

## **USO DEL ESTIMADOR DE MÍNIMOS CUADRADOS ORDINARIOS EN LA INFERENCIA CON DATOS DE SERIES DE TIEMPO EN MODELOS LINEALES USING ORDINARY LEAST SQUARES ESTIMATOR IN INFERENCE WITH TIME SERIES DATA IN LINEAR MODELS**

**Autores:** Reinaldo Alvarez Carrera

**Institución:** Empresa Agroindustrial Ceballos, Cuba

**Correo electrónico:** esp1\_inform@dcballos.co.cu

### **RESUMEN**

La econometría ha tomado un lugar muy importante para dar solución a múltiples problemas relacionados con la toma de decisiones. Una de sus aplicaciones consiste en descubrir conocimiento en procesos cuyo resultado depende del comportamiento de un grupo de variables que se miden en el tiempo. Para ello se plantean modelos que pueden ser estimados usando estimadores estadísticos, pero es necesario validar tales modelos, como también los resultados de su estimación para que puedan ser usados en la inferencia. Dentro de los estimadores, el de Mínimos Cuadrados Ordinarios tiene propiedades muy favorables cuando se aplica de forma correcta a la muestra seleccionada. Este trabajo tiene como objetivo, proponer una metodología para obtener resultados confiables al realizar inferencias usando el estimador Mínimos Cuadrados Ordinarios en datos de series de tiempo con modelos lineales o modelos que puedan ser transformados a modelos lineales. Aunque esta metodología puede ser aplicada en parte a modelos Box-Jenkins, esta investigación se centra en modelos comunes que especifican variables medidas. Como resultado se presenta un conjunto de pasos ordenados, que se pueden realizar para obtener uno o varios modelos aptos para realizar inferencia.

**Palabras claves:** Econometría, Inferencia estadística, Mínimos Cuadrados Ordinarios, Modelos lineales, Series temporales.

### **ABSTRACT**

Econometrics has taken a very important place to solve multiple problems related to decision making. One of its applications is to discover knowledge in processes whose result depends on the behavior of a group of variables measured over time. To achieve this, models that can be estimated using statistical estimators can be proposed, but it is necessary to validate such models, as also their estimation results in order that they can be used in the inference. Within the estimators, the Ordinary Least Squares (OLS)

has very favorable properties when applied correctly to the selected sample. The objective of this paper is to propose a methodology to obtain reliable results when making inference using the OLS estimator in time series data with linear models or models that can be transformed to linear models. Although this methodology can be applied in part to Box-Jenkins models, this research focuses on common models that specify measured variables. As a result, a set of ordered steps is presented, which can be performed to obtain one or more models suitable for inference.

**Keywords:** Econometrics, Linear models, Ordinary Least Squares, Statistical inference, Time series.

## INTRODUCCIÓN

El desarrollo en el campo de la estadística y la inteligencia artificial ha impulsado el incremento del uso de técnicas de estas ramas científicas para dar soluciones a numerosos y diversos problemas. Entre las técnicas estadísticas destaca la econometría<sup>1</sup> cuya aplicación más común es en el pronóstico de variables macroeconómicas tan importantes como las tasas de interés, de inflación y el producto interno bruto (Wooldridge, 2016).

Mediante el uso de la econometría se pueden plantear modelos econométricos, los cuales (Wooldridge, 2016) describe como una ecuación que relaciona la variable dependiente con un conjunto de variables explicativas y perturbaciones no observadas, donde los parámetros desconocidos de la población determinan el efecto *ceteris paribus*<sup>2</sup> de cada variable explicativa.

Estos modelos necesitan ser procesados con un estimador<sup>3</sup> que permita obtener resultados a partir de los datos de las observaciones correspondientes al problema que se quiere resolver. Entre los estimadores se destaca el de Mínimos Cuadrados Ordinarios (MCO) usado para estimar los parámetros de un modelo de regresión lineal múltiple y sus estimaciones se obtienen mediante la minimización de la suma de los residuales cuadrados (Wooldridge, 2016). Hay numerosas investigaciones en las cuales se ha utilizado el estimador de MCO para solucionar problemas, entre ellas se pueden mencionar (Carrillo, 2004; Vergara, 2006; Bus y Nicolini-Llosa, 2007; Olivera-

---

<sup>1</sup> La econometría se basa en el desarrollo de métodos estadísticos que se utilizan para estimar relaciones económicas, probar teorías económicas y evaluar e implementar políticas públicas y de negocios.

