occurs due to a large number of covariates in the model. When collinearity occurs, ridge estimator shrinks the ordinary least-squares (OLS) estimator ($\lambda = 0$) towards the origin and yields biased estimator with smaller variance, thus potentially leads to better estimation with small mean squared error (MSE). These have been demonstrated in the literature (Frank and Friedman (1993), Hocking (1996), Fu (1998)). However, not much result has been obtained on ridge estimator in singular designs. In this paper, we study ridge estimator in a special singular design case, where the design matrix has one-less than full rank. We point out that the ridge estimator lies in a sub-parameter space orthogonal to the null space of the design matrix X generated by the eigenvector of eigenvalue 0. We further point out that the ridge estimator converges to the intrinsic estimator, an estimable function of the singular design recently discovered (Fu and Rohan (1999)), as the shrinkage parameter λ tends to 0. An expansion of the ridge estimator is obtained for small $\lambda > 0$. We demonstrate with an application to the age-period-cohort (APC) analysis of cancer incidence rates that the tuning parameter λ can be selected via the generalized cross-validation (GCV) method (Craven and Wahba (1979)), and the ridge penalty determines a unique estimator in a singular design and performs well in the APC analysis of disease rates.

2. RIDGE ESTIMATOR IN SINGULAR DESIGN

Ridge estimator has been well studied in dealing with collinearity in non-singular linear models. It performs well by setting a penalty on model parameters to shrink least squares estimator. Here I study ridge estimator in a singular design. I assume throughout the paper that the regression matrix X of dimension $n \times p$ with $p < n$ has a rank of $p - 1$. The main results are as follows.

Theorem 1. The p-dimensional parameter space $\mathcal{B}$ can be decomposed as

$$\mathcal{B} = \mathcal{N} \oplus \mathcal{V},$$

where $\mathcal{N}$ is the null space of X spanned by the eigenvector of the unique eigenvalue 0. The ridge estimator $\hat{\beta}_{ridge}(\lambda)$ with $\lambda > 0$ lies in the subparameter space $\mathcal{V}$.

Fu and Rohan (1999) studied the structure of the infinite number of estimators of model (1) and pointed out that all estimators can be expressed as

$$\hat{\beta} = B + tB_0, \tag{3}$$

where B_0 is the eigenvector of X with eigenvalue 0, and t is an arbitrary real number. B is an estimable function and is named to be the intrinsic estimator.

Here, we point out a relationship between the ridge estimator and the intrinsic estimator.

Theorem 2. The ridge estimator $\hat{\beta}_{ridge}(\lambda)$ converges to the intrinsic estimator B as λ tends to 0.

The proofs of Theorems 1 and 2 are in the appendix. Theorems 1 and 2 apply as well to the generalized linear models in a singular design, where the ridge estimator is defined by minimizing an objective function with model deviance Dev as follows.

$$\hat{\beta}_{ridge} = arg\min_{\beta}(Dev + \lambda\beta^T\beta). \tag{4}$$

In a non-singular design, the convergence of the ridge estimator to the OLS estimator as λ tends to 0 is trivial by definition (2). However, the convergence is not trivial in a singular design since the matrix $(X^TX + \lambda I)^{-1}$ in (2) does not converge. Even if the convergence of the ridge estimator is known, it is still intricate to identify the limit of the ridge estimator among the infinite number of estimators (3) with $\lambda = 0$. The fact that the ridge estimator lies in the subparameter space $\mathcal{V}$ helps determine the convergence and the limit.

As demonstrated (Fu and Rohan (1999)), the intrinsic estimator B is obtained through regression in a sub-parameter space orthogonal to the null space of design matrix X. It can also be obtained through principal components by eliminating eigenvalue 0. The convergence of the ridge estimator to the intrinsic estimator provides insights to the relationship between the ridge penalty and the principal component analysis, two different approaches to regressions with collinearity.

In the following, I provide an expansion of the ridge estimator for small tuning parameter $\lambda > 0$ and obtain the variance of the ridge estimator.

