UNIVERSIDAD NACIONAL PEDRO RUÍZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

TESIS

"APLICACIÓN DE LOS MÍNIMOS CUADRADOS PENALIZADOS EN LA RELACIÓN ENTRE EL ÍNDICE GENERAL DE LA BOLSA DE VALORES DE LIMA Y LOS ÍNDICES BURSÁTILES MUNDIALES: UN CASO DE MULTICOLINEALIDAD"

PARA OPTAR TÍTULO PROFESIONAL DE:

LICENCIADO EN ESTADÍSTICA

AUTORES:

Bach. Racchumí Vela, Augusto Elmer

Bach. Valladares Chávez, Diana Gabriela

ASESOR:

Dr. Chung Alva, Víctor Manuel

LAMBAYEQUE, MAYO DEL 2018

UNIVERSIDAD NACIONAL PEDRO RUÍZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

TESIS

"APLICACIÓN DE LOS MÍNIMOS CUADRADOS PENALIZADOS EN LA RELACIÓN ENTRE EL ÍNDICE GENERAL DE LA BOLSA DE VALORES DE LIMA Y LOS ÍNDICES BURSÁTILES MUNDIALES: UN CASO DE MULTICOLINEALIDAD"

PARA OPTAR TÍTULO PROFESIONAL DE:

LICENCIADO EN ESTADÍSTICA

AUTORES:

Bach. Racchumí Vela, Augusto Elmer

Bach. Valladares Chávez, Diana Gabriela

ASESOR:

Dr. Chung Alva, Víctor Manuel

LAMBAYEQUE, MAYO DEL 2018

UNIVERSIDAD NACIONAL PEDRO RUÍZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

TESIS

"APLICACIÓN DE LOS MÍNIMOS CUADRADOS PENALIZADOS EN LA RELACIÓN ENTRE EL INDICE GENERAL DE LA BOLSA DE VALORES DE LIMA Y LOS INDICES BURSÁTILES MUNDIALES: UN CASO DE MULTICOLINEALIDAD"

Para optar el título de:

LICENCIADO EN ESTADÍSTICA

SUSTENTADO Y APROBADO ANTE LOS SIGUIENTES MIEMBROS DEL JURADO

Lic. Est. HUGO SAAVEDRA SAAVEDRA
PRESIDENTE

M. Sc MANUEL HURTADO SÁNCHEZ
SECRETARIO

Dr. JORGE ACOSTA PISCOYA
VOCAL

“APLICACIÓN DE LOS MÍNIMOS CUADRADOS PENALIZADOS EN LA RELACIÓN ENTRE EL INDICE GENERAL DE LA BOLSA DE VALORES DE LIMA Y LOS INDICES BURSÁTILES MUNDIALES: UN CASO DE MULTICOLINEALIDAD”

AUTORES:

Bach. AUGUSTO E. RACCHUMÍ VELA _____

Bach. DIANA G. VALLADARES CHÁVEZ _____

DIRIGIDA POR:

Dr. VICTOR MANUEL CHUNG ALVA _____
ASESOR

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS
DECANATO
Ciudad Universitaria - Lambayeque

ACTA DE SUSTENTACIÓN Nº 030-2018-D/FACFyM

(Sustentación Autorizada por Resolución N° 635-2018-D/FACFyM)

En la ciudad de Lambayeque, siendo las 5:30 PM. del día 18 de mayo del 2018 se reunieron en la videoteca del Laboratorio de física - FACFyM. los miembros del Jurado designados mediante Resolución Nº 656-2017-D/FACFyM, los docentes:

Lic. Est. Hugo Lorgio Saavedra Saavedra	Presidente
M.Sc. Manuel Francisco Hurtado Sánchez	Secretario
Dr. Jorge Antonio Acosta Piscoya	Vocal

Para recibir la tesis titulada:
Aplicación de los mínimos cuadrados penalizados en la relación entre el índice general de la Bolsa de Valores de Lima y los índices Bursátiles mundiales : un caso de multicolinealidad.

desarrollada por los Bachilleres en Estadística, **Racchumí Vela Augusto Elmer y Valladares Chávez Diana Gabriela.**

Después de escuchar la exposición y las respuestas a las preguntas formuladas por los miembros del Jurado, se acordó APROBAR el trabajo por UNANIMIDAD con el calificativo de Bueno

En consecuencia, los Bachilleres en referencia quedan aptos para recibir el Título Profesional de **Licenciado en Estadística,** de acuerdo a la Ley Universitaria, el Estatuto y Reglamento de la Universidad Nacional Pedro Ruiz Gallo de Lambayeque.

Observaciones:
ninguna.

Para constancia del hecho firman.

Lic. Est. Hugo Lorgio Saavedra Saavedra
Presidente

M.Sc. Manuel Francisco Hurtado Sánchez
Secretario

Dr. Jorge Antonio Acosta Piscoya
Vocal

DECLARACIÓN JURADA DE ORGINALIDAD

Nosotros, investigadores principales: Augusto Elmer Racchumí Vela y Diana Gabriela Valladares Chávez, y Víctor Manuel Chung Alva asesor del Trabajo de Investigación "Aplicación de los Mínimos Cuadrados Penalizados en la relación entre el Índice General de la Bolsa de Valores de Lima y los Índices Bursátiles Mundiales: Un caso de multicolinealidad"; declaramos bajo juramento que este trabajo no ha sido plagiado, ni contiene datos falsos. En caso se demostrara lo contrario, asumimos responsablemente la anulación de este informe y por ende el proceso administrativo a que hubiera lugar.

LAMBAYEQUE, MAYO 2018

Bach. Augusto E. Racchumí Vela
AUTOR

Bach. Diana G. Valladares Chávez
AUTOR

Dr. Víctor Manuel Chung Alva
ASESOR

DEDICATORIA

A Dios por ser el soporte en mi vida y la de mi familia; bendiciéndonos en cada momento y permitiéndome culminar una etapa más en mi vida académica.

A mi padre Augusto, a mi madre Zadith y a mi tía Angélica; quienes son el motor en mi vida.

Augusto Elmer Racchumí Vela

A Dios por guiarme durante toda mi vida universitaria, por todo y cuanto me ha dado y permitido, porque sin duda, sin él nada de esto hubiese sido posible.

A mis padres Enrique y Roxana, quienes me han heredado el tesoro más valioso que puede dársele a un hijo.

Diana Gabriela Valladares Chávez

AGRADECIMIENTOS

En primer lugar a Dios, porque una vida sin él, es una vida sin sentido.

Hay tantas personas en mi vida que pese a la distancia siempre estuvieron a mi lado; a mi padre que día a día una llamada lo cambiaba todo; a mi tía que cada semana siempre me da las bendiciones y a mi madre que está siempre a mi lado.

Y a mis buenos amigos de clase, a quienes los considero como una nueva familia, a los docentes de la carrera profesional de estadística y a mi asesor quien siempre me brindo su ayuda y confianza.

Augusto Elmer Racchumí Vela

Agradecer a Dios por haberme dado fuerza y valor para culminar esta etapa de mi vida. A mis padres, por el esfuerzo que hicieron para que me convirtiera en una persona de bien; a mi abuela Elisa, por su cariño y dedicación no sólo conmigo sino también con mis hermanos. A mis amigos, quienes pasaron a convertirse en parte de mi familia.

No puedo dejar de expresar mi gratitud a la facultad y los docentes que laboran en la escuela de estadística, por la formación académica que me entregaron durante mis años de estudio, convirtiéndome en el profesional que ahora soy.

Diana Gabriela Valladares Chávez

INDICE GENERAL

INDICE DE TABLAS

INDICE DE FIGURAS

RESUMEN

En el presente informe de tesis tuvo como finalidad determinar los estimadores más eficientes que expliquen el comportamiento de Índice General de la Bolsa de Valores de Lima (IGBVL) a partir de índices bursátiles mundiales. Los resultados aportaron información confiable con base científica del impacto positivo que trae consigo el uso de técnicas estadísticas modernas en función a al control de la multicolinealidad y su valor agregado en las estimaciones de los coeficientes de modelos econométricos que son utilizados en las planificaciones estratégicas y en la toma decisiones económicas de los países.

Para el presente estudio se tomó como muestra, los valores diarios recuperados de once series bursátiles entre los años 2000 y 2014, en donde cada una concentraba 3,773 observaciones.

Para el análisis se utilizó la regresión sesgada tales como: Regresión por Componentes Principales (RCP), Regresión Ridge y Regresión LASSO, y cuyo objetivo aplicativo fue comprobar la eficacia de los procedimientos sesgados poseen una mejor eficiencia respecto al estimador tradicional como es el de Mínimos Cuadrados Ordinarios (MCO).

Finalmente, se determinó que los métodos que genera menor varianza y superan a los indicadores MCO fueron: La Regresión por Componentes Principales (0.0000394) y Ridge (0.0003426).

Palabras clave: Estimadores sesgados, Regresión RCP, Ridge y LASSO.

ABSTRACT

The purpose of this thesis report was to determine the most efficient estimators that explain the behavior of the The IGBVL (Bolsa de Valores de Lima General Sector Index) based on world stock indices.

The results provided reliable scientific information on the positive impact the in use of modern statistical techniques has on the control of multicollinearity and its added value in the estimations of the coefficients of econometric models that are used in strategic planning and in It makes economic decisions of the countries.

For the present study, do used as sample the daily values recovered by the different stock series between 2000 and 2014, concentrated in 3,773 observations.

For the analysis we applied the used of biased estimations, such as Principal Components Regression (PCR), Ridge Regression and LASSO Regression and whose objetive was to demonstrate that the biased procedure has a better efficiency regarding the mean squared error and variance with respect traditional estimator the Ordinary Least Squares (OLS).

Finally, it was determined that the techniques that generate the least variance and surpass the OLS indicators were: The Regression by Principal Components (0.0000394) and Ridge (0.0003426).

Keywords: Biased Estimators, RCP Regression, Ridge, LASSO

INTRODUCCIÓN

El análisis de Regresión es un método estadístico comúnmente utilizado en las investigaciones econométricas cuando se consideran variables predictivas múltiples para estimar la asociación con la variable objeto de estudio. Sin embargo, la eficacia del análisis Regresión depende en gran medida de la estructura de correlación entre las variables predictivas, ya que para la realización de inferencia multivariante, supone que todas las variables predictivas no están correlacionadas, caso que no ocurre en la economía. Cuando las covariables del modelo no son independientes entre sí, en el análisis surgen problemas de colinealidad o multicolinealidad, lo que conduce a un aumento de la varianza de los estimadores de los coeficientes de regresión y la pérdida del poder predictivo.

La colinealidad conduce a estimaciones imprecisas de los coeficientes de Regresión con signos erróneos. Además, revela que un pequeño cambio en los datos puede conducir a grandes diferencias en los coeficientes de Regresión.

Las investigaciones basadas en modelamiento predictivo de indicadores macroeconómicos en un país, tienen impactos grandes en las decisiones económicas. Generalmente las variables macroeconómicas que intervienen en dichas investigaciones están muy correlacionadas entre sí; generando problemas estadísticos de multicolinealidad, y llevando como consecuencia a interpretaciones erróneas en resultados econométricos.

Dado este contexto y la necesidad de demostrar la importancia del uso de técnicas estadísticas para el control de los efectos de la multicolinealidad en modelos económicos y a su vez brindar información confiable con base científica del impacto positivo que trae consigo el uso de técnicas estadísticas modernas en función al control de la multicolinealidad y su valor agregado en las estimaciones de los coeficientes de modelos econométricos que son utilizados en las planificaciones estratégicas y en la toma

decisiones económicas de los países, e planteó investigar ¿Cómo enfrentar problemas generados por la presencia de multicolinealidad en modelos econométricos, para mejorar la precisión de sus estimaciones, Caso comportamiento del índice general de la bolsa de valores de Lima a partir de índices bursátiles?.

En base a lo anteriormente expuesto, se consideró el estudio de las variables bursátiles más representativas de las bolsas de valores internacionales tales como:

- SP500 y Nikkei: constituidas por las acciones de las mejores empresas norteamericanas y de Asía respectivamente.
- Dow Jones y Nasdaq: Constituida por las acciones más representativas del rubro tecnológico.

Con la finalidad de replicar un caso de multicolinealidad (identificación, consecuencias y tratamientos), se tomaron para el estudio la combinación de variables del rubro tecnológico tales como: Apple, Microsoft, International Business Machines, Amazon, AT&T y HP.

Posteriormente se investigó y aplicó la utilización de técnicas de Regresión modernas que ayuden a minimizar la inflación de varianza, como son la Regresión por Componentes Principales (RCP), Ridge y LASSO.

La finalidad de este trabajo de investigación consiste en encontrar un modelo que permita explicar el comportamiento del Índice General de la Bolsa de Valores de Lima (IGBVL) en función a los índices bursátiles mundiales; a partir de la determinación de estimadores que minimicen la varianza producida por la multicolinealidad.

I. DISEÑO TEÓRICO

1.1. Antecedentes

Carrasco (2016) en su tesis de grado titulado **"Técnicas de Regularización en Regresión: Implementación y Aplicación"** analizó los problemas de colinealidad en la estimación de los coeficientes de regresión por mínimos cuadrados ordinarios (MCO), con la finalidad de obtener estimaciones y predicciones fiables recurrió a la comparación de los métodos de regresión regularizada: Lasso, Ridge y Elastic Net. Se llegó a la conclusión que la Regresión Ridge posee errores menores y coeficientes contraídos a cero superando así a los resultados obtenidos en los métodos de Lasso y Elastic.