<sup>2</sup> si todos los demás factores relevantes permanecen constantes.

<sup>3</sup> Regla para combinar datos que produce un valor numérico para un parámetro poblacional; la forma de la regla no depende de la muestra particular obtenida.

Chávez et al., 2010; Molero Oliva, 2012; Novoa Hoyos, Sabogal Salamanca y Vargas Walteros, 2016; Salmerón Gómez y Rodríguez Martínez, 2017; Veloz Jaramillo y Cárdenas-Pérez, 2018; Barreto Rodríguez, 2018; Zambrano-Mohauad et al., 2019).

MCO también ha sido aplicado a problemas con datos de series de tiempo. Estos tipos de datos son definidos por (Madrigal Espinoza, 2011) como una sucesión ordenada de valores de una variable (generalmente aleatoria) observados en intervalos de tiempo igualmente espaciados. En trabajos actuales como (Veloz Jaramillo y Cárdenas-Pérez, 2018; Barreto Rodríguez, 2018; Velarde Moreno, Alarcón Osuna y Blanco Jiménez, 2019) se aplicó MCO a datos de series de tiempo.

Es relevante destacar, que a pesar de la existencia de trabajos recientes como (Correa Morales y Patricia Carmona, 2015) donde se comprueba que otros estimadores son más robustos ante la presencia de valores atípicos, (Wooldridge, 2016) expone que cuando MCO se aplica bajo el cumplimiento de los supuestos de Gauss-Markov es el Mejor Estimador Lineal Insesgado (MELI)<sup>4</sup> los coeficientes y sus estadísticos  $t$ ,  $F$  y sus intervalos de confianza son válidos para realizar inferencia estadística<sup>5</sup>. Por esta razón, es importante verificar el cumplimiento de todos los supuestos necesarios más otras comprobaciones adicionales para una interpretación confiable de los resultados. Este trabajo tiene como objetivo plantear una metodología para utilizar de forma precisa el estimador de MCO cuando se quiere realizar inferencia estadística usando datos de series de tiempo con modelos lineales.

Existen varios softwares que permiten el trabajo con series de tiempo, entre ellos se pueden mencionar: Statistical Package for the Social Sciences (SPSS), Statistical Analysis Software (SAS/STAT), R-Studio y Eviews. Por brevedad de esta investigación se ha seleccionado el Eviews (IHS Markit, 2018a), por ser un software de buen prestigio internacional y que está especializado en el trabajo con datos de corte transversal, de series de tiempo y de panel. Los métodos y pruebas citados en esta investigación son los que se encuentran implementados en el software Eviews en su versión 10.

## DESARROLLO

### Recolección de los datos

---

<sup>4</sup> Entre todos los estimadores lineales insesgados, el estimador con la menor varianza. Los estimadores de MCO son MELI, de forma condicional en los valores muestrales de las variables explicativas, bajo los supuestos de Gauss-Markov.

<sup>5</sup> El acto de probar hipótesis acerca de los parámetros poblacionales.

La recolección de datos es una de las etapas iniciales en un proceso de investigación de econometría. Como plantea (Wooldridge, 2016), es importante preguntarse si los datos que se pueden obtener son suficientes como para realizar un análisis *ceteris paribus* convincente. También es necesario que estos datos procedan de fuentes confiables y que comprendan el intervalo de tiempo que se desea analizar.

En series de tiempo es necesario que la frecuencia de los datos obtenidos sea la apropiada para formular el modelo econométrico, o que al menos, a partir de ella se pueda obtener la frecuencia deseada de la forma más confiable posible, en especial cuando se trabaje con datos no experimentales<sup>6</sup>. Por ejemplo, a partir de datos con frecuencia diaria se pueden obtener confiablemente otras frecuencias inferiores como: mensuales, trimestrales, anuales, etc. Pero a partir de frecuencias menores como la anual, no es posible obtener con precisión el comportamiento diario o mensual de cierto proceso no experimental, porque por lo general, el comportamiento de cada sección de tiempo de mayor frecuencia es variable con respecto a las demás secciones de la misma frecuencia.

