THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CAIRO UNIV., VOL., 41, No. 2, 1997

Quadratic Forms In Normal Variates Under Ridge Regression

By

Abdul-Mordy Hamed Azzam *

Abstract:

This paper concentrates on studying the quadratic forms in normal variates which appear when testing linear statistical hypothesis under ridge regression with positive non-stochastic biased factors $k_1, k_2, \dots, k_p$. Except for the correction factor $n\bar{y}^2$, it is shown that all other quadratic forms are not independent and do not follow central or non-central χ^2 distributions. Hence the classical F statistics are irrelevant under ridge. The results of Hoerl and Kennard (1990) are obtained as special cases when the biased factors are all positive and equal. Moreover, the classical ordinary least squares results are also obtained as special cases when all the biased factors are set to zero.

Key Words:

Ordinary Least Squares (OLS), Ordinary Ridge (OR), Generalized Ridge (GR), Sum of Squares of Regression under OLS (SSR_{ols}), Sum of Squares of Errors under OLS (SSE_{ols}), Sum of Squares of Interaction under OLS (SSI_{ols}), Sum of Squares of Regression under Ridge (SSR_r), Sum of Squares of Errors under Ridge (SSE_r), Sum of Squares of Interaction under Ridge (SSI_r), Mean Square Error (MSE), Orthogonal Projection Operator (OPO), Range space of a matrix [$\underline{R}(.)$], Null space of a matrix [$\underline{N}(.)$], trace of a matrix [$tr(.)$], Canonical Parametrization.

1. Introduction:

In the classical linear regression model

(*) Faculty of Commerce, Alexandria University, Egypt.

$$y = X\beta + \epsilon , \quad cov(y) = \sigma^2 I_n \tag{1.1}$$

let y be an $(n \times 1)$ vector of observations on the dependent variable, $X = [\ J_n : X_1\ ]$ be a known $(n \times p+1)$ full column rank matrix with J_n as an $(n \times 1)$ column vector of 1's and X_1 as an $(n \times p)$ matrix of standardized observations on p regressors such that $X_1' X_1$ is in correlation form, $\beta = [\beta_0, \beta_1, \beta_2, \ldots, \beta_p]' = [\beta_0 : \beta_1']'$ be a $(p+1 \times 1)$ vector of unknown parameters containing the intercept β_0 , and ϵ be an $(n \times 1)$ random error vector having a multivariate normal distribution with zero mean vector and covariance matrix $\sigma^2 I_n$.

Let G be a $(p+1 \times p+1)$ orthogonal matrix containing the normalized eigenvectors of $X' X$, and Λ be the $(p+1 \times p+1)$ diagonal matrix containing the corresponding eigenvalues of $X' X$. Hence $X' X$, G and Λ have the following form :

$$X' X = \left[\begin{array}{c|c} n & 0 \ldots 0 \\ \hline 0 & \\ \vdots & X_1' X_1 \\ 0 & \end{array}\right] , \quad G = \left[\begin{array}{c|c} 1 & 0 \ldots 0 \\ \hline 0 & \\ \vdots & G_1 \\ 0 & \end{array}\right] \quad \Lambda = \left[\begin{array}{c|c} n & 0 \ldots 0 \\ \hline 0 & \\ \vdots & \Lambda_1 \\ 0 & \end{array}\right] \tag{1.2}$$

where G_1 is the $(p \times p)$ orthogonal matrix of the normalized eigenvectors of $X_1' X_1$ and $\Lambda_1 = diag\{ \lambda_1 , \lambda_2 , \ldots , \lambda_p \}$ is the matrix of the corresponding eigenvalues of $X_1' X_1$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$. Hence, $X' X = G \Lambda G'$, $(X' X)^{-1} = G \Lambda^{-1} G'$, $G G' = G' G = I_{p+1}$ and $X_1' X_1 = G_1 \Lambda_1 G_1'$, $(X_1' X_1)^{-1} = G_1 \Lambda_1^{-1} G_1'$ with $G_1 G_1' = G_1' G_1 = I_p$. Now, let $Z_1 = X_1 G_1$ and $Z = [J_n : Z_1] = X G$. Hence, $X_1' J_n = Z_1' J_n = 0$. The canonical parametrization of model (1.1) is