Assume that matrix X^TX has eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{p-1} > \lambda_p =$

0. Let Λ be the diagonal matrix $diag(\lambda_1, \ldots, \lambda_p)$ and their corresponding eigen-vectors form the orthonormal $p \times p$ matrix Q, i.e. $X^T X = Q \Lambda Q^T$. Let $\hat{\alpha}_p(\lambda)$ be the ridge estimator of design matrix $Z = XQ$ under the orthonormal transformation Q. Then we have

$$\hat{\boldsymbol{\alpha}}_p(\lambda) = (Z^T Z + \lambda I)^{-1} Z^T \boldsymbol{y} = (\Lambda + \lambda I)^{-1} Z^T \boldsymbol{y}.$$

The ridge estimator $\hat{\boldsymbol{\beta}}_{ridge}(\lambda)$ of (1) satisfies

$$\hat{\boldsymbol{\beta}}(\lambda) = Q \hat{\boldsymbol{\alpha}}(\lambda)$$

with the lower index $_{ridge}$ omitted for simplicity. Denote by Q_{p-1} the first $p-1$ columns of matrix Q, and by $\hat{\boldsymbol{\alpha}}_{p-1}(\lambda)$ the vector of the first $p-1$ coordinates of $\hat{\boldsymbol{\alpha}}(\lambda)$. Let $Z_{p-1} = XQ_{p-1}$ and $\Lambda_{p-1} = diag(\lambda_1, \ldots, \lambda_{p-1})$. By Theorem 1, $\hat{\boldsymbol{\beta}}(\lambda) \in \mathcal{V}$, the subparameter space spanned by the column vectors of Q_{p-1},

$$\hat{\boldsymbol{\beta}}(\lambda) = Q \hat{\boldsymbol{\alpha}}(\lambda) = Q \begin{pmatrix} \hat{\boldsymbol{\alpha}}_{p-1}(\lambda) \\ 0 \end{pmatrix} = Q_{p-1} \hat{\boldsymbol{\alpha}}_{p-1}(\lambda).$$

We take the expansion of $\hat{\boldsymbol{\alpha}}_{p-1}(\lambda)$ for small $\lambda > 0$.

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{p-1}(\lambda) &= (Z_{p-1}^T Z_{p-1} + \lambda I_{p-1})^{-1} Z_{p-1}^T \boldsymbol{y} \\ &= (\Lambda_{p-1} + \lambda I_{p-1})^{-1} Z_{p-1}^T \boldsymbol{y} \end{aligned}$$

It can be derived that

$$(\Lambda_{p-1} + \lambda I_{p-1})^{-1} = \Lambda_{p-1}^{-1} - \lambda \Lambda_{p-1}^{-2} + O(\lambda^2).$$

Denote two $p \times p$ matrices

$$\tilde{\Lambda} = \begin{pmatrix} \Lambda_{p-1}^{-1} & \\ & 0 \end{pmatrix} \quad \text{and} \quad \tilde{\Lambda}^2 = \begin{pmatrix} \Lambda_{p-1}^{-2} & \\ & 0 \end{pmatrix}$$

Hence

$$\begin{aligned}\hat{\beta}(\lambda) &= Q_{p-1}\left[\Lambda_{p-1}^{-1} - \lambda\Lambda_{p-1}^{-2} + O(\lambda^2)\right]Q_{p-1}^T X^T y \\ &= Q_{p-1}\Lambda_{p-1}^{-1}Q_{p-1}^T X^T y - \lambda Q_{p-1}\Lambda_{p-1}^{-2}Q_{p-1}^T X^T y + O(\lambda^2) \\ &= Q\tilde{\Lambda}Q^T X^T y - \lambda Q\tilde{\Lambda}^2 Q^T X^T y + O(\lambda^2) \\ &= B - \lambda Q\tilde{\Lambda}Q^T B + O(\lambda^2) \\ &= \left[I - \lambda Q\tilde{\Lambda}Q^T\right]B + O(\lambda^2)\end{aligned}$$

where the intrinsic estimator $B = Q\tilde{\Lambda}Q^T X^T y$. Consequently, the variance of the ridge estimator can be derived as

$$\begin{aligned}var(\hat{\beta}(\lambda)) &= \left[I - \lambda Q\tilde{\Lambda}Q^T\right]var(B)\left[I - \lambda Q\tilde{\Lambda}Q^T\right] + O(\lambda^2) \\ &= var(B) - \lambda\left[Q\tilde{\Lambda}Q^T var(B) + var(B)Q\tilde{\Lambda}Q^T\right] + O(\lambda^2)\end{aligned}$$

To compare the variance of the ridge estimator with the intrinsic estimator, we have the following results.

$$\hat{\beta}(\lambda) = UB$$

$$var(\hat{\beta}(\lambda)) = U\,var(B)\,U$$

where

$$U = I - \lambda Q\tilde{\Lambda}Q^T + \lambda^2 Q\tilde{\Lambda}^2 Q^T + O(\lambda^3).$$

Then it is easy to show via simple algebra that $var(B) - var(\hat{\beta}(\lambda))$ is positive definite for small $\lambda > 0$. Hence we conclude with the following.

