The Coefficient of Determination in the Ridge Regression

Rodríguez Sánchez, A. (arsanchez@ugr.es) and García García, C. (garciaclaudia@ugr.es)

Phd. student at University of Granada

Salmerón Gómez, R. (romansg@ugr.es) and García García, C.B. (cbgarcia@ugr.es)

Quantitative Methods for Economics and Business

Campus Universitario de la Cartuja 18071 Granada (Spain)

RESUMEN

El coeficiente de determinación, R^2 , es una importante medida de la regresión lineal que representa la parte de variación de la variable dependiente que viene explicada por un conjunto de variables independientes, por lo tanto, mide la capacidad predictiva del modelo estimado. Para obtener este coeficiente en Mínimos Cuadrados Ordinarios (MCO), se utiliza la descomposición de suma de cuadrados. Por otro lado, cuando existe un problema de multicolinealidad en el modelo, es decir relación lineal entre las variables explicativas, la estimación por MCO aporta resultados inestables, por lo que se utilizan otros métodos de estimación alternativos, destacando la estimación Cresta (que consiste en la mitigación del problema desde un punto de vista numérico). En este trabajo se revisa si en el caso de la estimación Cresta la descomposición de suma de cuadrados se verifica y, en consecuencia, cómo se debería definir el coeficiente de determinación en este ambiente.

Palabras clave: [multicolinealidad, coeficiente de determinación, descomposición de suma de cuadrados, Regresión Cresta.]

Área Temática: [Métodos Cuantitativos para la Economía y la Empresa.]

ABSTRACT

The coefficient of determination, denoted R^2 , represents the proportion of total variation of dependent variable explained by the model. Thus, it is a relevant measure in linear regression that shows the predictive capacity of the estimated model. When the partition of the sum of squares (SS) holds in ordinary least squares (OLS) estimation, the R^2 can be expressed as the ratio of the explained variance to the total variance. However, when there is a problem of collinearity, the OLS estimation presents unstable results and others alternative estimation methods are applied such as the ridge regression. This work reviews if the partition of the sum of squares (SS) is verified in the ridge estimation and establishes how the coefficient of determination should be defined in this estimation method.

1 INTRODUCCIÓN

La regresión lineal se considera una técnica de gran interés, debido a que permite medir la relación lineal que existe entre una variable explicada denominada $\mathbf{Y}$ y un conjunto de variables explicativas denominadas $\mathbf{X}_1, \dots, \mathbf{X}_p$. Partiendo de Stock and Watson (2012) la pendiente de la recta que relaciona una variable independiente con la variable dependiente es el efecto de la variación en una unidad de $\mathbf{X}_i$ sobre $\mathbf{Y}$, es decir, la relación entre $\mathbf{X}_i$ e $\mathbf{Y}$.

Para medir la capacidad que tienen las variables independientes $\mathbf{X}_i$ para explicar a la variable dependiente $\mathbf{Y}$, se utiliza el coeficiente de determinación, R^2 , el cuál, según Wooldridge (2009), es un número que nos permite saber hasta qué punto la recta de regresión estimada a través de la estimación por Mínimos Cuadrados Ordinarios (MCO) se ajusta a los datos muestrales.

Sin embargo, cuando existe un problema de multicolinealidad, la estimación por MCO aporta resultados inestables, ya que teniendo en cuenta a Uriel et al. (1997) se define colinealidad como la existencia de relaciones lineales entre las variables explicativas del modelo lineal, cuando los estimadores obtenidos y la precisión de estos se ven sencillamente afectados. Por lo tanto, la medición de la relación de $\mathbf{X}_i$ sobre $\mathbf{Y}$ se ve afectada. Para solucionar este problema existen diferentes métodos alternativos de estimación, entre los cuáles uno de los más utilizados es el estimador Cresta. Este método se basa en solucionar el problema de la multicolinealidad desde un punto de vista puramente numérico.

Hoerl and Kennard (1970) propusieron la estimación cresta que consistía en mitigar los problemas de estimación que aparecen en un modelo de regresión lineal múltiple cuando las variables independientes no son ortogonales, consistente en la modificación de la matriz mediante la adición de una pequeña cantidad positiva en su diagonal.

