Banco Central de Chile
Documentos de Trabajo

Central Bank of Chile
Working Papers

N° 65

Marzo 2000

A NOTE ON THE MOMENTS OF STOCHASTIC SHRINKAGE PARAMETERS IN RIDGE REGRESSION

Luis Firinguetti   Hernán Rubio

BANCO CENTRAL DE CHILE

CENTRAL BANK OF CHILE

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Huérfanos 1175, primer piso.
Teléfono: (56-2) 6702475   Fax: (56-2) 6702231

Documentos de Trabajo
N° 65

Working Paper
N° 65

A NOTE ON THE MOMENTS OF STOCHASTIC SHRINKAGE PARAMETERS IN RIDGE REGRESSION

Luis Firinguetti
Departamento de Matemática
Universidad de Santiago

Hernán Rubio
Estadístico
Departamento de Cuentas Nacionales
Banco Central de Chile

Resumen

Un problema común en modelos econométricos y de regresión múltiple en general es el de la multicolinealidad, que produce efectos indeseados en los estimadores de Cuadrados Mínimos. Una posible solución a este problema es el estimador de Regresión "Ridge", propuesto por Hoerl y Kennard (1970), que ha sido aplicado en diversas áreas tales como la economía, el marketing y la calibración de instrumentos en procesos industriales. Sin embrago las propiedades de estos estimadores dependen de manera crucial de la elección de ciertos parámetros de sesgo que son estocásticos. Se han hecho varias propuestas al respecto y el propósito de este trabajo es derivar expresiones generales para los momentos de dichos parámetros estocásticos de sesgo. Con esto esperamos establecer condiciones bajo las cuales un estimador de Regresión Ridge es preferible a otros.

Abstract

A common problem in econometric models and multiple regression in general is multicollinearity, which produces undesirable effects on the Least Squares estimators. A possible solution to this problem is the "Ridge" Regression estimator proposed by Hoerl and Kennard (1970). Ridge Regression has been applied to such diverse areas as economics, marketing and the calibration of instruments in industrial processes. However, the properties of these estimators crucially depend upon the selection of certain biasing parameters which are stochastic. In this regard several proposals have been made and the purpose of this paper is to derive general expressions for the moments of the stochastic biasing parameters. With this knowledge we expect to stablish conditions under which a Ridge Regression estimator is better than others.

Forthcoming in *Communications in Statistics: Simulation and Computation*, 2000. This research was partially supported by FONDECYT-CHILE. Corresponding addresses: lfiringu@latinmail.com, hrubio@condor.bcentral.cl.

1 Introduction

Let us consider the Classical Linear Regression Model (CLRM)

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}\,, \tag{1.1}$$

where $\underline{y}$ is an $n \times 1$ vector of observations of the dependent variable; X is an $n \times p$ full rank matrix of nonstochastic observations of the explanatory variables; $\underline{\beta}$ is a $p \times 1$ vector of unknown coefficients and $\underline{\epsilon}$ is an $n \times 1$ vector of unobserved random disturbances, such that

$$\underline{\epsilon} \sim N(\underline{0}, \sigma^2 I)\,. \tag{1.2}$$

In this model the Ordinary Least Squares (OLS) estimator,

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{y}\,, \tag{1.3}$$

is known to have optimum properties. However, it is also well known that collinearity can have harmful effects on the OLS estimates: some elements of the parameter vector may be imprecisely estimated as their variances will be large, or the estimated coefficients may even have the wrong signs. This fact prompted Hoerl and Kennard (1970a, 1970b) to propose an alternative estimator which, although biased, may have a smaller Mean Square Error (MSE) than OLS. Let us define H and Q as the matrices of eigenvalues and eigenvectors of $X'X$, so that

$$Q'X'XQ = \Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)\; and\; Q'Q = QQ' = I\,. \tag{1.4}$$

Then the orthogonal version of the CLRM (1.1) is

$$\underline{y} = XQQ'\underline{\beta} + \underline{\epsilon} = Z\underline{\alpha} + \underline{\epsilon}\,, \tag{1.5}$$

where

$$Z = XQ\; and\; \underline{\alpha} = Q'\underline{\beta}\,. \tag{1.6}$$

The Generalized Ridge Regression (GRR) estimator proposed by Hoerl and Kennard (1970a, 1970b) is defined by

$$\underline{\tilde{\alpha}} = (\Lambda + K)^{-1}Z'\underline{y} = (\Lambda + K)^{-1}H\underline{\hat{\alpha}}\,, \tag{1.7}$$

where

$$K = diag(k_1, k_2, ..., k_p)\,,\; k_i > 0 \tag{1.8}$$

and

$$\underline{\hat{\alpha}} = \Lambda^{-1}Z'\underline{y}\,, \tag{1.9}$$

is the Ordinary Least Squares (OLS) estimator of α. Thus, according to (1.6) the GRR estimator of $\underline{\beta}$ is