Del Valle et al (2012) en su investigación **"La Multicolinealidad en modelos de Regresión Lineal Múltiple",** abordaron la problemática de la Multicolinealidad entre las variables regresoras en el Modelo de Regresión Lineal Múltiple aplicada a los ciencias agrícolas, en donde recrearon una situación de multicolinealidad en los datos con la finalidad de hacer comparaciones de efectividad de la Regresión Ridge y Componentes Principales con respecto al método de los mínimos cuadrados ordinarios(MCO), concluyéndose que Regresión Ridge y la Regresión sobre Componentes Principales, resultan efectivas para describir con exactitud y precisión los estimadores en el Modelo de Regresión Lineal Múltiple.

Dormann (2013) en su investigación titulada **"Collinearity: a review of methods to deal with it and a simulation study evaluating their performance."** Aborda el alcance del problema de la colinealidad en la ecología en base al desarrollo de diferentes enfoques de los métodos de variables latentes. Utilizando datos simulados con cinco relaciones predictor-respuesta de creciente complejidad y ocho niveles de colinealidad, se comparó formas de abordar la colinealidad con los métodos estándar de regresión múltiple y aprendizaje por máquina. Se concluyó que los métodos diseñados específicamente para la colinealidad, como los métodos de variables latentes y modelos basados en árboles, no superó el tradicional GLM.

Paccapelo (2015) en su investigación titulada **"Modelos de Selección Genómica para Caracteres Cuantitativos basados en Marcadores Moleculares**

aplicados al mejoramiento de maíz ",dicha investigación tuvo como objetivo principal realizar una selección de variables genómicas utilizando varios modelos de regularización, entre ellas se utilizaron a la regresión ridge clásico (RR), ridge bayesiano (BRR), regresión lasso (RL), lasso bayesiano.

Pariasca (1990); en su investigación titulada. **"Eficiencia de los estimadores sesgados en regresión, en presencia de multicolinealidad"** realiza una simulación de datos con el objetivo de realizar una comparación con un modelo mínimo cuadrático ordinario (MCO), llegando a la conclusión que los modelos MCO no son buenos para estimar los parámetros verdaderos de un modelo de regresión. Los estimadores sesgados como Componentes Principales, Ridge y Ridge Generalizado generan menor error cuadrático medio y varianza que un modelo clásico.

Pineda (2013) en su investigación titulada **"Una prueba sobre la especificación de un modelo de regresión"**, que surge debido a la explicación en la relación de la variable dependiente con las explicativas. La prueba RESET, utilizada para demostrar la especificación de un modelo de regresión, tiene una baja potencia debido a la presencia de multicolinealidad en el modelo. Con el fin de dar solución a este problema, se propuso la utilización de una prueba de especificación basado en multiplicadores de Lagrange y regresión Lasso, llegando a concluir que el método LASSO propuesto tuvo un buen desempeño para tamaños de muestra grande sobreestimando tamaños de muestra pequeños, mientras que la prueba RESET tiene un buen desempeño para cualquier tamaño de muestra.

Sopipan (2013) en su investigación titulada **"Forescating the financial returns for using multiple regression based on Principal Component Analysis"** el objetivo de este estudio fue predecir los rendimientos del Índice de la Bolsa de Tailandia (SET) añadiendo Algunas variables explicativas y el orden autoregresivo estacionario p (AR (p)) en la ecuación media de retornos. Además, se utilizó el Análisis de Componentes Principales (PCA) para eliminar posibles complicaciones causadas por Multicolinealidad. Los resultados mostraron que las regresiones múltiples basadas en PCA, tienen el mejor rendimiento.

1.2. Colinealidad y multicolinealidad

Para que un método de Mínimo Cuadrados Ordinarios (MCO) de Regresión genere buenas inferencias y pronósticos, es necesario que cumpla con varios supuestos establecidos, pero cuando faltan al cumplimiento de algunos de estos supuestos, las estimaciones de los estimadores del modelo de Regresión no son confiables. En este caso, se tratará de resolver el "problema de la multicolinealidad o dependencia casi lineal de las variables regresoras, los cuales son las columnas de la matriz **X,** por lo que es claro que una dependencia lineal exacta causaría una matriz $\mathbf{X'X}$ singular". (Montgomery, 2006).

La colinealidad hace referencia a una relación unívoca entre algunas de las variables explicativas de un modelo Regresión lineal, en cuanto a la multicolinealidad identifica la existencia de más de una combinación lineal no perfecta entre las variables explicativas.

Cuando no existe una relación lineal entre las variables regresoras de un modelo, se dice que éstos son ortogonales". (Montgomery, 2006).

$$\boldsymbol{X'X = I}$$

Dónde $\mathbf{X}$: es una matriz de diseño nxp.

$\boldsymbol{X'}$: es la transpuesta.

$\mathbf{I}$: es la matriz identidad.

1.2.1. Origen de la multicolinealidad

"El término de multicolinealidad hace mención la existencia de una relación perfecta entre algunas o todas las variables explicativas de un modelo de Regresión. Sea una modelo de Regresión de k variables, se dice que existe una relación lineal exacta si se satisface la siguiente condición". (Gujarati, 2004)

$$\gamma_1 X_1 + \gamma_2 X_2 + \cdots + \gamma_k X_k = 0$$

En donde $\gamma_1, \gamma_2, \ldots, \gamma_k$ son constantes tales que no todas de ellas son iguales a 0.

Si dicha condición es verídica, $(\mathbf{X'X})^{-1}$ no existe y se dice que existe dependencia lineal en la matriz $\mathbf{X'X}$; es decir que existe un problema de multicolinealidad.

Montgomery (2006) hace mención sobre 4 fuentes Principales causantes del problema de multicolinealidad:

a) **Método utilizado en la recolección de datos**. Esto es por la toma de la muestra en intervalos limitados de los valores de los predictores.
b) **Restricciones en el modelo o en la población.** Se inicia cuando se incluyen en el modelo a variables que de antemano ya tienen una relación, por consiguiente, la multicolinealidad está presente.
c) **Especificación del modelo.** Surge cuando se adiciona un término polinomial al modelo, produciendo un deterioro en la matriz $\mathbf{X'X}$¸ además si el rango de $\mathbf{X}$ es pequeño, al agregar un término $\boldsymbol{X^2}$, puede generar una multicolinealidad importante.

d) **Modelo sobre definido.** Es generado por la presencia de una mayor cantidad de variables regresoras frente a pocas observaciones disponibles.

1.2.2. Consecuencias

Las consecuencias que genera la presencia de multicolinealidad en un modelo de Regresión son de tipo explicativo y de estimación.

En primer lugar, el tamaño de la del coeficiente de determinación es limitado haciendo más difícil añadir una predicción explicatoria extra con variables adicionales. Y en segundo lugar, hace difícil determinar la contribución individual de cada variable predictora, a causa de que los efectos de cada variable explicativa son mixtos. (Tatham, 2004)

Debido a este problema, la multicolinealidad da como resultado porciones más grandes de varianza compartida y niveles más bajos de varianza única a partir de los cuales se pueden determinar los efectos de las variables independientes individuales. (Tatham, 2004)

En el caso de las estimaciones, la presencia de una multicolinealidad no perfecta, "puede tener un efecto sobre los signos de los coeficientes de regresión. Más específicamente, un valor de βi puede tener el signo opuesto de lo que se espera. Así como también causan efectos sustanciales en las pruebas de significancia estadística". (Mendenhall, 2012)

1.2.3. Métodos para detectar la multicolinealidad

Como se menciona en Tatham (2004), "un supuesto clave en la interpretación del valor teórico de la regresión es la correlación entre las variables predictoras. La

multicolinealidad se trata de un problema muestral, no un problema de especificación del modelo".

Es por eso, que la forma más sencilla para detectar la presencia de multicolinealidad en un modelo de regresión es mediante una matriz de correlación entre cada par de variables regresoras. Si uno o más valores de r obtenidos son cercanos a 1 o -1; las variables en cuestión están altamente correlacionadas, y por ende, una multicolinealidad severa; o examinando el condicionamiento en la matriz $\mathbf{X'X}$.

Según Novales (2000), habrá presencia de multicolinealidad si se detectan las siguientes situaciones como:

- Pequeños cambios en los datos como exclusión o adición a la muestra de un reducido número de observaciones que producen gran variación en la estimación mínimo cuadrático.
- Las desviaciones típicas obtenidas para cada coeficiente estimado son altas e individualmente son poco significativas, a pesar de ser conjuntamente significativas y tener el modelo un R^2 elevado.

Tabla 1. *Métodos para detectar la multicolinealidad*

MÉTODO	DESCRIPCIÓN	FORMULA	LIMITE
Valor absoluto de los coeficientes de correlación **(| r |)**	Si las correlaciones exceden un umbral (0.5 - 0.7) es alto. (Booth et al. 1994)	-	| **r** | >0.7
Determinante de la matriz de correlación **(D)**	Producto del autovalor; Si D es cercano a 0 la colinealidad es alta, si D es cercano a 1 no hay colinealidad en los datos.	-	D es cercano a 0
Número de condición ***k*** **(X**	Se define como la razón entre el máximo valor propio dividido entre el mínimo valor propio.	$K(X)=\frac{\lambda_{max}}{\lambda \min}$	$1{,}000> K(X)100$

Índice de Condición **(IC):**	Los IC es la raíz cuadrada entre el máximo valor propio dividido entre el mínimo valor propio.Todas los *IC* iguales o mayores de 30 Son "grandes" y críticas. (Gujarati, 2004)	$\mathrm{IC}(\lambda_i)=\sqrt{\frac{\lambda_{max}}{\lambda \min}}$	$30> \mathrm{IC}(\lambda_i) > 10$
Factor de inflación de varianza **(VIF)**		$VIF_i = \frac{1}{1-{R_i}^2}$	VIF_i>10
Valor de tolerancia **(TOL):**	Es la cantidad de variabilidad de las variables independientes seleccionadas no explicadas por el resto de las variables independientes". (Tatham, R., 2004)	$TOL = \frac{1}{VIF_i}$	$TOL < 0.1$
Test de Farrar – Glauber	En esta prueba e busca evaluar la otorgonalidad de los regresores sobre la base de la matriz de correlaciones por pares entre las series independientes.	$\chi^2 cal= \quad -\left[n - - \frac{(2k+5)}{6}\right] . Ln \lvert R \rvert$	H_O= Las X son ortogonales entre si. H_I=Las X no son ortogonales entre si.

Elaboración propia

1.2.4. Remedios para la multicolinealidad

Mendenhall (2012) menciona las siguientes recomendaciones para dar solución al problema de multicolinealidad.

1. Eliminar una o más de las variables independientes correlacionadas del modelo. Un procedimiento de selección como la Regresión escalonada es útil en determinando cuáles variables sacar.
2. Si decide mantener todas las variables independientes en el modelo:
   - Evitar hacer inferencias sobre los parámetros β individuales (tales como como establecer una relación de causa y efecto entre y las variables predictoras).

- Restringir las inferencias sobre los valores de E (y) y futuros y a los valores de las variables independientes que pertenecen a la región experimental.

3. Si el objetivo final es establecer una relación de causa y efecto entre **Y** y las variables predictoras, utilice un experimento diseñado.
4. Para reducir los errores de redondeo en los modelos de Regresión polinomial, codifique las variables independientes de modo que los términos de primer, segundo y de orden superior para una variable x particular no estén altamente correlacionados.
5. Para reducir los errores de redondeo y estabilizar los coeficientes de Regresión, utilice métodos de Regresión sesgada para estimar los parámetros β.

1.3. Métodos de Regresión sesgada

Son métodos que tienen la finalidad de minimizar la suma de cuadrados del error (SCE) sujeta a restricciones sobre valores posibles de los estimadores para la reducción de su varianza, generando predicciones más precisas.

1.3.1. Regresión por Componentes Principales

La Regresión de Componentes Principales es otra alternativa para la solución al problema de colinealidad. Es un método introducido por Massy (1965) que aplica mínimos cuadrados sobre un conjunto de variables latentes llamados Componentes Principales, obtenidos a partir de la matriz de correlación.

Consiste en eliminar de la consideración aquellas dimensiones del espacio X que están causando el problema de colinealidad. Esto es similar, en el concepto, a dejar caer una variable independiente del modelo cuando hay dispersión insuficiente en esa variable para aportar información significativa sobre Y. Sin embargo, en la Regresión de los Componentes Principales, la dimensión reducida de la consideración se define por una combinación lineal de las variables más que por una sola variable independiente. (Rawlings, 2005)

El objetivo del análisis de Componentes Principales es encontrar una transformación lineal de un conjunto de n variables **X** en un nuevo conjunto denotado por **P**, donde el nuevo conjunto tiene ciertas propiedades deseables. Mendieta (1992)

Estas propiedades, que proporcionan la razón para usar las p en lugar de las x originales, son:

- Los elementos de ***P*** no están correlacionados entre sí en la muestra (ortogonalidad); y
- Cada elemento de **P**, que avanza de $p_1 \dots p_n$, representa la mayor parte de la varianza combinada de las variables predictoras **X** como sea posible, consistente con ser ortogonal a los p anteriores.

Se considera al modelo de Regresión lineal estándar $Y = X\beta + \varepsilon$, donde Y es un vector respuesta de nx1, $X = [X_1, \dots, X_p]$ es nxp matriz de rango completo de variables predictoras, β es un vector de px1 parámetros, y ε es un vector n x 1 de errores aleatorios no correlacionados y normalmente distribuidos con la expectativa 0 y la varianza común.

La matriz de correlación de **X** definida como $X'X$, y $X'Y$ es el vector de correlación entre **X** y **Y**.

