Universidad Zaragoza

1542

# Trabajo Fin de Grado

## El contraste Reset. Un estudio de Monte Carlo.

## The Reset test. A Monte Carlo study.


Autor:

María Escalona Romeo

Director/es

María Isabel Ayuda

Inmaculada Villanúa

FACULTAD DE ECONOMÍA Y EMPRESA


2015/2016

**Autor del trabajo:** María Escalona Romeo.

**Dirección del trabajo:** María Isabel Ayuda. Inmaculada Villanúa.

**Título del trabajo:** El contraste Reset. Un estudio de Monte Carlo.

**Titulación a la que está vinculado:** Grado de Economía.

**Resumen:** Los modelos econométricos explican la influencia que unas variables pueden tener sobre otras, incluyendo además una perturbación aleatoria, que nos permite razonar en términos probabilísticos. Cuando especificamos un modelo con la intención de simplificar la realidad, es necesario dotarlo de una forma funcional, y es cuando nos preguntamos si dicha forma coincide con la que reflejarían los datos reales. Para responder a esta pregunta podemos servirnos, entre otras herramientas, del contraste Reset. A pesar de ser una herramienta muy utilizada, no sabemos a priori, si su funcionamiento es adecuado, es decir, si es capaz de detectar la correcta o incorrecta forma funcional de un modelo especificado y estimado. Para analizar el comportamiento del criterio Reset realizamos un estudio de Monte Carlo, prestando especial atención a cómo actúa Reset ante distintos modelos estimados, diferentes tamaños muestrales y distintas expresiones de la regresión auxiliar. Los resultados y conclusiones se desvelaran posteriormente.

**Palabras clave:** contraste Reset, forma funcional, modelo econométrico, tamaño muestral, regresión auxiliar, tamaño de un contraste y potencia.

**Abstract:** The econometric models explain the influence that some variables can have over others, including also a random shock, which allows us to rationalize in probabilistic terms. Whenever a model is specified with the intention to simplify the reality it is necessary to give it a functional form. This is when we ask ourselves if this mentioned form coincides with what the real data would show. In order to respond to this question we could use, among other tools, the Reset test. Although this tool is widely employed, we don't know a priori if its operation is adequate, or whether it is capable of detecting the correct or incorrect functional form of a specified or estimated model. In order to analyze the Reset test standard we carry out a Monte Carlo study, giving special attention on how the Reset test acts to different estimated models, those of different sample size and different linear regression expressions. The results and conclusions are revealed later.

**Key words:** Reset test, functional form, econometric model, simple size regression expression, contrast and power size.

# ÍNDICE

# 1. INTRODUCCIÓN.

A lo largo de este apartado abordaremos cuestiones tales como: la motivación que me ha llevado a elegir este tema de investigación y los objetivos que a priori he considerado que debía lograr. Por último, presentaremos brevemente las partes que componen el trabajo.

## 1.1. MOTIVACIONES.

Considero que la Econometría es una herramienta capaz de permitirnos conocer la relación e influencia que pueden tener diferentes aspectos, incluso de la vida cotidiana. Con esto me refiero a que tras la estimación de un modelo apropiado, podemos ser capaces de responder a múltiples preguntas tales como ¿cuánto influyen mis años de educación en el salario que esperaría obtener? ¿De qué depende el precio de la vivienda que deseo adquirir? ¿Qué factores influyen el mayor desarrollo de un país emergente? etc. Y es precisamente dicha utilidad de la econometría la que ha hecho de ella una de las ramas de la economía más atractivas a mi juicio.

Sin embargo, es imposible que durante los cuatro años que dura el Grado de Economía podamos ser capaces de conocer a fondo dicha rama. Por esta razón, he decido profundizar en ella con la realización de este trabajo. En particular voy a centrarme en una de las funciones de la econometría, la simulación. Tan solo había sido mencionada en las clases pero sin ahondar demasiado en ella. Como veremos más adelante, consiste en ensayar cambios en variables instrumento, que son las que influyen sobre las variables objeto, siendo el objetivo de la simulación generar la trayectoria que siguen dichas variables. Para llevar a cabo la simulación concreta he utilizado el Experimento de Monte Carlo. La labor del experimento, era en este caso, analizar si uno de los contrastes utilizados en econometría, el contraste Reset, funcionaba adecuadamente. Esto es, si era capaz de detectar cuando nos encontrábamos frente a una correcta o incorrecta forma funcional.

## 1.2. OBJETIVOS.

El objetivo principal de este trabajo es analizar el comportamiento del criterio Reset como instrumento para determinar si la forma funcional especificada en un modelo es la adecuada.

Decimos que un criterio se comporta adecuadamente, cuando la probabilidad de rechazar $H_0$ cuando es cierta, es baja, y próxima al tamaño nominal fijado por el investigador, y cuando la potencia es alta.

El instrumento utilizado para llevar a cabo este análisis de tamaño y potencia del contraste Reset son los Estudios de Monte Carlo. En este trabajo se realizarán diversos experimentos con objeto de investigar en tres aspectos fundamentales:

1- Analizar el tamaño y la potencia tanto en muestras pequeñas como en grandes, para ver si es un contraste válido para todo tamaño muestral.

2- Analizar el comportamiento del criterio Reset ante distintos tipos de formas funcionales de los modelos especificados.

3- El contraste Reset, tal y como explicamos en el trabajo, utiliza una regresión auxiliar que puede incluir distintos componentes, por lo que otro de nuestros objetivos va a ser conocer que regresión auxiliar es más apropiada y si existe o no una que podamos considerar óptima.

## 1.3. BREVE DESCRIPCIÓN DEL TRABAJO.

En primer lugar, hablaremos de la econometría ya que es el marco teórico en el que se desarrolla el experimento (apartado 2), posteriormente para contextualizar el estudio veremos qué son los modelos econométricos y las formas funcionales más utilizadas, los posibles errores que podemos cometer a la hora de especificar un modelo econométrico y el contraste Reset como instrumento para analizar la correcta o incorrecta forma funcional de un modelo especificado (apartado 3). Explicaremos después los estudios de Monte Carlo (apartado 4), para concluir mostrando los resultados y conclusiones que se derivan del estudio realizado con el objetivo de evaluar el funcionamiento del contraste Reset (apartado 5). El trabajo finaliza con una valoración general de los resultados (apartado 6).

# 2. LA ECONOMETRÍA.

Hablaremos a continuación de la econometría como ciencia social, así como de la forma en que ésta lleva a cabo sus funciones.

## 2.1. CONCEPTO Y MÉTODO DE LA ECONOMETRÍA.

Con el objeto de introducir este trabajo, es necesario en primer lugar, conocer en qué consiste la econometría. Dentro de la literatura económica encontramos una amplia variedad de definiciones, a continuación veremos algunas de ellas.

De forma genérica podemos definir la econometría como una disciplina científica que se basa en la utilización de métodos matemáticos y estadísticos para estimar, a partir de un conjunto de datos, la forma según la cual se pueden relacionar diferentes variables, basándonos en teorías económicas.

Una de las definiciones más completa de la econometría y su objeto de estudio es la ofrecida por Judge, Hill et al (1982): *" La Econometría usando la Teoría económica, las Matemáticas y la Inferencia estadística como fundamentos analíticos y los datos económicos como la base informativa proporciona una base para : 1) modificar, refinar o posiblemente refutar las conclusiones contenidas en el cuerpo de conocimientos conocido como Teoría económica, y 2) conseguir signos, magnitudes y afirmaciones de calidad para los coeficientes de las variables en las relaciones económicas, de modo que esta información puede usarse como base para la elección y toma de decisiones".*

Para llevar a cabo los procedimientos anteriormente nombrados, utilizamos modelos econométricos. Si entendemos por modelo económico una representación simplificada y en símbolos matemáticos de cierto conjunto de relaciones económicas, podemos definir los modelos econométricos como modelos económicos en los que la forma y la especificación nos permite realizar aplicaciones empíricas. (Trívez, 2004).

Distinguimos cuatro fases a lo largo de las cuales la econometría llevará a cabo su labor:

1- Especificación del modelo econométrico: consiste en formular un modelo tal como creemos que se adaptaría a los datos observados. Posteriormente, analizaremos algunas formas funcionales que pueden adoptar los modelos econométricos a la hora de ser especificados.

2- Estimación: consiste en cuantificar de forma aproximada los parámetros que aparecen en el modelo.

3- Validación: etapa a partir de la cual decidimos realmente si aceptamos o refutamos una teoría o modelo, sirviéndonos de herramientas econométricas.

4- Explotación: análisis estructural, predicción y simulación son las tres posibles explotaciones que podemos llevar a cabo.

