REPORTE DE INVESTIGACIÓN

1. Nombre del profesor

Dra. Lucía A. Ruiz Galindo

2. Proyectos registrados ante Consejo Divisional

607 Análisis Multivariado y de Series de Tiempo y

891 Modelos con fundamentos microeconómicos

3. Líneas de generación y/o aplicación de conocimiento

Econometría, Series de Tiempo y Modelos micro y macroeconómicos

4. Área o Grupo de Investigación

Grupo de Investigación de Modelación Económica Teórica y Aplicada
(en proceso de aprobación)

ECONOMETRÍA APLICADA USANDO R

ERROR DE ESPECIFICACIÓN

Por

LUCÍA A. RUIZ GALINDO

UNIVERSIDAD AUTÓNOMA METROPOLITANA
DIVISIÓN DE CIENCIAS SOCIALES Y HUMANIDADES
DEPARTAMENTO DE ECONOMÍA
GRUPO DE INVESTIGACIÓN MODELACIÓN ECONÓMICA
TEÓRICA Y APLICADA

Diciembre, 2015.

Error de Especificación
Lucía A. Ruiz Galindo[1]

1. Introducción

En la elaboración de un modelo econométrico es muy importante la evaluación económica y la econométrica del modelo estimado. En ambas se revisa si la información empírica incorporada al modelo, es decir, con la que éste se estimó, da evidencia a favor o en contra por un lado, de la teoría económica que lo sustentó y por el otro, de los supuestos tanto los que se hacen en su parte determinista como los que se plantean sobre el término estocástico.

A grandes rasgos, en la evaluación económica se revisa que los signos y las magnitudes de los parámetros estimados sean los propuestos por la teoría económica, mientras que la evaluación econométrica consiste de una variedad de pruebas estadísticas que permiten averiguar si se satisfacen todos los supuestos del modelo. Cuando la evaluación económica es exitosa, pero existen discrepancias entre el resultado de las pruebas y los supuestos del modelo, habrá indicios de que el modelo no está especificado correctamente, hay errores en su especificación.

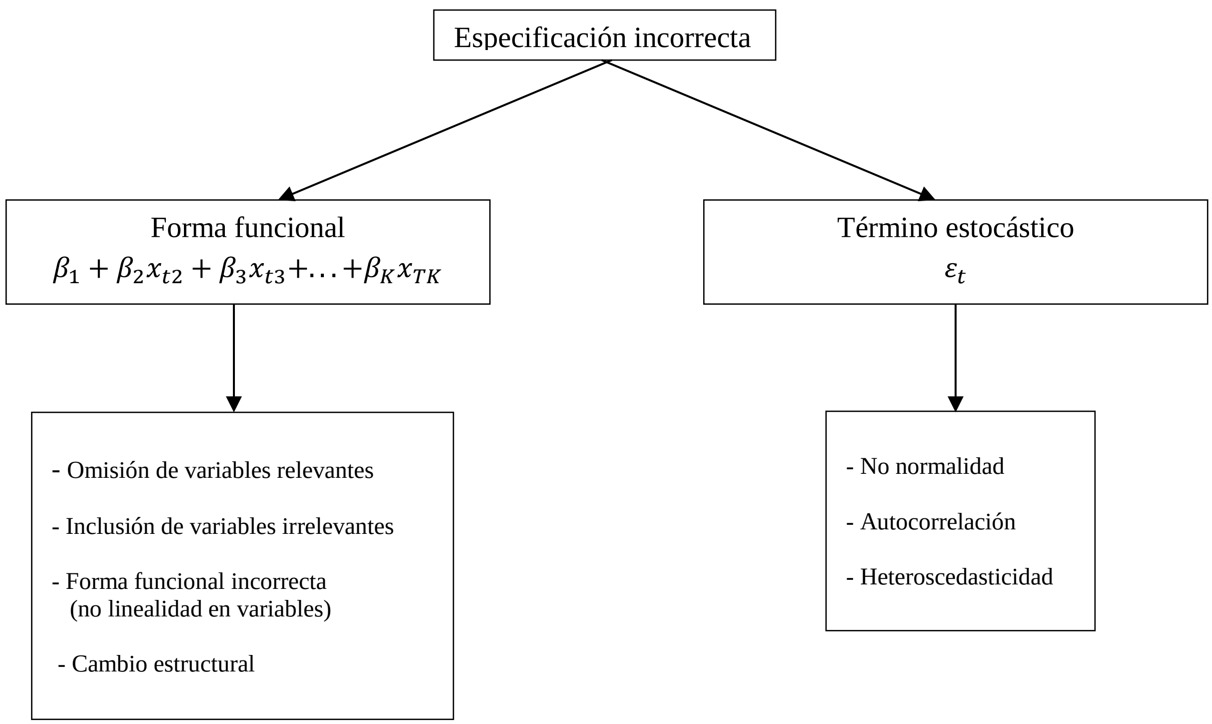
La especificación incorrecta del modelo puede deberse a una formulación no adecuada de la forma funcional o bien, a que se violan los supuestos del error aleatorio o incluso a la información empírica que se incorpora al modelo para su estimación. La Figura 1 muestra cómo están constituidas las dos primeras fuentes de especificación incorrecta, para cada una de sus componentes existen pruebas estadísticas que permiten decidir cuáles de los supuestos del modelo de regresión no se satisfacen dada la información empírica que se utiliza en su estimación.

