Banco Central de Chile
Documentos de Trabajo

Central Bank of Chile
Working Papers

N° 137

Enero 2002

THE DISTRIBUTION OF STOCHASTIC SHRINKAGE PARAMETERS IN RIDGE REGRESSION

Hernán Rubio Luis Firinguetti

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

Documento de Trabajo N° 137	Working Paper N° 137

THE DISTRIBUTION OF STOCHASTIC SHRINKAGE PARAMETERS IN RIDGE REGRESSION

Hernán Rubio
Departamento de Cuentas Nacionales
Banco Central de Chile

Luis Firinguetti
Gerencia de Información e Investigación Estadística
Banco Central de Chile

Resumen

En este trabajo se derivan las funciones de densidad y probabilidad acumulada de los parámetros de sesgo estocástico de tres conocidos estimadores de Regresión "Ridge" operacionales. El comportamiento de estos parámetros afecta las propiedades del estimador de Regresión "Ridge" resultante, por lo que un conocimiento de este tipo puede ser útil en la selección de la regla de encogimiento. También se presentan algunos cálculos numéricos para ilustrar el comportamiento de estas distribuciones. Estos resultados pueden a su vez ayudar a explicar el comportamiento de los estimadores.

Abstract

In this article we derive the density and distribution functions of the stochastic shrinkage parameters of three well-known operational Ridge Regression estimators by assuming normality. The stochastic behavior of these parameters is likely to affect the properties of the resulting Ridge Regression estimator, therefore such knowledge can useful in the selection of the shrinkage rule. Some numerical calculations are carried out to illustrate the behavior of these distributions, throwing light on the performance of the different Ridge Regression estimators.

E-mails: hrubio@bcentral.cl ; lfiringu@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 non-stochastic 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,

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

has well known optimal properties. Nevertheless, OLS can badly be affected by collinearity, a common condition in non-experimental time series data. With multicollinear data some coefficients may be statistically insignificant and may have the wrong signs. Working in the field of Engineering, Hoerl and Kennard (2000, 1970) found this to be a common occurrence, and led them to propose an alternative estimator which, although biased, may have a smaller Mean Square Error (MSE) than OLS.

Let Λ and Q be the matrices of eigenvalues and eigenvectors of $X'X$, then

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

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

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

The Generalized Ridge Regression (GRR) estimator proposed by Hoerl and Kennard (2000) is defined by

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

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

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

Hoerl and Kennard (2000) have shown that the values of k_i that minimize the MSE of $\underline{\tilde{\beta}}$ are given by

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

where α_i is the ith element of $\underline{\alpha}$. To yield an operational estimator, Hoerl and Kennard (2000, 1970) propose replacing σ^2 and α_i by their OLS estimates:

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

A simpler version of the estimator has also been discussed by Hoerl and Kennard (2000). The so called Ordinary Ridge Regression (ORR) estimator, which is obtained by setting $K = kI$:

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

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

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

From a Bayesian perspective, Lawless and Wang (1976) propose

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

as the estimator of k. Another Bayesian interpretation of the ORR estimator is provided by Frank and Friedman (1993), who also give an interesting discussion and comparison of Ridge Regression (RR) with other regression tools commonly used in chemometrics. For the new developments in RR techniques the reader is referred to Aldrin (1997), Elston and Proe (1995), Foucart (1999), Jang and Yoon (1997), Kibria (1996), Saleh and Kibria

(1993) and recently Shi and Wang (1999) among others. RR is also reviewed in a recent book by Gruber (1998).

A notorious fact about RR is that the original, 1970, article by Hoerl and Kennard has been republished in the Special 40th Anniversary Issue of *Technometrics*, being regarded by Gunst (2000) as a classical study that revolutionized the practice of regression analysis. But also, according to Gunst (2000): “Although ridge regression is widely used in the application of regression methods today, it remains as controversial as when it was first introduced”. Indeed, Gunst (2000) points out that RR methods have been criticized on two grounds: “Existence theorems do not apply to the usual setting where ridge parameters must be estimated from the data” and “assumptions needed for the ridge estimator to be optimal in a well-defined theoretical sense are unrealistic in practice, yet simulations often inadvertently impose these very assumptions”. From the outset, in this article we make no claim of universal validity of RR estimators, but remain convinced that if collinearity is present, particularly if the signal to noise ratio is not large, (it is not only the β coefficient that matters), RR is a useful alternative to OLS. We moreover take the view that the properties of RR estimators are strongly dependent on the stochastic shrinkage parameters and an effort should be made in studying their properties. Take for instance the case of $\hat{k}_i$, which according to (1.2), has no finite moments of any order, yielding 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 introducing much more bias than necessary to produce RR estimators with good MSE properties. Ultimately the selection of one among the many alternative operational RR estimators requires the knowledge of the properties of the estimators. Thus, Hemmerle and Carey (1984) derive some exact finite sample properties of GRR estimators, (see also Inoue (1999)), and Kozumi and Othani (1994) have obtained general expressions for the moments of the ORR estimator proposed by Lawless and Wang (1976). A perhaps more interesting paper from the practitioner's point of view is that of Crivelli, Firinguetti, Montaño and Muñoz (1995), which, apart from showing consistency, provides asymptotic confidence intervals based on the ORR due to Lawless and Wang. Unfortunately no general conclusions can be reached from these studies, but as argued earlier on the performance of all operational Ridge Regression estimators will crucially depend upon the distribution of the stochastic shrinkage parameters. The aim of this paper is to determine the probability density and the probability distribution function of the shrinkage parameters. In sections 2, 3 and 4 we set about to derive the distribution and density of $\hat{k}_i$, $\hat{k}_{LW}$ and $\hat{k}_{HKB}$ respectively;

in section 5 some numerical results are presented to compare the densities of these stochastic shrinkage parameters; finally a concluding remarks has been presented in section 6.