Theorem 3. The ridge estimator has a smaller variance than the intrinsic estimator.

3. APPLICATION TO AGE-PERIOD-COHORT MODEL

As an application, we apply the ridge estimator to the multiple classification model in APC analysis (Kupper et al.(1983), Kupper et al.(1985)),

TABLE I
Incidence rates of cervical cancer in Ontario women 1960-1994 (per 10^5 person-year)

Age \ Year	60-64	65-69	70-74	75-79	80-84	85-89	90-94
20-24	3.89	3.24	2.90	2.05	2.19	1.76	1.73
25-29	16.01	11.18	8.92	9.74	8.48	7.43	7.54
30-34	26.02	21.14	16.23	15.84	14.54	13.67	12.71
35-39	38.84	25.09	21.07	18.74	18.80	18.04	18.18
40-44	47.65	32.50	22.71	20.01	18.78	16.19	18.12
45-49	51.48	36.69	22.15	19.20	17.74	17.29	18.31
50-54	49.12	37.26	25.51	18.41	16.66	15.41	14.07
55-59	51.48	40.87	34.70	21.83	16.97	17.69	13.73
60-64	47.68	42.80	29.76	22.71	20.16	17.69	16.94
65-69	40.44	39.17	31.44	28.79	23.35	19.26	19.16
70-74	42.40	35.32	27.78	24.31	20.27	20.19	14.95
75-79	42.44	36.68	28.75	25.22	21.17	21.08	19.43
80-84	41.50	29.74	31.54	22.31	20.04	15.25	21.28
85+	30.79	32.43	37.10	19.81	16.42	14.87	12.06

where the design matrix is one-less than full rank.

In an APC analysis, we study chronic disease rates, especially cancer incidence and mortality rates, to determine the effects of age, period and cohort, and to identify trends in disease rates. See (Clayton and Schifflers (1987), Holford (1983, 1991), Osmond (1982), Robertson and Boyle (1998), Tarone and Chu (1996)).

Table I shows an APC data set of cervical cancer incidence rates of Ontario women 1960-1994. It has 14 rows of age groups, 7 columns of period groups and 20 diagonals of birth cohorts.

We fit a log-linear model to the incidence rate.

$$log(\tau_{ij}) = \mu + \alpha_i + \beta_j + \gamma_k \tag{5}$$

where τ_{ij} is the expected rate in cell (i, j). $i = 1, \ldots, A$, $j = 1, \ldots, P$, $k = 1, \ldots, (A+P-1)$ with $A = 14$ and $P = 7$. μ is the intercept, α_i the i-th row age effect, β_j the j-th column period effect, γ_k the k-th diagonal cohort effect with $k = A - i + j$.

Model (5) can usually be solved after reparametrization to center the parameters as follows.

$$\mu^* = \mu + \bar{\alpha} + \bar{\beta} + \bar{\gamma}, \quad \alpha_i^* = \alpha_i - \bar{\alpha}, \quad \beta_j^* = \beta_j - \bar{\beta}, \quad \gamma_k^* = \gamma_k - \bar{\gamma}, \tag{6}$$

where $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$ are the averages over the corresponding parameters. The parameters can then be estimated by regressing on the design matrix. However, an apparent temporal relationship **birth = period - age** induces a singular design matrix and yields infinite number of estimators. It precludes the identification of the age, period and cohort effects. This is called the identifiability problem, see (Kupper et al. (1983), Kupper et al. (1985), Holford (1983, 1991), Clayton and Schifflers (1987)).

To determine a unique estimator, we apply ridge penalty to the model deviance (4) and achieve a ridge estimator. Numerically, the ridge estimator is obtained for each fixed $\lambda > 0$ by solving for the ridge estimator in the IRLS procedure for the log-linear model. The tuning parameter λ is selected through the GCV. It minimizes the GCV through a grid search over a range of $\lambda \geq 0$.

$$GCV(\lambda) = \frac{Dev(\lambda)}{n(1 - p(\lambda)/n)^2}$$

where $Dev(\lambda)$ is the model deviance, $p(\lambda) = trace(W(W^T W + \lambda I)^{-1} W^T)$ is the effective number of parameters of the penalty model, and W is the regression matrix for the adjusted variable in the last step of the IRLS procedure before convergence.