Para el análisis de series de tiempo es necesario que los datos tengan una frecuencia uniforme (diaria, mensual, anual, etc.). No obstante, existen softwares como (IHS Markit, 2018a) que han implementado la regresión Mixed Data Sampling (MIDAS) (Ghysels, Santa-Clara y Valkanov, 2004; Ghysels, Sinko y Valkanov, 2007; Armesto, Engemann y Owyang, 2010). En este tipo de regresión la variable dependiente puede tener una frecuencia menor que uno o varios de los regresores y se pueden especificar regresores de diferentes frecuencias. Sin embargo, recomiendan tener cuidado cuando se use más de un regresor de alta frecuencia, porque por lo general conduce a problemas de multicolinealidad y en algunos casos de pesos no lineales (non-linear weighting), incrementa dramáticamente la complejidad de la estimación (IHS Markit, 2018b).

### **Preprocesamiento y limpieza de los datos**

Una vez que se han obtenido los datos es necesario aplicar una selección de datos para desechar los innecesarios. También puede ser necesario transformarlos para

---

<sup>6</sup> Datos no son obtenidos por medio de experimentos controlados. Pueden ser datos sobre fenómenos naturales, personas, empresas, etc.

obtener la frecuencia que se necesita. Por ejemplo, obtener datos anuales a partir de datos mensuales.

Teniendo los datos correctos, se procede a la detección de valores atípicos, los cuales (Galeano, 2011) define como observaciones que representan una parte pequeña de los datos y tienen un comportamiento diferente a la mayoría restante de los datos. Agrega también, que se puede ignorar la presencia de estos datos atípicos, pero entonces se puede perjudicar el análisis de múltiples formas, dependiendo del tipo de datos con los que se está trabajando. Para lidiar con este tipo de valores propone tres variantes:

- Suponer un modelo más complejo
- Uso de estimadores robustos
- Uso de procedimientos de detección de atípicos

Adicionalmente, (Galeano, 2011) explica las ventajas y desventajas de estas variantes y en (IHS Markit, 2018b) se presentan los métodos de Mínimos Cuadrados Robustos a valores atípicos que están implementados en Eviews.

Estos valores atípicos deben ser analizados cuidadosamente, pues pueden indicar problemas detectados durante el proceso de análisis o información valiosa para reestructurar el proceso de forma que aprovechen mejor las oportunidades. Por ejemplo, algunos valores de ventas elevados pueden indicar que hubo un tratamiento diferenciado para el producto en determinado momento y que pudo no haberse tomado en cuenta. El análisis de estos valores en la modelación del problema, determinará si se deben incluir o si se deben desechar para evitar los problemas mencionados por (Galeano, 2011).

Eviews implementa 2 mecanismos de detección de valores atípicos cuyos nombres son: Gráficos de apalancamiento (Leverage Plots) y Estadísticos de influencia (Influence Statistics) (IHS Markit, 2018b).

### **Planteamiento del modelo**

Para plantear buenos modelos es importante tener en cuenta toda la información que aporte a la descripción del proceso real y esto puede resultar en la inclusión y modelación de valores atípicos. Es necesario tratar de obtener la mejor representación posible de los factores que influyen en la variable dependiente, en especial cuando se tratan procesos que ocurren en ambientes no controlados, porque son influidos por factores que muchas veces no se pueden identificar o medir correctamente.

## Filtrado

Según (Gujarati y Porter, 2010) una serie de tiempo puede tener cuatro componentes: estacional, cíclico, tendencia y estrictamente aleatorio.