$$\tilde{\underline{\beta}} = Q\tilde{\underline{\alpha}}\,. \tag{1.10}$$

Hoerl and Kennard (1970a, 1970b) have shown that the values of k_i that minimize the MSE of $\tilde{\underline{\beta}}$ are given by

$$k_i = \frac{\sigma^2}{{\alpha_i}^2}\,. \tag{1.11}$$

where α_i is the ith element of $\underline{\alpha}$. Because σ^2 and α_i are unknown, Hoerl and Kennard (1970a, 1970b) propose replacing them by their OLS counterparts, yielding

$$\hat{k}_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}\,. \tag{1.12}$$

Hoerl and Kennard (1970a, 1970b) also proposed setting $K = kI$, which yields the Ordinary Ridge Regression (ORR) estimator of $\underline{\beta}$:

$$\tilde{\underline{\beta}}_k = (X'X + kI)^{-1}X'\underline{y}\,. \tag{1.13}$$

No explicit optimum value for k can be found. Nevertheless, several stochastic choices have been proposed for the shrinkage parameter k. For example, Hoerl, Kennard and Baldwin (1975) propose taking the armonic mean of the $\hat{k}_i$ in (1.12), yielding the following stochastic value of k:

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\underline{\beta}}'\hat{\underline{\beta}}}\,. \tag{1.14}$$

Also, from a Bayesian perspective, Lawless and Wang (1976) propose taking

$$\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\hat{\underline{\beta}}'X'X\hat{\underline{\beta}}}\,. \tag{1.15}$$

Another Bayesian interpretation of the ORR estimator can be found in Frank and Friedman (1993) who also provide an interesting discussion and comparison of several regression tools commonly used in chemometrics.

An important question arises regarding the alternative operational Ridge Regression (RR) estimators: which one is to be preferred? Clearly this will require the knowledge of the properties of the estimators. Hemmerle and Carey (1983) derive some exact finite sample properties of GRR estimators and Kozumi and Othani (1994) have obtained general expressions for the moments of the ORR estimator proposed by Lawless and Wang (1976). Further evidence on the behavior of these estimators may be found on a number of simulation studies (see for example Hoerl, Kennard and Baldwin (1975), Lawless and Wang (1976), Frank and Friedman (1993)). However from these and other studies no general conclusions may be reached. Nevertheless it must be clear that the performance of all the operational Ridge Regression estimators will depend on the properties of the stochastic shrinkage parameters. For instance, given (1.2), $\hat{k}_i$ has no finite moments of any order and therefore it is expected to yield values that are, on average, too large. Consequently, the resulting GRR estimator will shrink too much the estimates of $\underline{\alpha}$ and $\underline{\beta}$ towards zero. The effect of this will be to introduce much more bias than necessary to produce Ridge Regression estimates with good MSE properties. Thus, additional knowledge on the performance of the RR estimators may be gained if we study the moments of their stochastic shrinkage parameters. This is the aim of this note. In the next section we derive the moments of $\hat{k}_{LW}$; in section 3 we derive the moments of $\hat{k}_{HKB}$; in section 4 we present some numerical results to compare the mean and variance of these stochastic shrinkage parameters and finally, in section 5 we conclude.

2 The Moments of the Lawless and Wang Stochastic Shrinkage Parameter

Kadiyala(1981) derived $E(\hat{k}_{LW})$. Here we will obtain $E(\hat{k}_{LW}^r)$.

Theorem 2.1. *Under the conditions stated in equations (1.1) and (1.2) The rth moment of* $\hat{k}_{LW} = p\hat{\sigma}^2/\underline{\hat{\beta}}'X'X\underline{\hat{\beta}}$ *is given by*

$$E(\hat{k}_{LW}^r) = \left(\frac{p}{n-p}\right)^r \frac{\Gamma(\frac{n-p}{2}+r)}{\Gamma(\frac{n-p}{2})} \sum_{j=0}^{\infty} \frac{e^{-\theta/2}(\theta/2)^j}{j!} \frac{\Gamma(p/2+j-r)}{\Gamma(p/2+j)}, \quad j = 1, 2, \ldots, p, \tag{2.1}$$

provided $p > 2r$, *where*

$$\theta = \frac{\underline{\beta}'X'X\underline{\beta}}{\sigma^2}. \tag{2.2}$$

Proof. First, from the definition of $\hat{k}_{LW}$ and equation (1.2) we note that

$$\hat{k}_{LW} = \frac{p}{n-p}\frac{u}{v} \tag{2.3}$$

where

$$u = (n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(n-k)} \tag{2.4}$$

and

$$v = \underline{\hat{\beta}}' X' X \underline{\hat{\beta}}/\sigma^2 \sim \chi^2_{(n-k\,,\,\theta)} \tag{2.5}$$

