



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

**Una prueba sobre la especificación de un modelo de
regresión**

Amanda Lolita Pineda Norman

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:

MAESTRA EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO
2013

La presente tesis titulada: **Una prueba sobre la especificación de un modelo de regresión**, realizada por la alumna: **Amanda Lolita Pineda Norman**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRA EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO

Dr. Gustavo Ramírez Valverde

ASESOR

Dr. Benito Ramírez Valverde

ASESOR

M.E. Jorge Alberto Jiménez Ruiz

Una prueba sobre la especificación de un modelo de regresión

Amanda Lolita Pineda Norman

Colegio de Postgraduados, 2013

Resumen

Un modelo de regresión lineal múltiple tiene problemas de especificación incorrecta de la forma funcional cuando no explica en forma adecuada la relación entre la variable dependiente y las explicativas observadas. Una prueba sobre la especificación del modelo es la prueba RESET; sin embargo, la prueba ha demostrado tener baja potencia, quedando además poco clara la forma de definir las variables que serán incluidas en la regresión auxiliar que utiliza, presentando por construcción problemas de multicolinealidad. En el presente trabajo se propone una prueba de especificación basada en Multiplicadores de Lagrange y la regresión Lasso Adaptativo [Zou \(2006\)](#) a la cual le llamamos ML-ALasso. Al igual que la prueba RESET, la prueba ML-ALasso se basa en una regresión ampliada del modelo original. Además se propone un método para muestras pequeñas que utiliza la prueba de permutaciones al cual llamamos ML-ALasso-p. Mediante un estudio de simulación se analiza el comportamiento de las pruebas y se comparan con dos pruebas alternativas para verificar una correcta especificación del modelo. Se concluye que el método ML-ALasso tiene un buen desempeño para muestras grandes y el método ML-ALasso-p funciona bien para muestras pequeñas.

Palabras clave: RESET, Lasso Adaptativo, Multiplicador de Lagrange, Prueba de Permutaciones, Error de especificación, Forma funcional.

A test on the specification of a regression model

Amanda Lolita Pineda Norman

Colegio de Postgraduados, 2013

Abstract

A multiple linear regression model has problems of incorrect specification of the functional form when not adequately explain the relationship between the dependent variable and the explanatory observed. A test of the model specification is the RESET test, but the test has been shown to have low power while leaving unclear how to define the variables to be included in the auxiliary regression using, disclosing construction multicollinearity problems. In this paper a specification test based on Lagrange Multipliers and Adaptive Lasso [Zou \(2006\)](#) regression which we call ML-ALasso is proposed. Like the RESET test, the ML-ALasso test is based on an extended regression of the original model. In addition a method for small samples using the permutation test which we call ML-ALasso-p is proposed. By a simulation study test behavior is analyzed and compared with two alternative tests to verify proper specification of the model. We conclude that the ML-ALasso method has a good performance for large samples and ML-ALasso-p method works well for small samples.

Key words: RESET test, Adaptive Lasso, Lagrange Multiplier, Permutation Test, Misspecification, Functional form.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado durante la realización de mis estudios de postgrado.

Al Colegio de Postgraduados, por haberme brindado la oportunidad de seguir mi formación académica en sus aulas.

A los integrantes de mi Consejo Particular:

Muy especialmente al Dr. Gustavo Ramírez Valverde mi mayor agradecimiento por su interminable paciencia, sus enseñanzas y consejos para la realización de esta tesis.

Al Dr. Benito Ramírez Valverde, por su apoyo y colaboración en el presente trabajo.

Al M.E. Jorge Alberto Jiménez Ruiz, por sus acertadas sugerencias y ayuda incondicional.

A mis profesores, compañeros de clases y todos aquellos que de alguna u otra manera fueron copartícipes de esta tarea, a todos gracias.

DEDICATORIA

A mi querido y adorado hijo Mateo Jiménez Pineda

A mi amado esposo Jorge Alberto Jiménez Ruiz

A mis padres: Rosa Isabel Norman Quintero y Vicente Pineda Guzmán
(finado)

A mi papá: Andrés Rafael López García

A mis hermanos: Alfonso, Isabel, Vicente Alcino, Marta Esther y Andrés

A mi querida abuelita María Quintero por todas sus bendiciones

A mis tres preciosas sobrinas: Yihan, Jade y Deni

Índice

1. Introducción	1
2. Objetivos	3
3. Marco teórico	4
3.1. Modelo de Regresión Lineal	4
3.1.1. Supuestos del Modelo de Regresión Lineal	5
3.2. Pruebas de Especificación	7
3.2.1. Prueba RESET	8
3.2.2. Multiplicador de Lagrange (ML)	12
3.2.3. Examen de residuos	13
3.2.4. Estadístico d de Durbin-Watson	14
3.3. Regresión Lasso	16
3.3.1. Definición de regresión Lasso	16
3.3.2. Definición de regresión Lasso Adaptativo	17
3.4. Criterios para selección de modelos	18
3.4.1. Criterio R_p^2	19
3.4.2. Criterio F_p	20

3.4.3. Criterio $CME(p)$	21
3.4.4. C_p de Mallow	21
3.5. Prueba de permutaciones	22
3.5.1. Prueba de Permutación para un Modelo de Regresión Lineal Múltiple	22
4. Prueba $ML - ALasso$	25
4.1. Método $ML - ALasso - p$ para muestras pequeñas	28
5. Ejemplo de aplicación	30
6. Estudio de simulación	32
6.1. Factores evaluados en el estudio de simulación	32
6.1.1. Número de variables en el modelo bajo la hipótesis nula	33
6.1.2. Tamaño de muestra	34
6.1.3. Correlación entre las variables independientes	34
6.2. Propiedades evaluadas	34
6.2.1. Modelos generadores de los datos bajo la hipótesis nula	35
6.2.2. Potencia de la prueba	36
6.3. Prueba RESET	38
7. Resultados y Discusión	39
7.1. Tamaño de Prueba	39
7.1.1. Prueba $ML - ALasso$	39
7.1.2. Prueba $ML - ALasso - p$	41
7.2. Potencia de la prueba	44

Índice

7.2.1. Prueba $ML - ALasso$	44
7.2.2. Prueba $ML - ALasso - p$	48
8. Conclusiones y Recomendaciones	53
8.1. Conclusiones	53
8.1.1. Tamaño de prueba	53
8.1.2. Potencia de prueba	53
8.2. Recomendaciones	54
Referencias	54
Anexos	57
A. Código de la función Lasso Adaptativo.	58
B. Código de la prueba de permutaciones.	60
C. Código de la prueba $ML - ALasso$.	61
D. Código de la prueba $ML - ALasso - p$.	63
E. Código de la prueba RESET.	65

Índice de cuadros

6.1. Modelos generadores de los datos para las tres situaciones contempladas con el modelo 1 que tiene una sola variable explicativa	36
6.2. Modelos generadores de los datos para cuatro situaciones contempladas con el modelo 2 que tiene dos variables explicativas	37
7.1. Tamaño de la Prueba del Modelo 1: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i>	40
7.2. Tamaño de la Prueba del Modelo 2: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + 3\mathbf{x}_{2i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i>	41
7.3. Tamaño de la Prueba del Modelo 1: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i> – <i>p</i>	42
7.4. Tamaño de la Prueba del Modelo 2: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + 3\mathbf{x}_{2i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i> – <i>p</i>	43
7.5. Potencia de la Prueba del Modelo 1: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i>	45
7.6. Potencia de la Prueba del Modelo 2: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + 3\mathbf{x}_{2i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i>	47
7.7. Potencia de la Prueba del Modelo 1: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i> – <i>p</i>	49
7.8. Potencia de la Prueba del Modelo 2: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + 3\mathbf{x}_{2i} + \epsilon_i$ para <i>ML</i> – <i>ALasso</i> – <i>p</i>	51

Índice de figuras

3.1. Residuos \hat{u}_i obtenidos de las funciones de costo total de tipo <i>a</i>)lineal (3.15), <i>b</i>)cuadrática (3.14) y <i>c</i>) cúbica (3.13). Gujarati (2004).	15
---	----

Capítulo 1

Introducción

El modelo de regresión lineal es ampliamente usado en diversos campos y sirve para estudiar la relación que existe entre una variable dependiente y una o más variables independientes. Para estimar los parámetros de este modelo, se utiliza el método de mínimos cuadrados ordinarios (MCO), el cual busca minimizar la suma de cuadrados del error. Se requiere realizar inferencia de algunos supuestos al modelo. El modelo clásico de regresión lineal de acuerdo a [Gujarati \(2004\)](#) requiere de 10 supuestos. Uno de estos supuestos es que el modelo de regresión está correctamente especificado. Esto es, que el modelo incluye las variables correctas, sin omitir ninguna y sin incluir variable que no sean importantes. Cuando este supuesto, referente a forma funcional, no se cumple entonces nos enfrentamos con el problema de una mala especificación del modelo.

Han sido propuestas muchas pruebas para probar la especificación de un modelo de regresión. [Ramsey \(1969\)](#), [Thursby y Schmidt \(Sep., 1977\)](#), [Thursby \(1979\)](#) y [Engle \(1979\)](#) [Bera y Jarque \(1982\)](#) sugieren pruebas de especificación para la forma funcional, solo por mencionar algunas. La Prueba del Error de Especificación en Regresión (RESET, por sus siglas en inglés) es una prueba para conocer si se han cometido errores de especificación de una forma funcional, esta prueba de especificación es probablemente la prueba más aceptada. Sin embargo, RESET ha mostrado tener baja potencia, quedando además poco clara la forma de definir las variables que serán incluidas en la regresión auxiliar que utiliza, presentando por construcción problemas de multicolinealidad.

En este documento, el objetivo general fue proponer una prueba sobre la especificación

1. Introducción

de un modelo de regresión lineal y evaluar su desempeño en cuanto a la potencia y al tamaño de la prueba.

La metodología propone una prueba de especificación de un modelo de regresión lineal basada en los Multiplicadores de Lagrange y la regresión Lasso Adaptativo a la cual le llamamos *ML – ALasso*. Al igual que la prueba RESET, la prueba *ML – ALasso* se basa en una regresión ampliada del modelo original al cual se le prueba el siguiente juego de hipótesis, H_0 : El modelo está bien especificado vs H_a : El modelo no está bien especificado. Debido a que la prueba de Multiplicadores de Lagrange es asintótica, para muestras pequeñas se propone una prueba a la cual le llamamos *ML – ALasso – p* utilizando una prueba exacta en la regresión auxiliar.

En el capítulo 2 se mencionan los objetivos general y particulares del presente trabajo. En el capítulo 3 se definen los conceptos importantes relacionados a este trabajo. En el capítulo 4 se describen las pruebas propuestas *ML – ALasso* y *ML – ALasso – p*. En el capítulo 5 se muestra un ejemplo de aplicación referente al efecto del precio del tabaco sobre su consumo [Olivera~Chávez *et al.* (2010)]. En el capítulo 6 se describe el estudio de simulación evaluando el comportamiento de las pruebas propuestas. En el capítulo 7 se muestran los resultados y la discusión. Y finalmente en el capítulo 8 se enumeran las conclusiones y recomendaciones donde se muestra que la prueba *ML-ALasso* funciona bien para muestras grandes y *ML-ALasso-p* funciona bien para muestras pequeñas.

Capítulo 2

Objetivos

General

- Proponer una prueba sobre la especificación de un modelo de regresión lineal y evaluar su desempeño en cuanto al tamaño y potencia de la prueba.

Particulares

- Proponer una prueba de hipótesis sobre una correcta relación funcional basada en la Prueba del Error de Especificación en Regresión (RESET, por sus siglas en inglés) [[Ramsey \(1969\)](#)], Multiplicadores de Lagrange (ML) y la regresión Lasso adaptativo o ALasso [[Zou \(2006\)](#)].
- Evaluar el comportamiento de la prueba propuesta en cuanto a la potencia y al tamaño de prueba; y compararla con la prueba RESET, que es la prueba más aceptada.

Capítulo 3

Marco teórico

3.1. Modelo de Regresión Lineal

El modelo de regresión lineal es ampliamente usado en diversos campos y sirve para estudiar la relación que existe entre una variable dependiente y una o más variables independientes; de forma general se expresa de la siguiente forma:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u} \quad (3.1)$$

Donde: \mathbf{Y} es un vector de $\mathbf{nx1}$ con los valores de la variable dependiente, \mathbf{X} es una matriz $\mathbf{nx(k + 1)}$, donde la primer columna está asociada con el intercepto y está formada solamente por unos, \mathbf{k} es el número de variables independientes también llamadas explicativas o regresoras, después del intercepto, cada columna de las restantes contiene los valores de cada una de las variables independientes, β es un vector $(\mathbf{k + 1})\mathbf{x1}$ con los parámetros del modelo y \mathbf{u} es el término de error y es una variable aleatoria (estocástica).

