ECONOMETRÍA

JOSÉ ALBERTO MAURICIO

3

REGRESIÓN LINEAL MÚLTIPLE II

3.3 HETEROSCEDASTICIDAD

Departamento de Análisis Económico y Economía Cuantitativa
Universidad Complutense de Madrid

EctrGr-JAM-3-3.pdf

ucm.randomshock.com/ectrgr

Versión 2.4 - Enero 2024

REQUISITOS

En esta Sección 3.3 se suponen bien conocidos [1] los aspectos teóricos y prácticos del Análisis de Regresión Lineal cubiertos en los Temas 1 - 2 y en las Secciones 3.1 - 3.2 del Tema 3, y [2] todos los procedimientos descritos en las Secciones 1 - 15 de la guía *Introducción al Uso de EViews 4.1*.

BIBLIOGRAFÍA PARA LA SECCIÓN 3.3

Hill, Griffiths, Lim (2018): Capítulo 8.
Wooldridge (2020): Capítulo 8.

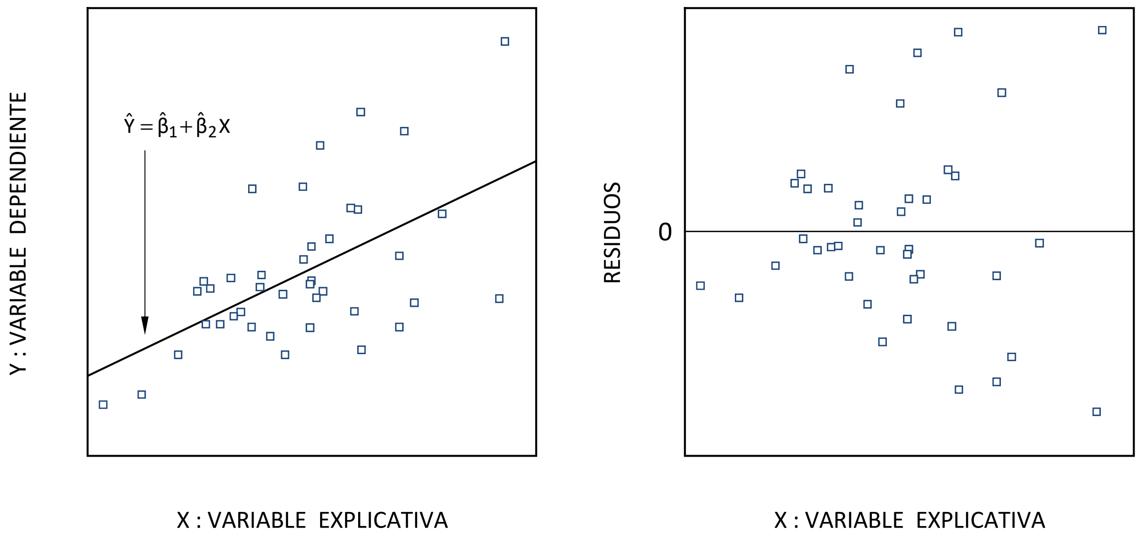
Sección 16 de la guía *Introducción al Uso de EViews 4.1*.

PARTE 1 - DEFINICIÓN - CAUSAS - CONSECUENCIAS

Un modelo de regresión presenta **heteroscedasticidad** cuando sus perturbaciones **no** tienen **varianza constante**. En la Figura 1 se ilustra esta posibilidad.

FIGURA 1
Modelo RLS Estimado por MCO con Síntomas de Heteroscedasticidad

Gasto en alimentación (Y) e ingresos (X) medios semanales (euros) de 40 familias en el año 2020.
Datos en SC09: Y = GALIM, X = INGR.