Hence

$$E(\hat{k}^r_{LW}) = \left(\frac{p}{n-p}\right)^r E(u^r) E(1/v^r) \tag{2.6}$$

Now from (2.4)

$$E(u^r) = 2^r \frac{\Gamma(\frac{n-p}{2}+r)}{\Gamma(\frac{n-p}{2})}\,, \tag{2.7}$$

and from (2.5), provided $p > 2r$,

$$\begin{aligned} E(1/v^r) &= \sum_{j=0}^{\infty} \frac{e^{-\theta/2}(\theta/2)^j}{j!} \int_0^{\infty} v^{-r} \frac{v^{p/2+j-1}e^{-v/2}}{2^{p/2+j}\Gamma(p/2+j)}\, dv \\ &= 2^{-r} \sum_{j=0}^{\infty} \frac{e^{-\theta/2}(\theta/2)^j}{j!} \frac{\Gamma(p/2+j-r)}{\Gamma(p/2+j)} \\ &\quad \times \int_0^{\infty} \frac{v^{p/2+j-r-1}e^{-v/2}}{2^{p/2+j-r}\Gamma(p/2+j-r)}\, dv \\ &= 2^{-r} \sum_{j=0}^{\infty} \frac{e^{-\theta/2}(\theta/2)^j}{j!} \frac{\Gamma(p/2+j-r)}{\Gamma(p/2+j)}\,, \end{aligned} \tag{2.8}$$

since the integral is a gamma distribution with parameters $p/2+j-r$ and 2. Then, replacing (2.7) and (2.8) in (2.6) we obtain the desired result. □

From (2.1) it is easily established that an upper bound for $E(\hat{k}^r_{LW})$ is given by

$$0 \le E(\hat{k}^r_{LW}) \le \left(\frac{p}{n-p}\right)^r \frac{\Gamma(\frac{n-p}{2}+r)}{\Gamma(\frac{n-p}{2})} \frac{\Gamma(p/2-r)}{\Gamma(p/2)} \tag{2.9}$$

The special case $r = 1$ gives

$$0 \le E(\hat{k}_{LW}) \le \frac{p}{p-2} \tag{2.10}$$

This result was first established by Kadiyala (1981).

3 The Moments of the Hoerl, Kennard and Baldwin Stochastic Shrinkage Parameter

According to (1.14) $\hat{k}_{HKB}$ may be written as

$$\hat{k}_{HKB} = \frac{p}{n-p}\frac{u}{w} \, , \tag{3.1}$$

where u is defined in (2.4) and

$$w = \frac{\underline{\hat{\beta}}'\underline{\hat{\beta}}}{\sigma^2} = \frac{\underline{y}'M\underline{y}}{\sigma^2} \, , \tag{3.2}$$

where

$$M = X(X'X)^{-2}X' \, . \tag{3.3}$$

Under assumption (1.2) u and w are independent, therefore

$$E(\hat{k}_{HKB}^r) = \left(\frac{p}{n-p}\right)^r E(u^r)E(1/w^r) \tag{3.4}$$

$E(u^r)$ is already known (see equation(2.7). However, to evaluate $E(1/w^r)$ we need the density function of $w = \underline{\hat{\beta}}'\underline{\hat{\beta}}/\sigma^2$, which is derived in the following theorem.

Theorem 3.1. *Let $\underline{\hat{\beta}}$ be the OLS estimator of $\underline{\beta}$, given in (1.3), then under conditions (1.1) and (1.2) the density function of $w = \underline{\hat{\beta}}'\underline{\hat{\beta}}/\sigma^2$ is given by*

$$f(w) = \sum_{j=0}^{\infty} \frac{a_j}{\Gamma(p/2+j)(2\Delta)^{p/2+j}} w^{p/2+j-1} e^{-w/(2\Delta)} \;\; , \;\; w > 0 \tag{3.5}$$

where

$$a_0 = e^{-d/2} \prod_{i=1}^{p} (\Delta/h_i)^{1/2} \;\; \text{and} \;\; a_j = (2j)^{-1} \sum_{i=0}^{j-1} b_{j-i} a_i \;\; , \;\; j > 0 \tag{3.6}$$

$$b_j = j\Delta \sum_{i=1}^{p} (\delta_i^2/h_i) c_i^{j-1} + \sum_{i=1}^{p} c_i^j \;\; , \;\; j > 0 \tag{3.7}$$

$$d = \sum_{i=1}^{p} \delta_i^2 \tag{3.8}$$

$$c_i = 1 - \Delta/h_i \tag{3.9}$$

and Δ is a number such that

$$| c_i | = | 1 - \Delta/h_i | < 1 \ , \ \ i = 1, 2, ..., p \tag{3.10}$$

and h_i is the ith eigenvalue of $M = X(X'X)^{-2}X'$.

