

ECONOMETRÍA

APLICADA UTILIZANDO R.

PAPIME PE302513 LIBRO ELECTRÓNICO Y COMPLEMENTOS DIDÁCTICOS EN MEDIOS COMPUTACIONALES, PARA EL FORTALECIMIENTO DE LA ENSEÑANZA DE LA ECONOMETRÍA

Capítulo 3.

El modelo de regresión múltiple.

Jorge Feregrino Feregrino



Identificación del modelo

- ❑ La identificación del objeto de investigación permitirá realizar una búsqueda exhaustiva de los datos para llevar a cabo una aproximación del comportamiento del fenómeno mediante los hechos estilizados.
- ❑ La búsqueda de la información de las variables, la relación teórica y la descripción estadística de estas será útil para determinar la metodología de análisis.
- ❑ La elección de la variable dependiente y las independientes conformarán una relación funcional múltiple para describir el fenómeno económico mediante la metodología econométrica propuesta.

La presentación del modelo

□ En el modelo de regresión múltiple las variables exógenas (X_j), asociadas a coeficientes lineales constantes (β_j), indican el efecto condicionado de cada variable independiente sobre la variable dependiente (Y), la especificación general del modelo con cinco variables independientes es la siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

Los errores

- ❑ La relación entre las variables es inexacta, por lo tanto, la evaluación se realiza en términos probabilísticos.
- ❑ La forma funcional reducida de la estimación de la regresión múltiple, al expresarse en términos probabilísticos debe incorporar un término de error (ε_i).

$$\hat{y}_i = b_0 + \sum_{j=1}^5 b_j x_{ji} + \varepsilon_i$$

Objetivos de la regresión

- 1) Estimar los valores de una variable independiente (\hat{y}) mediante una función lineal de un número (K) variables independientes observadas x_j , donde $j = 1, \dots, K$
- 2) Obtener los efectos estadísticos de cada variable independiente, mediante la estimación de los coeficientes b_j , sobre la variable dependiente (\hat{y}).
- 3) Estimar la exogeneidad débil, para mostrar que la distribución marginal de la variable independiente, al no contener información relevante para estimar los parámetros de interés, se puede eliminar.



Supuestos para la estimación de la regresión múltiple

- 1) Las variables independientes x_{ji} son números fijos o bien variables aleatorias X_j , independientes del término de error ε_i .
- 2) El valor esperado de la variable aleatoria (\hat{y}) es una función de las variables independientes X_j
- 3) Los términos de error ε_i son variables cuya media esperada es igual a cero y la varianza es constante σ^2 para todas las observaciones:

$$E[\varepsilon_i] = 0 \quad y \quad E[\varepsilon_i^2] = \sigma^2 \text{ para } (i = 1, \dots, n)$$

Supuesto para la estimación

4) Los términos de error aleatorios ε_i , no están correlación entre sí

$$E[\varepsilon_i \varepsilon_j] = 0 \quad \text{para todo } i \neq j$$

5) Es imposible hallar un conjunto de números que no sean iguales a cero tal que ,

$$c_0 + c_1 x_{1i} + c_2 x_{2i} + \dots + c_k x_{ki} = 0$$

Estimación mediante los MCO

- ❑ En modelo de Mínimos Cuadrados Ordinarios (MCO), los coeficientes se obtienen mediante la minimización de los errores o la suma de residuos explicados al cuadrado (SCE)
- ❑ Para minimizar la SCE se procede de la siguiente forma, matemática la SCE tiene la siguiente representación:

$$SCE = \sum_{i=1}^n e_i^2 = SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

El resultado obtenido al minimizar los errores en el MCO

- El desarrollo extenso del MCO es resultado de la aplicación de cálculo y se obtienen 3 ecuaciones lineales y 3 incógnitas, (b_0, b_1, b_2) :

$$= nb_0 + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i$$

$$b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i} x_{2i} = \sum_{i=1}^n x_{1i} y_i$$

$$b_0 \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{1i} x_{2i} + b_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n x_{2i} y_i$$

Inferencia estadística a partir de los resultados del MCO

- ❑ De la representación de la regresión múltiple se infiere, que en el caso de la primera variable independiente (x_{1i}, b_1), esta explicada por la misma variable al cuadrado, y en el caso del otro coeficiente (b_2) esta explicado por la asociación entre las variables independientes.
- ❑ En la regresión los dos coeficientes asociados a cada variable independiente explican el comportamiento de la variable dependiente de forma significativa.