2 Distribution of the Hoerl and Kennard Stochastic Shrinkage Parameter

We want to determine the distribution of $\hat{k}_i, \quad i = 1, 2, \cdots, p$, as defined by (1.12). This result is presented in the following theorem

Theorem 2.1. *Under the conditions stated in equations (1.1) and (1.2) the density function of* $\hat{k}_i = \hat{\sigma}^2/\hat{\alpha}_i^2$ *is given by*

$$\begin{aligned} f(\hat{k}_i) &= \frac{e^{-\theta_i/2}\,(n-p)^{(n-p)/2}}{B(1/2,(n-p)/2)}\;\frac{(1/\lambda_i)^{(n-p+1)/2}(1/\hat{k}_i)^{3/2}}{((n-p)/\lambda_i+1/\hat{k}_i)^{(n-p+1)/2}}\;\times \\ & \sum_{j=0}^{\infty}\left[\frac{\theta_i/2\hat{k}_i}{((n-p)\lambda_i+1/\hat{k}_i)}\right]^j \frac{\Gamma((n-p+1)/2+j)\,\Gamma(1/2)}{\Gamma(j+1)\Gamma((n-p+1)/2)\Gamma(j+1/2)}, \\ & \hat{k}_i > 0, \quad i = 1, 2, \cdots, p; \end{aligned} \tag{2.1}$$

where

$$\theta_i = \frac{\alpha_i^2\,\lambda_i}{\sigma^2}. \tag{2.2}$$

Proof. From the definition of $\hat{k}_i$ and (1.2) we note that

$$\hat{k}_i = \frac{\lambda_i}{n-p}\quad\frac{u}{v_i}, \tag{2.3}$$

where

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

and

$$v_i = \left(\frac{\hat{\alpha}_i}{\sigma/\sqrt{\lambda_i}}\right)^2 \sim \chi^2_{(1)}(\theta_i), \tag{2.5}$$

a non-central Chi-square distribution with one degree of freedom and non-central parameter θ_i. This last result follows since the normality of the disturbances implies

$$\hat{\alpha}_i \sim N(\alpha_i, \sigma^2\lambda_i^{-1}), \tag{2.6}$$

and $\hat{\sigma}^2$ and $\hat{\alpha}_i$ are independent. Then

$$y = \frac{v_i}{u/(n-p)} = \lambda_i \frac{\hat{\alpha}_i^2}{\hat{\sigma}^2} \sim F_{(1,n-p)}(\theta_i\,, 0), \tag{2.7}$$

that is y is a non-central F with 1 and $n-p$ degrees of freedom and with θ_i and 0 as first and second non-central parameters respectively. From Johnson and Kotz (1970, page 191):

$$\begin{aligned} f(y) &= \frac{e^{-\theta_i/2}(n-p)^{(n-p)/2}}{B(1/2,(n-p)/2)} \frac{y^{-1/2}}{((n-p)+y)^{(n-p+1)/2}} \times \\ & \sum_{j=0}^{\infty} \left[\frac{\theta_i\, y\,/2}{((n-p)+y)}\right]^j \frac{\Gamma((n-p+1)/2+j)\Gamma(1/2)}{\Gamma(j+1)\Gamma((n-p+1)/2)\Gamma(1/2+j)}, y > 0, \end{aligned}$$

where

$$\Gamma(a) = \int_0^{+\infty} t^{a-1}\, e^{-t}\, dt \quad \text{and} \quad B(a,b) \;=\; \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{2.8}$$

But since $\hat{k}_i \;=\; \hat{\sigma}^2/\hat{\alpha}_i^2 \;=\; \lambda_i/y$, from a change of variables we find

$$\begin{aligned} f(\hat{k}_i) &= \frac{e^{-\theta_i/2}\,(n-p)^{(n-p)/2}}{B(1/2,(n-p)/2)} \frac{(\lambda_i/\hat{k}_i)^{-1/2}\lambda_i/\hat{k}_i^2}{((n-p)+\lambda_i/\hat{k}_i)^{(n-p+1)/2}} \times \\ & \sum_{j=0}^{\infty} \left[\frac{(\theta_i/2)\,\lambda_i(1/\hat{k}_i)}{((n-p)+\lambda_i/\hat{k}_i)}\right]^j \frac{\Gamma((n-p+1)/2+j)\,\Gamma(1/2)}{\Gamma(j+1)\Gamma((n-p+1)/2)\Gamma(j+1/2)}, \end{aligned}$$

which proofs the theorem. □

Theorem 2.2. *Under the conditions stated in equations (1.1) and (1.2) the distribution function of $\hat{k}_i = \hat{\sigma}^2/\hat{\alpha}_i^2$ is given by*

$$F(x_i) = 1 - \sum_{j=0}^{\infty} \frac{(\theta_i/2)^j e^{-\theta_i/2}}{j!\, B(1/2+j\,,(n-p)/2)} B_{r_i}(j+1/2\,,\,(n-p)/2), \tag{2.9}$$

where

$$B_r(a,b) = \int_0^r t^{a-1}(1-t)^{b-1} dt \qquad \textit{(Incomplete Beta)} \tag{2.10}$$

and

$$r_i = \frac{\lambda_i}{x_i(n-p)+\lambda_i}. \tag{2.11}$$

Proof. According to (2.7)

$$\begin{aligned} F(x_i) &= I\!\!P(\hat{k}_i \le x_i) \\ &= I\!\!P(\lambda_i/y \le x_i) \\ &= 1 - I\!\!P(y \le \lambda_i/x_i), \end{aligned} \tag{2.12}$$

where $y \sim F_{(1,n-p)}(\theta_i\,,0) = (n-p)G_{(1,n-p)}(\theta_i\,,0)$ and $G_{(1,n-p)}(\theta_i\,,0) = \chi^2_{(1)}(\theta_i)/\chi^2_{(n-p)}$ (see Johnson and Kotz (1970), page 191). Hence

$$\begin{aligned} I\!\!P(y < \lambda_i/x_i) &= I\!\!P((n-p)G < \lambda_i/x_i) \\ &= I\!\!P(G < \lambda_i/x_i(n-p)) \\ &= \int_0^{\lambda_i/x_i(n-p)} \left\{ \sum_{j=0}^{\infty} \left(\frac{(\theta_i/2)^j e^{-\theta_i/2}}{j!} \right) \times \right. \\ &\qquad \left. \frac{g^{j-1/2}}{B(1/2+j,(n-p)/2)(1+g)^{(n-p+1)/2+j}} \right\} dg \\ &= \sum_{j=0}^{\infty} \left(\frac{(\theta_i/2)^j e^{-\theta_i/2}}{j!B(1/2+j,(n-p)/2)} \times \right. \\ &\qquad \left. \int_0^{\lambda_i/x_i(n-p)} \frac{g^{j-1/2}}{(1+g)^{(n-p+1)/2+j}} dg \right). \end{aligned} \tag{2.13}$$

To evaluate the integral we make the following change of variables:

$$z = \frac{g}{1+g}, \quad \text{hence} \quad g = \frac{z}{1-z}, \quad dg = \frac{dz}{(1-z)^2} \quad \text{and}$$

$$\begin{aligned} \int_0^{\lambda_i/x_i(n-p)} & \frac{g^{j-1/2}}{(1+g)^{(n-p+1)/2+j}}\, dg \\ &= \int_0^{r_i} \frac{(z/(1-z))^{j+1/2}}{(1+z/(1-z))^{(n-p+1)/2+j}} \frac{1}{(1-z)^2} \quad dz \\ &= \int_0^{r_i} z^{(j+1/2)-1} \quad (1-z)^{(n-p)/2-1} \quad dz \\ &= B_{r_i}(j+1/2,(n-p)/2), \end{aligned} \tag{2.14}$$

replacing this in (2.13) and the result in (2.12) we obtain (2.9) □

3 Distribution of the Hoerl, Kennard and Baldwin Stochastic Shrinkage Parameter

We now consider the density and distribution function of $\hat{k}_{HKB}$, which according to (1.14) 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

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

and also

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

The density function is presented in the following theorem.