3.1. Modelo de Regresión Lineal

Para estimar los parámetros de este modelo, un método es el de mínimos cuadrados ordinarios (MCO), el cual busca minimizar la suma de cuadrados del error.

El modelo requiere realizar inferencia de algunos supuestos, en la siguiente sección, se presentan en forma breve estos supuestos.

3.1.1. Supuestos del Modelo de Regresión Lineal

El modelo clásico de regresión lineal de acuerdo a [Gujarati \(2004\)](#) requiere de 10 supuestos, sin embargo, [Greene \(1999\)](#) los compacta en solo 6 supuestos.

A continuación se describen los 10 supuestos del modelo clásico de regresión lineal según [Gujarati \(2004\)](#):

Supuesto 1: El modelo de regresión es lineal en sus parámetros.

$$\mathbf{E}[\mathbf{Y}_i] = \beta_1 + \mathbf{X}_i\beta_2 \quad (3.2)$$

Supuesto 2: Señala que los valores de \mathbf{X} son fijos, es decir, \mathbf{X} es no estocástica.

Supuesto 3: El valor esperado del error aleatorio \mathbf{u}_i dado un valor de \mathbf{X} es cero.

$$E(u_i | X_i) = 0 \quad (3.3)$$

Lo anterior, también implica que $E(Y_i | X_i) = \beta_1 + \beta_2 X_i$.

Supuesto 4: Homocedasticidad o igual varianza de \mathbf{u}_i , esto quiere decir

3.1. Modelo de Regresión Lineal

que dado el valor de \mathbf{X} , la varianza de \mathbf{u}_i es la misma para todas las observaciones.

$$\begin{aligned} Var(u_i | X_i) &= E[u_i - E(u_i) | X_i]^2, \\ &= E(u_i^2 | X_i), \\ &= \sigma^2 \end{aligned} \tag{3.4}$$

Supuesto 5: No existe autocorrelación entre los errores o residuos. Es decir, dadas dos observaciones de \mathbf{X} , sean \mathbf{X}_i y \mathbf{X}_j , con $i \neq j$, entonces se debe cumplir que

$$\begin{aligned} Cov(u_i, u_j | X_i, X_j) &= E\{[u_i - E(u_i)] | X_i\} \{[u_j - E(u_j)] | X_j\}, \\ &= E(u_i | X_i)(u_j | X_j), \\ &= 0 \end{aligned} \tag{3.5}$$

Supuesto 6: La covarianza entre el error \mathbf{u}_i y la variable independiente \mathbf{X}_i es cero.

Supuesto 7: El número de observaciones \mathbf{n} debe ser mayor que el número de parámetros por estimar.

Supuesto 8: La variabilidad de los valores de \mathbf{X} que debe ser un número positivo finito.

Supuesto 9: El modelo de regresión está correctamente especificado. Esto es, que el modelo incluye las variables correctas, sin omitir ninguna y sin incluir variable que no sean importantes.

Supuesto 10: No hay multicolinealidad perfecta, esto es, ninguna variable independiente puede ser expresada como una combinación lineal de algún subconjunto del resto de las variables independientes.

3.2. Pruebas de Especificación

Cuando el supuesto 9, referente a forma funcional, no se cumple entonces nos enfrentamos con el problema de mala especificación del modelo. Por ejemplo, supóngase que el modelo verdadero es:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i$$

Pero por alguna razón el investigador postula el siguiente modelo:

$$Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 X_i^2 + u_i$$

Se presenta un error en la especificación ya que el modelo supuesto no incluye la variable X_i^3 .

3.2. Pruebas de Especificación

Han sido propuestas muchas pruebas para probar la especificación de un modelo de regresión. [Ramsey \(1969\)](#), [Thursby y Schmidt \(Sep., 1977\)](#), [Thursby \(1979\)](#) y [Engle \(1979\)](#) [Bera y Jarque \(1982\)](#) sugieren pruebas de especificación para forma funcional. [Russell y MacKinnon \(1999\)](#) mencionan una gran cantidad de pruebas: las pruebas t y F para variables omitidas y para parámetros de constancia, pruebas de heteroscedasticidad, pruebas de correlación serial, pruebas de las restricciones de los factores comunes usadas en modelos que sufren de mala especificación dinámica, prueba DWH, pruebas de restricciones de sobreidentificación, y las tres pruebas clásicas de los modelos estimados por máxima verosimilitud; además de la prueba RESET, prueba de momento condicional, pruebas de asimetría y curtosis, prueba de matriz de información, entre otros. [Maddala \(2001\)](#) menciona la prueba de error en especificación de Hausman y la prueba de diferenciación

3.2. Pruebas de Especificación

Plosser-Schwet-White (PSW) aplicada a series de tiempo. [Kleiber y Zeileis \(2008\)](#) mencionan algunas funciones para probar la forma funcional implementadas en R, como son, la prueba RESET (Ramsey 1969), la prueba arco iris (Utts 1982) y la prueba Harvey-Collier (Harvey y Collier 1977).

A continuación se describen las pruebas más importantes usadas para la especificación de un modelo de regresión.

3.2.1. Prueba RESET

La Prueba del Error de Especificación en Regresión es una prueba para conocer si se han cometido errores de especificación de una forma funcional, esta prueba es probablemente la más aceptada. La forma práctica de realizar ésta prueba es la siguiente [[Gujarati \(2004\)](#)]:

Suponga que se tiene el modelo siguiente:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u} \quad (3.6)$$

Donde: \mathbf{Y} es un vector de $\mathbf{nx1}$ con los valores de la variable dependiente, \mathbf{X} es una matriz de rango \mathbf{k} de $\mathbf{nx(k + 1)}$, donde la primer columna está asociada con el intercepto y está formada solamente por unos, \mathbf{k} es el número de variables independientes o regresoras, después del intercepto, cada columna de las restantes contiene los valores de cada una de las variables independientes, β es un vector de $(\mathbf{k + 1})\mathbf{x1}$ con los parámetros del modelo y \mathbf{u} es el vector de $\mathbf{nx1}$ y es el término de error que se distribuye $N(0, \sigma^2 I)$.

Se desea probar si este modelo está bien especificado, entonces las

3.2. Pruebas de Especificación

hipótesis a probar son:

H_0 : El modelo está bien especificado

vs

H_a : El modelo no está bien especificado

Primeramente se obtiene $\hat{\mathbf{Y}}$ que es el valor de \mathbf{Y} estimado utilizando los estimadores de mínimos cuadrados ordinarios en la ecuación original, es decir:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} \quad (3.7)$$

Donde: $\hat{\beta} = (X'X)^{-1}X'Y$.

En una segunda etapa se efectúa una regresión auxiliar, en donde el modelo ha sido aumentado con las potencias de los valores predichos en la etapa uno. El modelo de la regresión ampliada es:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\theta + \mathbf{u} \quad (3.8)$$

Donde: $\mathbf{Z} = (\hat{\mathbf{Y}}^2, \hat{\mathbf{Y}}^3, \dots, \hat{\mathbf{Y}}^k, \dots)$ con $k = 3$ o 4 . Para probar este modelo la hipótesis es $H_0 : \theta = 0$ vs $H_a : \theta \neq 0$, es equivalente a probar H_0 : El modelo está bien especificado vs H_a : El modelo no está bien especificado.

El estadístico de prueba es la prueba F reformulada sólo en términos de los valores R^2 , a saber:

$$F = \frac{(R_{nueva}^2 - R_{vieja}^2)/N}{(1 - R_{nueva}^2)/(n - P)} \quad (3.9)$$

Donde:

R_{nueva}^2 = coeficiente de determinación obtenido de (3.8)

3.2. Pruebas de Especificación

R_{vieja}^2 = coeficiente de determinación obtenido de (3.6)

N = número de regresoras agregadas para obtener el modelo extendido

P = número de parámetros en el modelo nuevo

El valor de F calculado se contrasta con F de tablas, a un nivel de significancia, para decidir si se acepta o rechaza la hipótesis nula.

Variantes de prueba RESET

En la literatura se encuentran reportadas algunas variantes de la prueba RESET, que se obtienen al modificar las variables contenidas en la regresión ampliada, como ejemplo tenemos los siguientes:

- Ramsey y Gilbert (1972), donde:

$$\mathbf{Z} = [\hat{\mathbf{Y}}^2, \hat{\mathbf{Y}}^3, \hat{\mathbf{Y}}^4]$$

- Thursby y Schmidt (1977), donde:

$$\mathbf{Z} = [\mathbf{X}^2, \mathbf{X}^3, \mathbf{X}^4]$$

- Más recientemente, Godfrey, McAleer, y McKenzie (1988) y Godfrey y Orme (1994), donde $\mathbf{Z} = \hat{\mathbf{Y}}^2$

Esto es, Ramsey y Gilbert (1972) pusieron como variables a probar las potencias segunda, tercera y cuarta de $\hat{\mathbf{Y}}$. En contraste, Thursby y Schmidt (1977) probaron el cuadrado, cubo y cuarta potencia de los elementos de \mathbf{X} como las variables de prueba y Godfrey, McAleer, y McKenzie (1988) y Godfrey y Orme (1994) prueban solo usando $\hat{\mathbf{Y}}^2$.

3.2. Pruebas de Especificación

Ambos, Thursby y Schmidt (1977) y Godfrey y Orme (1994) afirman que su propia versión de RESET es mejor que la prueba RESET de Ramsey y Gilbert (1972) detectando variables omitidas.

Respecto al resultado que obtuvieron [Thursby y Schmidt \(Sep., 1977\)](#) se puede mencionar lo siguiente. Realizaron la prueba RESET con dos modelos referentes a forma funcional incorrecta (uno con efecto aditivo y otro con efecto multiplicativo), con tamaños de muestra de 20, 30 y 50 y con niveles de significancia de 5 y 10 por ciento y usan las siguientes potencias de \hat{Y} :

1. \hat{Y}^2
2. \hat{Y}^2, \hat{Y}^3
3. $\hat{Y}^2, \hat{Y}^3, \hat{Y}^4$
4. $\hat{Y}^2, \hat{Y}^3, \hat{Y}^4, \hat{Y}^5$

Además usan potencias de X :

1. X^2
2. X^2, X^3
3. X^2, X^3, X^4

Los resultados que obtuvieron son que para el modelo con efecto aditivo los resultados fueron excelentes (cerca de 100 por ciento de rechazos) y para el modelo con efecto multiplicativo se obtuvo el 100 por ciento de rechazos para todas las pruebas.

A continuación se mencionan otras pruebas para probar especificación de un modelo, como son: Multiplicador de Lagrange (ML), Examen de Residuos y Estadístico d de Durbin-Watson.

3.2. Pruebas de Especificación

3.2.2. Multiplicador de Lagrange (ML)

Para ilustrar esta prueba Gujarati (2004) presenta el siguiente ejemplo:

Si se compara la función lineal de \mathbf{X} :

$$\mathbf{Y} = \beta_1 + \beta_2\mathbf{X} + \mathbf{u} \quad (3.10)$$

con la función cúbica de la misma variable:

$$\mathbf{Y} = \beta_1 + \beta_2\mathbf{X} + \beta_3\mathbf{X}^2 + \beta_4\mathbf{X}^3 + \mathbf{v} \quad (3.11)$$

Donde: \mathbf{Y} es un vector $n \times 1$ con los valores de la variable dependiente, \mathbf{X} es una matriz de rango \mathbf{k} de $n \times (\mathbf{k} + 1)$, donde la primer columna está asociada con el intercepto y está formada de unos, \mathbf{k} es el número de variables independientes, después del intercepto, cada columna de las restantes contiene los valores de cada una de las variables independientes o regresoras, β son los vectores $(\mathbf{k} + 1) \times 1$ con los coeficientes de las ecuaciones (3.10) y (3.11), y \mathbf{u} y \mathbf{v} son los vectores $n \times 1$ del término de error de las ecuaciones (3.10) y (3.11) respectivamente y se distribuyen $N(0, \sigma^2 I)$.

La primera es una versión restringida de la última.