$$y = Z\alpha + \epsilon , \quad cov(y) = \sigma^2 I_n \tag{1.3}$$

where $\alpha = G'\beta = [\alpha_0 , \alpha_1 , \ldots , \alpha_p]' = [\ \alpha_0 : \alpha_1'\ ]'$ with $\alpha_0 = \beta_0$

and $\alpha_1 = G'_1 \beta_1$. The OLS estimator $\hat{\alpha}$ of α is given by

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_0 \\ \\ \hat{\alpha}_1 \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \\ \Lambda_1^{-1} Z'_1 y \end{bmatrix} = \Lambda^{-1} Z' y \qquad (1.4)$$

which is unbiased with covariance matrix and mean square error

$$\begin{aligned} cov\ (\hat{\alpha}) &= \sigma^2 \Lambda^{-1} \\ \text{MSE}(\hat{\alpha}) &= (\sigma^2 / n) + \sigma^2 \sum_{i=1}^{p} \lambda_i^{-1} \\ &= \sigma^2\, tr(\Lambda^{-1}) \end{aligned} \qquad (1.5)$$

If the p regressors are highly correlated, some of the eigenvalues λ_i will be close to zero and hence $\hat{\alpha}_i$ and its MSE($\hat{\alpha}_i$) will be inflated. As a remedy, the ridge regression technique may be used to yield the ridge estimator

$$\hat{\alpha}_r = \begin{bmatrix} \hat{\alpha}_{0r} \\ \\ \hat{\alpha}_{1r} \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \\ \Delta_1^{-1} Z'_1 y \end{bmatrix} \qquad (1.6)$$

where $\Delta_1^{-1} = diag\ \{ 1/\ (\lambda_1 + k_1), \ldots, 1/\ (\lambda_p + k_p\ )\}$ is a $(p \times p)$ positive definite matrix with positive biased factors $k_1, k_2, \ldots, k_p$. It is also easy to see that

$$\hat{\alpha}_r = \begin{bmatrix} \hat{\alpha}_{0r} \\ \\ \hat{\alpha}_{1r} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_0 \\ \\ F_1\ \hat{\alpha}_1 \end{bmatrix} \qquad (1.7)$$

where $F_1 = diag\{\delta_1, \ldots, \delta_p\}$ is a $(p \times p)$ positive definite matrix with $\delta_i = \lambda_i/(\lambda_i + k_i)$. It is clear that $\hat{\alpha}_0 = \hat{\alpha}_{0r} = \bar{y}$ while $\hat{\alpha}_{ri} = \delta_i \hat{\alpha}_i$, $i = 1, 2, \ldots, p$.

In ridge regression literature, two different types of questions are considered. The first one addresses the problem of choosing "appropriate" values of the biase factors $k_1, k_2, \ldots, k_p$ such that the resulting MSE is minimized. ~~while~~ The second one addresses the inference problem about the parameter vector α . Obenchain (1977), Hoerl and Kennerd (1990) and Azzam (1996) began tackling the inference problem assuming that the biased factors are non-stochastic. Obenchain (1977) claimed that the F and t statistics are identical under both OLS and ridge. Azzam (1996) showed that this statement is only true when using the unbiased OLS estimator s^2 of σ^2 in the denominator of both F and t statistics. Hoerl and Kennard (1990) partitioned th total sum of squares and its degrees of freedom under OR regression.

In this paper, the results of Hoerl and Kennard (1990) are generalized to the case of GR regression. Moreover, it is shown that the resulting quadratic forms, except $n\bar{y}^2$, are not independent and do not follow central or non-central χ^2 distributions and hence the classical F statistics are irrelevant under ridge.

