



UNIVERSIDAD DEL BÍO-BÍO
FACULTAD DE CIENCIAS
Concepción- Chile

Departamento de Estadística
Departamento de Matemáticas

Magister en Matemática Mención en Estadística

**ESTUDIO COMPARATIVO DE MÉTODOS
DE SELECCIÓN DE VARIABLES
CON ESTIMACIÓN SESGADA**

Tesis presentada al programa de Magister en Matemática, Mención en Matemática Aplicada o Mención en Estadística como parte de los requisitos para la obtención del grado de Magister en Matemática Mención Estadística de la Universidad del Bío-Bío

Enrique Ávila Coloma

Profesor Guía

Dr. Luis Firinguetti Limone

Concepción

2017



FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA

**“ESTUDIO COMPARATIVO DE MÉTODOS DE
SELECCIÓN DE VARIABLES CON ESTIMACIÓN
SESGADA”**

Autor: Enrique Ávila Coloma

Tesis Presentada Para Optar al Grado de Magister en Matemática
Mención en Estadística

Profesor Guía:

Dr. Luis Firinguetti Limone

Índice general

1. Aspectos Generales.	4
1.1. Aplicaciones de la Selección de Variables.	5
1.2. Selección de Variables en Situaciones Adversas.	5
1.3. Ámbito de Estudio	6
2. Regresión Lineal Clásica.	7
2.1. Supuestos del Modelo de Regresión Clásico.	8
2.2. Estimador de Mínimos Cuadrados Ordinarios.	9
2.2.1. Propiedades del Estimador Mínimos Cuadrados.	9
2.3. Multicolinealidad.	10
2.3.1. Consecuencias de la Multicolinealidad.	11
2.3.2. Detección de la Multicolinealidad.	11
3. Estimación Sesgada con Regresión Ridge.	14
3.1. Regresión Ridge.	14
3.1.1. Propiedades de la Regresión Ridge.	15
3.1.2. Estimadores del Parametro de Sesgo.	18
3.1.3. Regresión Ridge Restringida.	21
3.1.4. Regresión Ridge con Restricciones Estocásticas.	22
4. Selección de Variables con Estimación Sesgada.	24

Magister en Matemática Mención Estadística

4.1.	Selección de Variables con Regresión Ridge.	26
4.1.1.	Selección de Variables con Criterio AIC Ridge Doble.	29
4.1.2.	Selección de Variables con Criterio C_p Ridge Doble.	30
4.1.3.	Selección de Variables Bajo Estadístico de C_p Modificado.	32
4.1.4.	Selección de Variables Mediante Intervalos de Confianza.	33
5.	LASSO.	35
5.1.	Algoritmos de Optimización Propuesto por Tibshirani.	37
5.1.1.	Algoritmo 1.	38
5.1.2.	Algoritmo 2.	39
5.2.	Algoritmo LAR.	40
6.	Elastic Net.	42
6.1.	Elección de Parámetros de Ajuste.	45
7.	Estudio de Simulación.	46
7.1.	Objetivo de Desarrollo.	46
7.1.1.	Conteo de Modelos Seleccionados.	47
7.1.2.	Indicador de Calidad de Selección de Variables.	50
7.1.3.	Estimación de Coeficientes.	52
7.1.4.	Error Cuadrático Medio de Predicción Promedio.	54
7.2.	Comparación entre Métodos de Selección.	54
7.2.1.	Coeficientes Definidos por Valores Propios.	57
7.2.2.	Coeficientes Definidos Arbitrariamente.	103
7.2.3.	Comparación Gráfica de Métodos.	150
8.	Conclusiones y Trabajo Futuro.	162
8.1.	Respecto del Vector de Coeficientes.	162
8.2.	Respecto de la Varianza.	162
8.3.	Respecto de las Observaciones.	163

Magister en Matemática Mención Estadística

8.4. Respecto de la Multicolinealidad.	163
8.5. Respecto del Indicador F.	163
8.6. Respecto del Error Cuadrático Medio Total (ECMT).	164
8.7. Respecto del Error Cuadrático Medio de Predicción (ECMP).	164
8.8. Respecto de los Métodos de Selección de Variables.	164
8.9. Respecto del Diseño e Implementación de la Simulación.	165
8.10. Recomendaciones para utilización de criterios.	166
8.11. Trabajo Futuro.	167
Bibliografía	168

Capítulo 1

Aspectos Generales.

En la actualidad el desarrollo de tecnologías para la recopilación de datos ha tenido avances significativos aportando soluciones relevantes desde la perspectiva de captura de datos, pero esto a su vez genera nuevos desafíos estadísticos. Específicamente, en el resumen y la extracción de información se han y se están desarrollando métodos de reducción de las dimensiones y/o de selección de variables, que destacan por la búsqueda de mejores predicciones, modelos parsimoniosos y mejores estimaciones a la vez que se descarta información que en realidad no es relevante estadísticamente.

La presente tesis se concentra en el problema de la selección de variables, vale decir, suponga que se tiene una matriz X de dimensiones $n \times P$, donde las P columnas son posibles variables explicativas y cada una de ellas posee n observaciones. Se desea encontrar un modelo parsimonioso para explicar el comportamiento de una variable Y , que es un vector de dimensiones $n \times 1$. En otras palabras se desea seleccionar las p variables más relevantes desde las P posibles variables explicativas, esto implica necesariamente que $p < P$. Para resolver este problema es necesario utilizar criterios que permitan descartar variables irrelevantes y que permitan incorporar solo las variables que aportan con información relevante.

1.1. Aplicaciones de la Selección de Variables.

Los métodos de selección de variables van más allá de la estadística y el modelamiento de datos. Resulta que existen aplicaciones en diversas áreas, por ejemplo en el sector industrial la preocupación va más por el lado de la información redundante y variables que en realidad incorporan ruido lo que implica mal uso de las herramientas estadísticas. Según Anzanello y Fogliatto (2014) en la industria química, por ejemplo, se ocupan en procesos de refinería, en industria de fertilizantes hasta en el reciclado de papel.

Otras aplicaciones industriales fueron mencionadas por Smits y otros (2006) y estas van por el lado de descubrir maneras de predecir variables de procesos (Salidas) difíciles de medir como emisiones de NOx (Oxidos Nitrosos), propiedades especiales de polímeros o medidas sobre Biomasa tomando en cuenta la información capturada por otras variables (Entradas) que son comunes y sencillas de medir como Flujo, Temperatura, Ph, Presión, etc.

Por otra parte técnicas más reconocidas como LASSO (técnica que será profundizada en el Capítulo 5), no solo buscan los mismos objetivos mencionados anteriormente, sino que además sus aplicaciones con diversos tipos de datos lo convierten en una técnica muy versátil. Chatterjee y otros (2012), por ejemplo, estudiaron las aplicaciones de LASSO en datos climatológicos, mostrando resultados significativamente mejores que el actual estado del arte.

1.2. Selección de Variables en Situaciones Adversas.

Ahora bien, dentro del problema de selección de variables se pueden encontrar varios sub-problemas que abordar. Por ejemplo, ¿Qué pasa cuando la cantidad de variables dis-

ponibles es superior a la cantidad de observaciones?, o ¿Qué tal si la cantidad de variables es menor que la cantidad de observaciones?, ¿Qué pasa si las variables están correlacionadas (Multicolinealidad)? o ¿Qué pasa si la correlación es solo entre variables relevantes o si la multicolinealidad incorpora a las variables irrelevantes?.

La verdad es que no existe una manera o método para solucionar todos los problemas, más bien, los métodos existentes se concentran en solucionar un problema particular de la mejor manera posible. La clave está en identificar claramente el problema para aplicar la solución adecuada, de modo de obtener el mejor resultado posible.

1.3. Ámbito de Estudio

La presente tesis es un estudio de simulación para evaluar el comportamiento de varios métodos de selección de variables, en distintos escenarios, que están caracterizados por la cantidad de variables disponibles, la correlación entre variables (multicolinealidad), la varianza y la cantidad de observaciones.

Los resultados son estudiados desde el punto de vista de las predicciones hechas por el modelo seleccionado, las estimaciones de los coeficientes y la capacidad de seleccionar el modelo adecuado.

Capítulo 2

Regresión Lineal Clásica.

Con frecuencia se consideran situaciones en las que se posee una cantidad importante de variables, con sus respectivas observaciones y es razonable esperar algún tipo de relación entre ellas. En ese contexto el Modelo de Regresión Lineal Múltiple (en adelante MRLM), que se caracteriza por poseer una estructura lineal (desde el punto de vista de los parámetros a estimar), permite establecer matemáticamente la relación existente entre la variable explicada (o variable endógena), con otras variables a las que se les llama variables explicativas (o variables exógenas).

El modelo descrito posee la siguiente estructura

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon, \quad (2.1)$$

donde $\beta_0, \beta_1, \dots, \beta_p$: son parámetros fijos desconocidos. X_1, \dots, X_p : son las variables explicativas y no son aleatorias. ϵ : es un vector de dimensiones $n \times 1$ con $\epsilon \sim N(\underline{0}, \sigma^2 I)$, la variable explicada Y es un vector de dimensiones $n \times 1$, cuya representación matricial tiene la siguiente estructura

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}, \quad (2.2)$$

donde X : es una matriz que se compone de las variables explicativas y es de dimensiones $n \times (p + 1)$. $\underline{\epsilon}$: es un vector de dimensiones $n \times 1$ con $\underline{\epsilon} \sim N(\underline{0}, \sigma^2 I)$. La variable explicada \underline{Y} es un vector de dimensiones $n \times 1$. El vector $\underline{\beta}$ es de dimensiones $(p + 1) \times 1$ y contiene los coeficientes desconocidos.

2.1. Supuestos del Modelo de Regresión Clásico.

Cuando se construye un MRLM, este debe cumplir con varios supuestos, de ese modo se garantizan las propiedades de los estimadores utilizados, estos supuestos deben ser siempre evaluados ya que el incumplimiento de alguno de ellos puede motivar la realización de transformaciones sobre el MRLM o incluso puede motivar otra estrategia de modelamiento. Los supuestos para el MRLM son:

- 1) La relación entre la variable explicada y las variables explicativas es lineal, pero desde el punto de vista de los parámetros.
- 2) La variable explicada \underline{Y} es aleatoria, en cambio las variables explicativas no lo son, más bien están preestablecidas.
- 3) No existe multicolinealidad perfecta entre las variables explicativas, es decir se requiere que ellas sean linealmente independientes, esto implica que el número de variables sea menor que la cantidad de observaciones, vale decir $p < n$
- 4) El modelo está correctamente especificado, es decir, que no posee variables irrelevantes, ni se han excluido variables relevantes.
- 5) El error es una perturbación aleatoria que debe cumplir con las siguientes condiciones: $E(\epsilon_i) = 0$ y la $Var(\epsilon_i) = \sigma^2$, es decir, su varianza es constante (Homocedasticidad),

además no debe estar autocorrelacionado ($Cov(\epsilon_i, \epsilon_j) = 0$) y por último es requerido que $\epsilon \sim N(\underline{0}, \sigma^2 I)$.

2.2. Estimador de Mínimos Cuadrados Ordinarios.

En este contexto se vuelve necesario un método para la estimación de los parámetros desconocidos del modelo, bajo algún determinado criterio. El Método de Mínimos Cuadrados Ordinarios (en adelante MCO) propone calcular los coeficientes que minimicen el cuadrado de la diferencia entre el valor observado y el estimado, vale decir

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij})^2 \right\}, \quad (2.3)$$

donde X_{ij} : es la observación i de la variable explicativa j . β_j : es el coeficiente estimado j , de la j -ésima variable explicativa. y_i : es la observación i de la variable explicada.

El resultado de este problema de minimización es una ecuación matricial cerrada, que calcula los coeficientes del modelo y que minimiza el error cuadrático, y que se le llama Estimador de Mínimos Cuadrados Ordinarios (en adelante EMCO) y está dada por

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y, \quad (2.4)$$

donde $\hat{\beta}_{MCO}$: es el EMCO y es un vector de dimensiones $p \times 1$. La matriz X contiene las variables explicativas y es de dimensiones $n \times (p + 1)$ porque además de las p variables agrega una columna que contiene unos (específicamente en la columna 1). La variable explicada Y es de dimensiones $n \times 1$.

2.2.1. Propiedades del Estimador Mínimos Cuadrados.

1) El EMCO ($\hat{\beta}_{MCO}$) es un estimador insesgado, es decir

$$E(\hat{\beta}_{MCO}) = \beta. \quad (2.5)$$

2) De acuerdo al Teorema de Gauss-Markov, dentro de la clase de estimadores lineales e insesgados, el EMCO posee varianza mínima lo que constituye una propiedad importante del estimador ya que garantiza un cálculo adecuado y además preciso de los coeficientes.

3) La varianza del vector de coeficientes $\hat{\beta}_{MCO}$ es

$$Var(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}. \quad (2.6)$$

4) La distribución del vector $\hat{\beta}_{MCO}$ es

$$\hat{\beta}_{MCO} \sim N_{p+1}(\beta, \sigma^2(X'X)^{-1}), \quad (2.7)$$

En particular se estudian las consecuencias de la multicolinealidad, en un contexto en el que se busca descartar variables irrelevantes y seleccionar aquellas variables que poseen información importante para construir un modelo. Lo que motiva la utilización de un estimador que mitigue las consecuencias de la multicolinealidad y permita elegir las variables que aportan más información a la variable explicada.

2.3. Multicolinealidad.

La multicolinealidad es un fenómeno que se da con mucha frecuencia, cuando se pretende establecer un modelo de Regresión Lineal Múltiple, esto debido a que si se poseen muchas variables explicativas, es posible que comience a manifestarse relaciones entre ellas y que estas sean incluso más fuertes que la relación que se espera con la variable explicada. Las consecuencias de la presencia de este fenómeno son variadas pero se vinculan principalmente a la varianza de las estimaciones lo que implica una falta de precisión considerable en las estimaciones, estas son mencionadas en Aparicio y otros (2003).

2.3.1. Consecuencias de la Multicolinealidad.

i. Cuando se presenta una multicolinealidad fuerte, las columnas de la matriz X no son linealmente independientes, esto conlleva a que la matriz $(X'X)$ no tenga rango igual a p (número de variables explicativas), implicando finalmente que $(X'X)^{-1}$ no existirá y por tanto $\hat{\beta}$ tampoco existe.

ii. La siguiente consecuencia de la multicolinealidad imperfecta es el aumento en las varianzas y covarianzas en los estimadores MCO.

iii. Los parámetros estimados pueden tener el signo incorrecto.

iv. El aumento en los errores estándar en las estimaciones, puede dar como resultado parámetros que no son significativamente distintos de cero.

v. Los intervalos de confianza tenderán a ser muy amplios, por lo que con frecuencia contienen el cero. Esto está directamente relacionado con la inflación de la varianza y la pérdida de precisión en las estimaciones.

vi. Es interesante mencionar que la multicolinealidad no afecta el ajuste global del modelo dado por el coeficiente de determinación (R^2).

2.3.2. Detección de la Multicolinealidad.

Para detectar este fenómeno, los análisis existentes se concentran en el estudio de los síntomas de la colinealidad o multicolinealidad, vale decir, como afecta las varianzas de las estimaciones, también se realizan estudios de la matriz de correlaciones y sus valores propios, etcétera. A continuación se mencionan algunas de las técnicas más populares para

la detección de esta situación:

i. La forma más elemental de detectar colinealidad entre pares de variables, es un **Gráfico de Dispersión** entre las variables explicativas, ya que un síntoma de la presencia de Colinealidad, es una correlación alta entre variables, lo que implica un gráfico de dispersión con tendencia lineal como muestra la Figura N°1.

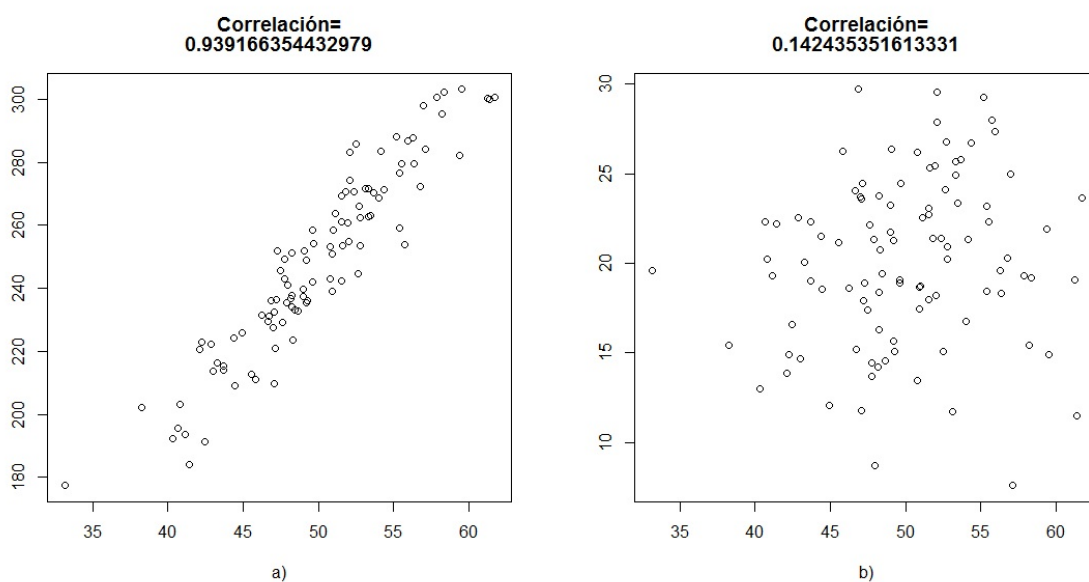


Figura 2.1: En ambos gráficos se contrasta la relación entre dos variables explicativas. En a) se aprecia una fuerte correlación que puede ser síntoma de colinealidad. Mientras que el gráfico en b) muestra una baja correlación lo que indica independencia entre variables, ambos gráficos fueron simulados y elaborados en software R, Fuente: Elaboración propia.

ii. Un método para detectar la presencia de multicolinealidad consiste en tomar los valores propios de la matriz $X'X$ para calcular el **Número de Condición**, que se define por

$$k = \frac{\lambda_{max}}{\lambda_{min}}. \tag{2.8}$$

donde: λ_{max} es el valor propio más grande mientras que λ_{min} es el valor propio más pe-

queño de la matriz $X'X$.

Si el número de condición es menor a 100 no hay problemas de multicolinealidad, si está entre 100 y 1000 implican multicolinealidad moderada, números de condición mayores a 1000 muestran una Muticolinealidad severa.

iii. Otra metodología muy ocupada es la del **Factor de Influencia de la Varianza** (FIV_j en español y VIF_j en inglés), que permite revisar los efectos de en la varianza de las estimaciones cuando existe multicolinealidad.

La ecuación es $VIF_j = FIV_j = \frac{1}{1-R_j^2}$, con $j = 1, 2, \dots, p$, R_j^2 el coeficiente de determinación para la regresión sobre las restantes $(p-1)$ covariables, si la correlación es alta entre el regresor y los demás $(p-1)$ regresores entonces $R_j^2 \approx 1$, lo que aumenta el valor del FIV.

Valores superiores a 10 del FIV indican multicolinealidad, pero este valor debe ser contrastado con el equivalente del modelo ajustado $\frac{1}{1-R^2}$, donde R^2 es el coeficiente de determinación del modelo, si el valor FIV es mayor al equivalente del modelo, indica que existe una relación más potente entre las variables explicativas que entre las variables explicativas y la Explicada.

Capítulo 3

Estimación Sesgada con Regresión Ridge.

Bajo la presencia de multicolinealidad severa, se originan problemas de precisión en las estimaciones y como se mencionó en el ítem 2.3.1 incluso podría darse el caso de la inviabilidad de calcular la matriz $(X'X)^{-1}$. En este escenario Hoerl y Kennard (1970) propusieron un estimador que soluciona el problema de la multicolinealidad pero que introduce un sesgo en la estimación pero que puede disminuir considerablemente la varianza.

3.1. Regresión Ridge.

El concepto de la Regresión Ridge (en adelante RR) radica en una idea similar al problema de minimización en mínimos cuadrados (ver ecuación 2.3), pero con la diferencia que se está sujeto a una restricción que limita el tamaño de los valores de los coeficientes estimados, dada por las siguientes expresiones

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \right\} \quad (3.1)$$

sujeta a la restricción

$$\sum_{j=1}^p \beta_j^2 \leq s, \quad (3.2)$$

donde x_{ij} es la observación i de la variable explicativa j . β_j es el Coeficiente de la variable explicativa j , β_0 es el Intercepto, y_j es la observación i de la variable explicada, n es la cantidad de observaciones, p es la cantidad de variables explicativas, y s es el tamaño de la restricción.

El resultado de este problema de minimización es una ecuación matricial cerrada, que calcula los coeficientes del modelo y que minimiza el error cuadrático, y a la cual se le llama Estimador de Regresión Ridge (en adelante ERR) y está dada por la siguiente ecuación

$$\hat{\beta}_R = [(X'X) + kI]^{-1}X'Y, \quad (3.3)$$

donde X es la matriz que contiene las variables explicativas y sus dimensiones son $n \times p$, Y es el vector que contiene las observaciones de la variable explicada y sus dimensiones son $n \times 1$, k es el parámetro de sesgo y $\hat{\beta}_R$ es el ERR.

3.1.1. Propiedades de la Regresión Ridge.

1) Cuando se propuso la RR, se discutieron las propiedades de este estimador. Hoerl y Kennard (1970) exponen un estimador para la **Matriz de Varianzas y Covarianzas** de $\hat{\beta}_R$ dada por

$$VAR(\hat{\beta}_R) = \sigma^2 Z(X'X)^{-1}Z', \quad (3.4)$$

Donde: $Z = [I_p + k(X'X)^{-1}]^{-1}$, σ^2 es la varianza del error e I_p es la matriz identidad de dimensiones $p \times p$.

Magister en Matemática Mención Estadística

Por simplicidad matemática ocuparemos una ecuación distinta, a la cual se puede llegar mediante algebra de matrices. Esta tiene la ventaja de evitar términos traspuestos lo que reduce los cálculos computacionales. La expresión para la varianza es:

$$VAR(\hat{\beta}_R) = \sigma^2(X'X + kI)^{-1}X'X(X'X + kI)^{-1}. \quad (3.5)$$

2) Como se demuestra en Tusell (2011) el **Error Cuadrático Medio** (en adelante ECM) se compone de la suma de la varianza y el cuadrado del sesgo, también posee una componente de la covarianza pero como se asume que no existe relación entre los errores, el valor de este componente es cero. La expresión para calcular el ECM cuando se trabaja con ERR (ver ecuación 3.6), fue demostrada por Hoerl y Kennard (1970), y de su análisis se desprenden características interesantes que debe poseer el Parámetro de Sesgo. El error cuadrático medio cuando se estima con RR es

$$ECM(\hat{\beta}_R) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^p \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2}, \quad (3.6)$$

donde λ_j es el valor propio j , asociado a la variable explicativa j , k es el parámetro de sesgo. p es la cantidad de coeficientes a estimar y σ^2 es la varianza del error. además $\alpha = V\beta$, donde V es una matriz con los vectores propios de $X'X$. Es de notar que $\lambda_j > 0$ para todo i y $k \geq 0$, y para cada elemento $\lambda_j + k$ es positivo.

Para facilidad de estudio expresaremos el $ECM(\hat{\beta}_R)$ como función de $\gamma_1(k)$ y $\gamma_2(k)$ dados por

$$ECM(\hat{\beta}_R) = \gamma_1(k) + \gamma_2(k), \quad (3.7)$$

donde

$$\gamma_1(k) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}, \quad (3.8)$$

$$\gamma_2(k) = \sum_{j=1}^p \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2}. \quad (3.9)$$

Magister en Matemática Mención Estadística

Si se considera la expresión de la varianza (dada por $\gamma_1(k)$), y se deriva respecto de k (ver ecuación 3.10) se evidencia que el valor es negativo, esto indica una pendiente negativa en la razón de cambio, es decir que la varianza disminuye en la medida que aumenta k .

$$\frac{d(\gamma_1(k))}{dk} = -2\sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^3}. \tag{3.10}$$

Por otra parte si se deriva la expresión del sesgo (dada por $\gamma_2(k)$) respecto de k (ver ecuación 3.11), se obtiene

$$\frac{d(\gamma_2(k))}{dk} = 2k \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^2} - 2k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^3}, \tag{3.11}$$

La Figura N°1 contrasta la evolución del ECM, el sesgo y la varianza a medida que aumenta el valor de k .

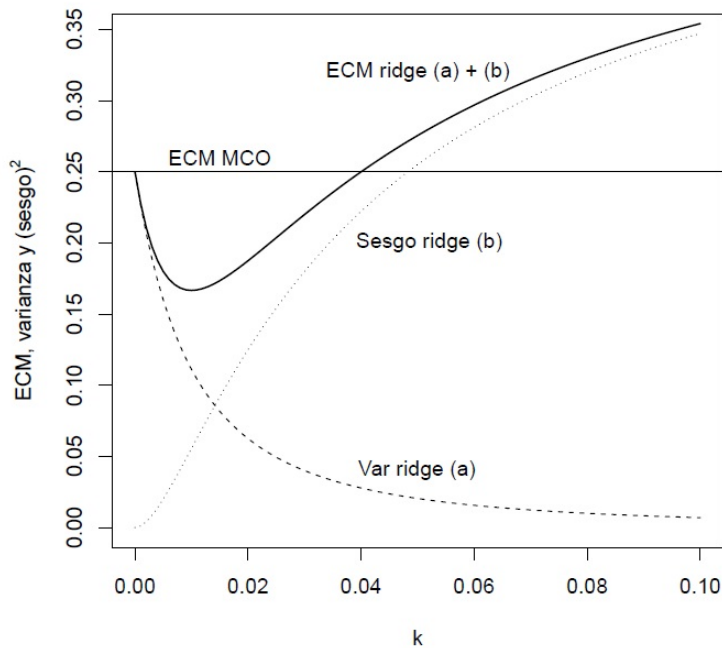


Figura 3.1: Componentes del $ECM(\hat{\beta}_R)$ en el ERR. Las líneas de trazos y puntos representan respectivamente la varianza y sesgo de $\hat{\beta}_R$ en función de k . La curva sólida representa $ECM(\hat{\beta}_R)$, La línea horizontal es la varianza y ECM del EMCO($\hat{\beta}_{MCO}$),Tusell(2011)

Lo que se desprende de este análisis es la necesidad de calcular un valor para el parámetro de sesgo que sea pequeño para no introducir demasiado sesgo y que a la vez minimice la varianza. Este criterio es formalizado por Hoerl y Kennard (1970) por medio de un teorema es establece que:

Teorema: "Siempre existe un valor de $k > 0$ tal que $ECM(\hat{\beta}_R) < ECM(\hat{\beta}_{MCO})$ "

(3.12)

Este teorema es sencillo de comprobar, primero tomaremos en cuenta que

$$ECM(\hat{\beta}_{MCO}) = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j} \tag{3.13}$$

y que el teorema establece que

$$ECM(\hat{\beta}_{MCO}) > ECM(\hat{\beta}_R), \tag{3.14}$$

vale decir,

$$\sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j} > \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^p \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2}. \tag{3.15}$$

Se puede ver que si $k = 0$, $ECM(\hat{\beta}_{MCO})=ECM(\hat{\beta}_R)$. Por otra parte, que para que se cumpla el teorema, se debe encontrar un valor de k que disminuya considerablemente la varianza de modo que el sesgo introducido sea manejable, esto implica una disminución considerable en el ECM.

3.1.2. Estimadores del Parametro de Sesgo.

Calcular el Parámetro de Sesgo es una tarea delicada, pues de él depende el buen funcionamiento del ERR. Se han propuesto muchas opciones de estimación desde la creación de la RR. Es importante señalar que está fuera del alcance de esta tesis evaluar todas las propuestas existentes, y en ese contexto se acota la evaluación del ERR a solo tres técnicas para calcular el parámetro de sesgo.

Para un mayor entendimiento de esta etapa, se considerará el artículo de Dorugade y Kashid (2010) donde se proponen algunos estimadores para la Regresión Ridge Generalizada y mediante un estudio de simulación evaluaron su comportamiento. Consideramos la Regresión Lineal Clásica modificada como

$$Y = Z\alpha + \epsilon, \quad (3.16)$$

donde: $Z = XV$, y X es la matriz que contiene las variables explicativas y V es una matriz ortogonal tal que $V'(X'X)V = \Lambda$ y Λ es una matriz diagonal que contiene a los valores propios de la matriz $X'X$ tal que $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$. Por su parte $\alpha = V^{-1}\beta$, donde β es el vector de parámetros desconocidos, es necesario establecer un estimador para α , la propuesta por MCO está dado por

$$\hat{\alpha}_{MCO} = (Z'Z)^{-1}Z'Y. \quad (3.17)$$

pero este estimador para α contiene los problemas generados por la multicolinealidad es por ello que se propone un estimador basado en RR para corregir el problema dado por

$$\hat{\alpha}_{RR} = (Z'Z + K)^{-1}Z'Y, \quad (3.18)$$

donde: $K = \text{diag}(k_1, k_2, \dots, k_p)$, $k_i > 0$, pero este estimador no es operacional ya que se desconocen los valores de k_i , a continuación se describen tres estimadores.

3.1.2.1. Propuesta de Hoerl, Kennard y Baldwin

Cuando se propuso la RR como alternativa a la Regresión Lineal para resolver los problemas originados por la Multicolinealidad, Hoerl y Kennard (1970) se dieron cuenta que necesitaban calcular de alguna manera el valor de k . Debido a que k es un parámetro desconocido el estimador propuesto es no operacional, por lo tanto k debe ser estimado.

Siguiendo el criterio del teorema 3.12, se construyó un estimador para el parámetro de sesgo que minimice el $ECM(\hat{\beta}_R)$ y que además permite que se cumpla que el $ECM(\hat{\beta}_{MCO}) >$

$ECM(\hat{\beta}_R)$. El estimador propuesto es

$$k_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{max}^2} \quad (3.19)$$

donde: α_{max} es el valor máximo contenido en el vector α .

El estimador propuesto k_{HK} es un estimador para la Regresión Ridge Ordinaria, es decir, se obtiene un único valor que ayuda a compensar el problema de la multicolinealidad.

Pero algunos años después Hoer, Kennard y Baldwin (1975) propusieron otro estimador, que se calcula a partir de la información existente en los coeficientes calculados por mínimos cuadrados y la varianza estimada del modelo, que básicamente es la media armónica entre los diferentes valores de k_i , dada por la siguiente expresión

$$k_{HKB} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \hat{\alpha}_i^2} = \frac{p\hat{\sigma}^2}{(\hat{\beta}_{MCO})'(\hat{\beta}_{MCO})} \quad (3.20)$$

Donde: α_i son elementos de α que se calculan con la ecuación 3.17, σ^2 es la varianza que se estima con $\hat{\sigma}^2 = \frac{1}{n-p} \sum \hat{e}_i^2$.

3.1.2.2. Propuesta de Dorugade.

La propuesta de Dorugade (2014) está basada en las ideas de Kibria (2003). El realizó un estudio basado en simulaciones computacionales, donde se contrasta 4 propuestas hechas por Dorugade v/s Las propuestas de otros autores, incluidas las propuestas de Kibria (2003) y Hoerl y Kennard (1970). De las 4 propuestas presentadas, la que dio mejores resultados fue un estimador basado en la media armónica, dada por

$$k_{DK} = \frac{2p}{\lambda_{max}} \sum_{i=1}^p \frac{\hat{\sigma}^2}{(\hat{\alpha}_i^2)} \quad (3.21)$$

3.1.2.3. Propuesta de Lawless y Wang.

La propuesta desarrollada por Lawless y Wang (1976), se construyo desde una perspectiva Bayesiana

$$k_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p (\lambda_i \hat{\alpha}_i^2)} \quad (3.22)$$

Actualmente es un reconocido estimador que incluso está implementado, junto al estimador de Hoerl, Kennard y Baldwin en R, para los paquetes que implementan la RR.

3.1.3. Regresión Ridge Restringida.

Como punto de partida se considera la propuesta de Özkale (2009), la que relaciona el EMCO y a través de él se construye el Estimador Ridge Restringido (en adelante ERRR).

El EMCO es insesgado y de mínima varianza, aunque posee estas características, cuando se presenta multicolinealidad el EMCO pierde considerable precisión, ya que la varianza asociada a la estimación de cada coeficiente aumenta. Para solucionar el problema se puede considerar la incorporación de información disponible en forma de restricciones exactas. Esto podría proveer de mejores estimaciones que con el EMCO. Consideremos las restricciones con la forma

$$r = R\beta \quad (3.23)$$

con r vector de dimensión $(l \times 1)$ y R es una matriz de dimensiones $(l \times p)$, con $l \leq p$, aplicando la restricción al EMCO se obtiene el estimador restringido de Mínimos Cuadrados (en adelante EMCR)

$$\hat{\beta}_{MCR} = \hat{\beta}_{MCO} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta}_{MCO}), \quad (3.24)$$

con $\hat{\beta}_{MCR}$ es el EMCR y $\hat{\beta}_{MCO}$ es el EMCO. X : Matriz que posee las variables explicativas y es de dimensiones $n \times p$. Y es la variable explicada y sus dimensiones son $n \times 1$.

Mientras que $\hat{\beta}_{MCR}$ solo reduce la varianza del estimador, el estimador $\hat{\beta}_R$ asegura un ECM menor. Considerando estas ideas, Sarkar(1992) propuso la Regresión Ridge Restringida (en adelante RRR). Para ello tomó el EMCR y lo modificó en línea con el ERR con lo que obtuvo el siguiente estimador para los parámetros

$$\hat{\beta}_{RRRS} = [I_p + k(X'X)^{-1}]^{-1}\hat{\beta}_{MCR}. \quad (3.25)$$

Groß (2003) declaró que $\hat{\beta}_{RRRS}$ no satisface $r = R\hat{\beta}_{RRRS}$ para todas las soluciones, por lo que el introduce un nuevo estimador restringido

$$\hat{\beta}_{RRRG} = \hat{\beta}_R + (X'X + kI)^{-1}R'(R(X'X + kI)^{-1}R')^{-1}(r - R\hat{\beta}_R); k > 0 \quad (3.26)$$

donde $\hat{\beta}_R$ está dada por la ecuación 2.4.

3.1.4. Regresión Ridge con Restricciones Estocásticas.

Cuando existe incertidumbre acerca de la información disponible, una alternativa es hacer uso de restricciones aleatorias con una estructura lineal, dadas por

$$r = R\beta + \phi, \quad (3.27)$$

con r vector de dimensiones $(l \times 1)$ y R es una matriz que contiene información a priori sobre el vector $\hat{\beta}$ y es de dimensiones $(l \times p)$, con $l \leq p$, además se incluye ϕ que es un vector aleatorio independiente de ϵ , con $E(\phi)=0$ y $Var(\phi)=(\sigma^2 A)$, donde A es una matriz conocida definida positiva de dimensiones $n \times n$.

Özkale (2009) cita a Trenkler (1984) para construir un estimador a partir de sus aportes que sea capaz de lidiar con este tipo de fenómenos, vale decir, cuando se presenta multicolinealidad y además se presentan restricciones con un cierto grado de incertidumbre, el estimador propuesto es:

$$\hat{\beta}_m(k) = \hat{\beta}_g(k) + (X'C^{-1}X + kI)^{-1}R'(A + R(X'C^{-1}X + kI)^{-1}R')^{-1}(r - R\hat{\beta}_g(k)); k > 0 \quad (3.28)$$

Magister en Matemática Mención Estadística

donde $\hat{\beta}_g(k) = (X'C^{-1}X + kI)^{-1}X'C^{-1}Y$ es el estimador de Trenkler con $k > 0$, C : es una matriz de dimensiones $n \times n$ conocida y simétrica, por su parte A es una matriz de dimensiones $n \times n$ conocida y definida positiva.

Capítulo 4

Selección de Variables con Estimación Sesgada.

Una de las consecuencias del uso del ERR es la contracción hacia cero de los coeficientes, como resultado directo de la utilización de la restricción (3.2). La figura 4.1 grafica este punto, suponiendo la estimación de sólo 2 coeficientes (β_1 y β_2), esta es una cualidad de RR, ya que los coeficientes crecen en valor absoluto ante la presencia de multicolinealidad.

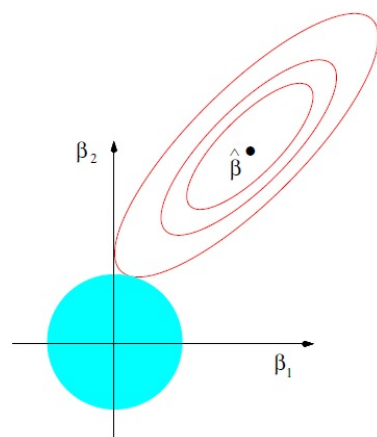


Figura 4.1: La región celeste representa la restricción $\beta_1^2 + \beta_2^2 < s$. Las elipses rojas son los contornos de la función de error de mínimos cuadrados. El punto negro representa la estimación con $(\hat{\beta}_{MCO})$ (Hastie y otros(2011), pag 71.)

Un ejemplo de esta cualidad de RR se puede visualizar en Hastie y otros (2011). Basado en los datos sobre el cáncer de próstata estudiados por Stamey y otros (1989), en la Figura N° 4. Se puede apreciar como los datos se mueven desde los coeficientes estimados por MCO a medida que evoluciona $df(k) = tr[X(X'X + kI_p)^{-1}X']$.

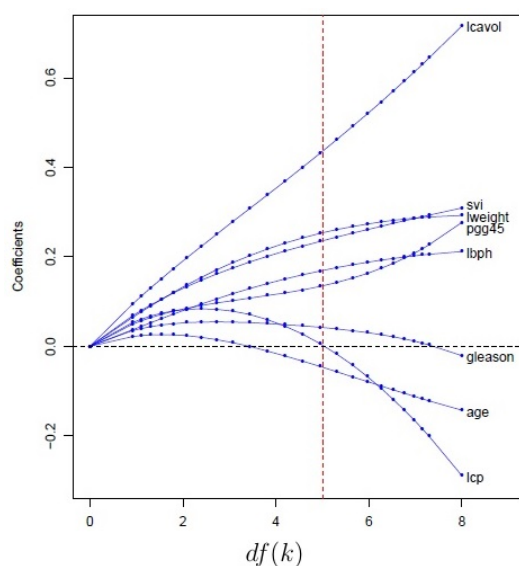


Figura 4.2: Coeficientes estimados con RR para el ejemplo del cáncer a la próstata, como función de la variación de $df(k)$, la línea punteada roja esta en $df(k)=5$ y fue seleccionada por validación cruzada (Hatie y otros(2011), pag 65)

Esta capacidad dista mucho de la selección de variables. Algunos autores para mantener la virtudes de la Regresión Ridge y a la vez incorporar la capacidad de seleccionar variables, utilizan algunos criterios existentes para selección de modelos y se basan en modificaciones que incorporan el ERR.

4.1. Selección de Variables con Regresión Ridge.

Las opciones consideradas por esta tesis no fueron diseñadas para un contexto de Regresión Lineal Multivariable (en adelante RLM) más bien fueron diseñadas para un contexto de Regresión Multivariada (en adelante RMa) que básicamente es un sistema de ecuaciones, desde esa perspectiva la RLM es un caso particular de la RMa.

Magister en Matemática Mención Estadística

Para contextualizar este tema consideraremos el problema descrito por Kubokawa y Srivastava (2012) en el contexto multivariado. Para obtener estimaciones estables de la matriz de varianzas y covarianzas se requiere que las dimensiones de esta sean considerablemente más pequeñas que el tamaño de la muestra, es decir $m < n$. Si no se cumple esta condición se compromete la estabilidad de las estimaciones de Σ .

Considerando un sistema de ecuaciones lineales descritas bajo una RMa con la siguiente estructura

$$\text{Modelo completo: } \tilde{Y} = \tilde{X}\beta_F + E, \tag{4.1}$$

donde \tilde{X} : es una matriz que contiene las variables explicativas y es de dimensiones $n \times P$, vale decir, contiene a las $x_{(1)}, \dots, x_{(P)}$ variables explicativas y que tienen n observaciones. La matriz \tilde{Y} contiene las variables explicadas y es de dimensiones $n \times m$, vale decir, la variables explicadas y_1, \dots, y_m . La matriz E es de dimensiones $n \times m$ y sus filas son iid como una normal multivariante cuyo vector de medias es cero. La matriz de varianzas y covarianzas Σ es de dimensiones $m \times m$. La matriz β_F contiene los parámetros desconocidos de dimensiones $P \times m$.

La idea es reducir la complejidad del modelo eliminando las variables que menos aportan al modelo y que estén causando problemas de muticolinealidad pero el modelo completo es un modelo que contempla información innecesaria, por lo que se dice está sobre especificado, entonces se desea especificar un modelo más reducido que se llamará Modelo Candidato y que posee x_1, \dots, x_p variables explicativas seleccionadas desde x_1, \dots, x_P variables, el modelo candidato es

$$\text{Modelo Candidato: } \tilde{Y} = X_{can}\beta_{can} + E, \tag{4.2}$$

cuya distribución de los datos de la matriz \tilde{Y} es

$$\tilde{Y} \sim N_{n,m}(X_{can}\beta_{can}, I_n, \Sigma) \tag{4.3}$$

donde $X = (x_1, \dots, x_p)$ y es una matriz de dimensiones $n \times p$. \tilde{Y} : es una matriz de dimensiones $n \times m$ que posee las variables explicadas y_1, \dots, y_m , E : es una matriz de dimensiones

Magister en Matemática Mención Estadística

$n \times m$ y sus filas son iid como una normal multivariante con vector de medias es cero, y la matriz de varianzas y covarianzas es Σ de dimensiones $m \times m$. β_{can} es la matriz de parámetros desconocidos de dimensiones $p \times m$.

Kubokawa y Srivastava (2012) propusieron estimadores para AIC y C_p cuando se presenta el problema de la multicolinealidad. Debido a las complicaciones que se pueden presentar en la matriz de varianzas y covarianzas se propuso un estimador tipo Ridge dado por

$$\hat{\Sigma}_{k_1} = \frac{1}{n}[(\tilde{Y} - \tilde{X}\tilde{\beta}_{MCO})(\tilde{Y} - \tilde{X}\tilde{\beta}_{MCO}) + k_1I], \quad (4.4)$$

donde $\hat{\Sigma}_{k_1}$: es una matriz con las estimaciones de las varianzas y covarianzas y es de dimensiones $P \times P$. $\tilde{\beta}_{MCO}$: es una matriz que posee las estimaciones de los coeficientes con EMCO y sus dimensiones son de $P \times m$. \tilde{Y} es una matriz que contiene las variables explicadas (y_1, \dots, y_m) , \tilde{X} es la matriz que posee las variables explicativas $x_{(1)}, \dots, x_{(P)}$ sus dimensiones son $n \times P$.

Al poseer muchas variables, la inversa de la matriz $\tilde{X}'\tilde{X}$ no es estable, esto implica que el EMCO ($\tilde{\beta}_{MCO}$) de β no se comporta bien debido a la imprecisión causada por la inflación de la varianza. Se plantea entonces el ERR para el contexto multivalente

$$\hat{\beta}_{Rk_2} = [(X'_{can}X_{can}) + k_2I]^{-1}X'_{can}\tilde{Y}, \quad (4.5)$$

donde k_2 es el parámetro de sesgo que es una constante no negativa. Elegir el valor de k_2 puede tomar relevancia en el cálculo, como alternativa se propone, en el marco de la selección de variables, elegir k_2 y X_{can} basados en AIC y/o C_p .

4.1.1. Selección de Variables con Criterio AIC Ridge Doble.

Kubokawa y Srivastava definen la Información de Akaike basado en estimadores de tipo doble de Ridge:

$$AI_{k_1, k_2} = -2E_{\tilde{Y}}[E_Z[\log f(Z|X_{can}\hat{\beta}_{Rk_2}(\tilde{Y}), \hat{\Sigma}_{k_1}(\tilde{Y}))]] \quad (4.6)$$

El criterio de Información de Akaike (AIC_{k_1, k_2}) es un estimador asintóticamente insesgado de AI_{k_1, k_2} basado en $-2\log f(\tilde{Y}; X_{can}\hat{\beta}_{Rk_2}, \hat{\Sigma}_{k_1})$, donde el sesgo queda definido como

$$\Delta_{k_1, k_2} = AI_{k_1, k_2} - E_Y[-2\log f(Z|X_{can}\hat{\beta}_{Rk_2}(\tilde{Y}), \hat{\Sigma}_{k_1})] \quad (4.7)$$

donde : AI_{k_1, k_2} es la Información de Akaike. Considerando $\rho_{k_2} = tr[X'_{can}X_{can}(X'_{can}X_{can} + k_2I)^{-1}]^2$ cuyo valor se estima con

$$\hat{\Delta}_{k_1, k_2} = A_1 + A_2(A_3)A_4 + A_5 \quad (4.8)$$

donde

$$A_1 = \frac{nm(m+1+p+(1-k_2^2)\rho_{k_2})}{n-p-m-1} \quad (4.9)$$

$$A_2 = \frac{c_n(n-p)}{m(n-p-m-1)} \quad (4.10)$$

$$A_3 = \frac{(n+(1-k_2^2)\rho_{k_2})(n-p)}{(n-m)^2} - 1 \quad (4.11)$$

$$A_4 = tr[\Sigma]tr[\Sigma^{-1}] \quad (4.12)$$

$$A_5 = O(n^{-\delta}) \quad (4.13)$$

el AIC_{k_1, k_2} tiene la siguiente estructura

$$AIC_{k_1, k_2} = -2\log f(\tilde{Y}; X_{can}\hat{\beta}_{Rk_2}, \hat{\Sigma}_{k_1}) + \hat{\Delta}_{k_1, k_2}. \quad (4.14)$$

y se calcula con

$$AIC_{k_1, k_2} = nm\log 2\pi + n\log|\hat{\Sigma}_{k_1}| + tr[\hat{\Sigma}_{k_1}^{-1}(Y - X\hat{\beta}_{Rk_2})'(Y - X\hat{\beta}_{Rk_2})] + \frac{nm(m+1+p+(1-k_2^2)\rho_{k_2})}{n-p-m-1} + \left\{ \frac{(n+(1-k_2^2)\rho_{k_2})(n-p)}{(n-m)^2} - 1 \right\} \hat{k}_1 tr[\hat{\Sigma}_{k_1}^{-1}] \quad (4.15)$$

En concreto, llego a obtener en varios escenarios el 100 % de los aciertos convirtiéndolo en un criterio bastante confiable.