Proof. Let H and P be the matrices of eigenvalues and eigenvectors of $M = X(X'X)^{-2}X'$, so that

$$Q'MQ = H = diag(h_1, h_2, ..., h_p) \text{ and } Q'Q = QQ' = I \ . \tag{3.11}$$

Note that because rank of M is p, $h_i = 0$ for $i = p+1, p+2, ..., n$. Now from equations (1.1) and (1.2) $\underline{y}'M\underline{y}$ may be written as:

$$\begin{aligned} \underline{y}'M\underline{y} &= \sigma^2(X\underline{\beta}/\sigma + \underline{\epsilon}/\sigma)'M(X\underline{\beta}/\sigma + \underline{\epsilon}/\sigma) \\ &= \sigma^2(P'X\underline{\beta}/\sigma + P'\underline{\epsilon}/\sigma)'H(P'X\underline{\beta}/\sigma + P'\underline{\epsilon}/\sigma) \\ &= \sigma^2 \sum_{i=1}^{n} h_i(u_i + \delta_i)^2 = \sum_{i=1}^{p} h_i(u_i + \delta_i)^2 \ , \end{aligned} \tag{3.12}$$

where

$$\underline{u} = (u_1 u_2 \cdots u_n)' = P'\underline{\epsilon}/\sigma \ \text{ and } \ \underline{\delta} = (\delta_1 \delta_2 \cdots \delta_n)' = P'X/\sigma \ . \tag{3.13}$$

Also note that

$$\underline{u} \sim N(\underline{0}, I) \ . \tag{3.14}$$

Then applying Theorem 1 in Ruben (1962) we obtain the desired result. □

Therefore

$$\begin{aligned} E(1/w^r) &= \sum_{j=0}^{\infty} \int_0^{\infty} w^{-r} \frac{a_j}{\Gamma(p/2+j)(2\Delta)^{p/2+j}} w^{p/2+j-1} e^{-w/(2\Delta)} \, dw \\ &= \sum_{j=0}^{\infty} \frac{a_j}{(2\Delta)^r} \frac{\Gamma(p/2+j-r)}{\Gamma(p/2+j)} \int_0^{\infty} \frac{w^{(p/2+j-r)-1} e^{-w/(2\Delta)}}{\Gamma(p/2+j-r)(2\Delta)^{p/2+j-r}} \, dw \\ &= \sum_{j=0}^{\infty} \frac{a_j}{(2\Delta)^r} \frac{\Gamma(p/2+j-r)}{\Gamma(p/2+j)} \ . \end{aligned} \tag{3.15}$$

The integral is equal to 1 since it is a gamma distribution with parameters $p/2+j-r$ and 2Δ. Note that this result is only valid if $p > 2r$.

We now prove a theorem concerning the moments of $\hat{k}_{HKB}$.

Theorem 3.2. *Let the density function of* $w = \underline{\hat{\beta}}'\underline{\hat{\beta}}/\sigma^2$ *be given by (3.5) and* $u = (n-p)\hat{\sigma}^2/\sigma^2$ *distributed according to (2.4). Then the moment of order* r *of* $\hat{k}_{HKB} = p\hat{\sigma}^2/\underline{\hat{\beta}}'\underline{\hat{\beta}}$ *is given by*

$$E(\hat{k}^r_{HKB}) = \left(\frac{p}{n-p}\right)^r \frac{1}{\Delta^r}\frac{\Gamma(\frac{n-p}{2}+r)}{\Gamma(\frac{n-p}{2})}\sum_{j=0}^{\infty} a_j \frac{\Gamma(p/2+j-r)}{\Gamma(p/2+j)}\ , \ r = 1, 2, ..., \qquad (3.16)$$

provided $p > 2r$. *The quantities* Δ *and* a_j *are as defined by equations (3.6) to (3.10).*

Proof. Follows straightforwardly by replacing (2.7) and (3.15) into (3.4). □

4 Some Numerical Results

Since the exact results are difficult to interpret, in this section we present some numerical calculations to compare the exact results obtained for the mean and variance of $\hat{k}_{LW}$ and $\hat{k}_{HKB}$.

The different model specifications can be characterized by the signal to noise ratio:

$$\theta = \frac{\beta' X' X \beta}{\sigma^2} = \frac{\delta' H \delta}{\sigma^2} \quad . \qquad (4.1)$$

This quantity is sensitive to the values of $X, \underline{\beta}$ and σ^2.
Hence the following factors were varied:

i) Degree and pattern of collinearity.

ii) The vector of coefficients $\underline{\beta}$.

iii) The error variance σ^2.

i) Four X matrices were specified, each with 5 explanatory variables, including a constant term, and 25 observations. To achieve different patterns and degrees of multicollinearity, the explanatory variables were generated using the following device:

$$x_{tj} = (1-a_j^2)^{1/2}\, z_{tj} + a_j z_{t,k} \qquad j = 1, \cdots, k-1 \quad ; \quad t = 1, \cdots, n.$$

where

$$z_{tj} \sim U(0,1) \qquad j = 1, \cdots\cdots, k \quad ; \quad t = 1, \cdots\cdots, n.$$