Esto es, la regresión restringida supone que $\beta_3 = \beta_4 = 0$. Por lo que las hipótesis a contrastar son:

H_0 : El modelo está bien especificado

vs

H_a : El modelo no está bien especificado

Para probar esto, la prueba ML se realiza de la siguiente manera:

3.2. Pruebas de Especificación

1. Estimar la regresión restringida (3.10) mediante MCO obteniendo los residuos $\hat{\mathbf{u}}$.
2. Si la regresión no restringida (3.11) resulta ser la verdadera regresión, los residuos obtenidos en (3.10) deben estar relacionados con \mathbf{X}^2 y \mathbf{X}^3 .
3. Esto sugiere que se debe efectuar la regresión de los $\hat{\mathbf{u}}$ obtenidos en el paso 1 sobre todas las regresoras:

$$\hat{\mathbf{u}} = \alpha_1 + \alpha_2\mathbf{X} + \alpha_3\mathbf{X}^2 + \alpha_4\mathbf{X}^3 + \mathbf{w} \quad (3.12)$$

Donde: w es un término de error con las propiedades usuales.

4. [Engle \(1982\)](#) demostró que para un tamaño de muestra (n) grande multiplicado por R^2 estimado en (3.12) sigue una distribución ji cuadrada (χ^2) con grados de libertad iguales al número de restricciones impuestas por la regresión restringida, es decir:

$$nR^2 \sim \chi^2_{(restricciones)} \quad (3.13)$$

5. Si el valor de ji cuadrada en (3.13) excede el valor ji cuadrada crítico obtenido en tablas, al nivel de significancia seleccionado, se rechaza que el modelo está bien especificado.

3.2.3. Examen de residuos

Los residuos pueden examinarse, especialmente en información de corte transversal, para la detección de errores de especificación en los modelos, tales como la omisión de una variable importante o la definición de una forma funcional incorrecta. Si estos errores existen, una gráfica de

3.2. Pruebas de Especificación

los residuos permitirá apreciarlos.

Para ilustrar lo anterior [Gujarati \(2004\)](#) presenta el siguiente ejemplo.

Suponiendo que el verdadero modelo es:

$$\mathbf{Y} = \beta_1 + \beta_2\mathbf{X} + \beta_3\mathbf{X}^2 + \beta_4\mathbf{X}^3 + \mathbf{u} \quad (3.14)$$

Pero un investigador ajusta la siguiente función cuadrática:

$$\mathbf{Y} = \beta_1 + \beta_2\mathbf{X} + \beta_3\mathbf{X}^2 + \mathbf{u}_2 \quad (3.15)$$

Y otro investigador ajusta la siguiente función lineal:

$$\mathbf{Y} = \beta_1 + \beta_2\mathbf{X} + \mathbf{u}_3 \quad (3.16)$$

A manera de ejemplo, la Figura 3.1 muestra que a medida que uno se mueve de izquierda a derecha, es decir, a medida que uno se acerca a la especificación correcta, no solamente los residuos son más pequeños (en valor absoluto) si no que también éstos no presentan los comportamientos cíclicos pronunciados asociados con modelos mal especificados.

La utilidad de examinar la gráfica de residuos es entonces clara: si hay errores de especificación, los residuos presentarán patrones distinguibles.

3.2.4. Estadístico d de Durbin-Watson

A continuación se explican los pasos para llevar a cabo la prueba del Estadístico d de Durbin-Watson. [Gujarati \(2004\)](#).

1. Se obtienen los residuos de MCO.

3.2. Pruebas de Especificación

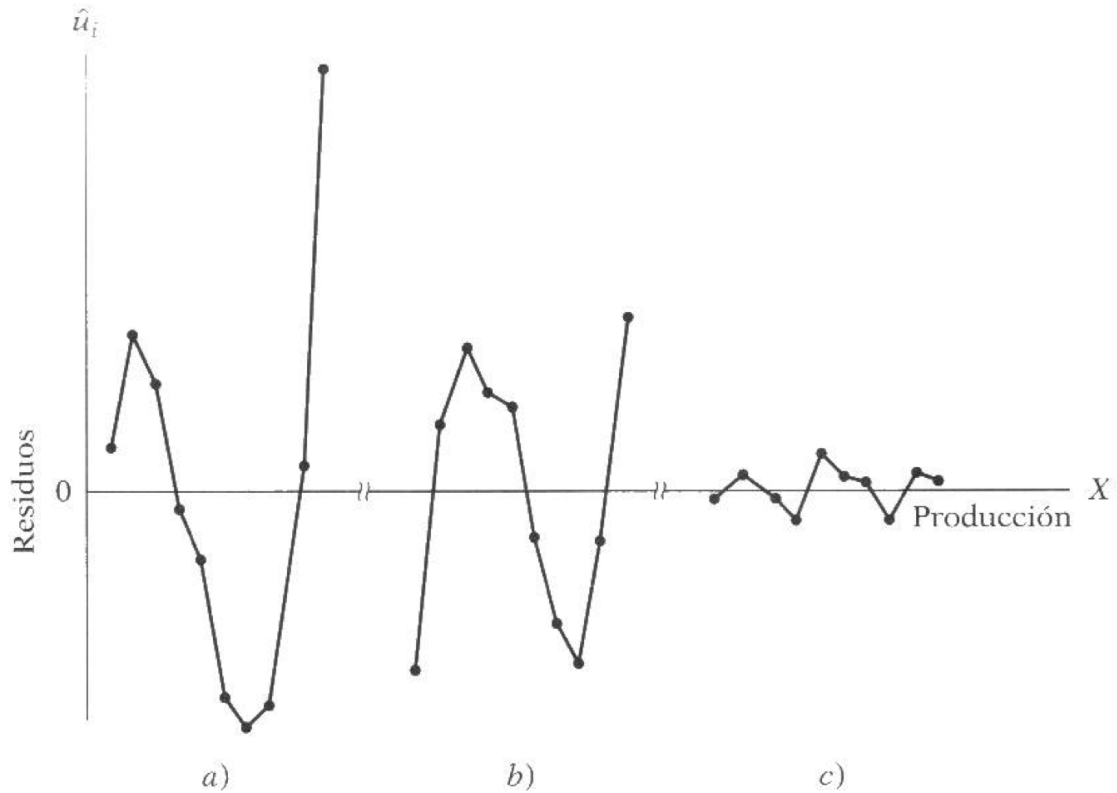


Figura 3.1: Residuos \hat{u}_i obtenidos de las funciones de costo total de tipo a) lineal (3.15), b) cuadrática (3.14) y c) cúbica (3.13). [Gujarati \(2004\)](#).

2. Si se cuenta con el supuesto de que el modelo está mal especificado porque excluye alguna variable explicativa relevante, por ejemplo Z , se ordenan los residuales obtenidos en el paso 1 de acuerdo con los valores crecientes de Z , la variable Z puede ser alguna regresora al cuadrado o al cubo.

3. A continuación se calcula el estadístico d con la fórmula siguiente:

$$d = \frac{\sum_{i=2}^n (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^n \hat{u}_i^2} \quad (3.17)$$

Donde: i es el índice de la observación.

4. De acuerdo a las tablas de Durbin-Watson, si el valor d estimado en

3.3. Regresión Lasso

3.17 es significativo, entonces podemos decir que rechazamos la hipótesis nula de que el modelo está bien especificado.

3.3. Regresión Lasso

3.3.1. Definición de regresión Lasso

Tibshirani (1996) propone un método de estimación en regresión al cual denominó "Least absolute shrinkage and selection operator" (Lasso, por sus siglas en inglés) y es una versión restringida de la regresión de MCO.

Los estimadores en este método, se obtienen minimizando la suma de cuadrados del error sujeta a que la suma del valor absoluto de los coeficientes sea menor a una constante. Debido a esta restricción se tiende a obtener algunos coeficientes exactamente iguales a cero.

Sea $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$, el estimador lasso $(\hat{\alpha}, \hat{\beta})$ definidos por

$$(\hat{\alpha}, \hat{\beta}) = \arg \min \left\{ \sum_{i=1}^N (y_i - \alpha - \sum_j \beta_j x_{ij})^2 \right\} \quad (3.18)$$

Sujeto a $\sum_j |\beta_j| \leq t$ Aquí $t \geq 0$ es un parámetro de ajuste. Ahora, para todo t , la solución para α es $\hat{\alpha} = \bar{y}$. Se puede asumir sin pérdida de generalidad que $\bar{y} = 0$ y por lo tanto omitir α .

La regresión Lasso logra además de estimar el modelo, seleccionar las variables importantes del mismo, a la vez que atenúa los efectos de la colinealidad.

Sin embargo, la regresión Lasso puede ser inconsistente debido a que

3.3. Regresión Lasso

no cumple las propiedades Oracle [Zou \(2006\)](#). La propiedad Oracle consiste en que el método identifica correctamente los componentes de β que son cero con probabilidad que tiende a 1 cuando el tamaño de muestra tiende a infinito y tiene la misma distribución límite del estimador de MCO de los componentes de β diferentes a cero.

3.3.2. Definición de regresión Lasso Adaptativo

El método Lasso Adaptativo fue introducido por [Zou \(2006\)](#) y es una variante de la regresión Lasso.

Lasso Adaptativo asigna diferentes pesos a los diferentes coeficientes y muestra que si los pesos son datos dependientes y son escogidos adecuadamente entonces se cumple con la propiedad Oracle, la cual se mencionó en la subsección 3.3.1 y es definida más adelante.

[Chatterjee y Lahiri \(2011\)](#) definen Lasso Adaptativo de la siguiente manera. Sea

$$y_i = \mathbf{x}_i' \beta + \epsilon_i, \quad i = 1, \dots, n \quad (3.19)$$

Donde y_i es la respuesta, $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})'$ es un vector covariante de dimensión p , $\beta = (\beta_1, \dots, \beta_p)'$ es el vector de parámetros, y $\{\epsilon_i : i = 1, \dots, n\}$ son los errores independientes e idénticamente distribuidos (i.i.d.). Asumimos que p está fijo.

Sea $\bar{\beta}_n$ el estimador preliminar genérico de β , como el estimador β de MCO, dado por

$$\bar{\beta}_n = \left[\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \mathbf{x}_i y_i \quad (3.20)$$

3.4. Criterios para selección de modelos

Entonces el estimador Lasso adaptativo de β se define como

$$\check{\beta}_n = \arg \min_{n \in \mathbb{R}^p} \sum_{i=1}^n (y_i - \mathbf{x}'_i \mathbf{u})^2 + \lambda_n \sum_{j=1}^p \frac{|u_j|}{|\bar{\beta}_{j,n}|^\gamma}$$

Donde $\bar{\beta}_n = (\bar{\beta}_{1,n}, \dots, \bar{\beta}_{p,n})$, $\lambda_n \geq 0$ es la penalización y $\gamma > 0$. Sea $\check{\beta}_{j,n}$ que denota la j th componente de $\check{\beta}_n$.

La propiedad Oracle, consta de dos puntos, [Zou \(2006\)](#) y [Chatterjee y Lahiri \(2011\)](#) los mencionan, y son:

1. $\mathbf{P}(B_n = \{p_0 + 1, \dots, p\}) \rightarrow 1, \quad n \rightarrow \infty,$
2. $\sqrt{n}(\check{\beta}_n^{(1)} - \beta^{(1)}) \xrightarrow{d} N(\mathbf{0}, \sigma^2 C_{11})$

Donde: C_{11} es una submatriz de C de tamaño $p_0 \times p_0$ que corresponde a la primer fila p_0 y columna p_0 y donde $\check{\beta}_n^{(1)}$ y $\beta^{(1)}$ es el vector de la primera p_0 -componente de $\check{\beta}_n$ y β respectivamente.

El primer punto de la propiedad Oracle quiere decir que la probabilidad de que los parámetros estimados que sean cero cuando realmente lo son tiende a uno cuando n tiende a infinito, y el segundo punto se refiere a que los estimadores son normales asintóticamente.

3.4. Criterios para selección de modelos

Para saber si un modelo tiene las regresoras correctas debemos llevar a cabo procedimientos estadísticos para conocer con que criterio nos quedamos.