4.1.1.1. Contextualización de AIC Ridge Doble.

El criterio AIC_{k_1, k_2} considera estimadores tipo Ridge para la matriz Σ para mitigar los problemas generados cuando la matriz Y es muy grande. Este problema no se presenta en nuestro estudio por lo que $k_1 = 0$. Además como la variable explicada Y es un vector de dimension $n \times 1$ por lo tanto $m = 1$, además como la varianza es un escalar, solo es un número que se debe estimar (para efectos de este estudio se utiliza un estimador basado en MCO). Haciendo uso de estas consideraciones la selección de variables con el método AIC doble presentado en la ecuación 4.15 toma la siguiente forma

$$AIC_{0,k} = n \log(2\pi) + n \log(\hat{\sigma}^2) + \frac{1}{\hat{\sigma}^2} (Y - X_{can} \hat{\beta}_R)' (Y - X_{can} \hat{\beta}_R) + \frac{n[2 + p + (1 - k^2)\rho_k]}{n - p - 2} \quad (4.16)$$

donde: $\hat{\sigma}^2 = \frac{1}{n-p} \sum \hat{e}_i^2$ y $\rho_k = tr[X'_{can} X_{can} (X'_{can} X_{can} + kI)^{-1}]^2$

4.1.2. Selección de Variables con Criterio C_p Ridge Doble.

El estadístico de Mallow puede ser extendido de manera similar a como se realizó con el criterio de AIC Ridge doble, entonces se debe considerar el ERR para la matriz de coeficientes dada por la ecuación 4.5 para construir

$$R_{PE}(\beta, \Sigma; \hat{\beta}_{k_1}) = np - PE_{k_2} \quad (4.17)$$

donde $PE_{k_2} = E_{\tilde{Y}}[tr[\Sigma^{-1}(\hat{\beta}_{Rk_2} - \beta)' X'_{can} X_{can} (\hat{\beta}_{Rk_2} - \beta)]]$ y PE_{k_2} es estimado por

$$PE_{k_2} = tr[\hat{\Sigma}_{k_1}^{-1}(\tilde{Y} - X_{can} \hat{\beta}_{k_2})' (\tilde{Y} - X_{can} \hat{\beta}_{k_2})] \quad (4.18)$$

con esto en cuenta el sesgo se calcula con:

$$\hat{\Delta}_{k_2} = P\hat{E}_{k_2} - E_{\tilde{Y}}[tr[\Sigma^{-1}(\hat{\beta}_{Rk_2} - \beta)'X'_{can}X_{can}(\hat{\beta}_{Rk_2} - \beta)]] \quad (4.19)$$

Kubokawa y Srivastava definen entonces el estimador para el sesgo como:

$$\hat{\Delta}_{k_2} = A_1 - A_2 + A_3 + A_4 \quad (4.20)$$

$$A_1 = n * tr[X'X(X'X + k_2I)^{-1}]^2 \quad (4.21)$$

$$A_2 = \frac{np(n - p - m - 1 + k_2^2)}{n - P - m - 1} \quad (4.22)$$

$$A_3 = \frac{c_n(n - P)}{(n - P - m - 1)} tr[\hat{\Sigma}]tr[\hat{\Sigma}_{-1}] \quad (4.23)$$

$$A_4 = O(1) \quad (4.24)$$

Entonces se define el estadístico Mallow Ridge doble como

$$C_{k_1, k_2} = tr[\tilde{\Sigma}_{k_1}^{-1}(\tilde{Y} - X\hat{\beta}_{Rk_2})'(\tilde{Y} - X_{can}\hat{\beta}_{Rk_2})] - \frac{nm(n - p - m - 1 + k_2^2\rho_{k_2})}{n - P - m - 1} + m\rho_{k_2} + \tilde{k}_1 tr[\tilde{\Sigma}^{-1}] \quad (4.25)$$

El valor óptimo del parámetro de sesgo k_2 y las variables optimas pueden ser seleccionadas simultánea y numéricamente con el fin de minimizar C_{k_1, k_2} considerando $0 \leq k_2 \leq k_2^0$.

En concreto C_{k_1, k_2} mostró muy buenos resultados, considerando variables con Multicolinealidad, para un $n = 22$, un $P = 7$ y un $m = 10$, sus resultados oscilaron entre el 99.3% y el 100% de seleccionar el modelo correcto. Cuando las variables no presentaban colinealidad la potencia bajaba un poco, logrando oscilar entre un 95.3% a un 97% de seleccionar el modelo adecuado.

4.1.2.1. Contextualización de C_p Ridge Doble.

La propuesta en la ecuación (4.25) considera estimadores basados en RR para la matriz de varianzas y covarianzas (Σ) y para la matriz de coeficientes (β). Para utilizar este criterio de selección en contexto univariante se consideraron las mismas modificaciones que para

el criterio de Akaike Doble, vale decir, la variable explicada Y es un vector de dimension $n \times 1$ por lo tanto $m = 1$, además como la varianza es un escalar y no es un matriz como el caso multivariado, solo es un número que se debe estimar (para efectos de este estudio se utiliza un estimador basado en MCO), entonces el Criterio de Mallow Doble para el contexto univariante es

$$C_{0,k} = \frac{1}{\hat{\sigma}^2} [(Y - X_{can}\hat{\beta}_R)'(Y - X_{can}\hat{\beta}_R)] - \frac{n(n-p-2+k^2\rho_k)}{n-P-2} + \rho_k \quad (4.26)$$

donde: $\hat{\sigma}^2 = \frac{1}{n-p} y \sum \hat{e}_i^2$ y $\rho_k = tr[X'_{can}X_{can}(X'_{can}X_{can} + kI)^{-1}]^2$

4.1.3. Selección de Variables Bajo Estadístico de C_p Modificado.

La siguiente propuesta en estudio la expuso Yanagihara y Satho (2010), bajo el mismo contexto multivariado expuesto desde el principio de este capítulo dados por las ecuaciones en 4.1 y considerando el hecho de que se presentan muchas variables explicativas o se presenta correlaciones altas entre variables explicativas, lo que implica estimaciones de los coeficientes con varianzas grandes. Los autores consideran al criterio C_p como un estimador de la función de riesgo basado en la MSE de predicción y proponen el siguiente estimador para MC_p :

$$MC_p = \left(1 - \frac{m+1}{n-p}\right) tr[WT^{-1}] + 2mtr[(X'_{can}X_{can} + k_2I)(X'_{can}X_{can})] + m(m+1). \quad (4.27)$$

donde: $W = (Y - \hat{Y}_{k_2})'(Y - \hat{Y}_{k_2})$ considerando que \hat{Y}_{k_2} son los valores predichos por el modelo utilizando el ERR para los coeficientes y T es un estimador insesgado de Σ bajo el modelo completo dado por $T = [Y'(I_n - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}')Y]/(n-p)$. I_n es una matriz idéntica de dimensiones $n \times n$.

Mediante un estudio de simulación de Monte Carlo, Yanagihara y Satho (2010) demostraron que el Criterio de Mallow Modificado (MC_p) logro los mejores resultados en comparación con el Criterio de Mallow tradicional, selección por Validación Cruzada y selección

por Validación Cruzada Generalizada. El MC_p logro elegir las variables adecuadas hasta en 779 ocasiones de las 1000 simulaciones realizadas.

4.1.3.1. Contextualización de C_p Modificado.

La propuesta del Criterio de Mallow Modificado, expuesto en la ecuación (4.27), está pensado para el contexto multivariado. Al igual que las propuestas de Kubokawa y Srivastava para contextualizarlas se considera que la variable explicada Y es un vector de dimension $n \times 1$ por lo tanto $m = 1$, además como la varianza es un escalar y no es un matriz como el caso multivariado, solo es un número que se debe estimar. Para efectos de este estudio utiliza un estimador basado en MCO. Entonces el Criterio de Mallow Mejorado para el contexto univariante es

$$MC_p = \left(1 - \frac{2}{n-p}\right) \frac{(Y - X_{can}\hat{\beta}_R)'(Y - X_{can}\hat{\beta}_R)}{\hat{\sigma}^2} + 2tr[(X'_{can}X_{can} + kI)^{-1}(X'_{can}X_{can})] + 2 \tag{4.28}$$

donde: $\hat{\sigma}^2 = \frac{1}{n-p}$ y $\sum \hat{e}_i^2$.

4.1.4. Selección de Variables Mediante Intervalos de Confianza.

Este concepto fue propuesto por Firinguetti y Bobadilla (2009). La idea es obtener intervalos de confianza asintóticos para los coeficientes del modelo. Basados en la RR considerando como estimador del parámetro de sesgo k la propuesta de Lawless y Wang (1976). Además de la Expansión de Edgeworth para obtener cuantiles asintóticos para la distribución del estimador de Lawless y Wang. La propuesta de ellos se traduce en una expresión matemática sencilla que se construye a través de las siguientes consideraciones:

Sea u el cuantil de una distribución $N(0, \sigma_0^2)$ y x el cuantil de

$$F_n(x) = P(\sqrt{ne}(q, w, n) < x), \tag{4.29}$$

Magister en Matemática Mención Estadística

Donde: q es un vector tal que $q = \frac{X'\epsilon}{n} \sim N(0, \sigma^2 \frac{X'X}{n})$, y w es un vector dimensiones $p \times 1$ tal que $w = \hat{\sigma}^2 - \sigma$, y n es el número de observaciones, y $e(q, w, n)$ es un función tal que $e(0, 0, n) = 0$ y $x = u - \frac{1}{4(n-p)}u + \frac{1}{4(n-p)}u^3 + O(n^{-\frac{3}{2}+\epsilon})$.

Considerando las anteriores ecuaciones Firinguetti y Bobadilla construyeron un intervalo de confianza asintótico al $100(1 - \alpha)\%$ para cada coeficiente estimado dado por

$$\hat{\beta}_{Ri} - x\sqrt{Var(\hat{\beta}_{Ri})} \leq \beta_i \leq \hat{\beta}_{Ri} + x\sqrt{Var(\hat{\beta}_{Ri})} \quad (4.30)$$

Donde la varianza ($Var(\hat{\beta}_{Ri})$) de cada coeficiente se determina con la ecuación (3.5).

El método de selección de variables fue puesto a prueba contra el EMCO, vale decir se construyeron intervalos de confianza (en adelante IC) considerando la varianza estimada con MCO. Como era de esperarse las varianzas estimadas por MCO producían IC grandes, que a menudo incluían el cero dentro. Por su parte el intervalo propuesto que incorpora las ventajas de la RR en general era mucho más reducido que el producido por MCO, lo que provocaba una selección de variables adecuada.

El método propuesto por Firinguetti y Bobadilla, está pensado para el contexto univariante en Regresión Líneal Múltiple por lo que será ocupado íntegramente sin modificaciones, por los que las ecuaciones 3.5 y 4.30 se programaron en R siguiendo las indicaciones del artículo y de la simulación que se realizó en la presente tesis.

Capítulo 5

LASSO.

LASSO (por su siglas en inglés Least Absolute Shrinkage and Selection Operator), es un método de Regresión Contraída propuesto por Tibshirani (1996), que al igual que RR y RLM busca minimizar la suma de los cuadrados de los residuos

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \right\}, \quad (5.1)$$

sujeta a la restricción

$$\sum_{j=1}^p |\beta_j| \leq s, \quad (5.2)$$

donde $s \geq 0$ y es un parámetro de sintonización que controla la contracción aplicada a las estimaciones. Para todo valor de s , $\beta_0 = \bar{y}$. Se puede asumir sin pérdida de generalidad que $\bar{y} = 0$, de este modo se omite β_0 , de ese modo la función a minimizar es:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 \right\}, \quad (5.3)$$

sujeta a la restricción

$$\sum_{j=1}^p |\beta_j| \leq s, \quad (5.4)$$

Debido al tipo de restricción, LASSO consigue un efecto similar al de Ridge contrayendo

Magister en Matemática Mención Estadística

los coeficientes hacia cero, ver figura 5.1 pero además logra reducir algunos coeficientes directamente hasta cero, de este modo se consigue seleccionar variables, ver Figura N°6.

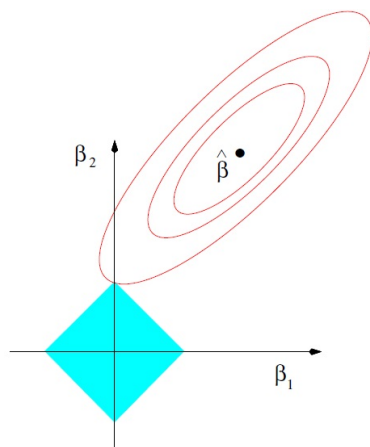


Figura 5.1: La región celeste representa la restricción $|\beta|_1 + |\beta|_2 < s$. Las elipses rojas son los contornos de la función de error de mínimos cuadrados. El punto negro representa la estimación con $(\hat{\beta}_{MCO})$, (Hastie y otros(2011))

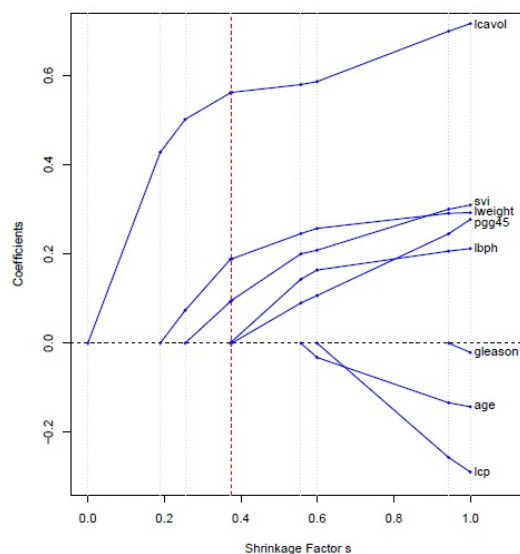


Figura 5.2: Coeficientes estimados con LASSO para el ejemplo del cáncer a la próstata, como función de la variación parámetro s , con $s = t / \sum_{j=1}^p |\beta_j|$, la línea punteada roja esta en $s=0.36$ y fue seleccionada por validación cruzada, (Hastie y otros(2011))

Tibshirani (1996) y Fu (1998) compararon la predicción de LASSO, Ridge y Regresión Bridge (Frank y Riedman, 1993) y encontraron que ninguno de ellos muestra resultados notablemente superiores a los otros dos, sin embargo Zuo y Hastie (2004) especificaron 3 escenarios en los que LASSO no funciona bien:

a) En el caso de que la cantidad de variables sea superior a la cantidad de observaciones, LASSO selecciona como máximo n variables, debido a la naturaleza del problema de optimización convexa. Esto es una característica limitante para este método de selección de variables.

b) Si hay un grupo de variables entre las cuales las correlaciones por pares son muy altas, entonces LASSO tiende a seleccionar cualquiera, y sólo una, de las variables de entre el grupo.

c) Para situaciones habituales donde la cantidad de observaciones es superior a la cantidad de variables, además de una alta correlación entre predictores, se ha observado empíricamente que las predicciones hechas con LASSO son dominadas por Regresión Ridge (Tibshirani, 1996).

5.1. Algoritmos de Optimización Propuesto por Tibshirani.

Minimizar el cuadrado de los errores dada la restricción (5.4), es un problema complejo. Tibshirani propuso dos algoritmos para resolverlo y así estimar los coeficientes.

5.1.1. Algoritmo 1.

El desarrollo del primer algoritmo, se basa en los aportes de Lawson y Hansen (1974) para solucionar problemas donde se minimiza el error cuadrático sujeto a restricciones del tipo:

$$G\beta \leq H \quad (5.5)$$

Donde G es una matriz de dimensiones $m \times p$ donde p es la cantidad de variables explicativas y H es un vector de dimensiones $m \times 1$, que contiene las restricciones. Para el caso particular de este problema de optimización $m = 2^p$, es fácil evidenciar que m crece exponencialmente en relación a p , lo que haría la aplicación del método algo ineficiente. Pero este problema se simplifica si se consideran secuencialmente las restricciones. Para entender el algoritmo se deben considerar lo siguiente:

$$1) g(\beta) = \sum_{i=1}^n \left(y_i - \sum_j \beta_j x_{ij} \right)^2$$

2) δ_i son las p duplas con la forma $(\pm 1, 1\pm, \dots, \pm 1)$ con $i = 1, 2, \dots, 2^p$,

3) La condición $\sum_{j=1}^p |\beta_j| \leq s$ es equivalente a $\delta'_i \beta \leq s$ para todo i .

4) Para un determinado β , se tiene $E = \{i : \delta'_i \beta = s\}$ y que $T = \{i : \delta'_i \beta < s\}$

5) Se denota G_E aquella matriz cuyas filas son los $\delta_i \in E$.

6) 1 es un vector de unos con la misma cantidad de filas que G_E .

El algoritmo es de carácter iterativo y repite un ciclo mientras se mantiene una determinada condición, pero antes debe hacer un paso previo.

Magister en Matemática Mención Estadística

Paso cero:

i) El algoritmo comienza con $E = i_0$, Se estima el vector de parámetros (β) con el EMCO $(\hat{\beta}_{MCO})$, se toman sus signos en δ_{i_0} para generar $\delta'_{i_0}\beta \leq s$.

ii) Luego si se cumple $\sum |\beta_j| \leq s$ el cálculo termina. Si no cumple la condición, las restricciones violadas se añaden a E y se continua con el proceso hasta que $\sum_{j=1}^p |\beta_j| \leq s$.

Paso uno:

Encontrar $\hat{\beta}$ que minimice $g(\beta)$ sujeto a $G_E\beta \leq s1$.

Paso dos:

Mientras $\sum |\hat{\beta}_j| > s$, agregar i al conjunto de E donde $\delta_i = \text{sign}(\hat{\beta})$, Encontrar $\hat{\beta}$ que minimice $g(\beta)$ sujeto a $G_E\beta \leq s1$

Este procedimiento siempre converge en un número determinado de pasos donde un elemento es agregado a E en cada paso.

5.1.2. Algoritmo 2.

La segunda propuesta se basa en los aportes de David Gay. Se escribe cada coeficiente β_j como $\beta_j^+ - \beta_j^-$, donde β_j^+ y β_j^- son no negativos. Luego se resuelve el problema de mínimos cuadrados con las siguientes restricciones

$$\beta_j^+ \geq 0, \tag{5.6}$$

$$\beta_j^- \geq 0, \tag{5.7}$$

$$\sum \beta_j^+ + \sum \beta_j^- \leq s, \quad (5.8)$$

Este, método transforma el problema de p variables y 2^p restricciones a uno de $2p$ variables y $2p + 1$ restricciones, finalmente ambos métodos dan el mismo resultado.

5.2. Algoritmo LAR.

Por sus siglas en inglés Least Angle Regression, es similar al método StepWise Forward, este último funciona de manera secuencial, incorporando una variable a la vez. Básicamente en cada paso se identifica la mejor variable a incorporar al modelo y en cada paso se recalcula el error cuadrático.

LAR ocupa una estrategia similar pero, como mucho, solo entra un predictor. Como primer paso se identifica la variable explicativa que posea mayor correlación con la variable respuesta, en lugar de adaptarse por completo a esa variable, LAR mueve al coeficiente de la variable continuamente hacia el valor el mínimo cuadrático (Causando con la evolución residual a decrecer en valor absoluto). Luego cuando entra una nueva variable se reestiman los coeficientes juntos.

El algoritmo se puede resumir en los siguientes pasos:

- 1) Se normalizan los datos con media cero y desviación estándar 1.
- 2) Se comienza con el residual $r = y - \bar{y}$.
- 3) Se considera que todos los coeficientes son igual cero ($\beta_1, \dots, \beta_p = 0$).
- 4) Se busca la variable explicativa x_j más correlacionada con la variable respuesta (r).

Magister en Matemática Mención Estadística

- 5) Mueve β_j desde cero hacia el coeficiente de Mínimos Cuadrados, hasta que otra variable β_k esté más correlacionada con el residual actual.

- 6) Mueve β_j y β_k desde sus puntos actuales hacia el coeficiente de Mínimos Cuadrados del residual actual sobre (x_j, x_k) , hasta que otra variable x_l esté más correlacionada con el residual actual.

- 7) Se sigue el procedimiento hasta que entran todas las variables. Al final se tiene la solución que Minimiza el error cuadrático.

Capítulo 6

Elastic Net.

La Regresión Elastic Net es un método propuesto por Zuo y Hastie (2005) para regularizar y a la vez de seleccionar variables mediante la minimización del error cuadrático sujeto a una restricción que se compone por una parte de la penalización L_1 referente a LASSO y la penalización L_2 referente a Ridge. En el artículo se compara este método con LASSO para contrastar sus ventajas por sobre este, evidenciando buenos resultados cuando la cantidad de variables es mucho más grande que la cantidad de observaciones.

La idea de esta propuesta es mezclar las mejores las ventajas de LASSO con la Regresión Ridge, sin capturar los defectos de estos métodos, vale decir, LASSO tiende a soslayar la correlación entre variables, mientras que Regresión Ridge contrae hacia cero todos los coeficientes.

Para un acercamiento más claro se considera un conjunto de datos con n observaciones y con p posibles variables explicativas que componen una matriz X de dimensiones $n \times p$, además se considera el vector Y de dimensiones $n \times 1$ como variable explicada y

Magister en Matemática Mención Estadística

que tiene n observaciones. Entonces se define un estimador Ingenuo para Elastic Net

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \right\}, \quad (6.1)$$

sujeta a la restricción:

$$\lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p (\beta_j)^2 \leq s, \quad (6.2)$$

donde: λ_1 y λ_2 son parámetros ajustables no negativos, la función de λ_1 es conseguir modelos parsimoniosos mientras que λ_2 promueve que variables altamente correlacionadas tengan coeficientes similares. De forma equivalente la restricción se puede reescribir como

$$(1 - \alpha) \sum_{j=1}^p |\beta_j| + \alpha \sum_{j=1}^p (\beta_j)^2 \leq s, \quad (6.3)$$

donde: $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$.

Es de notar que cuando $\alpha = 1$ el problema de minimización toma la forma de una Regresión Ridge, y cuando $\alpha = 0$ el problema toma forma de una Regresión LASSO.

La restricción proporcionada por Elastic Net representa una combinación lineal convexa de las penalizaciones aportadas por Ridge y LASSO, que por lo demás, a diferencia de LASSO es estrictamente convexa.

Las estimaciones y cálculos para resolver el problema de optimización asumen que las variables explicativas están estandarizadas y la variable explicada está centrada, vale decir:

$$\sum_{i=1}^n x_{ij} = 0, \text{ para } j = 1, \dots, p, \quad (6.4)$$

$$\sum_{i=1}^n x_{ij}^2 = 1, \text{ para } j = 1, \dots, p, \quad (6.5)$$

$$\sum_{i=1}^n y_i = 0. \quad (6.6)$$

Para resolver el problema de optimización se transforma la matriz X y el vector Y generando una matriz y un vector ficticio de acuerdo lo siguiente

$$X^* = \frac{1}{\sqrt{1 + \lambda_2}} \begin{pmatrix} X \\ \sqrt{\lambda_2} I \end{pmatrix}, \quad (6.7)$$

$$Y^* = \begin{pmatrix} Y \\ 0 \end{pmatrix}, \quad (6.8)$$

donde: X^* es una matriz de dimensiones $(n + p) \times p$ y el vector Y^* es de dimensiones $(n + p) \times 1$, esto permite resolver el problema de optimización como si fuera un problema tipo LASSO:

$$\hat{\beta}^* = \underset{\beta^*}{\text{mín}} |Y^* - X^* \beta^*| + \frac{\lambda_1}{\sqrt{1 + \lambda_2}} |\beta^*|. \quad (6.9)$$

Según Zuo y Hastie (2004) en sus simulaciones mostraron que el estimador Ingenuo para Elastic Net no lleva a cabo de manera satisfactoria la selección de variables a menos que sea muy cercano a una Regresión Ridge o LASSO es por eso que se le llamó estimador Ingenuo. Para corregir este problema propusieron una corrección a la estimación dada por

$$\hat{\beta}_{ing} = \frac{1}{\sqrt{1 + \lambda_2}} \hat{\beta}^*, \quad (6.10)$$

$$\hat{\beta}_{EN} = (1 + \lambda_2) \hat{\beta}_{ing}, \quad (6.11)$$

donde: $\hat{\beta}_{EN}$ es el vector de parámetros calculado con Elastic Net, $\hat{\beta}_{ing}$ es el vector de coeficientes calculado con el estimador ingenuo de Elastic Net, $(1 + \lambda_2)$ es el factor de corrección, este se implementa ya que la estructura de la restricción puede permitir la inclusión de sesgos que deben ser neutralizados.

6.1. Elección de Parámetros de Ajuste.

Hay buenas opciones para ajustar los parámetros de Elastic Net (λ_2 y s) (Hastie y otros (2001), capítulo 7). El método más popular y que está implementado en R, es la Validación Cruzada de 10 iteraciones, que en este caso es para dos dimensiones. Normalmente se escogen valores pequeños de λ_2 (ejemplo $\lambda_2 = (0, 0.001, 0.01, 0.1, 1, 10, 100)$). Luego para cada λ_2 el algoritmo produce soluciones completas para cada trayectoria de Elastic Net. Mientras que el otro parámetro de ajuste (λ_1 o S) es seleccionado para 10 iteraciones de la Validación Cruzada. El λ_2 seleccionado es el que menor error cuadrático medio de predicción.

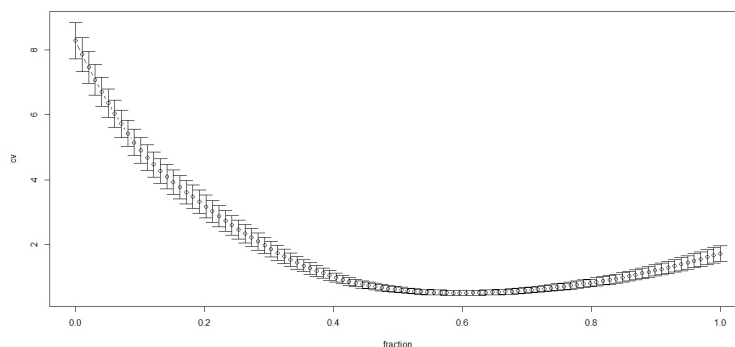


Figura 6.1: Error Cuadrático Medio de Predicción: Calculado con Validación Cruzada de 10 iteraciones, ejemplo con parámetro $\lambda_2 = 100$, Fuente: Elaboración Propia

Capítulo 7

Estudio de Simulación.

En este capítulo se describirá, en términos generales, la realización de la simulación (en adelante simulador). Si bien es cierto el simulador se pudo construir en otros lenguajes más livianos desde el punto de vista del costo computacional como C, Fortran u otros, no es menos cierto que trabajar con estas opciones tiene la complicación de que requieren más líneas código para conseguir el mismo objetivo. Considerando lo anterior, se escogió R como lenguaje de programación ya que algunas de las metodologías que se evalúan en esta tesis ya están implementadas en dicho lenguaje, lo que ahorra tiempo en la etapa de desarrollo del código del simulador.

7.1. Objetivo de Desarrollo.

El código se creó pensando estrictamente en los resultados que se desean obtener, por lo tanto se articularon algoritmos que entregan resultados estructurados en matrices y vectores, que posteriormente son resumidos en tablas y gráficos. En general se busca con esta tesis:

“Estudiar el comportamiento de los métodos de selección escogidos, desde el punto de vista de la selección de variables, de la estimación de coeficientes y de las predicciones que se pueden realizar con los modelos seleccionados”.

En resumen, se evalúa el comportamiento de estos métodos desde tres puntos de vista diferentes, que permitirán a futuros investigadores tener una idea más clara sobre las situaciones en que se puede ocupar estos métodos.

Para estudiar en detalle el comportamiento de estos métodos, se establecieron varios escenarios ante los cuales podrían llegar a estar expuestos. Estos escenarios (modelos teóricos) están dados por tres diferentes niveles de correlaciones (Baja, Moderada, Alta) y tres diferentes varianzas (0.01, 0.1 y 1). Además de considerar tres cantidades de variables disponibles para seleccionar (10, 15 y 20 variables). Adicionalmente se consideró la cantidad de observaciones disponibles como un factor determinante por lo que se evalúan todos los escenarios con tres cantidades diferentes de observaciones (25, 50 y 100).

7.1.1. Contéo de Modelos Seleccionados.

Desde este punto de vista, se evalúan los métodos cuantificando en términos porcentuales la cantidad de veces que un método determinado seleccionó el modelo correcto y si no lo hizo que tipo de modelo es el que ajusto, esto último es un aporte importante puesto que entrega nociones de cómo se están equivocando cuando no selecciona el modelo correcto.

Cabe mencionar que en los artículos citados se suele estudiar este problema desde un punto de vista todo o nada, sin entregar información sobre qué ocurre cuando no se selecciona el modelo correcto. Tomando en cuenta lo anterior se establecieron las siguientes categorías de conteo:

Magister en Matemática Mención Estadística

MB: Muy Bueno, se selecciona el modelo correcto, es decir se seleccionan todas las variables relevantes y se descartan todas las variables irrelevantes.

B: Bueno, se selecciona un modelo que incorpora todas las variables correctas pero se incorporan otras variables (variables irrelevantes).

Re: Regular, se seleccionan solo algunas de las variables relevantes en el modelo, pero no se incorporan variables irrelevantes.

Ma: Malo, se seleccionan solo algunas de las variables relevantes en el modelo y se incorporan algunas o todas las variables irrelevantes.

MMa: Muy Malo, No se selecciona ninguna variable relevante en el modelo y se incorporan todas las variables irrelevantes.

Ante cada escenario, se cuenta la cantidad de veces que se selecciona un modelo que corresponda a una de las categorías definidas, este conteo (en términos porcentuales) es almacenado en un matriz de resultados para la posterior lectura. A continuación se muestra un extracto de cómo se vería la matriz de resultados para 10 variables, con 100 observaciones con las 3 varianzas $\sigma_1^2 = 0,01$, $\sigma_2^2 = 0,1$ y $\sigma_3^2 = 1$, las correlaciones ρ_1 =Correlación Baja, ρ_2 =Correlación Moderada y ρ_3 =Correlación Alta y todos los métodos simulados.

Cuadro 7.1: Recuento de Modelos Seleccionados

	$P=10$ y $n=100$								
	σ_1^2			σ_2^2			σ_3^2		
	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3
AIC_{HKB}									
AIC_{LW}									
AIC_{DK}									
Cp_{HKB}									
Cp_{LW}									
Cp_{DK}									
MCp_{HKB}									
MCp_{LW}									
MCp_{DK}									
IC_{95}									
IC_{99}									
$LASSO$									
$E.N.$									

Considerando esta forma en la entrega de resultados se desarrolló un algoritmo cuyo propósito es el conteo de la cantidad de veces que seleccionó un modelo de las categorías definidas. El Cuadro (7.1) es un esquema representativo de cómo se exponen los resultados de la simulación en este aspecto, la matriz de resultados en (7.1) es solo un resumen, puesto que cada casilla se expande en las 5 categorías que se deben contar, en otras palabras la matriz de resultados no es de dimensiones 13×9 sino de 13×45 .

7.1.2. Indicador de Calidad de Selección de Variables.

Para la construcción de tablas de reporte que resuman el comportamiento de los métodos, se hizo evidente la necesidad de ocupar un indicador que de una noción de que tan bien se está seleccionado variables con algún método en particular.

En cada simulación se ocupó el indicador denotado por la letra F , el cual es un reconocido indicador para este tipo de estudios. Fue desarrollado a partir de los aportes de van Rijsbergen,(1981). Cuando F toma el valor 1 indica que se selecciona al modelo correcto. El 0 indica que se seleccionaron las variables irrelevantes, siendo el caso menos deseado. Valores entre cero y 1 indican imperfecciones en la selección. Entre más cercano a 1 menos errores al seleccionar variables, cuanto más cercano a cero más errores se cometen en la selección.

El indicador F se calcula mediante la media armónica entre P y R dada por la siguiente ecuación

$$F = \frac{2}{\left(\frac{1}{P} + \frac{1}{R}\right)}, \quad (7.1)$$

donde

$$0 \leq F \leq 1, \quad (7.2)$$

$$0 \leq P \leq 1, \quad (7.3)$$

$$0 \leq R \leq 1. \quad (7.4)$$

P se interpreta como un medidor de Precisión en el Modelo Recuperado. Cuando P toma el valor uno indica que se selecciona al modelo correcto. El cero indica que se seleccionaron las variables irrelevantes, siendo el caso menos deseable. Valores entre cero y uno indican selección de variables irrelevantes en el modelo, lo que constituye una situación no deseada. El Indicador P se define como

$$P = \frac{\text{Variables Correctamente Seleccionadas}}{\text{Total de Variables Seleccionadas}} \quad (7.5)$$

Magister en Matemática Mención Estadística

Por su parte R se interpreta como un medidor de Reconstrucción del Modelo Recuperado. Cuando R toma el valor uno indica que se selecciona al modelo correcto. El cero indica que se seleccionaron las variables irrelevantes, siendo el caso menos deseable. Valores entre cero y uno indican el descarte de variables relevantes para el modelo, lo que claramente es una situación inadecuada. El Indicador R se define como

$$R = \frac{\text{Variables Correctamente Seleccionadas}}{\text{Variables Relevantes Verdaderas}} \tag{7.6}$$

Con los medidores P y R se pueden configurar los siguientes casos:

	Variable Relevante	Variable Irrelevante
Recuperado	Verdadero Positivo (Vp)	Falso Positivo (Fp)
No Recuperado	Falso negativo (Fn)	Verdadero Negativo (Vn)

Entonces

$$P = \frac{Vp}{Vp + Fp}, \tag{7.7}$$

$$R = \frac{Vp}{Vp + Fn}. \tag{7.8}$$

Es importante mencionar que en el desarrollo de la tesis se encontró el caso del descarte de todas las variables, particularmente en los métodos de selección propuestos por Firinguetti y Bobadilla y Elastic Net. Sucede que este caso particular es considerado indeseable por lo que se incorporó una rutina en el programa que asigna un cero en F cuando se da este caso.

Es importante resaltar que para cada método se calcula en cada simulación el indicador F . Entonces al finalizar el simulador su operación, por cada método, se tienen 9000 datos(1000 simulaciones \times 3 niveles de colinealidad \times 3 niveles de varianza) que son almacenados en un vector, como ejemplo se presenta en el Cuadro 7.2 como es el vector que almacena los indicadores F calculados para AIC_{HKB} :

Cuadro 7.2: Vector de Indicador F para AIC_{HKB}

AIC_{HKB}	F_1	...	F_{9000}
-------------	-------	-----	------------

Para resumirlos se presentara en la tabla con el promedio de los indicadores F calculados para cada método dado por

$$\bar{F} = \frac{1}{1000} \sum_{j=1}^{1000} (F_j) \quad (7.9)$$

donde: F_j es el indicador F calculado en la simulación j y \bar{F} es el promedio de los indicadores F calculados. Es de notar que el promedio se calcula solo con 1000 datos puesto que dentro del vector hay nueve situaciones diferentes debido a los tres niveles de varianza y los tres niveles de correlación.

7.1.3. Estimación de Coeficientes.

Se consideró relevante la estimación de los coeficientes más allá de la selección de variables. En el cuadro 7.3 se puede ver a modo de ejemplo la matriz que se espera construir para el caso del criterio AIC_{ok} con parámetro de sesgo estimado con HKB con 10 variables disponibles.

Cuadro 7.3: Matriz de con Coeficientes Estimados por AIC_{HKB}

AIC_{HKB}			
Coeficiente	Simulación		
	1	...	9000
β_0	$\hat{\beta}_{0,1}$...	$\hat{\beta}_{0,9000}$
β_1	$\hat{\beta}_{1,1}$...	$\hat{\beta}_{1,9000}$
β_2	$\hat{\beta}_{2,1}$...	$\hat{\beta}_{2,9000}$
β_3	$\hat{\beta}_{3,1}$...	$\hat{\beta}_{3,9000}$
β_4	$\hat{\beta}_{4,1}$...	$\hat{\beta}_{4,9000}$
β_5	$\hat{\beta}_{5,1}$...	$\hat{\beta}_{5,9000}$
β_6	$\hat{\beta}_{6,1}$...	$\hat{\beta}_{6,9000}$
β_7	$\hat{\beta}_{7,1}$...	$\hat{\beta}_{7,9000}$
β_8	$\hat{\beta}_{8,1}$...	$\hat{\beta}_{8,9000}$
β_9	$\hat{\beta}_{9,1}$...	$\hat{\beta}_{9,9000}$
β_{10}	$\hat{\beta}_{10,1}$...	$\hat{\beta}_{10,9000}$

Si bien la matriz que guarda las estimaciones de los coeficientes en cada simulación es un resultado en si mismo, es conveniente resumir estos datos para evaluarlos de manera global, para ello se hace uso del Error cuadrático Medio Total dada por la siguiente expresión

$$TECM = \frac{1}{1000} \sum_{j=1}^{1000} (\hat{\beta}_{ji} - \beta_i)^2 \tag{7.10}$$

donde: $j = (1, \dots, 1000)$, $i = (0, \dots, p)$, $\hat{\beta}_{ji}$ es la estimación del coeficiente β_i en la simulación j con alguno de los métodos estudiados. Es de notar que el TECM se calcula solo con 1000 datos puesto que dentro de la Matriz hay nueve situaciones diferentes, debido a los tres niveles de varianza y los tres niveles de correlación.

7.1.4. Error Cuadrático Medio de Predicción Promedio.

Para evaluar los modelos desde el punto de vista de las predicciones que son capaces de realizar, se utilizó la siguiente estrategia. Cada vez que se genera la matriz X y el vector Y , se genera adicionalmente una observación más (ejemplo si uno de los Modelos teóricos incorpora 100 observaciones, entonces en realidad se generan 101 observaciones), este dato se deja fuera de los cálculos y estimaciones. Luego de haber seleccionado los modelos y haber estimado los coeficientes se realizaban las predicciones considerando los valores de la matriz X que fueron excluidos. Claramente estos cálculos se realizaban en cada simulación.

Posteriormente se comparan las estimaciones con el valor real de Y por medio del error cuadrático y resumían por medio de una sumatoria con lo que se conseguía un solo número que muestra la calidad de la estimación en cada simulación de acuerdo a la siguiente ecuación:

$$ECMP = \frac{1}{1000} \sum_{j=1}^{1000} (\hat{y}_{ji} - y_{ji})^2 \quad (7.11)$$

Donde: $j = (1, \dots, 1000)$, $i = (101)$, \hat{y}_{ji} es la estimación de la observación i del vector Y en la simulación j , y_{ji} es el valor real de la observación i del vector Y .

7.2. Comparación entre Métodos de Selección.

Con todas las consideraciones especificadas se definen en total 162 modelos teóricos diferentes, que deben ser estimados tomando en cuenta cada método de selección, para cada vector de coeficientes (81 modelos con coeficientes definidos a través de los Vectores Propios y 81 modelos con coeficientes definidos arbitrariamente).

Magister en Matemática Mención Estadística

Es importante describir cada una de estas situaciones considerando todos los elementos implicados. En este contexto un indicador importante es la Señal Ruido (SR): que se define como la razón entre el cuadrado de la señal explicada y la varianza, dada por

$$SR = \frac{\beta' X' X \beta}{\sigma^2} \tag{7.12}$$

Cuanto más pequeña es la Señal Ruido significa que la perturbación aleatoria (error) tiene una varianza σ^2 muy grande en relación a la Señal ($X\beta$). En cambio cuando la señal ruido es grande implica que la perturbación aleatoria (ϵ) tiene una varianza muy pequeña en relación a la señal. Contextualizando esta definición, una Señal Ruido pequeña indica poca información (pocas observaciones), correlaciones altas, varianzas grandes y muchas variables disponibles, lo que constituye una situación desfavorable para los métodos de selección y estimación. Por el contrario una Señal Ruido de valor grande se dará cuando hay más observaciones, correlaciones bajas, varianza pequeña y pocas variables disponibles; elementos que favorecen una selección de variables adecuada.

En el cuadro 7.4 se muestra una matriz con las Señal Ruido de todos los modelos a evaluar cuando se estima el vector de parámetros basado en los valores propios:

Cuadro 7.4: Señal Ruido, vector de coeficientes basado en vectores propios

	σ^2	0.01			0.1			1		
P	$n \backslash col$	<i>Baja</i>	<i>Moderada</i>	<i>Alta</i>	<i>Baja</i>	<i>Moderada</i>	<i>Alta</i>	<i>Baja</i>	<i>Moderada</i>	<i>Alta</i>
10	25	2064.28	1385.88	1324.23	206.43	138.59	132.42	20.64	13.86	13.24
	50	3278.93	3143.86	2826.32	327.89	314.39	282.63	32.79	31.44	28.26
	100	5926.46	5913.98	5575.93	592.65	591.40	557.59	59.26	59.14	55.76
15	25	1343.18	1245.49	892.54	134.32	124.55	89.25	13.43	12.45	8.93
	50	3364.84	2876.18	2897.37	336.48	287.62	289.74	33.65	28.76	28.97
	100	6451.03	4419.86	4298.29	645.10	441.99	429.83	64.51	44.20	42.98
20	25	1468.27	1164.71	676.82	146.83	116.47	67.68	14.68	11.65	6.77
	50	3530.08	1784.91	1189.87	353.01	178.49	118.99	35.30	17.85	11.90
	100	5653.98	5076.10	4216.54	565.40	507.61	421.65	56.54	50.76	42.17

Magister en Matemática Mención Estadística

En el cuadro 7.5 se muestra una matriz con las Señal Ruido de todos los modelos a evaluar cuando se estima el vector de parámetros con coeficientes definidos arbitrariamente:

Cuadro 7.5: Señal Ruido, Vector de Coeficientes Definido Arbitrariamente

	σ^2	0.01			0.1			1		
P	$n \backslash col$	<i>Baja</i>	<i>Moderada</i>	<i>Alta</i>	<i>Baja</i>	<i>Moderada</i>	<i>Alta</i>	<i>Baja</i>	<i>Moderada</i>	<i>Alta</i>
10	25	73789.2	71004.5	71330.0	7378.9	7100.5	7133.0	737.9	710.0	713.3
	50	139366.6	143104.8	145446.6	13936.7	14310.5	14544.7	1393.7	1431.0	1454.5
	100	259140.1	285019.8	282404.4	25914.0	28502.0	28240.4	2591.4	2850.2	2824.0
15	25	68466.7	63361.4	66323.6	6846.7	6336.1	6632.4	684.7	633.6	663.2
	50	123750.9	132912.6	122048.1	12375.1	13291.3	12204.8	1237.5	1329.1	1220.5
	100	246861.8	282030.8	255890.4	24686.2	28203.1	25589.0	2468.6	2820.3	2558.9
20	25	76641.6	65985.6	44918.2	7664.2	6598.6	4491.8	766.4	659.9	449.2
	50	144565.1	129176.6	108516.4	14456.5	12917.7	10851.6	1445.7	1291.8	1085.2
	100	268775.9	270225.5	249702.9	26877.6	27022.5	24970.3	2687.8	2702.3	2497.0

Antes de estudiar los resultados, es necesario presentar el cuadro 7.6, con las nomenclaturas que aparecerán en las tablas

Cuadro 7.6: Abreviaciones para Métodos de Selección

Sigla	Significado
AIC_{HKB}	AIC Ridge Doble con estimador de parametro de sesgo HKB
AIC_{LW}	AIC Ridge Doble con estimador de parametro de sesgo LW
AIC_{DK}	AIC Ridge Doble con estimador de parametro de sesgo DK
C_{HKB}	C_p Ridge Doble con estimador de parametro de sesgo HKB
C_{LW}	C_p Ridge Doble con estimador de parametro de sesgo LW
C_{DK}	C_p Ridge Doble con estimador de parametro de sesgo DK
MC_{PHKB}	C_p Modificado con estimador de parametro de sesgo HKB
MC_{PLW}	C_p Modificado con estimador de parametro de sesgo LW
MC_{pDK}	C_p Modificado con estimador de parametro de sesgo DK
IC_{95}	Selección con Intervalo de Confianza al 95 %
IC_{99}	Selección con Intervalo de Confianza al 99 %
$LASSO$	Selección con LASSO
EN	Selección con Elastic Net

7.2.1. Coeficientes Definidos por Valores Propios.

7.2.1.1. 10 Variables Disponibles.

7.2.1.1.1 25 Observaciones.

El cuadro 7.7 muestra en términos porcentuales la cantidad de veces que un método particular seleccionó una categoría de modelo (definidos en el ítem 7.1.1), por ejemplo cuando se simuló la selección de variables con el criterio AIC_{HKB} ante una colinealidad baja, con 10 variables disponibles, con 25 observaciones por variable y ante una varianza de 0.01 (varianza baja), del 100 % de las simulaciones (1000 simulaciones), solo en un 10.7 % se seleccionó el modelo correcto (MB).

Se puede ver en el cuadro 7.7 que seleccionar el modelo correcto (MB), en general es muy poco frecuente, incluso para métodos como LASSO y Elastic Net. Por otra parte los métodos basados en Regresión Ridge con Intervalos de Confianza al 95 % y 99 %, dieron buenos resultados en relación a los métodos basados en AIC Ridge, pero conforme aumenta la varianza, la capacidad de seleccionar disminuye con más rapidez en los métodos basados en Intervalos de Confianza. Además los peores resultados los dio el Criterio de Mallows basado en Ridge y el criterio de Mallows Mejorado, independiente del estimador utilizado para el parámetro de sesgo, aunque se puede ver una mejoría en la selección en la medida que crece la varianza. Sus mejores resultados se presentan en su mayoría al seleccionar modelos considerados regulares.

Respecto del cuadro 7.8, lo que más llama la atención es que el criterio para la selección de variables de Mallows Mejorado (MCp), para los tres estimadores del parámetro de sesgo solo consiguió, en el mejor de los casos, Modelos catalogados como regulares (Re). Ni una sola vez seleccionó modelos catalogados como buenos(B) o catalogados como una Correcta Selección (MB).