Different degrees and patterns of multicollinearity were achieved by specifying the following sets of a_j values:

$$\begin{aligned} A_1 &= (0.2, 0.3, 0.4, 0.5); \\ A_2 &= (-0.2, -0.3, -0.95, -0.95); \\ A_3 &= (0.2, 0.8, 0.95, 1); \\ A_4 &= (0.99, 0.95, 0.65, 0.6). \end{aligned}$$

ii) Three vectors of coefficients $\underline{\beta}$ were specified:

$$\underline{\beta}_1 = \frac{\sum_{i=1}^k i\underline{q}_i}{\sqrt{\sum_{i=1}^k i}}; \quad \underline{\beta}_2 = \frac{\sum_{i=1}^k (k-i+1)\underline{q}_i}{\sqrt{\sum_{i=1}^k i}}; \quad \underline{\beta}_3 = \frac{\sum_{i=1}^k \underline{q}_i}{\sqrt{k}},$$

where $Q = (\underline{q}_1, \underline{q}_2, \cdots\cdots, \underline{q}_k)$ is the matrix of eigenvectors of $X'X$. which are ordered according to the magnitude of their corresponding eigenvalues. Hence $\underline{\beta}_1$ is a weighted average of the eigenvectors with more weight given to the eigenvectors corresponding to the smallest eigenvalues; on the contrary $\underline{\beta}_2$ is a weighted average giving more weight to eigenvectors corresponding to the largest eigenvalues; finally $\underline{\beta}_3$ is a simple average of all eigenvectors. These vectors of coefficients are such that $\underline{\beta}'\underline{\beta} = 1$.

iii) The following values of σ^2 were considered $2.5, 5, 10, 20$.

Thus, a total of 48 different models were considered using all possible combinations of the factor specifications.

The results of the numerical calculations for the mean and variance are presented in Table I. From these results the behavior of $\hat{k}_{HKB}$ can be characterized by the following:

i) $V(\hat{k}_{HKB})$ increases with the value of σ^2, whereas the $E(\hat{k}_{HKB})$ increases very slowly (or remains constant).

ii) The mean and variance of $\hat{k}_{HKB}$ are largest for $\underline{\beta}_1$ and smallest for $\underline{\beta}_2$.

iii) $V(\hat{k}_{HKB})$ tends to decrease with increasing θ. However $E(\hat{k}_{HKB})$ is less sensitive to the size of θ.

TABLE I
Mean and Variance of the Hoerl, Kennard and Baldwin and of the Lawless and Wang Shrinkage Parameters

Collinearity Vector	$\underline{\beta}$	σ^2	θ	H K B	L W
A_1	$\underline{\beta}_1$	2.5	0.9427	$E(\hat{k}) = 2.8400$ $V(\hat{k}) = 34.2354$	$E(\hat{k}) = 1.3907$ $V(\hat{k}) = 4.3864$
		5	0.4714	$E(\hat{k}) = 3.0176$ $V(\hat{k}) = 40.0951$	$E(\hat{k}) = 1.5196$ $V(\hat{k}) = 5.2922$
		10	0.2357	$E(\hat{k}) = 3.1125$ $V(\hat{k}) = 43.4358$	$E(\hat{k}) = 1.5907$ $V(\hat{k}) = 5.8144$
		20	0.1178	$E(\hat{k}) = 3.1615$ $V(\hat{k}) = 45.2207$	$E(\hat{k}) = 1.6280$ $V(\hat{k}) = 6.0948$
	$\underline{\beta}_2$	2.5	12.6184	$E(\hat{k}) = 2.3624$ $V(\hat{k}) = 7.2152$	$E(\hat{k}) = 0.3613$ $V(\hat{k}) = 0.0898$
		5	6.3092	$E(\hat{k}) = 2.6943$ $V(\hat{k}) = 13.9611$	$E(\hat{k}) = 0.6351$ $V(\hat{k}) = 0.5811$
		10	3.1546	$E(\hat{k}) = 2.9180$ $V(\hat{k}) = 23.1505$	$E(\hat{k}) = 0.9610$ $V(\hat{k}) = 1.8465$
		20	1.5773	$E(\hat{k}) = 3.0535$ $V(\hat{k}) = 32.0590$	$E(\hat{k}) = 1.2413$ $V(\hat{k}) = 3.4118$
	$\underline{\beta}_3$	2.5	5.8568	$E(\hat{k}) = 2.5741$ $V(\hat{k}) = 13.8577$	$E(\hat{k}) = 0.6693$ $V(\hat{k}) = 0.6808$
		5	2.9284	$E(\hat{k}) = 2.8514$ $V(\hat{k}) = 23.4195$	$E(\hat{k}) = 0.9948$ $V(\hat{k}) = 2.0141$
		10	1.4642	$E(\hat{k}) = 3.0185$ $V(\hat{k}) = 32.4107$	$E(\hat{k}) = 1.2661$ $V(\hat{k}) = 3.5675$
		20	0.7321	$E(\hat{k}) = 3.1113$ $V(\hat{k}) = 38.8212$	$E(\hat{k}) = 1.4462$ $V(\hat{k}) = 4.7698$