Esto es, dado un criterio se ordenan los modelos desde el mejor hasta

3.4. Criterios para selección de modelos

el peor y así se toma la decisión. En este caso se consideran cuatro criterios citados por Kleinbaum *et al.* (1998) que son: R_p^2 , F_p , $CME(p)$, y C_p , los cuales se mencionan a continuación:

3.4.1. Criterio R_p^2

El coeficiente de determinación o R^2 se define de la siguiente forma para un modelo de regresión lineal:

$$R_p^2 = \frac{SCR}{SCTotal} \quad (3.21)$$

Donde:

SCR = Suma de Cuadrados de la Regresión

SCTotal = Suma de Cuadrados Total

La SCTotal puede definirse como la variación total de los valores reales de Y con respecto a su media muestral y se calcula de la siguiente forma:

$$SCTotal = \sum (Y_i - \bar{Y})^2 \quad (3.22)$$

Además, $SCR = SCTotal - SCError$

Donde:

SCError = Suma de Cuadrados del Error

Modelos con mayor R_p^2 serán preferidos. Se sabe que el máximo R_p^2 se obtiene al incluir todas las variables explicativas al modelo, pero lo ideal es encontrar el modelo tal que R_p^2 aumente marginalmente por lo

3.4. Criterios para selección de modelos

que en la práctica lo que se hace es graficar esta situación y así se ve claramente cual es el mejor modelo.

3.4.2. Criterio F_p

El estadístico de prueba F o F_p sirve para comparar el modelo completo con el reducido y puede expresarse en terminos de la Suma de Cuadrados del Error como se muestra a continuación:

$$\begin{aligned} F_p &= \frac{[SCE_{MR} - SCE_{MC}]/(gl_{MC} - gl_{MR})}{SCE_{MC}/(n - gl_{MC} - 1)} \\ &= \frac{[SCE_{MR} - SCE_{MC}]/(gl_{MC} - gl_{MR})}{CME_{MC}} \end{aligned} \tag{3.23}$$

Donde:

SCE_{MR} = Suma de cuadrados del error del modelo reducido

SCE_{MC} = Suma de cuadrados del error del modelo completo

gl_{MR} = grados de libertad del modelo reducido

gl_{MC} = grados de libertad del modelo completo

n = el número de observaciones

CME_{MC} = Cuadrado medio del error del modelo completo

Este criterio prueba si $SCE_{MR} - SCE_{MC}$ difiere significativamente de 0. Si F_p es no significativa, entonces se utiliza el modelo reducido (p variables) y se verá si logra más o menos la misma capacidad predictiva que arroja el modelo completo. Por lo tanto, una regla razonable para seleccionar las variables es retener p variables si F_p no es significativa y si p es tan pequeño como sea posible.

3.4. Criterios para selección de modelos

3.4.3. Criterio $CME(p)$

El cuadrado medio del error o $CME(p)$ se define como sigue:

$$CME(p) = \frac{SCE(p)}{(n - p - 1)} \quad (3.24)$$

Donde:

$CME(p)$ = Cuadrado medio del error del modelo con p variables

$SCE(p)$ = Suma de cuadrados del error del modelo con p variables

Lo que se busca son modelos con el mínimo $CME(p)$ ya que ésto también implica encontrar un modelo con poca varianza residual.

3.4.4. C_p de Mallow

Mallow propuso el criterio de selección para un modelo el cual implica a la $SCE(p)$ el cual se define a continuación:

$$C_p = \frac{SCE(p)}{CME(k)} - [n - 2(p + 1)] \quad (3.25)$$

Donde:

$SCE(p)$ = Suma de cuadrados del error del modelo con p variables

$CME(k)$ = Cuadrado medio del error del modelo completo

El criterio C_p ayuda a decidir cuantas variables deben ir en el modelo ya que obtiene un valor próximo a $p+1$ si el $CME(p)$ es aproximadamente igual al $CME(k)$, es decir, si el modelo correcto es de tamaño p . Esto implica que si se conoce el tamaño del modelo correcto entonces ayudaría a elegir el mejor modelo.

3.5. Prueba de permutaciones

La prueba basada en permutaciones o también llamada prueba aleatorizada, fue propuesta por Fisher en 1935 y Pitman en 1937 [Citados por [Minsan \(2010\)](#)]. En el caso de un modelo lineal la prueba basada en permutación no requiere de supuestos. El supuesto necesario y de mayor peso para realizar una prueba de permutación es la intercambiabilidad, esto significa que, bajo la hipótesis nula, la distribución conjunta de las observaciones es invariante bajo las permutaciones. Es importante notar que las observaciones pueden ser intercambiables incluso si no independientes, pero pueden tener la misma probabilidad de cualquier distribución particular de resultados sin importar el orden en el cual son consideradas las observaciones (Good, 2002: 243-247) [Citado por [Minsan \(2010\)](#)].

3.5.1. Prueba de Permutación para un Modelo de Regresión Lineal Múltiple

Considerando la siguiente ecuación lineal para regresión múltiple

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \boldsymbol{\epsilon} = \mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (3.26)$$

Donde \mathbf{y} es un vector $\mathbf{nx1}$ de la variable respuesta, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ es una matriz \mathbf{nxp} de regresoras de rango \mathbf{p} , donde \mathbf{X}_1 es tamaño \mathbf{nxq} de rango \mathbf{q} y \mathbf{X}_2 es de tamaño de $\mathbf{nx}(\mathbf{p} - \mathbf{q})$ de rango $\mathbf{p} - \mathbf{q}$. $\boldsymbol{\tau} = (\boldsymbol{\theta}, \boldsymbol{\beta})'$ es un vector $\mathbf{px1}$ de coeficientes de regresión de \mathbf{y} , donde $\boldsymbol{\theta}$ es un vector de $\mathbf{qx1}$ y $\boldsymbol{\beta}$ es un vector de $(\mathbf{p} - \mathbf{q})\mathbf{x1}$ de X_1 y X_2 respectivamente, y $\boldsymbol{\epsilon}$ es un vector i.i.d. $\mathbf{nx1}$ del término del error aleatorio. De la ecuación (3.26),

3.5. Prueba de permutaciones

las ecuaciones normales en forma matricial, se muestran a continuación.

$$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \theta \\ \beta \end{pmatrix} = \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix} \quad (3.27)$$

Resolviendo la primera ecuación normal para θ nos da $\hat{\theta} = (X_1'X_1)^{-1}X_1'(y - X_2\beta)$. Cuando esta es absorbida dentro de la segunda ecuación, entonces tenemos que

$$\mathbf{X}_2' \left[\mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \right] \mathbf{X}_2 \beta = \mathbf{X}_2' \left[\mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \right] \mathbf{y},$$

O bien

$$\mathbf{X}_2' \mathbf{M}_{\mathbf{X}_1} \mathbf{X}_2 \beta = \mathbf{X}_2' \mathbf{M}_{\mathbf{X}_1} \mathbf{y}, \quad (3.28)$$

Donde: $\mathbf{M}_{\mathbf{X}_1} = \mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'$.

Note que, si \mathbf{X}_2 y \mathbf{X}_1 son ortogonales, entonces $\mathbf{M}_{\mathbf{X}_1} \mathbf{X}_1 = \mathbf{0}$ y $\mathbf{M}_{\mathbf{X}_1} \mathbf{X}_2 = \mathbf{X}_2$. Los estimadores se obtienen por medio de MCO como se muestra a continuación:

$$\begin{aligned} \hat{\beta}_{MCO} &= \left(X_2' M_{X_1} X_2 \right)^{-1} X_2' M_{X_1} y, \\ &= \beta + \left(X_2' M_{X_1} X_2 \right)^{-1} X_2' M_{X_1} \epsilon. \end{aligned} \quad (3.29)$$

Se desea probar la siguiente hipótesis nula: $H_0 : \beta = 0$ (por ejemplo que el vector de coeficientes de regresión parcial para \mathbf{X}_2 no es significativamente diferente de cero). Sea $F(y, X_1, X_2)$ el estadístico F convencional usado para probar la hipótesis nula. F se define algebraicamente en la siguiente ecuación.

$$F = \frac{y' \left[X (X'X)^{-1} X' - X_1 (X_1'X_1)^{-1} X_1' \right] y / (p - q)}{y' \left[X (X'X)^{-1} X' \right] y / (n - p)} \quad (3.30)$$

3.5. Prueba de permutaciones

Los grados de libertad para F son $p - q$ en el numerador y $n - p$ en el denominador. Para el parámetro de error, (Lafleur (1996:6) (citado por Minsan (2010))) expresa la suma de cuadrados para la hipótesis ajustada, $R(\beta | \theta)$, como sigue:

$$R(X_2 | X_1) = (X_2\beta + \epsilon)' M_{X_1} X_2 (X_2' M_{X_1} X_2)^{-1} X_2' M_{X_1} (X_2\beta + \epsilon) \quad (3.31)$$

Manly(1991: 250-261)(Citado por Minsan (2010)) propuso una prueba basada en permutaciones para los parámetros de una regresión lineal, en este método primeramente se calcula el estadístico F para probar $\beta = 0$ y después se compara con el estadístico de F resultante de la permutación en y . El parámetro estimado esta dado por:

$$\begin{aligned} \hat{\beta} &= \left(X_2' M_{X_1} X_2 \right)^{-1} X_2' M_{X_1} S' y, \\ &= \left(X_2' M_{X_1} X_2 \right)^{-1} X_2' M_{X_1} S' (X_1\theta + X_2\beta + \epsilon), \end{aligned} \quad (3.32)$$

y

$$R(X_2|X_1) = (X_1\theta + X_2\beta + \epsilon)' S M_{X_1} X_2 \left(X_2' M_{X_1} X_2 \right)^{-1} X_2' M_{X_1} S' (X_1\theta + X_2\beta + \epsilon).$$

Donde: S es una matriz de permutación con la propiedad de $S' S = I$ y

$$\begin{aligned} M_{S' X_1} &= I - S' X_1 [(S' X_1)' S' X_1]^{-1} (S' X_1)', \\ &= S' S - S' X_1 (X_1' X_1)^{-1} (S' X_1)', \\ &= S' \left[I - X_1 (X_1' X_1)^{-1} X_1' \right] S, \\ &= S' M_{X_1} S. \end{aligned} \quad (3.33)$$

Capítulo 4

Prueba $ML - ALasso$

El presente trabajo propone una prueba de especificación de un modelo de regresión lineal basada en los Multiplicadores de Lagrange y la regresión Lasso Adaptativo, en adelante nos referiremos a este método como $ML - ALasso$.

Para mostrar la prueba propuesta suponga que se tiene un modelo al cual se le realizará la prueba de una correcta especificación y que es el siguiente:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u} \quad (4.1)$$

Donde \mathbf{Y} es un vector de $\mathbf{n} \times \mathbf{1}$ con los valores de la variable dependiente, \mathbf{X} es una matriz de rango \mathbf{k} de $\mathbf{n} \times (\mathbf{k} + \mathbf{1})$, donde la primer columna está asociada con el intercepto y está formado por unos, \mathbf{k} es el número de variables independientes o regresoras, después del intercepto, cada columna de las restantes contiene los valores de cada una de las variables independientes, β es un vector de $(\mathbf{k} + \mathbf{1}) \times \mathbf{1}$ con los coeficientes del modelo y \mathbf{u} es el término de error y además $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I})$.

Se desea probar que el modelo está correctamente especificado, entonces, las hipótesis a contrastar son:

H_0 : El modelo está bien especificado

vs

H_a : El modelo no está bien especificado

Al igual que la prueba RESET, $ML - ALasso$ se basa en una regresión ampliada del modelo original que consiste en el siguiente modelo:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\theta + \mathbf{u} \quad (4.2)$$

Donde \mathbf{Z} , igual que \mathbf{X} , es una matriz y sus columnas contienen los valores de las variables independientes, y se construye introduciendo $\hat{\mathbf{Y}}$ o \mathbf{X} en forma de alguna potencia o interacción.

En la regresión auxiliar si no se rechaza la hipótesis $H_0 : \theta = 0$ vs la hipótesis $H_a : \theta \neq 0$ entonces el modelo esta bien especificado.

Para llevar a cabo la prueba sobre $H_0 : \theta = 0$, utilizando Multiplicadores de Lagrange incluye realizar una prueba de χ_c^2 tal como se define en la sección 3.2.2 "Multiplicador de Lagrange", es decir:

$$\chi_c^2 \sim \mathbf{nR}^2 \quad (4.3)$$

Donde χ_c^2 es el estadístico de prueba calculado, \mathbf{n} es el tamaño de la muestra y \mathbf{R}^2 es el coeficiente de determinación y se define la siguiente forma (Ver subsección 3.4.1):

$$R^2 = \frac{SCR}{SCTotal} \quad (4.4)$$

Donde: SCR = Suma de Cuadrados de la Regresión y SCTotal = Suma

de Cuadrados Total.