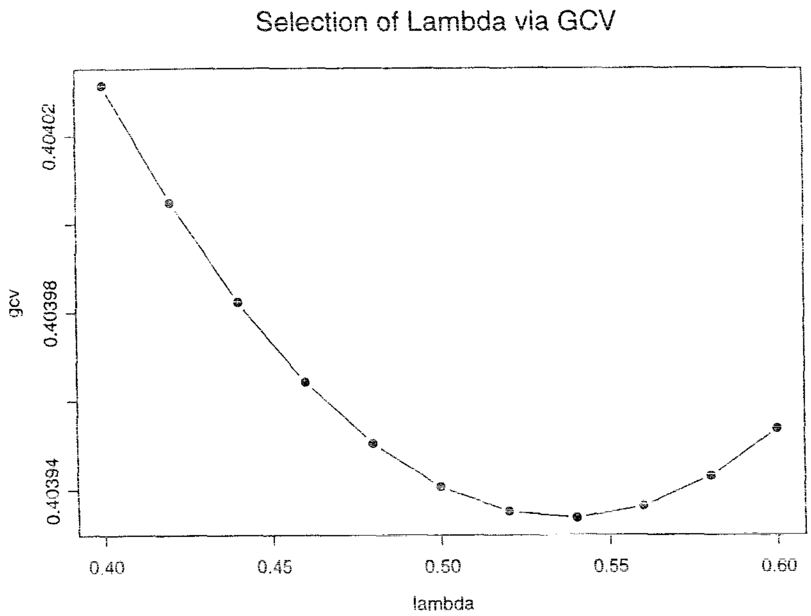

FIG 1. Selection of tuning parameter λ for ridge estimator via GCV.

The plot of $GCV(\lambda)$ is in Figure 1. It is shown that $GCV(\lambda)$ is decreasing then increasing with a minimum attained at $\lambda = 0.54$. The optimal ridge penalty model is selected with $\lambda = 0.54$. The ridge estimator $\hat{\beta}_{ridge}$ is then computed via the IRLS procedure with $\lambda = 0.54$. The standard errors are obtained through 1000 bootstrap samples of the residuals with ridge penalty model $\lambda = 0.54$ as shown in Table II.

We compare the ridge estimator with the intrinsic estimator in Table II. Since the optimal value $\lambda = 0.54$ is small, the ridge estimator presents slight difference from the intrinsic estimator. Smaller standard errors are observed for the ridge estimator due to the shrinkage effect. Similar patterns of the trends along age, period and cohort are also observed between ridge estimator

TABLE II
Comparison of estimators for incidence rates of Ontario cervical cancer
1960-1994

	Ridge	Intrinsic*		Ridge	Intrinsic*
Intercept	2.944(.010)	2.944(.024)			
Age			Cohort		
20-24	-1.815(.016)	-1.868(.111)	-1879	0.045(.070)	0.050(.090)
25-29	-0.504(.035)	-0.502(.058)	1876- 84	0.285(.047)	0.290(.062)
30-34	0.048(.037)	0.055(.048)	81- 89	0.309(.039)	0.312(.052)
35-39	0.309(.033)	0.317(.044)	86- 94	0.264(.037)	0.266(.048)
40-44	0.374(.033)	0.382(.042)	91- 99	0.155(.035)	0.154(.047)
45-49	0.343(.033)	0.350(.041)	1896-1904	0.197(.033)	0.196(.044)
50-54	0.234(.032)	0.241(.041)	1901- 09	0.184(.026)	0.182(.042)
55-59	0.294(.029)	0.300(.038)	06- 14	0.221(.032)	0.218(.043)
60-64	0.251(.030)	0.256(.038)	11- 19	0.168(.033)	0.164(.044)
65-69	0.241(.029)	0.245(.037)	16- 24	0.015(.033)	0.009(.047)
70-74	0.109(0.32)	0.111(.038)	21- 29	-0.119(.034)	-0.126(.049)
75-79	0.117(.030)	0.119(.038)	26- 34	-0.210(.038)	-0.217(.051)
80-84	0.031(.030)	0.030(.039)	31- 39	-0.230(.037)	-0.237(.053)
85+	-0.033(.033)	-0.035(.044)	36- 44	-0.233(.040)	-0.239(.057)
Period			41- 49	-0.180(.042)	-0.186(.062)
1960-64	0.471(.018)	0.469(.023)	46- 54	-0.113(.044)	-0.119(.068)
65-69	0.273(.019)	0.272(.024)	51- 59	-0.107(.049)	-0.112(.079)
70-74	0.094(.021)	0.094(.026)	56- 64	-0.169(.060)	-0.174(.102)
75-79	-0.103(.023)	-0.103(.028)	61- 69	-0.170(.029)	-0.171(.149)
80-84	-0.201(.024)	-0.201(.030)	66- 74	-0.312(.030)	-0.259(.377)
85-89	-0.263(.024)	-0.263(.031)			
90-94	-0.270(.025)	-0.269(.032)			