En este Capítulo se estudian la especificación incorrecta del modelo ocasionada por un planteamiento no apropiado de la forma funcional (parte determinista), esta situación generalmente se debe a que se han incluido variables irrelevantes (sobreparametrización),

[1] Departamento de Economía, Universidad Autónoma Metropolitana - Azcapotzalco.
laruizg@correo.azc.uam.mx

omitido variables relevantes (subparametrización) o bien, a que la forma funcional no es la correcta en lo que respecta a la manera en que se incorporan las variables independientes y en la literatura a esto se le conoce como errores de especificación. Cabe mencionar que los errores de especificación en la forma funcional, también se originan cuando existe cambio estructural en los parámetros, pero este tema no es objeto de estudio de este Capítulo.

Figura 1. Errores de especificación



Por su parte, la existencia de multicolinealidad y la de correlación entre las variables independientes y el término estocástico, también son fuente de especificaciones erróneas del modelo de regresión, pero ellas son debidas a la selección de la información empírica de las variables del modelo.

La especificación del modelo incluyendo sus supuestos, conducen a que los estimadores de los parámetros satisfagan propiedades estadísticas deseables, como son el insesgamiento, la eficiencia y la consistencia. Aquí se estudiará las consecuencias que tiene sobre las propiedades de los estimadores, la sub y sobreparametrización, además de formular y llevar a cabo en R, la prueba RESET, útil para saber entre otras cosas, si la especificación lineal en las variables es correcta o no.

Este Capítulo en su segunda Sección presenta una exposición sucinta del modelo de regresión lineal, sus supuestos y una breve explicación de la forma en que se incurre en una especificación incorrecta del mismo, en la tercera Sección se estudian las implicaciones que tiene sobre las propiedades de los estimadores de los parámetros, la sobreparametrización y subparametrización del modelo, la cuarta Sección se formula la prueba RESET para analizar si la forma funcional del modelo es correcta o no, en la siguiente Sección se explica la manera en que esa prueba se lleva a cabo en R y se muestran algunos ejemplos de su implementación, y en la sexta y última, se presentan algunas conclusiones.

2. Especificación y supuestos del modelo general de regresión lineal

En el desarrollo de este Capítulo se considera un modelo de regresión lineal en el que la variable dependiente y_t es explicada por *K*-1 variables independientes, esto es,

$$y_t = \beta_1 + \beta_2 x_{t2} + . . . + \beta_K x_{tK} + \varepsilon_t \tag{1}$$

donde $\beta_1, \ . \ . \ . \ , \ \beta_K$ son los parámetros del modelo, las x_{tk}'s son las variables independientes, $k = 2, \ . \ . \ . \ , K$, ε_t es el término o error estocástico y *t*, $t = 1, \ . \ . \ . \ , T$ es un índice que indica el número de la observación y *T* es el total de observaciones.[2]