Por tanto, será de interés obtener el coeficiente de determinación de la estimación cresta, debido a que nos permitirá medir la bondad de ajuste del modelo estimado a través del estimador Cresta, es decir, cómo se ajustan los datos estimados a los datos esperados. Si se desea calcular esta cantidad tal y como se hace en MCO, se tendrá que verificar la descomposición de suma de cuadrados, ya que el R^2 se obtiene a partir de esta descomposición.

El presente trabajo se estructura como sigue a continuación: en la sección 2 se presenta la descomposición de suma de cuadrados para la estimación cresta. En la sección 3 se obtiene el coeficiente de determinación para el estimador cresta y la monotonía de los factores que lo componen. En la sección 4 se presentan otros posibles problemas en la estimación cresta respecto a inferencia y, finalmente en la sección 5 se destacan las principales conclusiones del trabajo y futuras líneas de mejoras.

2 DESCOMPOSICIÓN DE SUMA CUADRADOS EN LA ESTIMACIÓN CRESTA

Partiendo de un modelo definido por n observaciones y p variables independientes, como sigue a continuación:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1)$$

donde $\mathbf{u}$ es la perturbación aleatoria (la cuál se presupone esférica), $\mathbf{X}_{n \times p}$ es una matriz de observaciones de la variable independiente (siendo $\mathbf{X}_1 = (1, 1, \dots, 1)^t$) y $\mathbf{Y}_{n \times 1}$ es una matriz de observaciones de la variable dependiente. El estimador Cresta de $\boldsymbol{\beta}$ viene dado por la siguiente expresión:

$$\hat{\boldsymbol{\beta}}(k) = (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t \mathbf{Y}, \quad k \geq 0. \quad (2)$$

Puesto que $\hat{\beta}(k) = \mathbf{Z}_k \hat{\beta}$, con $\mathbf{Z}_k = (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t \mathbf{X}$, se tiene que $\hat{\beta}(k)$ es un estimador sesgado de β cuando $k > 0$ y que su matriz de varianzas-covarianzas es:

$$\text{var}(\hat{\beta}(k)) = \text{var}(\mathbf{Z}_k \hat{\beta}) = \sigma^2 \cdot \mathbf{Z}_k (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{Z}_k.$$

Cuando $k = 0$, el estimador Cresta coincide con el de MCO.

Proposición 1 *Los residuos originados por el estimador Cresta, $\mathbf{e}(k) = \mathbf{Y} - \hat{\mathbf{Y}}(k) = \mathbf{Y} - \mathbf{X} \hat{\beta}(k)$, no tienen por qué sumar cero cuando $k > 0$.*

Demostración 1 *Por un lado, partiendo de un sistema de ecuaciones normales:*

$$(\mathbf{X}^t \mathbf{X} + k \mathbf{Id}) \hat{\beta}(k) = \mathbf{X}^t \mathbf{Y},$$

donde $\mathbf{X}_1 = (1, 1, \dots, 1)^t$, se tiene que la primera fila de $\mathbf{X}^t \mathbf{X} + k \mathbf{Id}$,

$$(n + k, \sum_{i=1}^n X_{2i}, \dots, \sum_{i=1}^n X_{pi}),$$

multiplicada por $\hat{\beta}(k) = (\hat{\beta}_1(k), \hat{\beta}_2(k), \dots, \hat{\beta}_p(k))$ es igual al primer elemento de $\mathbf{X}^t \mathbf{Y}$, $\sum_{i=1}^n Y_i$. Esto es:

$$\sum_{i=1}^n Y_i = (n + k) \hat{\beta}_1(k) + \hat{\beta}_2(k) \sum_{i=1}^n X_{2i} + \dots + \hat{\beta}_p(k) \sum_{i=1}^n X_{pi}.$$