Section 2 is devoted to the partitioning of the total sum of squares and its degrees of freedom under GR regression. The question of independence of the resulting quadratic forms and whether their distributions are central or non-central χ^2 or not is addressed in Section 3. Section 4 introduces the "ANOVA - Like" table under GR regression together with the classical "ANOVA" table under OLS which shows that both tables are identical when the biased factors $k_1 = k_2 = \dots = k_p = 0$.

2. Partitioning the Total Sum of Squares and its Degrees of Freedom under GR Regression:

In the classical OLS method, the total sum of squares $y'y$ is partitioned as follows:

$$y'y = n\bar{y}^2 + SSR_{ols} + SSE_{ols}$$
$$= y' A_1 y + y' P_{Z_1} y + y' (I_n - P_Z) y \qquad (2.1)$$

with

$$I_n = A_1 + P_{Z_1} + (I_n - P_Z) \qquad (2.2)$$

where $A_1 = J_n (J_n' J_n)^{-1} J_n'$, $P_{Z_1} = Z_1 (Z_1' Z_1)^{-1} Z_1'$, and $P_Z = A_1 + P_{Z_1}$ are $(n \times n)$ symmetric and idempotent matrices. Note that A_1 , P_{Z_1} and $(I_n - P_Z)$ are the OPO on $\underline{R}(J_n)$, the regression space $\underline{R}(Z_1)$, and the error space $\underline{N}(Z')$ respectively. Since $A_1 P_{Z_1}$ is the zero matrix, then $P_Z P_{Z_1} = P_{Z_1}$ and hence

$$(I_n - P_Z) P_{Z_1} = 0 \qquad (2.3)$$

So, the error and the regression spaces are orthogonal and hence the sum of squares due to their interaction (SSI_{ols}) is zero under OLS. Also, from (2.2) the degrees of freedom are decomposed as follows:

$$n = tr(I_n) = 1 + p + (n - p - 1)$$
$$= tr(A_1) + tr(P_{Z_1}) + tr(I_n - P_Z) \qquad (2.4)$$

Note that any of the components in the R.H.S. of (2.1),(2.2) and (2.4) may be obtained by subtraction. Note also that each of the total sum of squares and its degrees of freedom may be partitioned to four components one of them, SSI_{ols} and its degrees of freedom, is zero under OLS. Following the same line of thinking and using the fact that the error and regression spaces are not orthogonal under ridge, it is easy to see that

$$y'y = n\bar{y}^2 + SSR_r + SSE_r + SSI_r$$
$$= y'A_1 y + y'A_2 y + y'A_3 y + y'A_4 y \qquad (2.5)$$

such that

$$I_n = A_1 + A_2 + A_3 + A_4 \qquad (2.6)$$

and hence the degrees of freedom may be decomposed as:

$$n = tr(I_n) = tr(A_1) + tr(A_2) + tr(A_3) + tr(A_4) \qquad (2.7)$$

Recall that $A_1 = J_n (J_n' J_n)^{-1} J_n'$. The matrices A_2 and A_3 will be obtained via SSR_r and SSE_r respectively, while A_4 will be obtained using (2.6).

To determine A_2 , the symmetric matrix associated with the regression sum of squares under ridge, recall that SSR_{ols} may be expressed as:

$$SSR_{ols} = \hat{\alpha}_1' Z_1' Z_1 \hat{\alpha}_1 = y' P_{Z_1} y \qquad (2.8)$$