χ_c^2 bajo H_0 tienen una distribución χ^2 con grados libertad igual al número de coeficientes en θ .

Si el valor de χ_c^2 en la ecuación (4.3) excede el valor χ^2 crítico calculado de la función quantil al nivel de significancia seleccionado y grados de libertad igual al número de coeficientes que resultaron diferentes de cero, es decir, el número de coeficientes en θ ; si esto ocurre, entonces se rechaza la hipótesis nula.

Alternativamente, la prueba $ML - ALasso$ propone llevar a cabo tres acciones secuenciales:

- a) Obtener los residuos mediante MCO del modelo al que se le está probando una correcta especificación (ecuación 4.1).
- b) Realizar una regresión con Lasso Adaptativo al modelo auxiliar:

$$\mathbf{yn} = \mathbf{X}\beta + \mathbf{Z}\theta + \mathbf{u} \quad (4.5)$$

Donde \mathbf{yn} son los residuales obtenidos en el inciso **a**).

El parámetro de la regresión Lasso Adaptativo utilizado fue el valor del mínimo BIC (Criterio de Información de Bayes, por sus siglas en inglés). La prueba contempla llevar a cabo la regresión Lasso Adaptativo utilizando una modificación basada en el código **regresion.alasso** generado por Len Stefanski y Dennis Boos (Ver Anexo A), se requiere el paquete **lars** del software R para una correcta ejecución del programa.

Si todos los coeficientes θ son iguales a cero entonces decimos que la relación funcional es correcta, es decir, que se acepta la hipótesis nula;

4.1. Método $ML - ALasso - p$ para muestras pequeñas

de otra forma, si se obtienen valores diferentes de cero en al menos un coeficiente θ_i , entonces se realiza una prueba para saber si esos valores son significativos o no.

Si \mathbf{n} es grande, por la propiedad Oracle de la regresión Lasso Adaptativo de acuerdo a [Zou \(2006\)](#) (Ver subsección 3.3.2.) entonces \mathbf{n} asegura que la probabilidad de que los parámetros que sean cero realmente los haga cero tiende a uno cuando \mathbf{n} tiende a infinito y que los estimadores son normales asintóticamente.

c) Si se obtiene un valor diferente de cero en al menos un coeficiente θ_i , entonces se realiza una prueba para saber si esos valores son estadísticamente significativos o no.

A continuación se debe calcular el valor de χ^2 crítico al nivel de significancia seleccionado. Para muestras grandes se propone la prueba $ML - ALasso$.

4.1. Método $ML - ALasso - p$ para muestras pequeñas

Como se mencionó al principio de este capítulo la prueba de Multiplicadores de Lagrange es asintótica [[Engle \(1982\)](#)] por lo que se propone una prueba para muestras pequeñas a la cual le llamamos $ML - ALasso - p$.

Se llevan a cabo los incisos **a** y **b** de la prueba $ML - ALasso$.

c) Para realizar la inferencia sobre los coeficientes θ que resultan con valor diferente de cero, la propuesta es construir una prueba no paramétrica consistente en contratar el estadístico nR^2 con el cuantil re-

4.1. Método $ML - ALasso - p$ para muestras pequeñas

sultante de regresiones con permutaciones con intercambiabilidad en \mathbf{y} (Ver Anexo B), esto es, realizar regresiones Lasso Adaptativo al siguiente modelo auxiliar:

$$\mathbf{y}_{\mathbf{n}_p} = \mathbf{X}\beta + \mathbf{Z}\theta + \mathbf{u} \quad (4.6)$$

Donde $\mathbf{y}_{\mathbf{n}_p}$ son los residuales permutados que se obtuvieron en el inciso **a**).

Capítulo 5

Ejemplo de aplicación

Como ejemplo de aplicación de los dos métodos propuestos, **ML-ALasso** y **ML-ALasso-p**, se utilizan los datos del artículo titulado "El efecto del precio del tabaco sobre el consumo: un análisis de datos agregados para México" de [Olivera~Chávez *et al.* \(2010\)](#). El número de observaciones es $n=48$ y los datos son series trimestrales del año 1994 al 2005.

El modelo propuesto originalmente es el siguiente:

$$\ln c_{\mathbf{i}} = \beta_0 - \beta_1 \ln p_{\mathbf{i}} + \beta_2 \ln y_{\mathbf{i}} + T_{2i} + T_{3i} + T_{4i} + \epsilon_i$$

Donde: $\ln c$ es el logaritmo del consumo de cigarros por adulto, $\ln p$ es el logaritmo del precio trimestral real de los cigarros, $\ln y$ es el logaritmo del ingreso trimestral real por adulto, T_2 , T_3 y T_4 son variables indicadoras para los trimestres 2, 3 y 4 (las cuales se incluyen para tomar en cuenta la estacionalidad de la variable consumo) y ϵ es un error estacionario.

La regresión ampliada del modelo original es:

$$\ln c_{\mathbf{i}} = \beta_0 - \beta_1 \ln p_{\mathbf{i}} + \beta_2 \ln y_{\mathbf{i}} + T_{2i} + T_{3i} + T_{4i} + \mathbf{Z}_{\mathbf{i}}\theta + \epsilon_i$$

5. Ejemplo de aplicación

Donde: $\mathbf{Z}_i = [(\ln \mathbf{p}_i)^2, (\ln \mathbf{p}_i)^3, (\ln \mathbf{p}_i)^4, (\ln \mathbf{y}_i)^2, (\ln \mathbf{y}_i)^3, (\ln \mathbf{y}_i)^4, \ln \mathbf{p}_i * \ln \mathbf{y}_i]$.

Se busca probar el siguiente juego de hipótesis: $H_0 : \theta = 0$ vs $H_a : \theta \neq 0$, es decir, H_0 : El modelo está bien especificado vs H_a : El modelo no está bien especificado.

Al aplicar todas las pruebas los resultados son los siguientes:

Para ambos métodos, ML-ALasso y ML-ALasso-p, y las pruebas, RESET2 y RESET3 se obtiene que todas tienen un buen desempeño, esto es, coeficientes obtenidos al correr la regresión del inciso **b**) resultaron ser cero, por lo que se acepta la hipótesis nula y se concluye que el modelo está bien especificado a un nivel de significancia de $\alpha = 0.05$.

Capítulo 6

Estudio de simulación

Se realizó un estudio de simulación para evaluar el tamaño de la prueba y la potencia de las prueba de los métodos $ML - ALasso$ y $ML - ALasso - p$ (Ver Anexos C y D), y comparar su comportamiento con el que la prueba RESET, que es la prueba más aceptada para probar la especificación de un modelo.

6.1. Factores evaluados en el estudio de simulación

Los factores que se evaluaron en el estudio de simulación fueron:

- a) Número de variables en el modelo.
- b) Tamaño de muestra.
- c) Correlación entre las variables independientes (solo en el caso de 2 variables independientes)

El número de repeticiones para cada experimento varía dependiendo del método usado:

6.1. Factores evaluados en el estudio de simulación

Para evaluar el Método $ML - ALasso$ tenemos que el número de repeticiones en la simulación es de 10,000, para evaluar el Método $ML - ALasso - p$ el número de repeticiones fue de 100, esto, debido a que el tiempo que consume la simulación es muy alto.

6.1.1. Número de variables en el modelo bajo la hipótesis nula

Los modelos propuestos para probar una correcta especificación fueron dos, uno con una variable independiente y el otro con dos variables independientes.

La metodología para la generación de las matrices diseño se tomó del artículo de [Thursby \(1989\)](#). Las variables regresoras X_1 y X_2 se construyen de la siguiente forma:

$$\begin{aligned} X_{1t} &= \rho_1 X_{1t-1} + v_{1t} \\ X_{2t} &= \rho_2 X_{2t-1} + v_{2t} \end{aligned} \tag{6.1}$$

Donde $v_t = (v_{1t}, v_{2t})'$ son vectores i.i.d. y normales con media cero y matriz de varianzas y covarianzas $E(v_t v_t') = \{b_{ij}\}$ donde $b_{ij} = (1 - \rho_i^2)$ y $b_{12} = \lambda_{12}((1 - \rho_1^2)(1 - \rho_2^2))^{1/2}$. Esto es que X_{1t} y X_{2t} son un proceso autoregresivo y son normales con media cero, varianza uno, y coeficiente de correlación:

$$\rho_{12} = \lambda_{12}[(1 - \rho_1^2)(1 - \rho_2^2)]^{1/2} / (1 - \rho_1 \rho_2) \tag{6.2}$$

Los modelos propuestos para verificar una correcta especificación son:

Modelo 1. Con una variable independiente: $\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \epsilon$

6.2. Propiedades evaluadas

Modelo 2. Con dos variables independientes: $\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \epsilon$

Donde: \mathbf{y} es la variable dependiente, \mathbf{x}_1 es la variable explicativa construida X_{1t} , \mathbf{x}_2 es la variable explicativa construida X_{2t} , β son los coeficientes de los modelos y ϵ es el término de error.

Los errores en ambos modelos fueron generados i.i.d. $N(0, 2)$.

6.1.2. Tamaño de muestra

Para evaluar el comportamiento de las pruebas propuestas, se utilizaron tres tamaños de muestra. Los tamaños de muestra utilizados fueron: 20, 30 y 50.

6.1.3. Correlación entre las variables independientes

En el modelo con dos variables explicativas, se consideraron dos correlaciones entre las variables explicativas, los valores de las correlaciones fueron: de 0 y 0.9.

6.2. Propiedades evaluadas

Para los modelos a contrastar se evaluó el tamaño y la potencia de la prueba bajo diferentes situaciones que ordinariamente se presentan en modelos de regresión.

El tamaño de la prueba es la probabilidad de rechazar la hipótesis nula cuando ésta es verdadera y fue estimado con el porcentaje de

6.2. Propiedades evaluadas

rechazos en las situaciones en que los datos fueron generados con el mismo modelo que el que se tiene bajo la hipótesis nula.

La potencia de la prueba es la probabilidad de rechazar la hipótesis nula cuando ésta es falsa y se estima con la proporción de rechazos en los casos en que la hipótesis nula es falsa.

6.2.1. Modelos generadores de los datos bajo la hipótesis nula

En este caso, el modelo generador de los datos coincide con el modelo propuesto, al que se le probará una correcta especificación del modelo, esto es:

Modelo 1:

$$\mathbf{y} = \mathbf{1} + \mathbf{5x}_1 + \epsilon \quad (6.3)$$

El modelo generador de los datos es (6.3), y se le ajusta el modelo $y = \beta_0 + \beta_1 x_1 + \epsilon_i$

Modelo 2:

$$\mathbf{y} = \mathbf{1} + \mathbf{5x}_1 + \mathbf{3x}_2 + \epsilon \quad (6.4)$$

El modelo generador de los datos es (6.4), y se le ajusta el modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon_i$

El nivel de significancia para ambos modelos se realiza con un $\alpha = 0.05$.

6.2. Propiedades evaluadas

6.2.2. Potencia de la prueba

Para evaluar la potencia de la prueba en el caso del modelo con una variable explicativa, se tienen tres situaciones:

En la primera se estudio el caso en el que se omitió una variable explicativa que no tenía nada que ver con la variable explicativa presente en el modelo; el segundo caso corresponde a un modelo que omitió la variable explicativa en el modelo elevada al cuadrado, finalmente un modelo en donde se omitió la variable explicativa en el modelo elevada al cubo. En las tres situaciones, el modelo al que se le hizo la prueba de especificación corresponde al modelo $y = \beta_0 + \beta_1 x_1 + \epsilon_i$.

Los errores en fueron generados i.i.d. $N(0, 2)$. El cuadro 6.1 resume las tres situaciones.

Cuadro 6.1: Modelos generadores de los datos para las tres situaciones contempladas con el modelo 1 que tiene una sola variable explicativa

Modelo al que se le prueba la especificación del modelo	Situación	Modelo generador de los datos
$y = \beta_0 + \beta_1 x_1 + \epsilon_i$	a) Omisión de una variable que no tiene que ver con variable explicativa incluida en el modelo a probar	$y = 1 + 5\mathbf{x}_1 + 10\mathbf{x}_2 + \epsilon$
	b) Omisión de una variable que corresponde a la variable explicativa incluida en el modelo a probar al cuadrado	$y = 1 + 5\mathbf{x}_1 + 3.5\mathbf{x}_1^2 + \epsilon$
	c) Omisión de una variable que corresponde a la variable explicativa incluida en el modelo a probar al cubo	$y = 1 + 5\mathbf{x}_1 + 5\mathbf{x}_1^3 + \epsilon$

6.2. Propiedades evaluadas

En el caso del modelo con dos variables explicativas, el estudio de simulación contempló cuatro situaciones:

Un caso con la omisión del cuadrado de una de las variables explicativas, otros tres casos con la omisión del cubo de una de las variables explicativas, y otros tres casos se diferencian sólo en el valor de los coeficientes. En las tres situaciones, el modelo al que se le hizo la prueba de especificación corresponde al modelo $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \epsilon_i$.