Theorem 3.1. *Under the conditions stated in equations (1.1) and (1.2), the density function of* $\hat{k}_{HKB} = p\hat{\sigma}^2/\underline{\hat{\beta}}'\underline{\hat{\beta}}$ *is given by*

$$f(\hat{k}_{HKB}) = \sum_{j=0}^{\infty} \frac{\Gamma(j+n/2)a_j[(n-p)/p]^{(n-p)/2}\hat{k}_{HKB}^{(n-p)/2-1}}{\Gamma((n-p)/2)\Gamma(p/2+j)\Delta^{p/2+j}[(n-p)\hat{k}_{HKB}/p\,+1/\Delta]^{j+n/2}},$$
$$\hat{k}_{HKB} > 0, \tag{3.4}$$

where Δ *is a number such that*

$$\mid c_i \mid = \mid 1-\Delta/h_i \mid < 1 \;,\;\; i = 1,2,...,p, \tag{3.5}$$

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

Proof. It has been shown,(see Firinguetti and Rubio (2000)), that under the stated conditions, the density function of $w = \underline{\hat{\beta}}'\underline{\hat{\beta}}/\sigma^2$ is given by

$$f(w) = \sum_{j=0}^{\infty} \frac{a_j}{\Gamma(p/2+j)(2\Delta)^{p/2+j}} w^{p/2+j-1}e^{-w/2\Delta} \;,\;\; w>0, \tag{3.6}$$

with

$$a_0 = e^{-d/2}\prod_{i=1}^{p}(\Delta/h_i)^{1/2} \;\; \textit{and} \;\; a_j = (2j)^{-1}\sum_{i=0}^{j-1} b_{j-i}a_i \;,\;\; j>0, \tag{3.7}$$

$$b_j = j\Delta \sum_{i=1}^{p} (\delta_i^2/h_i) c_i^{j-1} + \sum_{i=1}^{p} c_i^j \;\; , \;\; j > 0, \tag{3.8}$$

$$d = \sum_{i=1}^{p} \delta_i^2, \tag{3.9}$$

Also, according to (2.4) $u = (n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(n-p)}$, and since u and w are independent, their joint distribution is:

$$f(u,w) = \sum_{j=0}^{\infty} \frac{a_j \quad u^{((n-p)/2-1)} \, w^{(p/2+j-1)} \, e^{-(u/2+w/2\Delta)}}{\Gamma((n-p)/2)\Gamma(p/2+j)(2\Delta)^{p/2+j} 2^{(n-p)/2}} \, , u, w > 0. \tag{3.10}$$

But from (3.1), $\hat{k}_{HKB} = (p/(n-p))u/w$. Hence $u = ((n-p)/p)\, \hat{k}_{HKB}\, x$, $x = w$, and the jacobian is $|J| = ((n-p)/p)x$. Therefore

$$\begin{aligned} f(x, \hat{k}_{HKB}) \;\; &= \;\; \sum_{j=0}^{\infty} \frac{a_j((n-p)/p)^{(n-p)/2} \hat{k}_{HKB}^{(n-p)/2-1} x^{j+(n/2)-1}}{\Gamma((n-p)/2)\Gamma(p/2+j)(2\Delta)^{p/2+j} 2^{(n-p)/2}} \times \\ & \quad e^{-x((n-p)\hat{k}_{HKB}/2p+1/2\Delta)} \, , \quad x, \hat{k}_{HKB} > 0. \end{aligned} \tag{3.11}$$

Now integrating out x we obtain

$$\begin{aligned} f(\hat{k}_{HKB}) &= \int_0^{+\infty} f(x, \hat{k}_{HKB}) dx \\ &= \sum_{j=0}^{\infty} \frac{a_j (n-p)/p)^{(n-p)/2} \hat{k}_{HKB}^{((n-p)/2)-1} \Gamma(p/2+j)}{\Gamma((n-p)/2)\Gamma(p/2+j)\, \Delta^{p/2+j} [(n-p)\hat{k}_{HKB}/p + 1/\Delta]^{j+n/2}} \\ & \quad \times \int_0^{+\infty} \frac{[x(n-p)/2p + 1/2\Delta]^{j+n/2}}{\Gamma(j+n/2)} \times \\ & \quad x^{j+(n/2)-1} e^{-[(n-p)\hat{k}_{HKB}/2p+1/2\Delta]x} dx \, . \end{aligned} \tag{3.12}$$

Noting that the integral is a Gamma distribution with parameters $(j+n/2)$ and $[(n-p)\,\hat{k}_{HKB}/p\,+1/\Delta]/2$, the theorem is proven. □

We now turn to the distribution function of $\hat{k}_{HKB}$

Theorem 3.2. *Under the conditions stated in equations (1.1) and (1.2) the distribution function of* $\hat{k}_{HKB} = p\hat{\sigma}^2/\underline{\hat{\beta}}'\underline{\hat{\beta}}$ *is given by*

$$F(x) = \sum_{j=0}^{\infty} \frac{a_j \Gamma(j+n/2) B_r((n-p)/2, j+p/2)}{\Gamma((n-p)/2)\Gamma(p/2+j)} \quad x > 0\,, \tag{3.13}$$

where

$$\Delta_* = \Delta^{(j+n/2)}\left((p/(n-p))\Delta\right)^{(n-p)/2}, \tag{3.14}$$

$$r = [(n-p)/p]\Delta x, \tag{3.15}$$

and $B_r(a,b)$ *is the incomplete Beta function defined in (2.10).*

Proof.

$$\begin{aligned} F(x) &= \int_0^x f(\hat{k}_{HKB})\, d\hat{k}_{HKB} \\ &= \sum_{j=0}^{\infty} \frac{a_j((n-p)/p)^{(n-p)/2}\,\Gamma(j+n/2)}{\Gamma((n-p)/2)\Gamma(p/2+j)\Delta^{p/2+j}} \times \\ &\quad \int_0^x \hat{k}_{HKB}^{(n-p)/2-1}[(n-p)\hat{k}_{HKB}/p + 1/\Delta]^{-(j+n/2)} d\hat{k}_{HKB}. \end{aligned} \tag{3.16}$$