Por otro lado:

$$\begin{aligned} e_i(k) &= Y_i - \hat{Y}_i(k) = Y_i - (\hat{\beta}_1(k) + \hat{\beta}_2(k) X_{2i} + \dots + \hat{\beta}_p(k) X_{pi}) \\ \sum_{i=1}^n e_i(k) &= \sum_{i=1}^n Y_i - (n \hat{\beta}_1(k) + \hat{\beta}_2(k) \sum_{i=1}^n X_{2i} + \dots + \hat{\beta}_p(k) \sum_{i=1}^n X_{pi}) = k \hat{\beta}_1(k). \end{aligned} \quad (3)$$

Es decir, $\sum_{i=1}^n e_i(k) = k \hat{\beta}_1(k)$ no tiene por qué ser cero. □

Corolario 1 *Como consecuencia de la Proposición 1, no se verifica que $\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i(k)$ cuando $k > 0$.*

Demostración 2 Como $\sum_{i=1}^n e_i(k) \neq 0$, se tiene que $\sum_{i=1}^n Y_i \neq \sum_{i=1}^n \hat{Y}_i(k)$ ya que:

$$Y_i = \hat{Y}_i(k) + e_i(k) \Rightarrow \sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i(k) + \sum_{i=1}^n e_i(k),$$

es decir, $\bar{Y} \neq \bar{\hat{Y}}(k)$. □

Corolario 2 Como consecuencia de la Proposición 1, no se verifica que $\sum_{i=1}^n (\hat{Y}_i(k) - \bar{Y})e_i(k) = 0$ cuando $k > 0$.

Demostración 3 Teniendo en cuenta que $\sum_{i=1}^n e_i(k) \neq 0$:

$$\sum_{i=1}^n (\hat{Y}_i(k) - \bar{Y})e_i(k) = \sum_{i=1}^n \hat{Y}_i(k)e_i(k) - \bar{Y} \sum_{i=1}^n e_i(k) \neq 0,$$

a no ser que $\hat{Y}_i(k) = \bar{Y}$, lo cual no tiene sentido. □

Teorema 1 A partir del Corolario 2 se puede afirmar que en la estimación Cresta no se verifica la descomposición de suma de cuadrados del MCO cuando $k > 0$.

Demostración 4

$$\begin{aligned} e_i(k) = Y_i - \hat{Y}_i(k) &\Rightarrow Y_i = \hat{Y}_i(k) + e_i(k) \Rightarrow Y_i - \bar{Y} = \hat{Y}_i(k) - \bar{Y} + e_i(k) \\ &\Rightarrow (Y_i - \bar{Y})^2 = (\hat{Y}_i(k) - \bar{Y})^2 + e_i(k)^2 + 2(\hat{Y}_i(k) - \bar{Y})e_i(k) \\ &\Rightarrow \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i(k) - \bar{Y})^2 + \sum_{i=1}^n e_i(k)^2 \\ &\quad + 2 \sum_{i=1}^n (\hat{Y}_i(k) - \bar{Y})e_i(k) \end{aligned} \quad (4)$$

como $\sum_{i=1}^n (\hat{Y}_i(k) - \bar{Y})e_i(k) \neq 0$, entonces no se cumple la descomposición de suma de cuadrados $SCT(k) = SCE(k) + SCR(k)$ análoga a la de MCO. □

3 COEFICIENTE DE DETERMINACIÓN EN LA ESTIMACIÓN CRESTA

El coeficiente de determinación, R^2 , es un valor muy utilizado en la regresión lineal ya que representa el ratio de variación de la variable explicada denominada $\mathbf{Y}$ explicado por un conjunto de variables explicativas denominadas $\mathbf{X}$ tras la estimación por MCO. En el modelo (1), este coeficiente se define como sigue:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 1 - \frac{SCR}{SCT}. \quad (5)$$

Si se parte de un modelo con solo una constante (conocido como modelo restringido) como el siguiente:

$$\mathbf{Y} = \beta_1 + \mathbf{v}, \quad (6)$$

donde $\mathbf{v}$ es la perturbación aleatoria (la cuál se presupone esférica), se verifica que su suma de cuadrados de los residuos, SCR_r , coincide con la suma de cuadrados totales, SCT , del modelo (1) ya que:

$$SCR_r = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Por lo tanto, se obtiene la siguiente expresión alternativa para la expresión (5):

$$R^2 = 1 - \frac{SCR}{SCR_r}. \quad (7)$$

A partir de esta expresión se tiene que:

- Si $R^2 \simeq 0$ se tiene que la SCR del modelo (1) es prácticamente igual a la del modelo restringido (6), por lo tanto, la inclusión de variables es innecesaria. Las variables no proporcionan información reciente, por lo que el modelo propuesto no es adecuado.