Los errores fueron generados i.i.d. $N(0, 2)$. El cuadro 6.2 presenta un resumen de estas cuatro situaciones.

Cuadro 6.2: Modelos generadores de los datos para cuatro situaciones contempladas con el modelo 2 que tiene dos variables explicativas

Modelo al que se le prueba la especificación del modelo	Situación	Modelo generador de los datos
$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \epsilon_i$	a) Omisión de una variable que corresponde a una de las variables explicativas incluidas en el modelo a probar al cuadrado	$y = 1 + 5\mathbf{x}_1 + 3\mathbf{x}_2 + 6\mathbf{x}_1^2 + \epsilon$
	b) Omisión de una variable que corresponde a una de las variables explicativas incluidas en el modelo a probar al cubo	$y = 1 + 5\mathbf{x}_1 + 3\mathbf{x}_2 + 8\mathbf{x}_1^3 + \epsilon$
	c) Omisión de una variable que corresponde a una de las variables explicativas incluidas en el modelo a probar al cubo	$y = 1 + 5\mathbf{x}_1 + 3\mathbf{x}_2 + 2\mathbf{x}_1^3 + \epsilon$
	d) Omisión de una variable que corresponde a una de las variables explicativas incluidas en el modelo a probar al cubo	$y = 1 + 5\mathbf{x}_1 + 3\mathbf{x}_2 + \mathbf{x}_1^3 + \epsilon$

6.3. Prueba RESET

La prueba RESET (Ver Anexo E) se realiza con ayuda de la función **resettest** que se encuentra dentro del paquete **lmtest** también del software R; se utilizaron dos versiones de la prueba RESET, una es con la segunda potencia y potencia cúbica de \mathbf{y} estimada (RESET3) y la otra con la segunda potencia de \mathbf{y} estimada (RESET2).

Capítulo 7

Resultados y Discusión

En esta sección se presentan los resultados del ejercicio de simulación para evaluar el comportamiento de las pruebas propuestas y su comparación con la prueba RESET.

7.1. Tamaño de Prueba

7.1.1. Prueba *ML – ALasso*.

La prueba ML-ALasso está diseñada para muestras grandes, en esta sección se evalúa la potencia de esta prueba comparada con las pruebas RESET2 y RESET3.

Tamaño de la Prueba del Modelo 1. En este caso, el modelo generador de los datos corresponde al modelo 1 con una sola variable explicativa ($y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$) y se prueba que este modelo esté correctamente especificado.

Como se aprecia en el cuadro 7.1 la prueba RESET presenta el mejor

7.1. Tamaño de Prueba

desempeño para todos los tamaños de muestra, manteniendo los tamaños de prueba estimados muy cercanos al valor nominal ($\alpha = 0.05$). La prueba ML - ALasso tiene un buen comportamiento cuando el tamaño de muestra es grande ($n=50$), pero tiende a sobreestimar el tamaño de prueba para valores pequeños y medianos del tamaño de muestra.

Cuadro 7.1: Tamaño de la Prueba del Modelo 1: $y_i = 1 + 5x_{1i} + \epsilon_i$ para *ML - ALasso*.

n	RESET3	RESET2	ML-ALasso
20	0.0499	0.049	0.0926
30	0.0454	0.0498	0.0855
50	0.047	0.0473	0.0405

Thursby (1989) realizó la prueba RESET con un modelo equivalente al modelo 1 compuesto por una sola variable explicativa donde prueba una correcta especificación. En este caso la \mathbf{Z} de su prueba RESET está compuesta de el cuadrado, cubo y cuarta potencia de \mathbf{X} . Los resultados son que para un tamaño de muestra de ($n = 50$) la prueba tuvo un buen desempeño ya que obtuvo valores dentro de dos desviaciones estándar del nivel crítico que fue de 0.05.

Tamaño de la Prueba del Modelo 2. En este caso, el modelo generador de los datos corresponde al modelo 2 con dos variables explicativas ($y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$) y se prueba que este modelo esté correctamente especificado.

En el cuadro 7.2 se puede observar que las pruebas RESET2 y RESET3 tienen un buen comportamiento al tener los valores del tamaño de prueba estimado muy cercanos al valor nominal ($\alpha = 0.05$), en este caso la prueba *ML - ALasso* funciona bien en tamaños de muestra

7.1. Tamaño de Prueba

grandes, tanto para $\rho_{12} = 0$ como para $\rho_{12} = 0.9$, pero al igual que el modelo con una variable, tiende a sobreestimar el tamaño de prueba para valores pequeños y medianos del tamaño de muestra.

Cuadro 7.2: Tamaño de la Prueba del Modelo 2: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + 3\mathbf{x}_{2i} + \epsilon_i$ para *ML - ALasso*.

n	ρ_{12}	RESET3	RESET2	ML-ALasso
20	0	0.0502	0.0488	0.1434
	0.9	0.0487	0.0517	0.0775
30	0	0.0488	0.0492	0.0923
	0.9	0.0505	0.0518	0.1016
50	0	0.0492	0.0498	0.0483
	0.9	0.0489	0.0513	0.064

Thursby y Schmidt (Sep., 1977) realizaron pruebas equivalentes a la RESET3 y RESET2 a un modelo semejante al modelo 2 que tiene dos variables explicativas donde se prueba una correcta especificación. Los tamaños de muestra son los mismos ($n = 20, 30$ y 50); con un nivel de significancia de $\alpha = 0.05$. Los resultados fueron satisfactorios ya que obtuvieron valores dentro de dos desviaciones estándar del nivel crítico que fue de 0.05.

7.1.2. Prueba *ML - ALasso - p*.

La prueba ML-Alasso-p está diseñada para muestras pequeñas; en esta sección se evalúa el comportamiento de esta prueba comparada con las pruebas RESET2 y RESET3.

Tamaño de la Prueba del Modelo 1. En este caso, el modelo generador de los datos corresponde al modelo 1 con una sola variable explicativa ($y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$) y se prueba que este modelo esté co-

7.1. Tamaño de Prueba

rectamente especificado.

Como se aprecia en el cuadro 7.3 la prueba RESET3 presenta buen desempeño para todos los tamaños de muestra, manteniendo los tamaños de prueba estimados muy cercanos al valor nominal ($\alpha = 0.05$). La prueba RESET2 tiene un comportamiento errático ya que para un tamaño de muestra de $n = 30$ se obtuvo un valor de 0.01 y para los tamaños de muestra pequeña y grande ($n = 20$ y 50) se sobreestima el tamaño de prueba. La prueba ML-ALasso-p tiene un buen comportamiento para los tres tamaños de muestra ya que mantiene los tamaños de prueba estimados semejantes al valor nominal ($\alpha = 0.05$) o menor en el caso del tamaño de muestra de $n = 20$ que es de 0.02.

Cuadro 7.3: Tamaño de la Prueba del Modelo 1: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + \epsilon_i$ para
ML - ALasso - p.

n	RESET3	RESET2	ML-ALasso-p
20	0.04	0.08	0.02
30	0.05	0.01	0.05
50	0.06	0.08	0.05

Thursby (1989) realizó la prueba RESET con un modelo semejante al modelo 1 compuesto por una sola variable explicativa donde prueba una correcta especificación. En este caso la \mathbf{Z} de su prueba RESET está compuesta de el cuadrado, cubo y cuarta potencia de \mathbf{X} . Los resultados son que para un tamaño de muestra de ($n = 50$) la prueba tuvo un buen desempeño ya que obtuvo valores dentro de dos desviaciones estándar del nivel crítico que fue de 0.05.

Tamaño de la Prueba del Modelo 2. En este caso, el modelo generador de los datos corresponde al modelo 2 con dos variables explicativas ($y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$) y se prueba que este modelo esté correc-

7.1. Tamaño de Prueba

tamente especificado.

En el cuadro 7.4 se puede observar que las pruebas RESET3 y RESET2 tienen un buen comportamiento al tener los valores del tamaño de prueba estimados cercanos al valor nominal ($\alpha = 0.05$), excepto en el tamaño de muestra $n = 20$ con $\rho_{12} = 0$ de la prueba RESET3 donde se tiende a sobreestimar el tamaño de prueba. En este caso la prueba *ML-ALasso-p* funciona bien tanto para $\rho_{12} = 0$ como para $\rho_{12} = 0.9$ en los tres tamaños de muestra.

Cuadro 7.4: Tamaño de la Prueba del Modelo 2: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + 3\mathbf{x}_{2i} + \epsilon_i$ para *ML-ALasso-p*.

n	ρ_{12}	RESET3	RESET2	ML-ALasso-p
20	0	0.1	0.03	0.03
	0.9	0.07	0.06	0.02
30	0	0.04	0.04	0.01
	0.9	0.06	0.04	0.07
50	0	0.03	0.03	0.03
	0.9	0.06	0.04	0.03

Thursby y Schmidt (Sep., 1977) realizaron pruebas equivalentes a la RESET3 y RESET2 a un modelo semejante al modelo 2 que tiene dos variables explicativas donde se prueba una correcta especificación. Los tamaños de muestra son los mismos ($n = 20, 30$ y 50) con un nivel de significancia de $\alpha = 0.05$. Los resultados fueron satisfactorios ya que obtuvieron valores dentro de dos desviaciones estándar del nivel crítico que fue de 0.05.

7.2. Potencia de la prueba

7.2.1. Prueba $ML - ALasso$.

La prueba ML-Alasso está diseñada para muestras grandes, en esta sección se evalúa la potencia de esta prueba comparada con las pruebas RESET2 y RESET3.

Potencia de la prueba del Modelo 1. En este caso, el modelo generador de los datos corresponde a alguno de los modelos generados al omitir una variable en el modelo (ver cuadro 6.1). Se prueba la hipótesis de que el modelo $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$ esté correctamente especificado; cuando en realidad omitió una variable, que dependiendo del caso puede ser una variable que no tiene que ver con la variable explicativa del modelo, o la variable que se omitió es el cuadrado o el cubo de la variable explicativa incluida en el modelo.

En el cuadro 7.5 se muestran los resultados para los tres modelos generadores y los tres diferentes tamaños de muestra, tanto para $\rho_{12} = 0.0$ como para $\rho_{12} = 0.9$.

En el primer modelo (inciso **a** del cuadro 7.5) se prueba una variable que no tiene que ver con la variable explicativa del modelo original y el resultado mostrado en el cuadro 7.5 nos dice que ninguna de las tres pruebas, ni RESET3, ni RESET2, ni ML-ALasso funcionan bien en este caso.

El segundo modelo (inciso **b** del cuadro 7.5) es el que contiene el cuadrado de la variable explicativa del modelo original y se tiene que la potencia estimada de la prueba es satisfactoria para las tres pruebas

7.2. Potencia de la prueba

con valor de 1 (100 por ciento de rechazos).

Y en el tercer modelo (inciso **c** del cuadro 7.5) que contiene la variable explicativa del modelo original al cubo se aprecia que el resultado tiene un buen comportamiento para las pruebas RESET3 y ML-ALasso con la potencia de prueba estimada con valor de 1, y la prueba RESET2 tiene un comportamiento errático.