Now let

$$\begin{aligned} I &= \int_0^x \hat{k}_{HKB}^{(n-p)/2-1}[(n-p)\hat{k}_{HKB}/p + 1/\Delta]^{-(j+n/2)} d\hat{k}_{HKB} \\ &= (1/\Delta)^{-(j+n/2)} \int_0^x \hat{k}_{HKB}^{(n-p)/2-1}\,[(n-p)\Delta\hat{k}_{HKB}/p + 1]^{-(j+n/2)} d\hat{k}_{HKB}. \end{aligned}$$

Taking $u = (n-k)\Delta\hat{k}_{HKB}/p$, we have $du = [(n-p)\Delta/p]d\hat{k}_{HKB}$ and

$$I = \Delta^{(j+n/2)} \int_0^{\Delta x(n-p)/p} \left(\frac{pu}{(n-p)\Delta}\right)^{(n-p)/2-1} (u+1)^{-(j+n/2)} \frac{p}{(n-p)\Delta} du$$

$$= \Delta_* \int_0^{\Delta x(n-p)/p} u^{(n-p)/2}(u+1)^{-(j+n/2)} du,$$

where Δ_* is given in (3.14). Let $z = u/(1+u)$, $0 < z < 1$, then $u = z/(1-z)$, $du = dz/(1-z)^2$. Letting $r = [(n-p)/p]\Delta x$, we have:

$$
\begin{aligned}
I &= \Delta_* \int_0^r \left(\frac{z}{(1-z)}\right)^{(n-p)/2-1} \left(\frac{z}{1-z}+1\right)^{-(j+n/2)} \frac{dz}{(1-z)^2} \\
&= \Delta_* \int_0^r z^{(n-p)/2-1}(1-z)^{j+n/2-(n-p)/2-1} dz \\
&= \Delta_* \int_0^r z^{(n-p)/2-1}(1-z)^{j+p/2-1} dz \\
&= \Delta_* B_r((n-p)/2, j+p/2)\,, \qquad (3.17)
\end{aligned}
$$

where $B_r((n-p)/2, j+p/2)$ is the incomplete beta function defined in (2.10). Then replacing (3.17) in (3.16) we obtain the desired result.

□

4 Distribution of the Lawless and Wang Stochastic Shrinkage Parameter

To obtain the density and the distribution function we recall from (1.15) that

$$
\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\underline{\hat{\beta}}' X' X \underline{\hat{\beta}}} = \frac{p}{n-p}\frac{u}{v}\,, \qquad (4.1)
$$

where, according to (2.4), u is a central Chi-square distribution with (n-p) degrees of freedom and

$$
v = \frac{\underline{\hat{\beta}}' X' X \underline{\hat{\beta}}}{\sigma^2} \sim \chi^2_{(p)}(\theta), \qquad (4.2)
$$

that is v is distributed as a non-central Chi-square distribution with p degrees of freedom and non-central parameter $\theta = \underline{\hat{\beta}}' X' X \underline{\hat{\beta}}/\sigma^2$.
We now state the following result:

Theorem 4.1. *Under the conditions stated in equations (1.1) and (1.2) the density function of $\hat{k}_{LW} = p\hat{\sigma}^2/\underline{\hat{\beta}}' X' X \underline{\hat{\beta}}$ is given by*

$$
\begin{aligned}
f(\hat{k}_{LW}) &= \frac{e^{-\theta/2} p^{p/2} (n-p)^{(n-p)/2}}{B(p/2, (n-p)/2)} \quad \frac{\hat{k}_{LW}^{(n-p)/2-1}}{(\hat{k}_{LW}(n-p)+p)^{n/2}} \times \\
& \sum_{j=0}^{\infty} \left(\frac{p\theta/2}{(\hat{k}_{LW}(n-p)+p)}\right)^j \frac{\Gamma(n/2+j)\Gamma(p/2)}{\Gamma(j+1)\Gamma(n/2)\Gamma(p/2+j)} \qquad (4.3)
\end{aligned}
$$

$$\theta = \underline{\hat{\beta}}' X' X \underline{\hat{\beta}} / \sigma^2. \tag{4.4}$$

Proof. Since u and v are independent it follows that

$$\hat{k}_{LW} \sim F_{(n-p,p)}(0, \theta). \tag{4.5}$$

That is, $\hat{k}_{LW}$ is distributed as a non-central F with (n-p) and p degrees of freedom and non-central parameters 0 and θ (see Johnson and Kotz (1970, page 191)), and the theorem is proven. □

Finally we derive the distribution function of $\hat{k}_{LW}$.

Theorem 4.2. *Under the conditions stated in equations (1.1) and (1.2) the distribution function of* $\hat{k}_{LW} = p\hat{\sigma}^2 / \underline{\hat{\beta}}' X' X \underline{\hat{\beta}}$ *is given by*

$$\begin{aligned} F(x) &= 1 - \sum_{j=0}^{\infty} \frac{(\theta/2)^j \, e^{-\theta/2}}{\Gamma(j+1)} \frac{\Gamma(n/2+j)}{\Gamma(p/2+j)\Gamma(p/2+j)\Gamma((n-p)/2)} \times \\ & \quad B_\gamma(p/2+j\,;(n-p)/2)\,, \end{aligned} \tag{4.6}$$

where $B_\gamma(p/2+j\,;(n-p)/2)$ *is the incomplete beta function defined in (2.10) with* $\gamma = p/[(n-p)x+p]$.

Proof. Since $\hat{k}_{LW} \sim F_{n-p,p}(0,\theta)$ it follows that $(\hat{k}_{LW})^{-1} \sim F_{p,n-p}(\theta, 0)$, (see Johnson and Kotz, 1970). Hence:

$$\begin{aligned} F(x) &= \mathbb{P}(\hat{k}_{LW}) \leq x) \\ &= \mathbb{P}((\hat{k}_{LW})^{-1} \geq 1/x) \\ &= 1 - \mathbb{P}((\hat{k}_{LW})^{-1} < 1/x). \end{aligned} \tag{4.7}$$