➪ Modelo RLS para los datos de la Figura 1 con **heteroscedasticidad proporcional**:

$$Y_i = \beta_1 + \beta_2 X_i + U_i \,, \text{ con } \mathrm{E}[U_i \mid \mathbf{X}] = 0, \tag{1}$$

$$\mathrm{Var}[U_i \mid \mathbf{X}] = \sigma^2 X_i \;\; (i = 1, ..., N) \;\; (\sigma^2 > 0). \tag{2}$$

[1]-[2] $\Rightarrow$ $$\mathrm{E}[Y_i \mid \mathbf{X}] = \beta_1 + \beta_2 X_i \,, \; \mathrm{Var}[Y_i \mid \mathbf{X}] = \sigma^2 X_i \;\; (i = 1, ..., N). \tag{3}$$

En muchas ocasiones (al igual que en este ejemplo), la causa de la heteroscedasticidad es la existencia de alguna relación entre la dispersión (varianza) de la variable dependiente y los valores de alguna/s variable/s explicativa/s en el modelo considerado. Esa relación, a su vez, suele tener que ver con los diferentes **tamaños** de las entidades observadas (familias, empresas, ciudades, ...) en una colección de datos de **sección cruzada**, por lo que la presencia de heteroscedasticidad siempre debe considerarse cuando se pretende elaborar un modelo con datos de ese tipo.

➪ Modelo RLM con **heteroscedasticidad general**:

$$Y_i = \beta_1 + \beta_2 X_{i2} + ... + \beta_K X_{iK} + U_i \,, \text{ con } \mathrm{E}[U_i \mid \mathbf{X}] = 0, \tag{4}$$

$$\mathrm{Var}[U_i \mid \mathbf{X}] = \sigma^2 \times v(X_{i2}, ..., X_{iK}) \;\; (i = 1, ..., N) \;\; (\sigma^2 > 0), \tag{5}$$

donde $v(\cdot)$ es alguna función tal que $v(X_{i2}, ..., X_{iK}) > 0$ para todo $i = 1, \ldots, N$.

El modelo RLM [4]-[5] se puede escribir como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}, \text{ con } \mathrm{E}[\mathbf{U} \mid \mathbf{X}] = \mathbf{0}, \quad [6]$$

$$\mathrm{Var}[\mathbf{U} \mid \mathbf{X}] \ (= \mathrm{E}[\mathbf{U}\mathbf{U}' \mid \mathbf{X}]) = \sigma^2 \mathbf{V}, \quad [7]$$

donde si $v_i = v(X_{i2}, ..., X_{iK})$, entonces $\mathbf{V}$ es la matriz **diagonal no escalar** siguiente:

$$\mathbf{V} = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_N \end{bmatrix} \neq \mathbf{I}. \quad [8]$$

Remplazando en el modelo RLM clásico la hipótesis de que $\mathrm{Var}[\mathbf{U} \mid \mathbf{X}] = \sigma^2 \mathbf{I}$ por [7]-[8], se puede comprobar que

$$\mathrm{E}[\hat{\boldsymbol{\beta}}_{\mathbf{W}}] = \boldsymbol{\beta},$$

$$\mathrm{Var}[\hat{\boldsymbol{\beta}}_{\mathbf{W}} \mid \mathbf{X}] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \neq \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}. \quad [9]$$

Ningún resultado basado en la expresión $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ para $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{\mathbf{W}} \mid \mathbf{X}]$ es aplicable ahora (en particular, el Teorema de Gauss-Markov). Además, cualquier cálculo basado en la estimación $\hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$ de la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\beta}}_{\mathbf{W}}$ es **incorrecto** (en particular, los errores estándar habituales de los estimadores MCO).

PARTE 2 - DETECCIÓN

➪ Gráficos de los residuos sobre cada variable explicativa (como en la Figura 1), o sobre los valores ajustados (que son una combinación lineal de todas las variables explicativas).

➪ **Contraste de White**:

1. Hipótesis nula (H_0): homoscedasticidad (varianzas constantes).
2. Hipótesis alternativa (H_1): heteroscedasticidad (varianzas no constantes).
3. Estadístico: $WH = N \times R^2_{WH}$, donde si $\hat{u}_1, \hat{u}_2, ..., \hat{u}_N$ son los residuos MCO de [4], entonces R^2_{WH} es el coeficiente de determinación en la **regresión auxiliar** de $\hat{u}_i^2$ sobre todas las variables explicativas de [4], sus cuadrados, y los productos cruzados no redundantes entre ellas.
4. Nivel de significación marginal: $\alpha^* = \Pr[\chi^2(M) \geq WH]$, donde M es el número de pendientes estimadas en la regresión auxiliar anterior. Si α^* es suficientemente pequeño (menor que el nivel de significación escogido), rechazar H_0 en favor de H_1.