- Si $R^2 \simeq 1$ se tiene que el SCR del modelo (1) es muy pequeño en contraste con el modelo restringido (6), en consecuencia, las variables introducidas proporcionan información que lo hacen preferible al modelo restringido.

Además, si en el modelo (1) hay término independiente se verifica que $0 \leq R^2 \leq 1$ y teniendo en cuenta la descomposición en suma de cuadrados, la expresión (5) se puede expresar como:

$$R^2 = \frac{SCE}{SCT}. \quad (8)$$

Puesto que en la estimación Cresta no se verifica la descomposición de la suma de cuadrados (ver Teorema 1), el coeficiente de determinación se ha de calcular a partir de la expresión (5) o (8). Teniendo en cuenta la interpretación realizada de la expresión (7), consideramos oportuno definirlo como:

$$R^2(k) = 1 - \frac{SCR(k)}{SCR_r(k)}, \quad k \geq 0. \quad (9)$$

Como es evidente, este coeficiente de determinación ha de depender del parámetro k y se ha de verificar que para $k = 0$ coincida con la expresión dada en MCO.

3.1 Monotonía de $R^2(k)$

En la presente subsección se va a estudiar la monotonía de $R^2(k)$ con respecto a k .

Proposición 2 $SCR_r(k) = \mathbf{Y}^t \mathbf{Y} - \frac{n+2k}{(n+k)^2} \cdot n^2 \bar{Y}^2$ y es creciente en k .

Demostración 5 *El estimador Cresta de β para el modelo restringido (6), depende del parámetro k y viene dado por la siguiente expresión:*

$$\hat{\beta}_r(k) = (\mathbf{1}^t \mathbf{1} + k)^{-1} \mathbf{1}^t \mathbf{Y}, \quad k \geq 0, \quad (10)$$

Teniendo en cuenta la expresión (10):

$$\begin{aligned}
 SCR_r(k) &= \mathbf{e}_r(k)^t \mathbf{e}_r(k) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_r(k))^t (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_r(k)) \\
 &= \mathbf{Y}^t \mathbf{Y} - 2\hat{\boldsymbol{\beta}}_r(k)^t \mathbf{1}^t \mathbf{Y} + \hat{\boldsymbol{\beta}}_r(k)^t \mathbf{1}^t \mathbf{1} \hat{\boldsymbol{\beta}}_r(k) \\
 &= \mathbf{Y}^t \mathbf{Y} - 2\mathbf{Y}^t \mathbf{1} (\mathbf{1}^t \mathbf{1} + k)^{-1} \mathbf{1}^t \mathbf{Y} \\
 &\quad + \mathbf{Y}^t \mathbf{1} (\mathbf{1}^t \mathbf{1} + k)^{-1} \mathbf{1}^t \mathbf{1} (\mathbf{1}^t \mathbf{1} + k)^{-1} \mathbf{1}^t \mathbf{Y} \\
 &= \mathbf{Y}^t \mathbf{Y} - 2 \left(\sum_{i=1}^n Y_i \right) (n+k)^{-1} \left(\sum_{i=1}^n Y_i \right) \\
 &\quad + \left(\sum_{i=1}^n Y_i \right) (n+k)^{-1} \cdot n \cdot (n+k)^{-1} \left(\sum_{i=1}^n Y_i \right) \\
 &= \mathbf{Y}^t \mathbf{Y} - 2 \cdot \frac{\left(\sum_{i=1}^n Y_i \right)^2}{(n+k)} + \frac{\left(\sum_{i=1}^n Y_i \right)^2}{(n+k)^2} \cdot n = \mathbf{Y}^t \mathbf{Y} + \frac{\left(\sum_{i=1}^n Y_i^2 \right) \cdot (-n-2k)}{(n+k)^2} \\
 &= \mathbf{Y}^t \mathbf{Y} - \frac{n+2k}{(n+k)^2} \cdot n^2 \bar{Y}^2.
 \end{aligned}$$