But $F_{p,n-p}(\theta,0) = ((n-p)/p) G_{p,n-p}(\theta,0)$ and $G_{(p,n-p)}(\theta,0) = \chi^2_{(p)}(\theta)/\chi^2_{(n-p)}$ (see Johnson and Kotz (1970), page 191). Hence:

$$\begin{aligned} \mathbb{P}(1/\hat{k}_{LW} < 1/x) &= \mathbb{P}(G < p/(n-p)x) \\ &= \int_0^{p/(n-p)x} \frac{e^{-\theta/2}}{B(p/2,(n-p)/2)} \frac{g^{p/2-1}}{(1+g)^{n/2}} \times \\ & \quad \sum_{j=0}^{\infty} \left[\frac{g^{\theta/2}}{1+g}\right]^j \frac{\Gamma(n/2+j)\,\Gamma(p/2)}{\Gamma(n/2)\,\Gamma(j+1)\,\Gamma(p/2+j)} dg \\ &= \frac{e^{-\theta/2}}{B(p/2,(n-p)/2)} \sum_{j=0}^{\infty} (\theta/2)^j \frac{\Gamma(n/2+j)\,\Gamma(p/2)}{\Gamma(n/2)\Gamma(p/2+j)\Gamma(j+1)} \\ & \quad \times \int_0^{p/(n-p)x} \frac{g^{p/2+j-1}}{(1+g)^{n/2+j}} dg. \end{aligned} \tag{4.8}$$

Now let $y = g/(1+g)$ then $dg = dy/(1-y)^2$. Consequently:

$$\int_0^{p/(n-p)x} \frac{g^{p/2+j-1}}{(1+g)^{n/2+j}} dg = \int_0^{p/[(n-p)x+p]} \frac{(y/(1-y))^{p/2+j-1}}{(1+y/(1-y))^{n/2+j}(1-y)^2} dy$$
$$= \int_0^{p/[(n-p)x+p]} \frac{y^{p/2+j-1}}{(1-y)^{-(n-p)/2+1}} dy.$$

But from (2.10)

$$\int_0^{p/(n-p)x} \frac{g^{p/2+j-1}}{(1+g)^{n/2+j}} dg = B_\gamma(p/2+j,(n-p)/2), \tag{4.9}$$

where $\gamma = p/[(n-p)x+p]$. Finally, from (4.7), (4.8) and (4.9) we obtain (4.6). □

5 Numerical Results

In this section we carry out numerical calculations to provide some empirical evidence on the behavior of the distribution of the stochastic shrinkage parameters, under different model set ups. The different model specifications can be conveniently summarized by the signal to noise ratio:

$$\theta = \frac{\beta' X' X \beta}{\sigma^2}, \tag{5.1}$$

which is sensitive to the values of $X, \underline{\beta}$ and σ^2. In fact, for a given length of the coefficient vector, $\underline{\beta}$, one would expect θ to vary with the orientation of $\underline{\beta}$ to the eigenvectors of $X'X$. Moreover, since it is desirable for the shrinkage parameters to be larger the higher the degree of collinearity and/or the smaller the signal to noise ratio, it would be useful to know whether these parameters vary with the degree of collinearity and size of the signal to noise ratio. Consequently, we produced different model setups by varying the following factors:

i) We specified two X matrices, each with 5 explanatory variables, including a constant term, and 25 observations. To achieve different degrees of collinearity, the explanatory variables were generated using the following device (see Firinguetti and Rubio (2000)):

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

where

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

We then specified the following sets of a_j values:

$$\begin{aligned} A_1 &= (0.20, 0.30, 0.40, 0.50)\,; \\ A_2 &= (0.99, 0.95, 0.65, 0.60)\,. \end{aligned}$$

ii) The vector of coefficients, $\underline{\beta}$, was specified by:

$$\underline{\beta} = \frac{\sum_{i=1}^{p} \underline{q}_i}{\sqrt{p}},$$

where $Q = (\underline{q}_1, \underline{q}_2, \cdots\cdots, \underline{q}_p)$ is the matrix of eigenvectors of $X'X$, which are ordered such that the corresponding eigenvalues are: $\lambda_1 > \lambda_2 > \cdots > \lambda_p$. That is, $\underline{\beta}$ is a simple average of all eigenvectors of $X'X$, such that $\underline{\beta}'\underline{\beta} = 1$. This choice of $\underline{\beta}$ is prompted by the fact that the properties of the RR estimators are affected by the orientation of the parameter vector to the eigenvectors of $X'X$, and in practice is more likely that $\underline{\beta}$ depends on all the eigenvectors rather than on any one in particular.

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

To obtain the density of $\hat{k}_{HKB}$ it was necessary to specify $\Delta = 1.99h_p$, where h_p is the smallest eigenvalue of $X'(X'X)^{-2}X$. This value was chosen to accelerate convergence.

The results for the density of $\hat{k}_i$ are presented in Figure I. From these results the density function is noted to be highly dependent on the size of the corresponding eigenvalue. In fact, the larger the corresponding eigenvalue is, the heavier the tail of the distribution. That is to say, there is a greater chance of getting a larger value of $\hat{k}_i$, the larger λ_i is, which is an undesirable result.

The results for the densities of $\hat{k}_{HKB}$ and $\hat{k}_{LW}$ are presented in Figure II. From these results the behavior of the density of $\hat{k}_{HKB}$can be characterized by the following: It is rather insensitive to the values of β, σ^2 and θ. The only factor that appears to significantly affect the the density is the degree of collinearity. In fact $\hat{k}_{HKB}$ tend to be smaller for high collinearity and vice versa, which is contrary to expectations, since one would hope to shrink more when collinearity is high, particularly if the signal to noise ratio is not large.

The results for the density of $\hat{k}_{LW}$ merit the following comments: firstly, there is a slight increase in variability with increasing σ^2 and a higher degree of collinearity. More importantly the density appears to be significantly affected by the values of θ. In fact the higher θ is, the more concentrated is the density, and the more likely this rule will produce less shrinkage; the opposite is also true. Thus, for reasonable values of θ (not too large, not too small) the density indicates that $\hat{k}_{LW}$ can produce, with high probability, values of which are large enough to attain significant reductions in the MSE of the estimator of β.

6 Concluding Remarks

In this paper we set about to derive the density and distribution function of the shrinkage parameters proposed by Hoerl and Kennard (2000), Hoerl, Kennard and Baldwin (1975) and by Lawless and Wang (1976).

From these limited numerical results we found evidence that $\hat{k}_{LW}$ has some advantages over $\hat{k}_i$ and $\hat{k}_{HKB}$: firstly, there is a larger probability that $\hat{k}_{LW}$ will produce smaller values when less collinearity is present and the signal to noise ratio is larger, which is commendable since OLS is most likely a superior estimator when θ is large. Secondly, the distribution of $\hat{k}_{LW}$ will tend to produce larger shrinkage values than the distribution of $\hat{k}_{HKB}$ whenever θ is small. In fact the distribution of $\hat{k}_{HKB}$ is rather insensitive to the value of θ. Thirdly, the distribution of $\hat{k}_{LW}$ is more concentrated around the mean and mode than that of $\hat{k}_{HKB}$. Thus, $\hat{k}_{LW}$ appears to be more reliable than the other rules of shrinking, as it will tend to shrink more when there is a greater chance of reducing the MSE of the estimator of β.