En este trabajo vamos a centrarnos en la simulación de modelos, con el objetivo de analizar la bondad de determinados instrumentos, por ello profundizaremos más adelante en esta etapa.

*"La simulación trata de generar la trayectoria de las variables objeto de estudio (variables objeto) ensayando cambios en las variables que las explican (variables instrumento). La simulación permite en consecuencia, mostrar los efectos temporales de ciertas medidas concretas." (Trívez, 2004)*

# 3. CONTEXTUALIZACIÓN DEL EXPERIMENTO.

A lo largo de este trabajo hemos llevado a cabo un Estudio de Monte Carlo, el cual hemos realizado con el objetivo de analizar el comportamiento del contraste Reset. Este contraste se usa en Econometría, principalmente, para comprobar si la forma funcional de un modelo que hemos especificado y estimado coincide con la del modelo verdadero. Por lo tanto, para comprender el funcionamiento que llevaremos a cabo en el estudio, debemos contextualizarlo previamente, y es lo que hacemos a lo largo de este apartado. Conoceremos qué son los modelos econométricos, así como las formas funcionales más habituales que pueden tomar éstos, a su vez veremos qué posibles errores pueden cometerse al especificar un modelo, y por último explicaremos la forma de proceder del contraste Reset para detectar si una forma funcional es o no correcta.

## 3.1. LOS MODELOS ECONOMÉTRICOS. LA FORMA FUNCIONAL.

Para especificar un modelo econométrico es necesario tener claro el factor económico que queremos estudiar y la variable que lo representa, así como los posibles elementos que influyen en ella y la forma en que éstos se relacionan. En segundo lugar, hemos de introducir un término aleatorio para que el modelo sea econométrico, y podamos razonar en términos probabilísticos. De hecho, este término que se conoce como perturbación aleatoria, es lo que diferencia los modelos econométricos de los económicos.

Nos centramos a continuación en analizar cuáles son las diferentes relaciones o **formas funcionales** más habituales en la literatura económica. Evidentemente existen muchas más, y más complejas.

Así pues, distinguimos en primer lugar entre **modelos lineales y no lineales**.

### 3.1.1 Modelo lineal.

Cuando la especificación de un modelo es lineal tanto en variables como en parámetros nos encontramos ante un modelo lineal. Este tipo de modelos nos son útiles porque la realidad económica, en muchas ocasiones puede adaptarse a esta especificación. Además las inferencias y métodos utilizados, pueden aplicarse también a modelos no lineales si estos son susceptibles de linealización. Si dicha transformación es posible, diremos que los modelos son intrínsecamente lineales.

Un modelo lineal en el que por simplificar, especificaremos con una única variable explicativa, adoptaría la siguiente forma:

$$Y_I=\beta_0+\beta_1X_{1i}+u_i$$

Planteamos como ejemplo de modelo lineal la versión más simplificada de la función de consumo keynesiana, la cual ha sido estimada a partir de datos proporcionados por el INE, utilizando una muestra compuesta por 50 familias españolas, y se analiza de ellas el consumo que han realizado a lo largo de 2006.

Ejemplo 1.

***Función de consumo keynesiana :***

$C_i=\beta_0+\beta_1Yd_i+u_i$

*Donde C e Yd representan el consumo y la renta disponible respectivamente. Además observamos que la función muestra que el consumo de cada individuo depende de una constante que es el consumo autónomo ($\beta_0$) y de la renta disponible ponderada por la propensión marginal a consumir ($\beta_1$*Yd). La inclusión de la perturbación aleatoria ($u_i$) hace de ella un modelo econométrico.*

***Estimación del modelo***

$\hat{C_i}= 539.42+0.73Y_{di}$

*Lo que significa que por cada unidad monetaria que se incrementa la renta disponible de los individuos, su consumo se incrementa 0.73 unidades monetarias. Además independientemente de la renta disponible de cada individuo, se consume de forma autónoma 539.42 unidades monetarias.*

A diferencia de los modelos lineales, en los no lineales la ausencia de linealidad puede ser en variables, en parámetros, o ambos. Si dicha ausencia se da únicamente en variables será posible transformar un modelo no lineal en lineal. Sin embargo, no en todos los casos en los que la ausencia de linealidad afecta a los parámetros será posible la transformación.

Cuando un modelo puede transformarse en lineal para aplicar sobre él los métodos e inferencias conocidos y utilizados en los modelos lineales, decimos que se trata de un modelo intrínsecamente lineal. Por el contrario, si no es posible la transformación bajo ningún concepto, hablaremos de modelos intrínsecamente no lineales, para los cuales será necesario aplicar métodos específicos. Tantas son las posibles transformaciones como la variedad de **modelos intrínsecamente lineales**, veremos a continuación los más utilizados.

### 3.1.2 Modelo doblemente logarítmico.

Una de las maneras que podríamos utilizar para obtener modelos lineales, si el modelo original lo permite, es la aplicación de logaritmos. Suponemos un modelo teórico que adopta la siguiente forma:

$$Y_i = AX_i^{\beta}$$

Si aplicamos logaritmos en la ecuación anterior el resultado sería:

$$\ln Y_i = \ln A + \beta \ln X_i$$

los logaritmos aparecen, como podemos ver, a ambos lados de la ecuación, afectando tanto a la variable endógena como a la explicativa. Cuando la forma funcional adopte esta expresión, el modelo será doblemente logarítmico.

La relación de las variables en los modelos doblemente logarítmicos, en el caso de existir una única variable explicativa, adoptan la forma que podemos ver en el grafico que presentamos a continuación:

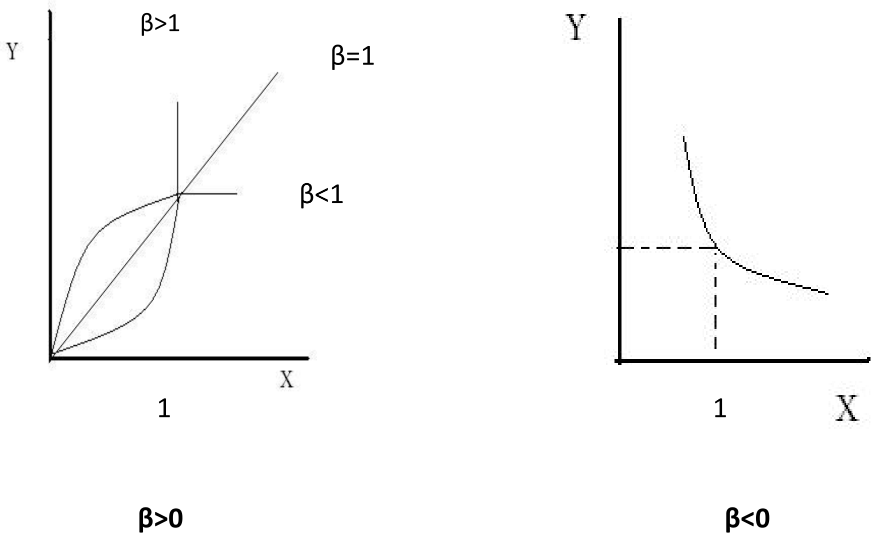


**Figura 1**. Representaciones gráficas del modelo doblemente logarítmico.

Para facilitar la comprensión de este tipo de modelos, seguiremos un ejemplo de Wooldridge (2006):

Ejemplo 2.

***Salario de director general y ventas de la compañía***

*Log* ($salary_i$)=$\beta_0$+ $\beta_1$log($sales_i$)+$u_i$

*Que explica la relación entre el salario del director general de una compañía (salary), medido en dólares, y las ventas de la misma (sales), medidas en millones de dólares.* $\beta_1$ *es la elasticidad de salary respecto de sales, lo que significa que si las ventas aumentan 1%, el salario del director general lo hará en* $\beta_1$*%.*

***Estimación del modelo***

$Log(\widehat{salary}_i) = 4.822+0.257\ log(sales_i)$

*Según el modelo estimado, si las ventas se ven incrementadas en un 1%, el salario del director general de la compañía aumentara un 0.275%.*

### 3.1.3 Modelos semilogarítmicos.

Otro tipo de modelos intrínsecamente lineales, ampliamente utilizados son los semilogarítmicos. Nos referimos a ellos cuando únicamente un lado de la ecuación se encuentra expresado en logaritmos, diferenciando así entre aquellos en los que tan solo la variable endógena aparece en logarítmicos y aquellos en los que son las explicativas y no la variable endógena a las que les afecta el neperiano.

Nos referiremos a partir de ahora al primero de los casos como modelo semilogarítmico 1, y al segundo como modelo semilogarítmico 2.