Los mecanismos de filtrado de datos permiten filtrar la serie para extraer información importante que se encuentra oculta en los datos, como es el caso de una tendencia con ciertas características o componentes cíclicos. Esta información puede hacer grandes aportes para modelar el proceso de forma más precisa. Eviews implementa varios mecanismos de filtrado. Entre ellos se pueden mencionar los siguientes:

- Ajuste estacional (elimina los movimientos estacionales cíclicos de una serie y extrae la tendencia)
- Filtro de Hodrick-Prescott (permite obtener un estimado alisado del componente de tendencia a largo plazo de una serie)
- Filtro Paso banda – aislar componentes cíclicos (permite aislar el componente cíclico de una serie especificando un rango para su duración del ciclo)
- Diferenciación (permite eliminar la tendencia aplicando la diferencia de orden 1 o en ocasiones una de orden superior)

## Detección de riesgo de regresión espuria

Un aspecto muy importante a tener en cuenta es comprobar que no exista riesgo de una regresión espuria<sup>7</sup>. Según (Wooldridge, 2016), esto ocurre cuando se efectúa una regresión de una serie  $I(1)$ <sup>8</sup>,  $y_t$ , sobre otra serie  $I(1)$ ,  $x_t$ . Hay un caso en que una regresión que implica variables  $I(1)$  no es espuria, y lo es cuando las series están cointegradas. Esto significa que una función lineal de las dos variables  $I(1)$  es  $I(0)$ <sup>9</sup>. Si  $y_t$  y  $x_t$  son  $I(1)$  pero  $y_t - x_t$  es  $I(0)$ ,  $y_t$  y  $x_t$  no pueden comportarse arbitrariamente para alejarse.

Para comprobar si una serie es  $I(1)$  se puede aplicar una prueba de raíz unitaria. Eviews tiene implementada las siguientes pruebas para comprobar la presencia de raíz unitaria: Augmented Dickey-Fuller (ADF), GLS transformed Dickey-Fuller

---

<sup>7</sup> Problema que surge cuando un análisis de regresión indica una relación entre dos o más procesos de series de tiempo no relacionados, simplemente debido a que cada uno tiene una tendencia, es una serie de tiempo integrada (como una caminata aleatoria), o ambas.

<sup>8</sup> Integrada de orden cero [ $I(0)$ ]: Proceso estacionario, débilmente dependiente de series de tiempo, que cuando se usa en el análisis de regresión, satisface la ley de los grandes números y el teorema del límite central.

<sup>9</sup> Integrada de orden uno [ $I(1)$ ]: Proceso de series de tiempo al que se necesita aplicar la primera diferencia con el fin de producir un proceso  $I(0)$ .

(DFGLS), Phillips-Perron (PP), Kwiatkowski, et al. (KPSS), Elliot, Richardson and Stock (ERS) Point Optimal, y Ng and Perron (NP). Con estas pruebas se puede determinar si la serie es estacionaria, o si lo es en su primera o segunda diferencia.

En caso de que ocurra una regresión que implique variables  $I(1)$  se puede comprobar si es debido a la cointegración y descartar el riesgo de regresión espuria. Para comprobar la existencia de cointegración e series de tiempo Eviews dispone de los test de Johansen, Engle-Granger o Phillips-Outlaris.

### **Detección de multicolinealidad**

En (Wooldridge, 2016) se le llama multicolinealidad a una correlación fuerte (pero no perfecta) entre dos o más variables independientes. También se muestra que la multicolinealidad puede introducir grandes varianzas en los parámetros estimados. Como consecuencia, los valores de los estadísticos  $t$  serán bajos, por lo que puede ocurrir que una regresión arroje un valor de prueba  $F$  significativa y que por otra parte, los estadísticos  $t$  sean insignificantes, lo que dificulta descubrir el efecto parcial debido a la presencia de multicolinealidad. Por esta razón, aunque la multicolinealidad no viola supuesto alguno, es importante tenerla en cuenta cuando se haga inferencia, pues el hecho de conocer si está presente puede ayudar a interpretar mejor los resultados del estimador.