EJEMPLO 1: Con los datos del archivo SC09 (GALIM, INGR),

$$\begin{array}{ccccc} \hat{Y} & = & 36.6908 & + & 0.1283X, \\ & & (19.9248) & & (0.0305) \\ & & [0.0734] & & [0.0002] \end{array} \quad [10]$$

$$N = 40, \quad WH = 14.5820 \ [0.0007].$$

En [10], $WH = 40 \times 0.36455$, donde 0.36455 es el coeficiente de determinación en la regresión con término constante de los cuadrados de los residuos sobre X, X^2 ($M = 2$). El nivel de significación marginal del contraste de White es

$$\alpha^* = \Pr[\chi^2(2) \geq 14.5820] = 0.0007.$$

EJEMPLO 2: Con los datos del archivo SC01 (PRECIO, SUPM2, FINCAM2, NDORM),

$$\begin{array}{ccccccccc} \hat{Y} & = & 7.6273 & + & 1.2785X_2 & + & 0.0214X_3 & + & 18.3902X_4, \\ & & (31.0867) & & (0.1459) & & (0.0070) & & (9.3628) \\ & & [0.8068] & & [0.0000] & & [0.0030] & & [0.0532] \end{array} \qquad [11]$$

$$N = 80, \quad WH = 31.2699 \ [0.0003].$$

En [11], $WH = 80 \times 0.39087$, donde 0.39087 es el coeficiente de determinación en la regresión con término constante de los cuadrados de los residuos sobre X_2, X_3, X_4, X_2^2, X_3^2, X_4^2, X_2X_3, X_2X_4, X_3X_4 ($M = 9$). El nivel de significación marginal del contraste de White es

$$\alpha^* = \Pr[\chi^2(9) \geq 31.2699] = 0.0003.$$

Observación 1: El contraste de **Breusch-Pagan** es idéntico al de White excepto por que en la regresión auxiliar sólo se incluyen las variables explicativas de [4] (ni sus cuadrados ni los productos cruzados entre ellas); en este caso, la distribución que se emplea para calcular α^* es una $\chi^2(K-1)$. Por otro lado, a veces se utiliza una versión del contraste de White en cuya regresión auxiliar sólo se incluyen (además del término constante) los valores ajustados MCO de [4] y sus cuadrados ($\hat{y}_i$, $\hat{y}_i^2$); en este caso, la distribución que se emplea para calcular α^* es una $\chi^2(2)$. Para más detalles, ver Wooldridge (2020): Sección 8-3.

PARTE 3 - UTILIZACIÓN ADECUADA DE MCO

El **estimador de White** es un estimador **robusto** (**consistente**) de [9] frente a cualquier tipo de heteroscedasticidad (con independencia de la forma funcional de $v(\cdot)$ en [5]):

$$\widehat{\mathrm{Var}}[\hat{\boldsymbol{\beta}}_{\mathbf{W}} \mid \mathbf{X}]_{\mathrm{WH}} = \tfrac{1}{N}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{S}}_{\mathrm{WH}}\hat{\mathbf{Q}}^{-1}, \text{ con}$$

$$\hat{\mathbf{Q}} = \tfrac{1}{N}\mathbf{X}'\mathbf{X} = \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i', \quad \hat{\mathbf{S}}_{\mathrm{WH}} = \tfrac{1}{N-K}\textstyle\sum_{i=1}^{N}\hat{u}_i^2\mathbf{x}_i\mathbf{x}_i',$$

donde $\mathbf{x}_i'$ es la fila i-ésima de la matriz $\mathbf{X}$.