En cuanto al modelo semilogaritmico 1, si el modelo teórico de partida se expresa:

$$Y_i = e^{\alpha+\beta Xi}$$

solucionaremos la no linealidad aplicando logaritmos, quedando como resultado:

$$\ln Y_i = \alpha + \beta X_{i.}$$

La representación grafica de los modelos semilogarítmicos se presenta a continuación:

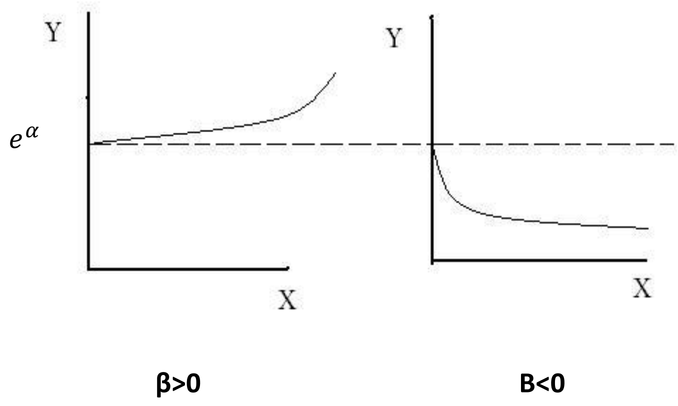


**Figura 1**. Representaciones gráficas del modelo semilogarítmico.

Es habitual que los llamados modelos de crecimiento adopten esta forma funcional, siendo la variable explicativa el tiempo. Sin embargo hemos optado por elegir un modelo más sencillo que nos sirva como ejemplo, tomado de Wooldridge (2006):

Ejemplo 3.

***Ecuación del salario por hora***

$log\ (wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 tenure_i + u_i$

*Donde la educación (educ), la experiencia laboral (exper) y la antigüedad en la empresa (tenure), medidas ambas en años, explican el salario de los trabajadores (wage) medido en dólares.*

*$\beta_1$ es la semieslasticidad del salario respecto de la educación, de tal forma que un año adicional de educación, haría aumentar el salario en ($\beta_1$*100)%. La misma explicación podemos ampliarla a la interpretación de $\beta_2$ y $\beta_3$.*

***Estimación del modelo***

$\log(\widehat{wage}) = 0.284 + 0.092 educ + 0.0041 exper + 0.022 tenure$

*La estimación determina que cada año adicional dedicado por un individuo a su educación supone un aumento de 9.2% de su salario; por cada año adicional que el trabajador adquiere experiencia adicional, su salario aumenta en 0.41%, y del mismo modo por cada año de permanencia en la empresa, su salario aumenta 2.2%.*

El segundo caso de los más conocidos modelos semilogarítmicos (modelo semilogarítmico 2, parte de un modelo teórico:

$$e^{Y_i}=AX_i^{\beta_1}$$

cuya linealización con la aplicación de logaritmos conlleva la obtención de un modelo tal que:

$$Y_i=\beta_0+\beta_1 \ln X_i$$

Para facilitar la comprensión de la importante interpretación de las semielasticidades, veremos un ejemplo siguiendo a Wooldridge (2006), quien especifica y estima un modelo para explicar los determinantes de la nota media en la universidad. Para la estimación se realizó una investigación en 1993 a 408 institutos de Michigan:

Ejemplo 4.

***Rendimiento escolar y tamaño de la escuela***

$Math10_i = \beta_0 + \beta_1 log(totcomp_i) + \beta_2 \, log\,(staff_i) + \beta_3 \, log(enroll_i) + u_i$

*Donde el rendimiento escolar se mide por el porcentaje de alumnos que aprueban el examen de matemáticas de décimo curso(math10); resultado que como muestra el modelo estará influido por la retribución media de los profesores (totcomp), la cantidad de personal del colegio por cada mil estudiante s(staff) y por el tamaño de los institutos que se mide por la matrícula estudiantil (enroll).* $B_1$ $B_2$ *y* $B_3$. *Representan las semielasticidades. Por lo que un aumento de 1% de totcomp, de staff o de enroll, hace aumentar el rendimiento de los estudiantes en* ($B_1$ /100), ($B_2$/100) *y* ($B_3$/100) *unidades respectivamente.*

***Estimación del modelo***

$\widehat{math}10 = -207.66+21.16 log(totcomp)+3.98 log(staff)-1.29 \, log(enroll)$

*Un incremento de 1% en la retribución media de los profesores, hace aumentar el porcentaje de alumnos que aprueban el examen de matemáticas e 0.21. Del mismo modo interpretaríamos las otras dos variables explicativas.*

### 3.1.4 Modelo mixto.

Puede ocurrir que tan solo algunas variables aparezcan en logaritmos sin cumplir exactamente las cualidades de los ya citados modelos doblemente logarítmicos o semilogarítmicos, hablamos entonces de modelos mixtos. La forma que puede adoptar este tipo de modelos es muy diversa.

Utilizaremos un modelo de precio hedónico como ejemplo, siguiendo a Wooldridge (2006). La estimación ha sido realizada a partir de 19 observaciones de casas que fueron vendidas en Massachusetts en el año 1990.

Ejemplo 5.

***Modelo de precio hedónico para la vivienda***

*Log (prize$_i$) =*$\beta_0 + \beta_1$ *log (sqrft$_i$)* + $\beta_2$ *bdrms$_i$* + $\beta_3$ *bthrms$_i$* + $u_i$

*Se trata de un modelo que explica el precio de una vivienda en función de sus características. En este caso la superficie medida en pies cuadrados (sqrft), el número de habitaciones (bdrms) y el número de cuartos de baño (bthrms), determinarán el valor monetario de la vivienda (prize). La interpretación de los coeficientes variará en función de si las variables a las que acompañan aparecen o no en logaritmos.*

***Estimación del modelo***

*Log* ($\widehat{price}_i$) = 7.46 + 0.634 *log (sqrft$_i$)* – 0.066*bdrms$_i$* + 0.158 *bthrms$_i$*

*Si el tamaño de la vivienda aumenta en 1%, el precio de la misma aumentara 0.634%. por cada habitación adicional que tenga la casa, el precio de la vivienda disminuirá 6.6 % (aunque parezca contra intuitiva la relación entre el precio y el número de habitaciones es necesario recordar que estamos suponiendo una situación ceteris paribus, es decir mismo tamaño y mismo número de cuartos de baño). Si añadimos un cuarto de baño a la vivienda, el precio de la misma se incrementará 15.8%.*

### 3.1.5 Modelos cuadráticos.

Los modelos cuadráticos son un tipo de especificación no lineal que se caracteriza porque una o varias variables explicativas aparecen elevadas al cuadrado.

Su aplicación es útil cuando lo que buscamos es captar efectos marginales crecientes o decrecientes, algo muy utilizado en economía.

A través de un sencillo ejemplo de un modelo cuadrático que aparece tomando a Wooldrige (2006), conoceremos como funcionan este tipo de modelos.

Ejemplo 6.

***Experiencia como determinante del salario***

$Wage_i = \beta_0 + \beta_1 exper_i + \beta_2\, exper_i^2 + u_i$

*La influencia que tienen los años de experiencia adquiridos por un individuo en el mercado laboral (exper) sobre el salario (wage), viene determinada por la derivada parcial de la experiencia con respecto del salario:*

$\frac{\partial exper}{\partial salary} = \beta_{1+}\ 2*\beta_{2*}\, exper.$

*Suponiendo que $\beta_1$ es positivo y $\beta_2$ es negativo, el modelo estaría mostrando que aunque los incrementos en los años de experiencia tengan un efecto positivo sobre el salario, cada vez tal efecto será menor (signo negativo del término cuadrático),*

***Estimación del modelo***

$\widehat{wage} = 3.73 + 0.298exper - 0.0061exper^2$

*El efecto marginal que supone sobre el salario un incremento de un año de la experiencia de un trabajador que tal solo lleva trabajando un año (exper=1), será de 0.286 (0.298-2*0.0061*1).*

## 3.2. ERRORES DE ESPECIFICACIÓN.

Tal y como enunció Davidson: "Debido a la naturaleza no experimental de la economía, nunca estamos seguros de la forma en que se generaron los datos observados (…)".

Es lo que ocurre a la hora de especificar un modelo, podemos cometer, entre otros, errores de especificación cuando la forma funcional de la regresión estimada difiere de la forma funcional de la regresión poblacional. Si se produce dicho error, el estimador del efecto parcial de un cambio de una de las variables será, en general, sesgado.