* Deviance = 14.90 on 60 degrees of freedom. Coefficient of dispersion = 0.25.

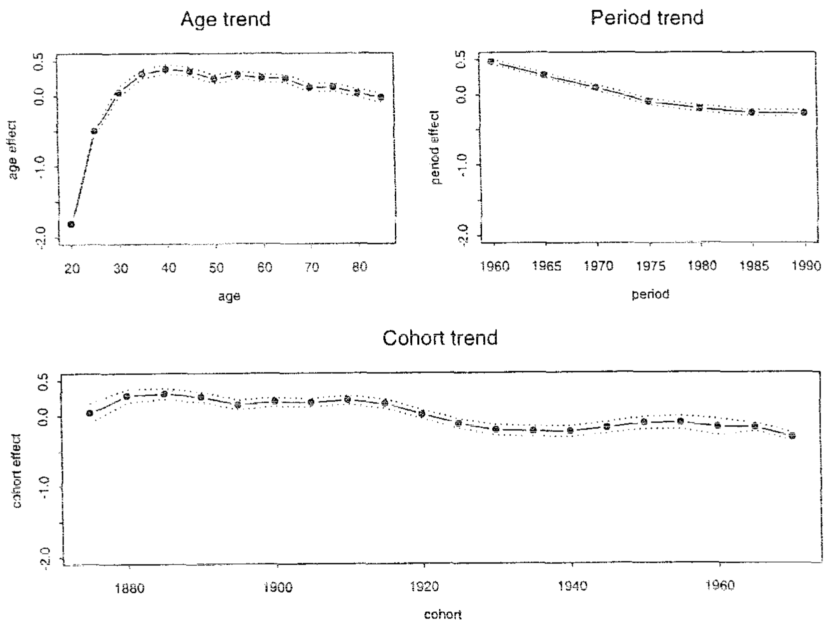

FIG 2. Ridge estimator and its 95% confidence bands for cervical cancer incidence rates in Ontario women. Upper left penal: age trend; Upper right panel: period trend; Lower panel: cohort trend with the mid year of each cohort.

in Figure 2 and the intrinsic estimator in Figure 3. Similar results have been observed (Arraiz, Wigle and Mao (1990)) for the trends in the cervical cancer incidence rates in Canada as a whole.

4. DISCUSSION

Ridge regression has been well studied in linear regression. It yields a shrinkage estimator by penalizing the parameters with ridge penalty and performs well in estimation and prediction when collinearity is present among