Además, se verifica que $SCR_r(k)$ incrementa en k puesto que:

$$\begin{aligned}
 \frac{dSCR_r(k)}{dk} &= \frac{-2n^2 \bar{Y}^2 (n+k)^2 + 2(n+k)(n+2k)n^2 \bar{Y}^2}{(n+k)^4} \\
 &= \frac{2n^2 \bar{Y}^2 (n+k)k}{(n+k)^4} > 0,
 \end{aligned}$$

ya que $k \geq 0$ y $n > 0$. □

Proposición 3 $SCR(k) = \mathbf{Y}^t \mathbf{Y} - 2 \sum_{i=1}^p \frac{\alpha_i^2}{\lambda_i + k} + \sum_{i=1}^p \frac{\alpha_i^2 \lambda_i}{(\lambda_i + k)^2}$ y es creciente en k .

Demostración 6 Siguiendo los pasos de McDonald (2010), descomponemos $\mathbf{X}^t \mathbf{X}$ en función de sus autovalores y autovectores, esto es, $\mathbf{X}^t \mathbf{X} = \mathbf{\Gamma} \mathbf{D}_\lambda \mathbf{\Gamma}^t$, con $\mathbf{\Gamma}^t = \mathbf{\Gamma}^{-1}$ y $\mathbf{D}_\lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$. Por tanto:

$$(\mathbf{X}^t \mathbf{X} + k\mathbf{Id}) = \mathbf{\Gamma}(\mathbf{D}_\lambda + k\mathbf{Id})\mathbf{\Gamma}^t = \mathbf{\Gamma} \mathbf{D}_{\lambda+k} \mathbf{\Gamma}^t.$$

Llamando $\gamma = \mathbf{X}^t \mathbf{Y}$ y $\alpha = \Gamma^t \gamma$, entonces:

$$\begin{aligned}
 SCR(k) &= \mathbf{e}(k)^t \mathbf{e}(k) = (\mathbf{Y} - \mathbf{X} \hat{\beta}(k))^t (\mathbf{Y} - \mathbf{X} \hat{\beta}(k)) \\
 &= \mathbf{Y}^t \mathbf{Y} - 2 \hat{\beta}(k)^t \mathbf{X}^t \mathbf{Y} + \hat{\beta}(k)^t \mathbf{X}^t \mathbf{X} \hat{\beta}(k) \\
 &= \mathbf{Y}^t \mathbf{Y} - 2 \mathbf{Y}^t \mathbf{X} (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t \mathbf{Y} \\
 &\quad + \mathbf{Y}^t \mathbf{X} (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t \mathbf{X} (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t \mathbf{Y} \\
 &= \mathbf{Y}^t \mathbf{Y} - 2 \gamma^t \mathbf{D}_{\frac{1}{\lambda+k}} \Gamma^t \gamma + \gamma^t \mathbf{D}_{\frac{1}{\lambda+k}} \Gamma^t \mathbf{D}_{\lambda} \Gamma^t \mathbf{D}_{\frac{1}{\lambda+k}} \Gamma^t \gamma \\
 &= \mathbf{Y}^t \mathbf{Y} - 2 \alpha^t \mathbf{D}_{\frac{1}{\lambda+k}} \alpha + \alpha^t \mathbf{D}_{\frac{1}{\lambda+k}} \mathbf{D}_{\lambda} \mathbf{D}_{\frac{1}{\lambda+k}} \alpha \\
 &= \mathbf{Y}^t \mathbf{Y} - 2 \sum_{i=1}^P \frac{\alpha_i^2}{\lambda_i + k} + \sum_{i=1}^P \frac{\alpha_i^2 \lambda_i}{(\lambda_i + k)^2} \\
 &= \mathbf{Y}^t \mathbf{Y} - \sum_{i=1}^p \frac{(\lambda_i + 2k) \alpha_i^2}{(\lambda_i + k)^2}.
 \end{aligned}$$