TABLE I continued

Collinearity Vector	$\underline{\beta}$	σ^2	θ	H K B	L W
A_2	$\underline{\beta}_1$	2.5	0.7146	$E(\hat{k}) = 1.9275$ $V(\hat{k}) = 20.7547$	$E(\hat{k}) = 1.4510$ $V(\hat{k}) = 4.8031$
		5	0.3573	$E(\hat{k}) = 2.0175$ $V(\hat{k}) = 23.5315$	$E(\hat{k}) = 1.5534$ $V(\hat{k}) = 5.5387$
		10	0.1787	$E(\hat{k}) = 2.0649$ $V(\hat{k}) = 25.0708$	$E(\hat{k}) = 1.6086$ $V(\hat{k}) = 5.9484$
		20	0.0893	$E(\hat{k}) = 2.0893$ $V(\hat{k}) = 25.8815$	$E(\hat{k}) = 1.6373$ $V(\hat{k}) = 6.1647$
	$\underline{\beta}_2$	2.5	8.5484	$E(\hat{k}) = 1.6163$ $V(\hat{k}) = 5.4389$	$E(\hat{k}) = 0.5034$ $V(\hat{k}) = 0.2775$
		5	4.2742	$E(\hat{k}) = 1.8171$ $V(\hat{k}) = 10.1901$	$E(\hat{k}) = 0.8184$ $V(\hat{k}) = 1.2094$
		10	2.1371	$E(\hat{k}) = 1.9488$ $V(\hat{k}) = 15.6751$	$E(\hat{k}) = 1.1287$ $V(\hat{k}) = 2.7383$
		20	1.0685	$E(\hat{k}) = 2.0263$ $V(\hat{k}) = 20.1893$	$E(\hat{k}) = 1.3590$ $V(\hat{k}) = 4.1726$
	$\underline{\beta}_3$	2.5	4.0556	$E(\hat{k}) = 1.7531$ $V(\hat{k}) = 10.0075$	$E(\hat{k}) = 0.8434$ $V(\hat{k}) = 1.3122$
		5	2.0278	$E(\hat{k}) = 1.9133$ $V(\hat{k}) = 15.6496$	$E(\hat{k}) = 1.1494$ $V(\hat{k}) = 2.8580$
		10	1.0139	$E(\hat{k}) = 2.0076$ $V(\hat{k}) = 20.2166$	$E(\hat{k}) = 1.3726$ $V(\hat{k}) = 4.2641$
		20	0.5069	$E(\hat{k}) = 2.0591$ $V(\hat{k}) = 23.1754$	$E(\hat{k}) = 1.5093$ $V(\hat{k}) = 5.2176$

TABLE I continued

Collinearity Vector	$\underline{\beta}$	σ^2	θ	H K B	L W
A_3	$\underline{\beta}_1$	2.5	0.7314	$E(\hat{k}) = 1.0690$ $V(\hat{k}) = 8.8818$	$E(\hat{k}) = 1.4464$ $V(\hat{k}) = 4.7712$
		5	0.3657	$E(\hat{k}) = 1.0985$ $V(\hat{k}) = 9.8665$	$E(\hat{k}) = 1.5509$ $V(\hat{k}) = 5.5202$
		10	0.1828	$E(\hat{k}) = 1.1138$ $V(\hat{k}) = 10.4083$	$E(\hat{k}) = 1.6073$ $V(\hat{k}) = 5.9385$
		20	0.0914	$E(\hat{k}) = 1.1216$ $V(\hat{k}) = 10.6927$	$E(\hat{k}) = 1.6364$ $V(\hat{k}) = 6.1595$
	$\underline{\beta}_2$	2.5	12.5533	$E(\hat{k}) = 0.9234$ $V(\hat{k}) = 2.3323$	$E(\hat{k}) = 0.3629$ $V(\hat{k}) = 0.0912$
		5	6.2766	$E(\hat{k}) = 1.0069$ $V(\hat{k}) = 3.9442$	$E(\hat{k}) = 0.6374$ $V(\hat{k}) = 0.5877$
		10	3.1383	$E(\hat{k}) = 1.0609$ $V(\hat{k}) = 5.9693$	$E(\hat{k}) = 0.9634$ $V(\hat{k}) = 1.8580$
		20	1.5692	$E(\hat{k}) = 1.0929$ $V(\hat{k}) = 7.8619$	$E(\hat{k}) = 1.2430$ $V(\hat{k}) = 3.4227$
	$\underline{\beta}_3$	2.5	5.7240	$E(\hat{k}) = 0.9876$ $V(\hat{k}) = 3.9919$	$E(\hat{k}) = 0.6800$ $V(\hat{k}) = 0.7136$
		5	2.8620	$E(\hat{k}) = 1.0508$ $V(\hat{k}) = 6.0977$	$E(\hat{k}) = 1.0050$ $V(\hat{k}) = 2.0663$
		10	1.4310	$E(\hat{k}) = 1.0877$ $V(\hat{k}) = 7.9894$	$E(\hat{k}) = 1.2735$ $V(\hat{k}) = 3.6146$
		20	0.7155	$E(\hat{k}) = 1.1079$ $V(\hat{k}) = 9.3090$	$E(\hat{k}) = 1.4507$ $V(\hat{k}) = 4.8014$