EJEMPLO 1 [cont.]: Errores estándar y p-valores calculados a partir del estimador de White:

$$\begin{array}{ccccc} \hat{Y} & = & 36.6908 & + & 0.1283X \\ & & (21.8880) & & (0.0392) \\ & & [0.1019] & & [0.0023] \end{array} \qquad [12]$$

EJEMPLO 2 [cont.]: Errores estándar y p-valores calculados a partir del estimador de White:

$$\begin{array}{ccccccccc} \hat{Y} & = & 7.6273 & + & 1.2785X_2 & + & 0.0214X_3 & + & 18.3902X_4 \\ & & (39.6450) & & (0.1861) & & (0.0134) & & (8.1771) \\ & & [0.8479] & & [0.0000] & & [0.1139] & & [0.0274] \end{array} \qquad [13]$$

Observación 2: En un contraste de heteroscedasticidad, interpretar el rechazo de H_0 como evidencia de heteroscedasticidad es correcto siempre que las restantes hipótesis clásicas del modelo considerado (especialmente las hipótesis de **regresores exógenos** y **ausencia de autocorrelación**) sean válidas. En particular, si $\mathrm{E}[U_i \,|\, \mathbf{X}] \neq 0$ debido a algún **error de especificación**, entonces se puede rechazar H_0 aun cuando $\mathrm{Var}[U_i \,|\, \mathbf{X}]$ sea constante. Por ejemplo, si se omite algún término cuadrático en el lado derecho del modelo, o si se omite el **logaritmo** en el lado izquierdo cuando debería haberse utilizado, entonces el estadístico de White y/o el de Breusch-Pagan puede/n resultar significativo/s. Por lo tanto, es recomendable contrastar primero la especificación del modelo (porque la pérdida de insesgadez suele considerarse más grave que la pérdida de eficiencia), y después, una vez resuelto cualquier error de especificación (que provoca pérdida de insesgadez), considerar la posible heteroscedasticidad (que "solamente" provoca ineficiencia).

EJEMPLO 2 [cont.]:

$$\begin{array}{ccccccccc} \widehat{\ln Y} & = & 4.9590 & + & 0.0033X_2 & + & 5.08\times10^{-5}X_3 & + & 0.0412X_4, \\ & & (39.6450) & & (0.1861) & & (1.83\times10^{-5}) & & (0.0245) \\ & & [0.0000] & & [0.0000] & & [0.0068] & & [0.0964] \\ \end{array} \quad [14]$$

$$N = 80, \;\; WH = 3.7205 \;\; [0.9288].$$

Observación 3: La **transformación logarítmica** de la **variable dependiente** en un modelo de regresión suele reducir notablemente (o eliminar por completo) la heteroscedasticidad. No obstante, no siempre es posible o tiene sentido práctico transformar logarítmicamente la variable dependiente de un modelo.

PARTE 4 - MÍNIMOS CUADRADOS PONDERADOS

➪ Modelo original **con** heteroscedasticidad (como [4]-[5]):

$$Y_i = \beta_1 + \beta_2 X_{i2} + \ldots + \beta_K X_{iK} + U_i, \text{ con } \mathrm{E}[U_i \mid \mathbf{X}] = 0, \quad [15]$$

$$\mathrm{Var}[U_i \mid \mathbf{X}] = \sigma^2 v_i \;\; (i = 1, \ldots, N), \quad [16]$$

➪ Modelo transformado **sin** heteroscedasticidad:

$$\frac{Y_i}{\sqrt{v_i}} = \beta_1 \frac{1}{\sqrt{v_i}} + \beta_2 \frac{X_{i2}}{\sqrt{v_i}} + \ldots + \beta_K \frac{X_{iK}}{\sqrt{v_i}} + \frac{U_i}{\sqrt{v_i}}, \text{ con } \mathrm{E}\left[\frac{U_i}{\sqrt{v_i}} \mid \mathbf{X}\right] = 0, \quad [17]$$

$$\mathrm{Var}\left[\frac{U_i}{\sqrt{v_i}} \mid \mathbf{X}\right] = \frac{1}{v_i} \times \mathrm{Var}[U_i \mid \mathbf{X}] = \sigma^2 \;\; (i = 1, \ldots, N). \quad [18]$$

➪ En el modelo transformado los parámetros son los mismos y se interpretan exactamente de la misma manera que en el modelo original.