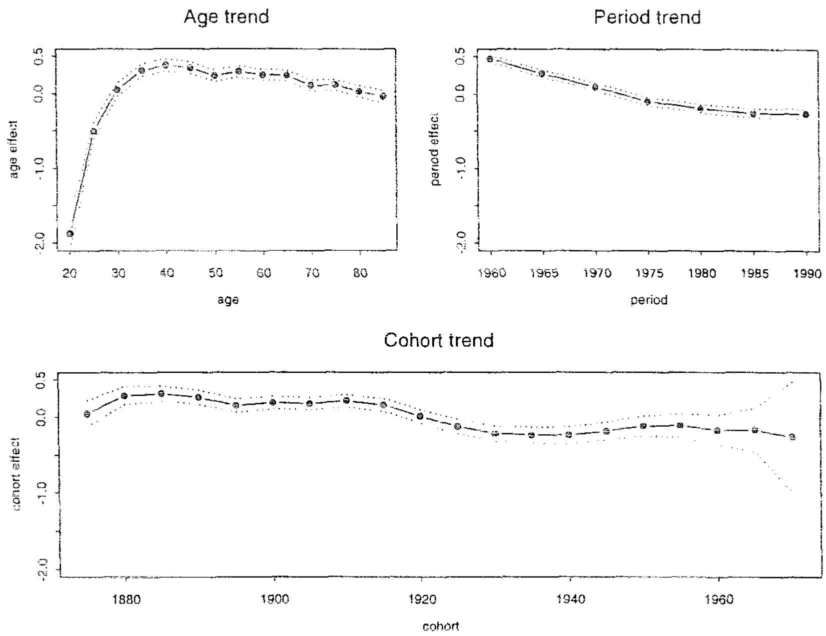

FIG 3. Intrinsic estimator and its 95% confidence bands for cervical cancer incidence rates in Ontario women. Upper left penal: age trend; Upper right panel: period trend; Lower panel: cohort trend with the mid year of each cohort.

covariates. In singular designs, ridge penalty transforms singular matrix X^TX into a non-degenerate matrix $(X^TX + \lambda I)$ by adding positive elements to the diagonal. It yields a unique estimator by shrinking the estimates towards 0. The tuning parameter $\lambda > 0$ can be selected via the GCV method.

In this paper, we have shown that ridge estimator lies in a sub-parameter space orthogonal to the null space of the design matrix X in a singular design. It also converges to the intrinsic estimator, a special estimable function of the singular design. With non-degenerate regression matrix, the convergence of the ridge estimator is trivial since the matrix $(X^TX + \lambda I)^{-1}$ converges as λ tends to 0. However, it is not trivial in a singular design since the matrix $(X^TX + \lambda I)$ becomes non-invertible when $\lambda = 0$. Even if the convergence of the ridge estimator is known, it is still intricate to identify the limit among the potential infinite number of estimators with $\lambda = 0$. Despite these difficulties, the fact that ridge estimator lies in the sub-parameter space helps establish the convergence and identify the limit of the ridge estimator. Since the intrinsic estimator is the result of applying the principal component analysis by eliminating the eigenvalue 0, the convergence of the ridge estimator to the intrinsic estimator reveals an important relationship between the ridge penalty method and the principal component analysis in singular design. The expansion of the ridge estimator not only presents the progression of the convergence, but also provides a variance estimate of the ridge estimator. It shows that the ridge estimator has a smaller variance than the intrinsic estimator, which is not trivial in singular designs compared to non-singular designs.

The application of the ridge estimator to the APC multiple classification model demonstrates that the ridge penalty in singular design yields a unique estimator and helps determine the estimates with no requirement of constraints on the parameters. The ridge estimator with the parameter $\lambda > 0$ selected via the GCV identifies a unique trend in the rates along age, period and cohort, and presents much sensible and useful results that have been confirmed by other studies and that would not be obtained otherwise due to the identifiability problem.

Overall, ridge estimator determines unique estimation in singular designs and yields unique trends in APC analysis of disease rates.

APPENDIX

We prove the convergence of the ridge estimator to the intrinsic estimator for linear regression model (1), the same results apply to generalized linear models as well.

Proof of Theorem 1.

Since matrix X is singular and has 1-less than full rank, the parameter space can be uniquely decomposed as

$$\mathcal{B} = \mathcal{N} \oplus \mathcal{V}$$

where $\mathcal{N}$ is the null space spanned by the unique eigen-vector B_0 of $X^T X$ corresponding to the eigen-value 0, $\mathcal{V}$ is the complementary sub-space of $\mathcal{N}$. Assume the ridge estimator has the decomposition

$$\hat{\boldsymbol{\beta}}(\lambda) = b_0 B_0 + B_1 \boldsymbol{b_1},$$

where B_1 is a matrix with column vectors to be the remaining eigen-vectors of $X^T X$, and $\boldsymbol{b_1}$ a coefficient vector. Since $\hat{\boldsymbol{\beta}}(\lambda)$ minimizes

$$(\boldsymbol{y} - X\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - X\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\beta}$$

for $\lambda > 0$, it must satisfy

$$(X^T X + \lambda I)\boldsymbol{\beta} = X^T \boldsymbol{y} \tag{7}$$

Plug in the decomposition of $\hat{\boldsymbol{\beta}}(\lambda)$,

$$(X^T X + \lambda I)(b_0 B_0 + B_1 \boldsymbol{b_1}) = X^T \boldsymbol{y}$$

Multiply both sides with B_0^T from the left and notice that $XB_0 = 0$ and $B_0^T B_1 = 0$,

$$\lambda b_0 B_0^T B_0 = 0$$

It follows that $b_0 = 0$, which implies that $\hat{\boldsymbol{\beta}}(\lambda) \in \mathcal{V}$.