Using (2.8) with $\hat{\alpha}_{1r}$ instead of $\hat{\alpha}_1$ together with equation (1.6), it is clear that

$$SSR_r = \hat{\alpha}_{1r}' Z_1' Z_1 \hat{\alpha}_{1r} = y' A_2 y \qquad (2.9)$$

where

$$A_2 = Z_1 B_1 Z_1' \qquad (2.10)$$

with $B_1 = diag\{\lambda_1/(\lambda_1+k_1)^2, \ldots, \lambda_p/(\lambda_p+k_p)^2\}$ as a $(p \times p)$ positive definite matrix. Note that if the biased factors $k_1 = k_2 = \ldots = k_p = 0$, then $B_1 = (Z_1' Z_1)^{-1}$, $A_2 = P_{Z_1}$, the OPO on $\underline{R}(Z_1)$, and hence equation (2.9) reduces to equation (2.8). Note also that A_2 may be expressed in terms of the design matrix Z as follows:

$$A_2 = Z[diag\{0, \lambda_1/(\lambda_1+k_1)^2, \ldots, \lambda_p/(\lambda_p+k_p)^2\}]Z' \qquad (2.11)$$

Now, let $\hat{y}_r = Z\hat{\alpha}_r$ and $\hat{\epsilon}_r = y - \hat{y}_r$ be the predicted and error vectors under ridge respectively. Using (1.6), then

$$\hat{\epsilon}_r = (I_n - Z\Delta^{-1}Z')y \qquad (2.12)$$

with $\Delta^{-1} = diag\ \{ 1/n\ ,\ 1/(\lambda_1 + k_1)\ ,\ \dots,\ 1/(\lambda_p + k_p\ )\ \}$ as a $(p+1 \times p+1)$ positive definite matrix. Hence . using (2.12) we have

$$SSE_r = y' A_3\ y \qquad (2.13)$$

where $A_3 = (I_n - Z\ \Delta^{-1}\ Z')^2$. With little algebra, it can be shown that

$$A_3 = (I_n - Z\ \ B_2\ Z') \qquad (2.14)$$

with $B_2 = diag\ \{1/n, (\lambda_1 + 2k_1)/(\lambda_1 + k_1)^2\ ,\ \dots,\ (\lambda_p + 2k_p)/(\lambda_p + k_p\ )^2\}$ as a $(p+1 \times p+1)$ positive definite matrix. If the biased factors are set to zero, then $B_2 = (\ Z'\ Z\ )^{-1}$, $A_3 = I_n - P_Z$, the OPO on $\underline{N}(Z')$, and hence SSE_r in (2.13) reduces to

$$SSE_{ols} = y' (\ I_n - P_Z\ )\ y \qquad (2.15)$$

Finally, the symmetric matrix A_4 associated with SSI_r is obtained via (2.6), i.e.,

$$\begin{aligned} A_4 &= I_n - (A_1 + A_2 + A_3\ ) \\ &= 2\ Z\ B_3\ Z' \qquad (2.16) \end{aligned}$$

with $B_3 = diag\ \{0,\ k_1/(\lambda_1 + k_1)^2\ ,\ \dots,\ k_p/(\lambda_p + k_p\ )^2\}$ as a $(p+1 \times p+1)$ positive definite matrix. Therefore,

$$SSI_r = y' A_4\ y \qquad (2.17)$$

If the biased factors are set to zero , then A_4 is the $(n \times n)$ zero matrix, and (2.16) reduces to (2.3) leading to the orthogonality of the regression and error spaces as it should be under OLS. In this case, the decomposition of the total sum of squares in (2.5) under GR coincides with the corresponding decomposition in (2.1) under OLS. Now, let us use (2.7) together with (2.10), (2.14), (2.16) and the fact that A_1 is the OPO on $\underline{R}(J_n)$ to decompose the degrees of freedom as follows:

$$tr(A_1) = 1$$

$$tr(A_2) = \sum_{i=1}^{p} \lambda_i^2 / (\lambda_i + k_i)^2 ,$$

$$tr(A_3) = (n-1) - \sum_{i=1}^{p} \lambda_i (\lambda_i + 2 k_i) / (\lambda_i + k_i)^2 , \qquad (2.18)$$