Magister en Matemática Mención Estadística

Cuadro 7.7: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,275, -0,121, -0,208, 0,428, 0,423, 0,617, -0,351, 0, 0, 0, 0)$															
<i>Colinealidad = Baja; P=10; n=25; S=1000</i>															
SR	2064.28					206.43					20.64				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
<i>AIC_{HKB}</i>	10.7	5.6	54.0	29.7	0.0	0.2	0.3	30.7	68.8	0.0	0.1	0.0	15.1	78.1	6.7
<i>AIC_{LW}</i>	9.4	5.1	56.7	28.8	0.0	0.2	0.1	54.5	43.5	1.7	0.0	0.0	45.1	37.8	17.1
<i>AIC_{DK}</i>	0.0	0.0	18.5	65.3	16.2	0.0	0.2	14.6	73.7	11.5	0.0	2.8	12.1	78.9	6.2
<i>C_{HKB}</i>	0.0	0.0	1.1	98.8	0.1	0.0	0.0	1.5	97.1	1.4	0.1	0.0	16.3	76.4	7.2
<i>C_{LW}</i>	0.0	0.0	16.0	0.0	84.0	0.0	0.0	38.9	0.0	61.1	0.0	0.1	53.5	6.1	40.3
<i>C_{DK}</i>	0.0	0.0	18.0	65.0	17.0	0.0	0.1	14.5	73.7	11.7	0.0	2.5	13.1	77.2	7.2
<i>MC_{pHKB}</i>	0.0	0.0	87.8	0.0	12.2	0.0	0.0	71.7	0.0	28.3	0.0	0.0	57.0	2.4	40.6
<i>MC_{pLW}</i>	0.0	0.0	16.0	0.0	84.0	0.0	0.0	38.4	0.0	61.6	0.0	0.0	53.4	0.0	46.6
<i>MC_{pDK}</i>	0.0	0.0	62.2	2.2	35.6	0.0	0.0	55.7	4.1	40.2	0.0	0.0	46.4	18.5	35.1
<i>IC₉₅</i>	0.1	85.3	0.0	14.6	0.0	0.5	23.6	4.7	71.2	0.0	0.1	0.6	25.5	59.6	6.6
<i>IC₉₉</i>	0.1	81.2	0.0	18.7	0.0	1.3	12.2	11.5	75.0	0.0	0.0	0.2	30.6	35.7	11.8
<i>LASSO</i>	0.0	0.0	21.2	78.8	0.0	0.2	12.1	7.2	80.5	0.0	0.0	67.3	0.0	32.7	0.0
<i>EN</i>	0.0	57.3	1.0	41.7	0.0	0.0	18.0	23.0	55.7	0.0	0.1	5.9	26.3	38.8	1.8

El conteo falla para *IC₉₅*, *IC₉₉* y *EN* ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.8: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (-0,211, 0,0559, -0,051, -0,717, -0,567, 0,205, 0,269, 0, 0, 0, 0)$															
<i>Colinealidad = Moderada; P=10; n=25; S=1000</i>															
SR	1385.88					138.59					13.86				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
<i>AIC_{HKB}</i>	1.1	0.6	63.7	34.6	0.0	0.0	0.4	36.2	63.0	0.4	0.0	0.0	20.3	71.4	8.3
<i>AIC_{LW}</i>	1.0	0.6	66.0	32.4	0.0	0.0	0.2	54.1	42.5	3.2	0.0	0.0	42.4	41.3	16.3
<i>AIC_{DK}</i>	0.0	0.0	37.2	59.4	3.4	0.2	0.8	18.8	74.3	5.9	0.1	1.5	18.3	73.1	7.0
<i>C_{HKB}</i>	0.0	0.0	8.7	90.2	1.1	0.0	0.0	6.7	91.3	2.0	0.0	0.0	21.6	69.8	8.6
<i>C_{LW}</i>	0.0	0.0	65.7	0.0	34.3	0.0	0.0	45.7	0.0	54.3	0.0	0.0	42.9	5.6	51.5
<i>C_{DK}</i>	0.0	0.0	24.2	71.4	4.4	0.1	0.7	18.5	73.7	7.0	0.1	1.1	18.7	72.8	7.3
<i>MC_{pHKB}</i>	0.0	0.0	81.7	0.0	18.3	0.0	0.0	62.5	0.0	37.5	0.0	0.0	50.3	3.9	45.8
<i>MC_{pLW}</i>	0.0	0.0	60.3	0.0	39.7	0.0	0.0	41.8	0.0	58.2	0.0	0.0	40.7	0.1	59.2
<i>MC_{pDK}</i>	0.0	0.0	58.5	1.0	40.5	0.0	0.0	54.4	4.1	41.5	0.0	0.0	45.3	21.6	33.1
<i>IC₉₅</i>	0.0	64.4	0.3	35.3	0.0	0.2	13.9	7.0	78.8	0.1	0.1	0.0	29.0	56.6	7.7
<i>IC₉₉</i>	0.2	55.0	0.6	44.2	0.0	0.2	7.4	12.9	79.1	0.2	0.0	0.0	34.3	29.3	11.3
<i>LASSO</i>	0.0	0.0	60.4	39.6	0.0	0.1	8.9	14.0	76.9	0.1	0.2	67.0	0.3	32.5	0.0
<i>EN</i>	0.0	36.9	7.3	55.8	0.0	0.0	14.4	48.1	34.7	0.1	0.0	7.8	30.0	33.5	0.6

El conteo falla para *IC₉₅*, *IC₉₉* y *EN* ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.9: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (0,145, 0,001, -0,475, 0,39, -0,662, -0,383, 0, 0, 0, 0)$														
		Colinealidad = Alta; P=10; n=25; S=1000														
SR		1324.23					132.42					13.24				
σ^2		0.01					0.1					1				
Categoría		MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}		0.6	0.5	66.1	32.8	0.0	0.1	0.0	51.4	47.9	0.6	0.1	0.1	41.4	48.4	10.0
AIC_{LW}		0.5	0.3	71.2	28.0	0.0	0.0	0.0	65.1	32.1	2.8	0.0	0.0	54.4	32.3	13.3
AIC_{DK}		0.1	0.1	52.8	47.0	0.0	0.0	0.4	38.6	59.9	1.1	0.2	3.7	27.9	60.0	8.2
C_{HKB}		0.0	0.0	93.0	4.0	3.0	0.0	0.0	42.7	48.3	9.0	0.0	0.0	43.4	46.1	10.5
C_{LW}		0.0	0.0	71.4	0.0	28.6	0.0	0.0	55.3	0.0	44.7	0.0	0.0	52.2	4.2	43.6
C_{DK}		0.0	0.1	48.1	51.3	0.5	0.1	0.3	29.3	63.8	6.5	0.2	2.8	25.6	62.9	8.5
MCp_{HKB}		0.0	0.0	94.5	0.0	5.5	0.0	0.0	71.5	1.2	27.3	0.0	0.0	62.7	6.0	31.3
MCp_{LW}		0.0	0.0	69.5	0.0	30.5	0.0	0.0	52.8	0.0	47.2	0.0	0.0	44.7	0.0	55.3
MCp_{DK}		0.0	0.0	90.0	1.3	8.7	0.0	0.0	62.7	13.0	24.3	0.0	0.0	57.2	17.4	25.4
IC_{95}		0.2	52.1	0.9	46.8	0.0	0.5	4.7	13.8	80.9	0.1	0.0	0.1	25.4	57.1	9.7
IC_{99}		0.7	39.5	2.8	57.0	0.0	0.4	1.8	28.3	68.9	0.1	0.0	0.0	30.6	31.0	12.3
LASSO		0.0	0.0	79.2	20.8	0.0	0.3	2.4	27.3	70.0	0.0	0.0	57.4	1.0	41.6	0.0
EN		3.5	37.4	19.5	39.6	0.0	0.3	11.8	49.5	37.7	0.0	0.1	6.8	38.9	33.6	0.4

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

En el cuadro 7.10 se estructuran tres columnas, la primera Error cuadrático Medio Total (ECMT), la segunda el Error Cuadrático Medio de Predicción (ECMP), y el indicador de calidad de la selección (F). Estas tres columnas se repiten para cada varianza estudiada. Se puede ver que cuando la varianza es pequeña, el método que consiguió una mejor estimación de los coeficientes y seleccionar mejor las variables fue AIC_{HKB} pero el que consiguió hacer mejores predicciones fue C_{HKB} . Por otro lado cuando la varianza es alta ($\sigma^2 = 1$) los métodos que consiguieron una mejor estimación de los coeficientes fueron AIC_{HKB} , C_{HKB} , y MCp_{DK} , pero el método que obtuvo un ECMP más pequeño fue MCp_{LW} , mientras que el que seleccionó mejor las variables fue LASSO con una notoria ventaja sobre los demás métodos.

Cuadro 7.10: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (0,275, -0,121, -0,208, 0,428, 0,423, 0,617, -0,351, 0, 0, 0, 0)$								
	Colinealidad = Baja; P=10; n=25; S=1000								
SR	2064.28			206.43			20.64		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00006	0.50031	0.87410	0.00044	0.47362	0.58943	0.00087	1.58770	0.39932
AIC_{LW}	0.00009	0.51838	0.87283	0.00049	0.51387	0.51241	0.00277	1.60475	0.31252
AIC_{DK}	0.00100	1.08635	0.32627	0.00100	1.18213	0.39019	0.00100	2.22605	0.47006
C_{HKB}	0.00115	0.36667	0.37868	0.00105	0.45709	0.40820	0.00087	1.59243	0.38716
C_{LW}	0.00108	0.46049	0.04571	0.00142	0.56954	0.11114	0.00098	1.57785	0.18400
C_{DK}	0.00100	1.09481	0.31031	0.00100	1.19756	0.37589	0.00100	2.23287	0.45636
MCp_{HKB}	0.00120	0.44224	0.25086	0.00129	0.54993	0.20486	0.00096	1.65654	0.17262
MCp_{LW}	0.00108	0.46035	0.04571	0.00142	0.56954	0.10971	0.00098	1.57066	0.15257
MCp_{DK}	0.00120	0.45701	0.18643	0.00119	0.55267	0.17344	0.00087	1.64973	0.20793
IC_{95}	0.00013	0.52621	0.76996	0.00080	0.61142	0.72072	0.00172	1.67883	0.38329
IC_{99}	0.00013	0.52613	0.77609	0.00064	0.60653	0.69961	0.00134	1.64995	0.26052
$LASSO$	0.00035	0.60771	0.60750	0.00044	0.54436	0.66307	0.00525	1.91474	0.72309
EN	0.00008	0.34825	0.73045	0.00104	0.57792	0.57282	0.00100	1.87426	0.34190
R_{HKB}	0.00012	0.48134	-	0.00045	0.51439	-	0.00175	1.63315	-
R_{LW}	0.00016	0.52542	-	0.00093	0.61114	-	0.00246	1.68400	-
R_{DK}	0.00064	0.41829	-	0.00088	0.56144	-	0.00097	1.63974	-
MCO	0.00014	0.53009	-	0.00138	0.65728	-	0.00761	2.07139	-

En el cuadro 7.11 se puede ver que cuando la varianza es pequeña, los métodos que consiguieron una mejor estimación de los coeficientes fueron IC_{95} , IC_{99} y EN sin mostrar diferencias significativa respecto R_{HKB} , R_{LW} y MCO , pero este empate no se dio en la predicción puesto que el mejor resultado se obtuvo bajo (C_{HKB}). Por otro lado el método que evidenció mejores resultados respecto de la selección de variables fue AIC_{HKB} . Otro detalle particular es que mientras los demás métodos muestran un deterioro en calidad de la selección en la medida que aumenta la varianza, LASSO es la excepción mostrando una mejora en la selección de variables conforme aumenta la varianza.

Cuadro 7.11: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (-0,211, 0,0559, -0,051, -0,717, -0,567, 0,205, 0,269, 0, 0, 0, 0)$								
	Colinealidad = Moderada; P=10; n=25; S=1000								
SR	1385.88			138.59			13.86		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00007	0.02214	0.77175	0.00081	0.15548	0.53015	0.00104	1.18882	0.37466
AIC_{LW}	0.00009	0.02217	0.77135	0.00101	0.18872	0.45884	0.00543	1.55145	0.31038
AIC_{DK}	0.00076	0.70453	0.45786	0.00099	0.95785	0.43317	0.00097	1.86436	0.44854
C_{HKB}	0.0011	0.1075	0.4528	0.0011	0.1765	0.4427	0.0010	1.1916	0.3686
C_{LW}	0.0011	0.0955	0.1877	0.0013	0.1775	0.1306	0.0013	1.1114	0.1477
C_{DK}	0.0010	0.8311	0.3822	0.0010	0.9812	0.4147	0.0010	1.8751	0.4410
MCp_{HKB}	0.0012	0.0912	0.2334	0.0012	0.1644	0.1786	0.0011	1.1801	0.1569
MCp_{LW}	0.0011	0.1002	0.1723	0.0013	0.1810	0.1194	0.0013	1.1153	0.1165
MCp_{DK}	0.0012	0.0962	0.1699	0.0012	0.1801	0.1704	0.0008	1.1914	0.2063
IC_{95}	0.0001	0.0321	0.7557	0.0026	0.2089	0.6745	0.0043	1.3680	0.3778
IC_{99}	0.0001	0.0318	0.7561	0.0037	0.2085	0.6389	0.0034	1.3020	0.2466
$LASSO$	0.0002	0.5491	0.4882	0.0006	0.3208	0.6276	0.0124	2.6588	0.7249
EN	0.0001	0.0726	0.7118	0.0008	0.3815	0.5544	0.0008	1.9602	0.3514
R_{HKB}	0.0001	0.0226	-	0.0009	0.1541	-	0.0034	1.4416	-
R_{LW}	0.0001	0.0319	-	0.0023	0.2121	-	0.0047	1.5067	-
R_{DK}	0.0006	0.0622	-	0.0009	0.2652	-	0.0009	1.2583	-
MCO	0.0001	0.0337	-	0.0034	0.2731	-	0.0124	2.6408	-

En el cuadro 7.12 se puede ver que cuando la varianza es pequeña los métodos que consiguieron una mejor estimación de los coeficientes fueron EN , R_{HKB} , siendo 4.5 veces más precisos que MCO , mientras que en la predicción el mejor resultado se obtuvo con IC_{99} . Por otro lado el método que evidenció mejores resultados respecto de la selección de variables fue EN . Además mientras los demás métodos muestran un deterioro en calidad de la selección en la medida que aumenta la varianza, $LASSO$ es la excepción mostrando una mejora en la selección de variables conforme aumenta la varianza.

Cuadro 7.12: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,145, 0,001, -0,475, 0,39, -0,662, -0,383, 0, 0, 0, 0)$									
<i>Colinealidad = Alta; P=10; n=25; S=1000</i>									
SR	1324.23			132.42			13.24		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00046	0.85115	0.71353	0.00099	1.19472	0.50316	0.00077	2.35948	0.37627
AIC_{LW}	0.00046	0.83769	0.69553	0.00231	1.19512	0.41718	0.00153	2.66727	0.31230
AIC_{DK}	0.00065	1.01074	0.58910	0.00071	1.10187	0.50321	0.00100	1.92723	0.46970
C_{HKB}	0.0015	2.0804	0.3012	0.0016	1.8209	0.3728	0.0008	2.3204	0.3649
C_{LW}	0.0019	2.4043	0.2040	0.0014	2.1185	0.1580	0.0015	2.6400	0.1734
C_{DK}	0.0010	1.3519	0.4527	0.0014	1.2660	0.4335	0.0010	1.9038	0.4648
MC_{PHKB}	0.0015	2.1383	0.2700	0.0013	1.8711	0.2085	0.0014	2.3979	0.2051
MC_{PLW}	0.0019	2.4132	0.1986	0.0014	2.1427	0.1509	0.0015	2.7017	0.1277
MC_{PDK}	0.0015	2.2275	0.2698	0.0014	1.8983	0.2412	0.0016	2.3738	0.2471
IC_{95}	0.0007	0.8274	0.7704	0.0020	1.1575	0.6381	0.0037	2.3930	0.3519
IC_{99}	0.0006	0.8266	0.7759	0.0013	1.1387	0.5901	0.0037	2.3735	0.2241
$LASSO$	0.0004	0.9642	0.6503	0.0005	1.4736	0.6193	0.0665	5.2838	0.7143
EN	0.0002	1.4491	0.7876	0.0007	1.4350	0.5749	0.0022	2.3477	0.3776
R_{HKB}	0.0002	0.8441	-	0.0011	1.0974	-	0.0250	3.2748	-
R_{LW}	0.0007	0.8293	-	0.0021	1.1447	-	0.0055	2.5605	-
R_{DK}	0.0004	1.0382	-	0.0008	1.2338	-	0.0008	2.0779	-
MCO	0.0009	0.8315	-	0.0040	1.4160	-	0.0827	7.5086	-

7.2.1.1.2 50 Observaciones.

Se puede ver en el cuadro 7.13 que seleccionar el modelo correcto (MB) en general es muy poco frecuente incluso para métodos como LASSO y Elastic Net, por otra parte AIC_{LW} obtuvo mejor puntuación con 3.4% de las simulaciones cuando la varianza es pequeña. Por otra parte, los métodos basados en Regresión Ridge conforme aumenta la varianza, la capacidad de seleccionar disminuye pero no queda claro cual tiene un mejor comportamiento. Otro detalle ha destacar es que a medida que aumenta la varianza LASSO parece mejorar su comportamiento no así *ElasticNet*, puesto su capacidad para seleccionar se ve afectada negativamente con el aumento de la varianza.

Magister en Matemática Mención Estadística

Cuadro 7.13: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,275, -0,121, -0,208, 0,428, 0,423, 0,617, -0,351, 0, 0, 0, 0)$ <i>Colinealidad = Baja; P=10; n=50; S=1000</i>															
SR	3278.93					327.89					32.79				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	3.3	2.9	53.4	40.4	0.0	0.1	2.1	35.7	62.1	0.0	0.1	0.0	8.5	88.6	2.8
AIC_{LW}	3.4	2.8	54.6	39.2	0.0	0.1	0.2	55.1	44.6	0.0	0.0	0.0	48.9	41.8	9.3
AIC_{DK}	0.2	0.2	43.0	56.6	0.0	0.0	0.0	26.5	66.1	7.4	0.0	0.2	24.6	66.0	9.2
C_{HKB}	0.0	0.0	75.6	23.4	1.0	0.0	0.0	0.5	98.8	0.7	0.0	0.0	8.7	88.4	2.9
C_{LW}	0.0	0.0	90.3	0.0	9.7	0.0	0.0	90.2	0.0	9.8	0.0	0.0	80.3	0.0	19.7
C_{DK}	0.0	0.0	43.7	56.3	0.0	0.0	0.0	22.0	67.3	10.7	0.0	0.1	24.7	65.4	9.8
MC_{pHKB}	0.0	0.0	99.6	0.0	0.4	0.0	0.0	77.5	0.0	22.5	0.0	0.0	52.8	0.0	47.2
MC_{pLW}	0.0	0.0	91.8	0.0	8.2	0.0	0.0	91.9	0.0	8.1	0.0	0.0	87.2	0.0	12.8
MC_{pDK}	0.0	0.0	91.4	0.4	8.2	0.0	0.0	56.4	3.0	40.6	0.0	0.0	49.4	12.3	38.3
IC_{95}	0.0	86.2	0.0	13.8	0.0	0.1	39.6	0.5	59.8	0.0	0.1	1.1	24.7	70.5	2.8
IC_{99}	0.0	82.0	0.0	18.0	0.0	0.6	26.8	1.7	70.9	0.0	0.0	0.2	37.0	48.3	6.4
$LASSO$	0.0	0.0	87.4	12.6	0.0	0.3	7.9	12.6	79.2	0.0	0.0	73.0	0.0	27.0	0.0
EN	1.3	62.6	9.1	27.0	0.0	0.5	25.4	35.9	36.8	0.0	0.0	13.7	28.2	33.4	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.14: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,076, 0,073, -0,332, -0,312, -0,556, -0,511, 0,461, 0, 0, 0, 0)$															
Colinealidad = Moderada; P=10; n=50; S=1000															
SR	3143.86					314.39					31.44				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	17.5	10.8	42.1	29.6	0.0	1.4	2.8	28.6	67.2	0.0	0.1	0.3	14.6	80.0	5.0
AIC_{LW}	17.5	10.0	44.4	28.1	0.0	1.3	0.8	50.0	47.7	0.2	0.1	0.0	46.1	38.3	15.5
AIC_{DK}	0.4	1.0	27.2	71.2	0.2	0.0	0.1	15.8	76.8	7.3	0.2	0.1	12.0	77.4	10.3
C_{HKB}	0.0	0.0	85.5	8.7	5.8	0.0	0.0	9.0	90.4	0.6	0.1	0.1	15.9	78.4	5.5
C_{LW}	0.0	0.0	92.8	0.0	7.2	0.0	0.0	26.6	0.0	73.4	0.0	0.0	17.3	0.0	82.7
C_{DK}	0.0	0.0	27.5	66.1	6.4	0.0	0.0	16.4	73.5	10.1	0.1	0.1	13.3	75.9	10.6
MC_{PHKB}	0.0	0.0	86.4	0.0	13.6	0.0	0.0	55.8	0.0	44.2	0.0	0.0	46.1	0.0	53.9
MC_{PLW}	0.0	0.0	90.2	0.0	9.8	0.0	0.0	21.6	0.0	78.4	0.0	0.0	10.4	0.0	89.6
MC_{PDK}	0.0	0.0	51.1	0.0	48.9	0.0	0.0	47.8	3.4	48.8	0.0	0.0	50.6	13.2	36.2
IC_{95}	0.1	92.2	0.0	7.7	0.0	1.4	48.3	0.9	49.4	0.0	0.2	0.3	29.9	60.7	5.4
IC_{99}	0.1	89.3	0.0	10.6	0.0	2.0	35.6	3.0	59.4	0.0	0.0	0.0	42.5	33.2	7.5
$LASSO$	0.0	0.0	53.6	46.4	0.0	1.5	5.9	16.7	75.9	0.0	0.0	65.9	0.0	34.1	0.0
EN	0.0	66.6	0.5	32.9	0.0	0.0	30.6	27.3	40.7	0.0	0.0	12.2	30.0	36.8	0.5

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.15: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,121, 0,191, -0,109, -0,565, -0,655, 0,295, -0,319, 0, 0, 0, 0)$															
Colinealidad = Alta; P=10; n=50; S=1000															
SR	2826.32					282.63					28.26				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	10.0	4.1	50.9	35.0	0	0.3	0.1	43.4	56.2	0.0	0.1	0.0	43.6	52.2	4.1
AIC_{LW}	9.5	3.1	55.0	32.4	0	0.1	0.0	58.7	40.8	0.4	0.0	0.0	54.8	32.6	12.6
AIC_{DK}	4.5	2.6	43.5	49.4	0	0.1	0.6	34.4	64.3	0.6	0.0	1.1	32.3	60.2	6.4
C_{HKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	82.0	16.8	1.2	0.0	0.1	44.9	50.0	5.0
C_{LW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.5	0.0	1.5	0.0	0.0	75.3	0.0	24.7
C_{DK}	0.0	0.0	100.0	0.0	0.0	0.0	0.1	82.4	16.0	1.5	0.0	1.2	36.1	55.3	7.4
MC_{PHKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	96.9	0.0	3.1	0.0	0.0	74.3	3.0	22.7
MC_{PLW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.6	0.0	1.4	0.0	0.0	75.1	0.0	24.9
MC_{PDK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	95.6	0.3	4.1	0.0	0.0	68.2	12.7	19.1
IC_{95}	0.5	78.3	0.2	21.0	0.0	1.2	13.5	11.2	74.1	0.0	0.0	0.0	47.3	30.8	5.6
IC_{99}	1.1	71.4	0.5	27.0	0.0	0.3	6.3	24.8	68.5	0.0	0.0	0.0	45.2	8.9	5.9
$LASSO$	0.0	0.0	85.6	14.4	0.0	1.1	2.8	31.0	65.1	0.0	0.1	57.9	0.3	41.7	0.0
EN	0.0	48.1	0.7	51.2	0.0	0.0	20.1	26.0	53.6	0.0	0.0	9.1	36.3	39.1	1.1

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

En el cuadro 7.16 se puede ver que cuando la varianza es pequeña los métodos que consiguieron una mejor estimación de los coeficientes fueron AIC_{HKB} , AIC_{LW} , IC_{95} , IC_{99} y no mostraron una diferencia significativa respecto MCO , pero este empate no se dio en la predicción, puesto que el mejor resultado se obtuvo realizando una Regresión Ridge (R_{DK}). Por otro lado el método que evidenció mejores resultados respecto de la selección de variables fue AIC_{LW} . Otro detalle particular es que mientras los demás métodos muestran un deterioro en calidad de la selección en la medida que aumenta la varianza, LASSO es la excepción mostrando una mejora en la selección de variables conforme aumenta la varianza.

Cuadro 7.16: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (0,274, 0,385, -0,049, -0,176, -0,310, 0,803, -0,050, 0, 0, 0, 0)$								
	Colinealidad = Baja; P=10; n=50; S=1000								
SR	3278.93			327.89			32.79		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00001	0.12580	0.80691	0.00032	0.22045	0.61425	0.00114	1.16780	0.44521
AIC_{LW}	0.00001	0.12905	0.80727	0.00027	0.25338	0.55816	0.00360	1.64886	0.36127
AIC_{DK}	0.00098	0.25435	0.48719	0.00067	0.41712	0.38207	0.00103	1.51777	0.38290
C_{HKB}	0.00095	0.15467	0.31496	0.00099	0.12274	0.44491	0.00114	1.16500	0.43863
C_{LW}	0.00113	0.19216	0.25800	0.00119	0.28466	0.25771	0.00127	1.35877	0.22943
C_{DK}	0.00098	0.31644	0.39378	0.00100	0.43659	0.33569	0.00112	1.52521	0.37314
$MC_{p_{HKB}}$	0.00113	0.19261	0.28457	0.00117	0.26187	0.22143	0.00113	1.28251	0.15086
$MC_{p_{LW}}$	0.00113	0.19207	0.26229	0.00119	0.28321	0.26257	0.00127	1.34502	0.24914
$MC_{p_{DK}}$	0.00086	0.14556	0.26214	0.00116	0.22896	0.17186	0.00112	1.26147	0.18775
IC_{95}	0.00001	0.14163	0.75768	0.00031	0.27800	0.73311	0.00612	1.64056	0.48013
IC_{99}	0.00001	0.14163	0.75951	0.00032	0.27715	0.72292	0.00601	1.60537	0.37016
LASSO	0.00025	0.26737	0.55085	0.00019	0.20967	0.67408	0.00305	1.68961	0.73383
EN	0.00002	0.11867	0.77695	0.00029	0.31606	0.61272	0.00100	1.70932	0.40462
R_{HKB}	0.00009	0.13009	-	0.00022	0.20907	-	0.00337	1.34196	-
R_{LW}	0.00009	0.14157	-	0.00026	0.27667	-	0.00578	1.66777	-
R_{DK}	0.00010	0.11047	-	0.00073	0.16739	-	0.00074	1.19357	-
MCO	0.00001	0.14203	-	0.00028	0.28367	-	0.00797	1.88288	-

Magister en Matemática Mención Estadística

En el cuadro 7.17 se puede ver que cuando la varianza es pequeña los métodos que consiguieron una mejor estimación de los coeficientes fueron AIC_{HKB} , AIC_{DK} , R_{HKB} , R_{DK} . Respecto de la predicción el mejor resultado se obtuvo realizando C_{LW} . Por otro lado el método que evidenció mejores resultados respecto de la selección de variables fue AIC_{LW} pero los métodos IC_{95} , IC_{99} , $LASSO$ y $ElasticNet$ tienen resultados similares. Además se evidencia que todos los métodos tienden a disminuir su capacidad para seleccionar variables, mientras que aumentan los errores al estimar los coeficientes y al hacer predicciones cuando aumenta la varianza.

Cuadro 7.17: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,076, 0,073, -0,332, -0,312, -0,556, -0,511, 0,461, 0, 0, 0, 0)$									
Colinealidad = Moderada; P=10; n=50; S=1000									
SR	3143.86			314.39			31.44		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	0.18381	0.89795	0.00068	0.30162	0.73039	0.00142	1.10518	0.42736
AIC_{LW}	0.00003	0.18578	0.89949	0.00038	0.32846	0.69322	0.00557	1.45227	0.33107
AIC_{DK}	0.00090	0.09552	0.53242	0.00099	0.21319	0.39670	0.00102	1.21511	0.39405
C_{HKB}	0.00135	0.01571	0.26522	0.00089	0.14551	0.37797	0.00142	1.09235	0.41884
C_{LW}	0.00163	0.01371	0.26514	0.00150	0.13700	0.07600	0.00170	1.12706	0.04943
C_{DK}	0.00101	0.09499	0.35617	0.00100	0.21070	0.34374	0.00102	1.21426	0.39086
MCp_{HKB}	0.00149	0.01581	0.24686	0.00185	0.10454	0.15943	0.00186	1.06895	0.13171
MCp_{LW}	0.00163	0.01501	0.25771	0.00150	0.14447	0.06171	0.00170	1.13752	0.02971
MCp_{DK}	0.00141	0.01449	0.14600	0.00187	0.10264	0.14636	0.00149	1.06942	0.19417
IC_{95}	0.00004	0.19174	0.76748	0.00058	0.30373	0.76739	0.00846	1.38884	0.43892
IC_{99}	0.00004	0.19169	0.77201	0.00041	0.30357	0.77027	0.00811	1.33938	0.31938
$LASSO$	0.00037	0.08096	0.75758	0.00055	0.23896	0.74247	0.00801	1.58842	0.73104
EN	0.00004	0.22660	0.76000	0.00051	0.31705	0.66566	0.01147	2.01246	0.42651
R_{HKB}	0.00003	0.19149	-	0.00068	0.28111	-	0.00423	1.29580	-
R_{LW}	0.00003	0.19177	-	0.00054	0.30284	-	0.00807	1.44377	-
R_{DK}	0.00004	0.15393	-	0.00091	0.21055	-	0.00080	1.16433	-
MCO	0.00004	0.19176	-	0.00054	0.30387	-	0.01041	1.51143	-

Magister en Matemática Mención Estadística

En el cuadro 7.18 se puede ver qué general la estimación de los coeficientes, las predicciones y la selección de variables empeora para todos los métodos en la medida que aumenta la varianza, sin embargo LASSO es capaz de mejorar las selección conforme aumenta la varianza. Mientras que *ElasticNet* logra realizar buenas estimaciones cuando la varianza es de nivel moderado ($\sigma^2 = 0,1$).

Cuadro 7.18: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,121, 0,191, -0,109, -0,565, -0,655, 0,295, -0,319, 0, 0, 0, 0)$									
Colinealidad = Alta; P=10; n=50; S=1000									
SR	2826.32			282.63			28.26		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00010	0.70201	0.83005	0.00138	0.97740	0.63003	0.00113	1.98051	0.42535
AIC_{LW}	0.00010	0.70378	0.82879	0.00195	1.00725	0.56389	0.00401	2.90959	0.32116
AIC_{DK}	0.00010	0.68030	0.72131	0.00101	0.77460	0.51929	0.00141	1.81557	0.44012
C_{HKB}	0.00199	0.91875	0.28571	0.00221	1.04903	0.31530	0.00113	1.96144	0.39773
C_{LW}	0.00207	0.96315	0.28571	0.00241	1.07897	0.28143	0.00180	2.06172	0.21514
C_{DK}	0.00206	0.95798	0.28571	0.00239	0.98994	0.33164	0.00141	1.74596	0.42044
MC_{PHKB}	0.00199	0.91875	0.28571	0.00221	1.01393	0.27686	0.00174	1.90392	0.22252
MC_{PLW}	0.00207	0.96315	0.28571	0.00241	1.07781	0.28171	0.00180	2.07819	0.21457
MC_{PDK}	0.00206	0.95798	0.28571	0.00239	1.08232	0.27535	0.00110	1.97324	0.25381
IC_{95}	0.00006	0.79602	0.77822	0.00203	1.10415	0.72879	0.00206	2.20126	0.32774
IC_{99}	0.00007	0.79563	0.78672	0.00203	1.02453	0.71180	0.00148	1.83839	0.19276
$LASSO$	0.00025	0.52241	0.66537	0.00068	0.93604	0.69460	0.00568	4.17173	0.72872
EN	0.00041	0.92083	0.75218	0.00057	1.10207	0.61091	0.00100	2.25118	0.41391
R_{HKB}	0.00007	0.76838	–	0.00118	0.93866	–	0.00249	2.58092	–
R_{LW}	0.00007	0.80001	–	0.00222	1.16533	–	0.00507	3.12398	–
R_{DK}	0.00064	0.80054	–	0.00088	0.95535	–	0.00096	1.95870	–
MCO	0.00007	0.80198	–	0.00244	1.25889	–	0.00786	6.18105	–

Magister en Matemática Mención Estadística

7.2.1.1.3 100 Observaciones.

Se puede ver en el cuadro 7.19 que seleccionar el modelo correcto (MB) en general es muy poco frecuente incluso para métodos como LASSO y Elastic Net, pero con bastante frecuencia se seleccionan modelos con más variables de las necesarias (B). Por su parte en el cuadro 7.20 y en el cuadro 7.21 se ve el mismo comportamiento, solo que con el aumento del nivel de colinealidad mejora calidad de la selección cuando las varianzas son pequeñas o medianas, esto último no se aprecia con claridad cuando la varianza es alta.

Cuadro 7.19: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,180, -0,935, 0,009, -0,176, -0,145, -0,178, 0,102, 0, 0, 0, 0)$ <i>Colinealidad = Baja; P=10; n=100; S=1000</i>															
SR	5926.46					592.65					59.26				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
<i>AIC_{HKB}</i>	7.7	5.8	45.1	41.4	0.0	0.5	17.0	33.1	49.4	0.0	0.0	0.4	4.0	94.4	1.2
<i>AIC_{LW}</i>	7.8	5.6	46.0	40.6	0.0	0.3	0.1	51.7	47.9	0.0	0.0	0.0	50.4	45.4	4.2
<i>AIC_{DK}</i>	2.1	3.8	35.9	49.0	9.2	0.0	0.2	18.4	65.2	16.2	0.0	0.0	19.4	66.5	14.1
<i>C_{HKB}</i>	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.2	4.1	94.4	1.3
<i>C_{LW}</i>	0.0	0.0	100.0	0.0	0.0	0.0	0.0	60.5	0.0	39.5	0.0	0.0	42.6	0.0	57.4
<i>C_{DK}</i>	0.0	0.0	17.7	69.9	12.4	0.0	0.1	15.0	65.2	19.7	0.0	0.1	19.4	65.8	14.7
<i>MC_{pHKB}</i>	0.0	0.0	39.5	0.0	60.5	0.0	0.0	37.8	0.0	62.2	0.0	0.0	57.7	0.0	42.3
<i>MC_{pLW}</i>	0.0	0.0	100.0	0.0	0.0	0.0	0.0	56.2	0.0	43.8	0.0	0.0	33.4	0.0	66.6
<i>MC_{pDK}</i>	0.0	0.0	42.7	0.0	57.3	0.0	0.0	41.8	0.0	58.2	0.0	0.0	50.7	8.7	40.6
<i>IC₉₅</i>	0.0	94.2	0.0	5.8	0.0	0.0	55.0	0.0	45.0	0.0	0.1	3.2	7.9	88.5	0.3
<i>IC₉₉</i>	0.0	92.8	0.0	7.2	0.0	0.1	45.5	0.0	54.4	0.0	0.1	0.6	18.9	79.7	0.6
<i>LASSO</i>	0.2	0.0	74.7	25.1	0.0	0.9	10.9	7.3	80.9	0.0	0.0	66.5	0.0	33.5	0.0
<i>EN</i>	0.1	60.4	3.2	36.3	0.0	0.3	22.8	41.1	35.6	0.0	0.0	16.4	29.1	39.0	0.0

El conteo falla para *IC₉₅*, *IC₉₉* y *EN* ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.20: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,286, -0,583, -0,575, -0,138, -0,102, -0,345, 0,317, 0, 0, 0, 0)$															
<i>Colinealidad = Moderada; P=10; n=100; S=1000</i>															
SR	5913.98					591.40					59.14				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	44.8	31.6	10.3	13.3	0.0	2.3	23.2	28.7	45.8	0.0	0.1	1.8	7.8	87.7	2.6
AIC_{LW}	45.1	30.9	10.8	13.2	0.0	2.1	2.7	55.6	39.6	0.0	0.0	0.0	50.1	44.0	5.9
AIC_{DK}	16.2	14.7	5.7	63.4	0.0	0.0	0.9	20.4	78.6	0.1	0.0	0.4	25.7	62.4	11.5
C_{HKB}	0.0	0.0	99.6	0.4	0.0	0.0	0.0	10.5	89.4	0.1	0.0	0.3	8.5	88.5	2.7
C_{LW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	97.8	0.0	2.2	0.0	0.0	71.6	0.0	28.4
C_{DK}	0.0	0.0	22.0	78.0	0.0	0.0	0.0	32.4	63.8	3.8	0.0	0.3	22.2	65.4	12.1
MC_{PHKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	92.3	0.0	7.7	0.0	0.0	56.3	0.0	43.7
MC_{PLW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.1	0.0	1.9	0.0	0.0	74.8	0.0	25.2
MC_{PDK}	0.0	0.0	99.8	0.0	0.2	0.0	0.0	82.6	0.0	17.4	0.0	0.0	51.3	8.5	40.2
IC_{95}	0.0	99.1	0.0	0.9	0.0	0.0	68.7	0.0	31.3	0.0	0.1	2.2	17.5	79.1	1.1
IC_{99}	0.0	98.7	0.0	1.3	0.0	0.3	59.7	0.2	39.8	0.0	0.0	0.4	32.2	63.3	2.3
$LASSO$	21.1	1.8	70.3	6.8	0.0	5.5	20.8	12.4	61.3	0.0	0.0	65.3	0.1	34.6	0.0
EN	0.0	94.3	0.6	5.1	0.0	1.2	32.0	36.3	30.4	0.0	0.1	15.3	32.5	39.1	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.21: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,003, -0,647, -0,276, -0,459, -0,314, 0,242, 0,371, 0, 0, 0, 0)$															
<i>Colinealidad = Alta; P=10; n=100; S=1000</i>															
SR	5575.93					557.59					55.76				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	52.1	44.3	1.2	2.4	0	4.3	6.0	27.6	62.1	0.0	0.0	1.2	36.3	61.4	1.1
AIC_{LW}	54.2	41.7	1.3	2.8	0	2.2	1.5	47.6	48.7	0.0	0.0	0.0	52.1	30.8	17.1
AIC_{DK}	20.3	21.2	34.3	24.2	0	0.1	1.7	30.6	67.4	0.2	0.1	1.0	33.9	57.8	7.2
C_{HKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	84.9	15.0	0.1	0.0	0.3	40.8	57.5	1.4
C_{LW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	97.3	0.0	2.7
C_{DK}	0.0	0.0	68.0	32.0	0.0	0.0	0.1	68.3	31.6	0.0	0.1	0.3	41.9	52.1	5.6
MC_{PHKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	88.7	0.4	10.9
MC_{PLW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.4	0.0	1.6
MC_{PDK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	82.1	5.6	12.3
IC_{95}	0.2	99.7	0.0	0.1	0.0	1.1	54.4	1.9	42.6	0.0	0.1	0.3	46.6	42.4	5.3
IC_{99}	0.2	99.7	0.0	0.1	0.0	3.2	41.3	4.8	50.7	0.0	0.0	0.0	50.9	18.1	5.1
$LASSO$	0.0	0.0	46.2	53.8	0.0	4.1	11.5	15.5	68.9	0.0	0.0	59.9	0.0	40.1	0.0
EN	0.0	76.7	2.5	20.8	0.0	0.3	28.8	37.5	33.3	0.0	0.1	12.4	36.6	38.9	0.1

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

En el cuadro 7.22 se puede ver que cuando la varianza es pequeña los métodos que consiguieron una mejor estimación de los coeficientes fueron Elastic Net seguido de muy cerca de AIC_{HKB} , AIC_{LW} , R_{HKB} , R_{LW} . Respecto de la predicción el mejor resultado se obtuvo realizando C_{HKB} . Por otro lado el método que evidenció mejores resultados respecto de la selección de variables fue AIC_{LW} pero los métodos AIC_{HKB} , IC_{95} , IC_{99} , LASSO y *ElasticNet* tienen resultados similares y bastante satisfactorios. Además se evidencia que todos los métodos tienden a disminuir su capacidad para seleccionar variables, mientras que aumentan los errores al estimar los coeficientes y los errores en las predicciones cuando aumenta la varianza.

Cuadro 7.22: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (0,180, -0,935, 0,009, -0,176, -0,145, -0,178, 0,102, 0, 0, 0, 0)$								
	Colinealidad = Baja; P=10; n=100; S=1000								
SR	5926.46			592.65			59.26		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00004	0.08298	0.86649	0.00014	0.18187	0.64503	0.00119	1.04305	0.48712
AIC_{LW}	0.00004	0.08562	0.86721	0.00015	0.20765	0.58581	0.00207	1.19506	0.39874
AIC_{DK}	0.00076	0.18409	0.55692	0.00118	0.32007	0.36563	0.00100	1.31801	0.34820
C_{HKB}	0.00110	0.18517	0.58799	0.00109	0.42199	0.53738	0.00119	1.29197	0.48314
C_{LW}	0.00167	0.38520	0.28571	0.00171	0.48086	0.17286	0.00229	1.30963	0.12171
C_{DK}	0.00101	0.68048	0.31941	0.00118	0.76606	0.29716	0.00100	1.44018	0.34159
$MC_{p_{HKB}}$	0.00150	0.22460	0.11286	0.00151	0.44657	0.10800	0.00155	1.30266	0.16486
$MC_{p_{LW}}$	0.00167	0.38520	0.28571	0.00171	0.47709	0.16057	0.00229	1.30143	0.09543
$MC_{p_{DK}}$	0.00136	0.31770	0.12200	0.00151	0.48888	0.11943	0.00147	1.31261	0.17425
IC_{95}	0.00004	0.30743	0.75720	0.00014	0.49775	0.73962	0.00221	1.39237	0.58822
IC_{99}	0.00004	0.30743	0.75971	0.00016	0.49808	0.73637	0.00221	1.38501	0.52499
LASSO	0.00009	0.26522	0.80507	0.00010	0.48591	0.71261	0.00244	1.40324	0.73271
EN	0.00002	0.31143	0.77023	0.00039	0.52008	0.61835	0.00216	1.56770	0.47197
R_{HKB}	0.00005	0.30180	-	0.00011	0.47201	-	0.00115	1.32314	-
R_{LW}	0.00005	0.30742	-	0.00014	0.49748	-	0.00233	1.40266	-
R_{DK}	0.00033	0.24127	-	0.00015	0.45739	-	0.00085	1.31238	-
MCO	0.00004	0.30752	-	0.00014	0.49813	-	0.00244	1.41026	-

En el cuadro 7.23 se puede ver que cuando la varianza es pequeña los métodos que consi-

Magister en Matemática Mención Estadística

guieron una mejor estimación de los coeficientes fueron AIC_{HKB} , AIC_{LW} , AIC_{DK} , R_{HKB} , IC_{95} , IC_{99} , EN , R_{HKB} , R_{LW} , R_{DK} y MCO . Respecto de la predicción el mejor resultado se obtuvo realizando C_{HKB} . Por otro lado el método que evidenció mejores resultados respecto de la selección de variables fue AIC_{HKB} y los métodos AIC_{LW} , AIC_{DK} , IC_{95} , IC_{99} tienen resultados similares y bastante buenos. Además se evidencia que todos los métodos tienden a disminuir su capacidad para seleccionar variables, mientras que aumentan los errores al estimar los coeficientes y los errores en las predicciones cuando aumenta la varianza..

Cuadro 7.23: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,286, -0,583, -0,575, -0,138, -0,102, -0,345, 0,317, 0,0,0,0)$									
Colinealidad = Moderada; P=10; n=100; S=1000									
SR	5913.98			591.40			59.14		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00004	0.19212	0.93578	0.00011	0.26239	0.78288	0.00141	1.06933	0.43283
AIC_{LW}	0.00003	0.19475	0.93660	0.00009	0.27261	0.77321	0.00110	1.18308	0.41313
AIC_{DK}	0.00090	0.16888	0.76399	0.00093	0.26870	0.56383	0.00099	1.27299	0.39909
C_{HKB}	0.00185	0.14953	0.28600	0.00116	0.58862	0.40076	0.00141	1.28735	0.42033
C_{LW}	0.00186	0.15021	0.28571	0.00185	0.40740	0.27943	0.00198	1.26663	0.20457
C_{DK}	0.00138	0.41323	0.38342	0.00102	0.57824	0.37621	0.00116	1.35348	0.36395
MCp_{HKB}	0.00185	0.14836	0.28571	0.00184	0.41170	0.26371	0.00191	1.26539	0.16086
MCp_{LW}	0.00186	0.15021	0.28571	0.00185	0.40741	0.28029	0.00198	1.26472	0.21371
MCp_{DK}	0.00185	0.15038	0.28514	0.00183	0.42828	0.23600	0.00195	1.26959	0.17440
IC_{95}	0.00003	0.82799	0.76386	0.00017	0.89551	0.76809	0.00092	1.52224	0.58740
IC_{99}	0.00003	0.82799	0.76866	0.00017	0.89548	0.77398	0.00043	1.49402	0.49791
$LASSO$	0.00022	0.51002	0.89784	0.00015	0.66250	0.80417	0.00077	1.46955	0.73608
EN	0.00004	0.78835	0.80143	0.00095	0.70649	0.71604	0.00100	1.56282	0.52325
R_{HKB}	0.00003	0.82200	-	0.00015	0.85280	-	0.00033	1.44697	-
R_{LW}	0.00003	0.82797	-	0.00017	0.89553	-	0.00092	1.54382	-
R_{DK}	0.00004	0.76096	-	0.00015	0.73497	-	0.00037	1.36279	-
MCO	0.00003	0.82807	-	0.00017	0.89649	-	0.00099	1.55256	-

En el cuadro 7.24 se puede ver que cuando la varianza es pequeña el método que consiguió menores errores en la estimación de los coeficiente fue AIC_{LW} , IC_{95} , IC_{99} , R_{RHK} , R_{LW} ,

Magister en Matemática Mención Estadística

MCO y los métodos C_{HKB} , C_{LW} , MCp_{HKB} , MCp_{LW} , MCp_{DK} configuran los peores resultados para estimación de coeficientes y selección de variables. Respecto de la predicción el mejor resultado se obtuvo realizando MCp_{HKB} . Por otro lado el método que evidenció mejores resultados respecto de la selección de variables fue AIC_{HKB} y AIC_{LW} .

Cuadro 7.24: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,003, -0,647, -0,276, -0,459, -0,314, 0,242, 0,371, 0, 0, 0, 0)$									
Colinealidad = Alta; P=10; n=100; S=1000									
SR	5575.93			557.59			55.76		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00001	3.33394	0.95320	0.00027	3.29507	0.74622	0.00118	3.58593	0.50154
AIC_{LW}	0.00001	3.38508	0.95564	0.00026	3.55619	0.70431	0.00257	4.80213	0.31786
AIC_{DK}	0.00001	2.67218	0.77232	0.00051	2.35403	0.56606	0.00098	3.14154	0.42652
C_{HKB}	0.00196	1.92762	0.28571	0.00230	1.89698	0.42248	0.00118	2.25306	0.46489
C_{LW}	0.00200	2.03991	0.28571	0.00232	2.06727	0.28571	0.00162	2.29897	0.27800
C_{DK}	0.00100	1.83924	0.43238	0.00232	1.85818	0.40036	0.00098	2.11627	0.42395
MCp_{HKB}	0.00196	2.00359	0.28571	0.00230	1.95726	0.28571	0.00157	2.25324	0.25440
MCp_{LW}	0.00200	2.03991	0.28571	0.00232	2.06848	0.28571	0.00162	2.30202	0.28114
MCp_{DK}	0.00200	2.02256	0.28571	0.00232	1.96890	0.28571	0.00157	2.25637	0.26353
IC_{95}	0.00003	2.41216	0.77766	0.00024	2.45607	0.78692	0.00345	2.60728	0.41157
IC_{99}	0.00003	2.41211	0.78644	0.00024	2.45371	0.79464	0.00323	2.46650	0.26925
$LASSO$	0.00030	2.12677	0.74168	0.00032	2.15583	0.76525	0.00358	2.69714	0.72932
EN	0.00002	2.36555	0.77703	0.00038	2.18473	0.64941	0.00118	2.51407	0.47221
R_{HKB}	0.00002	2.38729	-	0.00021	2.34237	-	0.00175	2.59495	-
R_{LW}	0.00003	2.41228	-	0.00024	2.45495	-	0.00401	2.86503	-
R_{DK}	0.00003	2.36690	-	0.00078	2.20198	-	0.00096	2.34561	-
MCO	0.00003	2.41276	-	0.00024	2.46022	-	0.00437	2.98156	-

7.2.1.2. 15 Variables Disponibles.

7.2.1.2.1 25 Observaciones.