Gujarati (1999) entre otros, nos explican los diferentes errores de especificación que pueden producirse. De aquellos que afectan a la parte sistemática, comentamos a continuación los más comunes, así como sus posibles consecuencias:

- **Inclusión de variables irrelevantes:** es decir, introducimos en la regresión variables cuyos coeficientes son cero en el modelo verdadero.
  Por ejemplo, un modelo se especifica como: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ , y sin embargo el modelo verdadero es $Y_i = \beta_1 + \beta_2 X_{2i} + u_i$. Necesariamente $\beta_3 = 0$ , lo que significa que $X_{3i}$ es una variable irrelevante que deberíamos excluir.

Las consecuencias de este error son estimadores menos eficientes, pero no hay sesgo en los parámetros de las variables relevantes.

- **La omisión de variables relevantes:** puede producirse que debido al desconocimiento de las mismas no se incluyan en el modelo algunas variables explicativas de la endógena. Como consecuencia nos encontramos ante una estimación sesgada e inconsistente de los coeficientes y de la varianza de las perturbaciones. Suele producirse el incumplimiento de las hipótesis de homocedasticidad y no autocorrelación. Además el sesgo será mayor cuanta más importancia tengan las variables explicativas omitidas en la explicación de la endógena; así como cuanto mayor sea la relación de las variables incluidas en el modelo y las relevantes excluidas.
- **Una forma funcional incorrecta:** se da cuando la forma funcional del modelo verdadero difiere de la del modelo estimado.

Nuestro trabajo se centra en analizar este último tipo de error.

### 3.3. CONTRASTE RESET DE RAMSEY.

Si queremos analizar si la forma funcional del modelo que hemos especificado coincide con el modelo verdadero, podemos observar en primer lugar, los gráficos de los residuos. Estos nos permitirán realizar una primera aproximación para conocer si existen o no errores de especificación, sin embargo, no sabremos si dichos errores tienen que ver con la forma funcional. Un análisis más riguroso sería el proporcionado por un contraste estadístico, utilizando el contraste Reset, el cual si que contrastaría los errores en la forma funcional del modelo estimado.

Vamos a suponer que el modelo para el cual queremos realizar el contraste, se especifica de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (1)$$

En el contraste Reset, la hipótesis nula y alternativa a contrastar son:

$$\begin{cases} H_0\text{: forma funcional correcta} \\ H_1\text{: forma funcional incorrecta} \end{cases}$$

Y los pasos para realizar el contraste son:

1º paso: estimamos el modelo para obtener la variable dependiente ajustada, es decir, obtenemos $\hat{Y}$.

2º paso: especificamos una nueva regresión en la cual incorporamos como variable explicativa, la variable endógena estimada ($\hat{Y}$ ) elevada al cuadrado, al cubo, etc tantas veces como consideremos oportuno.

$$Y_i = \beta_0+\beta_1 X_i+ \delta_1\widehat{Y_i^2}+\ldots+ \delta_p\widehat{Y_i^p}+ \varepsilon_i \qquad (2)$$

Si el modelo inicial está correctamente especificado los valores de los coeficientes δ serán 0. Es decir la hipótesis nula y alternativa en términos de esta regresión auxiliar tomaría la siguiente forma:

$$\begin{cases} H_0: \delta_2=\delta_3=\ldots\ldots=\delta_p=0 \\ H_1: \text{algún } \delta_p \neq 0 \end{cases}$$

3º paso: tomamos la decisión de aceptar o rechazar la hipótesis nula en función del estadístico F, que adopta la siguiente forma:

$$F_{\text{RESET}}= \frac{SR_0- SR_A}{SR_A}\,\frac{T-k}{p-1} \sim F\,(p\text{-}1,\ T\text{-}k)$$

Donde $SR_0$ es la suma residual del modelo estimado original (1), esto es, el modelo restringido bajo $H_0$ y $SR_A$ es la suma residual de la regresión auxiliar (2), es decir, del modelo amplio. El estadístico sigue una distribución F, donde p es el exponente máximo de $\hat{Y_i}$ en la regresión auxiliar, T el tamaño muestral y k el número de parámetros de la regresión auxiliar. Si aceptamos la hipótesis nula, estamos afirmando que la forma funcional inicial del modelo es correcta.

Este contraste tiene una fácil aplicación, lo que ocurre es que por el hecho de saber que un modelo está mal especificado no significa que hallemos también la solución. Es decir, el contraste Reset únicamente nos permite detectar que existe un problema de forma funcional, pero no soluciona el mismo.

A lo largo de este trabajo, hemos analizado si el funcionamiento del contarse Reset de Ramsey es adecuado, es decir, si nos permite extraer conclusiones verdaderas al aceptar o rechazar la hipótesis nula de forma funcional correcta. Para comprobarlo, realizamos un estudio de Monte Carlo, cuyo procedimiento explicaremos a continuación.

# 4. METODOLOGÍA DE ANÁLISIS: ESTUDIO DE MONTECARLO.

En este capítulo vamos a explicar cómo hemos llevado a cabo el experimento. Hemos considerado necesario, definir en primer lugar una serie de conceptos básicos que nos permitirán una mayor comprensión del estudio realizado. Posteriormente, se explica minuciosamente en qué consiste y el procedimiento seguido en el experimento de Monte Carlo. Por último, se presentan laos resultados y conclusiones derivadas del mismo.

## 4.1. CONCEPTOS DE INTERÉS

Los siguientes conceptos nos facilitarán la comprensión de la simulación y los experimentos de Monte Carlo.

1- Contraste de hipótesis: es un procedimiento estadístico utilizado para tratar de concluir si una característica propia de una población coincide con lo observado en una muestra. Se plantea una hipótesis nula y otra alternativa para tratar de dirimir cual de las dos es verdadera a partir de un estadístico de contraste, teniendo en cuenta un número determinado de experimentos.

2- Tipos de error

2.1- Error tipo I: se produce cuando rechazamos la hipótesis nula siendo ésta verdadera. El tamaño del error tipo I es la probabilidad que existe de cometerlo. Por ello, en todo contraste buscamos un tamaño nominal pequeño (0.01, 0.05 y 0.1).

2.2- Error tipo II: se produce cuando aceptamos la hipótesis nula siendo esta falsa. La probabilidad de cometer dicho error es el tamaño de error tipo II.

3- Potencia de un contraste: es la probabilidad de rechazar la hipótesis nula cuando ésta es falsa, es decir, 1 menos el tamaño de error tipo II. Un contraste de hipótesis funciona de forma adecuada cuando la potencia es elevada.

## 4.2. ESTUDIOS DE MONTE CARLO

La simulación es el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprobar el comportamiento del sistema o evaluar nuevas estrategias-dentro de los límites impuestos por un determinado criterio o conjunto de ellos- para el funcionamiento del sistema (Shannon, 1998).

Un estudio de Monte Carlo se basa en que cada simulación es la repetición de un proceso cuya base está en un modelo que presenta unas determinadas condiciones. El objetivo fundamental es la obtención de numerosos resultados de forma aleatoria en base a unos supuestos. Kennedy (2003) define el estudio de Monte Carlo como "un ejercicio de simulación diseñado para analizar las propiedades que en muestras pequeñas tienen los estimadores que utilizamos ante un problema de estimación dado". Además permite corroborar el cumplimiento de propiedades teóricas de un determinado estadístico o estimador.

Con carácter general, un experimento de Monte Carlo sigue la siguiente estrategia:

a) En primer lugar, se especifica un modelo "verdadero" o Proceso Generador de Datos (PGD), esto consiste en especificar la distribución del término de perturbación aleatoria, las variables explicativas, los coeficientes del modelo y el tamaño muestral, con el objeto de generar un conjunto de datos de la variable endógena. El término de la perturbación aleatoria irá cambiando en cada simulación cuando repetimos el proceso.

b) Se genera un conjunto de datos de la variable endógena utilizando dicho PGD.

c) Estimamos el modelo con esta muestra.

d) Calculamos la medida objeto de análisis

e) Se repiten las etapas b), c) y d) un elevado número de veces. Cada generación de un nuevo conjunto de datos se denomina réplica. En cada una de las simulaciones, las variables explicativas no cambian, por ello las consideramos fijas en muestras repetidas (exógenas).

e) Por último, se evalúa el resultado obtenido en el experimento planteado.

Evidentemente, esto son pautas generales, que sufren modificaciones para adaptarse en cada caso al objetivo perseguido. A continuación, profundizaremos en nuestra aplicación concreta del experimento de Monte Carlo.

# 5. ANÁLISIS DEL CONTRASTE RESET.

Nuestro análisis va a centrarse en aplicar el método de Monte Carlo para conocer como actúa el contraste Reset ante los errores en la forma funcional. Sabiendo de antemano qué forma funcional adopta el modelo verdadero, esto es el proceso generador de datos, nos preguntamos si dicho contraste es capaz de detectar los errores en la forma funcional de los modelos estimados. Queremos ver si funciona mejor según cuál sea el modelo verdadero, y el modelo estimado. También nos interesa conocer hasta qué valor de p (exponente de $\hat{Y}_i$) debe incluirse en la regresión auxiliar. Finalmente queremos analizar la influencia del tamaño muestral.