El estimador mínimo cuadrado de β es $\hat{\beta} = (X'X)^{-1}X'Y$

Sea $V = [V_1, \dots, V_p]$ la matriz de tamaño pxp cuyas columnas son los autovectores normalizados de $X'X$, y sea $\lambda_1, \dots, \lambda_p$ los correspondientes autovalores.

Sea $W = [W_1, \dots, W_p] = XV$. Entonces $W_j = XV_j$ es el j.ésimo componente principal de X. A continuación, algunas propiedades importantes de los Componentes Principales.

- $V'V = VV'$ *la matriz* V *es ortogonal*
- $W'W = \Lambda$, donde $\Lambda = diag(\lambda_1, \dots, \lambda_p)$, W es ortogonal y $|W| = \sqrt{\lambda_j}$
- $X = W'V$ y $X_j = \sum_{k=1}^{p} v_{jk} W_k$

Según lo visto, el modelo de regresión puede ser escrito de esta manera:

$$Y = X\beta + \varepsilon = XV$$

$$V'\beta + \varepsilon = W\gamma + \varepsilon, donde\ \gamma = V'\beta$$

Bajo esta formulación, el estimador de mínimos cuadrados de γ es:

$$\hat{\gamma} = (W'W)^{-1}W'Y = \Lambda^{-1}W'Y \qquad (2.1)$$

Y por lo tanto, el estimador de Componentes Principales de β se define por:

$$\tilde{\beta} = V\hat{\gamma} = V\Lambda^{-1}W'Y \qquad (2.2)$$

Si todos los Componentes Principales se usan en (2.1) entonces $\tilde{\beta} = \hat{\beta}$. Sin embargo, en la práctica, sólo un subconjunto $W_{(s)} = [W_1, \dots, W_s]$, de los Componentes Principales se utiliza en la estimación de γ .

$$\hat{\gamma} = (W_{(s)}{}' W_{(s)})^{-1} W_{(s)}{}' Y = \Lambda_{(s)}{}^{-1} W_{(s)}{}' Y,$$

Y por lo tanto, (2.2) se puede reescribir como:

$$\tilde{\beta} = V_{(s)} \Lambda_{(s)}{}^{-1} W_{(s)}{}' Y \qquad (2.3)$$

Se puede demostrar que $var(\hat{\beta}) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}$ y $var(\tilde{\beta}) = \sigma^2 \sum_{j=1}^{s} \frac{1}{\lambda_j}$, donde $var(\hat{\beta})$ y $var(\tilde{\beta})$ se refieren a la traza de las correspondiente matriz de varianza y covarianza. Se deduce que $var(\tilde{\beta}) < var(\hat{\beta})$, pero por otro lado, $E(\tilde{\beta}) \neq \beta$, es decir, el estimador de componente principal está sesgado. La clase de los denominados "métodos de Regresión sesgada". También se puede demostrar que el error cuadrático medio de lJ, (la varianza más el sesgo al cuadrado) está dado por:

$$MSE(\tilde{\beta}) = \sigma^2 \sum_{j=1}^{s} \frac{1}{\lambda_j} + \sum_{j=s+1}^{p} \gamma_j{}^2$$

1.3.2. Regresión Ridge:

La Regresión Ridge es método propuesto por Hoerl y Kennard (1970), usado para trabajar con modelos que presentan sesgo y que nos permiten detectar la multicolinealidad dentro de un modelo de regresión del tipo $Y = X\beta + \varepsilon$. (Piña, 2007)

Este método constituye una alternativa a la estimación mínimo-cuadrática, debido a que proporciona una evidencia gráfica de los efectos de la colinealidad en la

estimación de los coeficientes de regresión, dicho procedimiento se encuentra dentro del grupo de las regresiones parciales o sesgadas, consideradas como no lineales. (López, 1998)

Según James (2013), La Regresión Ridge es muy similar a los mínimos cuadrados, excepto que los coeficientes de Ridge son calculados minimizando una cantidad ligeramente diferente. (Encoge los coeficientes estimados hacia cero). Los coeficientes de Regresión Ridge son definidos como:

$$RSS\ (\lambda) = (y + X\beta)'\ \ (y + X\beta) +\ \lambda\beta'\beta$$

$$\hat{\beta}^{ridge} = (X'X +\ \lambda\, I)^{-1}\, X'Y$$

Donde: $\lambda\ \geq 0$, es un parámetro de ajuste, que controla la intensidad del término de penalización y I es la matriz identidad. Se observa que:

- cuando $\lambda = 0$, obtenemos la estimación de Regresión lineal
- cuando $\lambda = \infty$, obtenemos $\hat{\beta}^{ridge} = 0$
- Cuando λ se encuentra en el intermedio de los dos casos mencionado anteriormente, estamos equilibrando dos ideas, “ajustando un modelo lineal de Y sobre X y reduciendo su coeficiente”

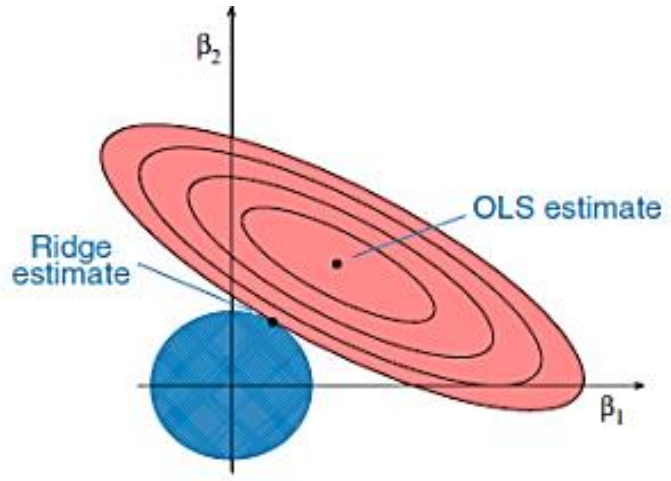

Figura 1. *Representación geométrica de la estimación MCO y Ridge*

Interpretación geométrica de la Regresión: Las elipses corresponden a los contornos de la suma residual de los cuadrados (RSS): la elipse interior tiene RSS más pequeño, y RSS se minimiza en las estimaciones MCO. (James, 2015)

En general, Regresión Ridge produce predicciones más precisas que los modelos obtenidos por MCO + selección "clásica" de variables.

Sin embargo, si bien al aumentar λ (mayor penalización) los coeficientes estimados se contraen hacia cero, ninguno de ellos vale exactamente cero por lo cual no se produce selección de variables. Todas las variables originales permanecen en el modelo final.

a) Selección de modelos

Estos indicadores ajustan el error de entrenamiento de acuerdo al tamaño del modelo y pueden ser usados para elegir entre un conjunto de modelos con diferente número de variable

- $\boldsymbol{R^2}$ **ajustado** : Para un modelo con d variables se define R^2 ajustado por:

$$R^2aj = 1 - \frac{RSS\ /\ (n-d-1)}{TSS\ /(n-1)}$$

A diferencia de los otros indicadores un valor grande de R^2aj indica un modelo con menor error de prueba. El R^2ajustado paga un precio por la inclusión de variables irrelevantes en el modelo. (López, 2014)

b) Validación cruzada:

Es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar cómo de preciso es un modelo que se llevará a cabo a la práctica. (Devijver, 1982). Es una técnica muy utilizada en proyectos de inteligencia artificial para validar modelos generados.

Validación cruzada de K iteraciones:

En la validación cruzada de K iteraciones o K-fold cross-validation los datos se dividen en K subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento. El proceso de validación cruzada es repetido durante K iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado. (FH Joanneum, 2006)

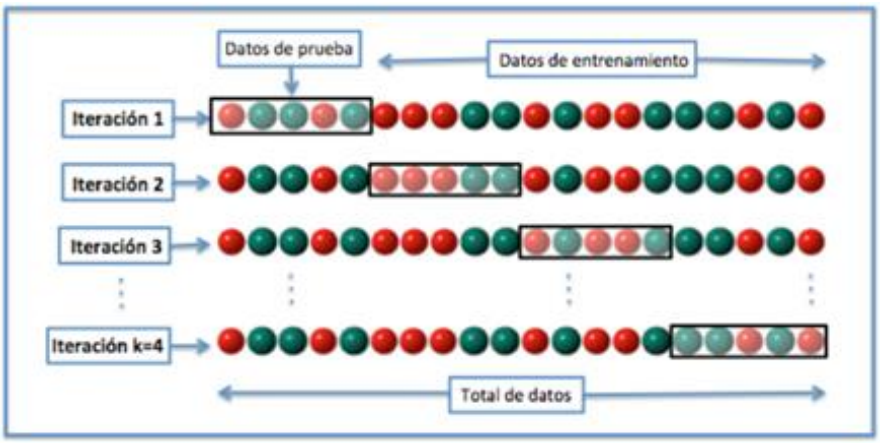

Figura 2. *Validación cruzada de K iteraciones con K=4*

Error de la validación cruzada de K iteraciones: En cada una de las k iteraciones de este tipo de validación se realiza un cálculo de error. El resultado final lo obtenemos a partir de realizar la media aritmética de los K valores de errores obtenidos, según la fórmula:

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i$$

Es decir, se realiza el sumatorio de los K valores de error y se divide entre el valor de K.

Validación cruzada aleatoria

Este método consiste al dividir aleatoriamente el conjunto de datos de entrenamiento y el conjunto de datos de prueba. Para cada división la función de aproximación se ajusta a partir de los datos de entrenamiento y calcula los valores de salida para el conjunto de datos de prueba. El resultado final se corresponde a la media aritmética de los valores obtenidos para las diferentes divisiones. (Moore,s.f.)

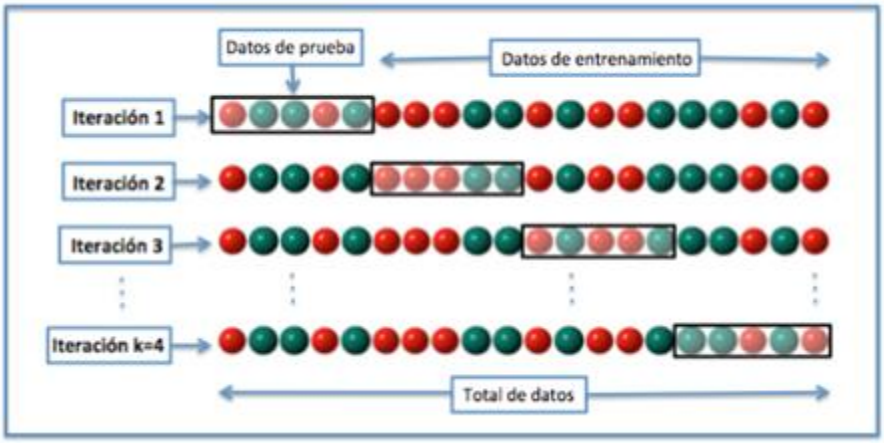

Figura 3. *Validación cruzada de K iteraciones con K=4*

Error de la validación cruzada aleatoria: En la validación cruzada aleatoria a diferencia del método anterior, cogemos muestras al azar durante k iteraciones, aunque de igual manera, se realiza un cálculo de error para cada iteración.

El resultado final también lo obtenemos a partir de realizar la media aritmética de los K valores de errores obtenidos, según la misma fórmula:

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i$$

Validación cruzada dejando uno fuera:

La validación cruzada dejando uno fuera o Leave-one-out cross-validation (LOOCV) implica separar los datos de forma que para cada iteración tengamos una sola muestra para los datos de prueba y todo el resto conformando los datos de entrenamiento.

La evaluación viene dada por el error, y en este tipo de validación cruzada el error es muy bajo, pero en cambio, a nivel computacional es muy costoso, puesto que se tienen que realizar un elevado número de iteraciones, tantas como N

muestras tengamos y para cada una analizar los datos tanto de entrenamiento como de prueba. (Elkan, 2011)

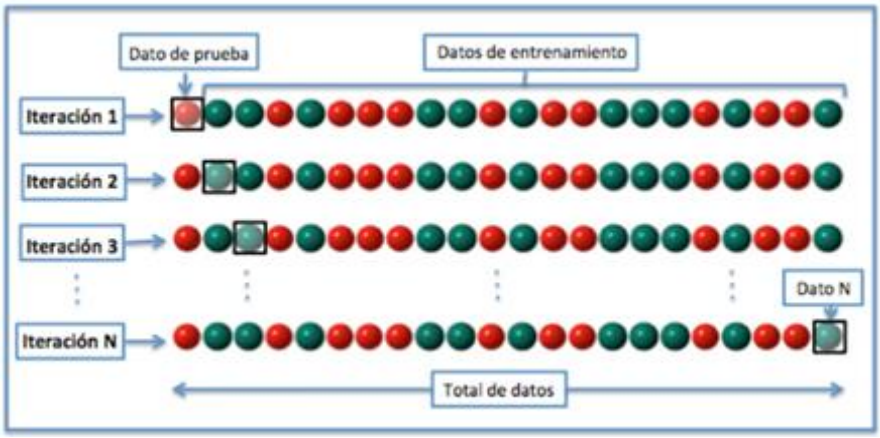

Figura 4. *Validación cruzada dejando uno fuera (LOOCV)*

Error de la validación cruzada dejando uno fuera: En la validación cruzada dejando uno fuera se realizan tantas iteraciones como muestras (N) tenga el conjunto de datos. De forma que para cada una de las N iteraciones se realiza un cálculo de error. El resultado final lo obtenemos realizando la media aritmética de los N valores de errores obtenidos, según la fórmula:

$$E = \frac{1}{N}\sum_{i=1}^{K} E_i$$

Donde se realiza el sumatorio de los N valores de error y se divide entre el valor de N.