Eviews posee tres mecanismos que puede ser utilizados para detectar multicolinealidad:

- Matriz de correlaciones entre variables. Se toma de ella el nivel de correlación entre variables dependientes.
- El test de Descomposición de la varianza de los coeficientes (Coefficient Variance Decomposition). Un valor de 0.9 puede considerarse alta.
- El test Factor Inflacionario de la Varianza (Variance Inflation Factor). Este test es recomendado por (Wooldridge, 2016) para una mejor percepción del impacto de la multicolinealidad en la varianza de los parámetros estimados. Un valor de 10 en adelante puede considerarse alta.

No obstante, (Wooldridge, 2016) aclara que estos valores no necesariamente garantizan que la multicolinealidad presente esté modificando la varianza de los parámetros estimados, por lo que observar directamente estos valores puede tener una utilidad limitada. También agrega que para corregir la multicolinealidad dado un

conjunto de datos, uno puede tratar de eliminar otras variables independientes del modelo con objeto de reducir la multicolinealidad, pero, eliminar una variable que pertenece al modelo poblacional puede llevar a sesgo. Por otra parte, en problemas donde ciertas variables estén, en teoría económica, muy relacionadas entre ellas, puede atenuarse el efecto de multicolinealidad recolectando más datos, aunque puede ser mejor modificar el enfoque del análisis intentando agruparlas en una sola variable dado que ya no se trataría de estimar el efecto de cada una.

### **Verificación del cumplimiento de los supuestos**

#### **Propiedades en muestras finitas de MCO bajo los supuestos clásicos**

Las propiedades que tiene MCO bajo los supuestos clásicos en muestras finitas que presenta (Wooldridge, 2016) son las siguientes:

##### **Supuesto ST.1 (Linealidad en los parámetros)**

El proceso estocástico  $\{(x_{t1}, x_{t2}, \dots, x_{tk}, y_t): t = 1, 2, \dots, n\}$  sigue el modelo lineal

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t$$

Donde  $\{u_t: t = 1, 2, \dots, n\}$  es una secuencia de errores o perturbaciones. Aquí  $n$  es el número de observaciones (períodos).

##### **Supuesto ST.2 (No hay colinealidad perfecta)**

En la muestra (y, por ende, en los procesos de series de tiempo subyacentes) no hay variables independientes que sean constantes ni que sean una combinación lineal perfecta de las otras.

Es válido destacar que este supuesto permite que las variables explicativas se correlacionen, pero descarta la correlación perfecta en la muestra.

##### **Supuesto ST.3 (Media condicional cero)**

Para cada  $t$ , dadas las variables explicativas para todos los periodos, el valor esperado del error  $u_t$  es cero. Matemáticamente,

$$E(u_t|X) = 0, t = 1, 2, \dots, n$$

Donde  $X$  denota la colección de todas las variables independientes para todos los períodos. Es útil considerar a  $X$  como una matriz, con  $n$  filas y  $k$  columnas, lo que refleja la manera en que se almacenan los datos de series de tiempo en el software de econometría: la  $t$ -ésima fila de  $X$  es  $x_t$ , que consta de todas las variables independientes en el periodo  $t$ .



Este supuesto implica que el error en el periodo  $t$ ,  $u_t$ , no se correlaciona con ninguna variable explicativa en *cada* uno de los períodos.

#### **Supuesto ST.4 (Homocedasticidad)**

La varianza de  $u_t$  condicional en  $X$ , es la misma para cualquier  $t$ :  $Var(u_t|X) = Var(u_t) = \sigma^2, t = 1, 2, \dots, n$ .

Este supuesto significa que  $Var(u_t|X)$  no puede depender de  $X$  (es suficiente con que  $u_t$  y  $X$  sean independientes) además que  $Var(u_t)$  debe ser constante en el tiempo. Cuando el supuesto ST.4 no es válido, se dice que los errores son heterocedásticos.

#### **Supuesto ST.5 (No hay correlación serial)**

Los errores, condicionales sobre  $X$ , en dos períodos distintos, no están correlacionados:  $Corr(u_t, u_s|X) = 0$ , para cualquier  $t \neq s$ .

La forma más sencilla de considerar este supuesto es ignorar el condicionamiento sobre  $X$ . Así, el supuesto ST.5 es sencillamente

$Corr(u_t, u_s) = 0$ , para cualquier  $t \neq s$ .