## 5.1. DISEÑO DEL EXPERIMENTO

Para realizar el estudio de Monte Carlo vamos a utilizar el programa Gretl.
En primer lugar, consideraremos seis PGD diferentes, y en todos ellos habrá dos variables explicativas que han sido generadas en este caso como una distribución uniforme, mientras que la perturbación aleatoria sigue una distribución normal de media 0 y varianza 2.

Trabajaremos con tres tamaños muestrales, T= 50, 100 ó 500 que responden a lo que podemos considerar como una muestra pequeña, mediana o grande, lo cual se hace con el objetivo de ver cómo varía el comportamiento del contraste a medida que aumenta el tamaño muestral. El número de réplicas que se llevaran a cabo son 1000, número que hemos considerado suficientemente grande como para extraer conclusiones fiables.

Para llevar a cabo nuestro estudio, hemos tenido en cuenta los siguientes aspectos:

a) Los PGD o modelos verdaderos que vamos a utilizar a lo largo de nuestro experimento adoptarán la forma de un modelo lineal, un modelo doblemente logarítmico un modelo semilogarítmico con logaritmos únicamente en la variable endógena (semilogarítmico 1), o únicamente con logaritmos en las variables exógenas (semilogarítmico 2), un modelo mixto y un modelo cuadrático. La forma concreta de cada uno de ellos es la siguiente:

***M1:Modelo lineal*** $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

***M2: Modelo doblemente logarítmico*** $\ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + u_i$

***M3: Modelo semilogarítmico (1)*** $\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

***M4: Modelo semilogarítmico (2)*** $Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + u_i$

***M5: Modelo mixto*** $\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 \ln X_{2i} + u_i$

***M6: Modelo cuadrático*** $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \beta_3 X_{2i} + u_i$

La razón por la cual hemos decidido investigar tomando como referencia este tipo de formas funcionales es porque son los más usados en economía.

b) Para cada uno de estos seis PGD, vamos a considerar diferentes especificaciones de modelos, que serán estimados, y para cada uno de los cuales calcularemos el estadístico Reset. De este modo, si el PGD y el modelo especificado coinciden, esperamos que el estadístico Reset que contrasta la hipótesis nula de forma funcional correcta, nos lleve a aceptar dicha hipótesis nula. Por el contrario, si el modelo especificado no coincide con el PGD, el estadístico Reset debería llevarnos a rechazar $H_0$, si realmente es un contraste adecuado para detectar este problema. No obstante, hay que tener en cuenta que la decisión de aceptar o rechazar se lleva a cabo para un nivel de significación, (habitualmente 5%), por lo que aunque el PGD y el modelo especificado coincidan, se espera que el porcentaje de réplicas en las que se rechaza $H_0$ esté alrededor del 5%.

Cuando la especificación no coincida con el PGD, Reset se comportará bien si rechaza $H_0$ el mayor número de veces posibles. El porcentaje de replicas en las que se rechaza $H_0$ cuando realmente es falsa, será la potencia del contraste, que esperamos se aproxime a 1 (100%).

Por lo tanto, nuestro objetivo será analizar si el contraste Reset capta adecuadamente errores en la forma funcional, viendo si experimenta grandes cambios en función del tamaño muestral utilizado. Como segundo objetivo, pretendemos abordar cuántos han de ser los elementos adicionales que deberíamos incluir en la regresión auxiliar del contraste, es decir, si los resultados son diferentes según cuál sea el valor de p, y si podemos concluir que existe un valor óptimo de p o no.

Por último, estudiaremos si existen diferencias en el comportamiento según cuál sea el PGD y cuál sea el modelo especificado y estimado.

## 5.2. RESULTADOS DEL EXPERIMENTO DE MONTE CARLO

A continuación presentaremos los resultados obtenidos en el experimento y las conclusiones que se derivan de ellos. De forma individual comenzaremos analizando cada uno de los PGD, para los cuales hemos especificado y estimado seis modelos diferentes. Veremos en cada caso el efecto que el tamaño muestral y la regresión auxiliar utilizada en el contraste tienen sobre las conclusiones.

Una vez hayamos realizado este proceso, podremos extraer conclusiones sobre el conjunto de resultados, y es lo que a su vez nos permitirá conocer el último de nuestros objetivos: sí existen diferencias en función del tipo de PGD y modelo estimado con el que trabajemos.

En las tablas aparece siempre la proporción de rechazo de $H_0$ entre el total de réplicas realizadas. Es necesario tener claro que si el modelo verdadero coincide con el modelo especificado y estimado, los resultados incluidos en las tablas reflejaran el tamaño de error tipo I. Sí por el contrario, no coinciden, los resultados estarían mostrando la potencia de cada contraste. Para llegar a considerar un buen funcionamiento del contraste Reset el tamaño del error tipo I que esperaríamos obtener estaría en torno al nivel de significación determinado en el experimento, en este caso 5%, mientras que la potencia, buscamos que sea lo más alta posible, que se acerque a 1 (100%).

### 5.2.1. PGD lineal.

Empezamos analizando aquella situación en la que el modelo verdadero, es lineal.Los parámetros que acompañan a las variables explicativas son en este caso iguales a 1. Los resultados que arroja el contraste Reset, son los que aparecen en la Tabla1.

En todas las tablas, vamos a incluir en primer lugar la estimación que corresponde a la misma especificación que tiene el modelo verdadero. Como se están mostrando el porcentaje de rechazos de $H_0$ y en esta primera estimación $H_0$ es cierta, la proporción de rechazos debería estar en torno al 5%, que corresponde al nivel de significación establecido en el experimento. Posteriormente, el resto de estimaciones corresponden a especificaciones distintas del PGD, por lo que las tablas proporcionan la potencia del contraste (la proporción de rechazos de Ho cuando esta es falsa. Empezamos analizando aquella situación en la que el modelo verdadero, es decir, el que ha generado los datos, es lineal.

**Tabla 1. PGD = Modelo lineal (M1).**

| PGD: LINEAL (M1) | | | |
|---|---|---|---|
| **Estimación de un modelo lineal (M1)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.05* | *0.05* | *0.058* |
| **100** | *0.049* | *0.059* | *0.051* |
| **500** | *0.052* | *0.063* | *0.063* |
| **Estimación de un modelo doblemente logarítmico (M2)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.507* | *0.577* | *0.625* |
| **100** | *0.635* | *0.653* | *0.66* |
| **500** | *0.857* | *0.878* | *0.883* |
| **Estimación de un modelo semilogarítmico 1 (M3)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.839* | *0.806* | *0.776* |
| **100** | *0.983* | *0.98* | *0.982* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo semilogarítmico 2 (M4)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *1* | *1* | *1* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo mixto (M5)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.485* | *0.384* | *0.619* |
| **100** | *0.63* | *0.511* | *0.656* |
| **500** | *0.852* | *0.802* | *0.882* |
| **Estimación de un modelo cuadrático (M6)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.051* | *0.049* | *0.059* |
| **100** | *0.05* | *0.057* | *0.049* |
| **500** | *0.05* | *0.059* | *0.061* |

Siendo el modelo verdadero lineal (M1) y el especificado y estimado un modelo con la misma forma funcional, las tres primeras filas de la Tabla.1, muestran un nivel de significación próximo al fijado (5%). Esto implica un buen comportamiento del contraste cuando Ho es cierta: se rechaza aproximadamente en el 5% de las simulaciones, incluso con un tamaño muestral pequeño. Podemos ver también, que el ajuste a este valor es bastante bueno con p=2, por lo que no sería necesario añadir exponentes más altos a la $\hat{Y}$ que incluimos en la regresión auxiliar del contraste.

En cuanto a la potencia, alcanza su máximo valor cuando el modelo estimado es un semilogarítmico con logaritmos en las variables explicativas (M4), y obtenemos dicho resultado independientemente de cuál sea el tamaño muestral. Además también es alta cuando el modelo estimado es un modelo semilogarítmico con logaritmos en la variable endógena (M3).

La potencia se hace intermedia para los casos en los que las estimaciones se hacen a partir de un modelo doblemente logarítmico o un modelo mixto. Este valor de la potencia no es demasiado alto teniendo en cuenta que el contraste Reset debería discriminar más a este tipo de modelos que son suficientemente distintos al lineal como para rechazar la hipótesis nula un número de veces mayor. La variable endógena aparece en logaritmos en el modelo estimado, mientras que en el modelo verdadero está en niveles. Las bajas potencias en estos dos casos podrían ser un indicio de un incorrecto funcionamiento del contraste.