TABLE I continued

Collinearity Vector	$\underline{\beta}$	σ^2	θ	H K B	L W
A_4	$\underline{\beta}_1$	2.5	0.6063	$E(\hat{k}) = 0.8836$ $V(\hat{k}) = 5.8878$	$E(\hat{k}) = 1.4810$ $V(\hat{k}) = 5.0150$
		5	0.3031	$E(\hat{k}) = 0.9034$ $V(\hat{k}) = 6.4097$	$E(\hat{k}) = 1.5699$ $V(\hat{k}) = 5.6598$
		10	0.1516	$E(\hat{k}) = 0.9135$ $V(\hat{k}) = 6.6920$	$E(\hat{k}) = 1.6172$ $V(\hat{k}) = 6.0132$
		20	0.0758	$E(\hat{k}) = 0.9187$ $V(\hat{k}) = 6.8388$	$E(\hat{k}) = 1.6417$ $V(\hat{k}) = 6.1982$
	$\underline{\beta}_2$	2.5	10.8179	$E(\hat{k}) = 0.7835$ $V(\hat{k}) = 1.7817$	$E(\hat{k}) = 0.4134$ $V(\hat{k}) = 0.1428$
		5	5.4089	$E(\hat{k}) = 0.8423$ $V(\hat{k}) = 2.8751$	$E(\hat{k}) = 0.7065$ $V(\hat{k}) = 0.7985$
		10	2.7045	$E(\hat{k}) = 0.8790$ $V(\hat{k}) = 4.1538$	$E(\hat{k}) = 1.0300$ $V(\hat{k}) = 2.1960$
		20	1.3522	$E(\hat{k}) = 0.9002$ $V(\hat{k}) = 5.2685$	$E(\hat{k}) = 1.2914$ $V(\hat{k}) = 3.7290$
	$\underline{\beta}_3$	2.5	4.9044	$E(\hat{k}) = 0.8293$ $V(\hat{k}) = 2.9371$	$E(\hat{k}) = 0.7529$ $V(\hat{k}) = 0.9586$
		5	2.4522	$E(\hat{k}) = 0.8724$ $V(\hat{k}) = 4.2529$	$E(\hat{k}) = 1.0722$ $V(\hat{k}) = 2.4217$
		10	1.2261	$E(\hat{k}) = 0.8968$ $V(\hat{k}) = 5.3540$	$E(\hat{k}) = 1.3208$ $V(\hat{k}) = 3.9199$
		20	0.6131	$E(\hat{k}) = 0.9100$ $V(\hat{k}) = 6.0882$	$E(\hat{k}) = 1.4791$ $V(\hat{k}) = 5.0014$

iv) With higher degrees of collinearity, particularly A_3 and A_4, the mean and variance of $\hat{k}_{HKB}$ decrease.

The results for $\hat{k}_{LW}$ are as follows:

i) The mean and variance of $\hat{k}_{LW}$ tend to increase with σ^2.

ii) The mean and variance of $\hat{k}_{LW}$ are largest for $\underline{\beta}_1$ and smallest for $\underline{\beta}_2$.

iii) The larger θ is, the smaller $E(\hat{k}_{LW})$ and $V(\hat{k}_{LW})$ are.

iv) There is a marginal increase in the mean and variance of $\hat{k}_{LW}$ with increasing collinearity.

v) $\hat{k}_{LW}$ has consistently smaller variance than $\hat{k}_{HKB}$. Also $\hat{k}_{LW}$ has a smaller mean than $\hat{k}_{HKB}$, unless collinearity is high (A_3, A_4) and the signal to noise ratio is low ($\theta < 2.8$).

5 Concluding Remarks

In this paper we have derived the moments of the shrinkage parameters proposed by Hoerl, Kennard and Baldwin (1975) and by Lawless and Wang (1976). These moments exist provided the order of the moment is less than half the number of coefficients in the equation. It has also been shown that $\hat{k}_{LW}$ has bounded moments.

From the numerical results we see that $\hat{k}_{LW}$ has some attributes that make it superior to $\hat{k}_{HKB}$: in the first place, the value of $\hat{k}_{LW}$ decreases with increasing signal to noise ratio, which is a desirable result since OLS tends to be unbeatable when θ is large. Secondly, $E(\hat{k}_{LW})$ will tend to be smaller than $\mathrm{E}(\hat{k}_{HKB})$ if there is less collinearity and smaller values of the signal to noise ratio. Thirdly, $\hat{k}_{LW}$ shows less variability than $\hat{k}_{HKB}$. Therefore it is likely that $\hat{k}_{LW}$ will introduce more bias than $\hat{k}_{HKB}$ on the estimated coefficient vector β only when there is a greater chance of reducing the MSE of the estimator. As a consequence of this the Lawless and Wang ORR estimator may be expected to have a smaller MSE than OLS more often than the Hoerl, Kennard and Baldwin ORR estimator. This may be the case even if on occasions the Hoerl, Kennard and Baldwin ORR estimator may outperform the Lawless and Wang estimator.

REFERENCES

Frank, I. and Friedman, J. (1993). "A Statistical View of Some Chemometrics Regression Tools," *Technometrics*, 35, 109-135.