Cuadro 7.5: Potencia de la Prueba del Modelo 1: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + \epsilon_i$ para *ML - ALasso*.

n	Término omitido	ρ_{12}	RESET3	RESET2	ML-ALasso
20	a) $10\mathbf{x}_{2i}$	0	0.005	0.0002	0.0035
		0.9	0.616	0.0002	0.4335
	b) $3.5\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	1	1	1
	c) $5\mathbf{x}_{1i}^3$	0	1	1	1
		0.9	1	0.6681	1
30	a) $10\mathbf{x}_{2i}$	0	0	0	0
		0.9	0.0081	0.0094	0.0028
	b) $3.5\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	1	1	1
	c) $5\mathbf{x}_{1i}^3$	0	1	0	1
		0.9	1	0	1
50	a) $10\mathbf{x}_{2i}$	0	0	0	0
		0.9	0.0001	0.0003	0.0004
	b) $3.5\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	1	1	1
	c) $5\mathbf{x}_{1i}^3$	0	1	0.9995	1
		0.9	1	0.0026	1

Thursby (1989) realizó la prueba RESET con un modelo equivalente al del inciso **a** donde se omite una variable que no tiene nada que ver con la variable explicativa del modelo. En este caso la \mathbf{Z} de su prueba RESET está compuesta de el cuadrado, cubo y cuarta potencia de \mathbf{X} . Los resultados son: para un tamaño de muestra de 50 con $\rho_{12} = 0$ y 0.9 la prueba RESET no tiene un buen desempeño. Lo anterior coincide con los resultados de nuestras pruebas.

7.2. Potencia de la prueba

Así mismo, Thursby (1989) realizó la prueba RESET con un modelo equivalente al inciso **b** donde la variable omitida es la variable explicativa incluida en el modelo a probar al cuadro. En este caso la \mathbf{Z} de su prueba RESET está compuesta de el cuadrado, cubo y cuarta potencia de \mathbf{X} . Los resultados son que para un tamaño de muestra $n=50$, la prueba obtuvo un buen desempeño ya que el porcentaje de rechazos fue de cien por ciento, coincidiendo con el resultado de este trabajo.

Potencia de la prueba del Modelo 2. En este caso, el modelo generador de los datos corresponde a alguno de los modelos generados al omitir una variable en el modelo (ver cuadro 6.2). Se prueba la hipótesis de que el modelo $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$ esté correctamente especificado; cuando en realidad omitió una variable, que dependiendo del caso puede ser la variable explicativa incluida en el modelo al cuadrado o al cubo.

En el cuadro 7.6 se muestran los resultados para los cuatro modelos generadores y los tres diferentes tamaños de muestra, tanto para $\rho_{12} = 0$ como para $\rho_{12} = 0.9$.

En el primer modelo (inciso **a** del cuadro 7.6) se observa que las tres pruebas tienen un buen comportamiento al tener los valores de la potencia de la prueba cercanos o iguales a 1 (100 por ciento de rechazos).

Los siguientes tres modelos generadores contienen la variable explicativa al cubo con la diferencia de tener un valor distinto en sus coeficientes.

En el caso del segundo modelo (inciso **b** del cuadro 7.6) las tres pruebas funcionan bien excepto la prueba RESET2 con tamaño de muestra de $n = 30$ con $\rho_{12} = 0$.

7.2. Potencia de la prueba

Cuadro 7.6: Potencia de la Prueba del Modelo 2: $y_i = 1 + 5x_{1i} + 3x_{2i} + \epsilon_i$ para *ML - ALasso*.

n	Término omitido	ρ_{12}	RESET3	RESET2	ML-ALasso
20	a) $6x_{1i}^2$	0	1	1	1
		0.9	1	1	1
	b) $8x_{1i}^3$	0	1	1	1
		0.9	1	1	1
	c) $2x_{1i}^3$	0	0.8355	0.3206	0.8645
		0.9	0.9921	0.5738	0.9659
	d) x_{1i}^3	0	0.9983	0.0001	0.9996
		0.9	0.1812	0.0453	0.1819
30	a) $6x_{1i}^2$	0	0.8913	0.9544	1
		0.9	1	1	1
	b) $8x_{1i}^3$	0	1	0	1
		0.9	1	0.8565	1
	c) $2x_{1i}^3$	0	1	0	1
		0.9	0.9832	0.5375	0.8979
	d) x_{1i}^3	0	0.619	0.0304	0.6445
		0.9	0.9997	0.72	0.9948
50	a) $6x_{1i}^2$	0	0.9998	0.9999	1
		0.9	1	1	1
	b) $8x_{1i}^3$	0	1	1	1
		0.9	1	1	1
	c) $2x_{1i}^3$	0	1	0	1
		0.9	1	0.0064	1
	d) x_{1i}^3	0	0.7524	0.048	0.9168
		0.9	1	0.3755	1

El tercer modelo (inciso **c** del cuadro 7.6) tiene un buen comportamiento para las pruebas RESET3 y ML-ALasso con valores de la potencia de la prueba cercanos a 1. La prueba RESET2 no funciona bien en ningún caso.

Para el cuarto modelo (inciso **d** del cuadro 7.6) se tiene que la prueba RESET3 tiene un comportamiento errático con algunos valores cercanos o iguales a 1 en los siguientes casos: tamaño de muestra de $n=20$ con $\rho_{12} = 0$, tamaño de muestra de $n=30$ con $\rho_{12} = 0.9$ y tamaño de muestra de $n=50$ con $\rho_{12} = 0.9$. La prueba RESET2 no es favorable

7.2. Potencia de la prueba

en ningún caso y la prueba ML-ALasso muestra valores de la potencia de la prueba cercanos o iguales a 1 en todos los casos excepto para los siguientes dos: tamaño de muestra de $n=20$ con $\rho_{12} = 0.9$ y tamaño de muestra de $n=30$ con $\rho_{12} = 0$.

7.2.2. Prueba *ML – ALasso – p*.

La prueba ML-Alasso-p está diseñada para muestras pequeñas, en esta sección se evalúa el comportamiento de esta prueba comparada con las pruebas RESET2 y RESET3.

Potencia de la Prueba del Modelo 1. En este caso, el modelo generador de los datos corresponde a alguno de los modelos generados al omitir una variable en el modelo (ver cuadro 6.1). Se prueba la hipótesis de que el modelo $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$ esté correctamente especificado; cuando en realidad omitió una variable, que dependiendo del caso puede ser una variable que no tiene que ver con la variable explicativa del modelo, o la variable que se omitió es el cuadrado o el cubo de la variable explicativa incluida en el modelo.

En el cuadro 7.7 se muestran los resultados para los tres modelos generadores y los tres diferentes tamaños de muestra, tanto para $\rho_{12} = 0.0$ como para $\rho_{12} = 0.9$.

El primer modelo (inciso **a** del cuadro 7.7) muestra que no se favorece a ninguna de las tres pruebas lo cual se esperaba debido a que la variable explicativa a probar no tiene ninguna relación con el modelo original.

El segundo modelo (inciso **b** del cuadro 7.7) es el que contiene el cuadrado de la variable explicativa del modelo original y se tiene que la

7.2. Potencia de la prueba

potencia de la prueba es satisfactoria para las tres pruebas con 100 por ciento de rechazos.

Cuadro 7.7: Potencia de la Prueba del Modelo 1: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + \epsilon_i$ para $ML - ALasso - p$.

n	Término omitido	ρ_{12}	RESET3	RESET2	ML-ALasso-p
20	a) $10\mathbf{x}_{2i}$	0	0	0	0
		0.9	0	0	0
	b) $3.5\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	1	1	1
	c) $5\mathbf{x}_{1i}^3$	0	1	1	1
		0.9	1	1	1
30	a) $10\mathbf{x}_{2i}$	0	0	0	0
		0.9	0	0	0.63
	b) $3.5\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	1	1	1
	c) $5\mathbf{x}_{1i}^3$	0	1	0.02	1
		0.9	1	0	1
50	a) $10\mathbf{x}_{2i}$	0	0	0	0
		0.9	0	0	0
	b) $3.5\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	1	1	1
	c) $5\mathbf{x}_{1i}^3$	0	1	1	1
		0.9	1	0.06	1

Y el tercer modelo (inciso **c** del cuadro 7.7) que contiene la variable explicativa del modelo original al cubo se aprecia que el resultado tiene un buen comportamiento para las pruebas RESET3 y ML-ALasso-p. RESET2 presenta un comportamiento errático con algunos resultados favorables con la potencia de prueba estimada con valor de 1 pero deja de ser favorable en los siguientes casos: tamaño de muestra $n = 30$ con $\rho_{12} = 0$ y $\rho_{12} = 0.9$ y tamaño de muestra $n = 50$ con $\rho_{12} = 0.9$.

Thursby (1989) realizó la prueba RESET con un modelo equivalente al del inciso **a** donde se omite una variable que no tiene nada que ver con la variable explicativa del modelo. En este caso la \mathbf{Z} de su prueba RESET está compuesta de el cuadrado, cubo y cuarta potencia de \mathbf{X} .

7.2. Potencia de la prueba

Los resultados son: para un tamaño de muestra de 50 con $\rho_{12} = 0$ y 0.9 la prueba RESET no tiene un buen desempeño. Lo anterior coincide con los resultados de nuestras pruebas.

Así mismo, Thursby (1989) realizó la prueba RESET con un modelo equivalente al inciso **b** donde la variable omitida es la variable explicativa incluida en el modelo a probar al cuadro. En este caso la \mathbf{Z} de su prueba RESET está compuesta de el cuadrado, cubo y cuarta potencia de \mathbf{X} . Los resultados son que para un tamaño de muestra $n=50$, la prueba obtuvo un buen desempeño ya que el porcentaje de rechazos fue de cien por ciento, coincidiendo con el resultado de este trabajo.

Potencia de la Prueba del Modelo 2. El modelo generador de los datos corresponde a alguno de los modelos generados al omitir una variable en el modelo (ver cuadro 6.2). Se prueba la hipótesis de que el modelo $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$ esté correctamente especificado; cuando en realidad omitió una variable, que dependiendo del caso puede ser la variable explicativa al cuadrado o al cubo.

En el cuadro 7.8 se muestran los resultados para los cuatro modelos generadores y los tres diferentes tamaños de muestra, tanto para $\rho_{12} = 0$ como para $\rho_{12} = 0.9$.

En el primer modelo (inciso **a** del cuadro 7.8) se observa que con tamaños de muestra medianos y grandes ($n = 30$ y 50) las tres pruebas tienen un buen comportamiento con valores de la potencia de la prueba cercanos o iguales a 1 (100 por ciento de rechazos). Para el tamaño de muestra $n = 20$ se tuvo un comportamiento errático ya que se obtuvieron los dos siguientes casos con un desempeño bajo: pruebas RESET3 y RESET2 con $\rho_{12} = 0$.

7.2. Potencia de la prueba

Cuadro 7.8: Potencia de la Prueba del Modelo 2: $\mathbf{y}_i = \mathbf{1} + 5\mathbf{x}_{1i} + 3\mathbf{x}_{2i} + \epsilon_i$ para $ML - ALasso - p$.

n	Término omitido	ρ_{12}	RESET3	RESET2	ML-ALasso-p
20	a) $6\mathbf{x}_{1i}^2$	0	0.3	0.56	1
		0.9	1	1	1
	b) $8\mathbf{x}_{1i}^3$	0	1	0.98	1
		0.9	1	1	1
	c) $2\mathbf{x}_{1i}^3$	0	1	0	1
		0.9	0.06	0.05	0.01
	d) \mathbf{x}_{1i}^3	0	0.52	0.09	0.2
		0.9	0.16	0.02	0.08
30	a) $6\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	1	1	1
	b) $8\mathbf{x}_{1i}^3$	0	1	0	1
		0.9	1	1	1
	c) $2\mathbf{x}_{1i}^3$	0	0.94	0.26	0.86
		0.9	1	0.18	0.99
	d) \mathbf{x}_{1i}^3	0	0.89	0.31	0.52
		0.9	0.59	0.04	0.27
50	a) $6\mathbf{x}_{1i}^2$	0	1	1	1
		0.9	0.99	1	1
	b) $8\mathbf{x}_{1i}^3$	0	1	0	1
		0.9	1	1	1
	c) $2\mathbf{x}_{1i}^3$	0	1	0.06	1
		0.9	1	1	1
	d) \mathbf{x}_{1i}^3	0	0.99	0.02	0.93
		0.9	1	0.7	1

Los siguientes tres modelos generadores contienen la variable explicativa al cubo con la diferencia de tener un valor distinto en sus coeficientes.

En el caso del segundo modelo (inciso **b** del cuadro 7.8) las tres pruebas tienen buen comportamiento excepto la prueba RESET2 que presenta un comportamiento errático en los siguientes casos, con tamaño de muestra de $n = 30$ con $\rho_{12} = 0$ y tamaño de muestra de $n = 50$ con $\rho_{12} = 0$.