➪ El estimador MCO de $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_K]'$ en el modelo transformado [17]-[18] se denomina el **estimador** de **Mínimos Cuadrados Ponderados** (**MCP**) de $\boldsymbol{\beta}$.

El estimador MCP debe su nombre a que es el estimador asociado con la estimación de $\boldsymbol{\beta}$ que resulta de minimizar con respecto a $\mathbf{b} = [b_1, b_2, \ldots, b_K]'$ la función

$$P(\mathbf{b}) = \textstyle\sum_{i=1}^{N} [\, p_i (y_i - b_1 - b_2 x_{i2} - \ldots - b_K x_{iK})\,]^2 = \sum_{i=1}^{N} [\, p_i (y_i - \mathbf{x}_i' \mathbf{b})\,]^2,$$

donde $p_i = 1/\sqrt{v_i}$ $(i = 1, \ldots, N)$. En $P(\mathbf{b})$, cada residuo ordinario $(y_i - \mathbf{x}_i'\mathbf{b})$ recibe una **ponderación** que es inversamente proporcional a la varianza (dispersión) de la distribución de probabilidad condicional de Y_i. De esta manera se aprovecha la heteroscedasticidad para obtener un estimador de $\boldsymbol{\beta}$ más preciso (más informativo) que el estimador MCO (que pondera a todas las observaciones por igual).

El modelo transformado [17]-[18] se puede escribir como (ver [6]-[8])

$$(\mathbf{PY}) = (\mathbf{PX})\boldsymbol{\beta} + (\mathbf{PU}), \text{ con} \tag{19}$$

$$\mathbf{P} = \text{diag}[p_1, ..., p_N] = \text{diag}\left[v_1^{-1/2}, \cdots, v_N^{-1/2}\right] \Rightarrow \mathbf{P}'\mathbf{P} = \mathbf{V}^{-1} \Rightarrow \mathbf{V} = (\mathbf{P})^{-1}(\mathbf{P}')^{-1},$$

$$\text{E}[\mathbf{PU} \mid \mathbf{X}] = \mathbf{0}, \ \text{Var}[\mathbf{PU} \mid \mathbf{X}] = \text{E}[\mathbf{PUU}'\mathbf{P}' \mid \mathbf{X}] = \sigma^2\mathbf{PVP}' = \sigma^2\mathbf{I}. \tag{20}$$

Si el modelo original [6]-[8] satisface todas las hipótesis clásicas, excepto por la presencia de heteroscedasticidad, entonces el modelo transformado satisface todas las hipótesis cásicas incluida la de homoscedasticidad. Por lo tanto, la manera óptima de hacer inferencia sobre $\boldsymbol{\beta}$ consiste en aplicar toda la teoría MCO al modelo transformado. En particular, el estimador MCP de $\boldsymbol{\beta}$ es

$$\begin{gathered}\ddot{\boldsymbol{\beta}}_{\mathbf{W}} = (\mathbf{X}'\mathbf{P}'\mathbf{PX})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{PY} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \text{ con} \\ \text{E}[\ddot{\boldsymbol{\beta}}_{\mathbf{W}}] = \boldsymbol{\beta}, \ \text{Var}[\ddot{\boldsymbol{\beta}}_{\mathbf{W}}] = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.\end{gathered} \tag{21}$$

- El problema práctico de todo lo anterior reside en que, salvo en casos excepcionales, no es posible conocer exactamente la forma de $v_i = v(X_{i2}, ..., X_{iK})$ en todo $i = 1, \ldots, N$, por lo que aplicar MCP en la práctica requiere utilizar estimaciones de $v_1, ..., v_N$.
- Un estimador MCP de $\boldsymbol{\beta}$ definido a través de algún estimador de $\mathbf{v} = [v_1, ..., v_N]'$ en vez de a través del verdadero valor de $\mathbf{v}$ (generalmente desconocido), se denomina un estimador de MCP **factibles** (MCPF).
- Un estimador MCPF **no** es, en general, insesgado (ni, por lo tanto, eficiente) porque utiliza un estimador de $\mathbf{v}$ en lugar de su verdadero valor. No obstante, si el estimador utilizado de $\mathbf{v}$ es un estimador adecuado (**consistente**), entonces el estimador MCPF es **consistente** y **asintóticamente relativamente más eficiente** que el estimador MCO. Para más detalles, ver Hayashi (2000): Sección 1.6.