El modelo en (1) se puede formular de manera matricial como sigue

$$y = X\beta + \varepsilon, \tag{2}$$

donde $y = (y_1, y_2, \ldots, y_T)'$,

$$X = \begin{pmatrix} 1 & x_{12} & x_{13} & \ldots & x_{1K} \\ 1 & x_{22} & x_{23} & \ldots & x_{2K} \\ 1 & x_{32} & x_{33} & \ldots & x_{3K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T2} & x_{T3} & \ldots & x_{TK} \end{pmatrix},$$

$\varepsilon = (\varepsilon_1, \varepsilon_2, . \ . \ \ ., \varepsilon_T)'$ y $\beta = (\beta_1, \beta_2, \ldots, \beta_K)'$. Observe que el vector *y* está constituido por las T observaciones de la variables dependiente, la matriz X de dimensión $T \times K$, por las

[2] En todo lo que sigue, sin pérdida de generalidad, se pensara que las variables están en series de tiempo y por tanto, *t* indica un periodo y hay observaciones para *T*. Es importante señalar que todo lo que se desarrolla en este Capítulo es válido también para cortes transversales, en cuyo caso, *t* representará un individuo.

variables independientes, el vector β de dimensión K, por los parámetros del modelo, y ε por los T términos estocásticos, uno por cada periodo.

El modelo está completamente especificado cuando se plantean sus supuestos. La forma funcional $\beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK}$, debe ser lineal en los parámetros, las variables x_{tk}'s, $k = 2, \ldots, K$, son las únicas que explican a y_t, ellas son linealmente independientes y por ello, la matriz X es de rango completo, y además, los parámetros no cambian en el periodo de estudio, esto es, hay permanencia estructural.[3]

Por su parte, el término estocástico ε_t, $t = 1, \ldots, T$, tiene media cero, es homoscedástico y no autocorrelacionado y se distribuye de manera normal, todos los supuestos de los errores aleatorios se pueden resumir diciendo que ellos son elegidos de manera no correlacionada de una distribución normal con media y varianza constante o equivalentemente,

$$\varepsilon \sim N(\mathbf{0}, \sigma^2 I),$$

donde $\mathbf{0}$ e I, son un vector de ceros de dimensión T y la matriz identidad de $T \times T$, respectivamente, y σ^2 es la varianza del término aleatorio, es decir, $V(\varepsilon_t) = \sigma^2$, $t = 1, \ldots, T$.

Una vez formulado el modelo de regresión, se procede a estimarlo usando datos de las variables dependiente e independientes, de forma que los estimadores de los parámetros dependen tanto de la especificación del modelo como de la información empírica que se incorpora a él. De manera que, errores en el modelo o incluso en los datos, conduccen errores de especificación.

La especificación correcta del modelo conduce a que los estimadores de las β_k's, $k = 1, \ldots, K$, las $\hat{\beta}_K$, son los mejores estimadores lineales e insesgados, MELI o BLUE por sus siglas en inglés (*Best Linear Unbaised Estimator*), es decir, dentro de los lineales e insesgados son los de mínima varianza, además de que son consistentes. Por su parte, el estimador mínimo cuadrático de σ^2 es insesgado, pero su varianza es mayor que la

[3] Los momentos poblacionales están condicionados a la información disponible de las variables en el modelo.

correspondiente al estimador máximo verosímil y éste a pesar de ser más eficiente, es sesgado.

A continuación, en la siguiente Sección, se estudia si estas propiedades prevalecen cuando en el modelo se excluyen variables importantes o cuando se incluyen variables irrelevantes.

3. Sobreparametrización y subparametrización, consecuencias sobre las propiedades de los estimadores

Considérense los siguientes modelos

$$\text{M1: } y = X_1\beta^1 + X_2\beta^2 + \varepsilon,\ \ \varepsilon \sim N(\mathbf{0}, \sigma^2 I)$$

y

$$\text{M2: } y = X\beta^1 + u,\ \ \ u \sim N(\mathbf{0}, \sigma_1^2 I),$$

donde $X = (X_1 \ \vdots\ X_2)$, $\beta = \begin{pmatrix} \beta^1 \\ \cdots \\ \beta^2 \end{pmatrix}$, $u = X_2\beta^2 + \varepsilon$, X_1 tiene las primeras K_1 variables de la matriz X , X_2 tiene las $K_2 = K - K_1$ restantes ($K = K_1 + K_2$), y el vector de parámetros β se plantea de acuerdo a esa división de las variables independientes en X.