Proof of Theorem 2.

To prove the existence of the limit of $\hat{\beta}(\lambda)$, notice that

$$(X^T X + \lambda I)B_1 b_1(\lambda) = X^T \boldsymbol{y}$$

from (7). It is easy to prove through simple calculation in linear algebra that the limit of $b_1(\lambda)$ exists as λ tends to 0. Denote the limit of $\hat{\beta}(\lambda)$ by $\hat{\beta}(0)$.

To prove that the limit is equal to the intrinsic estimator, take the limit as $\lambda \to 0$ on both sides of (7),

$$X^T X \hat{\beta}(0) = X^T \boldsymbol{y}$$

Since $\hat{\beta}(\lambda)$ lies in the subspace $\mathcal{V}$, so does its limit $\hat{\beta}(0)$. However, among the infinite number of estimators satisfying

$$X^T X \boldsymbol{\beta} = X^T \boldsymbol{y},$$

the intrinsic estimator is the only one which lies in the sub space $\mathcal{V}$. Therefore, $\hat{\beta}(0)$ is equal to the intrinsic estimator. This completes the proof.

ACKNOWLEDGEMENTS

The author would like to thank Profs. R. Tibshirani, S. Bull, N. Paneth and P. Pathak for very helpful discussions, and also thank Dr. T. Rohan, A. Moskowitz and the Ontario Cancer Registry for providing the Ontario cervical cancer data. This work is partly supported by a National Cancer Institute grant CA83010-01.

BIBLIOGRAPHY

Arraiz, G.A. Wigle, D.T. and Mao Y. (1996) "Is cervical cancer increasing among young women in Canada?" *Canadian Journal of Public Health* 81: 396-397.

Clayton, D. and Schifflers, E. (1987) "Models for temporal variation in cancer rates. II: age-period-cohort models," *Statistics in Medicine*, 6, 469-481.

Craven, P. and Wahba, G. (1979) "Smoothing noisy data with spline functions," *Numerische Mathematik* 31, 377-403.

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

Fu, W.J. (1998) "Penalized regressions: the Bridge versus the Lasso," *Journal of Computational and Graphical Statistist* 7, 397-416.

Fu, W.J. and Rohan, T.E. (1999) "Estimating the effective trends in age-period-cohort studies with applications to cervical cancer incidence and mortality rates," submitted.

Hocking, R.R. (1996) *Methods and applications of linear models: regression and the analysis of variance*, Wiley, New York.

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.

Holford, T.R. (1983) "The estimation of age, period and cohort effects for vital rates," *Biometrics* 39, 311-324.

Holford, T.R. (1991) "Understanding the effects of age, period and cohort on incidence and mortality rates," *Annual Review of Public Health* 12, 425-457.

Kupper, L.L. Janis, J.M. Salama, I.A. Yoshizawa, C.N. and Greenberg, B.G. (1983) "Age-period-cohort analysis: an illustration of the problems in assessing interaction in one observation per cell data," *Communication of Statistics -Theory and Methods* 12, 2779-2807.

Kupper, L.L. Janis, J.M. Karmous, A. and Greenberg, B.G. (1985) "Statistical age-period-cohort analysis: a review and critique," *Journal of Chronic Disease* 38, 811-830

Osmond, C. and Gardner, M.J. (1982) "Age, period and cohort models applied to cancer mortality rates," *Statistics in Medicine* 1, 245-259.

Robertson, C. and Boyle, P. (1998) "Age-period-cohort analysis of chronic disease rates I: modelling approac," *Statistics in Medicine* 17, 1305-1323

Tarone, R.E. and Chu, K.C. (1996) "Evaluation of birth cohort patterns in population disease rates," *American Journal of Epidemiology*, 143, 85-91.

Received September, 1999; Revised October, 1999.