Notación matricial

$$Y = X\beta + U$$

- La estimación objetivo del modelo, busca obtener los coeficientes estimados del modelo en relación a las variables independientes, para explicar la variable dependiente (\hat{Y}) y su notación es la siguiente:

$$\hat{Y} = X\hat{\beta}$$

Los resultados del MCO en la notación matricial

- Al respecto al vector de los coeficientes (β) tenemos la siguiente notación matricial reducida:

$$\frac{\partial s}{\partial \beta} = -X^T Y - X^T Y + 2 (X^T X \beta)$$
$$\frac{\partial s}{\partial \beta} = -2X^T Y + 2 (X^T X \beta) = \vec{0}$$

- Para obtener los coeficientes estimados despejamos β

$$\hat{\beta} = (X^T X)^{-1} - X^T Y$$

Las propiedades de errores para realizar inferencia estadística

1) La sumatoria de los errores en una serie son igual a cero:

$$\sum_{i=1}^n e_i x_{ij} = 0, j = 1 \dots k$$

2) La covarianza entre los errores y las variables explicativas a medida que aumenta el número de observaciones es igual cero:

$$Cova = (e_i, x_{ij}) = 0$$

3) La varianza residual de los errores, en ausencia de sesgo, debe estar entorno a la misma varianza:

$$s_r^2 = \frac{1}{n - (k + 1)} \sum_{i=1}^n e_i^2$$

La eficiencia del modelo y su poder explicativo

- ❑ La estabilidad de los coeficientes se analiza mediante su distribución estadística, al igual que en la regresión simple se distribuyen como una normal, es decir, la media es igual a cero y la desviación estándar es igual a uno.

$$\hat{\beta} \sim N(0,1)$$

- ❑ Esto comportamiento asegura que los coeficientes estimados sigan una trayectoria normal y no sigan un comportamiento errático

El poder explicativo de los coeficientes

- ❑ El análisis de probabilidad sobre los coeficientes, permite identificar la influencia de cada variable a partir de la hipótesis planteada desde el diseño del modelo y su forma funcional. El contraste de hipótesis, se construye mediante una t de Student con k grados de libertad:

- ❑ La hipótesis nula:

$$H_0: \beta_i = 0$$

- ❑ La hipótesis alternativa:

$$H_a: \beta_i \neq 0$$

La lectura de la significancia individual

- ❑ La distribución del valor de los coeficientes para $n > 30$ observaciones sigue una distribución t_{n-k-1} , con una probabilidad del 95% se encuentra en el intervalo $[-2,2]$.
- ❑ Si $t > 2$, se rechaza la hipótesis nula y se puede inferir estadísticamente que las variables independientes influyen en la variable dependiente, es decir, se acepta la hipótesis alternativa.

$$H_a: \beta_i \neq 0$$

El intervalo de confianza

- El criterio del intervalo de confianza está diseñado de la siguiente forma:

$$P \left(\hat{\beta}_i - t_{\frac{\alpha}{2}} SE(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + t_{\frac{\alpha}{2}} SE(\hat{\beta}_i) \right) = 1 - \alpha$$

- El criterio muestra la probabilidad de que el verdadero β_i se encuentra en el intervalo entre el coeficiente estimado ($\hat{\beta}_i$) y 2 desviaciones estándar (SE) a la derecha y a la izquierda.

El poder explicativo del modelo

- EL contraste F, muestra si las variables explicativas en conjunto explican las variaciones de la variable independiente. Se ha demostrado que los coeficientes $\beta_1 = \beta_2 = \dots = \beta_k = 0$ y además, siguen una distribución F dado la siguiente forma:

$$\frac{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{k}}{\frac{\sum_{i=1}^n e_i^2}{n - k - 1}} \sim F_{k, n-k-1}$$

Criterios de la prueba F

- Cuando se acepta la hipótesis nula se debe a dos factores: 1) las variables no influyen en la variable independiente, 2) existe dependencia no lineal entre la variable explicada y algún regresor.
- Cuando se rechaza la hipótesis nula en el contraste del test F, muestra que la variable dependiente esta explicada por alguna de las variables independientes.



Casos en el contraste de la prueba F

- 1) Cuando el contraste F es significativo y todos los coeficientes individuales de acuerdo al contraste de la t de student también son significativos, todas las variables independientes son significativas.
- 2) Si el contraste F es significativo y sólo algunos de los coeficientes individuales son significativos de acuerdo al contraste de la t de student, las variables no significativas deben ser eliminadas del modelo. Otra solución, es realizar una transformación y estimar nuevamente para verificar si la relación entre las variables no es lineal.
- 3) Cuando el contraste de F es significativo y ninguno de los coeficientes asociados a las variables es significativo, entonces podría estar presente un problema de multicolinealidad.

Casos en el contraste de la prueba F

- El Test F muestra la proporción en que la varianza de los errores determina el poder explicativo del modelo. La notación matricial de la prueba, muestra que la diagonal de la matriz conocida, arroja los valores de la varianza (σ^2):

$$D(X^T X)^{-1} \rightarrow \begin{bmatrix} d_{00} & & & \\ & d_{11} & & \\ & & d_{ii} & \\ & & & d_{kk} \end{bmatrix}$$

- De esta forma, la distribución de los coeficientes estimados es la siguiente:

$$\hat{\beta}_i \sim N(\beta_i, \sigma \sqrt{d_{ii}})$$

- La desviación entre el coeficiente estimado ($\hat{\beta}$) y el coeficiente (β_i) en proporción a la interacción en diagonal conocida se comportan como una normal.