Dados esos modelos, si el correcto o verdadero es M2 y se estima M1, entonces se están incorporando variables irrelevantes para la determinación de y, esto es, se está sobreparametrizando. Si por el contrario, M1 es el correcto y se estima M2, se están omitiendo en el modelo variables importantes, que pasan a formar parte del término estocástico, en este caso se está subparametrizando. Cada una de esas situaciones tienen consecuencias sobre las propiedades de los estimadores mismas que se plantearan a continuación.[4]

Al sobreparametrizar un modelo se están incluyendo variables que no son importantes en la determinación de y, de manera que el modelo adecuado es M2, pero el

[4] Un tratamiento riguroso de estos temas se puede estudiar en Kmenta (1997), Jhonston y Dinardo (1997), Davidson y MacKinnon (2004), y Greene (2007), por citar algunos.

que se estima M1. Observe que en este caso M2 se puede obtener de M1 haciendo $\beta^2 = 0$ y $\sigma^2 = \sigma_1^2$ y por ello, algunos autores no lo consideran un error de especificación o una forma incorrecta de especificación, pues solo no incorpora las restricciones mencionadas sobre los parámetros (Davidson y MacKinnon (2004), y Greene (2007)).

En esta situación los estimadores tanto de los parámetros β^1 y β^2, como el de σ^2 son insesgados y consistentes, y esta propiedad se satisface incluso cuando se imponen la restricciones $\beta^2 = 0$ y $\sigma^2 = \sigma_1^2$. Sin embargo, debe señalarse que inclusión de variables irrelevantes aumenta la varianza de los estimadores de las betas, de manera que ya no serán eficientes de manera relativa.

Por su parte, al subparametrizar un modelo se están omitiendo variables que son importantes en la determinación de la variable dependiente y. Si se supone que se dejan fuera K_2 variables, esto es, las variables en X_2, entonces el modelo verdadero es M1, pero se estima M2. En este caso se debe observar en primer lugar, que las variables excluidas se encuentran dentro del término estocástico, por ello su varianza no será estimada de manera correcta y en consecuencia, los intervalos de confianza y las pruebas de hipótesis conducirán a conclusiones erróneas, pues dependen de ese estimador, que además es sesgado. En segundo lugar el estimador del vector β^1 en el modelo M2 denotado por $\tilde{\beta}^1$, es segado y eficiente, es decir, su varianza es menor a la correspondiente a $\hat{\beta}^1$, que es el estimador de β^1 en M1 y por tanto, $\tilde{\beta}^1$ es más preciso que $\hat{\beta}^1$, pero no es insesgado (Davidson y MacKinnon (2004), y Greene (2007)).

4. Prueba RESET

Conocer los errores de especificación y en caso de incurrir en ellos, saber sus consecuencias, es importante en la elaboración de un modelo econométrico, igual relevancia tiene el averiguar si ellos se han cometido o no. En esta Sección se estudia la prueba de especificación de Ramsey, denominada RESET, por sus siglas en inglés *Regression Equation Specification Error Test*, debida a Ramsey (1969), que sirve para detectar errores de especificación ocasionados por la omisión de variables independientes, por la posible existencia de correlación entre las variables en X y ε o bien, porque la forma funcional de las variables independientes no es la apropiada.

Así pues, la prueba RESET se usa para analizar si el modelo está bien especificado o no, de manera que las hipótesis a probar son

H_0: Forma funcional correcta vs H_1: Forma funcional incorrecta.