$$tr(A_4) = 2 \sum_{i=1}^{p} \lambda_i k_i / (\lambda_i + k_i)^2$$

Note that the decomposition in (2.18) adds to $n = tr(I_n)$. Also, if the biased factors are set to zero, then the decomposition of the degrees of freedom in (2.18) under GR coincides with the corresponding decomposition in (2.4) under OLS. Finally, the results of Hoerl and Kennerd (1990) are obtained as special cases of the results of this section when all the biased factors are set equal to a positive constant k. As a final remark, the quadratic forms SSR_r, SSE_r and SSI_r depend on the choice of the biased factors k_1, k_2, ... , k_p and hence different choices of these biased factors may lead to different values of each of these quadratic forms.

3.Statistical Properties:

In this section,Theorems (3.1) and (3.3),PP.123-124, in [3] will be used together with the assumption of the normality of the random vector y to show that the quadratic forms SSR_r, SSE_r and SSI_r are not independent and do not follow any of the χ^2 distributions. Also, it is shown that $\hat{\alpha}_r$ and $\hat{\epsilon}_r$ are jointly normal and statistically dependent.

Lemma(3.1):

The quadratic forms SSR_r, SSE_r and SSI_r do not follow any of the χ^2 distributions.

Proof:

Using Theorem (3.1) in [3], it suffices to show that the symmetric matrices A_2 , A_3 and A_4 are not idempotent. Using (2.10), we find that

$$A_2^{\,2} = Z_1 [\ diag\ \{\ \lambda_1^3/(\lambda_1 + k_1)^4 ,\ \ldots,\ \lambda_p^3 / (\lambda_p + k_p)^4\ \}]\ Z_1'$$

Hence, $A_2 = A_2^{\,2}$ if and only if $\lambda_i^2 = (\lambda_i + k_i)^2$, i.e., if and only if $k_i = 0$ for all i, which is impossible under ridge. Also, using (2.14) we have

$$A_3^{\,2} = I_n - Z\ [\ diag\ \{\ a_0\ ,\ a_1\ ,\ \ldots,\ a_p\ \}]\ Z'$$

with $a_0 = 1/n$, $a_i = (\lambda_i + 2\ k_i) [\ (\lambda_i + k_i)^2 + k_i^2\ ] / (\lambda_i + k_i)^4$, $i = 1, 2, \ldots, p$. Hence A_3 is idempotant if and only if $k_1 = k_2 = \ldots = k_p = 0$ which is impossible under ridge. Finally, using (2.16), it is clear that $A_4 = A_4^{\,2}$ if and only if $k_1 = k_2 = \ldots = k_p = 0$ which is also impossible under ridge. □

Note that if the biased factors are set to zero, then the symmetric matrices A_2 and A_3 are idempotent and the corresponding quadratic forms follow χ^2 distributions, as it should be under OLS.

Lemma(3.2):

The quadratic forms SSR_r, SSE_r and SSI_r are not independent.

Proof:

Using Theorem (3.2) in [3], together with the fact that $cov\ (y) = \sigma^2\ I_n$, it suffices to show that the matrices $A_2\ A_3$, $A_2\ A_4$ and $A_3\ A_4$ do not equal the $(n \times n)$ zero matrix. Using equation (2.11) and (2.14), we find that

$$A_2\ A_3 = Z_1\ H_1\ Z_1' \qquad (3.1)$$

with $H_1 = diag\ \{\ \lambda_1 k_1^{\,2}/(\lambda_1 + k_1)^4 ,\ \ldots,\ \lambda_p\, k_p^2\ /\ (\lambda_p + k_p)^4\}$ as a