EJEMPLO 1 [cont.]: Estimación MCPF con $\hat{v}_i = x_i$ $(i = 1, ..., N)$ (ver [2] y [5]):

$$\begin{array}{rcccc} \hat{Y} & = & 28.7320 & + & 0.1410X, \\ & & (16.1875) & & (0.0270) \\ & & [0.0839] & & [0.0000] \\ \end{array} \tag{22}$$
$$N = 40, \ \ WH = 6.9907 \ [0.0303].$$

Observación 4: En presencia sólo de heteroscedasticidad, MCO y MCPF son ambos estimadores consistentes. Por su parte, la eficiencia asintótica relativa de MCPF se basa en que [i] el tamaño muestral sea suficientemente grande, y [ii] la estimación de la heteroscedasticidad sea "adecuada". Si [ii] falla, MCPF aún es consistente, pero su varianza asintótica ya no es necesariamente menor que la de MCO. Además, con muestras cortas, incluso si [ii] es aceptable, la aproximación asintótica puede funcionar peor para MCPF que para MCO, dado que MCPF requiere estimar más parámetros que MCO. En todo caso, MCPF puede ser una alternativa "razonable" a MCO (con muestras "razonablemente" grandes) cuando la heteroscedasticidad puede estimarse "razonablemente" bien. Para más detalles ver Wooldridge (2020): Sección 8-4.

PARTE 5 - RECOMENDACIONES PRÁCTICAS

A la hora de elaborar un modelo RLM, prestar especial atención al contenido del conjunto de **variables explicativas** y a la **forma funcional** de la relación entre la variable dependiente y las variables explicativas originales de interés.

Fundamentar ambos aspectos en (quizás) algún **modelo teórico** y (sobre todo) en el **sentido común** y en un análisis detallado de las **características muestrales** de los **datos** disponibles.

En general, comenzar con un **modelo sencillo** pero **sensato** para, en su caso, enriquecerlo paso a paso (*bottom-up/specific-to-general approach*). Aunque puede presentar ciertos inconvenientes, este enfoque suele ser preferible al opuesto (*top-down/general-to-specific approach*), que parte de una premisa difícilmente justificable (un modelo extenso que contiene como caso particular el "modelo correcto").

Recordar que la **omisión** de variables explicativas **relevantes** puede llevar a la estimación de relaciones espurias (MCO **sesgado**), mientras que la **inclusión** de variables **irrelevantes** "sólo" puede implicar pérdida de precisión (MCO **ineficiente**).

Analizar la posible presencia de **multicolinealidad** y de **observaciones influyentes** en los datos, cuyas consecuencias pueden distorsionar notablemente las conclusiones derivadas de cualquier modelo estimado.

En todo caso, recordar que la señal más evidente de una **mala especificación** o de alguna **deficiencia** en los **datos** suele ser la obtención de **estimaciones** con **valores** y/o **signos chocantes** (distintos de los razonablemente esperados).

Como ayuda para especificar un modelo, pueden ser de utilidad algunos procedimientos formales (contrastes de significación, coeficientes de determinación, criterios de información y de evaluación de previsiones), pero nunca como criterios únicos y fundamentales.

Recordar la diferencia entre significación **estadística** y significación **práctica**.

Evaluar la presencia de **heteroscedasticidad** sólo a partir de un modelo "razonablemente bien especificado", estimado con unos datos que no presenten deficiencias notables.

Para más detalles, consultar los dos artículos siguientes:

Kennedy, P.E. (2002), “Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics,” *Journal of Economic Surveys*, 16(4): 569-589.

Kennedy, P.E. (2005), “Oh No! I Got the Wrong Sign! What Should I Do?,” *Journal of Economic Education*, 36(1): 77-92.