Si la estimación se hace a partir de un modelo cuadrático, las potencias son mínimas (en torno al 0.05). Esto puede resultar lógico debido a la similitud con el modelo verdadero, ya que la única diferencia es el término cuadrático. Estos resultados, estarían indicando que a la hora de especificar el modelo hemos cometido el error de inclusión de una variable irrelevante.

### 5.2.2 PGD doblemente logarítmico.

Abordaremos de nuevo cada una de las cuestiones anteriormente vistas, teniendo en cuenta, en este caso, que el modelo verdadero es doblemente logarítmico (M2). Los coeficientes que acompañan a las variables explicativas asumen valores mayores que 1, concretamente $\beta_1$=8 y $\beta_2$=10. La razón por la cual los coeficientes elegidos no son iguales a 1, es porque si lo hicieran el modelo teórico de partida ya sería lineal. (ver punto 3).

**Tabla 2. PGD = Modelo doblemente logarítmico (M2).**

| PGD: DOBLEMENTE LOGARÍTMICO (M2) | | | |
|---|---|---|---|
| **Estimación de un modelo doblemente logarítmico (M2)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.043* | *0.052* | *0.051* |
| **100** | *0.05* | *0.06* | *0.052* |
| **500** | *0.055* | *0.058* | *0.058* |
| **Estimación de un modelo lineal (M1)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.907* | *0.864* | *0.854* |
| **100** | *0.96* | *0.932* | *0.926* |
| **500** | *0.999* | *0.998* | *0.997* |
| **Estimación de un modelo semilogarítmico 1 (M3)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.977* | *0.963* | *0.953* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo semilogarítmico 2 (M4)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.884* | *0.85* | *0.778* |
| **100** | *0.948* | *0.932* | *0.843* |
| **500** | *0.997* | *0.997* | *0.984* |
| **Estimación de un modelo mixto (M5)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.04* | *0.011* | *0.048* |
| **100** | *0.043* | *0.016* | *0.061* |
| **500** | *0.074* | *0.034* | *0.114* |
| **Estimación de un modelo cuadrático (M6)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.904* | *0.877* | *0.883* |
| **100** | *0.966* | *0.935* | *0.955* |
| **500** | *0.999* | *0.997* | *0.996* |

Inicialmente observamos que el tamaño del error tipo I, se aproxima al nivel de significación establecido (5%). Por ello podemos concluir una adecuada detección por parte del contraste Reset, ante una correcta forma funcional, y un p=2 sería el más adecuado.

Cuando el modelo verdadero es doblemente logarítmico y la estimación la hacemos con un M3, obtenemos potencias muy altas, incluso si T es pequeño.

El contraste Reset también discrimina mucho los modelos M1 y M6, que como observamos presentan potencias muy altas. Esto es debido a que existen suficientes diferencias entre el modelo verdadero y los estimados, por lo que resulta lógico que el contraste sea capaz de detectar la errónea forma funcional en un elevado número de veces de entre todas las réplicas realizadas. Una de las diferencias importantes que pueden ser razón de este alto valor en la potencia, es que la variable endógena aparece en los modelos M1 y M6 en niveles, mientras que en el PGD aparece en logaritmos. El modelo M4 también presenta potencias altas, sobre todo a partir de T=100.

Por el contrario, sabemos que el modelo mixto (M5) lo hemos especificado como $\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 \ln X_{2i} + u_{i,}$ por lo que esas diferencias que comentábamos anteriormente se convierte ahora en una similitud, ya que tanto en este modelo mixto como en el modelo doblemente logarítmico que es ente caso PGD, la variable endógena es especificada de la misma forma (lnY), y una de las explicativas del mixto también está en logarítmos. Como consecuencia el contraste Reset no es capaz de detectar la incorrecta forma funcional, pues tan solo alcanzamos potencias aproximadamente del 0.05.

Otra de las observaciones importantes, es que basta con utilizar p=2 pues los resultados no experimentan grandes diferencias al utilizar una u otra regresión auxiliar. Incluso hay casos en los que la potencia es mayor con p=2, que utilizando exponentes más altos.

Además en este caso los resultados proporcionados por el contraste no difieren de forma importante según utilicemos uno u otro tamaño muestral. Tal vez, observamos valores ligeramente superiores en T=50, aún siendo altos en M1, M4, M6 y M3. Por ello en los casos en lo que ya hemos comentado que el contraste actúa de acuerdo a lo previsto, lo hará tanto si utilizamos muestras grandes, medianas, o pequeñas.

En términos generales, el contraste sería potente, excepto en el modelo mixto, dado su alto grado de similitud con el modelo verdadero.

### 5.2.3 PGD: Modelo semilogarítmico 1 – logaritmos únicamente en la variable endógena.

El modelo verdadero adopta ahora una forma semilogarítmica, con logaritmos únicamente en la variable endógena (M3), los coeficientes que acompañan a las variables explicativas son mayores que 0, concretamente $\beta_1$=0.8 y $\beta_2$=0.3. Se han especificado y estimado los 6 modelos vistos también con anterioridad, y los resultados derivados de llevar a cabo el contraste Reset son los que aparecen en la tabla 3.

**Tabla 3. PGD = Modelo semilogarítmico 1 (M3).**

| PGD : SEMILOGARÍTMICO 1 (M3) | | | |
|---|---|---|---|
| **Estimación de un modelo semilogarítmico 1 (M3)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.009* | *0.011* | *0.014* |
| **100** | *0.003* | *0.009* | *0.012* |
| **500** | *0.005* | *0.015* | *0.021* |
| **Estimación de un modelo lineal (M1)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.492* | *0.44* | *0.42* |
| **100** | *0.575* | *0.494* | *0.47* |
| **500** | *0.895* | *0.836* | *0.801* |
| **Estimación de un modelo doblemente logarítmico (M2)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.412* | *0.311* | *0.269* |
| **100** | *0.676* | *0.583* | *0.519* |
| **500** | *1* | *0.997* | *0.997* |

| Estimación de un modelo semilogarítmico 2 (M4) | | | |
|---|---|---|---|
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.474* | *0.443* | *0.324* |
| **100** | *0.666* | *0.588* | *0.374* |
| **500** | *0.94* | *0.927* | *0.821* |
| **Estimación de un modelo mixto (M5)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.392* | *0.22* | *0.25* |
| **100** | *0.662* | *0.482* | *0.493* |
| **500** | *1* | *0.994* | *0.994* |
| **Estimación de un modelo cuadrático (M6)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.554* | *0.514* | *0.52* |
| **100** | *0.665* | *0.576* | *0.582* |
| **500** | *0.917* | *0.874* | *0.864* |

Si el modelo especificado y estimado es M3, coincidiendo con el PGD, esperamos una proporción muy baja de rechazos de la hipótesis nula. De hecho, debería encontrarse en torno al 0.05, valor que coincide con el nivel de significación establecido en el experimento, o lo que es lo mismo con el tamaño de error tipo I que estamos dispuestos a asumir. Lo que obtenemos en este caso son valores incluso por debajo de este 0.05. Esto implica la claridad con la que el contraste Reset es capaz de detectar que la forma funcional es correcta. El menor tamaño de error tipo I, 0.003 es obtenido para T=100 y p=2. O sea, de entre las 1000 simulaciones que realizamos solo rechazamos un 0.3%.

Una característica común que se produce cuando el PGD es M3, es que para que el contraste sea capaz de captar que existen errores en la forma funcional de los modelos especificados y estimados, ha de utilizarse un tamaño muestral grande, de al menos 500 observaciones. Sólo así podemos alcanzar potencias cercanas a 1. La diferencia si usamos uno u otro tamaño muestral es elevada. De hecho si T=50 tan solo se rechaza la hipótesis nula de forma funcional correcta entre el 20% y 50 % de las veces, valores que no aumentan demasiado si utilizamos una muestra algo mayor de 100 observaciones, pues de nuevo estos porcentajes de rechazo no llegarán a alcanzar a penas el 60%.

Encontramos además un valor excesivamente pequeño cuando estimamos un modelo mixto utilizando T=50 y p=3, pues la potencia es de 0.22.

Podríamos considerar lógico alcanzar potencias bajas cuando los modelos estimados son suficientemente similares al modelo verdadero, como ocurre con un doblemente logarítmico (M2), o un modelo mixto (M5), por especificar la variable endógena de la misma forma (lnY). Sin embargo a pesar de que para muestras pequeñas sí que las potencias son las más bajas, cuando T=500 y p=2 alcanzamos incluso la potencia máxima.