Si se comparan las tablas en los cuadros 7.25, 7.26, 7.27 con las tablas en los cuadros 7.7, 7.8, 7.9 se aprecia que la capacidad de seleccionar correctamente variables en general disminuye con la presencia de más variables disponibles. Ahora esto es manteniendo las consideraciones del caso ya que se estiman coeficientes de modelos diferentes.

Cuadro 7.25: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (-0,0331, -0,684, -0,047, -0,367, 0,013, -0,366, 0,509, 0, 0, 0, 0, 0, 0, 0, 0, 0)$														
		<i>Colinealidad = Baja; P=15; n=25; S=1000</i>														
SR	1343.18					134.32					13.43					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	0.2	1.3	31.5	67.0	0.0	0.0	0.1	4.1	91.6	4.2	0	0.0	2.0	88.5	9.5	
AIC_{LW}	0.2	1.0	36.0	62.8	0.0	0.0	0.0	26.3	72.4	1.3	0	0.0	15.1	64.1	20.8	
AIC_{DK}	0.0	0.2	1.9	87.5	10.4	0.0	2.9	1.0	90.3	5.8	0	7.9	0.7	87.0	4.4	
C_{HKB}	0.0	0.0	0.0	80.0	20.0	0.0	0.1	0.3	89.9	9.7	0.0	0.0	2.3	87.2	10.5	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.9	0.0	99.1	0.0	0.0	20.7	19.7	59.6	
C_{DK}	0.0	0.1	1.2	86.6	12.1	0.0	1.5	1.1	90.5	6.9	0.0	4.1	0.8	89.9	5.2	
MC_{PHKB}	0.0	0.0	1.3	0.0	98.7	0.0	0.0	30.9	0.0	69.1	0.0	0.0	39.9	6.2	53.9	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.5	0.0	99.5	0.0	0.0	7.7	0.0	92.3	
MC_{PDK}	0.0	0.0	33.2	0.4	66.4	0.0	0.0	42.7	2.2	55.1	0.0	0.0	31.0	16.7	52.3	
IC_{95}	0.0	72.1	0.0	27.9	0.0	0.0	17.7	0.4	81.9	0.0	0.0	1.5	6.6	82.5	8.2	
IC_{99}	0.0	64.1	0.0	35.9	0.0	0.0	9.6	2.5	87.9	0.0	0.0	0.3	13.8	62.6	14.8	
$LASSO$	0.1	0.1	52.7	47.1	0.0	0.2	29.6	0.7	69.5	0.0	0.0	74.5	0.0	25.5	0.0	
EN	0.0	23.0	0.5	76.5	0.0	0.0	8.6	25.6	56.7	0.0	0.0	3.6	21.6	42.5	0.8	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.26: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (0,142, 0,088, -0,294, -0,0818, -0,566, -0,712, 0,236, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Moderada; P=15; n=25; S=1000														
SR	1245.49					124.55					12.45					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	0.6	0.8	16.8	81.8	0.0	0	0.6	5.2	91.4	2.8	0	0.0	3.1	81.3	15.6	
AIC_{LW}	0.6	0.7	19.2	79.5	0.0	0	0.0	18.2	69.3	12.5	0	0.0	14.8	60.7	24.5	
AIC_{DK}	0.0	0.7	1.7	89.3	8.3	0	1.6	1.9	86.6	9.9	0	4.7	2.7	83.7	8.9	
C_{HKB}	0.0	0.0	0.0	99.8	0.2	0.0	0.0	0.9	91.4	7.7	0.0	0.0	4.3	79.1	16.6	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.2	0.0	99.8	0.0	0.0	16.2	25.1	58.7	
C_{DK}	0.0	0.2	2.0	88.9	8.9	0.0	0.6	2.1	87.1	10.2	0.0	3.0	2.8	83.8	10.4	
MC_{pHKB}	0.0	0.0	0.2	0.0	99.8	0.0	0.0	8.2	0.0	91.8	0.0	0.0	20.8	9.6	69.6	
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	7.9	0.1	92.0	
MC_{pDK}	0.0	0.0	24.7	1.3	74.0	0.0	0.0	27.3	2.9	69.8	0.0	0.0	24.6	20.1	55.3	
IC_{95}	0.0	64.7	0.0	35.3	0.0	0.0	14.2	0.8	85.0	0.0	0.0	0.8	5.4	83.6	9.2	
IC_{99}	0.0	57.3	0.0	42.7	0.0	0.0	8.1	2.9	88.7	0.2	0.0	0.2	13.5	60.5	17.6	
$LASSO$	0.0	0.0	1.7	98.3	0.0	0.0	12.3	0.4	87.3	0.0	0.0	67.0	0.0	33.0	0.0	
EN	0.0	25.9	3.6	70.5	0.0	0.0	8.5	27.9	61.3	0.0	0.0	3.6	21.7	45.5	2.8	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.27: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (0,234, 0,083, -0,383, -0,494, -0,679, -0,159, 0,245, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Alta; P=15; n=25; S=1000														
SR	892.54					89.25					8.93					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	0.1	0.8	21.6	77.5	0.0	0.0	0.1	10.1	84.4	5.4	0	0.0	12.1	60.4	27.5	
AIC_{LW}	0.1	0.6	24.8	74.5	0.0	0.0	0.0	25.6	62.3	12.1	0	0.0	17.3	58.5	24.2	
AIC_{DK}	0.0	0.0	12.1	87.3	0.6	0.0	0.9	7.9	78.3	12.9	0	3.8	5.2	78.4	12.6	
C_{HKB}	0.0	0.0	1.1	64.9	34.0	0.0	0.0	4.8	65.8	29.4	0.0	0.0	13.2	57.9	28.9	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	5.5	0.0	94.5	0.0	0.0	24.3	15.6	60.1	
C_{DK}	0.0	0.0	0.9	80.8	18.3	0.0	0.8	1.1	80.4	17.7	0.0	2.4	5.2	80.0	12.4	
MC_{pHKB}	0.0	0.0	2.1	0.0	97.9	0.0	0.0	16.4	0.5	83.1	0.0	0.0	30.5	7.9	61.6	
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	4.3	0.0	95.7	0.0	0.0	22.9	0.0	77.1	
MC_{pDK}	0.0	0.0	3.2	2.3	94.5	0.0	0.0	15.8	13.3	70.9	0.0	0.0	24.6	23.4	52.0	
IC_{95}	0.0	45.9	0.0	54.1	0.0	0.0	6.9	0.9	92.0	0.2	0.0	0.5	4.3	86.0	8.3	
IC_{99}	0.0	35.3	0.2	64.5	0.0	0.0	3.0	3.9	91.5	1.4	0.0	0.4	10.7	61.3	19.3	
$LASSO$	0.0	0.0	7.5	92.5	0.0	0.0	6.9	1.8	91.0	0.3	0.0	64.4	0.0	35.6	0.0	
EN	0.0	24.1	3.3	72.6	0.0	0.1	7.4	26.9	64.9	0.1	0.0	3.8	25.5	45.8	3.9	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.28, 7.29,7.30, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza, aunque hay casos como el de C_{LW} que falló en todas las simulaciones a pesar de que la Varianza pequeña y la colinealidad Baja. Ahora si se comparan las tablas desde una perspectiva global se puede ver que los resultados empeoran en la medida que aumenta el nivel de la colinealidad. También es importante comentar que los métodos C_{LW} y MCp_{LW} fallaron en todas las simulaciones para los 3 niveles de correlación cuando la Varianza es Baja.

Cuadro 7.28: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (-0,0331, -0,684, -0,047, -0,367, 0,013, -0,366, 0,509, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$									
Colinealidad = Baja; P=15; n=25; S=1000									
SR	1343.18			134.32			13.43		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00017	0.62136	0.75248	0.00081	0.65475	0.44120	0.00101	1.75196	0.31872
AIC_{LW}	0.00015	0.60869	0.75424	0.00210	0.81119	0.49880	0.00751	2.12579	0.28249
AIC_{DK}	0.00100	1.54656	0.36723	0.00100	1.65433	0.41128	0.00100	2.46428	0.44118
C_{HKB}	0.00112	0.56862	0.16703	0.00081	0.66905	0.30173	0.00101	1.75873	0.31135
C_{LW}	0.00134	0.72113	0.00000	0.00141	0.84708	0.00257	0.00157	1.71677	0.13633
C_{DK}	0.00099	1.55672	0.34873	0.00100	1.66146	0.39677	0.00100	2.47567	0.42335
MCp_{HKB}	0.00132	0.68658	0.00371	0.00119	0.78848	0.08829	0.00132	1.81138	0.13538
MCp_{LW}	0.00134	0.70511	0.00000	0.00141	0.82485	0.00143	0.00157	1.69367	0.02200
MCp_{DK}	0.00132	0.78487	0.09629	0.00119	0.86656	0.12814	0.00104	1.85151	0.14270
IC_{95}	0.00018	0.69409	0.59774	0.00142	1.08711	0.56289	0.00689	1.98530	0.37304
IC_{99}	0.00018	0.69306	0.60587	0.00112	1.07167	0.54666	0.00339	1.97723	0.28163
$LASSO$	0.00046	0.39102	0.81624	0.00141	0.78791	0.60525	0.03348	6.15424	0.55982
EN	0.00059	0.43457	0.63139	0.00100	0.89960	0.46185	0.00112	2.13466	0.27472
R_{HKB}	0.00021	0.75363	-	0.00080	0.92729	-	0.00655	2.32464	-
R_{LW}	0.00018	0.69846	-	0.00130	1.08009	-	0.00830	2.12210	-
R_{DK}	0.00062	0.63977	-	0.00088	0.83912	-	0.00097	1.79230	-
MCO	0.00019	0.68200	-	0.00223	1.41904	-	0.03348	7.41146	-

Magister en Matemática Mención Estadística

Cuadro 7.29: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (0,142, 0,088, -0,294, -0,0818, -0,566, -0,712, 0,236, 0,0, 0,0, 0,0, 0,0)$								
	Colinealidad = Moderada; P=15; n=25; S=1000								
SR	1245.49			124.55			12.45		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00034	0.04611	0.63480	0.00103	0.16530	0.41108	0.00101	1.24865	0.29290
AIC_{LW}	0.00038	0.06209	0.62344	0.00281	0.21996	0.35070	0.00291	1.79134	0.26691
AIC_{DK}	0.00104	0.64426	0.36954	0.00099	0.73607	0.38108	0.00100	1.61151	0.39209
C_{HKB}	0.00130	0.09311	0.37930	0.00103	0.20711	0.34326	0.00101	1.25822	0.28780
C_{LW}	0.00148	0.30039	0.00000	0.00157	0.40475	0.00057	0.00155	1.26219	0.14937
C_{DK}	0.00104	0.65194	0.36095	0.00099	0.74246	0.37307	0.00100	1.61709	0.37533
MCp_{HKB}	0.00144	0.25052	0.00057	0.00154	0.35885	0.02343	0.00101	1.32580	0.09166
MCp_{LW}	0.00148	0.30023	0.00000	0.00157	0.38822	0.00000	0.00159	1.26229	0.02282
MCp_{DK}	0.00141	0.25246	0.07446	0.00152	0.34715	0.08737	0.00141	1.32602	0.13243
IC_{95}	0.00032	0.06273	0.59698	0.00254	0.30791	0.54039	0.00393	1.62274	0.36617
IC_{99}	0.00039	0.06237	0.60605	0.00254	0.28894	0.52429	0.00233	1.43215	0.27661
$LASSO$	0.00056	0.26241	0.45430	0.00139	0.36363	0.52259	0.02310	4.58451	0.55677
EN	0.00019	0.14672	0.59508	0.00151	0.45757	0.45459	0.00157	1.70041	0.27877
R_{HKB}	0.00027	0.03923	-	0.00161	0.20528	-	0.00675	2.08925	-
R_{LW}	0.00032	0.06229	-	0.00245	0.31253	-	0.00620	1.89654	-
R_{DK}	0.00086	0.09534	-	0.00093	0.25101	-	0.00091	1.19647	-
MCO	0.00033	0.06667	-	0.00287	0.50044	-	0.02691	5.34542	-

Cuadro 7.30: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,234, 0,083, -0,383, -0,494, -0,679, -0,159, 0,245, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$									
Colinealidad = Alta; P=15; n=25; S=1000									
SR	892.54			89.25			8.93		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00057	0.33177	0.66120	0.00073	0.32690	0.40448	0.00125	1.28458	0.24452
AIC_{LW}	0.00071	0.37464	0.64970	0.00105	0.44320	0.33589	0.00650	1.58047	0.25761
AIC_{DK}	0.00100	0.69677	0.39307	0.00100	0.84428	0.36231	0.00098	1.74757	0.36425
C_{HKB}	0.00142	0.07631	0.16149	0.00148	0.21537	0.22993	0.00125	1.28864	0.23819
C_{LW}	0.00168	0.04886	0.00000	0.00163	0.18135	0.01571	0.00157	1.17806	0.13048
C_{DK}	0.00100	0.77194	0.26622	0.00100	0.88349	0.31681	0.00098	1.77130	0.36055
MCp_{HKB}	0.00157	0.09594	0.00600	0.00166	0.24374	0.04829	0.00139	1.31213	0.11400
MCp_{LW}	0.00168	0.04886	0.00000	0.00163	0.17698	0.01229	0.00157	1.18914	0.06543
MCp_{DK}	0.00139	0.07725	0.01487	0.00161	0.23573	0.08569	0.00137	1.32073	0.14421
IC_{95}	0.00130	0.42606	0.59716	0.00168	0.58836	0.52200	0.00471	1.34617	0.36270
IC_{99}	0.00125	0.42487	0.60825	0.00081	0.54925	0.49133	0.00414	1.31112	0.26001
$LASSO$	0.00072	0.27443	0.56617	0.00069	0.44146	0.52517	0.04157	9.60867	0.55595
EN	0.00066	0.19378	0.62523	0.00101	0.54107	0.44999	0.00100	1.63150	0.28985
R_{HKB}	0.00055	0.28434	-	0.00168	0.54807	-	0.01200	2.84479	-
R_{LW}	0.00123	0.42358	-	0.00211	0.60489	-	0.00489	1.53793	-
R_{DK}	0.00091	0.18791	-	0.00099	0.32512	-	0.00099	1.22839	-
MCO	0.00162	0.49634	-	0.00947	1.82093	-	0.05327	13.36645	-

7.2.1.2.2 50 Observaciones.

Como se aprecia en los cuadros 7.31, 7.32, 7.33, es muy difícil que un método de selección capture el modelo correcto, el mejor resultado en este sentido lo alcanzo AIC_{HKB} seleccionando el modelo correcto en un 25.3% de las ocasiones, y fue el único método que lo logro para cualquier nivel de correlación y cualquier nivel de varianza. También es interesante observar en estos cuadros es que LASSO para cualquiera de los niveles de Colinealidad mejora su capacidad para seleccionar a medida que aumentaba la varianza. Se puede observar que AIC con cualquiera de los estimadores del parámetro de sesgo es muy competitivo respecto de LASSO.

Magister en Matemática Mención Estadística

Cuadro 7.31: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,136, -0,265, -0,255, -0,37774128, -0,791, 0,182, 0,212, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Baja; P=15; n=50; S=1000															
SR	3364.84					336.48					33.65				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	25.3	61.8	1.0	11.9	0.0	0.2	7.2	7.8	84.8	0.0	0	0.1	0.7	90.5	8.7
AIC_{LW}	27.2	59.7	1.1	12.0	0.0	0.2	1.8	25.8	72.2	0.0	0	0.1	22.8	63.7	13.4
AIC_{DK}	0.1	0.1	4.3	66.5	29.0	0.0	0.0	3.7	70.4	25.9	0	0.8	3.8	80.9	14.5
C_{HKB}	0.0	0.0	0.0	85.0	15.0	0.0	0.0	0.0	93.3	6.7	0.0	0.1	0.8	90.0	9.1
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	0.0	0.0	8.8	0.2	91.0
C_{DK}	0.0	0.0	2.9	65.2	31.9	0.0	0.0	3.2	68.3	28.5	0.0	0.6	3.7	80.4	15.3
MC_{pHKB}	0.0	0.0	8.0	0.0	92.0	0.0	0.0	24.0	0.0	76.0	0.0	0.0	30.0	0.0	70.0
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.2	0.0	99.8	0.0	0.0	3.1	0.0	96.9
MC_{pDK}	0.0	0.0	31.2	0.1	68.7	0.0	0.0	23.9	0.6	75.5	0.0	0.0	30.5	9.0	60.5
IC_{95}	0.0	99.3	0.0	0.7	0.0	0.0	42.4	0.0	57.6	0.0	0.0	0.6	4.3	89.4	4.7
IC_{99}	0.0	98.9	0.0	1.1	0.0	0.0	29.4	0.1	70.5	0.0	0.0	0.3	17.0	67.6	10.7
$LASSO$	0.0	0.0	92.6	7.4	0.0	0.5	19.5	0.6	79.4	0.0	0.0	75.5	0.0	24.5	0.0
EN	0.0	75.1	3.5	21.4	0.0	0.0	18.4	27.1	53.7	0.0	0.0	8.6	21.6	45.5	0.7

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.32: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,119, 0,144, -0,876, -0,06, 0,054, -0,182, -0,397, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Moderada; P=15; n=50; S=1000															
SR	2876.18					287.62					28.76				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	1.1	2.4	26.5	70.0	0.0	0.2	2.6	14.4	82.8	0.0	0	0.2	1.8	84.2	13.8
AIC_{LW}	1.3	1.9	28.2	68.6	0.0	0.2	0.1	28.0	71.7	0.0	0	0.1	21.9	62.1	15.9
AIC_{DK}	0.0	0.3	9.8	68.1	21.8	0.0	0.0	6.8	75.7	17.5	0	0.5	5.4	79.4	14.7
C_{HKB}	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	95.1	4.9	0.0	0.1	2.0	83.8	14.1
C_{LW}	0.0	0.0	0.5	0.0	99.5	0.0	0.0	2.8	0.0	97.2	0.0	0.0	8.3	0.3	91.4
C_{DK}	0.0	0.0	8.7	63.9	27.4	0.0	0.1	6.3	72.8	20.8	0.0	0.2	5.3	78.9	15.6
MC_{pHKB}	0.0	0.0	49.3	0.0	50.7	0.0	0.0	32.0	0.0	68.0	0.0	0.0	24.3	0.0	75.7
MC_{pLW}	0.0	0.0	0.3	0.0	99.7	0.0	0.0	2.0	0.0	98.0	0.0	0.0	4.7	0.0	95.3
MC_{pDK}	0.0	0.0	25.7	0.2	74.1	0.0	0.0	25.0	1.8	73.2	0.0	0.0	28.8	10.8	60.4
IC_{95}	0.0	82.3	0.0	17.7	0.0	0.0	27.6	0.2	72.2	0.0	0.0	0.4	7.6	82.1	8.8
IC_{99}	0.0	75.9	0.0	24.1	0.0	0.1	16.5	0.3	83.1	0.0	0.0	0.2	18.4	54.9	16.2
$LASSO$	0.0	0.0	78.7	21.3	0.0	0.2	11.5	0.8	87.5	0.0	0.0	72.1	0.0	27.9	0.0
EN	0.0	38.5	1.7	59.8	0.0	0.0	11.9	13.9	73.7	0.0	0.0	7.0	18.2	52.2	2.4

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.33: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,148, 0,245, -0,018, 0,012, -0,789, -0,206, -0,501, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Alta; P=15; n=50; S=1000															
SR	2897.37					289.74					28.97				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	0	0.8	23.0	76.2	0.0	0	0.4	13.8	85.7	0.1	0	0.1	5.1	69.9	24.9
AIC_{LW}	0	0.6	24.9	74.5	0.0	0	0.1	32.6	66.6	0.7	0	0.0	23.3	54.2	22.5
AIC_{DK}	0	0.3	13.2	86.5	0.0	0	0.4	14.6	80.3	4.7	0	0.8	9.6	68.9	20.7
C_{HKB}	0.0	0.0	0.1	12.0	87.9	0.0	0.0	0.7	70.6	28.7	0.0	0.1	4.9	68.5	26.5
C_{LW}	0.0	0.0	10.7	0.0	89.3	0.0	0.0	35.7	0.0	64.3	0.0	0.0	37.3	0.5	62.2
C_{DK}	0.0	0.0	0.0	73.9	26.1	0.0	0.1	2.0	69.9	28.0	0.0	0.8	8.0	68.4	22.8
MC_{pHKB}	0.0	0.0	0.3	0.0	99.7	0.0	0.0	14.0	0.0	86.0	0.0	0.0	32.6	1.2	66.2
MC_{pLW}	0.0	0.0	12.1	0.0	87.9	0.0	0.0	39.2	0.0	60.8	0.0	0.0	38.1	0.0	61.9
MC_{pDK}	0.0	0.0	0.3	0.1	99.6	0.0	0.0	8.3	5.6	86.1	0.0	0.0	23.8	17.6	58.6
IC_{95}	0.0	63.5	0.0	36.5	0.0	0.0	9.2	0.2	90.6	0.0	0.0	0.0	12.9	70.8	12.6
IC_{99}	0.0	52.4	0.0	47.6	0.0	0.0	3.0	1.4	95.6	0.0	0.0	0.0	23.9	35.9	20.0
$LASSO$	0.0	0.0	82.1	17.9	0.0	0.0	4.2	1.8	94.0	0.0	0.0	67.3	0.0	32.7	0.0
EN	0.0	23.2	0.4	76.4	0.0	0.0	9.9	15.1	74.5	0.0	0.0	7.2	20.4	50.1	2.8

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Respecto de los cuadros 7.34, 7.35,7.36, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados empeoran en la medida que aumenta el nivel de la colinealidad. Además los métodos AIC_{HKB} , AIC_{LW} son muy competitivos respecto a $LASSO$ en la selección de variables, llegando a ser superiores que *ElasticNet* incluso, pero $LASSO$ se ve afectado en menor medida cuando aumenta la varianza

Magister en Matemática Mención Estadística

Cuadro 7.34: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (0,136, -0,265, -0,255, -0,37774128, -0,791, 0,182, 0,212, 0,0, 0,0, 0,0, 0,0)$								
	Colinealidad = Baja; P=15; n=50; S=1000								
SR	3364.8			336.48			33.65		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00001	0.17277	0.89352	0.00053	0.32969	0.61155	0.00091	1.09791	0.31882
AIC_{LW}	0.00001	0.17669	0.89667	0.00054	0.44182	0.59451	0.00495	1.92953	0.32247
AIC_{DK}	0.00099	0.24999	0.22326	0.00085	0.33245	0.24178	0.00099	1.27072	0.32574
C_{HKB}	0.00115	0.01255	0.15920	0.00113	0.09858	0.24212	0.00091	1.09808	0.31338
C_{LW}	0.00148	0.07797	0.00000	0.00143	0.16067	0.00114	0.00234	1.11075	0.02609
C_{DK}	0.00099	0.25913	0.18567	0.00101	0.33723	0.22574	0.00099	1.27448	0.31953
MCp_{HKB}	0.00145	0.08349	0.02286	0.00128	0.15823	0.06857	0.00161	1.11557	0.08571
MCp_{LW}	0.00148	0.07797	0.00000	0.00143	0.16199	0.00057	0.00234	1.11958	0.00886
MCp_{DK}	0.00143	0.05118	0.08939	0.00147	0.12675	0.06979	0.00152	1.09838	0.11483
IC_{95}	0.00004	0.16599	0.59883	0.00056	0.35044	0.59308	0.00666	1.83125	0.40854
IC_{99}	0.00004	0.16595	0.60627	0.00054	0.34994	0.59688	0.00660	1.75945	0.33166
$LASSO$	0.00022	0.16367	0.73121	0.00027	0.41061	0.63700	0.01283	2.40510	0.56336
EN	0.00004	0.20165	0.65132	0.00062	0.34605	0.58101	0.00105	2.41458	0.34588
R_{HKB}	0.00004	0.15744	-	0.00024	0.25067	-	0.00422	1.50090	-
R_{LW}	0.00005	0.16606	-	0.00050	0.35045	-	0.00918	2.00252	-
R_{DK}	0.00039	0.06863	-	0.00088	0.15622	-	0.00077	1.13058	-
MCO	0.00004	0.16637	-	0.00056	0.36191	-	0.01479	2.43904	-

Magister en Matemática Mención Estadística

Cuadro 7.35: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,119, 0,144, -0,876, -0,06, 0,054, -0,182, -0,397, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Moderada; P=15; n=50; S=1000									
SR	2876.18			287.62			28.76		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00006	0.07992	0.73864	0.00043	0.17962	0.53548	0.00129	1.08158	0.30064
AIC_{LW}	0.00007	0.08056	0.73967	0.00051	0.19111	0.50253	0.00780	1.36091	0.30454
AIC_{DK}	0.00101	0.24833	0.27975	0.00100	0.35914	0.28705	0.00100	1.34342	0.31894
C_{HKB}	0.00111	0.01904	0.28820	0.00115	0.10215	0.29031	0.00129	1.08383	0.29561
C_{LW}	0.00147	0.04154	0.00143	0.00157	0.12268	0.00800	0.00157	1.09413	0.02492
C_{DK}	0.00101	0.27523	0.20813	0.00100	0.37240	0.25918	0.00100	1.34463	0.30823
MCp_{HKB}	0.00146	0.04387	0.14086	0.00141	0.11875	0.09143	0.00149	1.09163	0.06943
MCp_{LW}	0.00147	0.04199	0.00086	0.00157	0.12177	0.00571	0.00157	1.08279	0.01343
MCp_{DK}	0.00146	0.03433	0.07393	0.00142	0.12063	0.07636	0.00149	1.09062	0.11503
IC_{95}	0.00008	0.08278	0.58916	0.00080	0.20745	0.57104	0.00765	1.32488	0.36357
IC_{99}	0.00008	0.08275	0.59346	0.00080	0.20811	0.56746	0.00765	1.28490	0.26755
$LASSO$	0.00009	0.02934	0.65722	0.00028	0.15602	0.58003	0.00580	1.78580	0.56553
EN	0.00086	0.06394	0.62554	0.00104	0.20636	0.45912	0.00307	1.71317	0.32284
R_{HKB}	0.00008	0.08054	-	0.00053	0.17878	-	0.00466	1.26826	-
R_{LW}	0.00008	0.08278	-	0.00085	0.20915	-	0.00840	1.44166	-
R_{DK}	0.00035	0.05704	-	0.00086	0.13270	-	0.00074	1.13700	-
MCO	0.00008	0.08287	-	0.00089	0.21238	-	0.01097	1.57051	-

Magister en Matemática Mención Estadística

Cuadro 7.36: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,148, 0,245, -0,018, 0,012, -0,789, -0,206, -0,501, 0, 0, 0, 0, 0, 0, 0)$									
<i>Colinealidad = Alta; P=15; n=50; S=1000</i>									
SR	2897.37			289.74			28.97		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00005	0.01952	0.65101	0.00153	0.14292	0.50669	0.00097	1.36493	0.24633
AIC_{LW}	0.00005	0.01954	0.65030	0.00213	0.14906	0.46404	0.00432	1.40859	0.26384
AIC_{DK}	0.00080	0.33984	0.57880	0.00100	0.66898	0.43775	0.00100	1.74814	0.30241
C_{HKB}	0.00120	0.36899	0.02341	0.00124	0.31062	0.19435	0.00097	1.36598	0.23494
C_{LW}	0.00152	0.38425	0.03057	0.00149	0.40535	0.10200	0.00162	1.41613	0.10752
C_{DK}	0.00099	0.69958	0.21171	0.00100	0.75152	0.24008	0.00100	1.77421	0.28495
MC_{PHKB}	0.00150	0.40512	0.00086	0.00148	0.47049	0.04000	0.00156	1.46028	0.09649
MC_{PLW}	0.00152	0.38111	0.03457	0.00149	0.39323	0.11200	0.00162	1.37406	0.10886
MC_{PDK}	0.00152	0.39530	0.00111	0.00147	0.45513	0.03812	0.00149	1.43193	0.11981
IC_{95}	0.00012	0.02678	0.58763	0.00197	0.18544	0.54426	0.00417	1.36403	0.31898
IC_{99}	0.00013	0.02655	0.59081	0.00180	0.17551	0.53034	0.00151	1.37010	0.20029
$LASSO$	0.00016	0.72087	0.61798	0.00046	0.32026	0.53957	0.00532	2.40726	0.56040
EN	0.00072	0.13883	0.55570	0.00109	0.47958	0.43425	0.00129	1.81190	0.30594
R_{HKB}	0.00013	0.02492	-	0.00106	0.15175	-	0.00293	1.39507	-
R_{LW}	0.00012	0.02689	-	0.00207	0.19898	-	0.00606	1.49876	-
R_{DK}	0.00041	0.09204	-	0.00096	0.26562	-	0.00098	1.29263	-
MCO	0.00012	0.02717	-	0.00245	0.22827	-	0.01026	2.26487	-

7.2.1.2.3 100 Observaciones.

Como se aprecia en los cuadros 7.37,7.38, 7.39, es muy difícil que un método de selección capture el modelo correcto, el mejor resultado en este sentido lo alcanzó AIC_{HKB} , AIC_{LW} seleccionando el modelo correcto en un 2.7% de las simulaciones cuando la varianza y la colinealidad son bajas, también se aprecia en los mismos cuadros que medida que aumenta la varianza los resultados van empeorando, alejándose del modelo correcto y acercándose a modelos compuestos por variables irrelevantes.

Cuadro 7.37: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (-0,047, -0,306, -0,316, -0,831, 0,214, 0,005, 0,261, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Baja; P=15; n=100; S=1000														
SR	6451.03					645.10					64.51					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	2.7	11.7	21.1	64.5	0.0	0.9	46.4	4.1	48.6	0.0	0	0.2	0.1	94.6	5.1	
AIC_{LW}	2.7	11.3	21.9	64.1	0.0	0.9	3.6	21.9	73.6	0.0	0	0.0	21.0	73.4	5.6	
AIC_{DK}	0.0	0.2	11.1	57.1	31.6	0.0	0.0	6.3	66.0	27.7	0	0.1	8.3	69.5	22.1	
C_{HKB}	0.0	0.0	0.0	99.2	0.8	0.0	0.0	0.0	97.8	2.2	0.0	0.2	0.0	94.4	5.4	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.7	0.0	97.3	0.0	0.0	17.0	0.0	83.0	
C_{DK}	0.0	0.0	4.8	53.5	41.7	0.0	0.0	4.1	61.2	34.7	0.0	0.1	8.5	68.6	22.8	
MC_{pHKB}	0.0	0.0	0.8	0.0	99.2	0.0	0.0	24.1	0.0	75.9	0.0	0.0	39.6	0.0	60.4	
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.0	0.0	98.0	0.0	0.0	12.2	0.0	87.8	
MC_{pDK}	0.0	0.0	28.6	0.0	71.4	0.0	0.0	22.5	0.0	77.5	0.0	0.0	32.5	7.0	60.5	
IC_{95}	0.0	94.8	0.0	5.2	0.0	0.0	74.4	0.0	25.6	0.0	0.0	4.9	0.6	94.3	0.2	
IC_{99}	0.0	92.7	0.0	7.3	0.0	0.0	67.0	0.0	33.0	0.0	0.0	1.1	3.3	94.9	0.6	
$LASSO$	0.3	0.9	46.8	52.0	0.0	0.0	35.5	0.1	64.4	0.0	0.0	79.0	0.0	21.0	0.0	
EN	0.0	48.7	0.1	51.2	0.0	0.0	23.8	28.3	47.2	0.0	0.0	15.0	22.0	46.0	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.38: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,388, -0,349, -0,539, -0,531, 0,047, -0,369, 0,131, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Moderada; P=15; n=100; S=1000															
SR	4419.86					441.99					44.20				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	9.4	22.7	16.9	51.0	0.0	0.3	21.5	6.9	71.3	0.0	0	0.6	0.7	81.4	17.3
AIC_{LW}	9.8	22.2	17.3	50.7	0.0	0.5	1.6	24.3	73.6	0.0	0	0.0	22.0	68.9	9.1
AIC_{DK}	0.8	1.7	26.3	71.2	0.0	0.0	0.1	19.8	72.0	8.1	0	0.3	6.2	69.7	23.8
C_{HKB}	0.0	0.0	0.3	88.4	11.3	0.0	0.0	0.0	97.2	2.8	0.0	0.0	0.7	81.2	18.1
C_{LW}	0.0	0.0	4.9	0.0	95.1	0.0	0.0	9.5	0.0	90.5	0.0	0.0	15.7	0.0	84.3
C_{DK}	0.0	0.0	0.3	38.9	60.8	0.0	0.0	1.7	57.8	40.5	0.0	0.1	4.9	65.1	29.9
MC_{pHKB}	0.0	0.0	19.0	0.0	81.0	0.0	0.0	27.5	0.0	72.5	0.0	0.0	23.7	0.0	76.3
MC_{pLW}	0.0	0.0	4.9	0.0	95.1	0.0	0.0	8.6	0.0	91.4	0.0	0.0	11.4	0.0	88.6
MC_{pDK}	0.0	0.0	6.9	0.0	93.1	0.0	0.0	19.6	0.0	80.4	0.0	0.0	21.7	9.8	68.5
IC_{95}	0.0	95.2	0.0	4.8	0.0	0.0	64.1	0.0	35.9	0.0	0.0	2.6	1.0	94.8	1.6
IC_{99}	0.0	93.9	0.0	6.1	0.0	0.0	56.2	0.0	43.8	0.0	0.0	0.4	6.8	87.1	5.3
$LASSO$	0.0	0.0	42.5	57.5	0.0	0.0	20.4	0.1	79.5	0.0	0.0	69.0	0.0	31.0	0.0
EN	0.8	62.0	7.8	29.4	0.0	0.1	31.6	25.0	43.1	0.0	0.0	12.4	20.6	55.5	0.4

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.39: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (0,497, 0,174, -0,452, -0,099, 0,071, 0,708, -0,053, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Alta; P=15; n=100; S=1000															
SR	4298.29					429.83					42.98				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	1.6	2.8	20.3	75.3	0.0	0.0	0.7	13.8	85.5	0.0	0	0.2	4.3	87.2	8.3
AIC_{LW}	1.7	2.4	22.3	73.6	0.0	0.0	0.1	27.0	72.6	0.3	0	0.0	21.9	58.7	19.4
AIC_{DK}	0.5	2.2	20.8	62.8	13.7	0.1	0.6	34.1	36.7	28.5	0	0.0	21.2	53.1	25.7
C_{HKB}	0.0	0.0	2.5	0.0	97.5	0.0	0.1	8.9	85.4	5.6	0.0	0.0	4.7	86.6	8.7
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.3	0.0	98.7
C_{DK}	0.0	0.0	15.8	21.7	62.5	0.1	0.1	46.4	19.9	33.5	0.0	0.1	23.1	50.9	25.9
MC_{pHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	7.3	0.0	92.7	0.0	0.0	17.0	0.0	83.0
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.6	0.0	99.4
MC_{pDK}	0.0	0.0	24.5	0.0	75.5	0.0	0.0	43.8	0.0	56.2	0.0	0.0	35.7	6.2	58.1
IC_{95}	0.0	82.6	0.0	17.4	0.0	0.0	21.3	0.0	78.7	0.0	0.0	0.0	6.9	82.0	9.6
IC_{99}	0.0	77.2	0.0	22.8	0.0	0.0	11.9	0.2	87.9	0.0	0.0	0.0	18.5	53.0	18.7
$LASSO$	0.0	0.0	96.0	4.0	0.0	0.0	18.4	0.2	81.4	0.0	0.0	79.2	0.0	20.8	0.0
EN	0.0	42.1	24.6	33.3	0.0	0.0	17.1	44.0	32.0	0.0	0.0	12.8	26.0	35.4	0.1

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.40,7.41,7.42, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados empeoran en la medida que aumenta el nivel de la colinealidad. Se puede observar que AIC_{HKB} Y AIC_{LW} con cualquiera de los estimadores del parámetro de sesgo es muy competitivo respecto de LASSO para varianzas moderadas y altas en cuanto a selección de variables se trata.

Cuadro 7.40: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (-0,047, -0,306, -0,316, -0,831, 0,214, 0,005, 0,261, 0, 0, 0, 0, 0, 0, 0, 0)$								
	Colinealidad = Baja; P=15; n=100; S=1000								
SR	6451.03			645.1			64.51		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	0.39072	0.82790	0.00032	0.45976	0.65468	0.00120	1.20163	0.34243
AIC_{LW}	0.00003	0.39467	0.82922	0.00020	0.51525	0.72389	0.00188	1.58516	0.38614
AIC_{DK}	0.00046	0.83088	0.28727	0.00099	0.97877	0.24251	0.00103	2.03001	0.27364
C_{HKB}	0.00112	0.21518	0.21102	0.00103	0.26997	0.26782	0.00120	1.20137	0.34142
C_{LW}	0.00133	0.38795	0.00000	0.00142	0.41449	0.00771	0.00141	1.40266	0.04857
C_{DK}	0.00100	0.96715	0.15811	0.00099	1.05471	0.18767	0.00103	2.03420	0.27209
$MC_{p_{HKB}}$	0.00127	0.20037	0.00229	0.00127	0.38418	0.06886	0.00145	1.38829	0.11314
$MC_{p_{LW}}$	0.00133	0.38606	0.00000	0.00142	0.40473	0.00571	0.00141	1.37955	0.03486
$MC_{p_{DK}}$	0.00126	0.36639	0.08171	0.00126	0.45164	0.06429	0.00146	1.38454	0.11332
IC_{95}	0.00004	0.39727	0.58450	0.00031	0.52160	0.60375	0.00272	1.70409	0.51804
IC_{99}	0.00004	0.39729	0.58803	0.00029	0.52278	0.61570	0.00265	1.67092	0.48319
LASSO	0.00009	0.20536	0.82647	0.00019	0.39712	0.66431	0.00254	1.78740	0.56860
EN	0.00022	0.42244	0.63364	0.00025	0.53723	0.56881	0.00357	1.89012	0.39247
R_{HKB}	0.00003	0.38799	-	0.00032	0.44870	-	0.00106	1.38541	-
R_{LW}	0.00003	0.39723	-	0.00031	0.52142	-	0.00273	1.72610	-
R_{DK}	0.00003	0.27268	-	0.00091	0.34431	-	0.00094	1.38527	-
MCO	0.00004	0.39741	-	0.00031	0.52352	-	0.00297	1.76994	-

Magister en Matemática Mención Estadística

Cuadro 7.41: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (0,388, -0,349, -0,539, -0,531, 0,047, -0,369, 0,131, 0, 0, 0, 0, 0, 0, 0, 0)$								
	Colinealidad = Moderada; P=15; n=100; S=1000								
SR	4419.86			441.99			44.2		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00004	0.39096	0.84296	0.00027	0.44724	0.64998	0.00126	1.01693	0.27402
AIC_{LW}	0.00004	0.39447	0.84494	0.00017	0.51478	0.71502	0.00118	1.38167	0.35370
AIC_{DK}	0.00093	0.10586	0.62648	0.00105	0.21380	0.45742	0.00100	1.15876	0.27133
C_{HKB}	0.00125	0.06971	0.19871	0.00118	0.16506	0.24034	0.00126	1.01412	0.26636
C_{LW}	0.00170	0.01885	0.01400	0.00162	0.11087	0.02714	0.00169	0.99238	0.04486
C_{DK}	0.00118	0.08361	0.09802	0.00105	0.22141	0.16228	0.00100	1.17152	0.22753
MCp_{HKB}	0.00168	0.01842	0.05429	0.00169	0.11361	0.07857	0.00168	0.99751	0.06771
MCp_{LW}	0.00170	0.01867	0.01400	0.00162	0.11384	0.02457	0.00169	1.02160	0.03257
MCp_{DK}	0.00170	0.02056	0.01971	0.00166	0.11726	0.05600	0.00167	0.99995	0.08851
IC_{95}	0.00005	0.39163	0.58692	0.00026	0.51742	0.60384	0.00158	1.51010	0.47229
IC_{99}	0.00005	0.39165	0.59253	0.00026	0.51601	0.61647	0.00073	1.40797	0.40771
$LASSO$	0.00013	0.12676	0.74684	0.00026	0.35232	0.65056	0.00158	1.66423	0.56443
EN	0.00003	0.40590	0.69528	0.00027	0.39062	0.63979	0.00039	1.89331	0.42515
R_{HKB}	0.00004	0.38053	-	0.00027	0.43155	-	0.00057	1.28151	-
R_{LW}	0.00005	0.39163	-	0.00026	0.51696	-	0.00176	1.63431	-
R_{DK}	0.00005	0.32806	-	0.00096	0.24033	-	0.00076	1.03195	-
MCO	0.00005	0.39182	-	0.00026	0.51930	-	0.00194	1.69172	-

Magister en Matemática Mención Estadística

Cuadro 7.42: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,497, 0,174, -0,452, -0,099, 0,071, 0,708, -0,053, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Alta; P=15; n=100; S=1000									
SR	4298.29			429.83			42.98		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00005	0.01581	0.67446	0.00027	0.11393	0.48822	0.00094	1.09339	0.32768
AIC_{LW}	0.00005	0.01695	0.67437	0.00041	0.11795	0.45698	0.00226	1.16546	0.29081
AIC_{DK}	0.00100	0.17512	0.53074	0.00084	0.45440	0.26762	0.00095	1.58204	0.25297
C_{HKB}	0.00082	0.07654	0.00714	0.00081	0.14508	0.36756	0.00094	1.09402	0.32367
C_{LW}	0.00079	0.08547	0.00000	0.00079	0.15593	0.00000	0.00083	1.08489	0.00371
C_{DK}	0.00100	0.21782	0.12313	0.00084	0.40379	0.20831	0.00095	1.57139	0.25176
MC_{PHKB}	0.00082	0.05137	0.00000	0.00084	0.14601	0.02086	0.00083	1.10560	0.04857
MC_{PLW}	0.00079	0.08499	0.00000	0.00079	0.15455	0.00000	0.00083	1.07291	0.00171
MC_{PDK}	0.00074	0.05683	0.07000	0.00088	0.14964	0.12514	0.00084	1.10947	0.11969
IC_{95}	0.00015	0.01973	0.58843	0.00043	0.13403	0.55699	0.00259	1.17564	0.34902
IC_{99}	0.00015	0.01975	0.59274	0.00043	0.13450	0.55175	0.00240	1.12891	0.25029
$LASSO$	0.00018	0.34801	0.44088	0.00020	0.22669	0.54788	0.00330	1.60142	0.56397
EN	0.00006	0.12526	0.62878	0.00075	0.33171	0.42890	0.00100	1.68748	0.32671
R_{HKB}	0.00012	0.01719	-	0.00020	0.11362	-	0.00121	1.13603	-
R_{LW}	0.00015	0.01972	-	0.00041	0.13434	-	0.00342	1.28318	-
R_{DK}	0.00041	0.03236	-	0.00090	0.16648	-	0.00098	1.14542	-
MCO	0.00015	0.01978	-	0.00044	0.13585	-	0.00419	1.35439	-

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.46, 7.47,7.48, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados aunque numéricamente diferentes mantienen cierta estabilidad, lo que indicaría que el nivel de la colinealidad no sería determinante en este caso.

Cuadro 7.46: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (-0,051, 0,187, -0,804, 0,236, -0,498, 0,046, -0,101, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Baja; P=20; n=25; S=1000									
SR	1468.27			146.83			14.68		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	0.02291	0.64116	0.00104	0.11413	0.24868	0.00106	1.11719	0.28134
AIC_{LW}	0.00011	0.03100	0.65147	0.00140	0.19435	0.43273	0.00306	1.61417	0.29048
AIC_{DK}	0.00100	0.43042	0.34887	0.00100	0.52689	0.37836	0.00099	1.39177	0.41916
C_{HKB}	0.00108	0.01412	0.16095	0.00104	0.11271	0.22490	0.00106	1.12318	0.27177
C_{LW}	0.00148	0.05181	0.07029	0.00149	0.15271	0.12371	0.00166	1.12798	0.25441
C_{DK}	0.00100	0.43826	0.33450	0.00099	0.53590	0.36688	0.00099	1.40214	0.41243
MC_{PHKB}	0.00143	0.06041	0.23571	0.00118	0.14433	0.16114	0.00133	1.13663	0.15396
MC_{PLW}	0.00148	0.05244	0.06286	0.00149	0.15568	0.11743	0.00168	1.19007	0.13371
MC_{PDK}	0.00133	0.03265	0.11193	0.00116	0.12784	0.10664	0.00123	1.17166	0.13985
IC_{95}	0.00126	0.10177	0.48670	0.00166	0.38360	0.45050	0.00112	1.75498	0.34856
IC_{99}	0.00126	0.09868	0.49520	0.00156	0.35921	0.44922	0.00089	1.58840	0.28790
$LASSO$	0.00015	0.04541	0.55056	0.00165	0.62290	0.46622	0.06875	10.70754	0.45557
EN	0.00017	0.06490	0.52971	0.00100	0.26078	0.38739	0.00102	1.80888	0.23836
R_{HKB}	0.00059	0.04542	-	0.00241	0.31597	-	0.00860	3.01569	-
R_{LW}	0.00109	0.09874	-	0.00191	0.37702	-	0.00569	2.02079	-
R_{DK}	0.00088	0.04649	-	0.00089	0.17318	-	0.00092	1.13881	-
MCO	0.00145	0.13343	-	0.02235	1.16176	-	0.07630	11.42704	-

Magister en Matemática Mención Estadística

Cuadro 7.47: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (-0,344, -0,313, -0,288, 0,036, -0,675, 0,369, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$								
	Colinealidad = Moderada; P=20; n=25; S=1000								
SR	1164.71			116.47			11.65		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00025	0.02008	0.68699	0.00106	0.29598	0.23637	0.00090	1.40007	0.25086
AIC_{LW}	0.00029	0.02355	0.66561	0.00184	0.18234	0.39417	0.00737	1.49776	0.29849
AIC_{DK}	0.00100	0.90611	0.32512	0.00100	0.99658	0.36330	0.00100	1.85467	0.39104
C_{HKB}	0.00100	0.15005	0.24067	0.00106	0.30646	0.22005	0.00090	1.41702	0.24321
C_{LW}	0.00107	0.12623	0.17171	0.00131	0.24145	0.12686	0.00100	1.15509	0.26710
C_{DK}	0.00100	0.91984	0.31073	0.00100	1.01118	0.34437	0.00100	1.86866	0.37558
MCp_{HKB}	0.00101	0.07721	0.04857	0.00103	0.22713	0.07171	0.00093	1.41913	0.12975
MCp_{LW}	0.00107	0.06123	0.05486	0.00131	0.18811	0.06171	0.00101	1.15385	0.08349
MCp_{DK}	0.00101	0.16592	0.06743	0.00082	0.29906	0.06743	0.00092	1.52013	0.12662
IC_{95}	0.00067	0.05575	0.49832	0.00380	0.27699	0.46100	0.01024	1.50557	0.34713
IC_{99}	0.00067	0.05457	0.51177	0.00396	0.25955	0.44951	0.00880	1.40553	0.28322
$LASSO$	0.00045	0.16666	0.53623	0.00272	0.55420	0.46626	0.03230	6.09260	0.45893
EN	0.00082	0.17542	0.58976	0.00080	0.51285	0.40810	0.00092	1.68877	0.22575
R_{HKB}	0.00035	0.04097	-	0.00177	0.25974	-	0.01703	2.52234	-
R_{LW}	0.00069	0.05532	-	0.00317	0.28468	-	0.01248	1.81268	-
R_{DK}	0.00094	0.11589	-	0.00083	0.34230	-	0.00099	1.35331	-
MCO	0.00157	0.06521	-	0.00602	0.60232	-	0.04879	6.06215	-

Magister en Matemática Mención Estadística

Las tablas expuestas en los cuadros 7.52, 7.53 y 7.54, muestran una notable mejoría en todos los indicadores respecto de las tablas en los cuadros 7.46, 7.47 y 7.48, debido al aumento de las observaciones. Pero a diferencia de cuando se trabaja con menos variables, el nivel de la colinealidad no tiene el efecto deteriorante de las primeras tablas. Ahora bien la varianza sigue siendo el factor que más perjudica a la selección de variables y estimación de coeficientes.