Esta prueba se realiza una vez que se ha estimado el modelo planteado en (1) y que se ha calculado su ajuste, dado por

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_{t2} + \ldots + \hat{\beta}_K x_{tK} \tag{3}$$

y consiste en agregar al modelo inicial, potencias de sus valores ajustados y analizar la significancia estadística conjunta de los parámetros asociados a las potencias de la variable ajustada. De esta manera, el modelo que se debe estimar para efectuar la prueba RESET es

$$y_t = \beta_1 + \sum_{k=2}^{K} \beta_k x_{tk} + \sum_{i=2}^{m+1} \alpha_i \hat{y}_t^i + v_t, \tag{4}$$ [5]

en el que se han incorporado m potencias de la variable ajustada $\hat{y}_t$.

Observe que bajo H_0, los parámetros $\alpha_i = 0$, para toda $i = 2, \ldots, m+1$, y bajo H_1, al menos uno de esos parámetros es diferente de cero, en cuyo caso la especificación del modelo no es correcta. Con estas consideraciones, la prueba se puede plantear como

$$H_0: \alpha_i = 0, \ \ \forall i = 2, \ldots, m+1$$

vs

$$H_1: \alpha_i \neq 0, \text{ para al menos una } i = 2, \ldots, m+1.$$

Bajo H_0 el estadístico de prueba se distribuye como una $F_{(m,T\text{-}K\text{-}m)}$.[6] El número m, que también representa el número de restricciones lineales bajo la hipótesis nula, se puede determinar usando los criterios de información de Akaike, Schwarz o Hannan-Quinn, utilizados comúnmente para seleccionar entre modelos alternativos, en los que la variable dependiente debe ser la misma.

5. Prueba RESET en R

[5] v_t denota el error estocástico de este modelo y por tanto, tiene los mismos supuestos del modelo de regresión en (1).

[6] En el Capítulo 5 se explican de manera sucinta, las pruebas de significancia conjunta.

La prueba RESET en R, requiere del paquete lmtest y se efectúa una vez que el modelo ha sido estimado. Mediante la instrucción

```
> library(lmtest)
```

se carga el paquete lmtest, que dicho sea de paso, contiene varias pruebas que son importantes en la evaluación econométrica de un modelo de regresión. En seguida y ya que se dispone de los datos, se estima el modelo y hasta entonces, se hace la prueba RESET introduciendo

```
> resettest(vdep)
```

en donde el argumento vdep es el nombre del objeto donde se guarda el resultado de la estimación. Es importante indicar que esta instrucción introduce por *default* la segunda y tercera potencia de la variable ajustada $\hat{y}_t$, de manera que el modelo en la que se basa la prueba RESET es

$$y_t = \beta_1 + \sum_{k=2}^{K} \beta_k x_{tk} + \alpha_2 \hat{y}_t^2 + \alpha_3 \hat{y}_t^3 + v_t.$$

Si se requieren potencias superiores a 3, se introduce la instrucción

> resettest(vdep,power=2:*m*)

y si sólo se desea introducir la segunda potencia, se escribe

```
> resettest(vdep,power=2:2)
```

El resultado de la prueba presenta el nombre del objeto en `data`, el estadístico de prueba en `RESET`, los grados de libertad del numerador (*m*), en `df1` y los del denominador (*T*-*K*-*m*), en `df2` y el mínimo nivel de significancia al que se rechaza la hipótesis nula, en `p-value`.

Ejemplo 1.

La información anual de 1953 a 2004 contenida en el archivo Gasolina.txt es usada para estimar un modelo para la demanda de gasolina en USA (Greene, 2003). Se plantea una regresión log-log, en la que se modela la demanda per-cápita en función del ingreso per-cápita, del índice de precios de la gasolina y el de los autos nuevos. Estimado el modelo se analiza si la forma funcional es correcta mediante dos pruebas RESET, la primera incorpora de la segunda a la cuarta potencia del ajuste y la segunda, solo la segunda potencia.[7]

```
> library(lmtest)

> Gasolina <- read.csv("Gasolina.txt")
>   View(Gasolina)
> attach(Gasolina)
> cons<-lm(log(G/Pobl)~log(Y)+log(Pg)+log(Pan))
> resettest(cons,power=2:4)

        RESET test

data:  cons
RESET = 34.05, df1 = 3, df2 = 38, p-value = 7.347e-11

> resettest(cons,power=2:2)

        RESET test

data:  cons
RESET = 90.541, df1 = 1, df2 = 40, p-value = 7.933e-12
```

Los resultados de ambas pruebas indican que la forma funcional no es correcta ya que en ambas el *p-value* es menor que cualquiera de los niveles de significancia, comúnmente utilizados, de manera que el modelo debe ser reespecificado.