Cuando lo anterior no se cumple, se dice que los errores en el supuesto ST.1 tienen correlación serial o autocorrelación, debido a que se correlacionan a lo largo del tiempo.

#### **Supuesto ST.6 (Normalidad)**

Los errores  $u_t$  son independientes de  $X$  y son independientes e idénticamente distribuidos como  $Normal(0, \sigma^2)$ .

El supuesto ST.6 comprende los supuestos ST.3, ST.4 y ST.5, pero es más fuerte debido a los supuestos de independencia y normalidad.

De estos supuestos se concluye que:

- Bajo los supuestos ST.1, ST.2 y ST.3, los estimadores de MCO son insesgados condicionales sobre  $X$ .
- Bajo los supuestos ST.1 a ST.5, los estimadores de MCO son los mejores estimadores lineales insesgados condicionales sobre  $X$ .
- Cuando los supuestos ST.1 a ST.6 son válidos, los estadísticos  $t$  pueden emplearse para probar la significancia estadística de las variables explicativas individuales y los estadísticos  $F$  pueden utilizarse para probar la significancia conjunta. También es válida la construcción usual de los intervalos de confianza y se puede realizar inferencia.

Al verificar el cumplimiento de estos supuestos, puede encontrarse la violación de algunos de ellos. A continuación, se presentan algunas soluciones ante la violación de los supuestos:

**Violación del supuesto ST.1:** Se detecta a simple vista, cuando los coeficientes (parámetros) se encuentran como exponentes en la ecuación. Para solucionarlo (Gujarati y Porter, 2010) muestran que se puede realizar una transformación logarítmica a la ecuación para hacer que el modelo sea lineal en los parámetros, pero no siempre es posible encontrar una transformación logarítmica apropiada.

**Violación del supuesto ST.2:** El Eviews tiene un mecanismo automático para detectar esta anomalía y muestra un mensaje de alerta. Para dar solución (Wooldridge, 2016) elimina una de las variables independientes de la ecuación para verificar si se soluciona el problema.

**Violación del supuesto ST.3:** La detección se puede realizar mediante la comprobación de las correlaciones entre las variables de la ecuación y el término de error. Según (Wooldridge, 2016) Si  $u_t$  es independiente de  $X$  y  $E(u_t) = 0$ , el supuesto ST.3 es válido de manera automática. Por tanto, es suficiente con chequear la matriz de correlaciones que computa el Eviews para las variables independientes y el término de error de la regresión y comprobar que este último no esté correlacionado con las variables.

**Violación del supuesto ST.4:** Para este caso (Wooldridge, 2016) recomienda probar primero la correlación serial, con ayuda de una prueba robusta a la heterocedasticidad, si se sospecha de la existencia de ésta. Una vez que se han realizado las acciones pertinentes para corregir la correlación serial, se pone a prueba la heterocedasticidad. Eviews brinda 5 pruebas para comprobar la existencia de heterocedasticidad: Breusch-Pagan-Godfrey (BPG), Harvey, Glejser, ML para Heterocedasticidad Condicional Autorregresiva (ARCH) y White. Para dar solución a este problema, se puede reestimar la ecuación usando el estimador de matriz de covarianzas robusto a la heterocedasticidad White o el de Newey-West implementados en Eviews.

**Violación del supuesto ST.5:** Se puede detectar usando los Correlogramas y estadísticos Q o la prueba de correlación serial basada en Multiplicadores de

Lagrange. Para corregirla se puede diferenciar<sup>10</sup> la serie o usar un estimador de matriz de covarianzas robusto a la autocorrelación como el Newey-West en Eviews.

**Violación del supuesto ST.6:** Este es el supuesto más restrictivo, se puede detectar su violación aplicando la prueba de Jarque-Bera a los residuales de la regresión. En caso de violación hay que recurrir al uso de los supuestos para muestras grandes con el objetivo de justificar la normalidad de forma asintótica. Para esto es necesario contar con suficientes observaciones. Según (Wooldridge, 2016) algunos econométristas piensan que  $n = 30$  es satisfactorio, pero esto puede no ser suficiente para todas las distribuciones de  $u$ . Una solución es adquirir más datos o usar datos de mayor frecuencia y remodelar el problema para validarlo bajo los supuestos asintóticos presentados a continuación.