Cuadro 7.52: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

		$\beta = (0,026, 0,407, -0,223, -0,37, -0,387, -0,651, 0,272, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$								
		Colinealidad = Baja; P=20; n=50; S=1000								
SR	3530.08			353.01			35.30			
σ^2	0.01			0.1			1			
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F	
AIC_{HKB}	0.00001	0.70354	0.87697	0.00055	0.65772	0.53732	0.00103	1.32614	0.26426	
AIC_{LW}	0.00001	0.72582	0.88370	0.00063	0.79269	0.56572	0.00414	1.96669	0.29917	
AIC_{DK}	0.00099	0.07885	0.20652	0.00100	0.16283	0.22252	0.00098	1.11686	0.29865	
C_{HKB}	0.00102	0.30529	0.08838	0.00096	0.37698	0.17127	0.00103	1.32484	0.26259	
C_{LW}	0.00141	0.24809	0.02800	0.00134	0.31786	0.04771	0.00172	1.37041	0.09000	
C_{DK}	0.00100	0.08074	0.16750	0.00100	0.16211	0.21338	0.00098	1.11792	0.29324	
MCp_{HKB}	0.00133	0.27455	0.00257	0.00132	0.36398	0.03600	0.00149	1.34735	0.09086	
MCp_{LW}	0.00141	0.24967	0.02429	0.00134	0.32046	0.04286	0.00172	1.39507	0.07429	
MCp_{DK}	0.00110	0.25797	0.07343	0.00129	0.31130	0.07300	0.00109	1.30719	0.09146	
IC_{95}	0.00007	0.73164	0.49293	0.00077	0.88336	0.51958	0.01001	1.97353	0.35388	
IC_{99}	0.00007	0.73137	0.50426	0.00083	0.87453	0.53980	0.00700	1.78187	0.28236	
$LASSO$	0.00036	0.19348	0.73237	0.00046	0.55819	0.54900	0.01836	2.88945	0.45970	
EN	0.00006	0.51808	0.56588	0.00104	0.47693	0.49455	0.02113	2.30563	0.29175	
R_{HKB}	0.00005	0.67728	-	0.00042	0.69068	-	0.00600	1.87588	-	
R_{LW}	0.00006	0.73156	-	0.00092	0.88394	-	0.01257	2.23191	-	
R_{DK}	0.00065	0.36702	-	0.00086	0.33285	-	0.00097	1.33744	-	
MCO	0.00007	0.73390	-	0.00110	0.91428	-	0.02153	2.67125	-	

Magister en Matemática Mención Estadística

Cuadro 7.53: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,017, -0,167, 0,328, -0,003, -0,642, 0,147, -0,656, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$									
Colinealidad = Moderada; P=20; n=50; S=1000									
SR	1784.91			178.49			17.85		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00010	0.22474	0.73758	0.00054	0.23074	0.51311	0.00112	1.08964	0.22345
AIC_{LW}	0.00010	0.25096	0.74017	0.00078	0.32444	0.51599	0.01060	1.54903	0.27549
AIC_{DK}	0.00102	0.14702	0.27232	0.00100	0.24485	0.25020	0.00100	1.21002	0.28712
C_{HKB}	0.00123	0.14374	0.26814	0.00114	0.19448	0.21746	0.00112	1.09257	0.21923
C_{LW}	0.00145	0.27100	0.27171	0.00150	0.31870	0.16600	0.00152	1.24016	0.09683
C_{DK}	0.00102	0.15952	0.21105	0.00100	0.24863	0.22152	0.00100	1.21288	0.27906
MCp_{HKB}	0.00133	0.24925	0.19000	0.00133	0.26791	0.08314	0.00137	1.15244	0.08333
MCp_{LW}	0.00145	0.27071	0.27000	0.00150	0.31828	0.15714	0.00144	1.24174	0.08743
MCp_{DK}	0.00132	0.17033	0.08904	0.00131	0.22876	0.07162	0.00136	1.13377	0.08570
IC_{95}	0.00014	0.25315	0.48491	0.00122	0.37276	0.48536	0.00725	1.44604	0.31541
IC_{99}	0.00014	0.25307	0.49357	0.00172	0.36668	0.49033	0.00781	1.33481	0.23946
$LASSO$	0.00022	0.14744	0.62253	0.00053	0.24857	0.51022	0.00577	2.21746	0.45924
EN	0.00006	0.14787	0.55056	0.00066	0.23478	0.43537	0.00736	1.61579	0.27870
R_{HKB}	0.00011	0.19898	-	0.00054	0.21859	-	0.00491	1.36132	-
R_{LW}	0.00014	0.25297	-	0.00140	0.37875	-	0.00956	1.62812	-
R_{DK}	0.00093	0.03118	-	0.00094	0.12597	-	0.00088	1.09534	-
MCO	0.00015	0.25561	-	0.00172	0.40817	-	0.01326	2.00285	-

Magister en Matemática Mención Estadística

Cuadro 7.54: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (-0,018, -0,0325, -0,733, -0,423, -0,383, 0,109, 0,3512, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$									
Colinealidad = Alta; P=20; n=50; S=1000									
SR	1189.87			118.99			11.90		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00025	0.27660	0.71720	0.00161	0.50742	0.42562	0.00103	1.31629	0.18639
AIC_{LW}	0.00014	0.29218	0.72521	0.00249	0.64326	0.38419	0.00750	2.15566	0.20787
AIC_{DK}	0.00100	0.06462	0.53198	0.00099	0.14071	0.36213	0.00100	1.08792	0.24996
C_{HKB}	0.00189	0.77833	0.28179	0.00125	0.50212	0.23674	0.00103	1.29564	0.17786
C_{LW}	0.00200	0.86268	0.28200	0.00170	0.53326	0.10143	0.00167	1.34987	0.05016
C_{DK}	0.00101	0.60945	0.29408	0.00100	0.28286	0.28308	0.00098	1.10531	0.23833
MC_{PHKB}	0.00189	0.78375	0.28400	0.00160	0.59596	0.16657	0.00134	1.30652	0.06801
MC_{PLW}	0.00200	0.85756	0.28000	0.00170	0.47922	0.08057	0.00297	1.31455	0.02429
MC_{PDK}	0.00199	0.86004	0.28514	0.00166	0.65393	0.17182	0.00141	1.31764	0.09535
IC_{95}	0.00061	0.29308	0.48821	0.00253	0.53338	0.45452	0.00611	1.84604	0.26395
IC_{99}	0.00061	0.29260	0.49952	0.00249	0.54733	0.44446	0.00455	1.64609	0.16734
$LASSO$	0.00040	0.06329	0.54029	0.00055	0.40910	0.46628	0.01160	4.07357	0.45395
EN	0.00067	0.16671	0.50315	0.00108	0.28708	0.39814	0.00105	1.63319	0.26155
R_{HKB}	0.00040	0.25030	-	0.00134	0.39858	-	0.00514	2.05337	-
R_{LW}	0.00059	0.29240	-	0.00253	0.51429	-	0.00771	2.09909	-
R_{DK}	0.00090	0.25125	-	0.00060	0.31397	-	0.00100	1.23722	-
MCO	0.00063	0.29760	-	0.00317	0.60608	-	0.02222	3.84576	-

Magister en Matemática Mención Estadística

Las tablas expuestas en los cuadros 7.58, 7.59 y 7.60, muestran una notable mejoría en todos los indicadores respecto de las tablas en los cuadros 7.46, 7.47 y 7.48, debido al aumento de las observaciones. Pero a diferencia de cuando se trabaja con menos variables, el nivel de la colinealidad no tiene el efecto deteriorante de las primeras tablas. Ahora bien la varianza sigue siendo el factor que más perjudica a la selección de variables y estimación de coeficientes. Es de notar que cuando la colinealidad es alta pero la varianza es baja, los Métodos basados en Mallows se ven severamente perjudicados llegando incluso a marcar cero con el indicador F.

Cuadro 7.58: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,433, 0,807, -0,217, -0,096, 0,14, 0,106, -0,272, 0,0, 0,0, 0,0, 0,0, 0,0, 0,0, 0,0, 0,0, 0,0)$									
Colinealidad = Baja; P=20; n=100; S=1000									
SR	5653.98			565.4			56.54		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00004	0.01171	0.82572	0.00039	0.11016	0.50584	0.00064	1.00743	0.27909
AIC_{LW}	0.00004	0.01179	0.82905	0.00026	0.11162	0.56444	0.00258	1.12097	0.33591
AIC_{DK}	0.00099	0.11638	0.54603	0.00097	0.26101	0.29549	0.00100	1.18589	0.23605
C_{HKB}	0.00083	0.01254	0.04486	0.00098	0.10828	0.27043	0.00064	1.00657	0.27605
C_{LW}	0.00088	0.01344	0.00286	0.00095	0.10139	0.02057	0.00096	0.99773	0.06429
C_{DK}	0.00088	0.21679	0.22156	0.00097	0.29608	0.20973	0.00100	1.18510	0.23191
MCp_{HKB}	0.00088	0.01527	0.16200	0.00095	0.10121	0.11543	0.00036	0.99934	0.10914
MCp_{LW}	0.00088	0.01341	0.00171	0.00095	0.10147	0.01914	0.00096	0.99884	0.05200
MCp_{DK}	0.00083	0.01148	0.11943	0.00093	0.10187	0.08457	0.00090	0.99857	0.08935
IC_{95}	0.00004	0.01249	0.47873	0.00053	0.11919	0.48449	0.00551	1.16937	0.41920
IC_{99}	0.00004	0.01250	0.48421	0.00053	0.11930	0.49216	0.00551	1.14252	0.39233
$LASSO$	0.00010	0.03104	0.79193	0.00022	0.12762	0.50909	0.00657	1.19758	0.46050
EN	0.00001	0.01758	0.61117	0.00084	0.14112	0.45166	0.00056	1.43600	0.30757
R_{HKB}	0.00023	0.01224	-	0.00039	0.10957	-	0.00257	1.05244	-
R_{DK}	0.00023	0.01249	-	0.00051	0.11925	-	0.00593	1.18354	-
R_{LW}	0.00029	0.01961	-	0.00090	0.12652	-	0.00095	1.02529	-
MCO	0.00004	0.01249	-	0.00052	0.11964	-	0.00687	1.21049	-

Magister en Matemática Mención Estadística

Cuadro 7.59: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,045, -0,092, 0,329, -0,551, -0,743, 0,095, -0,129, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$									
<i>Colinealidad = Moderada; P=20; n=100; S=1000</i>									
SR	5076.1			507.61			50.76		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00005	0.01191	0.79399	0.00039	0.12231	0.52518	0.00118	1.11490	0.23682
AIC_{LW}	0.00005	0.01186	0.79752	0.00022	0.11949	0.57415	0.00065	1.13003	0.31862
AIC_{DK}	0.00078	0.52897	0.38024	0.00100	0.79067	0.23435	0.00100	1.79960	0.23282
C_{HKB}	0.00111	0.10603	0.25178	0.00115	0.18482	0.22405	0.00118	1.11699	0.23073
C_{LW}	0.00146	0.05997	0.26229	0.00146	0.16122	0.16057	0.00151	1.05985	0.10543
C_{DK}	0.00098	0.73244	0.22684	0.00100	0.84582	0.17937	0.00100	1.80405	0.22200
$MC_{p_{HKB}}$	0.00146	0.05903	0.28343	0.00144	0.15408	0.17000	0.00148	1.06622	0.10057
$MC_{p_{LW}}$	0.00146	0.06028	0.25914	0.00146	0.16198	0.15486	0.00151	1.07134	0.09286
$MC_{p_{DK}}$	0.00144	0.05340	0.16229	0.00137	0.15196	0.08568	0.00141	1.08745	0.08039
IC_{95}	0.00007	0.01246	0.48093	0.00055	0.11942	0.48485	0.00089	1.17238	0.38165
IC_{99}	0.00007	0.01247	0.48669	0.00054	0.11949	0.49276	0.00041	1.15824	0.33411
$LASSO$	0.00013	0.25182	0.73904	0.00036	0.20253	0.51798	0.00182	1.36322	0.45736
EN	0.00007	0.09273	0.59257	0.00053	0.31404	0.49360	0.00100	1.70071	0.33878
R_{HKB}	0.00006	0.01294	-	0.00038	0.11988	-	0.00059	1.07285	-
R_{LW}	0.00007	0.01247	-	0.00048	0.11970	-	0.00179	1.18858	-
R_{DK}	0.00079	0.08626	-	0.00074	0.23273	-	0.00094	1.21771	-
MCO	0.00007	0.01247	-	0.00049	0.11998	-	0.00203	1.22016	-

Magister en Matemática Mención Estadística

Cuadro 7.60: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (0,083, -0,095, -0,65, 0,29, -0,559, 0,167, -0,369, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$									
<i>Colinealidad = Alta; P=20; n=100; S=1000</i>									
SR	4216.54			421.65			42.17		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00013	0.53685	0.76817	0.00036	0.59932	0.51502	0.00103	1.29591	0.16656
AIC_{LW}	0.00014	0.54002	0.77125	0.00024	0.65613	0.54221	0.00241	1.99282	0.23870
AIC_{DK}	0.00087	0.13924	0.57898	0.00100	0.15533	0.43939	0.00237	1.07597	0.24689
C_{HKB}	0.00154	0.50682	0.00000	0.00118	0.38297	0.15782	0.00103	1.26094	0.13877
C_{LW}	0.00156	0.55820	0.00000	0.00173	0.42251	0.02257	0.00156	1.31424	0.04629
C_{DK}	0.00156	0.11016	0.14426	0.00154	0.21972	0.14212	0.00100	1.07640	0.18851
MC_{PHKB}	0.00154	0.51699	0.00000	0.00150	0.43354	0.00457	0.00146	1.24575	0.03463
MC_{PLW}	0.00156	0.55729	0.00000	0.00173	0.41677	0.02286	0.00156	1.36568	0.04371
MC_{PDK}	0.00156	0.54200	0.00000	0.00154	0.44392	0.00971	0.00153	1.26223	0.05261
IC_{95}	0.00017	0.58299	0.48153	0.00065	0.74446	0.48084	0.00303	1.88069	0.27323
IC_{99}	0.00017	0.58258	0.48936	0.00060	0.72896	0.48684	0.00225	1.68332	0.17664
$LASSO$	0.00022	0.13394	0.65454	0.00020	0.50854	0.50603	0.00533	2.60791	0.45329
EN	0.00065	0.33924	0.55922	0.00067	0.36591	0.44084	0.00100	1.67613	0.29503
R_{HKB}	0.00017	0.58627	-	0.00027	0.66003	-	0.00156	1.78690	-
R_{LW}	0.00017	0.58345	-	0.00066	0.75786	-	0.00491	2.28467	-
R_{DK}	0.00100	0.40349	-	0.00090	0.39172	-	0.00095	1.27439	-
MCO	0.00017	0.58320	-	0.00072	0.76339	-	0.00624	2.69704	-

7.2.2. Coeficientes Definidos Arbitrariamente.

7.2.2.1. 10 Variables Disponibles.

7.2.2.1.1 25 Observaciones.

Como se aprecia en los cuadros 7.61, 7.62, 7.63, es muy difícil que un método de selección capture el modelo correcto, el mejor resultado en este sentido lo alcanzo AIC_{LW} seleccionando el modelo correcto en un 71.6% de las simulaciones seguido muy de cerca de AIC_{HKB} con un 71,4% cuando la varianza es baja y la colinealidad es Baja, otro método que logro capturar el modelo correctamente una cantidad de veces significativa fue $LASSO$ con 29.3% de las simulaciones cuando la varianza es moderada y la colinealidad es baja. También se logra ver que el aumento de la varianza es el factor que perjudica más la selección de variables, seguido de la colinealidad pero con un impacto más difícil de distinguir por la limitada de información ($n = 25$).

Cuadro 7.61: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$														
		Colinealidad = Baja; P=10; n=25; S=1000														
SR	73789.2					7378.9					737.9					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	71.4	28.6	0.0	0.0	0	51.7	22.7	15.5	10.1	0.0	0.1	2.5	40.9	56.1	0.4	
AIC_{LW}	71.6	28.4	0.0	0.0	0	47.9	19.4	21.2	11.5	0.0	0.0	0.0	55.8	43.3	0.9	
AIC_{DK}	0.0	0.0	24.3	75.7	0	0.3	1.0	31.5	67.2	0.0	0.2	2.2	26.6	68.2	2.8	
C_{HKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.8	0.0	1.2	0.0	0.0	84.9	2.1	13.0	
C_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.6	0.0	99.4	0.0	0.0	24.0	0.0	76.0	
C_{DK}	0.0	0.0	29.4	70.6	0.0	0.3	0.5	35.9	63.3	0.0	0.1	2.1	29.1	66.0	2.7	
MC_{PHKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.7	0.0	1.3	0.0	0.0	85.4	0.0	14.6	
MC_{PLW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.6	0.0	99.4	0.0	0.0	23.8	0.0	76.2	
MC_{PDK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.0	0.0	1.0	0.0	0.0	81.9	0.5	17.6	
IC_{95}	0.1	99.9	0.0	0.0	0.0	8.7	85.3	1.0	5.0	0.0	0.0	0.1	51.4	29.0	2.4	
IC_{99}	0.2	99.8	0.0	0.0	0.0	16.2	72.9	3.4	7.5	0.0	0.0	0.0	47.6	15.2	1.1	
$LASSO$	29.3	70.7	0.0	0.0	0.0	9.1	89.4	0.2	1.3	0.0	0.4	40.6	0.0	59.0	0.0	
EN	0.0	100.0	0.0	0.0	0.0	0.8	69.1	17.7	12.4	0.0	0.3	9.5	70.0	15.1	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.62: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$														
		Colinealidad = Moderada; P=10; n=25; S=1000														
SR	71004.5					7100.5					710.0					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	0.0	100.0	0.0	0.0	0.0	5.0	28.3	4.9	61.8	0.0	0.0	0.5	35.2	64.0	0.3	
AIC_{LW}	0.0	100.0	0.0	0.0	0.0	4.0	21.3	6.3	68.4	0.0	0.0	0.0	44.8	53.3	1.9	
AIC_{DK}	0.0	0.0	0.0	45.9	54.1	0.0	0.7	4.9	75.1	19.3	0.1	2.5	16.5	73.7	7.2	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	30.2	5.2	64.6	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	8.7	0.0	91.3	
C_{DK}	0.0	0.0	0.0	45.3	54.7	0.0	0.5	5.3	73.1	21.1	0.1	1.3	16.3	74.2	8.1	
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	99.9	0.0	0.0	29.4	0.0	70.6	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	8.3	0.0	91.7	
MC_{PDK}	0.0	0.0	10.4	0.0	89.6	0.0	0.0	37.5	0.0	62.5	0.0	0.0	47.6	0.1	52.3	
IC_{95}	0.4	99.6	0.0	0.0	0.0	9.4	69.6	2.6	18.4	0.0	0.0	0.1	48.0	20.4	3.2	
IC_{99}	0.9	99.1	0.0	0.0	0.0	11.5	53.2	9.7	25.6	0.0	0.0	0.0	39.0	9.5	1.8	
$LASSO$	2.8	97.2	0.0	0.0	0.0	7.8	85.5	0.1	6.6	0.0	0.5	24.6	1.2	73.7	0.0	
EN	0.0	99.6	0.0	0.4	0.0	0.0	53.4	17.8	28.8	0.0	0.0	8.6	74.9	16.0	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.63: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$														
		Colinealidad = Alta; P=10; n=25; S=1000														
SR	71330.0					7133.0					713.3					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	27.0	67.0	0.0	6.0	0.0	5.7	3.4	23.9	67.0	0.0	0.0	0.0	39.1	55.4	5.5	
AIC_{LW}	28.9	64.0	0.0	7.1	0.0	3.6	2.4	28.5	65.5	0.0	0.0	0.0	44.3	44.3	11.4	
AIC_{DK}	2.2	5.7	4.9	87.2	0.0	0.2	1.9	10.1	87.6	0.2	0.2	0.7	20.8	73.8	4.5	
C_{HKB}	0.0	0.0	53.8	0.0	46.2	0.0	0.0	64.2	0.0	35.8	0.0	0.0	71.3	0.7	28.0	
C_{LW}	0.0	0.0	0.1	0.0	99.9	0.0	0.0	17.8	0.0	82.2	0.0	0.0	40.4	0.0	59.6	
C_{DK}	0.0	0.0	69.0	31.0	0.0	0.0	0.3	38.3	59.9	1.5	0.0	1.2	26.2	68.8	3.8	
MC_{PHKB}	0.0	0.0	37.5	0.0	62.5	0.0	0.0	54.1	0.0	45.9	0.0	0.0	60.4	0.0	39.6	
MC_{PLW}	0.0	0.0	0.1	0.0	99.9	0.0	0.0	17.4	0.0	82.6	0.0	0.0	39.7	0.0	60.3	
MC_{PDK}	0.0	0.0	91.7	0.0	8.3	0.0	0.0	80.6	0.0	19.4	0.0	0.0	71.2	2.5	26.3	
IC_{95}	1.9	97.3	0.0	0.8	0.0	3.7	9.8	36.6	49.9	0.0	0.0	0.0	39.2	4.4	0.9	
IC_{99}	5.0	93.4	0.1	1.5	0.0	2.0	4.0	59.9	34.1	0.0	0.0	0.0	21.8	1.1	0.5	
$LASSO$	3.7	94.9	0.0	1.4	0.0	9.6	45.8	2.3	42.3	0.0	0.8	14.5	3.6	81.1	0.0	
EN	0.0	80.1	0.0	19.9	0.0	0.0	30.9	12.1	57.0	0.0	0.0	6.4	73.5	20.1	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.64, 7.65, 7.66, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados empeoran conforme aumenta la colinealidad. Un detalle importante es que para las tres tablas se repite el mismo fenómeno, que los métodos basados en AIC con Ridge y los de Intervalos de Confianza son muy competitivos contra LASSO en lo que a selección de variables se refiere (cuando la varianza es baja o moderada) pero este último tiene alto desempeño cuando la varianza es alta siendo notoriamente mejor que los demás métodos.

Cuadro 7.64: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$								
		Colinealidad = Baja; P=10; n=25; S=1000								
SR	73789.2			7378.9			737.9			
σ^2	0.01			0.1			1			
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F	
AIC_{HKB}	0.00007	6.72959	0.97566	0.00039	6.95351	0.94415	0.00528	9.44170	0.61894	
AIC_{LW}	0.00007	6.72033	0.97599	0.00048	6.94209	0.93644	0.00603	10.04452	0.52935	
AIC_{DK}	0.00630	39.87216	0.60267	0.00677	39.69592	0.59094	0.00689	41.68574	0.51890	
C_{HKB}	0.01421	14.26961	0.28571	0.01450	14.32031	0.28229	0.01181	15.36243	0.25034	
C_{LW}	0.02223	15.92314	0.00000	0.02475	16.01085	0.00171	0.02077	17.48855	0.06857	
C_{DK}	0.00630	39.43581	0.59905	0.00677	39.07708	0.56534	0.00689	41.71105	0.51613	
MCp_{HKB}	0.01421	14.26961	0.28571	0.01450	14.32022	0.28200	0.01181	15.39262	0.24400	
MCp_{LW}	0.02223	15.92314	0.00000	0.02475	16.01085	0.00171	0.02077	17.48998	0.06800	
MCp_{DK}	0.01411	14.32192	0.28571	0.01441	14.50057	0.28286	0.01133	15.55537	0.23568	
IC_{95}	0.00014	6.72424	0.78274	0.00094	6.84991	0.85148	0.00880	11.58965	0.38048	
IC_{99}	0.00014	6.72336	0.79343	0.00085	6.84105	0.87452	0.00880	13.35902	0.26473	
$LASSO$	0.00032	7.65422	0.93246	0.00119	7.86983	0.88233	0.00373	9.88971	0.74060	
EN	0.00058	7.48470	0.82960	0.01136	9.37585	0.73788	0.03593	15.77759	0.41561	
R_{HKB}	0.00108	6.68783	-	0.00075	6.68322	-	0.00237	8.16613	-	
R_{LW}	0.00110	6.72369	-	0.00136	6.85400	-	0.00684	8.51067	-	
R_{DK}	0.00294	9.85242	-	0.00231	13.64823	-	0.00176	17.53655	-	
MCO	0.00014	6.72409	-	0.00138	6.85842	-	0.00761	8.57874	-	

Magister en Matemática Mención Estadística

Cuadro 7.65: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$									
<i>Colinealidad = Moderada; P=10; n=25; S=1000</i>									
SR	71004.5			7100.5			710.0		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00013	0.10495	0.86896	0.00053	0.25145	0.78465	0.01220	1.67942	0.56988
AIC_{LW}	0.00014	0.11885	0.86869	0.00132	0.30045	0.75669	0.01610	1.94169	0.48909
AIC_{DK}	0.00700	31.94194	0.18561	0.00585	33.54832	0.35046	0.00774	35.19136	0.46938
C_{HKB}	0.01860	2.07485	0.00000	0.01800	2.08514	0.00000	0.01798	2.06737	0.10358
C_{LW}	0.01875	1.97225	0.00000	0.01818	2.05751	0.00000	0.01840	2.69329	0.02486
C_{DK}	0.00700	32.32186	0.17782	0.00585	33.95557	0.33121	0.00685	35.50496	0.45100
MCp_{HKB}	0.01860	2.07485	0.00000	0.01800	2.15531	0.00029	0.01798	2.42465	0.08400
MCp_{LW}	0.01875	1.97225	0.00000	0.01818	2.05751	0.00000	0.01840	2.71472	0.02371
MCp_{DK}	0.01709	1.02238	0.02971	0.01210	1.28983	0.10714	0.01235	2.13106	0.13646
IC_{95}	0.00014	0.12317	0.79246	0.00415	0.33949	0.84565	0.01895	1.82218	0.28490
IC_{99}	0.00014	0.12261	0.80440	0.00415	0.33508	0.85434	0.01933	1.57636	0.18114
$LASSO$	0.00010	0.39278	0.89764	0.00043	0.77839	0.87177	0.00588	3.75984	0.72612
EN	0.00042	0.34648	0.84063	0.01529	2.67623	0.71040	0.00903	12.78434	0.41019
R_{HKB}	0.00012	0.10662	–	0.00090	0.23286	–	0.00475	1.56380	–
R_{LW}	0.00013	0.12345	–	0.00332	0.35678	–	0.01195	2.57923	–
R_{DK}	0.00218	2.35849	–	0.00184	4.73021	–	0.00226	8.02742	–
MCO	0.00014	0.12366	–	0.00343	0.36074	–	0.01245	2.76723	–

Magister en Matemática Mención Estadística

Cuadro 7.66: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$									
<i>Colinealidad = Alta; P=10; n=25; S=1000</i>									
SR	48214.2			4821.4			482.1		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00011	94.73094	0.92593	0.00307	92.83081	0.76822	0.01738	92.63375	0.45725
AIC_{LW}	0.00013	95.06737	0.92692	0.00422	93.21040	0.74430	0.01983	94.08853	0.37401
AIC_{DK}	0.00587	68.98147	0.67927	0.00606	61.27303	0.59320	0.00616	65.31357	0.49051
C_{HKB}	0.01643	122.21649	0.15371	0.02452	107.71025	0.18343	0.01558	101.57466	0.20561
C_{LW}	0.02854	119.32135	0.00029	0.02606	119.73722	0.05086	0.01727	113.41258	0.11543
C_{DK}	0.00650	76.98717	0.34679	0.00377	68.75317	0.41297	0.01672	66.03614	0.47408
MC_{PHKB}	0.01643	123.00000	0.10714	0.02452	112.58512	0.15457	0.01558	105.98808	0.17257
MC_{PLW}	0.02854	119.32135	0.00029	0.02606	119.80220	0.04971	0.01727	113.75598	0.11343
MC_{PDK}	0.01713	89.02003	0.26200	0.01428	97.16854	0.23029	0.01672	99.39683	0.21177
IC_{95}	0.00086	95.74558	0.82411	0.00223	94.33411	0.78196	0.01146	98.29731	0.15727
IC_{99}	0.00086	95.67926	0.84472	0.00223	94.62354	0.75269	0.01727	98.47437	0.07750
$LASSO$	0.00032	91.14804	0.89800	0.00187	92.62175	0.85298	0.01711	96.54871	0.71329
EN	0.00482	94.29172	0.77676	0.00232	95.33797	0.66331	0.01374	103.52268	0.40597
R_{HKB}	0.00033	94.79735	-	0.00232	93.75676	-	0.02650	95.34089	-
R_{LW}	0.00086	95.76195	-	0.00397	96.76979	-	0.06895	101.05081	-
R_{DK}	0.00158	87.20334	-	0.00110	83.43996	-	0.00201	82.03868	-
MCO	0.00087	95.77624	-	0.00402	96.92252	-	0.08271	103.12213	-

7.2.2.1.2 50 Observaciones.

Como se aprecia en los cuadros 7.67, 7.68,7.69, es muy difícil que un método de selección capture el modelo correcto, el mejor resultado en este sentido lo alcanzo AIC_{LW} seleccionando el modelo correcto en un 60.3% de las simulaciones cuando la varianza es baja y la colinealidad es baja, otro método que logro capturar el modelo correctamente una cantidad de veces significativa fue AIC_{HKB} con 60.1% de las simulaciones cuando la varianza es baja y la colinealidad es baja. También se logra ver que el aumento de la varianza es el factor que más perjudica una adecuada selección de variables. También se aprecia una mejoría en la selección respecto de los cuadros 7.61, 7.62 y 7.63 debido al mayor cantidad de observaciones disponibles.

Cuadro 7.67: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$															
<i>Colinealidad = Baja; P=10; n=50; S=1000</i>															
SR	139366.6					13936.7					1393.7				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	60.1	39.9	0.0	0.0	0.0	56.1	43.8	0.1	0.0	0.0	3.8	37.9	19.8	38.5	0.0
AIC_{LW}	60.3	39.7	0.0	0.0	0.0	59.5	40.4	0.1	0.0	0.0	0.8	0.8	53.3	45.1	0.0
AIC_{DK}	0.1	0.0	37.9	62.0	0.0	0.3	0.4	40.9	58.4	0.0	0.4	0.5	48.1	50.0	1.0
C_{HKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	91.5	0.0	8.5
C_{LW}	0.0	0.0	89.9	0.0	10.1	0.0	0.0	76.8	0.0	23.2	0.0	0.0	90.4	0.0	9.6
C_{DK}	0.0	0.0	69.4	30.6	0.0	0.1	0.2	55.9	43.8	0.0	0.3	0.4	51.8	46.2	1.3
MCp_{HKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	94.1	0.0	5.9
MCp_{LW}	0.0	0.0	91.4	0.0	8.6	0.0	0.0	77.3	0.0	22.7	0.0	0.0	90.6	0.0	9.4
MCp_{DK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	92.3	0.0	7.7
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.7	99.3	0.0	0.0	0.0	2.2	7.1	31.2	59.5	0.0
IC_{99}	0.0	100.0	0.0	0.0	0.0	2.4	97.6	0.0	0.0	0.0	0.7	2.3	54.8	41.1	0.3
$LASSO$	35.1	64.9	0.0	0.0	0.0	12.8	87.2	0.0	0.0	0.0	0.6	74.0	0.0	25.4	0.0
EN	0.0	100.0	0.0	0.0	0.0	1.6	89.4	7.3	1.7	0.0	0.7	24.5	59.5	12.6	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.68: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$														
		Colinealidad = Moderada; P=10; n=50; S=1000														
SR	143104.8					14310.5					1431.0					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	62.7	37.3	0.0	0.0	0.0	56.5	42.3	0.4	0.8	0.0	3.5	27.2	20.3	49.0	0.0	
AIC_{LW}	63.2	36.8	0.0	0.0	0.0	59.4	39.3	0.4	0.9	0.0	0.1	0.4	49.9	49.6	0.0	
AIC_{DK}	0.0	0.0	19.2	46.2	34.6	0.0	0.4	23.0	49.6	27.0	0.2	0.7	15.8	69.1	14.2	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	6.6	0.1	93.3	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	99.9	
C_{DK}	0.0	0.0	10.5	41.3	48.2	0.0	0.1	16.6	49.1	34.2	0.2	0.5	15.0	68.3	16.0	
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.0	0.0	98.0	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	99.9	
MC_{PDK}	0.0	0.0	5.5	0.0	94.5	0.0	0.0	26.5	0.0	73.5	0.0	0.0	39.9	0.1	60.0	
IC_{95}	0.1	99.9	0.0	0.0	0.0	2.6	97.4	0.0	0.0	0.0	0.5	2.2	48.2	48.5	0.2	
IC_{99}	0.1	99.9	0.0	0.0	0.0	5.5	94.4	0.0	0.1	0.0	0.1	0.4	65.9	29.0	0.8	
$LASSO$	16.7	83.3	0.0	0.0	0.0	10.7	89.2	0.0	0.1	0.0	0.4	58.6	0.1	40.9	0.0	
EN	0.0	100.0	0.0	0.0	0.0	0.7	79.6	17.8	1.9	0.0	0.7	18.0	66.8	14.0	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.69: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$														
		Colinealidad = Alta; P=10; n=50; S=1000														
SR	145446.6					14544.7					1454.5					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	64.8	35.2	0.0	0.0	0	49.2	34.4	5.7	10.7	0.0	2.5	3.1	35.6	58.8	0.0	
AIC_{LW}	65.5	34.5	0.0	0.0	0	52.3	28.5	8.1	11.1	0.0	0.1	0.0	49.4	50.3	0.2	
AIC_{DK}	0.7	0.6	29.6	69.1	0	0.4	1.1	24.9	73.3	0.3	0.1	0.9	20.2	78.0	0.8	
C_{HKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	94.5	0.0	5.5	0.0	0.0	84.5	0.0	15.5	
C_{LW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	97.7	0.0	2.3	0.0	0.0	70.7	0.0	29.3	
C_{DK}	0.0	0.0	22.3	77.7	0.0	0.0	0.2	43.7	55.1	1.0	0.0	0.6	37.9	56.5	5.0	
MC_{PHKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	94.7	0.0	5.3	0.0	0.0	79.7	0.0	20.3	
MC_{PLW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	97.7	0.0	2.3	0.0	0.0	70.8	0.0	29.2	
MC_{PDK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	94.3	0.0	5.7	0.0	0.0	85.0	0.0	15.0	
IC_{95}	0.7	99.3	0.0	0.0	0.0	12.9	64.4	3.5	19.2	0.0	0.0	0.0	79.0	8.9	0.3	
IC_{99}	1.6	98.4	0.0	0.0	0.0	18.5	48.0	11.0	22.5	0.0	0.0	0.0	64.2	2.2	0.6	
$LASSO$	22.4	77.6	0.0	0.0	0.0	17.9	76.6	0.1	5.4	0.0	2.1	26.8	2.7	68.4	0.0	
EN	0.0	99.7	0.0	0.3	0.0	0.0	60.7	5.0	34.3	0.0	0.0	12.4	57.7	29.9	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.70, 7.71 y 7.72, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados empeoran conforme aumenta la colinealidad. Un detalle importante es que para las tres tablas se repite el mismo fenómeno, que los métodos basados en AIC con Ridge, los de Intervalos de Confianza y *ElasticNet* son muy competitivos contra LASSO en lo que respecta a selección de variables cuando la varianza es baja o moderada pero este último tiene un mejor desempeño cuando la varianza es alta siendo notoriamente mejor que los demás métodos, además LASSO alcanza un indicador $F=0.93595$ cuando la varianza es pequeña y la colinealidad es baja, por otra parte los resultados muestran que la estimación de coeficientes muestra un deterioro considerable en todos los métodos al punto que el efecto de la varianza reduce la calidad de los métodos hasta obtener resultados similares a *MCO*.

Magister en Matemática Mención Estadística

Cuadro 7.70: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$ <i>Colinealidad = Baja; P=10; n=50; S=1000</i>								
SR	139366.6			13936.7			1393.7		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00001	0.05694	0.96361	0.00012	0.17860	0.95961	0.00570	1.47038	0.78411
AIC_{LW}	0.00001	0.05792	0.96383	0.00012	0.19284	0.96305	0.00735	1.83206	0.70779
AIC_{DK}	0.00487	15.42966	0.62604	0.00659	16.53845	0.60368	0.00673	17.96329	0.51355
C_{HKB}	0.01362	10.80288	0.28571	0.01360	10.88630	0.28571	0.01655	11.76343	0.26143
C_{LW}	0.01379	10.87801	0.25686	0.02020	10.97278	0.21943	0.02043	12.50288	0.25829
C_{DK}	0.01414	9.96833	0.47624	0.01438	14.08942	0.51550	0.00673	17.49041	0.49723
MCp_{HKB}	0.01362	10.80288	0.28571	0.01360	10.88630	0.28571	0.01655	11.90228	0.26886
MCp_{LW}	0.01379	10.87449	0.26114	0.02020	10.96904	0.22086	0.02043	12.49831	0.25886
MCp_{DK}	0.01414	4.92771	0.28571	0.01438	4.90284	0.28571	0.01243	6.71918	0.26371
IC_{95}	0.00001	0.05873	0.76980	0.00028	0.19781	0.81198	0.00908	1.85559	0.72173
IC_{99}	0.00001	0.05871	0.77626	0.00028	0.19749	0.83220	0.00499	2.41058	0.64275
$LASSO$	0.00009	0.01608	0.93595	0.00027	0.15192	0.89041	0.00436	1.64633	0.78371
EN	0.00014	0.10328	0.83369	0.00044	1.65164	0.81667	0.04561	15.93706	0.51802
R_{HKB}	0.00101	0.05836	–	0.00025	0.18424	–	0.00440	1.34015	–
R_{LW}	0.00101	0.05873	–	0.00028	0.19806	–	0.00789	1.69685	–
R_{DK}	0.00102	0.18974	–	0.00038	1.12161	–	0.00170	4.47905	–
MCO	0.00001	0.05873	–	0.00028	0.19814	–	0.00797	1.70433	–

Magister en Matemática Mención Estadística

Cuadro 7.71: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$ <i>Colinealidad = Moderada; P=10; n=50; S=1000</i>								
SR	143104.8			14310.5			1431.0		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	0.33893	0.96636	0.00039	0.44998	0.95940	0.01333	1.70988	0.76596
AIC_{LW}	0.00003	0.33932	0.96687	0.00043	0.45385	0.96256	0.01654	1.74905	0.68041
AIC_{DK}	0.00623	12.56679	0.35863	0.00666	13.43792	0.37510	0.00586	16.06630	0.41517
C_{HKB}	0.01851	0.27266	0.00000	0.01847	0.36537	0.00000	0.01806	1.40379	0.01908
C_{LW}	0.01856	0.26869	0.00000	0.01853	0.36628	0.00000	0.01846	1.38646	0.00029
C_{DK}	0.00623	12.44996	0.24306	0.00666	13.63071	0.29647	0.00691	16.10378	0.38419
MCp_{HKB}	0.01851	0.27266	0.00000	0.01847	0.34953	0.00000	0.01851	1.30423	0.00571
MCp_{LW}	0.01856	0.26869	0.00000	0.01853	0.36628	0.00000	0.01846	1.38509	0.00029
MCp_{DK}	0.01886	0.72428	0.01571	0.01303	0.60596	0.07571	0.01155	1.70144	0.11425
IC_{95}	0.00004	0.33939	0.77400	0.00046	0.45641	0.82414	0.01288	1.74903	0.65336
IC_{99}	0.00004	0.33940	0.78089	0.00043	0.45522	0.84749	0.01255	1.78731	0.53313
$LASSO$	0.00002	0.22476	0.90383	0.00045	0.30261	0.88022	0.00516	1.72119	0.77979
EN	0.00004	0.28662	0.83365	0.00126	2.55296	0.75221	0.01870	13.00828	0.49382
R_{HKB}	0.00003	0.34303	–	0.00039	0.48447	–	0.00575	1.69878	–
R_{LW}	0.00004	0.33936	–	0.00054	0.45770	–	0.01033	1.74013	–
R_{DK}	0.00014	0.89452	–	0.00217	1.54277	–	0.00217	3.27320	–
MCO	0.00004	0.33934	–	0.00054	0.45754	–	0.01041	1.74086	–

Magister en Matemática Mención Estadística

Cuadro 7.72: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$									
<i>Colinealidad = Alta; P=10; n=50; S=1000</i>									
SR	145446.6			14544.7			1454.5		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00004	51.84172	0.96748	0.00167	51.88932	0.94336	0.01359	53.10042	0.64782
AIC_{LW}	0.00004	51.83152	0.96828	0.00201	51.73899	0.94593	0.01418	51.83342	0.55477
AIC_{DK}	0.00672	39.17270	0.68196	0.00483	38.90628	0.62501	0.00430	39.72643	0.54743
C_{HKB}	0.01820	55.92926	0.28571	0.01719	54.53233	0.27000	0.01530	49.40884	0.24143
C_{LW}	0.01869	56.45194	0.28571	0.01761	56.94274	0.27914	0.02247	61.08043	0.20200
C_{DK}	0.00587	47.11014	0.45588	0.01045	40.34677	0.46012	0.01574	41.09867	0.42603
MC_{PHKB}	0.01820	55.95986	0.28571	0.01719	56.72555	0.27057	0.01530	53.40315	0.22771
MC_{PLW}	0.01869	56.45194	0.28571	0.01761	56.94274	0.27914	0.02247	61.10194	0.20229
MC_{PDK}	0.01863	56.33150	0.28571	0.02570	50.69416	0.26943	0.01574	48.40540	0.24286
IC_{95}	0.00006	51.87986	0.79836	0.00246	52.81664	0.86705	0.01392	46.89443	0.40880
IC_{99}	0.00006	51.86247	0.81394	0.00246	53.02245	0.88691	0.01392	40.75910	0.25729
$LASSO$	0.00008	51.44038	0.92203	0.00069	51.92182	0.90112	0.00808	55.93831	0.76479
EN	0.00008	51.59045	0.80953	0.00227	52.43694	0.73705	0.00723	46.93521	0.44825
R_{HKB}	0.00009	52.17323	–	0.00200	53.70205	–	0.00388	56.35595	–
R_{LW}	0.00007	51.86091	–	0.00244	52.46010	–	0.00775	56.83339	–
R_{DK}	0.00027	54.74948	–	0.00103	55.72243	–	0.00195	55.58365	–
MCO	0.00007	51.85919	–	0.00244	52.44593	–	0.00786	56.93028	–

7.2.2.1.3 100 Observaciones.

Como se aprecia en los cuadros 7.73, 7.74, 7.75, es muy difícil que un método de selección capture el modelo correcto, el mejor resultado en este sentido lo alcanzó AIC_{LW} seleccionando el modelo correcto en un 55.7% de las simulaciones cuando la varianza es moderada y la colinealidad es baja, también se logra ver que el aumento de la varianza es el factor que más influye en la selección de variables. También se aprecia una mejoría en la selección respecto de los cuadros 7.61, 7.62 y 7.63 debido al mayor cantidad de observaciones disponibles (n=100).

Cuadro 7.73: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$															
<i>Colinealidad = Baja; P=10; n=100; S=1000</i>															
SR	259140.1					25914.0					2591.4				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	53.8	46.2	0.0	0.0	0	53.4	46.6	0.0	0.0	0.0	3.0	92.2	1.2	3.6	0.0
AIC_{LW}	54.2	45.8	0.0	0.0	0	55.7	44.3	0.0	0.0	0.0	20.3	19.2	29.3	31.2	0.0
AIC_{DK}	0.0	0.0	47.4	52.6	0	0.0	0.0	54.7	45.3	0.0	0.3	0.6	47.6	50.3	1.2
C_{HKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100.0	0.0	0.0	49.1	0.0	50.9
C_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.5	0.0	99.5	0.0	0.0	13.9	0.0	86.1
C_{DK}	0.0	0.0	59.9	40.1	0.0	0.3	0.0	48.5	51.2	0.0	0.1	0.4	49.2	49.0	1.3
MCp_{HKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100.0	0.0	0.0	37.1	0.0	62.9
MCp_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.3	0.0	99.7	0.0	0.0	13.3	0.0	86.7
MCp_{DK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	91.4	0.0	8.6
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	6.5	71.6	2.4	19.5	0.0
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.2	99.8	0.0	0.0	0.0	13.0	52.0	8.4	26.6	0.0
$LASSO$	16.4	83.6	0.0	0.0	0.0	11.1	88.9	0.0	0.0	0.0	0.2	96.3	0.0	3.5	0.0
EN	0.0	100.0	0.0	0.0	0.0	1.1	94.6	3.9	0.4	0.0	1.2	43.9	45.5	8.8	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.74: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$															
Colinealidad = Moderada; P=10; n=100; S=1000															
SR	285019.8					28502.0					2850.2				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	57.3	42.7	0.0	0.0	0.0	56.0	44.0	0.0	0.0	0.0	4.6	79.5	4.8	11.1	0.0
AIC_{LW}	57.3	42.7	0.0	0.0	0.0	58.5	41.5	0.0	0.0	0.0	7.7	6.6	41.0	44.7	0.0
AIC_{DK}	0.1	0.0	16.3	47.1	36.5	0.4	0.5	26.2	54.5	18.4	0.5	0.2	25.2	62.1	12.0
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	17.1	0.0	82.9
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	13.5	0.0	86.5
C_{DK}	0.0	0.0	0.2	56.2	43.6	0.0	0.0	6.7	66.9	26.4	0.1	0.1	12.7	70.0	17.1
MCp_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	15.4	0.0	84.6
MCp_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	13.5	0.0	86.5
MCp_{DK}	0.0	0.0	24.8	0.0	75.2	0.0	0.0	29.3	0.0	70.7	0.0	0.0	34.4	0.0	65.6
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	9.3	38.4	10.1	42.2	0.0
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.5	99.5	0.0	0.0	0.0	8.2	19.3	28.3	44.2	0.0
$LASSO$	17.6	82.4	0.0	0.0	0.0	9.2	90.8	0.0	0.0	0.0	0.7	88.4	0.0	10.9	0.0
EN	0.0	100.0	0.0	0.0	0.0	0.5	92.3	7.1	0.1	0.0	2.7	37.1	51.6	8.5	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.75: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$															
Colinealidad = Alta; P=10; n=100; S=1000															
SR	282404.4					28240.4					2824.0				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	55.4	44.6	0.0	0.0	0.0	48.7	50.4	0.2	0.7	0.0	4.8	25.1	20.3	49.8	0.0
AIC_{LW}	55.7	44.3	0.0	0.0	0.0	53.4	45.4	0.4	0.8	0.0	0.7	0.9	45.5	52.9	0.0
AIC_{DK}	1.4	1.6	23.6	73.4	0.0	0.9	0.8	30.7	67.6	0.0	0.1	0.6	21.2	77.7	0.4
C_{HKB}	0.0	0.0	96.1	0.0	3.9	0.0	0.0	74.2	0.0	25.8	0.0	0.0	69.8	0.0	30.2
C_{LW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.0	0.0	2.0
C_{DK}	0.0	0.0	6.6	30.8	62.6	0.0	0.0	15.1	45.2	39.7	0.0	0.2	29.2	47.8	22.8
MCp_{HKB}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	89.1	0.0	10.9	0.0	0.0	78.4	0.0	21.6
MCp_{LW}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	98.1	0.0	1.9
MCp_{DK}	0.0	0.0	17.2	0.0	82.8	0.0	0.0	43.2	0.0	56.8	0.0	0.0	62.3	0.0	37.7
IC_{95}	0.2	99.8	0.0	0.0	0.0	3.3	95.6	0.0	1.1	0.0	0.3	0.5	62.4	36.7	0.0
IC_{99}	0.2	99.8	0.0	0.0	0.0	8.1	89.7	0.2	2.0	0.0	0.1	0.2	82.8	15.4	0.1
$LASSO$	25.0	75.0	0.0	0.0	0.0	14.5	85.2	0.0	0.3	0.0	1.4	53.5	0.7	44.4	0.0
EN	0.0	100.0	0.0	0.0	0.0	0.0	92.4	5.0	2.6	0.0	0.0	24.1	49.6	26.3	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto del cuadro 7.76, se aprecia que AIC_{HKB} , AIC_{LW} , AIC_{DK} con una colinealidad Baja mantiene estable su capacidad selectiva a pesar del aumento de la varianza, en esta misma línea LASSO y *ElasticNet* mantienen su capacidad selectiva entre la varianza baja y moderada. Además es interesante observar que IC_{95} y IC_{99} mejoraron su desempeño conforme aumento la varianza. Con respecto del cuadro 7.77, se repitió el comportamiento del cuadro 7.76, pero es destacable que AIC_{LW} fue el que realizó mejor la tarea de selección cuando la varianza es pequeña. Por ultimo tomando en cuenta la tabla en el cuadro 7.78 se puede observar un deterioro en la capacidad selectiva de todos los métodos pero el impacto es amortiguado por la cantidad de observaciones, lo que permite realizar mejores estimaciones.