1.3.3. Least Absolute Shrinkage and Selection Operator (LASSO)

Con la finalidad de encontrar una técnica de Regresión lineal que brinde un modelo parsimonioso de fácil interpretación y que a su vez tenga la propiedad de contraer los coeficientes del modelo, logrando estabilizar las estimaciones y

predicciones, Tibshirani (1996) propuso una técnica que combina la selección de variables y la penalización de los parámetros llamada LASSO.

LASSO, al igual que Ridge, encoge algunos coeficientes pero a su vez establece otros a 0, con lo cual hace una especie de selección de variables en forma continua debido a la norma L_1 , reteniendo las buenas características.

En los últimos años en el avance de la investigación y aplicación de las técnicas LASSO, se debe principalmente a la existencia de problemas donde p >> n. (Tibshirani, 2011)

LASSO resuelve el problema de mínimos cuadrados con restricción sobre la norma L1 del vector de coeficientes.

$$\hat{\beta}^{lasso} = min\left\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2\right\} \quad s.a \sum_{j=1}^{p}|\beta_j| \leq s$$

O de forma equivalente, minimizando

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

Siendo s y $\lambda \geq 0$ los parámetros de penalización por complejidad. Para la determinar λ se utilizará la técnica de validación cruzada (cross validation). A medida que λ aumenta, la varianza disminuye y el sesgo aumenta. Además si $\lambda \to 0$, la estimación será muy parecida a la de mínimos cuadrados.

La diferencia entre la Regresión de Ridge y la LASSO se puede ver desde la forma diferente de las funciones de penalidad representadas. La Regresión Ridge

impone una penalización cuadrática que tiene un impacto muy fuerte en los valores de los grandes coeficientes pero una pequeña penalización para los valores cercanos a cero. Por el contrario, la penalización del valor absoluto para el LASSO aumenta a una velocidad más lenta para valores de coeficientes grandes, pero se aleja de cero más rápido para coeficientes cercanos a cero. Por consiguiente, esperamos que el comportamiento deseable sea que los coeficientes pequeños se contraigan más fuertemente hacia cero, mientras que los coeficientes más grandes serán menos afectados por la penalización. (Fahrmeir, 2013)

a. La propiedad de selección de variable LASSO

En la figura 1 la solución de mínimos cuadrados está marcada como β, mientras que el diamante y el círculo azul, representan las restricciones de Regresión de LASSO y Ridge respectivamente. Si s es suficientemente grande, entonces las regiones de restricción contendrán β, por lo que las estimaciones de Regresión de Ridge y LASSO serán las mismas que las estimaciones de mínimos cuadrados. (James, 2015)

Sin embargo, en las estimaciones de mínimos cuadrados están fuera del diamante y del círculo, por lo que las estimaciones de mínimos cuadrados no son las mismas que las estimaciones de LASSO y de Regresión de Ridge.

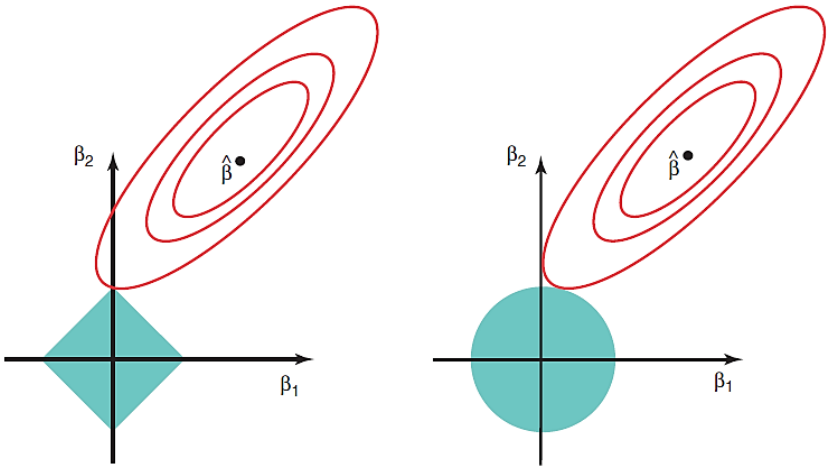

Figura 5. *Representación geométrica de la Regresión Ridge y Lasso*

Contornos de las funciones de error y restricción para LASSO (izquierda) y Regresión de Ridge (derecha) para p=2. Las áreas azul sólidas son las regiones de restricción,$|\beta_1| + |\beta_2| \leq s$ y $\beta_1^2 + \beta_2^2 \leq s$, mientras que las elipses rojas son los contornos del RSS. (James, 2015)

Las elipses que están centradas alrededor de $\hat{\beta}$ representan regiones de constante RSS. En otras palabras, todos los puntos de una elipse comparten un valor común del RSS. A medida que las elipses se expanden lejos de las estimaciones del coeficiente de mínimos cuadrados, el RSS aumenta.

Desde que la Regresión Ridge tiene una restricción circular sin puntos afilados, esta intersección no generalmente ocurrirá en un eje, y por lo que las estimaciones de los coeficientes de Regresión Ridge serán exclusivamente distinto de cero. Sin embargo, la restricción del LASSO tiene esquinas en cada uno de los ejes, por lo que la elipse a menudo intersecará la región de restricción en un eje. Cuando esto ocurre, uno de los coeficientes será igual a cero. (James, 2015)

1.4. Variables de estudio

1.4.1. Bolsa de Valores

El BCRP (2011) define a la bolsa de valores como "Mercado organizado en el que se negocia públicamente la compra y la venta de títulos de renta fija y variable (acciones, obligaciones, etc.), bienes, materias primas, etc. Las bolsas facilitan y regulan los cambios comerciales y ofrecen un magnífico medio para conocer las condiciones del mercado. Los bienes que se negocian en las bolsas deben reunir las características de estandarización, fungibilidad y abundancia como para negociarlos con fluidez."

a) Funciones

La Bolsa de Valores de Lima tiene, entre otros, las siguientes funciones:

- Proporcionar a los participantes del mercado los locales, sistemas y mecanismos que les permitan, en sus diarias negociaciones, disponer de información transparente de las propuestas de compra y venta de los valores, la imparcial ejecución de las órdenes respectivas y la liquidación eficiente de sus operaciones.
- Fomentar las negociaciones de valores, realizando las actividades y brindando los servicios para ello, de manera de procurar el desarrollo creciente del mercado.
- Inscribir, con arreglo a las disposiciones legales y reglamentarias, valores para su negociación en Bolsa, y registrarlos.

- Ofrecer información al público sobre los Agentes de Intermediación y las operaciones bursátiles.
- Divulgar y mantener a disposición del público información sobre la cotización de los valores, así como de la marcha económica y los eventos trascendentes de los emisores. (BVL, 2017)

b) Importancia de las Bolsas de Valores

Los recursos invertidos por medio de las Bolsas de Valores permiten tanto a las empresas como a los gobiernos, financiar proyectos productivos y de desarrollo que generan empleos y riqueza para un país. Los oferentes de recursos de estos recursos reciben a cambio la oportunidad de invertir en una canasta de instrumentos que les permite diversificar su riesgo, optimizando sus rendimientos.

Es importante destacar que las Bolsas de Valores son mercados complementarios al Sistema Financiero tradicional. (Valle, 2011)

c) Factores que influyen en la bolsa de valores

Calzada (2016) nos dice que existen muchos factores que influyen el movimiento de la bolsa de valores, pero principalmente son factores económicos y políticos los que mueven los indicadores de los mercados, tal es el caso de la bolsa de New York y la de China, ya que un movimiento relevante que se de en Wall Street repercutirá en las bolsas a nivel global.

Algunas de las variables más importantes y que tienen un peso específico en el movimiento de las bolsas de valores y son las siguientes: PIB, Empleo, Inflación, Tipo de cambio, Tasa de interés y Precio de la materia prima.

1.4.2. Índice General de la Bolsa de Valores de Lima (IGBVL)

El índice muestra el comportamiento a largo plazo de los precios de las principales acciones, las cuales están inscritas en Bolsa, considerando una cartera seleccionada. En la actualidad esta cartera está conformada por 32 activos más negociados del mercado.

Para obtener este indicador se toma en cuenta las variaciones en los precios y los dividendos o acciones liberadas repartidas, así como la suscripción de acciones. La base para el cálculo es el 30 de diciembre de 1991 = 100. El 2 de enero y el primero de julio de realiza una revisión constante con el objetivo de mantener actualizada la cartera (BVL, 2017).

1.4.3. Los índices bursátiles

Según el Banco Central de Reserva del Perú (2011), los índices bursátiles son indicadores que expresan la tendencia promedio de los valores más representativos de un mercado bursátil.

a) Función

Vallle (2011) nos dice que la función de los índices bursátiles es medir el comportamiento del mercado al que representan y compararlo con la evolución de un valor o una cartera de valores determinada.

Los índices bursátiles son una referencia cada vez más importante para los gestores de cartera. Lo son también en la oferta de nuevos productos, sobre todo en depósitos y fondos. Se habla más de ellos que de los mercados a los que representan.

b) Los más importantes

- **Standard & Poor´s 500 (S&P 500):** Precio de cierre diario ponderado basado en las 500 compañías más representativas de la Bolsa neoyorquina.
- **National Association of Securities Dealers Automated Quotation (NASDAQ):** Precio de cierre diario ponderado compuesto por compañías del rubro tecnológico.
- **Dow Jones:** Precio de cierre diario ponderado compuesto por 30 de las acciones más significativas de compañías industriales (servicios financieros, tecnología, minoristas, entretenimiento y bienes del consumidor), que cotizan en la bolsa de New York.
- **Nihon Keizai Shinbun (NIKKEI 225):** Precio de cierre diario ponderado, compuesto por 225 acciones que cotizan en la bolsa de Tokio.

Además se consideraron otras variables para el estudio por tratarse de acciones correspondientes a las principales empresas del rubro tecnológico y que están vinculados linealmente a la rentabilidad al mercado al cual pertenecen.

- **Apple:** Precio de cierre diario ponderado de las acciones de la empresa multinacional estadounidense que diseña produce equipos electrónicos.
- **Microsoft:** Precio de cierre diario ponderado de las acciones de la empresa estadounidense, dedicada al sector software y hardware.
- **International Business Machines (IBM):** Precio de cierre diario ponderado de las acciones de la empresa estadounidense de tecnología y consultoría relacionado a la informática.

- **Amazon:** Precio de cierre diario de las acciones de la compañía estadounidense dedicada al comercio electrónico.
- **AT&T:** Precio de cierre diario de las acciones de la compañía estadounidense de telecomunicaciones.
- **Hewlett Packard (HP):** Precio de cierre de las acciones de la compañía dedicada a la tecnología de información.

II. MÉTODOS Y MATERIALES

2.1. Tipo de investigación

Investigación de tipo analítico - correlacional.

2.2. Diseño de investigación

Diseño no experimental

2.3. Población y muestra

2.3.1. Población

La población está constituida por las series temporales de los precios diarios de los índices bursátiles y el Índice Bursátil de la Bolsa de Valores de Lima. Se consideraron 11 procesos estocásticos.

2.3.2. Muestra

Conformada por los valores diarios registrados en las distintas series bursátiles en el periodo comprendido entre los años 2000 al 2014, la cual está constituida por 3,773 observaciones.

Tal como se mencionó en el apartado del marco teórico 1.4.2 y 1.4.3, las variables tomadas para presente investigación fueron las siguientes:

Tabla 2. *Variables de estudio.*

N°	Abreviatura	Variables de estudio	Tipo de variable
1	IGVBL*	Índice General de la Bolsa de Valores de Lima	Respuesta
2	-	Apple	Explicativa
3	-	Microsof	Explicativa
4	-	Amazon	Explicativa
5	AT&T	American Telephone and Telegraph	Explicativa
6	IBM	International Business Machines	Explicativa
7	AP500	Standard & Poor´s 500	Explicativa
8	DOW	Dow Jones	Explicativa
9	NIKKEI	Nihon Keizai Shinbun	Explicativa
10	HP	Hewlett Packard	Explicativa
11	NASDAQ	National Association of Securities Dealers Automated Quotation	Explicativa

Elaboración propia

2.4. Técnicas e instrumentos de recolección de datos

No se aplicó instrumentos de recolección de datos, puestos que los mismos fueron obtenidos de fuentes disponibles de la red.

Los indicadores bursátiles diarios se obtuvieron de los registros financieros recuperados de la página web: finance.Yahoo.com. y el indicador del IGBVL se obtuvo a partir de la página web de la Bolsa de Valores de Lima.

2.5. Análisis estadístico de los datos

Inicialmente se realizó el análisis exploratorio con la finalidad de encontrar las relaciones lineales existentes entre las variables de estudio para la cual se hizo uso de gráficos de matriz de dispersión y matriz de correlaciones.

Luego, mediante indicadores como el Factor de Inflación de la Varianza (FIV), índice de tolerancia e índice de condición, se confirmó la presencia de multicolinealidad en las variables de estudio.

Posteriormente, se construyeron modelos en función de los índices bursátiles, tales como el modelo clásico de mínimos cuadrados ordinarios, Regresión por Componentes Principales, Regresión Ridge y LASSO.

Finalmente se determinó cuál de los modelos propuestos genera una menor varianza ante presencia de multicolinealidad y supera al modelo clásico de MCO.

Para el procesamiento de los datos se utilizó el software R- Studio y para la edición de tablas de resultados, la hoja de cálculo de Excel 2013.

III. RESULTADOS

Con la finalidad de crear el modelamiento lineal en relación de los índices bursátiles mundiales y el Índice General de Bolsa de Valores de Lima (IGBVL) se procedió a determinar la existencia de colinealidad entre las variables regresoras.