Ejemplo 2

En este ejemplo se presenta una versión del modelo estático para la elasticidad de sustitución Armington para México.[8] La variable explicada en el modelo es la demanda relativa (DRel), que resulta del cociente entre las importaciones totales y la demanda doméstica (diferencia entre el valor bruto de la producción y las exportaciones, ambas a precios de mercado) y la variable explicativa es el precio relativo (PRel), que se obtiene de

[7] Las variables del archivo son
Año: 1953-2004,
G: Gasto total en gasolina,
Pobl: Población
Pg: Indice de precio de la gasolina,
Y: Ingreso disponible per-cápita,
Pan: Indice de precios de los autos nuevos,
Pau: Indice de precios de los autos usados,
Ptp: Indice de precios del transporte público,
Pd: Indice de precios agregado del consumo de bienes durables,
Pnd: Indice de precios agregado del consumo de bienes no durables,
Ps: Indice de precios agregado para el consumo de servicios.
Fuente: http://people.stern.nyu.edu/wgreene/Text/econometricanalysis.htm

[8] Un análisis detallado de este modelo desde sus microfundamentos hasta la especificación final de un modelo dinámico es presentado en Casares, Ruiz-Galindo y Sobarzo (por publicarse).

dividir el índice de precios de la demanda relativa entre el correspondiente a las importaciones.[9]

Una vez estimado el modelo se prueba si la forma funcional es la correcta mediante la prueba de Ramsey, RESET. Ella se realiza considerando primero la segunda y tercera potencia y en seguida se efectúa solo para la segunda potencia, tal y como se muestra a continuación.

```
> library(lmtest)
Loading required package: zoo

Attaching package: ‘zoo’

The following objects are masked from ‘package:base’:

    as.Date, as.Date.numeric

> Elast <- read.csv("C:/Users/Atzimba/Desktop/Elast.txt")
>   View(Elast)
> attach(Elast)

> model<-lm(log(DRel)~log(PRel))
> resettest(model)

        RESET test

data:  model
RESET = 0.32523, df1 = 2, df2 = 80, p-value = 0.7233

> resettest(model,power=2:2)

        RESET test

data:  model
RESET = 0.25329, df1 = 1, df2 = 81, p-value = 0.6161
```

En ambas pruebas no se rechaza la hipótesis nula, el *p-value*>α (α=1%, 5% o 10%), y por tanto hay evidencia a favor de que la forma funcional es correcta. Debe observarse que en la primera prueba RESET de este ejemplo, no se introduce el comando *power* que indica las potencias que se desean incorporar de la variable ajustada y el modelo en (4) se estima con la segunda y tercera potencia, puesto que como ya se mencionó, esas potencias son las que se introducen por *default*.

[9] La estimación incorpora información trimestral del INEGI para el periodo que comprende del primer trimestre de 1993 al primero del 2013, a precios constantes del 2008.

Bibliografía

Casares, E. R., L. A. Ruiz-Galindo y H. Sobarzo, (por publicarse). "Short and Long Run Armington Elasticities for the Mexican Economy" en A. Pinto y D. Zilberman (editors), Modeling, Dynamics, Optimization and Bioeconomics II, en la serie Springer Proceedings in Mathematics an Statistics.

Davidson R. y J. G. MacKinnon, (2004). Ed. Oxford University Press, New York.

Greene, W. H., (2007). Econometric Analysis. Ed. New York University, New York.

Johnston, J. y J. Dinardo, (1997). *Econometrics Methods* Ed. McGraw-Hill, Singapur.

Kmenta, J., (1997). *Elements of Econometrics*. Ed. University of Michigan Press

Ramsey, J. B., (1969). "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis, *Journal of the Royal Statistical Society, Series B.*, vol. 31, 2, pp 350-371.