Cuadro 7.76: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

		$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)$								
		Colinealidad = Baja; P=10; n=100; S=1000								
SR		259140.1			25914.0			2591.4		
σ^2		0.01			0.1			1		
Indicador		TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	4.90394	0.95748	0.00008	4.99234	0.95563	0.00174	5.28087	0.80421	
AIC_{LW}	0.00003	4.90422	0.95785	0.00009	4.99452	0.95831	0.00251	5.97725	0.87830	
AIC_{DK}	0.00977	20.60304	0.54465	0.00445	21.32756	0.58612	0.00659	23.26961	0.51169	
C_{HKB}	0.01737	5.03381	0.00000	0.01748	5.09198	0.00000	0.01328	7.60522	0.14029	
C_{LW}	0.01742	5.08322	0.00000	0.01753	5.17539	0.00143	0.01693	6.87947	0.03971	
C_{DK}	0.00437	21.74485	0.64944	0.00622	21.24946	0.57767	0.00659	23.04280	0.50870	
MCp_{HKB}	0.01737	5.03381	0.00000	0.01748	5.09198	0.00000	0.01689	5.61501	0.10600	
MCp_{LW}	0.01742	5.08322	0.00000	0.01753	5.16806	0.00086	0.01693	6.83993	0.03800	
MCp_{DK}	0.01373	10.63863	0.28571	0.01326	9.39487	0.28571	0.01262	9.46711	0.26114	
IC_{95}	0.00004	4.89901	0.76195	0.00014	4.99003	0.78848	0.00266	6.01754	0.85231	
IC_{99}	0.00004	4.89898	0.76569	0.00014	4.99063	0.80163	0.00266	6.03061	0.87022	
LASSO	0.00007	5.03341	0.89687	0.00011	5.09861	0.87734	0.00216	5.63657	0.79419	
EN	0.00004	4.99168	0.82218	0.00021	4.63043	0.81113	0.01309	6.37654	0.59077	
R_{HKB}	0.00102	4.89146	-	0.00014	4.91651	-	0.00174	5.46413	-	
R_{LW}	0.00102	4.89903	-	0.00014	4.98874	-	0.00244	6.01507	-	
R_{DK}	0.00357	4.85084	-	0.00015	5.27341	-	0.00135	6.82265	-	
MCO	0.00004	4.89905	-	0.00014	4.98895	-	0.00244	6.01755	-	

Magister en Matemática Mención Estadística

Cuadro 7.77: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$ <i>Colinealidad = Moderada; P=10; n=100; S=1000</i>									
SR	285019.8			28502.0			2850.2		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00002	3.07514	0.96006	0.00010	3.18207	0.95846	0.00085	4.06175	0.81753
AIC_{LW}	0.00002	3.07459	0.96006	0.00011	3.17633	0.96127	0.00009	4.01953	0.83066
AIC_{DK}	0.00522	17.02145	0.27293	0.00661	17.03772	0.41909	0.00444	18.28922	0.44025
C_{HKB}	0.01865	3.48255	0.00000	0.01894	3.54593	0.00000	0.01792	4.11152	0.04886
C_{LW}	0.01868	3.47249	0.00000	0.01898	3.54681	0.00000	0.01796	4.00520	0.03857
C_{DK}	0.00522	17.66745	0.16385	0.00670	17.95571	0.27133	0.00700	19.05415	0.34696
MCp_{HKB}	0.01865	3.48255	0.00000	0.01894	3.40169	0.00000	0.01792	3.65786	0.04400
MCp_{LW}	0.01868	3.47249	0.00000	0.01898	3.54681	0.00000	0.01796	4.01605	0.03857
MCp_{DK}	0.01334	4.77261	0.07086	0.01833	4.26245	0.08371	0.01290	4.59812	0.09829
IC_{95}	0.00003	3.07385	0.76462	0.00017	3.18171	0.79875	0.00095	3.98792	0.83904
IC_{99}	0.00003	3.07385	0.76970	0.00017	3.18180	0.81577	0.00031	4.00370	0.83150
$LASSO$	0.00002	3.15133	0.89946	0.00011	3.35202	0.87752	0.00067	4.21173	0.79711
EN	0.00019	3.06846	0.84026	0.00066	2.90711	0.81396	0.00948	4.70173	0.60438
R_{HKB}	0.00003	3.07514	–	0.00015	3.19370	–	0.00093	4.04751	–
R_{LW}	0.00003	3.07383	–	0.00017	3.18194	–	0.00099	3.98178	–
R_{DK}	0.00005	3.92996	–	0.00177	4.46060	–	0.00144	6.26019	–
MCO	0.00003	3.07383	–	0.00017	3.18191	–	0.00099	3.98160	–

Magister en Matemática Mención Estadística

Cuadro 7.78: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)$									
<i>Colinealidad = Alta; P=10; n=100; S=1000</i>									
SR	282404.4			28240.4			2824.0		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00001	98.77088	0.95911	0.00013	98.51768	0.95095	0.00346	100.15198	0.78065
AIC_{LW}	0.00001	98.76390	0.95973	0.00016	98.55054	0.95697	0.00347	102.10995	0.70423
AIC_{DK}	0.00268	69.46656	0.71546	0.00492	72.25024	0.68087	0.00460	69.42569	0.56441
C_{HKB}	0.01810	124.12188	0.27457	0.02338	124.93790	0.21200	0.01623	118.93509	0.19943
C_{LW}	0.01831	124.37514	0.28571	0.01741	124.25912	0.28571	0.02039	124.93639	0.28000
C_{DK}	0.02512	115.60598	0.16269	0.02351	105.63120	0.24738	0.01648	93.29280	0.32452
MC_{PHKB}	0.01810	124.07486	0.28571	0.01722	124.50361	0.25457	0.01623	120.31648	0.22400
MC_{PLW}	0.01831	124.37514	0.28571	0.01741	124.25912	0.28571	0.02039	124.93706	0.28029
MC_{PDK}	0.01830	125.76984	0.04914	0.02351	125.96601	0.12343	0.01648	119.83919	0.17800
IC_{95}	0.00003	98.88197	0.77795	0.00024	98.87201	0.83800	0.00356	98.00892	0.68016
IC_{99}	0.00003	98.88352	0.78667	0.00024	98.79022	0.86397	0.01148	98.96573	0.58761
$LASSO$	0.00002	98.32509	0.91660	0.00015	98.17646	0.89565	0.00262	101.79553	0.78889
EN	0.00003	98.48305	0.82034	0.00044	99.99220	0.79879	0.02111	126.90114	0.53538
R_{HKB}	0.00003	98.97562	-	0.00026	99.64651	-	0.00244	102.26397	-
R_{LW}	0.00003	98.87872	-	0.00024	98.99230	-	0.00435	101.97528	-
R_{DK}	0.00010	99.48829	-	0.00183	99.18912	-	0.00172	97.74271	-
MCO	0.00003	98.87847	-	0.00024	98.99007	-	0.00437	101.97724	-

7.2.2.2. 15 Variables Disponibles.

7.2.2.2.1 25 Observaciones.

Como se aprecia en los cuadros 7.79, 7.80 y 7.81, los resultados empeoran respecto de las tablas mostradas en los cuadros 7.61, 7.62 y 7.63 debido a la presencia de más variables irrelevantes. Se logra también apreciar en las tablas como la capacidad de selección disminuye en la medida que aumenta la varianza, también si se comparan los resultados entre tablas para visualizar el efecto del aumento de la colinealidad en la selección de variables pero el resultado es poco concluyente. Se puede destacar que además que los métodos C_{HKB} , C_{LW} , MC_{PHKB} y MC_{PLW} fallaron en todas las simulaciones cuando la colinealidad es baja y la varianza es baja, y aunque tuvieron mejorías a medida que aumento la varianza son más bien despreciables. Mientras que AIC_{HKB} obtuvo el mejor resultado consiguiendo seleccionar el modelo correcto en un 2.7% de las simulaciones cuando la colinealidad es baja y la varianza es pequeña.

Cuadro 7.79: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Baja; P=15; n=25; S=1000															
SR	68466.7					6846.7					684.7				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	2.7	97.3	0.0	0.0	0.0	8.6	55.5	1.5	34.4	0.0	0.1	10.7	6.7	82.2	0.3
AIC_{LW}	2.6	97.4	0.0	0.0	0.0	8.0	41.2	3.2	47.6	0.0	0.0	0.1	26.4	72.3	1.2
AIC_{DK}	0.0	0.0	1.7	98.0	0.3	0.0	0.7	4.4	89.9	5.0	0.0	6.0	3.3	85.3	5.4
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.0	0.0	98.0	0.0	0.0	7.6	58.8	33.6
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.9	0.0	97.1
C_{DK}	0.0	0.0	2.6	97.1	0.3	0.0	0.4	5.1	88.8	5.7	0.1	3.6	4.1	85.6	6.6
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	6.0	0.0	94.0
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.6	0.0	97.4
MC_{PDK}	0.0	0.0	39.0	0.0	61.0	0.0	0.0	50.1	0.0	49.9	0.0	0.0	49.1	0.1	50.8
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.6	51.6	3.6	44.2	0.0	0.0	0.7	24.9	38.3	7.2
IC_{99}	0.0	100.0	0.0	0.0	0.0	1.0	37.8	9.9	51.3	0.0	0.0	0.3	23.1	21.2	6.6
LASSO	2.6	97.4	0.0	0.0	0.0	0.4	96.7	0.0	2.9	0.0	0.0	39.4	0.0	60.6	0.0
EN	0.0	97.0	0.0	3.0	0.0	0.0	32.8	20.1	47.1	0.0	0.0	5.5	75.6	16.9	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.80: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Moderada; P=15; n=25; S=1000														
SR	63361.4					6336.1					633.6					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	26.2	73.8	0.0	0.0	0.0	10.5	28.6	4.2	56.7	0.0	0	4.4	7.7	87.6	0.3	
AIC_{LW}	27.2	72.7	0.0	0.1	0.0	9.8	21.6	6.4	62.2	0.0	0	0.0	18.3	80.0	1.7	
AIC_{DK}	0.0	0.0	0.7	94.3	5.0	0.0	1.1	1.4	89.1	8.4	0	3.7	2.3	83.3	10.7	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	2.7	97.3	0.0	0.0	0.7	76.7	22.6	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
C_{DK}	0.0	0.0	0.0	94.5	5.5	0.0	0.8	0.2	89.4	9.6	0.0	2.4	2.3	83.3	12.0	
MC_{pHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.5	0.0	99.5	
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	99.9	
MC_{pDK}	0.0	0.0	0.3	0.0	99.7	0.0	0.0	2.4	0.0	97.6	0.0	0.0	17.3	0.1	82.6	
IC_{95}	0.0	100.0	0.0	0.0	0.0	2.0	56.4	5.0	36.6	0.0	0.0	0.2	22.1	29.8	8.0	
IC_{99}	0.0	100.0	0.0	0.0	0.0	3.3	42.2	11.9	42.6	0.0	0.0	0.0	15.9	16.0	5.5	
$LASSO$	0.5	99.5	0.0	0.0	0.0	1.0	88.7	0.0	10.3	0.0	0.0	32.6	0.0	67.4	0.0	
EN	0.0	90.4	0.1	9.5	0.0	0.0	28.9	16.8	54.3	0.0	0.0	4.1	75.5	20.2	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.81: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Alta; P=15; n=25; S=1000														
SR	66323.6					6632.4					663.2					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	28.8	71.0	0.0	0.2	0.0	3.7	7.6	11.7	77.0	0.0	0	0.4	11.4	85.2	3.0	
AIC_{LW}	32.0	67.8	0.0	0.2	0.0	2.7	5.1	17.7	74.5	0.0	0	0.0	24.4	67.6	8.0	
AIC_{DK}	0.3	0.2	3.3	96.2	0.0	0.0	1.0	3.6	91.1	4.3	0	1.3	6.8	73.8	18.1	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.4	20.6	77.0	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.3	0.0	99.7	
C_{DK}	0.0	0.0	0.0	92.2	7.8	0.0	0.2	0.0	75.3	24.5	0.0	1.1	2.2	71.3	25.4	
MC_{pHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.8	0.0	99.2	
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	
MC_{pDK}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.3	0.0	99.7	0.0	0.0	10.5	1.9	87.6	
IC_{95}	0.1	98.3	0.0	1.6	0.0	0.1	9.3	14.9	75.4	0.1	0.0	0.0	19.0	14.4	5.7	
IC_{99}	0.6	95.9	0.1	3.4	0.0	0.2	4.8	27.3	66.3	0.3	0.0	0.0	10.5	6.6	4.1	
$LASSO$	0.0	70.3	0.0	29.7	0.0	0.1	25.4	0.3	74.2	0.0	0.0	20.0	0.0	80.0	0.0	
EN	0.0	37.4	0.0	62.6	0.0	0.0	10.0	21.4	68.6	0.0	0.0	2.7	74.7	22.4	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.82, 7.83 y 7.84 se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados se mantienen relativamente estables conforme aumenta la colinealidad lo que evidencia no ser un factor relevante para la selección, por otro lado conforme aumenta la colinealidad se ve afectada negativamente las estimaciones de los coeficientes y las predicciones. También es de notar que cuando el indicador F marco cero, específicamente se refiere esto a C_{HKB} y C_{LW} , MCp_{HKB} y MCp_{LW} es porque el método falló en todas las simulaciones.

Cuadro 7.82: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ <i>Colinealidad = Baja; P=15; n=25; S=1000</i>								
SR	68466.7			6846.7			684.7		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00015	0.91613	0.88949	0.00051	0.84581	0.83314	0.00316	2.32816	0.56366
AIC_{LW}	0.00014	0.94945	0.88972	0.00375	0.87967	0.78900	0.00826	2.96927	0.47505
AIC_{DK}	0.00689	26.25430	0.43236	0.00613	26.04479	0.41301	0.00683	27.71610	0.44555
C_{HKB}	0.01636	3.72314	0.00000	0.01679	4.22462	0.00571	0.00591	3.90581	0.22014
C_{LW}	0.01681	3.93305	0.00000	0.01711	4.52408	0.00000	0.01615	6.75304	0.00829
C_{DK}	0.00689	26.44920	0.42908	0.00613	26.28002	0.40950	0.00683	27.84612	0.43125
MCp_{HKB}	0.01658	4.78029	0.00000	0.01679	4.84562	0.00000	0.01569	5.85593	0.01714
MCp_{LW}	0.01681	3.97569	0.00000	0.01711	4.55517	0.00000	0.01604	6.71299	0.00743
MCp_{DK}	0.01180	5.22531	0.11143	0.01187	5.43620	0.14314	0.01044	6.65231	0.14054
IC_{95}	0.00019	0.99881	0.62566	0.00239	1.66740	0.70719	0.00823	6.48106	0.25651
IC_{99}	0.00018	0.99767	0.64410	0.00205	1.71076	0.72986	0.00823	7.55876	0.17105
$LASSO$	0.00030	1.33276	0.84227	0.00094	1.75155	0.77225	0.01146	4.45145	0.57336
EN	0.00093	1.51668	0.73361	0.00599	3.83745	0.57212	0.01240	10.92150	0.35533
R_{HKB}	0.00121	0.96585	-	0.00129	1.04564	-	0.00666	2.62005	-
R_{LW}	0.00119	1.00005	-	0.00215	1.57552	-	0.02692	5.89121	-
R_{DK}	0.00338	2.96358	-	0.00356	6.35078	-	0.00331	9.56895	-
MCO	0.00019	1.00040	-	0.00223	1.60289	-	0.03348	7.78359	-

Magister en Matemática Mención Estadística

Cuadro 7.83: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Moderada; P=15; n=25; S=1000								
SR	63361.4			6336.1			633.6		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00019	0.58714	0.91599	0.00124	0.91226	0.78793	0.01517	2.84343	0.52704
AIC_{LW}	0.00025	0.58099	0.91712	0.00153	0.97566	0.75865	0.01662	3.82113	0.42243
AIC_{DK}	0.00542	32.34851	0.38741	0.00668	32.33787	0.37500	0.00656	33.69588	0.38302
C_{HKB}	0.01631	20.39365	0.00000	0.01424	18.72052	0.00492	0.00738	6.89337	0.18994
C_{LW}	0.01762	14.07875	0.00000	0.01683	14.26158	0.00000	0.01998	16.42454	0.00000
C_{DK}	0.00594	32.70370	0.38362	0.00668	32.73462	0.35000	0.00656	34.05106	0.36809
MCp_{HKB}	0.01707	13.59161	0.00000	0.01643	12.49842	0.00000	0.01635	13.27090	0.00143
MCp_{LW}	0.01762	14.07875	0.00000	0.01683	14.22801	0.00000	0.01998	16.25394	0.00029
MCp_{DK}	0.01712	7.68275	0.00086	0.01487	11.03361	0.00686	0.01579	12.97052	0.04968
IC_{95}	0.00033	0.61025	0.63368	0.00291	1.03047	0.71910	0.01650	12.74976	0.20366
IC_{99}	0.00033	0.60793	0.65470	0.00291	1.11109	0.74051	0.02070	15.97924	0.11958
$LASSO$	0.00023	1.51336	0.81886	0.00083	2.08127	0.78012	0.01497	5.90788	0.57393
EN	0.00173	3.11239	0.74508	0.00030	12.01938	0.53278	0.02436	20.78692	0.34320
R_{HKB}	0.00021	0.55264	–	0.00189	0.72895	–	0.00865	2.70519	–
R_{LW}	0.00033	0.60952	–	0.00283	1.03502	–	0.02459	5.09391	–
R_{DK}	0.00204	2.08343	–	0.00228	5.57836	–	0.00351	10.27661	–
MCO	0.00033	0.61043	–	0.00287	1.05223	–	0.02691	5.93043	–

Magister en Matemática Mención Estadística

Cuadro 7.84: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)$ <i>Colinealidad = Alta; P=15; n=25; S=1000</i>									
SR	66323.6			6632.4			663.2		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00054	3.98113	0.92014	0.00751	3.33383	0.72084	0.01318	5.61638	0.43068
AIC_{LW}	0.00084	4.10113	0.92450	0.00806	3.48316	0.69666	0.01542	6.80832	0.35301
AIC_{DK}	0.00700	32.64017	0.53888	0.00693	32.46620	0.49071	0.00837	34.04261	0.35720
C_{HKB}	0.01763	3.55808	0.00000	0.01770	3.67424	0.00000	0.01765	5.64545	0.06320
C_{LW}	0.01866	2.97470	0.00000	0.01884	3.08015	0.00000	0.01907	4.36674	0.00086
C_{DK}	0.00620	33.87873	0.26790	0.00716	33.06994	0.25459	0.00698	35.56599	0.27729
MC_{PHKB}	0.01763	3.55808	0.00000	0.01770	3.67424	0.00000	0.01765	5.24996	0.00229
MC_{PLW}	0.01866	2.97470	0.00000	0.01884	3.08015	0.00000	0.01907	4.36123	0.00114
MC_{PDK}	0.01723	3.79858	0.00000	0.01735	4.16977	0.00086	0.01745	6.52424	0.03496
IC_{95}	0.00162	4.24657	0.66628	0.00716	3.75712	0.61424	0.01526	9.20479	0.11166
IC_{99}	0.00162	4.23487	0.69577	0.00473	3.54131	0.58451	0.01511	10.01380	0.05541
$LASSO$	0.00080	3.54350	0.71756	0.00304	5.09152	0.68237	0.01161	9.99671	0.55206
EN	0.00127	8.30829	0.57539	0.00886	14.82156	0.39610	0.01113	18.65748	0.31388
R_{HKB}	0.00066	3.58589	–	0.00327	3.21395	–	0.01387	5.84944	–
R_{LW}	0.00159	4.24417	–	0.00907	5.06171	–	0.04432	10.88747	–
R_{DK}	0.00141	5.37845	–	0.00437	8.19147	–	0.00370	12.20781	–
MCO	0.00162	4.25798	–	0.00947	5.28688	–	0.05327	17.94093	–

7.2.2.2.2 50 Observaciones.

Como se aprecia en los cuadros 7.85, 7.86 y 7.87, los resultados mejoran respecto de las tablas mostradas en los cuadros 7.79, 7.80 y 7.81 debido a la presencia de más observaciones. Se logra también apreciar en las tablas como la capacidad de selección disminuye en la medida que aumenta la varianza, también si se comparan los resultados entre tablas para visualizar el efecto del aumento de la colinealidad se aprecia más bien un impacto negativo muy leve. Se puede destacar que además que los métodos C_{HKB} , C_{LW} , MC_{PHKB} , MC_{PLW} y MC_{PDK} fallaron en todas las simulaciones para todos los niveles de colinealidad y con varianza baja y moderada y aunque tuvieron mejorías a medida que aumento la varianza son más bien despreciables. Mientras que AIC_{LW} obtuvo el mejor resultado consiguiendo seleccionar el modelo correcto en un 33.3% de las simulaciones cuando la colinealidad es moderada y la varianza es pequeña.

Cuadro 7.85: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Baja; P=15; n=50; S=1000														
SR	123750.9					12375.1					1237.5					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	33.1	66.9	0.0	0.0	0.0	28.8	71.2	0.0	0.0	0.0	0.6	53.8	3.0	42.6	0.0	
AIC_{LW}	33.3	66.7	0.0	0.0	0.0	33.2	66.8	0.0	0.0	0.0	1.0	1.0	24.1	73.8	0.1	
AIC_{DK}	0.0	0.0	25.6	74.4	0.0	0.0	0.1	28.2	69.7	2.0	0.0	0.7	16.8	75.7	6.8	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	6.8	3.6	89.6	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
C_{DK}	0.0	0.0	20.2	79.8	0.0	0.0	0.0	27.2	70.7	2.1	0.0	0.4	17.2	75.0	7.4	
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	6.3	0.0	93.7	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
MC_{PDK}	0.0	0.0	96.9	0.0	3.1	0.0	0.0	83.6	0.0	16.4	0.0	0.0	60.7	0.0	39.3	
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	99.9	0.0	0.1	0.0	0.4	3.5	16.2	79.2	0.4	
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.1	99.8	0.0	0.1	0.0	0.0	0.8	35.0	60.1	1.1	
$LASSO$	3.8	96.2	0.0	0.0	0.0	0.8	99.2	0.0	0.0	0.0	0.0	66.5	0.0	33.5	0.0	
EN	0.0	100.0	0.0	0.0	0.0	0.0	87.9	3.7	8.4	0.0	0.1	17.8	56.2	24.8	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.86: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Moderada; P=15; n=50; S=1000														
SR	132912.6					13291.3					1329.1					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	19.5	80.5	0.0	0.0	0.0	20.2	78.5	0.0	1.3	0.0	0.3	39.8	3.7	56.2	0.0	
AIC_{LW}	20.1	79.9	0.0	0.0	0.0	22.6	75.7	0.0	1.7	0.0	0.0	0.8	21.1	78.1	0.0	
AIC_{DK}	0.0	0.0	1.2	29.5	69.3	0.0	0.0	4.4	67.2	28.4	0.0	0.3	5.9	72.7	21.1	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.7	2.6	95.7	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
C_{DK}	0.0	0.0	0.0	23.2	76.8	0.0	0.0	1.2	65.4	33.4	0.0	0.2	4.8	71.4	23.6	
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.2	0.0	99.8	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
MC_{PDK}	0.0	0.0	0.5	0.0	99.5	0.0	0.0	7.9	0.0	92.1	0.0	0.0	18.7	0.0	81.3	
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.2	99.8	0.0	0.0	0.0	0.0	2.4	24.3	71.4	0.7	
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.5	99.3	0.0	0.2	0.0	0.0	0.5	44.0	45.9	1.7	
$LASSO$	2.6	97.4	0.0	0.0	0.0	0.5	99.4	0.0	0.1	0.0	0.0	59.5	0.0	40.5	0.0	
EN	0.0	100.0	0.0	0.0	0.0	0.0	72.0	7.8	20.2	0.0	0.0	12.2	62.4	25.1	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.87: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Alta; P=15; n=50; S=1000														
SR	122048.1					12204.8					1220.5					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	28.7	71.3	0.0	0.0	0.0	18.9	60.5	1.6	19.0	0.0	0.3	4.8	9.6	84.9	0.4	
AIC_{LW}	30.0	70.0	0.0	0.0	0.0	20.2	53.5	3.2	23.1	0.0	0.1	0.0	21.6	77.0	1.3	
AIC_{DK}	0.1	0.1	29.1	70.7	0.0	0.4	0.9	15.9	82.8	0.0	0.1	1.0	9.5	84.2	5.2	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	3.0	0.2	96.8	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	4.5	0.0	95.5	
C_{DK}	0.0	0.0	0.0	73.4	26.6	0.0	0.0	0.0	68.8	31.2	0.0	0.6	0.6	69.2	29.6	
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.9	0.0	98.1	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	5.0	0.0	95.0	
MC_{PDK}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	5.0	0.0	95.0	
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.5	74.5	0.0	25.0	0.0	0.0	0.0	40.0	35.9	5.8	
IC_{99}	0.0	100.0	0.0	0.0	0.0	1.6	59.6	1.2	37.6	0.0	0.0	0.0	35.3	13.2	5.4	
$LASSO$	0.2	99.8	0.0	0.0	0.0	0.5	85.4	0.0	14.1	0.0	0.0	31.4	0.0	68.6	0.0	
EN	0.0	98.1	0.0	1.9	0.0	0.0	38.2	5.8	56.0	0.0	0.0	6.5	59.9	33.6	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.88, 7.89 y 7.90, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados se deterioran conforme aumenta la colinealidad. Por otro lado conforme aumenta la colinealidad se ve afectada negativamente las estimaciones de los coeficientes y las predicciones. También es de notar que cuando el indicador F marco cero, específicamente se refiere esto a C_{HKB} y C_{LW} , MC_{PHKB} y MC_{PLW} es porque el método en todas las simulaciones seleccionó las variables irrelevantes. Se observa además que los métodos de selección AIC_{HKB} , AIC_{LW} y AIC_{DK} independiente del nivel de colinealidad se mantienen más o menos estables en su capacidad para seleccionar variables.

Cuadro 7.88: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$								
		Colinealidad = Baja; P=15; n=50; S=1000								
SR	123750.9			12375.1			1237.5			
σ^2	0.01			0.1			1			
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F	
AIC_{HKB}	0.00001	2.68872	0.92140	0.00025	2.74484	0.91579	0.00586	3.18653	0.65616	
AIC_{LW}	0.00001	2.69587	0.92173	0.00011	2.83060	0.92310	0.00783	4.13825	0.65846	
AIC_{DK}	0.00688	10.79634	0.43046	0.00699	10.36274	0.40162	0.00682	12.62189	0.38622	
C_{HKB}	0.01546	3.08016	0.00000	0.01594	3.17457	0.00000	0.01448	3.94114	0.02912	
C_{LW}	0.01558	2.99811	0.00000	0.01621	2.97836	0.00000	0.01456	4.14924	0.00000	
C_{DK}	0.00688	11.24291	0.39245	0.00700	10.51551	0.37991	0.00682	12.68008	0.37805	
MC_{PHKB}	0.01546	3.08016	0.00000	0.01594	3.17457	0.00000	0.01448	4.39400	0.01800	
MC_{PLW}	0.01558	2.99811	0.00000	0.01621	2.99153	0.00000	0.01456	4.16764	0.00000	
MC_{PDK}	0.01232	0.43548	0.27686	0.01109	1.30405	0.23886	0.01440	2.69409	0.17343	
IC_{95}	0.00004	2.71315	0.59920	0.00062	2.89401	0.66044	0.00923	4.11189	0.59699	
IC_{99}	0.00004	2.71339	0.60715	0.00060	2.89640	0.69191	0.00923	3.86943	0.52163	
$LASSO$	0.00008	2.05399	0.83294	0.00034	1.78374	0.77205	0.00548	3.66506	0.61076	
EN	0.00001	2.24861	0.72429	0.00814	2.71721	0.70727	0.03787	13.06477	0.44414	
R_{HKB}	0.00103	2.66631	-	0.00040	2.55317	-	0.00496	3.21211	-	
R_{LW}	0.00103	2.71255	-	0.00056	2.89634	-	0.01424	4.63518	-	
R_{DK}	0.00283	1.85898	-	0.00291	1.90633	-	0.00548	3.19163	-	
MCO	0.00004	2.71282	-	0.00056	2.89933	-	0.01479	4.68323	-	

Magister en Matemática Mención Estadística

Cuadro 7.89: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Moderada; P=15; n=50; S=1000									
SR	132912.6			13291.3			1329.1		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	1.40143	0.89959	0.00059	1.51334	0.89705	0.00877	2.69130	0.65906
AIC_{LW}	0.00004	1.40162	0.90095	0.00068	1.51409	0.90378	0.01010	2.81160	0.59678
AIC_{DK}	0.00543	8.95258	0.09791	0.00611	10.59103	0.25778	0.00611	12.84597	0.30017
C_{HKB}	0.01647	0.65478	0.00000	0.01752	0.68475	0.00000	0.01673	1.46657	0.00935
C_{LW}	0.01770	1.31724	0.00000	0.01761	1.38776	0.00000	0.01871	1.96509	0.00000
C_{DK}	0.00699	8.89803	0.06276	0.00742	10.93341	0.20953	0.00611	13.04589	0.27660
MCp_{HKB}	0.01647	0.23836	0.00000	0.01688	0.39695	0.00000	0.01673	1.44616	0.00057
MCp_{LW}	0.01770	1.31724	0.00000	0.01761	1.38776	0.00000	0.01871	1.96559	0.00000
MCp_{DK}	0.01737	1.08324	0.00143	0.01623	0.82180	0.02257	0.01614	1.89095	0.05343
IC_{95}	0.00008	1.40978	0.60350	0.00086	1.55245	0.67623	0.01230	2.58889	0.53970
IC_{99}	0.00008	1.40979	0.61333	0.00086	1.54793	0.70997	0.01230	2.32803	0.42928
$LASSO$	0.00011	1.14225	0.81745	0.00050	1.05478	0.78753	0.00786	2.90026	0.61250
EN	0.00034	1.28237	0.72495	0.00568	2.46809	0.66192	0.02857	12.91780	0.40962
R_{HKB}	0.00008	1.41538	-	0.00065	1.58449	-	0.00620	2.77502	-
R_{LW}	0.00008	1.40979	-	0.00089	1.55440	-	0.01086	2.91551	-
R_{DK}	0.00077	1.58545	-	0.00362	1.83280	-	0.00433	3.25254	-
MCO	0.00008	1.40976	-	0.00089	1.55418	-	0.01097	2.91815	-

Magister en Matemática Mención Estadística

Cuadro 7.90: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Alta; P=15; n=50; S=1000									
SR	122048.1			12204.8			1220.5		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00005	5.47785	0.91535	0.00190	5.69068	0.88078	0.00649	6.97816	0.56639
AIC_{LW}	0.00005	5.46765	0.91747	0.00220	5.66213	0.88281	0.00779	7.15283	0.49036
AIC_{DK}	0.00617	29.08670	0.64192	0.00687	30.07065	0.57329	0.00675	33.01178	0.44105
C_{HKB}	0.01955	22.51740	0.00000	0.01980	21.95596	0.00000	0.01801	20.06732	0.00892
C_{LW}	0.02000	22.44663	0.00000	0.02034	21.96597	0.00000	0.01856	19.82152	0.01286
C_{DK}	0.01986	33.53977	0.18990	0.01432	33.13579	0.19226	0.00496	34.42582	0.23425
MC_{PHKB}	0.01955	22.51740	0.00000	0.01980	22.02444	0.00000	0.01801	20.33318	0.00543
MC_{PLW}	0.02000	22.44663	0.00000	0.02034	21.97609	0.00000	0.01856	19.73644	0.01429
MC_{PDK}	0.01986	22.47081	0.00000	0.02019	21.73957	0.00000	0.01833	19.66484	0.01429
IC_{95}	0.00013	5.46035	0.62123	0.00250	5.71835	0.71196	0.00957	12.65921	0.29170
IC_{99}	0.00013	5.46090	0.63801	0.00243	5.75914	0.74876	0.01011	17.20431	0.16218
$LASSO$	0.00009	6.42474	0.76602	0.00239	7.57014	0.75423	0.00688	9.52786	0.58225
EN	0.00030	6.35025	0.66577	0.00487	13.22994	0.53822	0.01217	22.61205	0.33147
R_{HKB}	0.00012	5.51180	-	0.00194	5.76857	-	0.00486	6.69980	-
R_{LW}	0.00012	5.46086	-	0.00244	5.68304	-	0.01004	7.35277	-
R_{DK}	0.00066	6.86092	-	0.00395	9.18970	-	0.00471	11.60824	-
MCO	0.00012	5.46049	-	0.00245	5.68049	-	0.01026	7.40950	-

7.2.2.2.3 100 Observaciones.

Como se aprecia en los cuadros 7.91, 7.92 y 7.93, los resultados mejoran respecto de las tablas mostradas en los cuadros 7.79, 7.80 y 7.81 debido a la presencia de más observaciones. Se logra también apreciar en las tablas como la capacidad de selección disminuye en la medida que aumenta la varianza, también si se comparan los resultados entre tablas para visualizar el efecto del aumento de la colinealidad se aprecia más bien un impacto negativo muy leve. Se puede destacar que además que los métodos C_{HKB} , C_{LW} , MCp_{HKB} , MCp_{LW} y MCp_{DK} tuvieron los resultados más desfavorables para todos los niveles de colinealidad y con varianza baja y moderada y aunque tuvieron mejorías a medida que aumento la varianza son más bien despreciables. Mientras que AIC_{LW} obtuvo el mejor resultado consiguiendo seleccionar el modelo correcto en un 24.7% de las simulaciones cuando la colinealidad es moderada y la varianza es pequeña.

Cuadro 7.91: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Baja; P=15; n=100; S=1000															
SR	246861.8					24686.2					2468.6				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	24.6	75.4	0.0	0.0	0	23.6	76.4	0.0	0.0	0	0.0	97.4	0.1	2.5	0.0
AIC_{LW}	24.7	75.3	0.0	0.0	0	25.8	74.2	0.0	0.0	0	9.2	28.4	12.8	49.6	0.0
AIC_{DK}	0.0	0.0	56.1	43.9	0	0.2	0.0	40.1	59.7	0	0.1	0.2	40.9	58.3	0.5
C_{HKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100	0.0	0.0	15.9	0.0	84.1
C_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100	0.0	0.0	0.2	0.0	99.8
C_{DK}	0.0	0.0	42.1	57.9	0.0	0.3	0.0	39.4	60.3	0.0	0.1	0.2	41.3	57.7	0.7
MCp_{HKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100	0.0	0.0	22.7	0.0	77.3
MCp_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100	0.0	0.0	0.1	0.0	99.9
MCp_{DK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	92.5	0.0	7.5
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	1.0	73.3	0.0	25.7	0.0
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	2.9	57.8	1.8	37.5	0.0
$LASSO$	3.9	96.1	0.0	0.0	0.0	0.5	99.5	0.0	0.0	0.0	0.0	95.9	0.0	4.1	0.0
EN	0.0	100.0	0.0	0.0	0.0	0.0	95.9	3.0	1.1	0.0	0.2	35.9	46.5	16.9	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.92: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = ((1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0))$														
		Colinealidad = Moderada; P=15; n=100; S=1000														
SR	282030.8					28203.1					2820.3					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	27.4	72.6	0.0	0.0	0.0	24.3	75.7	0.0	0.0	0.0	0.1	88.9	0.5	10.5	0.0	
AIC_{LW}	27.6	72.4	0.0	0.0	0.0	27.2	72.8	0.0	0.0	0.0	2.8	8.1	18.4	70.7	0.0	
AIC_{DK}	0.0	0.0	5.2	58.8	36.0	0.1	0.0	6.8	50.4	42.7	0.0	0.1	5.2	60.2	34.5	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.9	0.0	99.1	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	
C_{DK}	0.0	0.0	0.0	57.9	42.1	0.0	0.0	1.2	48.3	50.5	0.0	0.1	3.1	57.8	39.0	
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.2	0.0	99.8	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	
MC_{PDK}	0.0	0.0	13.5	0.0	86.5	0.0	0.0	10.4	0.0	89.6	0.0	0.0	17.2	0.0	82.8	
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.3	37.5	1.3	60.9	0.0	
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	1.3	17.4	8.4	72.9	0.0	
$LASSO$	4.1	95.9	0.0	0.0	0.0	0.5	99.5	0.0	0.0	0.0	0.0	86.5	0.0	13.5	0.0	
EN	0.0	100.0	0.0	0.0	0.0	0.0	91.1	5.3	3.6	0.0	0.0	27.9	44.2	27.8	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.93: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = ((1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0))$														
		Colinealidad = Alta; P=15; n=100; S=1000														
SR	255890.4					25589.0					2558.9					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	24.0	76.0	0	0.0	0.0	20.4	78.3	0.0	1.3	0.0	0.2	38.5	3.4	57.9	0.0	
AIC_{LW}	24.3	75.7	0	0.0	0.0	23.4	75.0	0.1	1.5	0.0	0.2	1.5	18.5	79.8	0.0	
AIC_{DK}	2.0	1.0	17	80.0	0.0	0.4	1.9	22.3	75.4	0.0	0.1	0.4	13.8	85.3	0.4	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
C_{DK}	0.0	0.0	0.0	6.2	93.8	0.0	0.0	0.0	36.3	63.7	0.0	0.2	0.0	48.8	51.0	
MC_{PHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.2	0.0	99.8	
MC_{PLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
MC_{PDK}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	98.4	0.0	1.6	0.0	0.0	0.4	15.7	83.7	0.1	
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.2	96.6	0.0	3.2	0.0	0.0	0.0	38.8	58.4	0.9	
$LASSO$	1.9	98.1	0.0	0.0	0.0	0.8	98.5	0.0	0.7	0.0	0.0	50.2	0.0	49.8	0.0	
EN	0.0	100.0	0.0	0.0	0.0	0.0	81.8	4.1	14.1	0.0	0.0	17.9	41.1	41.0	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.94, 7.95, y 7.96, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados se deterioran conforme aumenta la colinealidad. En concreto conforme aumenta la colinealidad se ve afectada negativamente las estimaciones de los coeficientes y las predicciones. Ahora si se comparan los resultados de la tablas en los cuadros 7.82, 7.83, 7.84, 7.88, 7.89, y 7.90, se puede contemplar una mejoría en las estimaciones debido al incremento de observaciones. También es de notar que cuando el indicador F marco cero, específicamente se refiere esto a C_{HKB} y C_{LW} , MC_{pHKB} y MC_{pLW} es porque el método en todas las simulaciones seleccionó las variables irrelevantes. Se observa además que los métodos de selección AIC_{HKB} , AIC_{LW} y AIC_{DK} independiente del nivel de colinealidad se mantienen más o menos estables en su capacidad para seleccionar variables.

Magister en Matemática Mención Estadística

Cuadro 7.94: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Baja; P=15; n=100; S=1000									
SR	246861.8			24686.2			2468.6		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	0.91055	0.90773	0.00024	1.01011	0.90456	0.00167	1.69362	0.62527
AIC_{LW}	0.00003	0.91034	0.90800	0.00025	1.01193	0.91021	0.00146	2.13855	0.82483
AIC_{DK}	0.00535	19.00139	0.60685	0.00566	19.91603	0.52962	0.00692	24.92382	0.47039
C_{HKB}	0.01590	3.80655	0.00000	0.01590	3.88687	0.00000	0.01584	5.90181	0.04543
C_{LW}	0.01997	8.45487	0.00000	0.01938	8.52538	0.00000	0.02087	8.70401	0.00057
C_{DK}	0.00535	19.83632	0.51473	0.00477	22.15518	0.51594	0.00692	24.97391	0.46977
MCp_{HKB}	0.01590	3.80655	0.00000	0.01590	3.88687	0.00000	0.01584	6.43896	0.06486
MCp_{LW}	0.01997	8.45487	0.00000	0.01938	8.52538	0.00000	0.02087	8.65641	0.00029
MCp_{DK}	0.01315	5.64273	0.28571	0.01270	6.56576	0.28571	0.01272	7.68279	0.26429
IC_{95}	0.00004	0.91090	0.58828	0.00031	1.02428	0.62525	0.00240	2.16153	0.73071
IC_{99}	0.00004	0.91087	0.59328	0.00032	1.02296	0.64482	0.00240	2.12927	0.77026
$LASSO$	0.00003	1.06069	0.83898	0.00016	1.23447	0.76835	0.00200	2.17542	0.62432
EN	0.00005	1.02202	0.74788	0.00014	1.28272	0.73313	0.02566	10.97376	0.50994
R_{HKB}	0.00103	0.90497	–	0.00028	0.97032	–	0.00179	1.82015	–
R_{LW}	0.00103	0.91093	–	0.00031	1.02555	–	0.00296	2.19370	–
R_{DK}	0.00343	1.40286	–	0.00296	2.64604	–	0.00330	5.15518	–
MCO	0.00004	0.91095	–	0.00031	1.02573	–	0.00297	2.19628	–

Magister en Matemática Mención Estadística

Cuadro 7.95: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$ <i>Colinealidad = Moderada; P=15; n=100; S=1000</i>									
SR	282030.8			28203.1			2820.3		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00004	2.39020	0.91154	0.00013	2.46054	0.90437	0.00118	2.82983	0.65194
AIC_{LW}	0.00004	2.39317	0.91192	0.00014	2.49804	0.91078	0.00137	3.51993	0.76847
AIC_{DK}	0.00698	14.04468	0.24153	0.00535	13.62943	0.21408	0.00687	15.07844	0.22696
C_{HKB}	0.01746	1.69968	0.00000	0.01758	1.73487	0.00000	0.01715	2.66811	0.00257
C_{LW}	0.01749	1.32909	0.00000	0.01761	1.38590	0.00000	0.01722	2.64614	0.00114
C_{DK}	0.00698	14.47690	0.18047	0.00646	14.25209	0.14804	0.00687	15.57003	0.19136
MCp_{HKB}	0.01746	1.69968	0.00000	0.01758	1.78708	0.00000	0.01715	3.45698	0.00057
MCp_{LW}	0.01749	1.36300	0.00000	0.01761	1.40309	0.00000	0.01722	2.67298	0.00114
MCp_{DK}	0.01789	1.40895	0.03857	0.01786	1.35528	0.02971	0.01760	2.29708	0.04914
IC_{95}	0.00005	2.40529	0.59053	0.00026	2.54523	0.63563	0.00171	3.71788	0.71701
IC_{99}	0.00005	2.40521	0.59698	0.00027	2.54509	0.65753	0.00062	3.59268	0.73047
$LASSO$	0.00003	2.14264	0.82147	0.00015	1.94116	0.78444	0.00115	3.35356	0.62558
EN	0.00003	2.21760	0.74591	0.00198	3.67133	0.71792	0.00397	22.19159	0.49029
R_{HKB}	0.00004	2.39041	-	0.00024	2.40790	-	0.00145	3.01175	-
R_{LW}	0.00005	2.40526	-	0.00026	2.54838	-	0.00194	3.89666	-
R_{DK}	0.00004	1.99920	-	0.00059	1.78115	-	0.00217	2.90501	-
MCO	0.00005	2.40530	-	0.00026	2.54877	-	0.00194	3.90201	-

Magister en Matemática Mención Estadística

Cuadro 7.96: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)$ <i>Colinealidad = Alta; P=15; n=100; S=1000</i>									
SR	255890.4			25589.0			2558.9		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00010	2.11703	0.90763	0.00038	2.17886	0.89755	0.00237	3.08157	0.66839
AIC_{LW}	0.00011	2.11897	0.90842	0.00038	2.19982	0.90675	0.00320	3.36839	0.65883
AIC_{DK}	0.00700	20.56788	0.63313	0.00370	22.32662	0.65856	0.00699	28.15300	0.52940
C_{HKB}	0.02273	5.65736	0.00000	0.02215	5.67930	0.00000	0.02317	6.28374	0.00114
C_{LW}	0.02301	5.64301	0.00000	0.02244	5.68761	0.00000	0.02364	6.38117	0.00000
C_{DK}	0.02299	7.87809	0.01542	0.02241	19.23837	0.09694	0.00662	23.55354	0.14850
MC_{PHKB}	0.02273	5.65736	0.00000	0.02215	5.69457	0.00000	0.02317	6.35568	0.00057
MC_{PLW}	0.02301	5.64301	0.00000	0.02244	5.68761	0.00000	0.02364	6.39833	0.00000
MC_{PDK}	0.02299	5.64429	0.00000	0.02241	5.67073	0.00000	0.02360	6.21661	0.00114
IC_{95}	0.00015	2.12069	0.60290	0.00044	2.20394	0.67259	0.00598	3.23359	0.57484
IC_{99}	0.00015	2.12056	0.61220	0.00044	2.20390	0.70686	0.01034	3.37334	0.50088
$LASSO$	0.00003	2.31582	0.80024	0.00019	2.71263	0.77576	0.00264	4.29105	0.60950
EN	0.00009	2.17438	0.69676	0.00039	3.09555	0.67399	0.02361	12.66856	0.46106
R_{HKB}	0.00014	2.10884	–	0.00041	2.11741	–	0.00241	3.06791	–
R_{LW}	0.00015	2.12083	–	0.00044	2.20400	–	0.00416	3.40016	–
R_{DK}	0.00012	2.23939	–	0.00090	3.35189	–	0.00356	7.35157	–
MCO	0.00015	2.12086	–	0.00044	2.20436	–	0.00419	3.40580	–

Magister en Matemática Mención Estadística

7.2.2.3. 20 Variables Disponibles.