En la **tabla 3** se muestran los indicadores generales; la determinante de la matriz de diseño $X'X$ dio como resultado 0.000. En este caso la matriz $X'X$ es casi singular, es decir que su determinante no es 0 pero es muy pequeña. Como para invertir una matriz hay que dividir por su determinante, en esa situación surgen problemas de precisión en la estimación de los coeficientes. Por consiguiente, a la hora de plantear modelos de Regresión conviene estudiar previamente la existencia de colinealidad.

Otro indicador general de multicolinealidad es el índice de condición, que se define como la raíz cuadrada del cociente del máximo valor propio entre el mínimo valor propio; puesto que el resultado se encuentra entre los límites de 20 y 30, se considera la presencia de multicolinealidad moderada.

Y para los indicadores individuales de multicolinealidad, el VIF tiene como límite establecido el valor de 10 por lo cual se muestra que las únicas variables que no superan el límite indicado son Microsoft, Dow Jones y Nikkei.

Los resultados encontrados muestran que sería un problema realizar estimaciones de modelos de Regresión de Mínimos Cuadrados usando estas variables que son altamente colineales y como consecuencia genera problemas en la inflación de la varianza y en las estimaciones de los parámetros, llevando como consecuencia a una mala inferencia estadística.

Tabla 3

Indicadores para la determinación de multicolinealidad en las variables regresoras

Variables regresoras	Indicadores individuales		Indicadores generales	
	VIF	**Tolerancia**	**Determinante (D)**	**Índice de condición IC (λ_i)**
Apple	17.375*	0.058*	0.0000*	24.823
Microsoft	6.139	0.163		
AT&T	14.094*	0.071*		
IBM	19.614*	0.051*		
Amz	13.369*	0.075*		
Sp500	52.586*	0.019*		
Nasdaq	28.749*	0.035*		
Dow Jones	8.631	0.116		
Nikkei	6.560	0.152		
Hp	13.421*	0.075*		

Elaboración propia

En consecuencia del análisis realizado de la prueba de multicolinealidad, se detectó que las variables están altamente correlacionadas (ver anexo 1), por ende este efecto causaría problemas en la estimación de los coeficientes de Regresión.

3.1. Mínimo Cuadrados Ordinarios (MCO)

En la **tabla 4** se observa las estimaciones realizadas por el método de mínimos cuadrados ordinarios. Todas las estimaciones de los coeficientes de Regresión resultaron ser significativas a excepción de la variable SP500 el cual obtuvo un p valor de 0.06250.

La presencia de multicolinealidad tiende a aumentar las estimaciones de los betas y sus respectivas varianzas, otra consecuencia es un $\boldsymbol{R^2}$ muy alto, en este caso se obtuvo un valor de 0.94690 y esto es resultado del efecto de inflación de varianza que tienen las variables regresoras. Por otra parte la medida general para comparar la variabilidad del modelo MCO se consideró a la varianza efectiva que es una medida de generalización de la media geométrica, siendo su valor de 0.00037.

Tabla 4

Estimación de coeficientes mediante la Regresión Mínimos Cuadrados Ordinarios

Variables regresoras	Coeficiente	Std. Error	R^2	MSE	Varianza efectiva
Constante	11.725	0.46			
Apple	0.804	0.011			
Microsoft	-1.179	0.043			
AT&T	0.654	0.051			
IBM	-0.453	0.048			
Amazon	0.283	0.015	0.9469	0.06225	0.00037
SP500	0.266	0.143			
Nasdaq	-1.218	0.069			
Dow Jones	0.534	0.037			
Nikkei	0.489	0.043			
HP	-0.282	0.021			

Elaboración propia

3.2. Métodos de contracción o regresión penalizada

3.2.1. Regresión por Componentes Principales (RCP)

Una de las propuestas planteadas para el tratamiento y control de las variables colineales; es la Regresión por Componentes Principales (RCP). Un previo al análisis al RCP es la identificación del número de componentes a utilizar en el modelo.

La decisión para la selección de número de Componentes está en base al análisis de autovalores mayores a 1 y tal como se puede observar en la **tabla 5**, el componente PC1 posee un autovalor de 6.7471 siendo esta variable una combinación lineal de las variables regresoras originales que a su vez explica un 67.47% de la variabilidad total de las predictoras. Así mismo se observa en general que el número de Componentes es igual al número de variables que contribuyen al modelo; cada una con su respectiva varianza explicada y acumulada.

Tabla 5

Componentes Principales y varianza explicada

Componentes	Autovalor	Varianza	Varianza acumulada
PC1	6.7471	0.6747	0.6747
PC2	1.7317	0.1732	0.8479
PC3	0.6335	0.0634	0.9113
PC4	0.3820	0.0382	0.9495
PC5	0.2388	0.0239	0.9733
PC6	0.1257	0.0126	0.9859
PC7	0.0551	0.0055	0.9914
PC8	0.0417	0.0042	0.9956
PC9	0.0334	0.0033	0.9989
PC10	0.0109	0.0011	1.0000

Elaboración propia

Para la estimación del modelo con Componentes Principales, se consideró a utilizar dos Componentes como variables explicativas para el modelo. Los dos Componentes elegidos rescatan el 84.79% de la información de las variables explicativas originales.

En la **tabla 6** se observa la estimación de los coeficientes de Regresión para los Componentes Principales. El análisis individual de los coeficientes de las variables y la constante resultaron ser muy significativas. El modelo tiene un coeficiente de determinación alto, verificando que los Componentes explican en un 71.27% la variabilidad del IGBVL; y su varianza efectiva es de 3.93666E-05.

Tabla 6

Estimación de coeficientes mediante el método de Regresión por Componentes Principales

Componentes	Coeficiente	Std. Error	R^2	MSE	Varianza efectiva
Constante	8.8174	0.0095			
PC1	0.3039	0.0036	0.7127	0.3368	3.93666E-05
PC2	-0.3501	0.0072			

Elaboración propia

3.2.2. Regresión Ridge

El primer paso para realizar la Regresión Ridge, es encontrar aquel valor de lambda que minimiza la suma de cuadros del modelo de Regresión ordinario. La técnica utilizada para encontrar dicho valor es la validación cruzada.

En el **Figura 6** se muestra la secuencia de un conjunto de lambdas calculados en logaritmo y el efecto de contracción que tienen en los coeficientes a medida que el log lambda aumenta.

Se puede apreciar que a medida que lambda aumenta, los coeficientes de Regresión tienden a converger hacia cero. Pese a esto, los coeficientes de la Regresión Ridge nunca son iguales a cero.

En el **Figura 7** se muestra la selección de lambda que minimiza el error cuadrático de la validación cruzada, siendo su valor de λ =0.1109698.

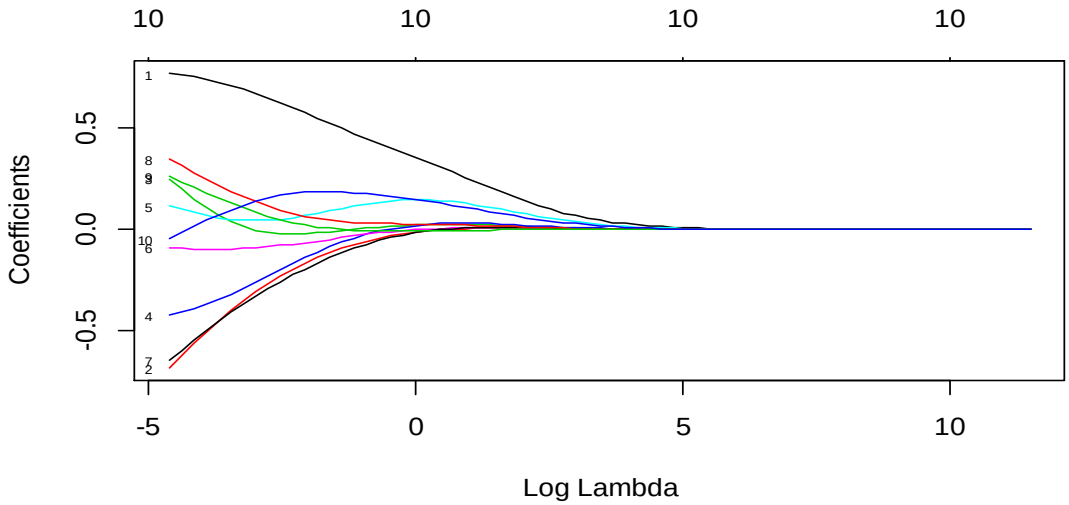

Figura 6. *Encogimiento de coeficientes de Regresión según la función de lambda*

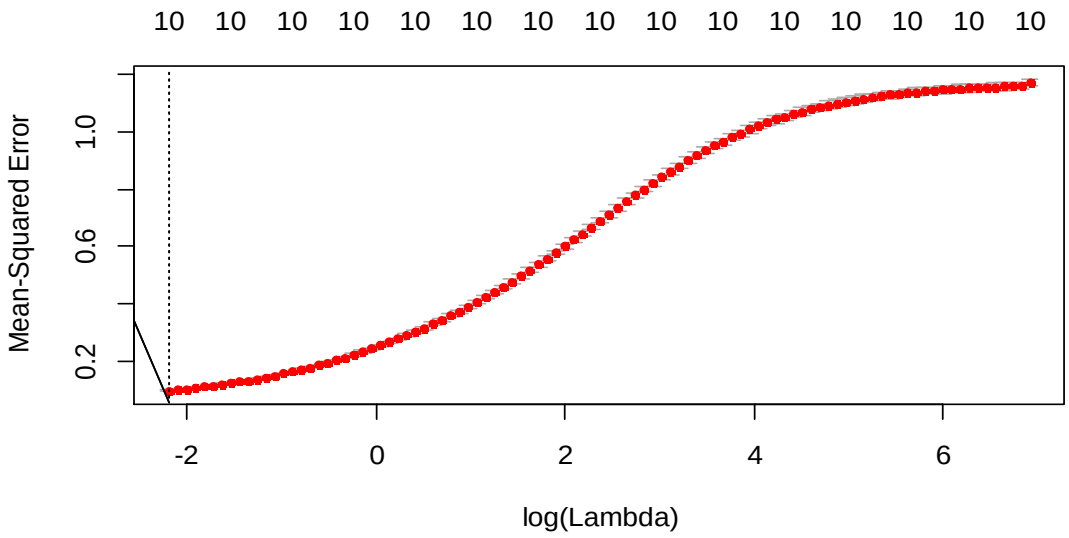

Figura 7. *Selección de lambda mediante validación cruzada*.

Al principio, el error cuadrático medio es muy alto y luego a medida que el valor de lambda disminuye, el error cuadrático también lo hace hasta en algún punto donde se hace mínimo. En la parte superior de la **Figura 7** se muestra el número de variables distinto de cero que hay en el modelo.

Después de haberse encontrado el valor de lambda λ =0.1109698, se procede a la estimación del modelo, en la **Tabla 7** se muestra los resultados de la estimación de coeficientes de Regresión Ridge. A diferencia de las estimaciones obtenidas por mínimos cuadrados, Ridge contrae a las estimaciones como a sus errores estándar.

El coeficiente de determinación obtenida por Ridge es de 90.28% y una varianza efectiva del 0.00034. Puesto que las estimaciones de mínimo cuadrado se hicieron con variables que están afectados por la multicolinealidad; el sesgo que introduce Ridge hace que las estimaciones de los coeficientes se contraigan, sucediendo lo mismo con las varianzas de las estimaciones; este efecto hace que el $\boldsymbol{R^2}$ sufra ese mismo cambio.

Tabla 7

Estimación de coeficientes mediante el método de la Regresión Ridge

Variables regresoras	Coeficiente	Std. Error	R^2	MSE	Varianza efectiva
Constante	9.484	0.327	0.90284	0.11385	0.00034
Apple	0.593	0.011			
Microsoft	-0.182	0.042			
AT&T	-0.02	0.042			
IBM	-0.154	0.043			
Amazon	0.065	0.015			
SP500	-0.067	0.119			
Nasdaq	-0.209	0.061			
Dow Jones	0.075	0.032			
Nikkei	0.03	0.041			
HP	0.183	0.021			

Elaboración propia

3.2.3. Regresión LASSO

Del mismo modo que la Regresión Ridge, para la estimación de la Regresión LASSO también es necesario de encontrar el valor de lambda mediante validación cruzada.

En la **Figura 8** se muestra la relación de log lambda y el encogimiento de los coeficientes de Regresión; a diferencia de Ridge, LASSO puede hacer selección de variables, esto es, algunos coeficientes se hacen cero.

En la **Figura 9**, se muestra que el error cuadrático medio de la validación cruzada comienza siendo demasiado alto, y a medida que el lambda va disminuyendo, llega a un punto en donde el MSE es el mínimo y ese valor está asociado a un valor de lambda óptimo ($\lambda = 0.003160459$). Además se observa en la parte superior del gráfico que el modelo realiza una selección de variables.

A continuación, se observa las estimaciones obtenidas por la Regresión LASSO. Puesto que LASSO realiza una penalización en la norma L1, y esto da la posibilidad de que un coeficiente se haga cero.