Hemmerle, W. J. and Carey, M. B. (1984) "Some Properties of Generalized Ridge Estimators," *Communications in Statistics-Simulation and Computation*, B12, 239-256.

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.

Hoerl, A. E. and Kennard, R. W. and Baldwin, K. F. (1975). "Ridge Regression: Some Simulations," *Communications in Statistics*, 4, 105-123.

Hocking, R.R., Speed, F.M. and Lynn, M.J. (1976). "A Class of Biased Estimators in Linear Regression," *Technometrics*, 18, 425-437.

Kadiyala, K. (1981) "Bounds for the Biasing Parameter in Ridge Regression," *Communications in Statistics-Theory and Methods*, A10, 2369-2372.

Kozumi, H. and Othani, K. (1994) "The General Expressions for the Moments of the Lawless and Wang's Ordinary Ridge Regression Estimator," *Communications in Statistics-Theory and Methods*, A23, 2755-2774.

Lawless, J. F. and Wang P. (1976) "A Simulation Study of Some Ridge and Other Regression Estimators," *Communications in Statistics-Theory and Methods*, A5, 307-323.

Ruben, H. (1962) "Probability Content of Regions Under Spherical Normal Distributions, IV: The Distribution of Homogeneous and Non-Homogeneous Quadratic Functions of Normal Variables," *The Annals of Mathematical Statistics*, 33, 542-570.

Documentos de Trabajo	**Working Papers**
Banco Central de Chile	**Central Bank of Chile**
NÚMEROS ANTERIORES	PAST ISSUES

La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: http://www.bcentral.cl/Estudios/DTBC/doctrab.htm. Existe la posibilidad de solicitar una copia impresa con un costo de $500 si es dentro de Chile y US$12 si es para fuera de Chile. Las solicitudes se pueden hacer por fax: (56-2) 6702231 o a través de correo electrónico: bcch@condor.bcentral.cl

Working Papers in PDF format can be downloaded free of charge from: http://www.bcentral.cl/Estudios/DTBC/doctrab.htm. Hard copy versions can be ordered individually for US$12 per copy (for orders inside Chile the charge is Ch$500.) Orders can be placed by fax: (56-2) 6702231 or email: bcch@condor.bcentral.cl

DTBC-64 Febrero 2000
El Encaje, los Flujos de Capitales Y el Gasto: Una Evaluación Empírica
Guillermo Le Fort y Sergio Lehmann

DTBC-63 Febrero 2000
Household Saving in Chile: Microeconomic Evidence
Andrea Butelmann y Francisco Gallego

DTBC-62 Enero 2000
Bank Concentration: Chile and International Comparisons
Ross Levine

DTBC-61 Enero 2000
Optimal Monetary Policy Rules under Inflation Range Targeting
Juan Pablo Medina y Rodrigo Valdés

DTBC-60 Diciembre 1999
Comovement and Macroeconomic Interdependence: Evidence for Latin America, East Asia, and Europe
Norman Loayza, Humberto Lopez y Angel Ubide

DTBC-59 Diciembre 1999
Capital Controls in Chile: Effective? Efficient?
Francisco Gallego, Leonardo Hernández y Klaus Schmidt-Hebbel

DTBC-58 Diciembre 1999
Demand for Reserves under International Capital Mobility
Pablo García

DTBC-57 Diciembre 1999
Origins and Resolution of a Banking Crisis: Chile 1982-86
Edgardo Barandiarán y Leonardo Hernández

DTBC-56 Diciembre 1999
Financial Intermediation and Growth: Causality and Causes
Thorsten Beck, Ross Levine y Norman Loayza

DTBC-55 Diciembre 1999
Inflation Targets and Stabilization in Chile
Oscar Landerretche, Felipe Morandé y Klaus Schmidt-Hebbel

DTBC-54 Diciembre 1999
Income Inequality and the Real Exchange Rate
Pablo García

DTBC-53 Diciembre 1999
Series de Ahorro e Ingreso por Agente Económico en Chile, 1960-1997
Herman Bennett, Klaus Schmidt-Hebbel y Claudio Soto

DTBC-52 Diciembre 1999
Integración Financiera y Coordinación Macroeconómica en el Mercosur
Carlos Budnevich y Roberto Zahler

DTBC-51 Noviembre 1999
Determinants of Current Account Deficits in Developing Countries
César Calderón, Alberto Chong y Norman Loayza

DTBC-50 Noviembre 1999
The Effect of Capital Controls on Interest Rate Differentials
Luis Oscar Herrera y Rodrigo Valdés

DTBC-49 Noviembre 1999
Cuenta Corriente y Desvíos Transitorios en Términos de Intercambio y Volúmenes de Exportaciones: Chile 1985-1999
Jaime Guajardo y Guillermo Le Fort

DTBC-48 Noviembre 1999
Do Depositors Punish Banks for "Bad" Behavior?: Examining Market Discipline in Argentina, Chile, and Mexico
María Soledad Martínez y Sergio Schmukler