$(p \times p)$ positive definite matrix. Also, using (2.11) and (2.16), we have

$$A_2 A_4 = 2 Z_1 H_2 Z_1' \quad (3.2)$$

with $H_2 = diag \{\lambda_1^2 k_1 / (\lambda_1 + k_1)^4, \ldots, \lambda_p^2 k_p / (\lambda_p + k_p)^4\}$ as a $(p \times p)$ positive definite matrix. Finally, using (2.14) and (2.16), we find that

$$A_3 A_4 = 2 Z_1 H_3 Z_1' \quad (3.3)$$

with $H_3 = diag \{ k_1^3 / (\lambda_1 + k_1)^4, \ldots, k_p^3 / (\lambda_p + k_p)^4 \}$ as a $(p \times p)$ positive definite matrix. Since each of (3.1) - (3.3) represents a group of positive definite quadratic forms, then each of them equals the $(n \times n)$ zero matrix if and only if Z_1 is the zero matrix, which is impossible. □

Note that if the biased factors are set to zero, then H_1 equals the $(p \times p)$ zero matrix and the corresponding quadratic forms are independent, which is the case under OLS.

Theorem (3.1):

The F statistic is irrelevant under ridge.

Proof:

Use lemma (3.1) or lemma (3.2).□

Lemma(3.3):

The quadratic form $n\bar{y}^2$ is independent of each of the quadratic forms SSR_r, SSE_r and SSI_r .

Proof:

It suffices to show that the matrices $A_1 A_2$, $A_1 A_3$ and $A_1 A_4$ are equal to the $(n \times n)$ zero matrix. Using the fact that $A_1 J_n = J_n$, $A_1 Z_1 = 0$ and $A_1 Z = [J_n : 0]$, it is easy to show that $A_1 A_2 = A_1 A_3 = A_1 A_4 = 0$. □

Note that the quadratic form $n\bar{y}^2$ has a non-central χ^2 distribution with one degree of freedom and a non-centrality

ISSR, CAIRO UNIV., VOL., 41, No. 2, 1997 PP, 95-109

parameter $n\,\alpha_0^2$.

Lemma(3.4):

$\hat{\alpha}_r$ and $\hat{\epsilon}_r$ are jointly normal and statistically dependent.

Proof:

Using lemma (1.6) and (2.12) we fine that $\hat{\alpha}_r$ and $\hat{\epsilon}_r$ are linear combinations of the random vector y of the form

$$\hat{\alpha}_r = A\,y \quad , \quad \hat{\epsilon}_r = (\,I_n - B\,)\,y$$

with $A = \Delta^{-1} Z$, $B = Z\,\Delta^{-1}\,Z'$ and

$\Delta^{-1} = diag\,\{\,1/n\,,\,1/(\lambda_1 + k_1)\,,\,\ldots,\,1/(\lambda_p + k_p)\,\}$. Hence

$[\hat{\alpha}'_r \;\; \hat{\epsilon}'_r]' = Wy \sim N_{n+p+1}\,(\,W\,Z\,\alpha\,,\,\sigma^2\,W\,W'\,)$ with

$$W = \begin{bmatrix} A \\ \ldots\ldots \\ I_n - B \end{bmatrix}$$ as an $(n+p+1 \times n)$ matrix and

$$W\,W' = \begin{bmatrix} A\,A' & A\,(\,I_n - B) \\ (I_n - B)\,A' & (\,I_n - B)^2 \end{bmatrix}$$

as an $(n+p+1 \times n+p+1)$ matrix . Note that

$$cov\,(\hat{\alpha}_r\,,\,\hat{\epsilon}_r) = \sigma^2\;A\,(\,I_n - B) = \sigma^2 \begin{bmatrix} 0 \quad 0 \ldots 0 \\ \ldots\ldots\ldots \\ H\,Z_1' \end{bmatrix}$$

with $H = diag\,\{\,k_1\,/\,(\lambda_1 + k_1)^2\,,\,\ldots,\,k_p\,/\,(\lambda_p + k_p)^2\,\}$ as a $(p \times p)$ positive definite matrix. Hence $cov\,(\hat{\alpha}_r\,,\,\hat{\epsilon}_r) = 0$ if and only if $H\,Z_1' = 0$. Since H is a positive definite matrix, then $cov\,(\hat{\alpha}_r\,,\,\hat{\epsilon}_r) = 0$ if and only if Z_1 is the zero matrix which is impossible , and the result holds. □

Note that $\hat{\alpha}_r$ and $\hat{\epsilon}_r$ are marginally normal and that the OLS results concerning independence holds if $k_1 = \ldots = k_p = 0$.