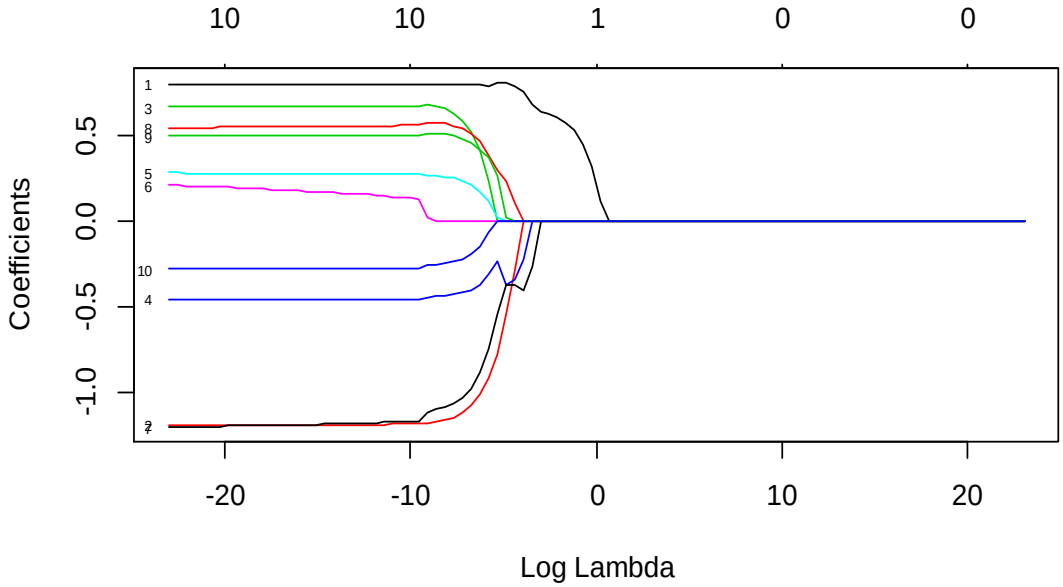

Figura 8. *Encogimiento de coeficientes de Regresión según la función de lambda*

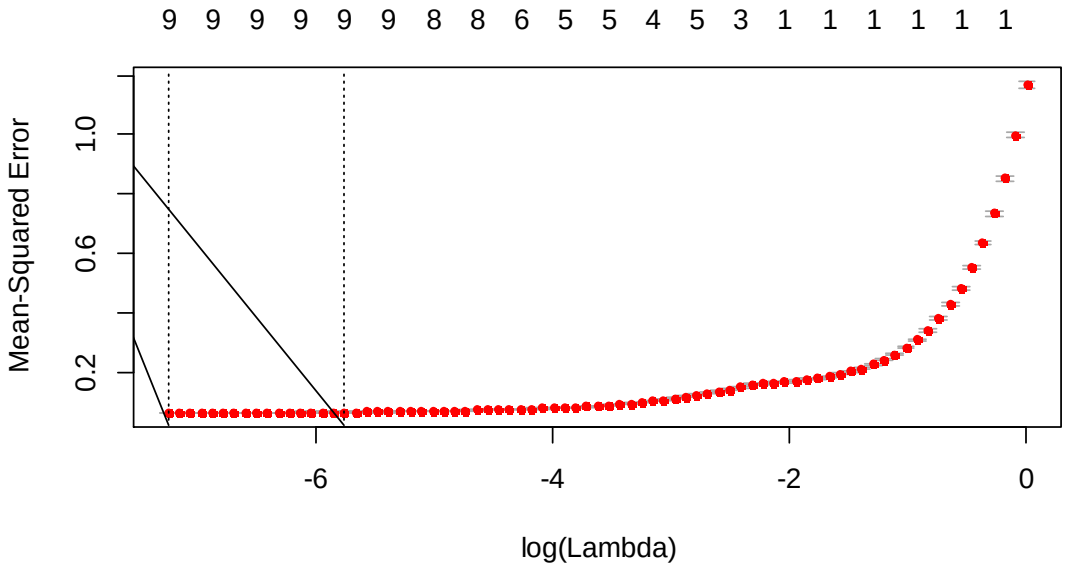

Figura 9. *Selección de lambda mediante validación cruzada para LASSO*

En el análisis grafico se determinó que LASSO hace una selección de variable, en este caso, la variable que queda fuera del modelo es SP500, siendo esta variable la que

obtuvo mayor VIF y ser poco significativa. **(Ver Tabla 3)**. El coeficiente de determinación obtenido por el modelo es de 94.12%, siendo este valor muy parecido al que se obtuvo por el MCO, y un valor de varianza efectiva de 0.00072.

Tabla 8

Estimación de coeficientes mediante el método de la Regresión LASSO

Variables regresoras	Coeficiente	Std. Error	R^2	MSE	Varianza efectiva
Constante	11.197	0.407	0.94118	0.06893	0.00072
Apple	0.791	0.011			
Microsoft	-0.896	0.043			
AT&T	0.188	0.047			
IBM	-0.294	0.046			
Amazon	0.097	0.015			
Nasdaq	-0.7	0.066			
Dow Jones	0.371	0.035			
Nikkei	0.355	0.042			
HP	-0.045	0.021			

Elaboración propia

3.3. Comparación de modelos

Finalmente en la **Tabla 9** se observa las estimaciones de cada coeficiente por los distintos métodos propuestos en la presente investigación. Los resultados por MCO fueron obtenidos en presencia del problema de multicolinealidad; para dar solución a ese problema se aplicó la primera técnica propuesta que es la estimación Ridge el cual abordó el problema contrayendo los coeficientes de regresión observándose así mismo la reducción de su error estándar y su R^2.

La segunda técnica aplicada fue LASSO, el cual al igual que Ridge, también generó una reducción en los coeficientes de estimación, sumándose también la propiedad de selección de variables, que en este caso la variable SP500 quedo fuera del modelo debido a su poca contribución.

Y por último la Regresión por Componentes Principales redujo dimensionalidad en las variables generando un modelo más parsimonioso con tal solo dos variables regresoras más la constante; siendo estas variables independientes entre sí cumpliéndose la propiedad de ortogonalidad y podría decirse eliminando el problema de multicolinealidad.

Tabla 9

Comparación de estimaciones de los modelos MCO, Ridge, LASSO y RCP

Estimación	MCO		RIDGE		LASSO		RCP		
	Coeficiente	**Std. Error**	**Coeficiente**	**Std. Error**	**Coeficiente**	**Std. Error**	**Estimación**	**Coeficiente**	**Std. Error**
Constante	11.7252	0.4602	9.4841	0.3274	11.1967	0.4069	Constante	8.8174	0.0095
Apple	0.8037	0.0111	0.5933	0.0110	0.7905	0.0111	PC1	0.3039	0.0036
Microsoft	-1.1794	0.0432	-0.1816	0.0424	-0.8955	0.0429	PC2	-0.3501	0.0072
AT&T	0.6538	0.0507	-0.0200	0.0423	0.1875	0.0473	-	-	-
IBM	-0.4526	0.0480	-0.1539	0.0433	-0.294	0.0461	-	-	-
Amazon	0.2834	0.0148	0.0649	0.0147	0.0967	0.0148	-	-	-
SP500	0.2664	0.1430	-0.0670	0.1194	-	-	-	-	-
Nasdaq	-1.2179	0.0694	-0.2093	0.0611	-0.6996	0.0664	-	-	-
Dow Jones	0.5339	0.0366	0.0752	0.0316	0.3711	0.0347	-	-	-
Nikkei	0.4895	0.0428	0.0301	0.0409	0.3554	0.0421	-	-	-
HP	-0.2824	0.0210	0.1826	0.0207	-0.0447	0.0209	-	-	-
$\boldsymbol{R^2}$	0.9469		0.90258		0.9412		0.7127		
Varianza efectiva	0.0003669		0.0003426		0.0007242		0.0000394		

Elaboración propia

IV. CONCLUSIONES

- Con respecto a los resultados finales de la investigación se llega a la conclusión que los estimadores penalizados o sesgados (Ridge, LASSO y Componentes Principales) generan menor estimación de varianza que los generados por estimados Mínimo Cuadrático Ordinario (MCO) corroborándose los antecedentes de investigaciones pasadas.
- Las variables que corresponde a los Índices Bursátiles mundiales en relación con los Índices General de la Bolsa de Valores de Lima (IGBVL) son datos que están altamente correlacionados; por ende, existe la presencia de multicolinealidad aproximada que podría generar problemas en la inflación de la varianza en los coeficientes de Regresión.
- La estimación por mínimos cuadrados generó un modelo que explica en un 94.69% la variación del Índice general de la Bolsa de Valores de Lima (IGBVL) y con una varianza efectiva de 0.0003669.
- La estimación por Componentes Principales generó un modelo que explica en un 71.27% al Índice general de la Bolsa de Valores de Lima (IGBVL) y obteniendo una varianza efectiva de 0.0000394; además simplificó la complejidad del modelo a tres coeficientes de Regresión. Se tomó como base a dos Componentes Principales que explica un 84.79% de variabilidad del total de variables explicativas, siendo poca la perdida de información.

- Las estimaciones de Ridge y LASSO contraen a los coeficientes de Regresión de tal manera que tratan de controlar el efecto de colinealidad en el modelo minimizando las estimaciones de los betas y sus varianzas. Las estimaciones de Ridge y LASSO generaron un coeficiente de determinación de 90.26% y 94.12% respectivamente y varianza efectiva de 0.0003426 y 0.0007242. Se añade que la Regresión de LASSO generó una selección de variables dejando como resultado un modelo con diez coeficientes de Regresión.
- La técnica que genera menor varianza fue la Regresión por Componentes Principales (0.0000394), Ridge (0.0003426), Mínimo Cuadrados (0.0003669) y por último LASSO (0.0007242).

V. RECOMENDACIONES

El estudio de técnicas de análisis de datos a grandes dimensiones y en particular en estudios econométricos es considerada una de las áreas más dinámicas de investigación en los últimos años, por lo cual en medida de complementar las técnicas expuestas, se recomienda el estudio de la técnica elastic net, que combina los beneficios de la Regresión Ridge y LASSO.

Por otro lado, para personas interesadas en profundizar los temas propuestos en esta investigación, se recomienda considerar el estudio de técnicas de estimaciones robustas del tipo bayesiana, tales como Ridge Generalizado, Ridge Bayesiano, LASSO Bayesiano o Regresión por garrotes negativos, con la finalidad de prevenir la sensibilidad de los mínimos cuadrados ordinarios ante observaciones atípicas.

REFERENCIAS BIBLIOGRÁFICAS

- Aguirre Solís, Jesús J.; Rodríguez Medina, Manuel A.; Piña Monarrez, Manuel R.; (2007). Regresión Ridge y la distribución central t. Ciencia Ergo Sum, julio-octubre, 191-196.
- Andrew W. Moore, Cross-validation for detecting and preventing over fitting, Carnegie Mellon University
- BCRP (2011) "Glosario de términos económicos del Banco Central de Reserva del Perú", recuperado de: http://www.bcrp.gob.pe/publicaciones/glosario/i.html
- Booth G. D. et al. 1994. Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation. US Dept of Agriculture, Forest Service.
- BVL (2017) "Web del la Bolsa de Valores de Lima", recuperado de: http://www.bvl.com.pe/
- Calzada, H. (2016) "¿Qué Factores mueven la bolsa de valores", recuperado de: https://www.rankia.mx/blog/como-comenzar-invertir-bolsa/3127530-que-factores-mueven-bolsa-valores
- Carrasco Carrasco, María (2016). Técnicas de Regularización en Regresión: Implementación y Aplicaciones. (Tesis de pregrado) Universidad de Sevilla, España.
- Charles Elkan, Evaluating Classifiers University of California, San Diego, 18 de enero de 2011.
- Del Valle Moreno, Juan; Guerra Bustillo, C. Walkiria; (2012). La Multicolinealidad en modelos de Regresión Lineal Múltiple. Revista Ciencias Técnicas Agropecuarias, Octubre-Diciembre, 80-83.

- Devijver, P. A., and J. Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall, Londres, 1982.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. and Lautenbach, S. (2013), Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36: 27–46.
- Fahrmeir, Ludwig; Kneib, Thomas; Lang, Stefan; Marx, Brian D. (op. 2013): Regression. Models, methods and applications. Berlin, Heidelberg: Springer.
- FH Joanneum, Cross-Validation Explained, Institute for Genomics and Bioinformatics, 2005-2006.
- Gujarati, Damodar N.; Guerrero, Demetrio Garmendia; Medina, Gladys Arango (2005): Econometría. 4. ed. México [u.a.]: McGraw-Hill.
- Hair, Joseph F.; Gómez Suárez, Mónica (1999): Análisis multivariante. 5ª ed. Madrid: Pearson Prentice Hall.
- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome H. (op. 2016): The elements of statistical learning. Data mining, inference, and prediction. 2nd ed., corrected at 11th printing. New York: Springer (Springer series in statistics).
- James, Gareth (2015): An introduction to statistical learning. With applications in R. Corr. at 6. printing. New York, NY [u.a.], New York, NY [u.a.]: Springer (Springer texts in statistics).
- Judge, George G. (1985): The theory and practice of econometrics. 2. ed. New York [u.a.]: Wiley (Wiley series in probability and mathematical statistics).
- López González, Emelina; (1998). Tratamiento de la colinealidad en Regresión múltiple. Psicothema, 491-507.

- Massy, W.F. (1965) “Principal Components Regression in Exploratory Statistical Research”, Journal of the American Statistical Association, 60: 234–246.
- Mendenhall, William; Sincich, Terry (2012): A second course in statistics. Regression analysis. 7th ed. Boston: Prentice Hall.
- Mendieta, G. (1992) "Regression Modeling Using Principal Components", recuperado de: http://newprairiepress.org/cgi/viewcontent.cgi?article=1408&context=agstatconference.
- Montgomery, Douglas C.; Peck, Elizabeth A.; Vining, G. Geoffrey (2001): Introduction to linear regression analysis. 3rd ed. New York: Wiley (Wiley series in probability and mathematical statistics. Applied probability and statistics).
- Novales Cinca, Alfonso (DL 1993); Econometría. 2ª ed. Madrid: McGraw – Hill
- Tibshirani,R. (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. Series B (Methodological) , 58(1): págs. 267288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the LASSO. A retrospective, Journal of the Royal Statistical Society: Series B (Methodological), 73(3): págs. 273282.
- Paccapelo Valeria, María; (2015). Modelos de Selección Genómica para Caracteres Cuantitativos basados en Marcadores Moleculares aplicados al mejoramiento de maíz. (Tesis de postgrado) Universidad Nacional de Córdova, Argentina.
- Pariasca Luciano, Juan; (1999). Eficiencia de los Estimadores Segados en Regresión, en presencia de Multicolinealidad. (Tesis de pregrado) Universidad Nacional de Ingeniería, Perú.
- Pineda Norman, Amanda; (2013). Una prueba sobre la especificación de un modelo de Regresión. (Tesis de postgrado) Colegio de Postgraduados, México.