En los casos en los que estimamos un modelo lineal o cuadrático (M1,M6), que presentan enormes diferencias con respecto al PGD, nos sorprende que el contraste no sea capaz de hacer una discriminación mayor usando un tamaño muestral pequeño, y que aun siendo T=500 las potencias a pesar de ser suficientemente grandes, están por debajo de los modelos que comentábamos justo anteriormente que eran más similares (M2, M5).

Otra de las cuestiones abordadas, la regresión auxiliar a utilizar, veremos que basta con utilizar p=2, de hecho es más conveniente pues las potencias alcanzadas son mayores, por tanto, no será necesario incluir la estimación de la variable endógena ni al cubo ni a la cuarta.

Lo que sí podemos concluir con claridad es un adecuado funcionamiento del contraste cuando debe ser capaz de determinar que se encuentra ante una correcta forma funcional, tal como vemos en las tres primeras líneas de la Tabla 3. Siendo además este el único caso en el que los resultados son muy similares usando uno u otro tamaño muestral.

### 5.2.4 PGD : Modelo semilogarítmico (2) – logaritmos únicamente en las variables exógenas.

El modelo verdadero adopta en este caso una forma semilogaritmica, incluyendo logaritmos únicamente en las variables exógenas (M4) cuyos coeficientes son mayores que 1, concretamente $\beta_1$=8 y $\beta_2$= 10. He elegido coeficientes altos con el objetivo de que las semielasticidades que han de ser divididas por 100 en su interpretación tengan sentido.

Se especificarán y estimarán, tal y como veníamos haciendo hasta ahora las seis formas funcionales con las que estamos trabajando. La Tabla 4 , mostrará los resultados obtenidos en cuento al tamaño de error tipo I, y las potencias.

**Tabla 4. PGD = Modelo semilogarítmico 2 (M4)**

| PGD : SEMILOGARÍTMICO 2 (M4) | | | |
|---|---|---|---|
| **Estimación de un modelo semilogarítmico 2 (M4)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.043* | *0.052* | *0.012* |
| **100** | *0.05* | *0.06* | *0.019* |
| **500** | *0.055* | *0.058* | *0.021* |
| **Estimación de un modelo lineal (M1)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.999* | *0.993* | *0.996* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo doblemente logarítmico (M2)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.994* | *0.99* | *0.985* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo semilogarítmico 1 (M3)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *1* | *1* | *1* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo mixto (M5)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.994* | *0.978* | *0.984* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo cuadrático (M6)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.999* | *0.994* | *0.997* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |

En este caso las potencias son muy altas, sea cual sea el modelo estimado. Esto no solo es un aspecto positivo, sino que además es razonable, por lo claramente diferente que es este modelo respecto a cualquier otro. Además las potencias son altas incluso utilizando muestras pequeñas.

En cuanto a la regresión auxiliar, podemos concluir que bastaría con la utilización de p = 2, puesto que usar regresiones más amplias como es el hecho de p = 3 ó 4, no hacen variar los resultados de forma significativa y aun así estas lograrían valores en las potencias menores.

Tal y como venía ocurriendo con el resto de casos, el contraste Reset sigue siendo capaz de detectar correctamente cuando la forma funcional es correcta, minimizando el tamaño del error tipo I, que coincide con el nivel de significación (0.05).

### 5.2.5 PGD: Modelo mixto.

El modelo verdadero adopta en este caso la forma $\ln Y_i = 1 + X_{1i} + \ln X_{2i} + u_i$. Si especificamos y estimamos los seis modelos con los que estamos trabajando, los resultados en cuanto a la proporción de veces que se rechaza $H_0$, se muestran en la Tabla 5.

**Tabla 5. PGD= Modelo mixto (M1).**

| PGD : MIXTO (M5) | | | |
|---|---|---|---|
| **Estimación de un modelo mixto (M5)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.052* | *0.014* | *0.044* |
| **100** | *0.045* | *0.016* | *0.053* |
| **500** | *0.053* | *0.021* | *0.062* |
| **Estimación de un modelo lineal (M1)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.195* | *0.189* | *0.19* |
| **100** | *0.192* | *0.169* | *0.16* |
| **500** | *0.299* | *0.234* | *0.233* |

| Estimación de un modelo doblemente logarítmico (M2) | | | |
|---|---|---|---|
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.055* | *0.049* | *0.049* |
| **100** | *0.049* | *0.066* | *0.057* |
| **500** | *0.059* | *0.073* | *0.069* |

| Estimación de un modelo semilogarítmico 1 (M3) | | | |
|---|---|---|---|
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.009* | *0.009* | *0.015* |
| **100** | *0.007* | *0.013* | *0.017* |
| **500** | *0.041* | *0.054* | *0.067* |

| Estimación de un modelo semilogarítmico 2 (M4) | | | |
|---|---|---|---|
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.166* | *0.173* | *0.121* |
| **100** | *0.182* | *0.181* | *0.114* |
| **500** | *0.394* | *0.316* | *0.177* |

| Estimación de un modelo cuadrático (M6) | | | |
|---|---|---|---|
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.223* | *0.222* | *0.249* |
| **100** | *0.217* | *0.208* | *0.245* |
| **500** | *0.306* | *0.247* | *0.279* |

En este caso las potencias son bajas en todas las estimaciones, incluso en el cuadrático (M6), el lineal (M1) y el semilogarítmico 2 (M4) que son los más diferentes, y hubiésemos esperado por ello, valores más altos.

En general, podemos observar que los resultados se alejan de lo esperado, salvo en el primer caso. Cuando el modelo estimado coincide con el verdadero, el contraste si será capaz de detectar esa igualdad, puesto que minimiza el tamaño de error tipo I, que se encuentra tal como esperábamos en torno al 0.05.

Sin embargo el resto de los modelos especificados y estimados, que no coinciden en ningún caso con el modelo verdadero arrojan resultados en sus potencias muy alejados del valor máximo que buscamos (1). Aunque no deba sorprendernos demasiado para los casos en que los modelos estimados son M2 y M3, puesto que la variable endógena adopta la misma forma que en el PGD ($\ln Y_i$). Por ello, el contraste no es capaz de detectar en absoluto las diferencias, llegado a rechazar la hipótesis nula de forma funcional correcta tan solo en torno al 0.05% de las veces.

Sin embargo, resulta más difícil de explicar por qué el contraste no es capaz de discriminar más modelos tales como el lineal o el cuadrático, con los que tan solo obtenemos potencias inferiores al 0.3 incluyo utilizando muestras grandes, a pesar de las diferencias existentes entre ambos.

El doblemente logarítmico (M2) y el semilogarítmico 1 (M3), son considerados por el contraste como modelos muy similares al modelo verdadero, por lo que el porcentaje de rechazos de la hipótesis nula se aproximan más al nivel de significación que a una potencia alta.

En todos los casos en los que el modelo estimado difiere del verdadero, a medida que utilizamos un tamaño muestral mayor, conseguimos incrementar la potencia, pero ésta sigue siendo insuficiente como para permitirnos concluir un buen funcionamiento del contraste, principalmente en aquellos casos en los que tal y como comentábamos resulta complicado entender el bajo porcentaje de rechazo de $H_0$.

En cuanto a la regresión auxiliar utilizada, vemos que no es necesario utilizar muchos exponentes, puesto que bastará con p=2, es decir tan solo deberíamos incluir la estimación de la variable endógena ($\hat{Y}_i$) elevada al cuadrado.

### 5.2.6 PGD: Modelo cuadrático.

Cuando el modelo verdadero o proceso generador de datos es un modelo cuadrático que adopta la siguiente forma: $Y_i = 1 + X_{1i} + 0.3X_{1i}^2 + X_{2i} + u_i$ y especificamos y estimamos modelos con la misma y distintas formas funcionales, la proporción de veces en la que la hipótesis nula es rechazada, es la que muestra la Tabla 6.