ACKNOWLEDGEMENTS

This research was partially supported by FONDECYT-CHILE.

BIBLIOGRAPHY

Aldrin, M. (1997). "Length Modified Ridge Regression," *Computational Statistics and Data Analysis*, 25, 377-398.

Crivelli, A., Firinguetti, L., Montaño, R. and Muñoz, M. (1995). "Confidence Intervals in Ridge Regression By Bootstrapping the Dependent Variable: A Simulation Study", *Communications in Statistics: Simulation and Computation*, 24, 631-652.

Elston, D. and Proe, M. (1995). "Smoothing Regression Coefficients in an Overspecified Regression Model with Interrelated Explanatory Variables," *Applied Statistics*, 44, 395-406.

Firinguetti, L. and Rubio, H. (2000). "A Note on the Moments of Stochastic Shrinkage Parameters in Ridge Regression," *Communications in Statistics-Simulation and Computation*, 29, 955-970.

Foucart, T. (1999). "Stability of the Inverse Correlation Matrix. Partial Ridge Regression," *Journal of Statistical Planning and Inference*, 77, 141-154 .

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

Gruber, M. (1998)."*Improving Efficiency by Shrinkage,*" Marcel Dekker Inc., New York.

Gunst, R. (2000). "Classical Studies that Revolutionized the Practice of Regression Analysis," *Technometrics*, 42, 80-86.

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. (1970). "Ridge Regression: Applications to Nonorthogonal Problems," *Technometrics*, 12, 69-82.

Hoerl, A.E. and Kennard, R. W. (2000). "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 42, 80-86.

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

Inoue, Takakatsu (1999). "Density Function and Relative Efficiency of the Modified Generalized Ridge Regression Estimators," *Journal of the Japanese Statistical Society*, 29, 39-54.

Jang, D. and Yoon, M. (1997). "Graphical Methods for Evaluating Ridge Regression Estimator in Mixture Experiments," *Communications in Statistics-Simulation and Computation*, 26, 1049-1061 .

Johnson, N. and Kotz, S. (1970). "*Distributions in Statistic: Continous*

Univariate Distributions - 2," John Wiley and Sons, Inc., New York.

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

Kibria, B. (1996). "On Preliminary Test Ridge Regression Estimator for the Restricted Linear Model with Non-Normal Disturbances," *Communications in Statistics-Theory and Methods*, A25, 2349-2369.

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.

Saleh, A. and Kibria, B. (1993). "Performances of Some New Preliminary Test Ridge Regression Estimators and their properties," *Communications in Statistics-Theory and Methods*, A22, 2747-2764.

Shi, L. and Wang, X. (1999). "Local Influence in Ridge Regression," *Computational Statistics and Data Analysis*, 31, 341-353.

figure I

Density Function of the Hoerl and Kennard Estimator of the Generalized Shrinkage Parameter

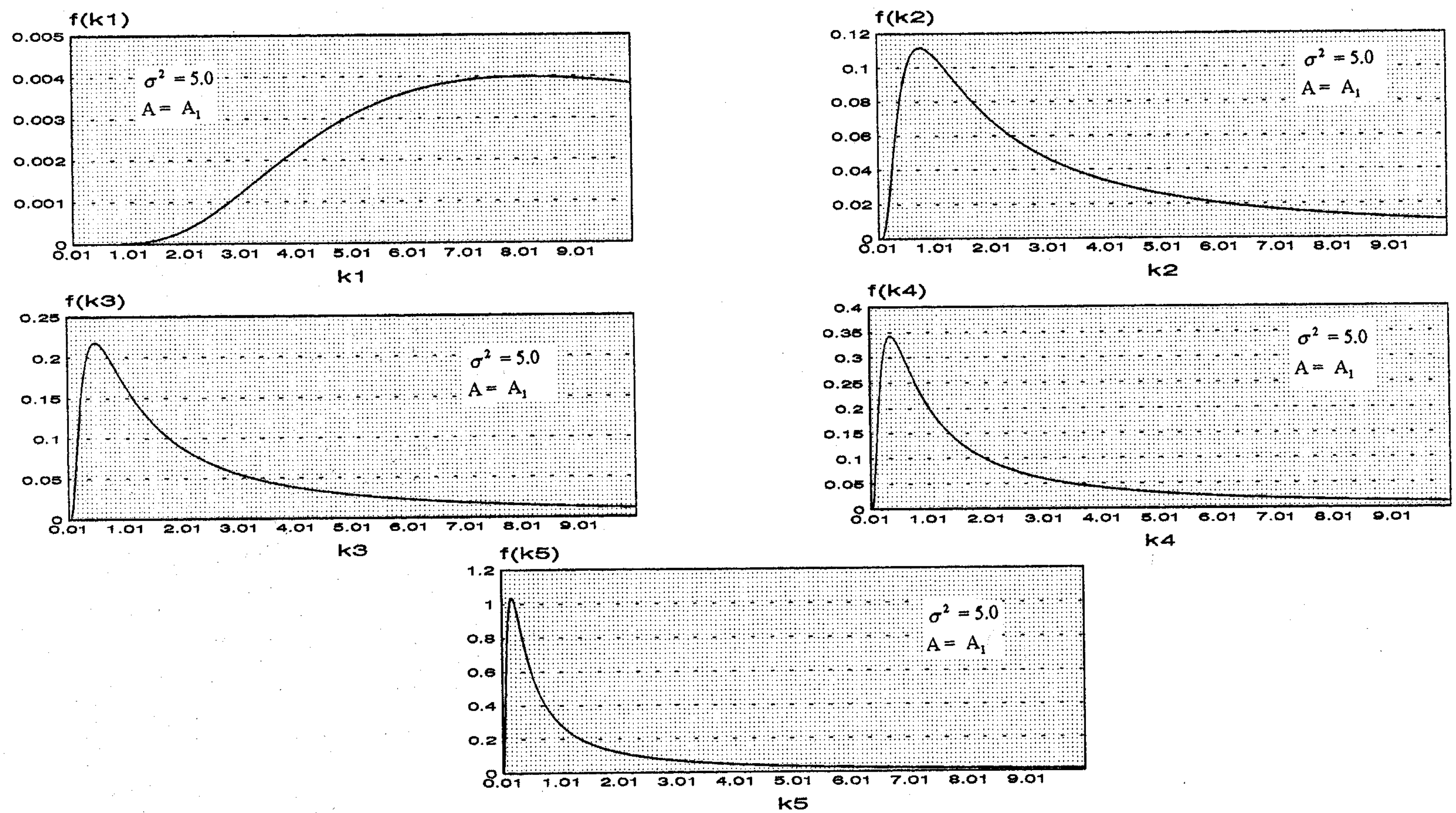


figure I continued

Density Function of the Hoerl and Kennard Estimator of the Generalized Shrinkage Parameter