- Sopipan (2013): Forescating the financial returns for using multiple regression based on Principal Component Analysis. En: Journal of Mathematics and Statistics 9 (1), pág. 65–71. DOI: 10.3844/jmssp.2013.65.71.
- Rawlings, John O.; Pantula, Sastry G.; Dickey, David A. (2005], c1998): Applied regression analysis. A research tool. 2nd ed. New York: Springer (Springer texts in statistics).
- Valle, M. (2011) "La Bolsa de valores y los principales índices del mundo" - Universidad de Veracruz. Recuperado de: https://www.uv.mx/personal/mvalle/files/2011/08/BOLSA-DE-VALORES-E-INDICES-BURSATILES.pdf

ANEXOS

Anexo 1

Tabla 10

Resumen de los valores y vectores propios obtenidos de la matriz de correlación

VALORES PROPIOS										
[1]	6.74708	1.73175	0.63355	0.38197	0.23880	0.12573	0.05512	0.04168	0.03339	0.01094

VECTORES PROPIOS										
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	-0.33364	-0.30211	-0.07333	0.38876	0.12121	-0.11372	0.15899	0.70970	-0.23908	0.16479
[2,]	-0.34329	0.17831	0.09859	-0.37643	0.15418	-0.81484	-0.02138	0.01097	0.10258	0.01029
[3,]	-0.33917	0.00005	0.42862	0.15119	-0.60435	0.04498	0.33028	0.02213	0.36597	-0.26003
[4,]	-0.33687	-0.28511	0.22253	-0.28816	-0.17329	0.11297	0.06517	-0.33193	-0.70804	0.11893
[5,]	-0.32986	-0.31115	-0.07748	0.02683	0.54095	0.18414	0.41593	-0.36727	0.37388	0.11595
[6,]	-0.34220	0.32041	-0.01235	-0.10723	-0.15663	0.27789	-0.28338	0.07196	0.21907	0.73031
[7,]	-0.34494	0.20812	0.24940	-0.21524	0.37524	0.39524	-0.33889	0.25574	-0.01781	-0.50158
[8,]	-0.26893	0.22045	-0.77921	-0.24313	-0.24253	0.11532	0.26706	0.05370	-0.07123	-0.24780
[9,]	-0.13659	0.65230	0.07494	0.55410	0.17825	-0.06090	0.19730	-0.26897	-0.31040	-0.00999
[10,]	-0.32625	-0.28065	-0.26557	0.42021	-0.13308	-0.15483	-0.61849	-0.32379	0.08687	-0.17631

Elaboración propia

Anexo 2

Tabla 11

Matriz de correlaciones de los índices bursátiles y el IGVBL

Variables Regresoras	APPLE	MICR.	AT&T	IBM	AMAZON	SP500	NASDAQ	DOW	NIKKEI	HP	IGVBL
APPLE	1.00000										
MICR.	0.63400	1.00000									
AT&T	0.74800	0.76500	1.00000								
IBM	0.84400	0.72700	0.83200	1.00000							
AMAZON	0.91600	0.66100	0.67000	0.86800	1.00000						
SP500	0.57800	0.87300	0.79300	0.63400	0.57100	1.00000					
NASDAQ	0.63300	0.88300	0.78800	0.72500	0.68600	0.92000	1.00000				
DOW	0.48600	0.65600	0.43000	0.43300	0.49100	0.76600	0.58300	1.00000			
NIKKEI	0.04700	0.45400	0.33800	-0.05900	-0.01900	0.63900	0.52600	0.40000	1.00000		
HP	0.93800	0.60400	0.70700	0.80000	0.86600	0.59000	0.57100	0.57300	0.05200	1.00000	
IGVBL (*)	0.93400	0.44100	0.62000	0.68800	0.82400	0.44000	0.44400	0.45600	0.01600	0.88400	1.00000

Elaboración propia

(*) Variable dependiente

Anexo 3

Tabla 12

Matriz de varianza y covarianza de los estimadores de Mínimos Cuadrados Ordinarios

Variables regresoras	Cons	Apple	Microsoft	AT&T	IBM	Amazon	SP500	Nasdaq	Dow	Nikkei	HP
Cons	0.21177	-0.00063	0.00485	0.01816	-0.01345	-0.00080	-0.04765	0.01992	0.01134	-0.00748	0.00170
Apple	-0.00063	0.00012	0.00000	-0.00020	0.00003	-0.00006	0.00049	-0.00015	-0.00006	-0.00006	-0.00015
Microsoft	0.00485	0.00000	0.00186	0.00000	-0.00051	0.00001	-0.00009	-0.00053	-0.00028	-0.00031	0.00007
AT&T	0.01816	-0.00020	0.00000	0.00257	-0.00144	0.00013	-0.00430	0.00160	0.00114	-0.00052	0.00012
IBM	-0.01345	0.00003	-0.00051	-0.00144	0.00231	-0.00015	0.00060	-0.00101	-0.00015	0.00148	-0.00014
Amazon	-0.00080	-0.00006	0.00001	0.00013	-0.00015	0.00022	0.00059	-0.00047	-0.00011	0.00003	-0.00007
SP500	-0.04765	0.00049	-0.00009	-0.00430	0.00060	0.00059	0.02044	-0.00783	-0.00435	-0.00147	-0.00101
Nasdaq	0.01992	-0.00015	-0.00053	0.00160	-0.00101	-0.00047	-0.00783	0.00482	0.00160	-0.00051	0.00065
Dow	0.01134	-0.00006	-0.00028	0.00114	-0.00015	-0.00011	-0.00435	0.00160	0.00134	0.00019	0.00003
Nikkei	-0.00748	-0.00006	-0.00031	-0.00052	0.00148	0.00003	-0.00147	-0.00051	0.00019	0.00183	-0.00003
HP	0.00170	-0.00015	0.00007	0.00012	-0.00014	-0.00007	-0.00101	0.00065	0.00003	-0.00003	0.00044

Elaboración propia

VARIANZA TOTAL	VARIANZA GENERALIZADA	VARIANZA EFECTIVA
0.24772	1.62E-38	0.00037

Anexo 4

Tabla 13

Matriz de varianza y covarianza de los estimadores de Componentes Principales

	Cons	PC1	PC2
Cons	8.93136E-05	3.46605E-21	9.03636E-20
PC1	3.46605E-21	1.32409E-05	-1.26386E-20
PC2	9.03636E-20	-1.26386E-20	0.000051588

Elaboración propia

VARIANZA TOTAL	VARIANZA GENERALIZADA	VARIANZA EFECTIVA
0.000154142	6.10E-14	3.93666E-05

	1 comps	2 comps	3 comps	4 comps	5 comps	6 comps	7 comps	8 comps	9 comps	10 comps
X	67.47	84.79	91.12	94.94	97.33	98.59	99.14	99.56	99.89	100
IGBVL	53.15	71.27	74.99	90.21	90.22	90.23	92.42	94.55	94.56	94.7

Anexo 5

Tabla 14

Matriz de varianza y covarianza de los estimadores del método Ridge

Variables regresoras	Cons	Apple	Microsoft	AT&T	IBM	Amazon	SP500	Nasdaq	Dow	Nikkei	HP
Cons	0.10722	-0.00027	0.00260	0.00913	-0.00700	-0.00036	-0.02309	0.00963	0.00555	-0.00407	0.00077
Apple	-0.00027	0.00012	0.00000	-0.00016	0.00001	-0.00006	0.00039	-0.00011	-0.00004	-0.00007	-0.00015
Microsoft	0.00260	0.00000	0.00180	-0.00019	-0.00036	0.00001	0.00038	-0.00073	-0.00039	-0.00022	0.00005
AT&T	0.00913	-0.00016	-0.00019	0.00179	-0.00089	0.00016	-0.00218	0.00071	0.00064	-0.00023	0.00004
IBM	-0.00700	0.00001	-0.00036	-0.00089	0.00188	-0.00017	-0.00081	-0.00041	0.00018	0.00123	-0.00009
Amazon	-0.00036	-0.00006	0.00001	0.00016	-0.00017	0.00022	0.00047	-0.00042	-0.00008	0.00003	-0.00007
SP500	-0.02309	0.00039	0.00038	-0.00218	-0.00081	0.00047	0.01425	-0.00526	-0.00291	-0.00214	-0.00077
Nasdaq	0.00963	-0.00011	-0.00073	0.00071	-0.00041	-0.00042	-0.00526	0.00374	0.00100	-0.00021	0.00055
Dow	0.00555	-0.00004	-0.00039	0.00064	0.00018	-0.00008	-0.00291	0.00100	0.00100	0.00035	-0.00003
Nikkei	-0.00407	-0.00007	-0.00022	-0.00023	0.00123	0.00003	-0.00214	-0.00021	0.00035	0.00167	-0.00001
HP	0.00077	-0.00015	0.00005	0.00004	-0.00009	-0.00007	-0.00077	0.00055	-0.00003	-0.00001	0.00043

Elaboración propia

VARIANZA TOTAL	VARIANZA GENERALIZADA	VARIANZA EFECTIVA
0.13411	7.62E-39	0.00034

Anexo 6

Tabla 15

Matriz de varianza y covarianza de los estimadores de LASSO

Variables regresoras	Cons	Apple	Microsoft	AT&T	IBM	Amazon	SP500	Nasdaq	Dow	Nikkei	HP
Cons	0.16558	-0.00049	0.00379	0.01420	-0.01051	-0.00063	-0.03726	0.01558	0.00886	-0.00585	0.00133
Apple	-0.00049	0.00012	0.00000	-0.00018	0.00002	-0.00006	0.00046	-0.00014	-0.00006	-0.00006	-0.00015
Microsoft	0.00379	0.00000	0.00184	-0.00009	-0.00045	0.00001	0.00015	-0.00063	-0.00034	-0.00027	0.00006
AT&T	0.01420	-0.00018	-0.00009	0.00223	-0.00119	0.00014	-0.00341	0.00122	0.00093	-0.00038	0.00009
IBM	-0.01051	0.00002	-0.00045	-0.00119	0.00212	-0.00016	-0.00006	-0.00073	0.00000	0.00138	-0.00012
Amazon	-0.00063	-0.00006	0.00001	0.00014	-0.00016	0.00022	0.00055	-0.00045	-0.00010	0.00003	-0.00007
Nasdaq	0.01558	-0.00014	-0.00063	0.00122	-0.00073	-0.00045	-0.00685	0.00441	0.00137	-0.00035	0.00062
Dow	0.00886	-0.00006	-0.00034	0.00093	0.00000	-0.00010	-0.00379	0.00137	0.00120	0.00028	0.00001
Nikkei	-0.00585	-0.00006	-0.00027	-0.00038	0.00138	0.00003	-0.00184	-0.00035	0.00028	0.00177	-0.00002
HP	0.00133	-0.00015	0.00006	0.00009	-0.00012	-0.00007	-0.00093	0.00062	0.00001	-0.00002	0.00044

Elaboración propia

VARIANZA TOTAL	VARIANZA GENERALIZADA	VARIANZA EFECTIVA
0.18312	3.97E-32	0.00072

Anexo 7

Tabla 16

Comparación de varianzas estimadas por cada modelo de Regresión

Estadísticos	**TÉCNICAS DE REGRESIÓN**			
	Mínimos Cuadrados Ordinarios (MCO)	Componentes Principales (RCP)	Ridge	LASSO
Varianza total	0.24772	0.000154142	0.13411	0.18312
Varianza generalizada	1.62E-38	6.10075E-14	7.62E-39	3.97E-32
Varianza efectiva	0.00037	3.93666E-05	0.00034	0.00072

Elaboración propia

Anexo 8

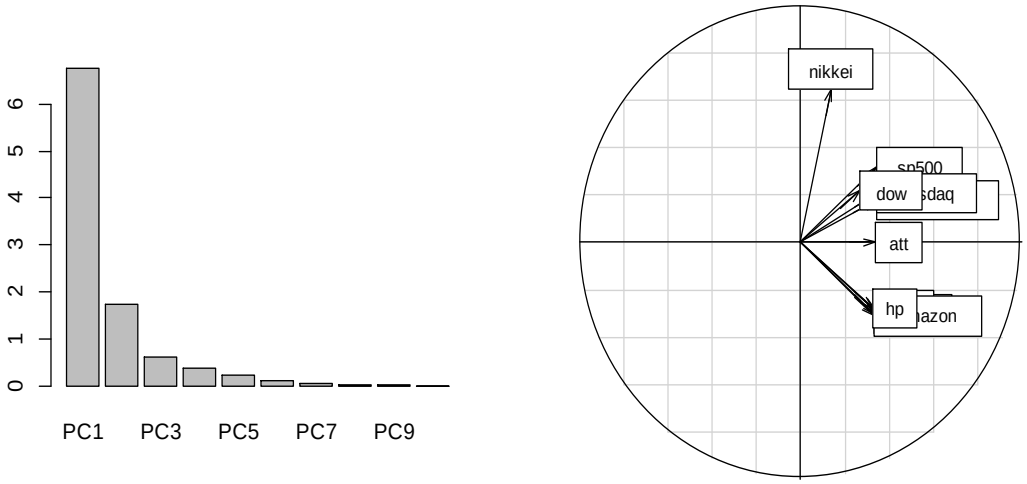

Figura 10. *Elección del número de Componentes y relación de Componentes y sus variables*