4. The "ANOVA-Like" Table:

In this section, the expected values of the quadratic forms SSR_r, SSE_r and SSI_r are obtained and the "ANOVA-Like" table is constructed assuming that the biased factors are positive and different. These expected values are obtained using the well known result

$$E(y'A\,y) = \mu' A \mu + tr\,(\textstyle\sum A\,) \qquad (4.1)$$

with $\mu = E(y)$ and $\sum = cov\,(y)$. Applying (4.1) to the canonical parametrization (1.3) and using equations (2.10),(2.14) and (2.16) , it is easy to show that

$$E(SSR_r) = \sum_{i=1}^{p} \lambda_i^2\,(\lambda_i \alpha_i^2 + \sigma^2)\;/\;(\lambda_i + k_i)^2 \qquad (4.2)$$

$$E(SSE_r) = (n-1)\,\sigma^2 + \sum_{i=1}^{p} \lambda_i\,[k_i^2 \alpha_i^2 - \sigma^2(\lambda_i + 2\,k_i)]\;/\;(\lambda_i + k_i)^2 \qquad (4.3)$$

11

$$E(SSI_r) = 2\sum_{i=1}^{p} \lambda_i\,k_i\,(\lambda_i \alpha_i^2 + \sigma^2)\;/\;(\lambda_i + k_i)^2 \qquad (4.4)$$

Recalling that

$$E(n\bar{y}^2) = n\,\alpha_0^2 + \sigma^2 \qquad (4.5)$$

and adding (4.2)-(4.5), we have

$$E(y'\,y) = \alpha'\,Z'\,Z\,\alpha + n\,\sigma^2 \qquad (4.6)$$

as it should be. Also, if the biased factors are set to zero, then (4.2) and (4.3) coincide with the well known OLS results [see Table (4.2)] while (4.4) reduces to zero.

In Tables (4.1) and (4.2), the "ANOVA-Like" table under ridge and the "ANOVA" table under OLS respectively, two different null hypotheses are considered. These null hypotheses are H_0: $\alpha_0 = 0$ and H_0: $\alpha_1 = \dots = \alpha_p = 0$. The sums of squares in Table (4.1) are obtained using (2.9), (2.13) and (2.17) while the degrees of freedom are obtained using (2.18).

Examining the "ANOVA-Like" table under ridge and the "ANOVA" table under OLS , the following remarks can be drawn:

1. Both tables coincide when $k_1 = k_2 = \dots = k_p = 0$.

2. The quadratic forms SSR_r, SSE_r and SSI_r and their degrees of freedom depend on the choice of the non-stochastic biased factors.

3. The expected value of SSE_r depends on the null hypothesis H_0: $\alpha_1 = \dots = \alpha_p = 0$, while the expected value of SSE_{ols} is $\sigma^2 (n-p-1)$ whether H_0 is true or not.

Finally, remarks 2 and 3 together with the results of Section 3 assure that the published F-tables are irrelevant for inference under ridge regression.

Table (4.1)