7.2.2.3.1 25 Observaciones.

Como se aprecia en los cuadros 7.97, 7.98 y 7.99, los resultados empeoran respecto de cuando se disponían 10 y 15 variables puesto que el aumento de variables irrelevantes entorpece la selección de variables. Se logra también apreciar en las tablas como la capacidad de selección disminuye en la medida que aumenta la varianza, también si se comparan los resultados entre tablas para visualizar el efecto del aumento de la colinealidad se aprecia más bien un impacto negativo muy leve. Se puede destacar que además que los métodos C_{HKB} , C_{LW} , MCp_{HKB} , MCp_{LW} y MCp_{DK} tuvieron lo resultados más desfavorables para todos los niveles de colinealidad con varianza baja y moderada y aunque tuvieron mejorías a medida que aumento la varianza son más bien despreciables. Mientras que AIC_{HKB} obtuvo el mejor resultado consiguiendo seleccionar el modelo correcto en un 5 % de las simulaciones cuando la colinealidad es baja y la varianza es moderada.

Cuadro 7.97: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Baja; P=20; n=25; S=1000														
SR	76641.6					7664.2					766.4					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	4.3	95.7	0.0	0.0	0.0	5.0	50.8	1.0	43.2	0.0	0	20.2	0.1	79.3	0.4	
AIC_{LW}	4.6	94.2	0.0	1.2	0.0	5.3	28.2	2.2	64.3	0.0	0	0.0	11.2	82.5	6.3	
AIC_{DK}	0.0	0.4	0.2	99.4	0.0	0.0	4.2	1.2	93.7	0.9	0	13.8	1.5	83.4	1.3	
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.5	9.8	88.7	0.0	0.7	2.1	95.1	2.1	
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	0.0	0.0	15.4	0.0	84.6	
C_{DK}	0.0	0.4	0.2	99.4	0.0	0.0	2.1	1.6	96.0	0.3	0.0	7.7	1.7	88.7	1.9	
MCp_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	6.8	0.0	93.2	0.0	0.0	47.7	0.0	52.3	
MCp_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6	0.0	0.0	15.3	0.0	84.7	
MCp_{DK}	0.0	0.0	99.8	0.0	0.2	0.0	0.0	91.2	0.1	8.7	0.0	0.0	60.5	1.4	38.1	
IC_{95}	0.0	99.6	0.0	0.4	0.0	0.1	32.8	7.2	59.9	0.0	0.0	2.0	14.3	41.5	6.7	
IC_{99}	0.1	99.3	0.0	0.6	0.0	0.1	24.7	14.5	60.4	0.1	0.0	1.2	13.7	28.7	6.3	
$LASSO$	0.3	99.7	0.0	0.0	0.0	0.1	96.3	0.0	3.6	0.0	0.0	58.3	0.0	41.7	0.0	
EN	0.0	69.7	11.4	18.9	0.0	0.0	24.1	60.5	15.4	0.0	0.0	4.3	82.0	10.3	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.98: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Moderada; P=20; n=25; S=1000															
SR	65985.6					6598.6					659.9				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	29.6	70.4	0.0	0.0	0.0	10.6	38.3	2.6	48.5	0.0	0	13.0	0.0	86.8	0.2
AIC_{LW}	32.1	67.9	0.0	0.0	0.0	10.6	22.4	11.4	55.6	0.0	0	0.1	12.0	85.7	2.2
AIC_{DK}	0.0	0.1	0.0	94.2	5.7	0.0	1.3	0.0	93.9	4.8	0	10.0	0.8	85.2	4.0
C_{HKB}	0.0	0.0	0.0	98.4	1.6	0.0	0.0	0.0	83.5	16.5	0.0	0.0	0.0	84.8	15.2
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	6.6	0.0	93.4
C_{DK}	0.0	0.1	0.0	92.5	7.4	0.0	0.7	0.0	93.4	5.9	0.0	4.9	0.8	89.2	5.1
MC_{pHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	13.2	0.0	86.8
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	8.3	0.0	91.7
MC_{pDK}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	5.2	0.0	94.8	0.0	0.0	20.2	0.0	79.8
IC_{95}	0.1	99.6	0.0	0.3	0.0	0.0	28.6	9.3	62.0	0.0	0.0	2.1	13.5	37.3	8.4
IC_{99}	0.5	98.8	0.1	0.6	0.0	0.0	19.1	16.8	63.1	0.0	0.0	1.7	11.2	23.8	6.4
LASSO	1.0	99.0	0.0	0.0	0.0	0.2	86.5	0.0	13.3	0.0	0.0	49.4	0.0	50.6	0.0
EN	0.0	73.0	0.2	26.8	0.0	0.0	27.6	33.3	39.1	0.0	0.0	3.6	79.9	15.8	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.99: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Alta; P=20; n=25; S=1000															
SR	44918.2					4491.8					6449.2				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	26.9	71.6	0.2	1.3	0.0	0.9	4.1	8.3	86.7	0.0	0	0.6	0.5	97.3	1.6
AIC_{LW}	32.1	65.9	0.2	1.8	0.0	0.6	1.5	17.4	80.5	0.0	0	0.0	11.8	80.7	7.5
AIC_{DK}	0.0	0.2	1.7	98.1	0.0	0.0	6.8	3.7	89.3	0.2	0	13.6	2.3	79.8	4.3
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	3.6	96.4	0.0	0.0	0.0	51.0	49.0
C_{LW}	0.0	0.0	0.5	0.0	99.5	0.0	0.0	12.9	0.0	87.1	0.0	0.0	20.4	0.0	79.6
C_{DK}	0.0	0.3	0.0	87.9	11.8	0.0	2.3	0.0	82.7	15.0	0.0	5.8	0.4	84.2	9.6
MC_{pHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.9	0.0	98.1	0.0	0.0	15.9	0.0	84.1
MC_{pLW}	0.0	0.0	8.2	0.0	91.8	0.0	0.0	24.7	0.0	75.3	0.0	0.0	18.5	0.0	81.5
MC_{pDK}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.3	99.7	0.0	0.0	8.1	6.0	85.9
IC_{95}	0.2	95.5	0.8	3.5	0.0	0.0	7.3	16.2	71.3	1.1	0.0	0.8	8.9	28.9	12.7
IC_{99}	1.5	90.5	1.6	6.4	0.0	0.0	4.5	19.5	59.6	2.8	0.0	0.5	5.4	18.8	9.6
LASSO	0.2	88.6	0.0	11.2	0.0	0.0	24.1	0.0	75.9	0.0	0.0	39.0	0.0	60.9	0.1
EN	0.0	34.6	1.3	64.1	0.0	0.0	8.4	42.4	49.2	0.0	0.0	3.2	80.5	15.8	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Respecto de los cuadros 7.100, 7.101,7.102, se aprecia en cada uno de ellos que los resultados tanto para el ECMT, ECMP y F empeoran conforme aumenta la varianza. Por otra parte si se comparan las tablas desde una perspectiva global se puede ver que los resultados se deterioran conforme aumenta la colinealidad. Por otro lado la colinealidad afecta negativamente las estimaciones de los coeficientes, las predicciones y el indicador F. También es de notar que cuando el indicador F marco cero, específicamente se refiere esto a C_{LW} , MCp_{HKB} , MCp_{LW} y MCp_{DK} es porque el método en todas las simulaciones seleccionó las variables irrelevantes.

Cuadro 7.100: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$								
		Colinealidad = Baja; P=20; n=25; S=1000								
SR	76641.6			7664.2			766.4			
σ^2	0.01			0.1			1			
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F	
AIC_{HKB}	0.00022	0.83684	0.83614	0.00123	0.90870	0.78438	0.00257	1.59896	0.50788	
AIC_{LW}	0.00018	0.86832	0.83238	0.00389	1.05192	0.73193	0.00583	2.69792	0.38427	
AIC_{DK}	0.00697	30.19688	0.41998	0.00700	30.09682	0.42507	0.00698	31.39745	0.42522	
C_{HKB}	0.01766	0.87260	0.00000	0.01839	0.89342	0.02332	0.00676	1.33945	0.34391	
C_{LW}	0.01837	0.73210	0.00000	0.02041	0.84307	0.00114	0.01706	1.72288	0.04400	
C_{DK}	0.00697	30.61442	0.42249	0.00700	30.54167	0.42870	0.00698	32.08220	0.41538	
MCp_{HKB}	0.01766	0.87260	0.00000	0.01839	0.99421	0.01943	0.01181	1.88727	0.13629	
MCp_{LW}	0.01837	0.73210	0.00000	0.02041	0.84307	0.00114	0.01706	1.74220	0.04371	
MCp_{DK}	0.01406	0.53443	0.28514	0.01222	0.61244	0.26082	0.01638	1.70870	0.17700	
IC_{95}	0.00145	0.95604	0.53314	0.00975	1.44537	0.59137	0.00844	2.51296	0.22887	
IC_{99}	0.00145	0.95442	0.55800	0.00766	1.29493	0.60367	0.01359	2.13265	0.16631	
$LASSO$	0.00023	0.44234	0.77412	0.00118	0.51293	0.68030	0.03198	9.13751	0.46737	
EN	0.00023	1.08243	0.68482	0.01234	2.99632	0.43162	0.00624	16.43610	0.32315	
R_{HKB}	0.00201	0.96424	-	0.00279	1.27310	-	0.00762	3.86641	-	
R_{LW}	0.00232	0.96552	-	0.01817	1.83366	-	0.03590	7.34177	-	
R_{DK}	0.00208	1.62447	-	0.00388	3.99931	-	0.00393	7.62663	-	
MCO	0.00145	0.96629	-	0.02235	1.93029	-	0.07630	12.14726	-	

Magister en Matemática Mención Estadística

Cuadro 7.101: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Moderada; P=20; n=25; S=1000									
SR	65985.6			6598.6			659.9		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00023	1.43980	0.91381	0.00417	1.54993	0.81933	0.00465	2.26490	0.51595
AIC_{LW}	0.00027	1.44378	0.91808	0.00555	1.61032	0.81058	0.01837	3.22189	0.41259
AIC_{DK}	0.00629	14.26177	0.30578	0.00687	14.40879	0.33162	0.00678	15.36476	0.36789
C_{HKB}	0.00864	1.23702	0.14305	0.00790	1.17590	0.12077	0.00856	1.47589	0.16718
C_{LW}	0.01794	2.14852	0.00000	0.01650	2.16726	0.00000	0.01782	2.34679	0.01886
C_{DK}	0.00669	14.40557	0.29090	0.00687	14.60588	0.31864	0.00699	15.76285	0.35050
MCp_{HKB}	0.01707	3.36008	0.00000	0.01586	2.93773	0.00000	0.01181	2.80165	0.03771
MCp_{LW}	0.01794	4.31677	0.00000	0.01650	3.66270	0.00000	0.01751	3.64365	0.02371
MCp_{DK}	0.01515	1.75932	0.00000	0.01503	1.40722	0.01486	0.01144	1.85487	0.05771
IC_{95}	0.00101	1.46196	0.54869	0.00678	1.57831	0.58432	0.01829	2.07328	0.19573
IC_{99}	0.00101	1.45725	0.57819	0.00550	1.42977	0.57698	0.01942	1.69193	0.13446
$LASSO$	0.00013	0.97467	0.77818	0.00125	0.92639	0.68922	0.01945	7.75174	0.46783
EN	0.00034	0.71249	0.69435	0.00629	3.25739	0.46615	0.00731	8.69053	0.31686
R_{HKB}	0.00062	1.58587	–	0.00197	1.92910	–	0.01858	4.03833	–
R_{LW}	0.00152	1.45788	–	0.00564	1.93225	–	0.04216	5.34427	–
R_{DK}	0.00295	1.68504	–	0.00416	2.28832	–	0.00467	4.11370	–
MCO	0.00157	1.45579	–	0.00602	1.95189	–	0.04879	7.13068	–

Magister en Matemática Mención Estadística

Cuadro 7.102: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Alta; P=20; n=25; S=1000									
SR	44918.2			4491.8			6449.2		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00019	2.69726	0.90828	0.00261	2.78037	0.69555	0.00853	3.76302	0.42831
AIC_{LW}	0.00031	2.69776	0.91547	0.00167	2.85002	0.67995	0.01350	4.89532	0.33814
AIC_{DK}	0.00550	14.27532	0.47869	0.00699	14.42691	0.47830	0.00698	15.72310	0.39516
C_{HKB}	0.01475	4.13444	0.00000	0.01270	3.81789	0.00574	0.00863	4.33237	0.11558
C_{LW}	0.01734	3.99105	0.00143	0.01725	4.41099	0.03686	0.01020	5.38693	0.05829
C_{DK}	0.00686	14.76659	0.26560	0.00695	14.83521	0.27469	0.00659	15.99183	0.32320
MC_{PHKB}	0.01477	4.76070	0.00000	0.01486	4.95946	0.00543	0.00933	6.78616	0.04543
MC_{PLW}	0.01734	4.51440	0.02343	0.01725	4.63318	0.07057	0.01020	4.10585	0.05286
MC_{PDK}	0.01202	5.34845	0.00000	0.01493	5.16721	0.00075	0.00942	6.56830	0.03895
IC_{95}	0.00102	2.71554	0.59734	0.01113	3.26150	0.43862	0.01070	7.04474	0.13176
IC_{99}	0.00102	2.71328	0.63332	0.01113	3.54347	0.37185	0.01287	7.46473	0.08362
$LASSO$	0.00029	3.72800	0.73060	0.00331	3.50749	0.61016	0.01619	8.30344	0.45185
EN	0.00229	4.34781	0.60570	0.00835	7.46212	0.40559	0.00638	10.53852	0.30525
R_{HKB}	0.00065	2.64523	-	0.00784	2.97901	-	0.02765	6.44127	-
R_{LW}	0.00586	2.72476	-	0.01461	3.42584	-	0.04004	7.07540	-
R_{DK}	0.00393	2.74983	-	0.00462	3.95973	-	0.00487	6.30627	-
MCO	0.00719	2.72761	-	0.01653	3.52355	-	0.20833	10.51678	-

7.2.2.3.2 50 Observaciones.

Se aprecia en la tabla del cuadro 7.103 que seleccionar el modelo correcto es poco frecuente el mejor resultado lo alcanzo AIC_{LW} con un 20.5% de la simulaciones cuando la varianza es pequeña y la colinealidad es baja. En el cuadro 7.104 el mejor resultado lo alcanzo AIC_{LW} seleccionando el modelo correcto solo un 19.9% de las simulaciones y en el 74.8% de las simulaciones seleccionó modelos que contenían el modelo correcto. Por su parte en el cuadro 7.105 el método que seleccionó más veces el modelo correcto fue AIC_{LW} con un 19.9% de las simulaciones cuando la varianza es Baja. Por su parte LASSO seleccionó el modelo correcto 0.1% de las simulaciones y en el resto de las simulaciones selecciono modelos que contenían el modelo correcto cuando la varianza es baja.

Cuadro 7.103: INDICADORES DE SELECCIÓN DE MODELOS

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$														
		Colinealidad = Baja; P=20; n=50; S=1000														
SR	144565.1					14456.5					1445.7					
σ^2	0.01					0.1					1					
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	
AIC_{HKB}	19.8	80.2	0.0	0.0	0	14.9	84.8	0.0	0.3	0.0	0.1	56.5	0.7	42.7	0.0	
AIC_{LW}	20.5	79.5	0.0	0.0	0	18.3	80.8	0.0	0.9	0.0	0.2	0.8	12.8	86.1	0.1	
AIC_{DK}	0.0	0.0	49.2	50.8	0	0.0	0.4	25.3	74.3	0.0	0.0	1.2	14.6	80.8	3.4	
C_{HKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	6.9	0.0	93.1	0.0	0.0	42.3	1.5	56.2	
C_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100.0	0.0	0.0	5.6	0.0	94.4	
C_{DK}	0.0	0.0	22.1	77.9	0.0	0.0	0.1	17.3	82.6	0.0	0.0	0.8	13.9	81.8	3.5	
MC_{pHKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	15.3	0.0	84.7	0.0	0.0	48.6	0.0	51.4	
MC_{pLW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100.0	0.0	0.0	5.5	0.0	94.5	
MC_{pDK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	0.0	0.1	0.0	0.0	76.1	0.0	23.9	
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	99.7	0.0	0.3	0.0	0.0	1.5	16.2	79.0	1.9	
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.0	99.4	0.0	0.6	0.0	0.0	0.1	33.3	53.4	3.6	
LASSO	1.4	98.6	0.0	0.0	0.0	0.2	99.8	0.0	0.0	0.0	0.0	65.7	0.0	34.3	0.0	
EN	0.0	100.0	0.0	0.0	0.0	0.1	70.8	11.1	18.0	0.0	0.0	12.9	71.8	13.6	0.0	

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.104: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Moderada; P=20; n=50; S=1000															
SR	1129176.6					12917.7					1291.8				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	18.7	81.3	0.0	0.0	0.0	15.9	79.5	0.1	4.5	0.0	0	39.1	0.9	60.0	0.0
AIC_{LW}	18.9	81.1	0.0	0.0	0.0	19.9	74.8	0.1	5.2	0.0	0	0.0	13.6	86.4	0.0
AIC_{DK}	0.0	0.0	0.3	72.2	27.5	0.0	0.0	0.5	74.4	25.1	0	0.1	3.2	73.8	22.9
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.8	18.6	79.6
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.7	0.0	99.3	0.0	0.0	15.4	0.0	84.6
C_{DK}	0.0	0.0	0.0	70.9	29.1	0.0	0.0	0.0	72.2	27.8	0.0	0.1	2.3	71.5	26.1
MC_{pHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	5.3	0.0	94.7
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.7	0.0	99.3	0.0	0.0	15.2	0.0	84.8
MC_{pDK}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.4	0.0	98.6	0.0	0.0	11.7	0.0	88.3
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	98.0	0.0	2.0	0.0	0.0	0.5	21.5	69.1	3.9
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.0	96.8	0.0	3.2	0.0	0.0	0.0	31.1	40.0	5.7
$LASSO$	0.9	99.1	0.0	0.0	0.0	0.3	99.5	0.0	0.2	0.0	0.0	55.2	0.0	44.8	0.0
EN	0.0	99.9	0.0	0.1	0.0	0.0	63.0	2.9	34.1	0.0	0.0	10.6	68.5	20.7	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.105: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Alta; P=20; n=50; S=1000															
SR	108516.4					10851.6					1085.2				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	18.5	81.5	0.0	0.0	0.0	5.3	42.6	3.7	48.4	0.0	0	3.0	3.5	92.6	0.9
AIC_{LW}	19.9	80.1	0.0	0.0	0.0	7.3	33.9	6.4	52.4	0.0	0	0.0	12.9	84.5	2.6
AIC_{DK}	0.1	0.8	6.9	92.2	0.0	0.0	1.1	10.5	88.3	0.1	0	1.5	6.8	89.7	2.0
C_{HKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100	0.0	0.0	1.2	2.0	96.8
C_{LW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.8	0.0	99.2
C_{DK}	0.0	0.0	0.0	15.8	84.2	0.0	0.0	0.0	42.0	58	0.0	0.1	0.0	56.2	43.7
MC_{pHKB}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100	0.0	0.0	1.6	0.0	98.4
MC_{pLW}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.7	0.0	99.3
MC_{pDK}	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100	0.0	0.0	1.4	0.1	98.5
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	50.9	0.5	48.6	0.0	0.0	0.0	19.7	26.7	14.7
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.4	35.6	2.1	61.9	0.0	0.0	0.0	12.6	8.2	9.6
$LASSO$	0.1	99.9	0.0	0.0	0.0	0.0	79.3	0.0	20.7	0.0	0.0	29.4	0.0	70.6	0.0
EN	0.0	85.1	0.0	14.9	0.0	0.0	29.7	8.1	62.2	0.0	0.0	5.8	65.9	28.3	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Se aprecia en la tabla del cuadro 7.106 que en cuanto selección de variables los resultados fueron bastante competitivos entre AIC_{HKB} , AIC_{LW} , $LASSO$, $ElasticNet$ cuando la varianza es baja y moderada. En esta misma tabla el peor desempeño lo mostraron C_{HKB} , C_{LW} , MCp_{HKB} , MCp_{LW} para varianzas pequeñas y moderadas. En la tabla del cuadro 7.107 aunque numéricamente los valores mostraron un peor desempeño producto de que aumento la colinealidad, lo resultados no varían demasiado de los resultados del cuadro 7.106. En el cuadro 7.108 se muestran los resultados para una colinealidad alta, aunque desde el punto de vista del indicador F los resultados no varían demasiado por lo que se ve una estabilidad en la capacidad de seleccionar, pero si se ven resultados más desfavorables en las predicciones y en la estimación de los coeficientes.

Cuadro 7.106: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$								
		Colinealidad = Baja; P=20; n=50; S=1000								
SR	144565.1			14456.5			1445.7			
σ^2	0.01			0.1			1			
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F	
AIC_{HKB}	0.00001	5.72352	0.88796	0.00018	5.76291	0.87377	0.00490	6.21769	0.57631	
AIC_{LW}	0.00001	5.72681	0.88957	0.00024	5.81968	0.88440	0.00713	6.35614	0.58882	
AIC_{DK}	0.00493	9.12508	0.53189	0.00699	10.88044	0.48008	0.00603	12.85398	0.38057	
C_{HKB}	0.01866	3.14919	0.00000	0.01911	3.23815	0.01971	0.01619	3.92479	0.12448	
C_{LW}	0.01888	3.06042	0.00000	0.01944	3.13870	0.00000	0.01755	3.79681	0.01600	
C_{DK}	0.00493	10.47591	0.46513	0.00699	11.11932	0.44848	0.00603	12.95602	0.37784	
MCp_{HKB}	0.01866	3.14919	0.00000	0.01911	3.23234	0.04371	0.01725	4.19604	0.13886	
MCp_{LW}	0.01888	3.06042	0.00000	0.01944	3.14219	0.00000	0.01755	3.80863	0.01571	
MCp_{DK}	0.01408	5.22529	0.28571	0.01405	4.10286	0.28543	0.01284	4.18068	0.21743	
IC_{95}	0.00007	5.72382	0.49318	0.00106	5.86050	0.57784	0.01310	5.66179	0.48432	
IC_{99}	0.00007	5.72449	0.50449	0.00091	5.86408	0.62107	0.01000	4.57149	0.39121	
$LASSO$	0.00004	4.96553	0.77633	0.00036	4.19214	0.71157	0.00805	6.93963	0.49531	
EN	0.00035	5.03793	0.72942	0.00372	5.52389	0.68438	0.02382	24.09845	0.38795	
R_{HKB}	0.00106	5.66562	-	0.00089	5.50563	-	0.00678	6.39018	-	
R_{LW}	0.00106	5.72350	-	0.00110	5.85727	-	0.02059	7.54264	-	
R_{DK}	0.00240	4.78491	-	0.00398	4.36192	-	0.00481	4.57243	-	
MCO	0.00007	5.72381	-	0.00110	5.86054	-	0.02153	7.59844	-	

Magister en Matemática Mención Estadística

Cuadro 7.107: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Moderada; P=20; n=50; S=1000									
SR	129176.6			12917.7			1291.8		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00003	1.05139	0.88475	0.00076	1.17438	0.87155	0.00761	2.41451	0.57271
AIC_{LW}	0.00004	1.05259	0.88552	0.00076	1.18995	0.88268	0.01098	2.72503	0.56283
AIC_{DK}	0.00697	12.34419	0.21282	0.00654	12.68871	0.22663	0.00697	13.64240	0.24661
C_{HKB}	0.01670	2.01654	0.00000	0.01662	3.09717	0.00000	0.01624	4.06477	0.03356
C_{LW}	0.01748	6.91040	0.00000	0.01744	6.28766	0.00200	0.01637	6.64022	0.04400
C_{DK}	0.00697	12.34343	0.20714	0.00680	12.77008	0.21218	0.00697	13.72988	0.22223
MCp_{HKB}	0.01676	6.46128	0.00000	0.01730	5.17163	0.00000	0.01624	5.03654	0.01514
MCp_{LW}	0.01748	6.91040	0.00000	0.01744	6.35893	0.00200	0.01637	6.66858	0.04343
MCp_{DK}	0.01689	4.69454	0.00000	0.01679	3.52596	0.00400	0.01678	4.46136	0.03343
IC_{95}	0.00015	1.05304	0.49908	0.00150	1.19630	0.59509	0.01471	4.26454	0.38595
IC_{99}	0.00015	1.05282	0.51152	0.00147	1.19620	0.64239	0.01748	5.17079	0.27256
$LASSO$	0.00004	0.76945	0.77352	0.00081	0.67704	0.72711	0.00756	3.03877	0.49253
EN	0.00094	0.79302	0.69571	0.00743	1.92928	0.64090	0.02039	19.49392	0.38672
R_{HKB}	0.00014	1.06321	-	0.00085	1.24056	-	0.00659	2.48501	-
R_{LW}	0.00015	1.05295	-	0.00170	1.19718	-	0.01305	2.74705	-
R_{DK}	0.00087	1.32650	-	0.00508	1.66058	-	0.00448	3.14514	-
MCO	0.00015	1.05289	-	0.00172	1.19673	-	0.01326	2.75753	-

Magister en Matemática Mención Estadística

Cuadro 7.108: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

	$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$								
	Colinealidad = Alta; P=20; n=50; S=1000								
SR	108516.4			10851.6			1085.2		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00023	9.39840	0.88451	0.00374	9.31676	0.80126	0.00664	10.45818	0.47249
AIC_{LW}	0.00025	9.43201	0.88826	0.00410	9.62984	0.80657	0.00757	12.60224	0.42075
AIC_{DK}	0.00672	2.08774	0.58743	0.00326	2.20744	0.54830	0.00320	3.53606	0.42319
C_{HKB}	0.01929	4.52745	0.00000	0.01863	5.59287	0.00000	0.01129	6.08716	0.00670
C_{LW}	0.02177	4.80862	0.00000	0.02238	4.44035	0.00000	0.01964	4.73894	0.00229
C_{DK}	0.01974	3.83695	0.03571	0.01910	3.23968	0.10860	0.00698	4.02279	0.15988
MCp_{HKB}	0.02088	4.86194	0.00000	0.01863	5.36098	0.00000	0.01129	5.90854	0.00457
MCp_{LW}	0.02177	4.80199	0.00000	0.02238	4.43471	0.00000	0.01964	4.61577	0.00200
MCp_{DK}	0.01974	4.54927	0.00000	0.01910	5.97483	0.00000	0.01161	6.48276	0.00425
IC_{95}	0.00063	9.41483	0.51727	0.00306	9.68820	0.61715	0.01008	8.56145	0.14993
IC_{99}	0.00063	9.41532	0.53830	0.00292	9.47778	0.65864	0.01872	6.54580	0.06234
$LASSO$	0.00011	8.06002	0.71140	0.00219	6.75896	0.69058	0.00796	10.70622	0.46084
EN	0.00052	7.31613	0.60360	0.00219	4.64191	0.50235	0.02209	12.51064	0.32114
R_{HKB}	0.00046	9.31144	-	0.00248	9.18271	-	0.00667	10.17905	-
R_{LW}	0.00063	9.41541	-	0.00316	9.83383	-	0.02034	12.55692	-
R_{DK}	0.00092	5.78092	-	0.00396	4.31202	-	0.00452	4.02532	-
MCO	0.00063	9.41619	-	0.00317	9.84330	-	0.02222	12.81222	-

7.2.2.3.3 100 Observaciones.

Se aprecia en la tabla del cuadro 7.109 que seleccionar el modelo correcto es poco frecuente el mejor resultado lo alcanzo AIC_{LW} con un 13.1% de la simulaciones cuando la varianza es pequeña y la colinealidad es baja. En el cuadro 7.110 el mejor resultado lo alcanzo AIC_{LW} seleccionando el modelo correcto solo un 14.9% de las simulaciones cuando la varianza es moderada y en el 85.8% de las simulaciones seleccionó modelos que contenían el modelo correcto. En el cuadro 7.111 el método que seleccionó más veces el modelo correcto fue AIC_{LW} con un 14.2% de las simulaciones cuando la varianza es pequeña, por su parte LASSO seleccionó el modelo correcto 0.4% de las simulaciones y en el resto de las simulaciones seleccionó modelos que contenían el modelo correcto cuando la varianza es baja.

Cuadro 7.109: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ Colinealidad = Baja; P=20; n=100; S=1000															
SR	268775.9					26877.6					2687.8				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
AIC_{HKB}	12.8	87.2	0.0	0.0	0	11.9	88.1	0.0	0.0	0.0	0.0	97.4	0.0	2.6	0.0
AIC_{LW}	13.1	86.9	0.0	0.0	0	13.6	86.4	0.0	0.0	0.0	3.6	27.2	7.3	61.9	0.0
AIC_{DK}	0.0	0.0	30.8	69.2	0	0.0	0.0	34.0	65.9	0.1	0.0	0.1	21.4	76.5	2.0
C_{HKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100.0	0.0	0.0	20.2	0.0	79.8
C_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.3	0.0	98.7
C_{DK}	0.0	0.0	38.2	61.8	0.0	0.1	0.0	31.7	68.2	0.0	0.0	0.2	22.7	75.1	2.0
MCp_{HKB}	0.0	0.0	0.0	0.0	100	0.0	0.0	1.3	0.0	98.7	0.0	0.0	52.3	0.0	47.7
MCp_{LW}	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.3	0.0	98.7
MCp_{DK}	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	91.3	0.0	8.7
IC_{95}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.2	61.7	0.1	38.0	0.0
IC_{99}	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.7	45.0	2.0	52.3	0.0
LASSO	1.4	98.6	0.0	0.0	0.0	0.1	99.9	0.0	0.0	0.0	0.0	95.2	0.0	4.8	0.0
EN	0.0	100.0	0.0	0.0	0.0	0.1	93.3	4.0	2.6	0.0	0.2	31.6	55.0	12.5	0.0

El conteo falla para IC_{95} , IC_{99} y EN ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Cuadro 7.110: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Moderada; P=20; n=100; S=1000															
SR	270225.5					27022.5					2702.3				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
<i>AIC_{HKB}</i>	14.4	85.6	0.0	0.0	0.0	13.2	86.8	0.0	0.0	0.0	0.0	88.5	0.0	11.5	0.0
<i>AIC_{LW}</i>	14.7	85.3	0.0	0.0	0.0	14.9	85.1	0.0	0.0	0.0	1.2	7.0	8.6	83.2	0.0
<i>AIC_{DK}</i>	0.0	0.0	0.1	51.1	48.8	0.0	0.0	0.6	49.5	49.9	0.0	0.0	1.9	58.0	40.1
<i>C_{HKB}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1.8	0.0	98.2
<i>C_{LW}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.4	0.0	97.6
<i>C_{DK}</i>	0.0	0.0	0.0	48.7	51.3	0.0	0.0	0.0	46.4	53.6	0.0	0.0	1.0	56.8	42.2
<i>MC_{pHKB}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	4.8	0.0	95.2
<i>MC_{pLW}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	2.4	0.0	97.6
<i>MC_{pDK}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.3	0.0	99.7	0.0	0.0	8.6	0.0	91.4
<i>IC₉₅</i>	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	23.2	0.4	76.4	0.0
<i>IC₉₉</i>	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.2	9.9	6.0	83.9	0.0
<i>LASSO</i>	1.2	98.8	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	85.1	0.0	14.9	0.0
<i>EN</i>	0.0	100.0	0.0	0.0	0.0	0.0	91.4	4.0	4.6	0.0	0.0	22.8	49.7	27.3	0.0

El conteo falla para *IC₉₅*, *IC₉₉* y *EN* ya que no se contabiliza el caso en que se eliminan todas las variables.

Cuadro 7.111: INDICADORES DE SELECCIÓN DE MODELOS

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$															
Colinealidad = Alta; P=20; n=100; S=1000															
SR	249702.9					24970.3					2497.0				
σ^2	0.01					0.1					1				
Categoría	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa	MB	B	Re	Ma	MMa
<i>AIC_{HKB}</i>	13.7	86.3	0	0.0	0.0	7.9	78.5	1.2	12.4	0.0	0	20.2	1.8	78.0	0.0
<i>AIC_{LW}</i>	14.2	85.8	0	0.0	0.0	11.7	71.9	1.6	14.8	0.0	0	0.2	10.4	89.4	0.0
<i>AIC_{DK}</i>	0.4	0.5	20	79.1	0.0	0.2	0.8	17.7	81.3	0.0	0	0.4	10.3	89.2	0.1
<i>C_{HKB}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0
<i>C_{LW}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6
<i>C_{DK}</i>	0.0	0.0	0.0	6.1	93.9	0.0	0.0	0.0	37.7	62.3	0.0	0.0	0.0	47.7	52.3
<i>MC_{pHKB}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	99.9
<i>MC_{pLW}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.4	0.0	99.6
<i>MC_{pDK}</i>	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0
<i>IC₉₅</i>	0.0	100.0	0.0	0.0	0.0	0.0	95.0	0.0	5.0	0.0	0.0	0.0	8.9	86.3	4.0
<i>IC₉₉</i>	0.0	100.0	0.0	0.0	0.0	0.0	91.2	0.0	8.8	0.0	0.0	0.0	24.5	56.7	9.7
<i>LASSO</i>	0.4	99.6	0.0	0.0	0.0	0.2	97.3	0.0	2.5	0.0	0.0	41.0	0.0	59.0	0.0
<i>EN</i>	0.0	99.5	0.0	0.5	0.0	0.0	50.3	6.9	42.8	0.0	0.0	10.3	50.7	39.0	0.0

El conteo falla para *IC₉₅*, *IC₉₉* y *EN* ya que no se contabiliza el caso en que se eliminan todas las variables.

Magister en Matemática Mención Estadística

Se aprecia en la tabla del cuadro 7.112 que en cuanto selección de variables los mejores resultados los obtuvo AIC_{HKB} , AIC_{LW} , $LASSO$, $ElasticNet$ cuando la Varianza es Baja y Moderada. En esta misma tabla el peor desempeño lo mostro C_{HKB} , C_{LW} , MCp_{HKB} , MCp_{LW} para varianzas pequeñas y moderadas. En la tabla del cuadro 7.113 aunque numéricamente los valores mostraron un peor desempeño producto de que aumento la colinealidad, los resultados no varían demasiado de los resultados del cuadro 7.112. En el cuadro 7.114 se muestran los resultados para una colinealidad alta, aunque desde el punto de vista del indicador F los resultados no varían demasiado por lo que se ve una estabilidad en la capacidad de seleccionar, pero si se ven resultados más desfavorables en las predicciones y en la estimación de los coeficientes.

Cuadro 7.112: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

		$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$								
		Colinealidad = Baja; P=20; n=100; S=1000								
SR	268775.9			26877.6			2687.8			
σ^2	0.01			0.1			1			
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F	
AIC_{HKB}	0.00004	1.15399	0.86519	0.00040	1.24530	0.85966	0.00173	1.92589	0.52481	
AIC_{LW}	0.00004	1.15378	0.86588	0.00042	1.24115	0.86910	0.00267	2.27057	0.76761	
AIC_{DK}	0.00675	26.65980	0.54353	0.00610	24.50705	0.51865	0.00664	26.72017	0.41411	
C_{HKB}	0.01776	4.96836	0.00000	0.01742	5.03813	0.00000	0.01393	5.56220	0.05771	
C_{LW}	0.01784	5.01287	0.00000	0.01749	5.08692	0.00000	0.01976	5.90591	0.00371	
C_{DK}	0.00676	25.31557	0.60237	0.00663	25.20320	0.52179	0.00664	26.65946	0.41473	
MCp_{HKB}	0.01776	4.96836	0.00000	0.01742	5.05926	0.00371	0.01393	6.56013	0.14943	
MCp_{LW}	0.01784	5.01287	0.00000	0.01749	5.08692	0.00000	0.01976	5.90393	0.00371	
MCp_{DK}	0.01358	3.29440	0.28571	0.01343	4.42190	0.28571	0.01342	5.53601	0.26086	
IC_{95}	0.00004	1.15429	0.47965	0.00054	1.26253	0.52185	0.00596	2.29451	0.64183	
IC_{99}	0.00004	1.15430	0.48555	0.00054	1.26288	0.54357	0.00596	2.32666	0.69315	
$LASSO$	0.00002	1.32566	0.76984	0.00028	1.56494	0.68695	0.00523	2.46828	0.50964	
EN	0.00021	1.33743	0.71854	0.00059	1.83248	0.70760	0.03973	9.44345	0.46114	
R_{HKB}	0.00104	1.15248	-	0.00050	1.24102	-	0.00357	2.05718	-	
R_{LW}	0.00104	1.15434	-	0.00052	1.26264	-	0.00683	2.29580	-	
R_{DK}	0.00306	2.12908	-	0.00213	3.06791	-	0.00386	6.41249	-	
MCO	0.00004	1.15434	-	0.00052	1.26270	-	0.00687	2.29717	-	

Magister en Matemática Mención Estadística

Cuadro 7.113: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ <i>Colinealidad = Moderada; P=20; n=100; S=1000</i>									
SR	270225.5			27022.5			2702.3		
σ^2	0.01			0.1			1		
Indicador	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00005	1.26385	0.86959	0.00013	1.36968	0.86107	0.00182	2.65835	0.54047
AIC_{LW}	0.00005	1.26313	0.87050	0.00014	1.36326	0.87085	0.00012	2.39818	0.70439
AIC_{DK}	0.00620	32.37400	0.13331	0.00698	33.41830	0.14537	0.00697	33.96512	0.18087
C_{HKB}	0.01779	2.15384	0.00000	0.01786	1.95377	0.00000	0.01679	2.08141	0.00514
C_{LW}	0.01782	2.13316	0.00000	0.01790	2.22682	0.00000	0.01814	2.99898	0.00686
C_{DK}	0.00620	32.72059	0.12601	0.00698	33.84940	0.12749	0.00697	34.19375	0.17003
MCp_{HKB}	0.01779	2.15384	0.00000	0.01786	2.24852	0.00000	0.01807	2.94798	0.01371
MCp_{LW}	0.01782	2.13316	0.00000	0.01790	2.22682	0.00000	0.01814	2.99570	0.00686
MCp_{DK}	0.01760	2.11211	0.00000	0.01743	1.93456	0.00086	0.01707	2.56306	0.02457
IC_{95}	0.00007	1.26002	0.48346	0.00050	1.36147	0.53305	0.00109	2.43015	0.61467
IC_{99}	0.00007	1.26008	0.48962	0.00049	1.36157	0.55781	0.00186	2.41126	0.63449
$LASSO$	0.00004	1.41609	0.75994	0.00020	1.83785	0.70905	0.00136	2.95786	0.50684
EN	0.00007	1.26860	0.68636	0.00209	1.41273	0.66719	0.00812	12.84867	0.45288
R_{HKB}	0.00007	1.26709	–	0.00041	1.42428	–	0.00194	2.60220	–
R_{LW}	0.00007	1.25995	–	0.00049	1.36307	–	0.00203	2.45515	–
R_{DK}	0.00007	2.23780	–	0.00140	3.87804	–	0.00687	7.73406	–
MCO	0.00007	1.25993	–	0.00049	1.36289	–	0.00203	2.45461	–

Magister en Matemática Mención Estadística

Cuadro 7.114: EVALUACIÓN DE MÉTODOS DE SELECCIÓN

$\beta = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$									
<i>Colinealidad = Alta; P=20; n=100; S=1000</i>									
SR	249702.9			24970.3			2497.0		
σ^2	0.01			0.1			1		
Indicadores	TECM	ECMP	F	TECM	ECMP	F	TECM	ECMP	F
AIC_{HKB}	0.00010	4.10807	0.86620	0.00043	4.17863	0.84007	0.00310	5.30868	0.54962
AIC_{LW}	0.00011	4.11755	0.86733	0.00007	4.25155	0.85214	0.00316	5.95862	0.55390
AIC_{DK}	0.00685	2.47623	0.66927	0.00635	2.84314	0.61291	0.00436	4.39078	0.48856
C_{HKB}	0.02222	0.79388	0.00000	0.02159	1.47841	0.00000	0.02168	3.71979	0.00000
C_{LW}	0.02682	0.78185	0.00000	0.02191	1.26429	0.00000	0.02493	2.89398	0.00114
C_{DK}	0.02251	1.05306	0.01394	0.02188	2.74388	0.09085	0.00698	4.64520	0.12571
MCp_{HKB}	0.02222	0.79388	0.00000	0.02159	1.61450	0.00000	0.02168	3.86221	0.00029
MCp_{LW}	0.02682	0.78111	0.00000	0.02191	1.27663	0.00000	0.02493	2.84767	0.00114
MCp_{DK}	0.02251	0.79449	0.00000	0.02188	1.42332	0.00000	0.02215	3.73480	0.00000
IC_{95}	0.00017	4.12337	0.49344	0.00072	4.34601	0.56888	0.00851	5.53461	0.38933
IC_{99}	0.00017	4.12362	0.50406	0.00064	4.32497	0.60677	0.01914	5.76496	0.29835
$LASSO$	0.00004	3.54251	0.72124	0.00047	2.90468	0.70029	0.00339	5.78464	0.48003
EN	0.00005	3.43690	0.63376	0.00298	3.28289	0.59133	0.02466	11.80578	0.40933
R_{HKB}	0.00016	4.10229	-	0.00084	4.20391	-	0.00281	5.35754	-
R_{LW}	0.00017	4.12314	-	0.00072	4.34192	-	0.00617	6.19196	-
R_{DK}	0.00024	3.77455	-	0.00364	3.57729	-	0.00438	3.85416	-
MCO	0.00017	4.12321	-	0.00072	4.34275	-	0.00624	6.21604	-

7.2.3. Comparación Gráfica de Métodos.

Las figuras a continuación expuestas fueron desarrolladas pensando en contrastar la mayor cantidad de información posible considerando las limitaciones de graficar en 2 dimensiones y sin saturar de información la figura, en otras palabras el objetivo de estas figuras es desplegar mucha información sin ser difícil de interpretar.

Debido a que se mantienen fijos algunos aspectos, la comparación de resultados entre gráficos da una visión del efecto de estos aspectos en la selección de variables.

Para comenzar el análisis se tomará en cuenta las figuras 7.1, 7.2 y 7.3, ya que las tres figuras mantienen fijo el vector de coeficientes y la cantidad de variables disponibles, entre ellos solo varía el nivel de la colinealidad de menos a más. En estas tres figuras se aprecia que el método de selección de variables que tuvo los peores resultados fue MCP con cualquiera de los estimadores del parámetro de sesgo. Por otra parte se puede ver que los métodos basados en AIC (con estimadores del parámetro de sesgo LW, HKB) dan los mejores resultados cuando la varianza es baja o moderada siendo muy competitivos con LASSO e incluso mejores. Si ahora se toma en cuenta las mismas figuras pero solo en la parte que la varianza es alta, se ve claramente en el caso de los métodos basados en Intervalos de Confianza que el aumento de las observaciones contrarresta el problema causado por la varianza.

Si se considera ahora las figuras 7.4, 7.5 y 7.6 para ver el efecto de la colinealidad, primero se debe precisar que se tienen disponible 15 variables. En estas tres figuras claramente el factor que más perjudica a la capacidad de seleccionar variables es el nivel de la varianza de los residuos. Además para las tres figuras se puede ver que cuando la varianza es alta el aumento de las observaciones compensa la falta de precisión en la selección. Por otra parte cuando la colinealidad es alta y la varianza es baja o moderada los resultados muestran que

AIC con parámetro de sesgo LW y HKB dan mejores resultados que LASSO y Elastic Net.

Por su parte cuando se estudian los gráficos de las figuras 7.7, 7.8 y 7.9, se ve que en cada uno de ellos, cuando la varianza es baja o moderada, los resultados de AIC con LW son tan buenos como los de LASSO incluso mejores aun cuando se tienen pocas observaciones. Dentro de las 3 figuras los peores resultados se dieron cuando la varianza es alta, la colinealidad es alta y las observaciones eran pocas.

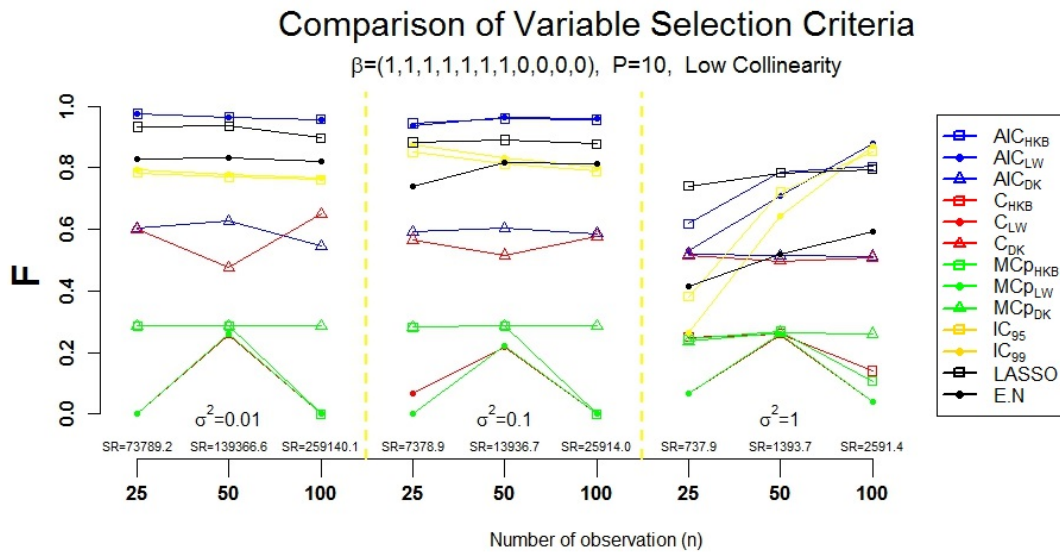


Figura 7.1: Fuente: Elaboración propia.

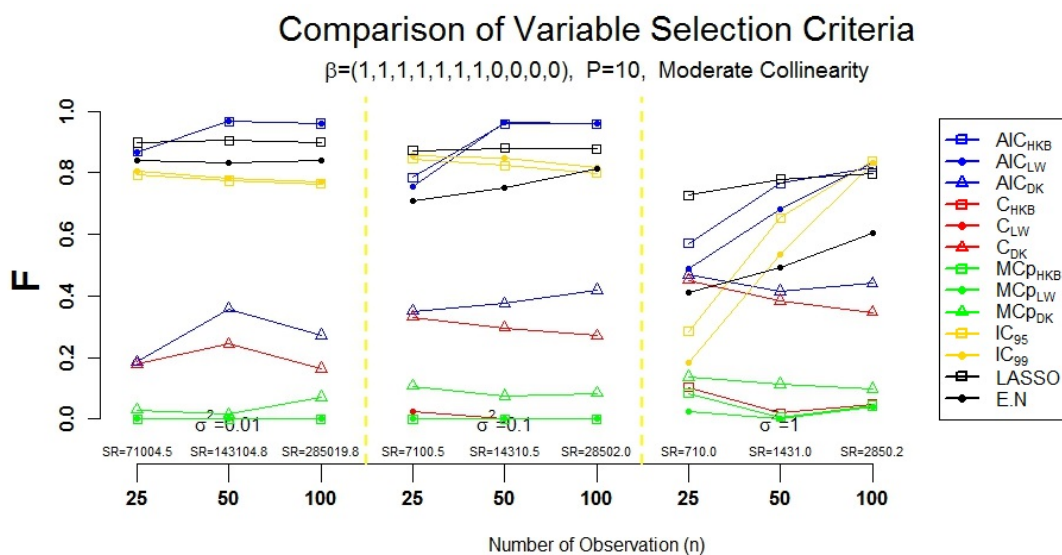


Figura 7.2: Fuente: Elaboración propia.