Anexo 9

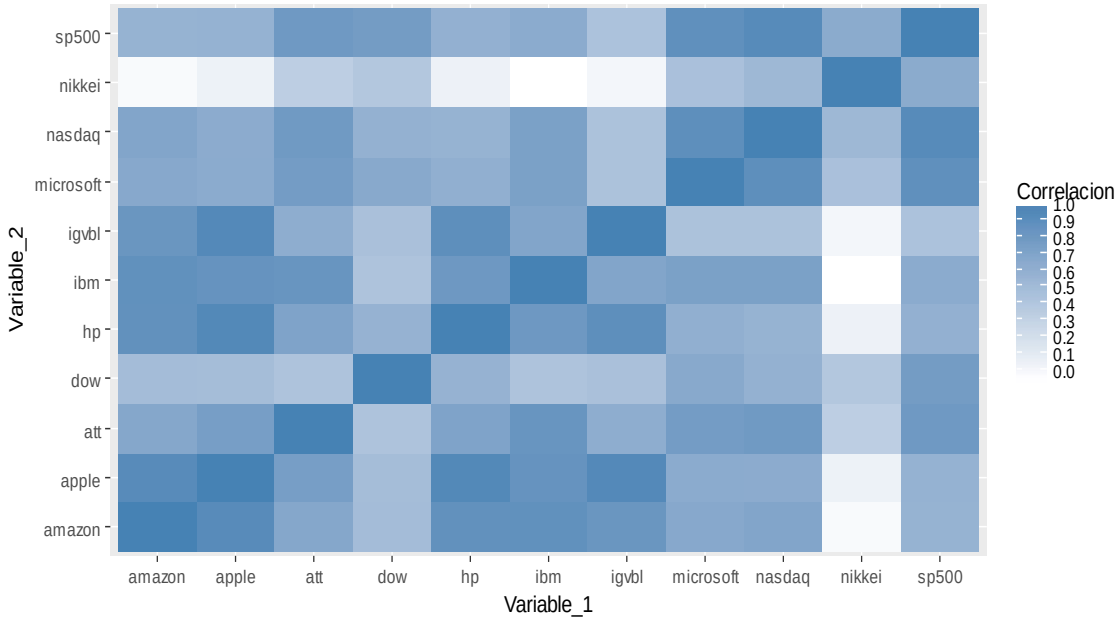

Figura 11. *Gráfico de densidad de la matriz de correlaciones de las variables predictoras*

Anexo 10

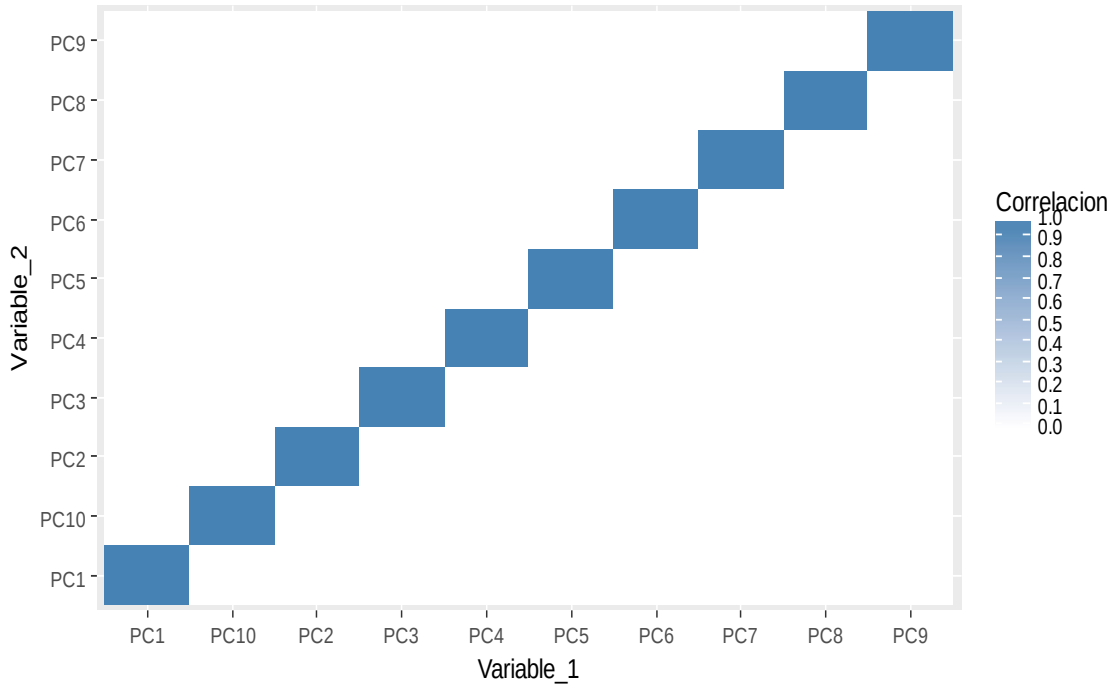

Figura 12. *Gráfico de densidad de la matriz de correlaciones de los Componentes Principales*

Anexo 11

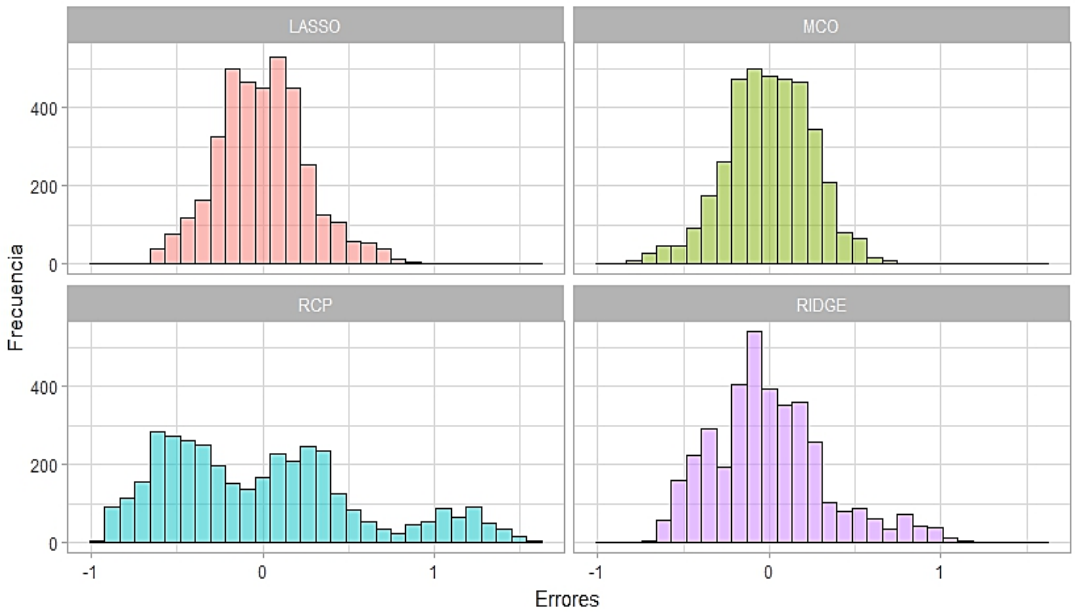

Figura 13. *Distribución de los errores para cada método utilizado*

Anexo 12

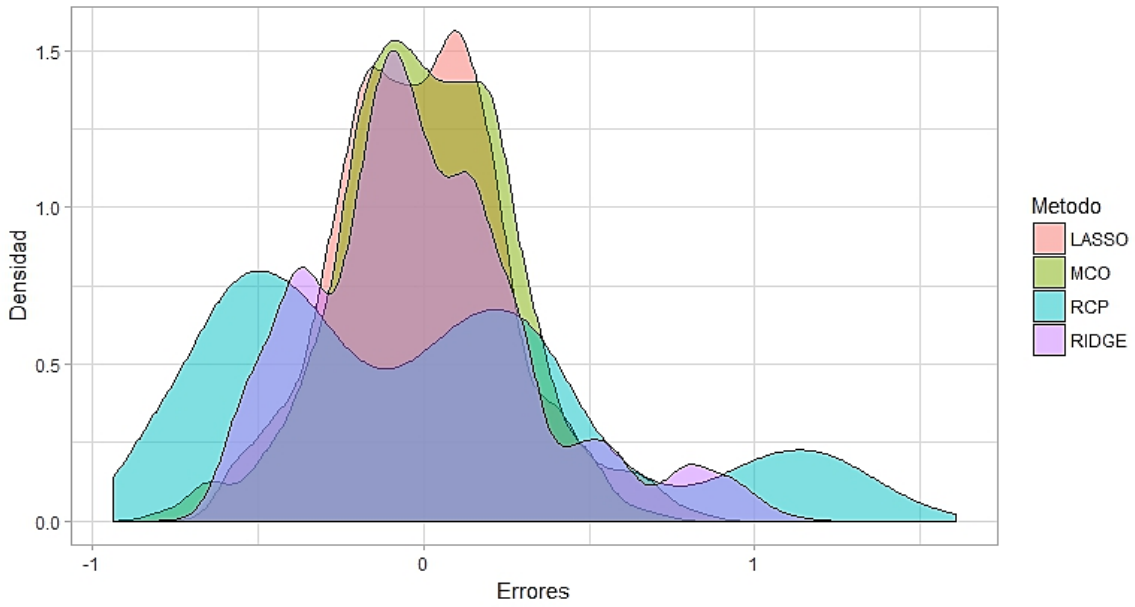

Figura 14. *Comparación de densidades para cada método de Regresión*

Anexo 13

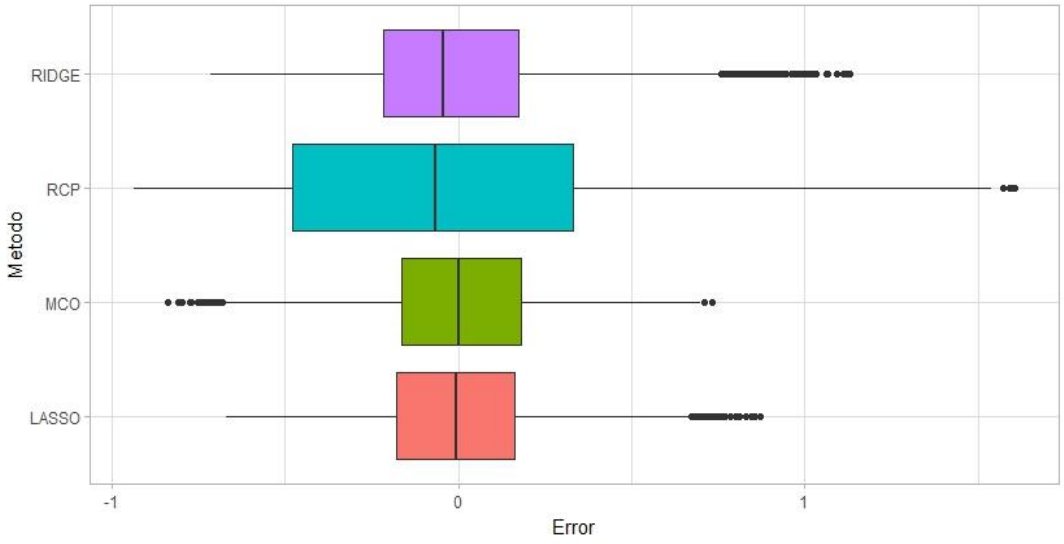

Figura 15. *Boxplot de los errores para cada modelo de Regresión*

Anexo 14

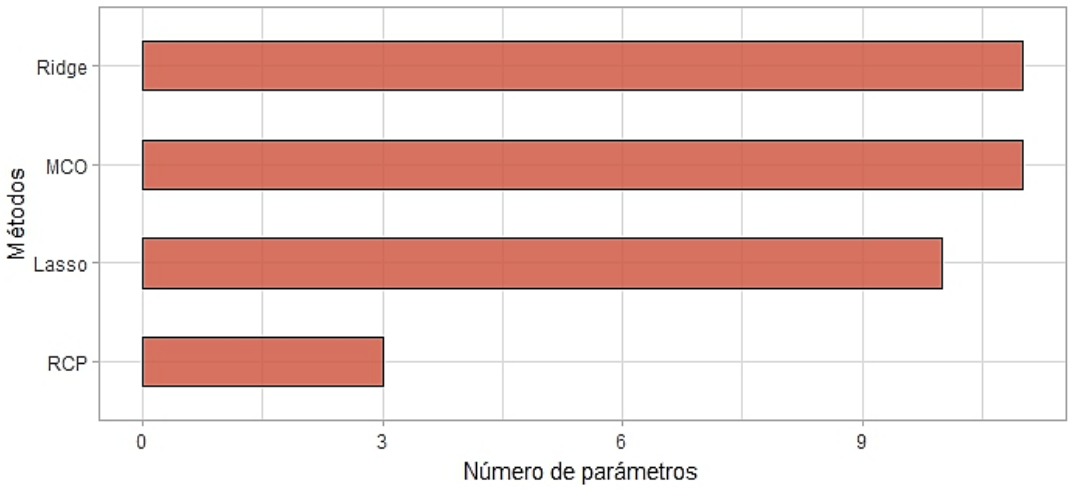

Figura 16. *Número de parámetros para modelo de Regresión*