The "ANOVA-Like" Table Under Ridge

Source	SS	DF	MSS	E(SS)	H_0	E(SS) Under H_0	F-Like Statistics
Mean	$y'A_1 y = n\bar{y}^2$	$tr(A_1) = 1$	$n\bar{y}^2 / tr(A_1)$	$n\,\alpha_0^2 + \sigma^2$	$\alpha_0 = 0$	σ^2	$[y'A_1 y/tr(A_1)] / [y'A_3 y/tr(A_3)]$
Regression	$y'A_2 y = \hat{\alpha}_{1r}' Z_1' Z_1 \hat{\alpha}_{1r}$	$tr(A_2) = \sum_{i=1}^{p} \lambda_i^2 / (\lambda_i+k_i)^2$	$y'A_2 y/tr(A_2)$	$\sum_{i=1}^{p} \lambda_i^2 (\lambda_i \alpha_i^2 + \sigma^2)/(\lambda_i+k_i)^2$	$\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$	$\sigma^2 \sum_{i=1}^{p} \lambda_i^2 / (\lambda_i+k_i)^2$	$[y'A_2 y/tr(A_2)] / [y'A_3 y/tr(A_3)]$
Interaction	$y'A_4 y$	$tr(A_4) = 2 \sum_{i=1}^{p} \lambda_i k_i / (\lambda_i+k_i)^2$	$y'A_4 y/tr(A_4)$	$2 \sum_{i=1}^{p} \lambda_i k_i (\lambda_i \alpha_i^2 + \sigma^2)/(\lambda_i+k_i)^2$		$2\,\sigma^2 \sum_{i=1}^{p} \lambda_i k_i / (\lambda_i+k_i)^2$	
Error	$y'A_3 y$	$tr(A_3) = (n-1) - \sum_{i=1}^{p} \lambda_i(\lambda_i+2k_i) / (\lambda_i+k_i)^2$	$y'A_3 y/tr(A_3)$	$(n-1)\sigma^2 + \sum_{i=1}^{p} \lambda_i [\, k_i^2 \alpha_i^2 - \sigma^2 (\lambda_i + 2k_i)] / (\lambda_i+k_i)^2$		$\sigma^2[(n-1) - \sum_{i=1}^{p} \lambda_i(\lambda_i+2k_i)/(\lambda_i+k_i)^2]$	
Total	$y'y$	$tr(I_n) = n$		$\alpha' Z' Z \alpha + n\sigma^2$		$n\,\sigma^2$	

Table (4.2)

The "ANOVA" Table Under OLS

Source	SS	DF	MSS	E(SS)	H_0	E(SS) Under H_0	F-Statistic
Mean	$y'A_1 y = n\bar{y}^2$	$tr(A_1) = 1$	$n\bar{y}^2 / tr(A_1)$	$n\,\alpha_0^2 + \sigma^2$	$\alpha_0 = 0$	σ^2	$[y'A_1 y/tr(A_1)] / [y'(I_n-P_Z)y/tr((I_n-P_Z))]$
Regression	$y'P_{Z_1} y = \hat{\alpha}_1' Z_1' Z_1 \hat{\alpha}_1$	$tr(P_{Z_1}) = p$	$y'P_{Z_1} y/tr(P_{Z_1})$	$\alpha_1' Z_1' Z_1 \alpha_1 + p\,\sigma^2$	$\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$	$p\,\sigma^2$	$[y'P_{Z_1} y/tr(P_{Z_1})] / [y'(I_n-P_Z)y/tr((I_n-P_Z))]$
Error	$y'(I_n-P_Z)y$	$tr(I_n-P_Z) = (n-p-1)$	$y'(I_n-P_Z)y/tr(I_n-P_Z)$	$(n-p-1)\sigma^2$		$(n-p-1)\sigma^2$	
Total	$y'y$	$tr(I_n) = n$		$\alpha' Z' Z \alpha + n\sigma^2$		$n\,\sigma^2$	

References:

1. Azzam, A. H. (1996). Inference in Linear Models with Non-Stochastic Biased Factors.

2. Hoerl, A. E. and Kennard, R. W. (1990). Ridge Regression: Degrees of Freedom in the Analysis of Variance. Commun. Statist. Simula., 19(4), PP.1485 - 1495.

3. Morrison, D. F. (1983). Applied Linear Statistical Models. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 07632.

4. Obenchain, R. L. (1977). Classical F- test and Confidence Regions for Ridge Regression. Technometrics, 19, PP. 429 - 439.