Referencias electrónicas

Datos (Greene, 2007), http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm

INEGI (2013a), "Banco de Información Económica", http://dgcnesyp.inegi.gob.mx

Archivos de las bases de datos de este Capítulo

Gasolina.txt

Elast.txt

ECONOMETRÍA
APLICADA UTILIZANDO R.

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

CAPÍTULO 4
Error de especificación

Lucía A. Ruiz Galindo*

*** Departamento Economía, UAM-A**

FES ACATLÁN dgapa SAREE

Error de especificación

Modelo General de Regresión Lineal

- Especificación

$$y_t = \beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK} + \varepsilon_t,$$

donde $\beta_1, \ldots, \beta_K$ son los parámetros del modelo,

y_t es la variable dependiente,

x_{tk}'s, $k = 2, \ldots, K$, son las variables independientes,

ε_t es el término o error estocástico,

t, $t = 1, \ldots, T$, es un índice que indica el número de la observación y

T es el total de observaciones

Error de especificación

Modelo General de Regresión Lineal

- Supuestos en la forma funcional

S1. Linealidad en los parámetros.

S2. Las K-1 variables independientes son las únicas que explican a la dependiente.

S3. El número de observaciones T, es mucho mayor que el de parámetros K.

S4. Las variables explicativas son linealmente independientes de manera que ninguna es combinación lineal de otra o de otras y por tanto el rango de X es K.

S5. Los parámetros no cambian en la muestra, es decir, hay permanencia estructural

Error de especificación

Modelo General de Regresión Lineal

- Supuestos Gauss-Markov

SGM1. $E(\varepsilon_t) = 0,\ \forall t = 1,\ ...,T.$

SGM2. $\{\varepsilon_1, \varepsilon_2,\ ..., \varepsilon_T\}$ y $\{x_{1k}, x_{2k},\ ..., x_{Tk}\}$ son independientes $\forall k = 2,\ ...,K.$

SGM3. $V(\varepsilon_t) = \sigma^2,\ \forall t = 1,\ ...,T.$

SGM4. $Cov(\varepsilon_t, \varepsilon_s) = 0,\ \forall t, s = 1,\ ...,T,\ t \neq s.$

SGM5. ε_t se distribuye Normal, $\forall t = 1,\ ...,T.$

Error de especificación

Fuentes de errores en la especificación

- Forma funcional (parte determinista)
  - Omisión de variables relevantes (subparametrización)
  - Inclusión de variables irrelevantes (sobreparametrización)
  - Forma funcional incorrecta en las variables
  - Cambio en los parámetros

Errores especificación

Fuentes de errores en la especificación

- Error Estocástico
  - No normalidad
  - Autocorrelación
  - Heteroscedasticidad
- Información emprírica
  - Multicolinealidad
  - No independencia entre variables independientes y término estocástico

Error de especificación

Consecuencias de la sobre y subparametrización en las propiedades estimadores del modelo

- Sobreparametrización
  - Los estimadores de las betas son inseesgados y consistentes, pero no son eficientes
- Subparametrización
  - La varianza del error no se estima de manera correcta, ya que existen variables importantes en él.
  - Los estimadores de las betas son eficientes, pero sesgados.

Error de especificación

Uso de la Prueba RESET

- Sobreparametrización o subparametrización
- Las variables independientes no se especifican de manera lineal
- No independencia entre variables independientes y errores estocásticos

Error de especificación

Prueba RESET en R

- Cargar paquete lmtest

```
> library(lmtest)
```

- Disponer del objeto al que se le aplicara la prueba
- Aplicar la prueba

```
> resettest(vdep)

> resettest(vdep,pwr=2:m)
```

Error de especificación

Ejemplo de la Prueba RESET en R

- Modelo para la demanda de gasolina

```
> library(lmtest)
> Gasolina <- read.csv("Gasolina.txt") >   View(Gasolina)
> attach(Gasolina)
> cons<-lm(log(G/Pobl)~log(Y)+log(Pg)+log(Pan))
> resettest(cons,power=2:4)

        RESET test

data:  cons

RESET = 34.05, df1 = 3, df2 = 38, p-value = 7.347e-11

> resettest(cons,power=2:2)

        RESET test

 data:  cons
RESET = 90.541, df1 = 1, df2 = 40, p-value = 7.933e-12
```