Además, se verifica que $SCR(k)$ incrementa con k :

$$\frac{dSCR(k)}{dk} = - \frac{2\alpha_i^2(\lambda_i + k)^2 - 2(\lambda_i + 2k)\alpha_i^2(\alpha_i + k)}{(\lambda_i + k)^4} = \frac{2\alpha_i^2(\lambda_i + k)k}{(\lambda_i + k)^4} > 0,$$

ya que $k \geq 0$ y $\lambda_i > 0$ al ser los autovalores de una matriz simétrica. $\square$

Puesto que tanto $SCR(k)$ y $SCR_r(k)$ crecen conforme aumenta k , la monotonía de $R^2(k)$ según la expresión dada en (9) depende de la *velocidad* con que lo haga cada una. Por tanto, se hace necesario un estudio más profundo sobre este aspecto.

4 INFERENCIA EN LA ESTIMACIÓN CRESTA

Siguiendo el mismo razonamiento que en la estimación por MCO del modelo (1) y que $\mu \sim N(0, \sigma^2 \mathbf{Id})$ se tiene que $\hat{\beta}(k) \sim N(\mathbf{Z}_k \beta, \sigma^2 \mathbf{Z}_k (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{Z}_k)$ donde $\mathbf{Z}_k = (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t \mathbf{X}$ ya que $\hat{\beta}(k) = \mathbf{Z}_k \hat{\beta}$.

Además, $\mathbf{e}(k) = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X} \hat{\beta}(k) = \mathbf{Y} - \mathbf{X} (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t \mathbf{Y} = \mathbf{M}(k) \mathbf{Y}$, donde $\mathbf{M}(k) = \mathbf{Id} - \mathbf{X} (\mathbf{X}^t \mathbf{X} + k \mathbf{Id})^{-1} \mathbf{X}^t$ es simétrica pero no es idempotente y $\mathbf{M}(k) \mathbf{X} \neq 0$.

Por tanto, no es posible reproducir la estimación de σ^2 como:

$$\hat{\sigma}^2(k) = \frac{\mathbf{e}^t(k)\mathbf{e}(k)}{n - p},$$

y tampoco deducir una distribución asociada.

5 CONCLUSIONES

El coeficiente de determinación es un valor ampliamente utilizado para analizar la bondad de ajuste de una estimación en un modelo de regresión lineal, y de esta forma poder observar cómo se ajustan los datos estimados a los datos esperados. La expresión de este coeficiente en MCO se obtiene a partir de la descomposición de suma de cuadrados $SCT = SCE + SCR$.

Por otro lado, cuando existe un problema de multicolinealidad en el modelo de regresión se hace necesario utilizar métodos de estimación alternativos, debido a que en esta situación al estimar por MCO se pueden obtener resultados inestables. Uno de los métodos alternativos más utilizados es la regresión Cresta, por esta razón en este trabajo se obtiene su coeficiente de determinación.

Como principales conclusiones del presente trabajo se tiene que:

- En la estimación Cresta no se cumple la descomposición de suma de cuadrados, utilizada para obtener el R^2 de MCO.
- Por este motivo, no se puede generalizar la obtención del R^2 usada en MCO a la estimación Cresta.

Como futuras líneas de mejora del presente trabajo destaca la realización de una simulación para analizar la monotonía de $R^2(k)$.

References

- Hoerl, A. and R. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55–67.
- McDonald, G. (2010). Tracing ridge regression coefficients. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(6), 695–703.
- Stock, J. and M. Watson (2012). *Introducción a la Econometría (3ª Edición)*. Madrid: Ed. Pearson.
- Uriel, E., A. Periró, D. Contreras, and M. Moltó (1997). *Econometría: El Modelo Lineal*. Madrid: Ed. Alfa Centauro.
- Wooldridge, J. (2009). *Introductory Econometrics: A modern approach*. Canada: South-Western Cengage Learning.