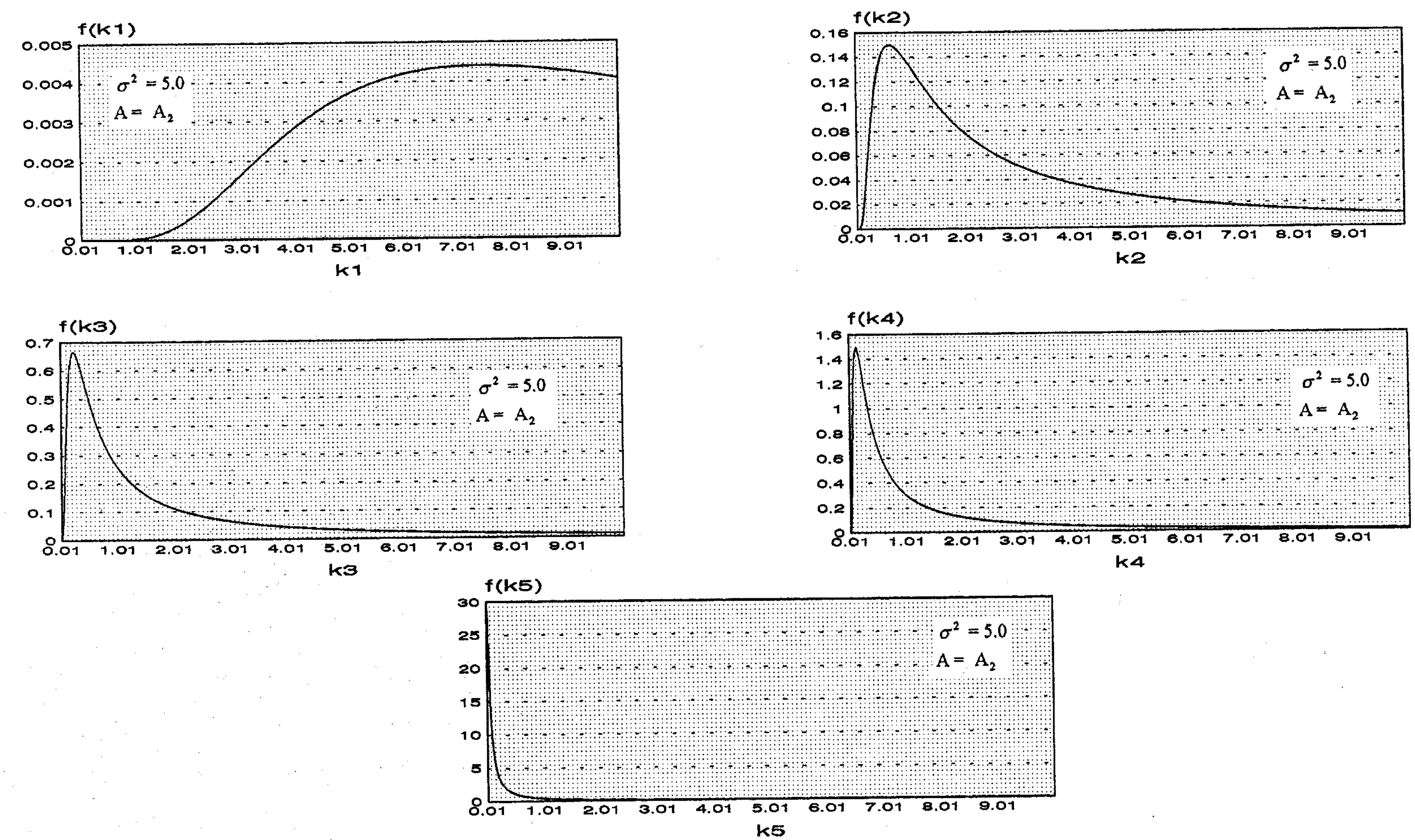
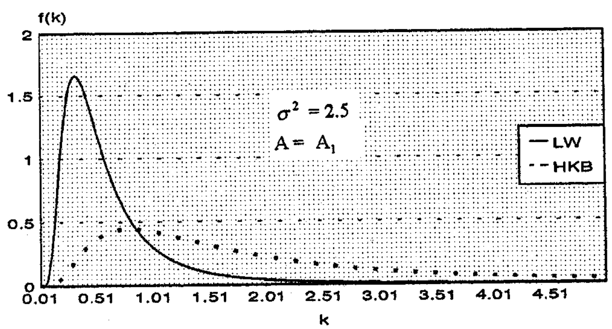
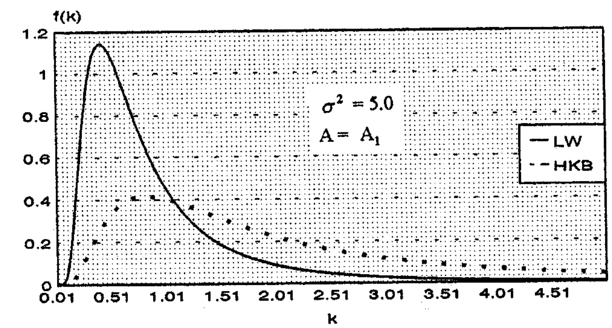


figure II

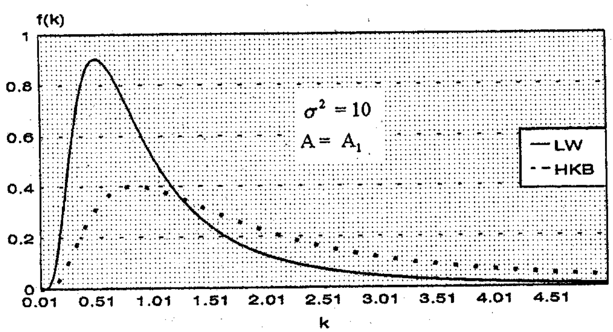
Density Function of the HKB and the LW Estimators of the Shrinkage Parameter



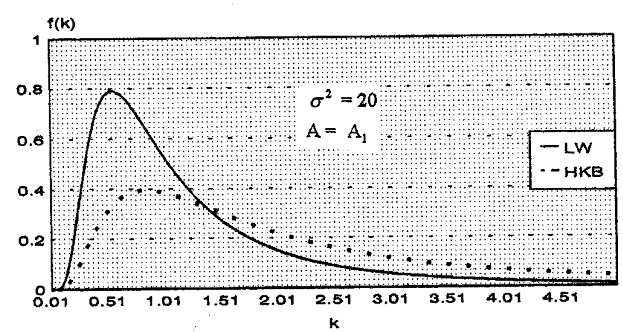
E(HKB_k)=2.57 E(LW_k)=0.67 V(HKB_k)=13.86 V(LW_k)=0.68 Theta=5.86