**Tabla 6. PGD = Modelo cuadrático (M6).**

| PGD: CUADRÁTICO (M6) | | | |
|---|---|---|---|
| **Estimación de un modelo cuadrático (M6)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.051* | *0.044* | *0.052* |
| **100** | *0.049* | *0.048* | *0.051* |
| **500** | *0.056* | *0.052* | *0.053* |
| **Estimación de un modelo lineal (M1)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.052* | *0.042* | *0.051* |
| **100** | *0.047* | *0.048* | *0.049* |
| **500** | *0.056* | *0.054* | *0.056* |
| **Estimación de un modelo doblemente logarítmico (M2)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.498* | *0.582* | *0.606* |
| **100** | *0.601* | *0.608* | *0.641* |
| **500** | *0.784* | *0.788* | *0.849* |
| **Estimación de un modelo semilogarítmico (1) (M3)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.825* | *0.791* | *0.74* |
| **100** | *0.984* | *0.98* | *0.977* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo semilogarítmico (2) (M4)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *1* | *1* | *0.998* |
| **100** | *1* | *1* | *1* |
| **500** | *1* | *1* | *1* |
| **Estimación de un modelo mixto (M5)** | | | |
| **Tamaño muestral** | **p=2** | **p=3** | **p=4** |
| **50** | *0.445* | *0.335* | *0.578* |
| **100** | *0.569* | *0.438* | *0.649* |
| **500** | *0.762* | *0.673* | *0.853* |

De nuevo, las tres primeras líneas nos están mostrando si el contraste es capaz de asumir que existe una correcta forma funcional, y al ver que alcanza valores en torno al 0.05, podemos afirmar el buen funcionamiento de Reset cuando ha de determinar que la forma funcional de un modelo corresponde a la del modelo verdadero. Conclusión que podemos extrapolar independientemente de cuál sea el tamaño muestral.

Si el modelo estimado es M4, vamos a alcanzar potencias máximas para todos los tamaños muestrales y regresiones utilizadas salvo si T=50 y p=4, que a pesar de no llegar a 1, prácticamente alcanza dicho valor (0.998). Y no es de extrañar esta discriminación tan elevada, pues se trata de modelos muy diferentes ya que las variables explicativas cambian. Así como si M3 es el modelo estimado, que también alcanzaremos máximas potencias aunque necesariamente con la utilización de muestras grandes.

Cuando la estimación la hacemos a partir de un modelo doblemente logarítmico o mixto, los valores de las potencias no son tan altos como habríamos esperado. De hecho cuando la muestra utilizada es pequeña ni si quiera se rechaza $H_0$ un 50% de las veces. Este aspecto resulta difícil de explicar al considerar que existen grandes diferencias entre el modelo verdadero y los estimados. Además aunque en el resto de los casos convenía la utilización de una regresión auxiliar que utilizase p=2, para la estimación de estos dos últimos modelos los resultados son más claros al utilizar regresiones auxiliares más amplias con p=4. Podemos concluir que cuando el modelo verdadero es M6, y estimamos M2 ó M5 no podemos afirmar que el contraste Reset sea capaz de detectar una errónea forma funcional cuando la muestra utilizada es pequeña.

Al igual que vimos cuando el PGD era lineal y la estimación la hacíamos a partir de un modelo cuadrático, el contraste no es capaz tampoco de detectar las leves diferencias que existen entre el modelo verdadero, en este caso M6 y el modelo estimado M1. Puesto que tan solo el término cuadrático con un coeficiente de 0.3 es lo que diferencia a ambos.

En este caso el contraste los considera tan similares que el porcentaje de rechazo se aproxima al nivel de significación (5%).

Se presenta a continuación, una tabla que resume los resultados obtenidos en el Estudio de Monte Carlo:

**Tabla 7. Resumen de los resultados obtenidos.**

| Modelo estimado | Potencias frente a diferentes modelos verdaderos |
|---|---|
| **M1: Modelo lineal** | M2 alta<br>M3 media-baja<br>M4 muy alta<br>M5 media<br>M6 baja |
| **M2: Modelo doblemente logarítmico** | M1 media<br>M3 alta<br>M4 muy alta<br>M5 muy baja<br>M6 media |
| **M3: Modelo semilogarítmico (logaritmo en la endógena)** | M1 alta<br>M2 muy alta<br>M4 muy alta<br>M5 muy baja<br>M6 alta |
| **M4: Modelo semilogarítmico (logaritmo en las explicativas)** | M1 muy alta<br>M2 alta<br>M3 alta<br>M5 baja<br>M6 muy alta |

| | |
|---|---|
| **M5: Modelo mixto** | M1 media-alta<br>M2 muy baja<br>M3 alta<br>M4 muy alta<br>M6 media |
| **M6: Modelo cuadrático** | M1 baja<br>M2 alta<br>M3 alta<br>M4 muy alta<br>M5 baja |

Se concluye que si estimo un modelo M3 (modelo semilogarítimico 1) un M4 (modelo semilogarítimico 2), el contraste Reset presenta potencias altas si el modelo verdadero es cualquiera de los presentados, excepto M5 (modelo mixto). No obstante, la baja potencia puede deberse a cierta similitud en este caso entre el modelo estimado y el verdadero.

# 6. CONCLUSIONES GENERALES

Con el estudio de Monte Carlo pretendíamos analizar el funcionamiento del contraste Reset en cuanto a tamaño y potencia, prestando especial atención a como se comportaba ante la utilización de diferentes tamaños muestrales (T=50,100 o 500), y al emplear distintas regresiones auxiliares en las que determinábamos p = 2, 3 o 4. Además queríamos comprobar si influía en los resultados el tipo de modelo que actuaba como PGD y los modelos estimados.

Para ello, realizamos 1000 replicas sobre cada uno de los PGD considerados y las estimaciones de los modelos con diferentes formas funcionales. Los resultados mostraban la proporción de veces en la que la hipótesis nula de forma funcional correcta era rechazada por el contraste Reset. Esperábamos por tanto, que cuando el modelo estimado coincidiese con el modelo verdadero, el resultado, que representa el tamaño de error tipo I, se aproximase al nivel de significación que habíamos establecido en el experimento (0.05). Y por el contrario, si el modelo que estimábamos no coincidía con el PGD los resultados debían arrojar valores cercanos a 1, que representa la máxima potencia. Es decir, un buen funcionamiento del contraste Reset implicaría rechazar $H_0$ el menor número de veces posibles (con un margen correspondiente al nivel de significación) si esta es cierta, esto es, si PGD y modelo estimado coinciden; o bien, rechazarla el mayor número de veces si esta es falsa, esto es si el PGD y el modelo estimado difieren.

Un modelo lineal, uno doblemente logarítmico, dos tipos de semilogarítmicos (uno de ellos incluía únicamente la variable endógena en logaritmos, y el otro eran las variables explicativas las que aparecían en logaritmos), un modelo mixto y uno cuadrático, actuaban en cada uno de los casos como PGD. Y se contrastaba la hipótesis nula de forma funcional correcta para cada uno de los modelos anteriormente nombrados que eran especificados y estimados.

De los resultados podemos extraer las siguientes conclusiones:

1- El tamaño del error tipo I, que obtenemos cuando el modelo estimado coincide con el modelo verdadero, se aproxima siempre al nivel de significación establecido en el experimento (0.05). Lo que significa que el contraste Reset actúa de forma adecuada cuando a de detectar una forma funcional correcta.

2- En la mayoría de los casos, hemos podido observar que pasar de un tamaño muestral de 50 observaciones a 100 observaciones supone un aumento importante de la potencia, que no es tal, cuando pasamos de 100 observaciones a 500. Esto significaría que no es necesario utilizar muestras tan grandes. Sin embargo, la excepción se da cuando el PGD es M3, en el que necesariamente, para poder llegar a conclusiones fiables, necesitábamos usar muestras grandes.

3- Al analizar, que regresión auxiliar es más adecuada, hemos llegado a conocer que bastaría con utilizar p = 2, puesto que el comportamiento del contraste Reset es similar a cuando se incluyen valores de p más altos.

4- De la comparación de cada PGD con cada uno de los modelos estimados podemos concluir que si estimamos un modelo semilogarítmo M3 ó M4, es cuando el contraste Reset funciona adecuadamente, ya que en ambos casos las potencias obtenidas son altas, independientemente del modelo verdadero.

## BIBLIOGRAFÍA

GUJARATI, D. (1997): *Basic Econometrics.* Santafé de Bogotá : McGraw-Hill Companies.

KENNEDY, P.E. (2003). *A Guide to Econometrics, Fifth Edition*. MIT Press.

- Definición de Econometría por JUDGE, HILL et al (1982).

MUR, J. (1992). *"Contrastes de autocorrelación espacial. Un estudio de Monte Carlo"* – Estadística Española, vol 34, Núm. 130, págs.285 a 307.

RAMSEY, F.P. (1969). *"Test for specification error and classical leant*-*speak Regression Analysis."* – Journal of the Royal Stadistical Sociality serie B, vol 31, Núm 2, págs. 350 a 371.

SHANON, R.E. (1998). *Introduction to the art and science of simulation*.

TRÍVEZ, F.J. (2004). *Introducción a la Econometría*. Madrid : Pirámide.

WOOLDRIDGE, J.M. (2006). *Introducción a la Econometría- Un enfoque moderno.* Madrid: Paraninfo. (Obra original: Introductory Econometrics A modern approach).