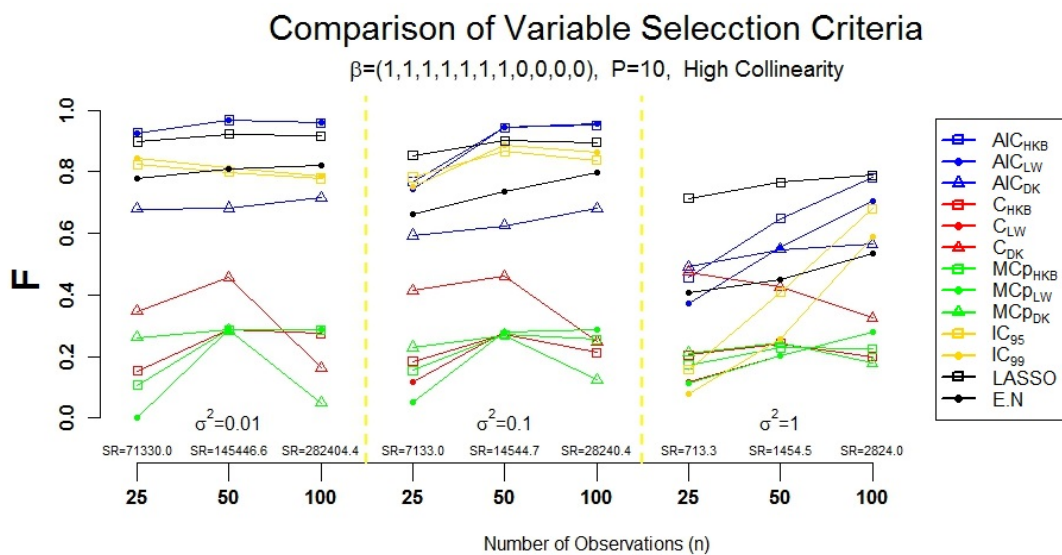


Figura 7.3: Fuente: Elaboración propia.

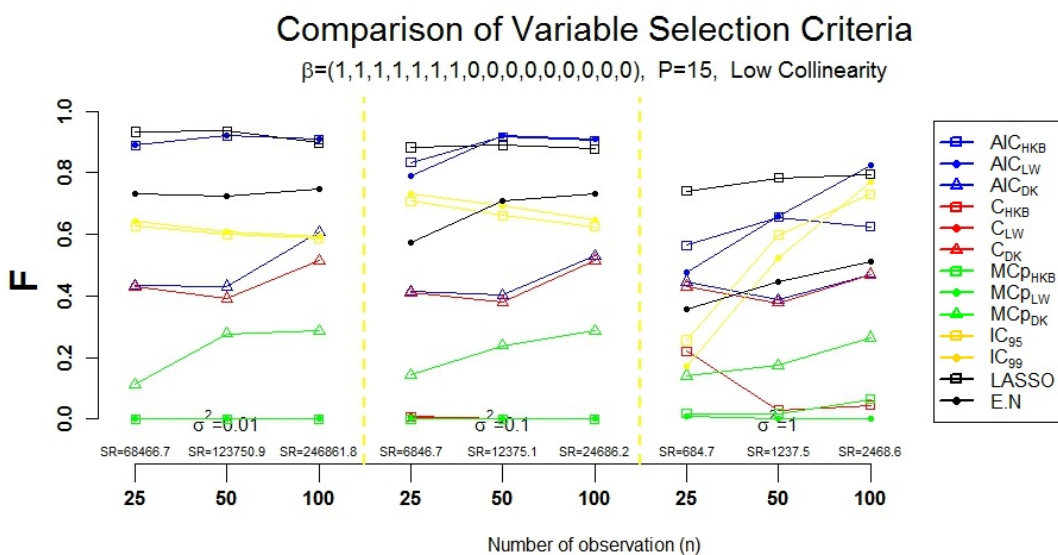


Figura 7.4: Fuente: Elaboración propia.

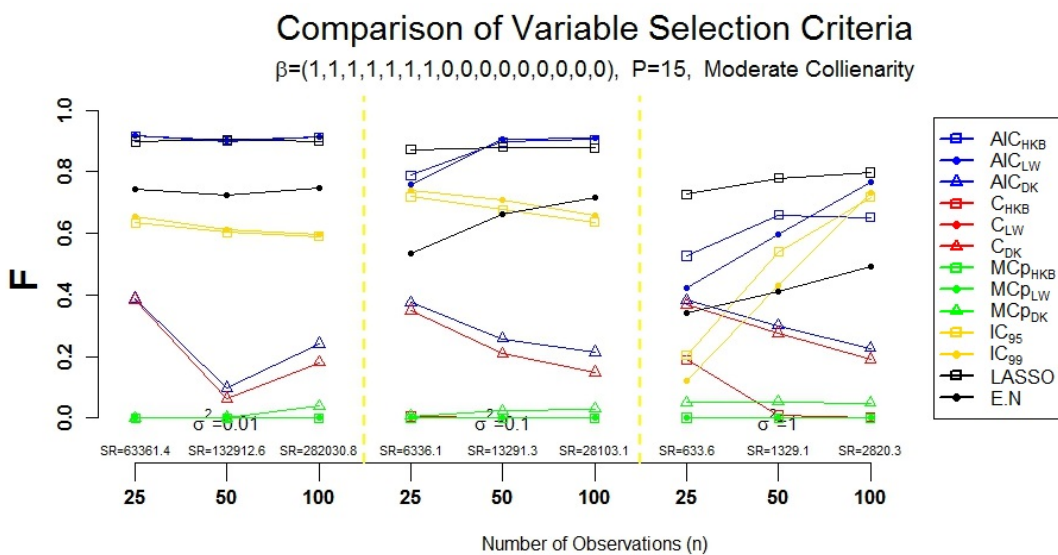


Figura 7.5: Fuente: Elaboración propia.

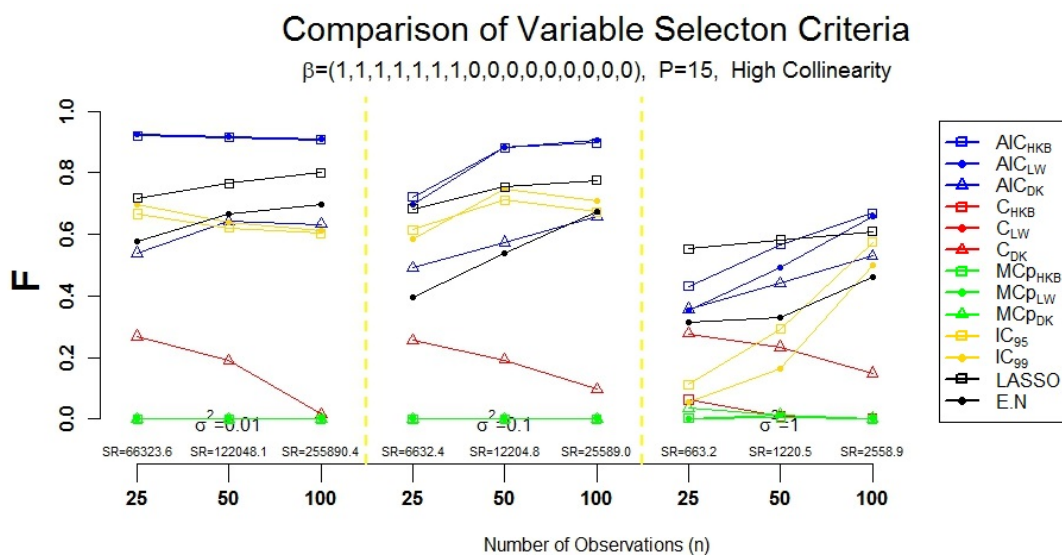


Figura 7.6: Fuente: Elaboración propia.

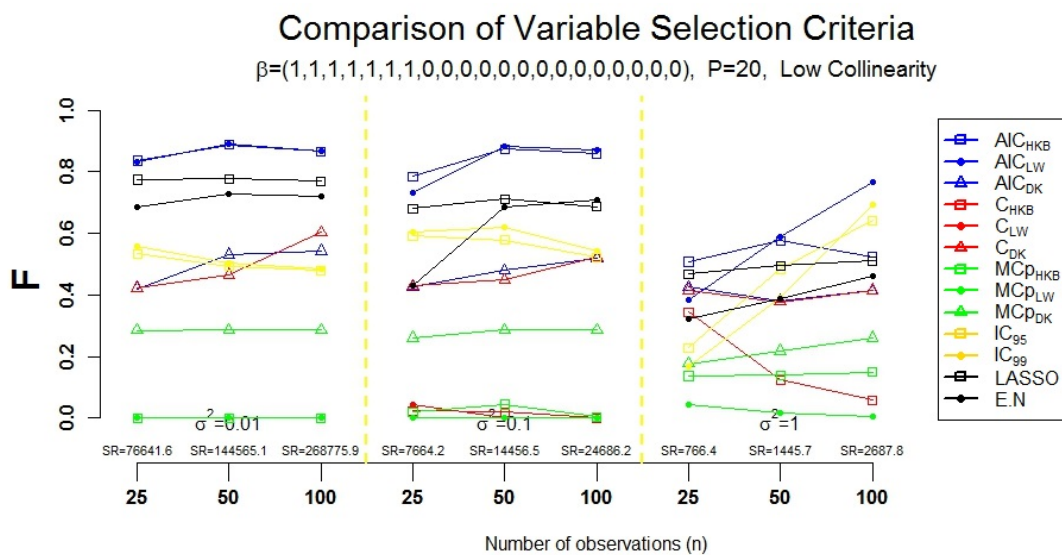


Figura 7.7: Fuente: Elaboración propia.

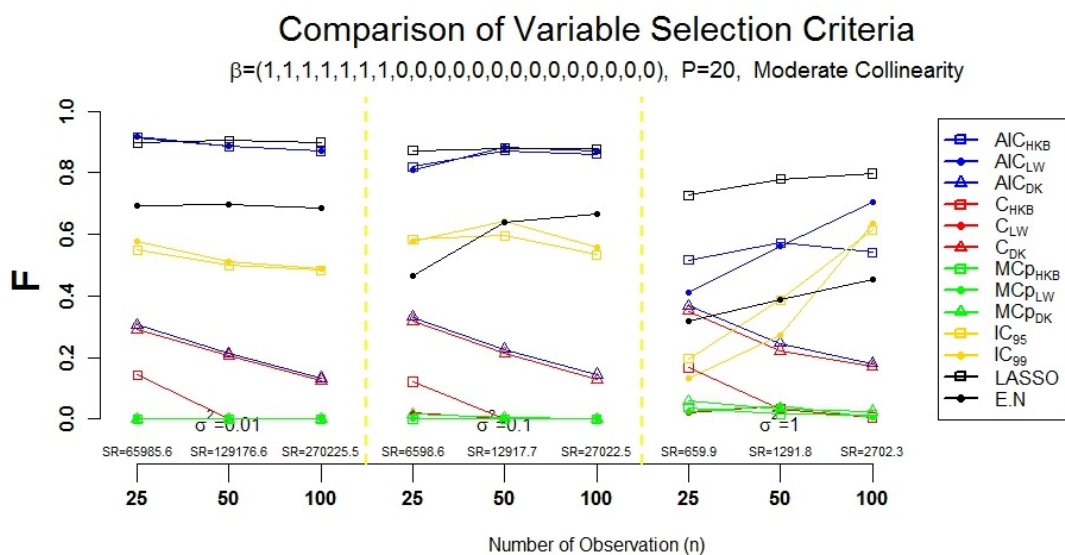


Figura 7.8: Fuente: Elaboración propia.

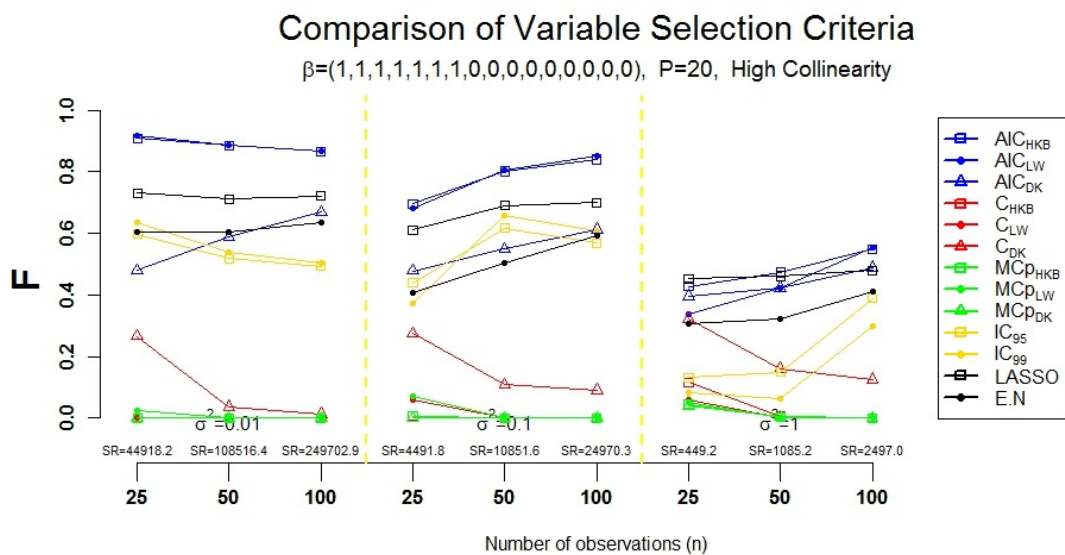


Figura 7.9: Fuente: Elaboración propia.

Magister en Matemática Mención Estadística

A continuación las Figuras que se estudian consideran el vector de coeficientes teórico calculado a partir de los valores propios de la matriz que considera solo las variables relevantes. Para comenzar el análisis se tomará en cuenta las figuras 7.10, 7.11 y 7.12, ya que las tres figuras mantienen fijo el vector de coeficientes y la cantidad de variables disponibles, entre ellos solo varía el nivel de la colinealidad de menos a más. En estas tres figuras se aprecia que el método de selección de variables que tuvo los peores resultados fue MCp con cualquiera de los estimadores del parámetro de sesgo. Por otra parte se puede ver que los métodos basados en AIC pero con estimadores del parámetro de sesgo LW, HKB dan los mejores resultados cuando la varianza es baja siendo muy superiores a LASSO y Elastic Net, por su parte los métodos basados en Intervalos de Confianza tuvieron un comportamiento similar a Elastic net. Si ahora se toma en cuenta las mismas figuras pero solo en la parte en que la varianza es alta, se ve claramente que LASSO muestra resultados superiores a los demás métodos.

Si se considera ahora las figuras 7.13, 7.14 y 7.15 para ver el efecto de la colinealidad, primero se debe precisar que se tienen disponible 15 variables. En estas tres figuras claramente el factor que más perjudica a la capacidad de seleccionar variables es el nivel de la varianza de los residuos. Además para las tres figuras se puede ver que cuando la varianza es alta el aumento de las observaciones no tiene un impacto significativo en la selección. Por otra parte cuando la colinealidad es alta y la varianza es alta los resultados muestran que LASSO es muy superior a los demás métodos independiente de la cantidad de observaciones.

Por su parte cuando se estudian los gráficos de las figuras 7.16, 7.17 y 7.18, se ve que en cada uno de ellos que cuando la varianza aumenta los resultados se deterioran considerablemente siendo LASSO el que mejor responde, pero no se ve un impacto positivo de las observaciones tan notorio como en las figuras 7.7, 7.8 y 7.9.

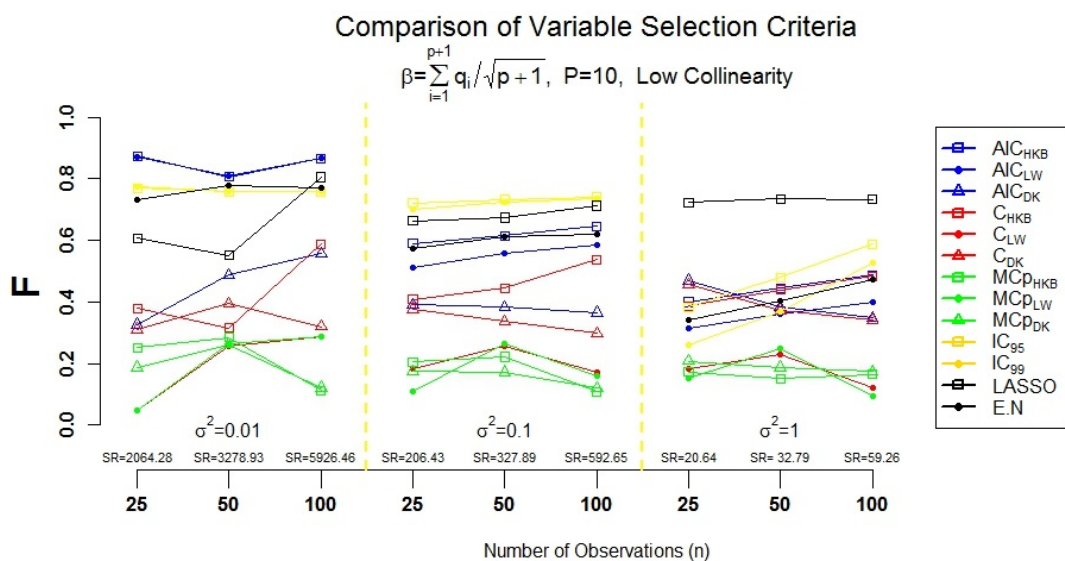


Figura 7.10: Fuente: Elaboración propia.

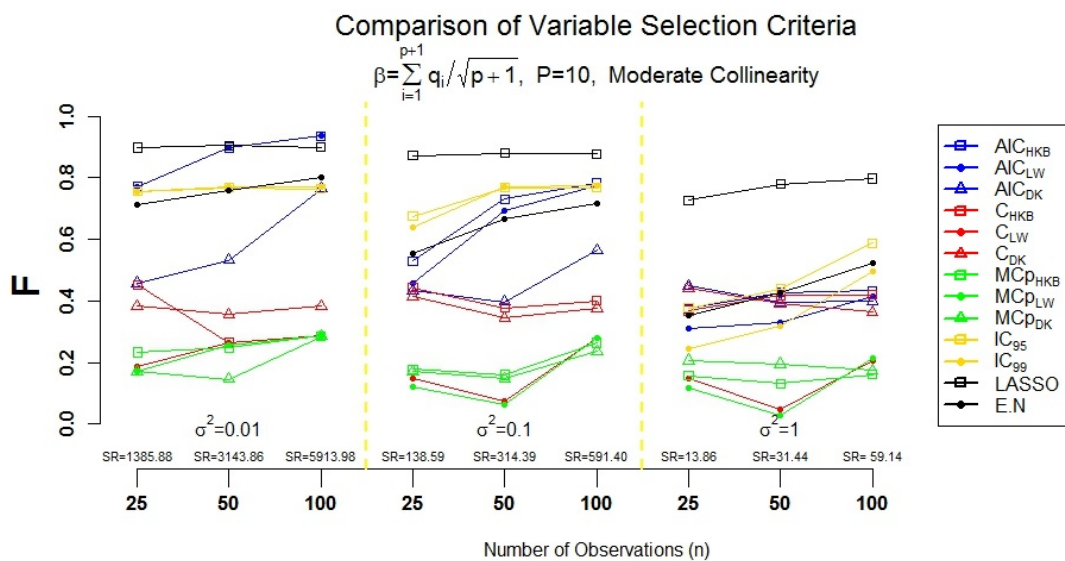


Figura 7.11: Fuente: Elaboración propia.

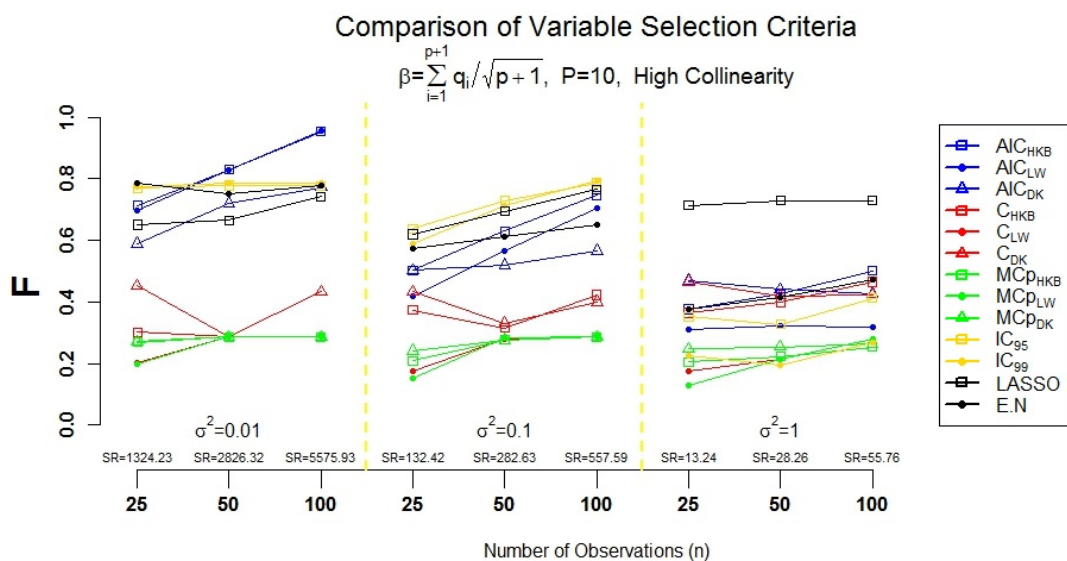


Figura 7.12: Fuente: Elaboración propia.

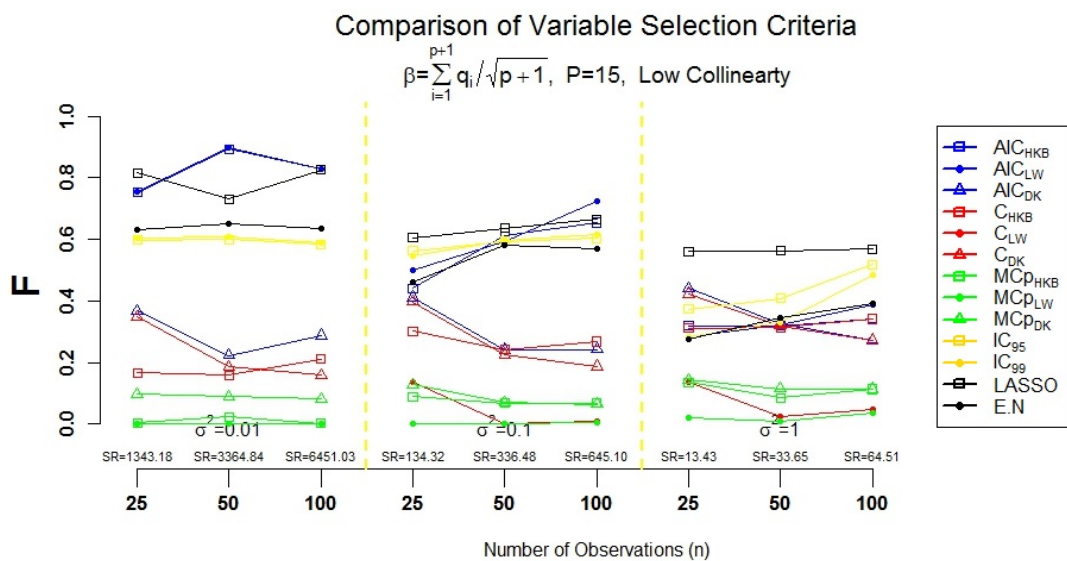


Figura 7.13: Fuente: Elaboración propia.

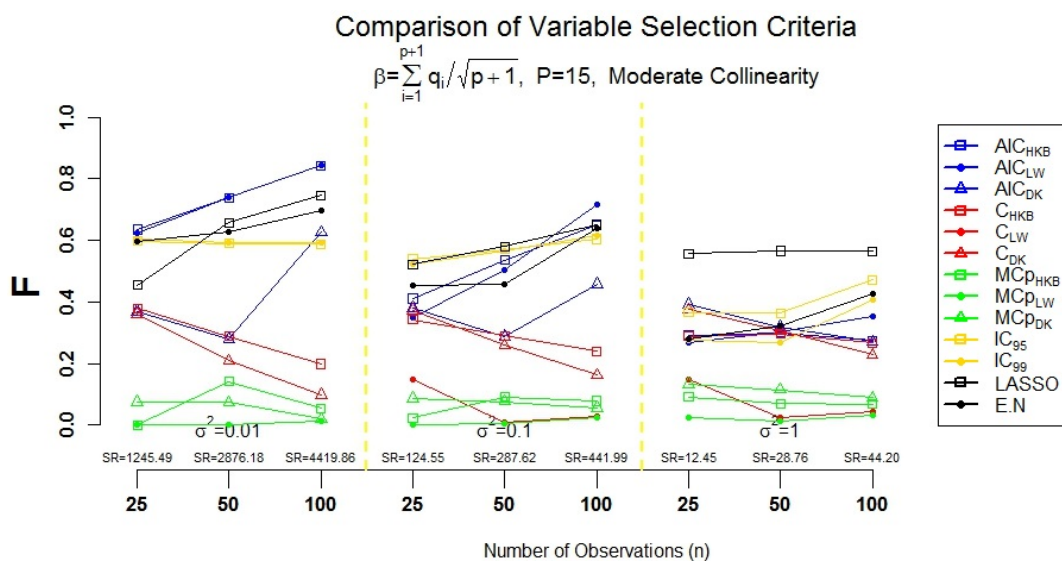


Figura 7.14: Fuente: Elaboración propia.

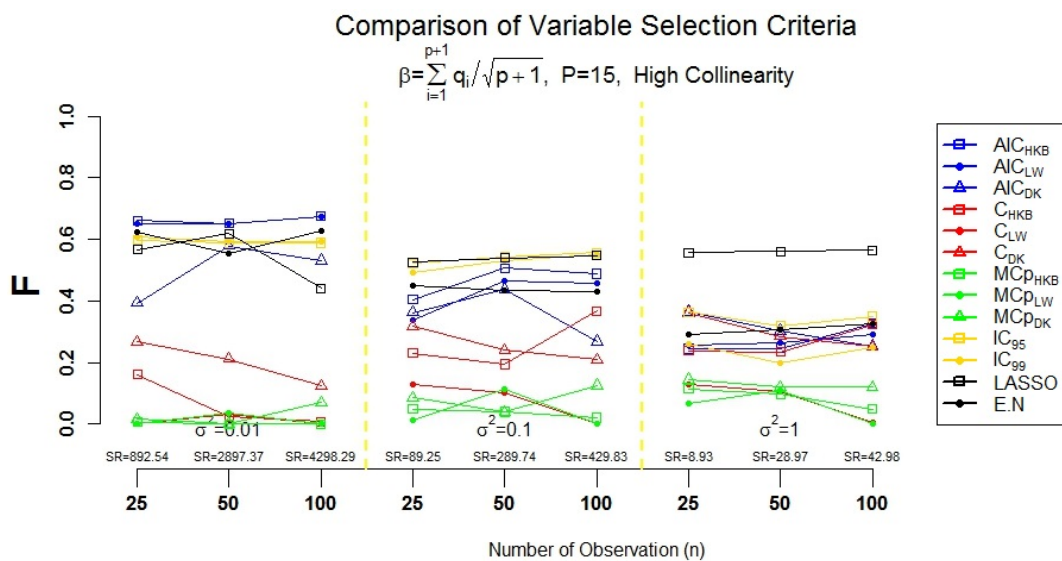


Figura 7.15: Fuente: Elaboración propia.

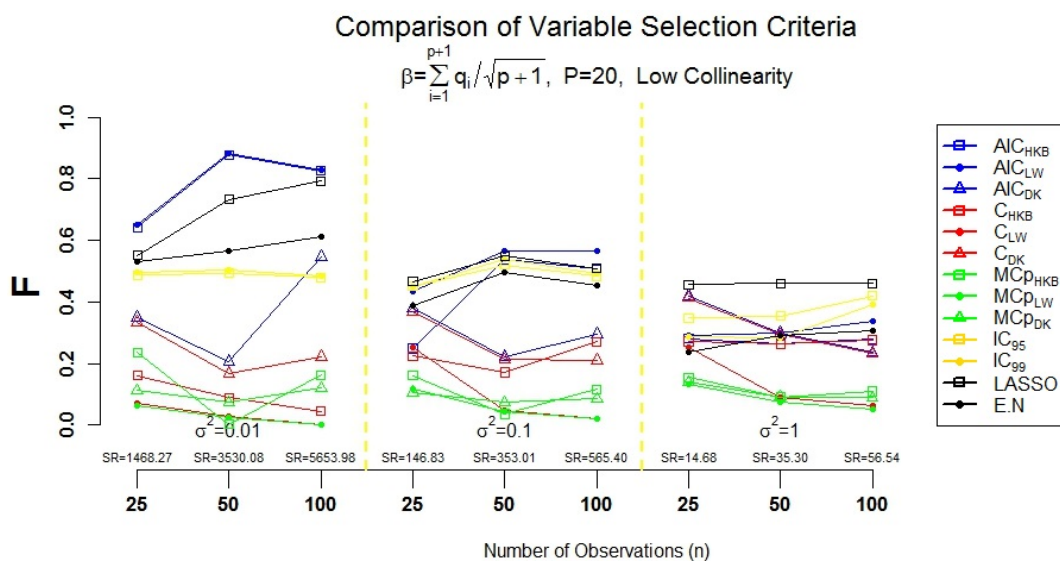


Figura 7.16: Fuente: Elaboración propia.

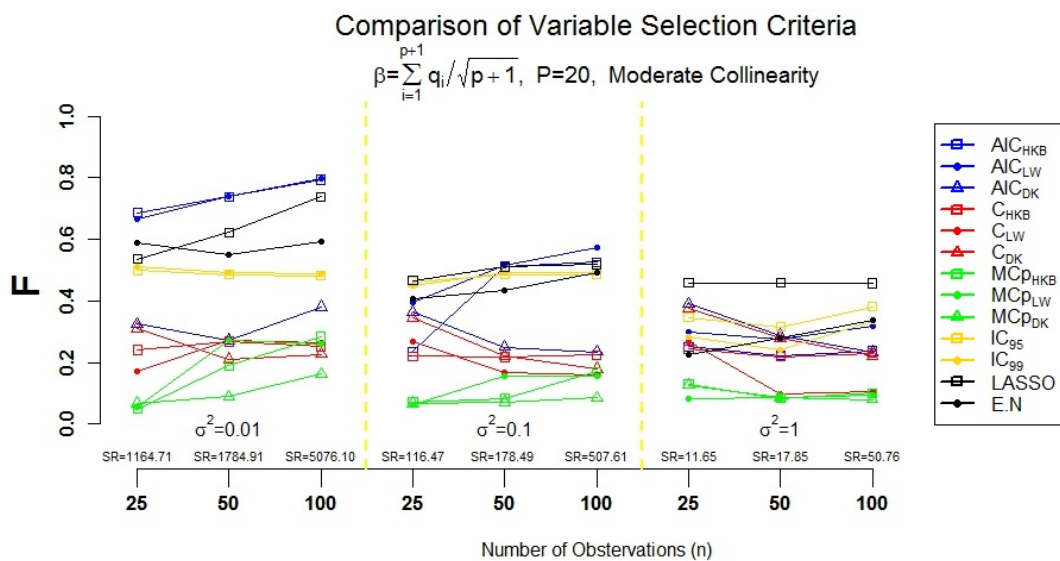


Figura 7.17: Fuente: Elaboración propia.

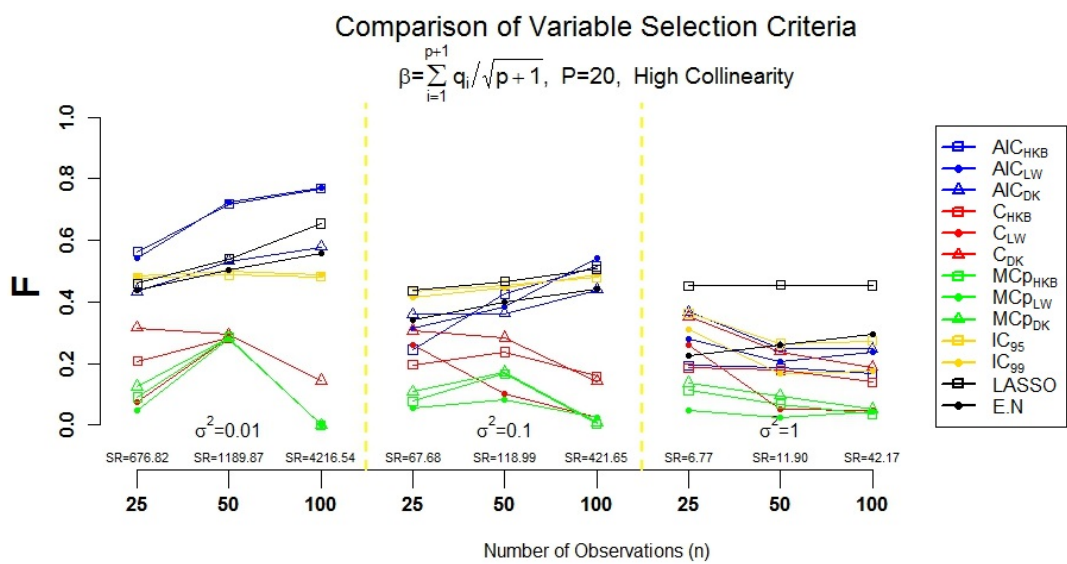


Figura 7.18: Fuente: Elaboración propia.

Capítulo 8

Conclusiones y Trabajo Futuro.

La tesis desarrollada analiza los resultados entregados por el simulador considerando el hecho de que este se configuró para estudiar varios modelos. Ahora bien, para una fácil comprensión a continuación se enumeran las conclusiones según temática.

8.1. Respecto del Vector de Coeficientes.

Se debe destacar que el vector de coeficientes es un elemento que causa un efecto en cómo se seleccionará variables, claramente se distinguen resultados más desfavorables para los métodos cuando se estimaba el vector de coeficientes calculado a partir de los valores propios, pero las tendencias seguían iguales a excepción del efecto positivo del aumento de las observaciones.

8.2. Respecto de la Varianza.

Con los resultados de este estudio queda claro que el factor que más repercute en la selección de variables es la varianza y esto es independiente de la colinealidad, de la

cantidad de variables disponibles, de la cantidad de observaciones y el tipo de vector de coeficientes.

8.3. Respecto de las Observaciones.

El aumento de las observaciones mostraron un efecto positivo en la selección de variables para los métodos AIC y los basados en Intervalos de Confianza principalmente cuando el vector de parámetros es definido arbitrariamente ya que cuando este es definido por los valores propios, el aumento de las observaciones no muestra tener efectos positivos significativos, más bien tienen un efecto neutral.

8.4. Respecto de la Multicolinealidad.

El nivel de la colinealidad tiene un efecto deteriorante en la selección de variables pero es menos notorio que el efecto producido por el nivel de la varianza y que se percibe aun menos cuando la cantidad de variables disponible es pequeña, su efecto se ve con mayor intensidad cuando aumenta la cantidad de variables disponibles, concretamente sus efectos se pueden notar comparando tablas y figuras pero cuando se disponen de 15 o 20 variables.

8.5. Respecto del Indicador F.

Resultó ser un indicador fácil de interpretar, implementar y utilizar aunque se tuvo que considerar una modificación en el cálculo del mismo puesto que no se contempla el caso de que se descarten todas las variables, situación que ocurre con los métodos basados en intervalos de confianza, se decidió finalmente evaluar esta situación como algo indeseado, de tal modo que cuando se daba esta situación F toma el valor cero.

8.6. Respecto del Error Cuadrático Medio Total (ECMT).

Es un indicador fácil de leer, se espera que sea pequeño, lo que indica mayor precisión en la estimación de los coeficientes. En general funcionó bastante bien permitiendo una fácil lectura, pero se descubrió que los resultados tenían escalas extremadamente diferentes, lo que impide la construcción de gráficos de fácil interpretación. Además es claro que el factor que más perjudica las estimaciones de los coeficientes es la varianza.

Los resultados muestran que la estimación de los coeficientes no necesariamente está ligada a la correcta selección de variables ya que el ECMT no siempre coincide con el indicador F.

8.7. Respecto del Error Cuadrático Medio de Predicción (ECMP).

Es un indicador fácil de leer, se espera que sea pequeño, lo que indica mayor precisión en la estimación de las predicciones. En general funcionó bastante bien permitiendo una fácil lectura, pero tiene el mismo problema que el ECMT, los resultados tenían escalas extremadamente diferentes, lo que impide la construcción de gráficos de fácil interpretación. Además es claro que el factor que más perjudica las predicciones es la varianza.

8.8. Respecto de los Métodos de Selección de Variables.

En general se evaluaron 6 métodos diferentes (AIC, C_p , MC_p , Intervalos de Confianza, LASSO y Elastic Net), pero al considerar los estimadores del parámetro de sesgo y los 2

intervalos de confianza (al 95 % y al 99 %) se constituyeron en total 13 métodos.

El estimador del parámetro de sesgo se constituyó como otro factor que influye en la selección de variables. En este sentido los estimadores que dieron mejores resultados *HKB*, *LW*.

Los métodos basados en intervalos de confianza se ven notoriamente beneficiados con la presencia de más observaciones, considerando varianzas altas y el vector de coeficientes definido arbitrariamente, pero cuando el vector de coeficientes se definió a través de los valores propios el aumento de observaciones no evidenció efectos significativos en este método.

Por su parte LASSO mostró ser particularmente resistente al efecto de la varianza. Si bien es cierto que se deteriora su rendimiento con el aumento de la varianza, no es menos cierto que su rendimiento, cuando la varianza es alta, es notoriamente mejor que los demás métodos pero este efecto se destaca con mayor facilidad cuando el vector de coeficientes se define por los valores propios de la matriz con variables relevantes.

Cabe mencionar que uno de los aportes de este estudio radica en guardar y catalogar la selección de modelos, ya que el estado del arte suele remitirse a la selección correcta o incorrecta, dejando de lado otras alternativas analizadas en esta tesis.

8.9. Respecto del Diseño e Implementación de la Simulación.

En el futuro se debe considerar de manera más exhaustiva y detallada el diseño y desarrollo del simulador, porque si bien es cierto lo importante es la interpretación de los resultados,

no es menos cierto que el grueso de este tipo de trabajos es la programación por lo que se requiere considerar seriamente las restricciones físicas y computacionales de la implementación.

8.10. Recomendaciones para utilización de criterios.

En esta tesis se estudió el desempeño de los métodos de selección de acuerdo a tres indicadores, se puede mencionar que en ningún escenario se dio que un método fuera satisfactorio para los tres indicadores lo que concuerda con el actual estado del arte.

Recomendar el uso de algún determinado criterio es inherente a la situación en la que se está intentando seleccionar, vale decir, depende de la colinealidad, la varianza, las observaciones y la cantidad de variables disponibles, aun así los resultados mostraron ciertos patrones:

- 1) Si la colinealidad es baja y la varianza es baja es recomendable el uso de AIC con estimador LW o HKB para seleccionar variables.
- 2) Si la varianza es alta, LASSO es el que funciona mejor en la selección de variables, aunque si se disponen de muchas observaciones (desde 100) el método por Intervalo de Confianza compite con LASSO.
- 3) Para varianzas moderadas no hay diferencias importantes entre LASSO y AIC con estimador LW o HKB en la capacidad de seleccionar variables.
- 4) Respecto de la estimación de los Coeficientes, es dominada por el estimador RR, sobre todo para varianzas, además la estimación de coeficientes con los modelos seleccionados por AIC con estimador LW o HKB son competitivos con RR.

5) Respecto de las predicciones el resultado no es concluyente, los resultados están dispersos lo que impide notar un patrón, pero se puede indicar que en este aspecto, funcionaron bien RR con estimador DK , LASSO y el métodos AIC con estimadores HKB, LW y DK.

8.11. Trabajo Futuro.

1) Los niveles que se implementaron para este trabajo, para la multicolinealidad, la varianza y las observaciones se reducen a tres. Si las características computacionales lo permiten, en futuras investigaciones, sería útil elevar la cantidad de niveles de la multicolinealidad, es decir, realizar simulaciones con niveles Bajos, Moderado-bajos, Moderados, Moderados-Altos y Altos (5 niveles de Multicolinealidad). La mismo principio sería útil de incorporarlo en la varianza y las observaciones, con la finalidad de identificar tendencias.

2) La presencia de valores atípicos y sus efectos en la selección de variables, en la estimación de coeficientes y la predicción, en este proyecto no fue abordado, mediante algunas modificaciones al algoritmo del simulador se podría realizar un estudio en este tema.

3) Un estimador de dos etapas basado en LASSO y Regresión Ridge, vale decir, seleccionar variables con LASSO y estimar los coeficientes con Ridge, podría ser una alternativa a Elastic Net en el contexto de que $p < n$. Mediante un estudio de simulaciones se podría evaluar esta propuesta y contrastarla con LASSO y Elastic Net .

Bibliografía

ALHEETY, M. I., KIBRIA, B. G. (2014). *A generalized stochastic restricted ridge regression estimator*. Communications in Statistics-Theory and Methods, 43(20), 4415-4427

ANZANELLO, M. J., FOGLIATTO, F. S. (2014). *A review of recent variable selection methods in industrial and chemometrics applications*. European Journal of Industrial Engineering, 8(5), 619-645. APARICIO, J., MARTINEZ, M., MORALES, J. (2004). *Modelos Lineales Aplicados en R*. Dto. Estadística, Matemáticas e Informática.

CASELLA, G., BERGER, R. L. (2002). *Statistical inference.*, (Vol. 2). Pacific Grove, CA: Duxbury,239.

CHATTERJEE, S., STEINHAEUSER, K., BANERJEE, A., CHATTERJEE, S., GANGULY, A. (2012, APRIL). *Sparse group lasso: Consistency and climate applications*. In Proceedings of the 2012 SIAM International Conference on Data Mining (pp. 47-58). Society for Industrial and Applied Mathematics.

DORUGADE, A., KASHID, D., (2010). *Alternative Method for Choossing Ridge*. Applied Mathematical Science, 4 (9),447-456.

DORUGADE, A. V. (2014). *New ridge parameters for ridge regression*. Journal of the Association of Arab Universities for Basic and Applied Sciences, 15, 94-99.

DORUGADE, A. V. (2014). *On comparison of some ridge parameters in Ridge Regression*. Sri Lankan Journal of Applied Statistics, 15(1).

DORUGADE, A. V. (2016). *Adjusted ridge estimator and comparison with Kibria's method in linear regression*. Journal of the Association of Arab Universities for Basic and Applied Sciences, 21, 96-102.

FIRINGUETTI, L., BOBADILLA, G. (2011). *Asymptotic confidence intervals in ridge regression based on the edgeworth expansion*. Statistical Papers, 52(2), 287-307.

FRANK, L. E., FRIEDMAN, J. H. (1993). *A statistical view of some chemometrics regression tools*. Technometrics, 35(2), 109-135.

FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R. (2001). *The elements of statistical learning*. (Vol. 1, pp. 241-249). New York: Springer series in statistics.

FU, W. J. (1998). *Penalized regressions: the bridge versus the lasso*. Journal of computational and graphical statistics, 7(3), 397-416.

GROß, J. (2003). *Restricted ridge estimation*. Statistics probability letters, 65(1), 57-64.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, 2^oed, USA, Springer, 2009, pag 76-74

Magister en Matemática Mención Estadística

HOERL A., KENNARD R. (1970). *Ridge regression: biased estimation for nonorthogonal problems.*, Technometrics12:55-67

HOERL, A. E., KANNARD, R. W., BALDWIN, K. F. (1975). *Ridge regression: some simulations.* Communications in Statistics-Theory and Methods, 4(2), 105-123.

KIBRIA, B. G. (2003). *Performance of some new ridge regression estimators.* Communications in Statistics-Simulation and Computation, 32(2), 419-435

KUBOKAWA, T., SRIVASTAVA, M. S. (2012). *Selection of variables in multivariate regression models for large dimensions.* Communications in Statistics-Theory and Methods, 41(13-14), 2465-2489.

LAWLESS, J., WANG P. (1976). *A simulation study of ridge and other regression estimators.* Communications in Statistics, Theory and Methods, 14 , pp. 1589-1604

LAWSON, C. L., HANSON, R. J. (1995). *Solving least squares problems.* Society for Industrial and Applied Mathematics.

LUO, J. ZUO, Y. (2011). *A New Test for Large Dimensional Regression Coefficients.*, Open Journal of Statistics,1, 212-216

ÖZKALE, M. R. (2009). *A stochastic restricted ridge regression estimator.* Journal of Multivariate Analysis, 100(8), 1706-1716.

SAKALLIOĞLU, S., KAÇIRANLAR, S. (2008). *A new biased estimator based on ridge estimation.* Statistical Papers, 49(4), 669-689.

SARKAR, N. (1992). *A new estimator combining the ridge regression and the restricted least squares methods of estimation*. Communications in statistics-theory and methods, 21(7), 1987-2000.

SMITS, G., KORDON, A., VLADISLAVLEVA, K., JORDAAN, E., KOTANCHEK, M. (2006). *Variable selection in industrial datasets using pareto genetic programming*. GENETIC PROGRAMMING SERIES, 9, 79.

STAMEY, T. A., KABALIN, J. N., FERRARI, M., YANG, N. (1989). *Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate*. IV. Anti-androgen treated patients. The Journal of urology, 141(5), 1088-1090.

TRENKLER, G. (1984). *On the performance of biased estimators in the linear regression model with correlated or heteroscedastic errors*, Journal of Econometrics, 25, 179-190

TIBSHIRANI, R. (1996). *Regression shrinkage and selection via the lasso*. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.

TUSELL, F. (2011). *Análisis de regresión. Introducción teórica y práctica basada en R*. Adolescence. An age of opportunity.

TUTZ, G., BINDER, H. (2007). *Boosting ridge regression*. Computational Statistics Data Analysis, 51(12), 6044-6059.

YANAGIHARA, H., SATOH, K. (2010). *An unbiased Cp criterion for multivariate ridge regression*. Journal of Multivariate Analysis, 101(5), 1226-1238.

Magister en Matemática Mención Estadística

VAN RIJSBERGEN, C. J. (1981). *Retrieval effectiveness*. Progress in communication sciences, 1, 91-118.

ZOU, H., HASTIE, T. (2005). *Regularization and variable selection via the elastic net*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.