E(HKB_k)=2.85 E(LW_k)=0.99 V(HKB_k)=23.41 V(LW_k)=2.01 Theta=2.93



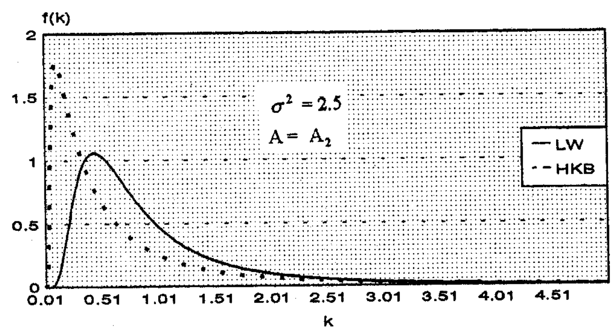
E(HKB_k)=3.01 E(LW_k)=1.27 V(HKB_k)=32.41 V(LW_k)=3.57 Theta=1.46



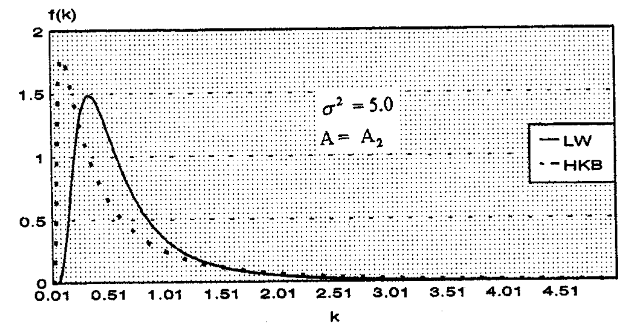
E(HKB_k)=3.11 E(LW_k)=1.45 V(HKB_k)=38.82 V(LW_k)=4.77 Theta=0.73

figure II continued

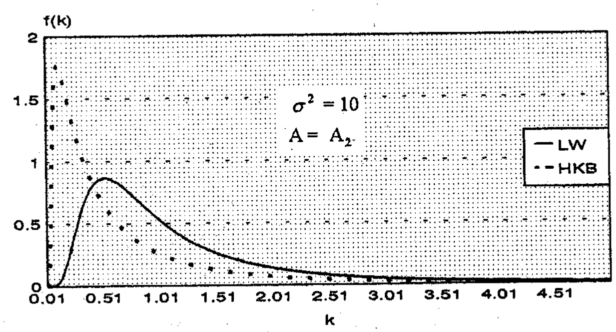
Density Function of the HKB and the LW Estimators of the Shrinkage Parameter



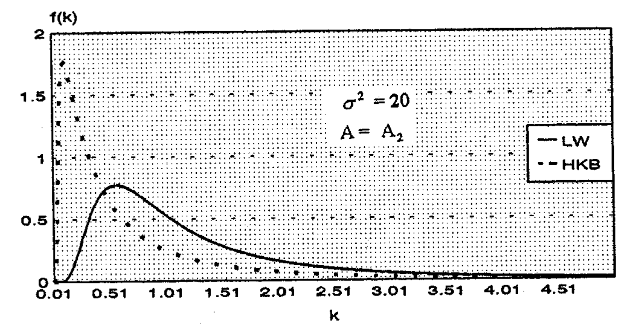
E(HKB_k)=0.87 E(LW_k)=1.07 V(HKB_k)=4.25 V(LW_k)=2.42 Theta=2.45



E(HKB_k)=0.83 E(LK_k)=0.75 V(HKB_k)=2.94 V(LW_k)=0.96 Theta=4.90



E(HKB_k)=0.90 E(LW_k)=1.32 V(HKB_k)=5.35 V(LW_k)=3.92 Theta=1.23



E(HKB_k)=0.91 E(LW_k)=1.48 V(HKB_k)=6.09 V(LW_k)=5.00 Theta=0.61

Documentos de Trabajo Banco Central de Chile	Working Papers 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. Printed 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 e-mail: bcch@condor.bcentral.cl

DTBC-136 Enero 2002
Value at Risk: Teoría y Aplicaciones
Christian A. Johnson

DTBC-135 Enero 2002
Agency Problems in the Solutions of Banking Crises
Gonzalo I. Sanhueza

DTBC-134 Enero 2002
On the Determinants of the Chilean Economic Growth
Rómulo A. Chumacero y J. Rodrigo Fuentes

DTBC-133 Diciembre 2001
Cálculo del Stock de Capital para Chile, 1985-2000
Ximena Aguilar M. y María Paz Collinao

DTBC-132 Diciembre 2001
Políticas de Estabilización en Chile durante los Noventa
Carlos José García T.

DTBC-131 Noviembre 2001
Ten Years of Inflation Targeting: Design, Performance, Challenges
Norman Loayza y Raimundo Soto

DTBC-130 Noviembre 2001
Trends and Cycles in Real-Time
Rómulo A. Chumacero y Francisco A. Gallego

DTBC-129 Noviembre 2001
Alternative Monetary Rules in the Open-Economy: A Welfare-Based Approach
Eric Parrado y Andrés Velasco

DTBC-128 Noviembre 2001
Price Inflation and Exchange Rate Pass-Through in Chile
Carlos José García y Jorge Enrique Restrepo

DTBC-127 Noviembre 2001
A Critical View of Inflation Targeting: Crises, Limited Sustainability, and Agregate Shocks
Michael Kumhof

DTBC-126 Noviembre 2001
Overshootings and Reversals: The Role of Monetary Policy
Ilan Goldfajn y Poonam Gupta

DTBC-125 Noviembre 2001
New Frontiers for Monetary Policy in Chile
Pablo S. García, Luis Oscar Herrera y Rodrigo O. Valdés

DTBC-124 Noviembre 2001
Monetary Policy under Flexible Exchange Rates: An Introduction to Inflation Targeting
Pierre-Richard Agénor

DTBC-123 Noviembre 2001
Targeting Inflation in an Economy with Staggered Price Setting
Jordi Galí

DTBC-122 Noviembre 2001
Market Discipline and Exuberant Foreign Borrowing
Eduardo Fernández-Arias y Davide Lombardo

DTBC-121 Noviembre 2001
Japanese Banking Problems: Implications for Southeast Asia
Joe Peek y Eric S. Rosengren

DTBC-120 Noviembre 2001
The 1997-98 Liquidity Crisis: Asia versus Latin America
Roberto Chang y Andrés Velasco

DTBC-119 Noviembre 2001
Politics and the Determinants of Banking Crises: The Effects of Political Checks and Balances
Philip Keefer