El tercer modelo (inciso **c** del cuadro 7.8) tiene un buen comporta-

7.2. Potencia de la prueba

miento para las pruebas RESET3 y ML-ALasso-p con valores cercanos o iguales a 1 excepto para los siguientes casos en que se aprecia un comportamiento errático: prueba RESET3 con tamaño de muestra de $n = 20$ con $\rho_{12} = 0.9$ y prueba ML-ALasso-p con tamaño de muestra de $n = 20$ con $\rho_{12} = 0.9$. La prueba RESET2 no funciona bien en ningún caso excepto para tamaño de muestra de $n = 50$ con $\rho_{12} = 0$.

Para el cuarto modelo (inciso **d** del cuadro 7.8) se tiene que la prueba RESET3 tiene valores cercanos o iguales a 1 solo en los siguientes casos: tamaño de muestra de $n=30$ con $\rho_{12} = 0$ y tamaño de muestra de $n=50$ con $\rho_{12} = 0$ y $\rho_{12} = 0.9$. La prueba RESET2 no es favorable en ningún caso y la prueba ML-ALasso-p muestra un comportamiento errático ya que se obtienen valores de la potencia de la prueba cercanos o iguales a 1 solo en los siguientes dos casos: tamaño de muestra de $n=50$ con $\rho_{12} = 0$ y $\rho_{12} = 0.9$.

Capítulo 8

Conclusiones y Recomendaciones

8.1. Conclusiones

8.1.1. Tamaño de prueba

- El método ML-ALasso propuesto tuvo un buen desempeño para tamaños de muestra grande ($n=50$) y lo sobreestima en tamaños de muestra medianos y pequeños.
- El método ML-ALasso-p tuvo un buen desempeño para los tres tamaños de muestras.
- RESET2 y RESET3 tuvieron buen comportamiento en cuanto al tamaño de prueba en ambos modelos para todos los tamaños de muestra.

8.1.2. Potencia de prueba

- Para el caso en que se omite una variable que no tiene nada que ver con el modelo, ninguna de las pruebas, RESET2, RESET3,

8.2. Recomendaciones

ML-ALasso y ML-ALasso-p, tienen un buen comportamiento.

- En el caso es que la variable omitida es el cuadrado de la variable explicativa incluida en el modelo, la prueba RESET2 (que contempla el cuadrado de \mathbf{y} estimada), RESET3, ML-ALasso y ML-ALasso-p, en general, se comportan excelente para ambos modelos y todos los tamaños de muestra. En el caso en que se tienen dos variables en el modelo, en tamaños de muestra pequeña ($n=20$) ML-ALasso-p tiene mejor comportamiento que RESET2 y RESET3.
- Cuando la variable omitida es el cubo de la variable explicativa incluida en el modelo, la prueba RESET2 tiene un comportamiento totalmente errático en ambos modelos, mientras que RESET3, ML-ALasso y ML-ALasso-p, en general, tienen un buen comportamiento, sin haber un dominio total en ninguna. En el caso del modelo con dos variables, para tamaño de muestra grande $n=50$ la prueba ML-ALasso se comporta mejor que la prueba RESET3.

8.2. Recomendaciones

- Las conclusiones del método ML-ALasso-p se deben tomar con reserva ya que solo se hicieron 100 repeticiones porque el tiempo que consume la simulación es muy alto.
- Se recomienda probar más modelos como por ejemplo la omisión de la interacción entre las variables incluidas en el modelo.
- En el caso en que la variable no tiene que ver con la variable explicativa incluida en el modelo a probar, es conveniente seguir en la búsqueda de una prueba que lo detecte.

Referencias

- Bera, A. K. y Jarque, C. M. (1982). Model Specification Test: A Simultaneous Approach. *Journal of Econometric*, 20, 59–82.
- Chatterjee, A. y Lahiri, S. N. (2011). Bootstrapping Lasso Estimators. *Journal of the American Statistical Association*, 106:494, 608–625.
- Engle, R. F. (1982). A General Approach to Lagrangian Multiplier Model Diagnostics. *Journal of Economic*, 20, 83–104.
- Greene, W. H. (1999). *Análisis Económico*. Prentice Hall Iberia.
- Gujarati, D. N. (2004). *Econometría*. United States Military Academy, West Point.
- Kleiber, C. y Zeileis, A. (2008). *Applied Econometrics with R*. Springer.
- Kleinbaum, D. G., Kupper, L. L., Muller, K. E. y Nizam, A. (1998). *Applied Regression Analysis and Other Multivariable Methods*. Duxbury Press.
- Maddala, G. S. (2001). *Introduction to Econometrics*. John Wiley & Sons.
- Minsan, P. (2010). *A Permutation Test for Partial Regression Coefficients on First-Order Autocorrelation*. Proyecto Fin de Carrera, School of Applied Statistics National Institute of Development Administration.
- Olivera Chávez, R. I., Cermeño Bazán, R., Sáenz de Miera Juárez, B., Jiménez Ruiz, J. A. y Reynales Shigematsu, L. M. (2010). El efecto del precio del tabaco sobre el consumo: un análisis de datos agregados para México. *Salud Pública de México*, vol. 52, suplemento 2, S197–S205.
- Ramsey, J. B. (1969). Test por specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, Volume 31. Issue 2, 350–371.
- Russell, D. y MacKinnon, J. (1999). *Econometric Theory and Methods*. Oxford University Press.

Referencias

- Thursby, J. G. (1989). A Comparison of Several Specification Error Tests for a General Alternative. *International Economic Review*, 30 No. 1, 217–230.
- Thursby, J. G. y Schmidt, P. (Sep., 1977). Some Properties of Test for Specification Error in a Linear Regression Model. *Journal of the American Statistical Association*, Vol. 72, No. 359, pp. 635–641.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, Volume 58, Issue 1, 267–288.
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101:476, 1418–1429.

Anexos

Anexo A

Código de la función Lasso Adaptativo.

La función lasso adaptativo **regresion.lasso** para R fue creado por Len Stefanski y Dennis Boos, ambos pertenecen al Departamento de Estadística de la Universidad de Carolina del Norte de los Estados Unidos de América. La adaptación se muestra a continuación.

```
regresion.lasso<-function(x,y){

# lasso adaptativo de lars con Criterio de Información de Bayes (BIC) como
criterio de selección
# éste usa la versión de "varianza conocida"
# el BIC con la Suma de Cuadrados del Error (RSS) /(Error del Cuadrado Medio
del modelo completo)
# se debe utilizar una versión reciente de R para poder usar normalize=FALSE
de lars

require(lars)
ok<-complete.cases(x,y)
x<-x[ok,] # elimina los na's
y<-y[ok] m<-ncol(x)
n<-nrow(x)
x<-as.matrix(x) #en caso de que x no sea una matriz

# estandarizar las variables como lars lo hace
one <- rep(1, n)
```

A. Código de la función Lasso Adaptativo.

```
meanx <- drop(one%*% x)/n
xc <- scale(x, meanx, FALSE)           # primero resta la media
normx <- sqrt(drop(one%*% (xc^2)))
names(normx) <- NULL
xs <- scale(xc, FALSE, normx)         # ahora reescala con norm (no sd)

out.ls=lm(y~xs)                       # MCO estimados con las variables estandarizadas
beta.ols=out.ls$coeff[2:(m+1)]         # MCO excepto para el intercepto
w=abs(beta.ols)                        # pesos para el lasso adaptativo
xs=scale(xs,center=FALSE,scale=1/w)    # xs veces los pesos
object=lars(xs,y,type="lasso",normalize=FALSE)

# obtención de mínimo BIC
# bic=log(n)*object$df+n*log(as.vector(object$RSS)/n) # versión RSS/n
sig2f=summary(out.ls)$sigma^2          #mse para el modelo completo
bic2=log(n)*object$df+as.vector(object$RSS)/sig2f
step.bic2=which.min(bic2)              # mínimo BIC

fit=predict.lars(object,xs,s=step.bic2,type="fit",mode="step")$fit
coeff=predict.lars(object,xs,s=step.bic2,type="coef",mode="step")$coefficients
coeff=coeff*w/normx                   # regresar a la escala original
st=sum(coeff !=0)                      # número distinto de cero
mse=sum((y-fit)^2)/(n-st-1)            # 1 para el intercepto
msemc=sum((y-fit)^2)

if(st>0) x.ind<-as.vector(which(coeff !=0)) else x.ind<-0
intercept=as.numeric(mean(y)-meanx%*%coeff)
return(list(fit=fit,st=st,mse=mse,x.ind=x.ind,
coeff=coeff,intercept=intercept,object=object,
bic2=bic2,step.bic2=step.bic2,msemc=msemc))
}
```

Anexo B

Código de la prueba de permutaciones.

```
permuta.func <- function(x,y,B,np)
  jiout <- matrix(NA,nrow=B,ncol=1)
  for(i in 1:B)
  {
    y_sample <- sample(y)
    moda <- lm(y_sample~x[,1:np])
    yne <- y_sample-fitted(moda)
    Reg_permuta <- regresion.lasso(x=x,y=yne)
    if(Reg_permuta$st==0) jiout[i,] <- 0
    if(Reg_permuta$st>0)
    scte <- sum((yne-mean(yne))^2)
    scre <- scte-Reg_permuta$msemc
    ji2calp <- n*(scre/scte)
    jiout[i,] <- ji2calp
    quaji2 <- quantile(jiout,probs=.95,na.rm=T)
    return(list(jiout=jiout,n=n,quaji2=quaji2))
  }
```

Anexo C

Código de la prueba

ML – ALasso.

Para realizar la prueba se genera la variable independiente x a la cual llamamos x_1 y la variable dependiente y ; además el término del error u se distribuye normal con media cero y varianza 2. A continuación se definen las demás variables de prueba:

```
x2 <- x^2
x3 <- x^3
x4 <- x^4
```

A modo de ejemplo se muestra el código del modelo 1 para probar una correcta especificación.

```
# MODELO 1 ORIGINAL
y <- 1 + 5*x1 + u                #modelo original o restringido
mod <- lm(y~x1)
yn <- y-fitted(mod)             #residuales modelo original
x <- cbind(x1,x2,x3,x4)         # x del modelo completo
x <- as.matrix(x)
y <- as.matrix(y)

#PRUEBA ML-ALASSO
Reg <- regresion.lasso(x=x,y=yn) #función Lasso adaptativo
Coef <- Reg$coeff
st <- sum(Coef !=0)             # coeficientes diferentes a cero
if(st > 0){
```


C. Código de la prueba $ML - ALasso$.

```
sct <- sum((yn-mean(yn))^2)           #suma de cuadrados total
scr <- sct-Reg$msemc                 #suma de cuadrados regresión
jical2 <- n*(scr/sct)               #ji cuadrada calculada
}
```

Anexo D

Código de la prueba

ML – ALasso – p.

Para realizar la prueba se genera la variable independiente x a la cual llamamos x_1 y la variable dependiente y ; además el término del error u se distribuye normal con media cero y varianza 2. A continuación se definen las demás variables de prueba:

```
x2 <- x^2
x3 <- x^3
x4 <- x^4
```

A modo de ejemplo se muestra el código del modelo 1 para probar una correcta especificación.

```
# MODELO 1 ORIGINAL
```

```
np <- 1                                #número de regresoras en la versión restringida
y <- 1 + 5*x1 + u                       #modelo original o restringido
mod <- lm(y~x1)
yn <- y-fitted(mod)                     #residuales modelo original
x <- cbind(x1,x2,x3,x4)                 # x del modelo completo
x <- as.matrix(x)
y <- as.matrix(y)
```

```
#PRUEBA ML-ALASSO-P
```

```
Reg <- regresion.lasso(x=x,y=yn)        #función Lasso adaptativo
Coef <- Reg$coeff
st <- sum(Coef !=0)                     # coeficientes diferentes a cero
```

D. Código de la prueba $ML - ALasso - p$.

```
if(st > 0){
  sct <- sum((yn-mean(yn))^2)      #ji cuadrada calculada
  scr <- sct-Reg$msemc           #suma de cuadrados de la regresión
  ji2cal <- n*(scr/sct)          #suma de cuadrados total
  Repr <- permuta.func(x=x,y=y,  #función de prueba de permutaciones
  B=500,np=np)
}}
```

Anexo E

Código de la prueba RESET.

```
# PRUEBA RESET
RT <- resettest(mod,power=3,type="fit")           #RESET3
RT2 <- resettest(mod,power=2,type="fit")         #RESET2
if (RT$p.value<.05) {
  reset <- reset+1                               #rechazo la hipótesis nula
  reseti <- 1
}
if (RT2$p.value<.05) {
  reset2 <- reset2+1                             #rechazo la hipótesis nula
  reset2i <- 1
}
```