### **Propiedades en muestras grandes de MCO bajo los supuestos asintóticos**

#### **Supuesto ST.1' (Linealidad y dependencia débil)**

El proceso estocástico  $\{(x_{t1}, x_{t2}, \dots, x_{tk}, y_t): t = 1, 2, \dots, n\}$  sigue el modelo lineal

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t$$

donde  $\{u_t: t = 1, 2, \dots, n\}$  es la secuencia de errores o perturbaciones. Aquí,  $n$  es el número de observaciones (periodos).

#### **Supuesto ST.2' (No hay colinealidad perfecta)**

En la muestra (y por tanto en los procesos de series de tiempo subyacentes), ninguna variable dependiente es constante ni es una combinación perfecta de las otras.

#### **Supuesto ST.3' (Media condicional cero)**

Las variables explicativas son *contemporáneamente exógenas*, es decir,  $E(u_t | x_{t1}, x_{t2}, \dots, x_{tk}) = 0$ . Hay que tener en cuenta que ST.3' es notablemente más débil que el supuesto de exogeneidad estricta ST.3'.

#### **Teorema Consistencia de MCO**

Bajo los supuestos ST.1', ST.2' y ST.3', los estimadores de MCO son consistentes:

$$plim \hat{\beta}_j = \beta_j, j = 0, 1, \dots, k.$$

Bajo los supuestos ST.1' a ST.3' los estimadores de MCO son consistentes, pero no necesariamente insesgados.

---

<sup>10</sup> Aplicar diferencia de primer orden o de orden superior.

Para usar procedimientos de inferencia estándar, es necesario imponer versiones de los supuestos de homocedasticidad y errores sin correlación serial.

#### **Supuesto ST.4' (Homocedasticidad)**

Los errores son *contemporáneamente homocedásticos*, es decir,  $Var(u_t|X_t) = \sigma^2$ , donde  $X_t$  es la abreviatura de  $(x_{t1}, x_{t2}, \dots, x_{tk})$ .

#### **Supuesto ST.5' (No hay correlación serial)**

Para toda  $t \neq s$ ,  $E(u_t, u_s|X_t, X_s) = 0$ .

Del cumplimiento de estos supuestos se concluye que:

Bajo los supuestos ST.1' a ST.5', los estimadores de MCO tienen distribuciones asintóticamente normales. Además, los errores estándares usuales de MCO, los estadísticos  $t$ , los estadísticos  $F$  y los estadísticos de multiplicador de Lagrange ( $ML$ ) son asintóticamente válidos.

Los supuestos ST.4' y ST.5' implican que es posible usar los intervalos de confianza usuales, los estadísticos  $t$  y los estadísticos  $F$  como si fueran aproximadamente válidos en las muestras grandes.

La forma en que verifica si hay violación de estos supuestos y se le da solución, es la misma que para los supuestos ST.1 – ST.5 (muestras finitas).

Una vez que se ha verificado el cumplimiento de los supuestos, es necesario realizar otro grupo de verificaciones importantes:

#### **Valores esperados**

La magnitud de los coeficientes debe ser verificada, pues una variable no debe mostrar una influencia mayor o menor que la esperada y que no tenga una justificación correcta. Tampoco las variables deben mostrar un signo contrario al esperado, pues hay variables que influyen de forma positiva y otras de forma negativa. Aunque existen variables que pueden influir de forma positiva durante una etapa y durante otra etapa pueden hacerlo de forma negativa.

#### **Significación individual de los coeficientes**

La significación individual de los coeficientes indica que la variable analizada es importante en la ecuación. Esta significación está dada por el estadístico  $t$  y a un nivel de significación de un 5%, entonces para valores inferiores a 0.05, no hay evidencia para rechazar la hipótesis nula de no significación, y se puede considerar mantener tal variable en la ecuación.

### **Significación conjunta de los coeficientes**

La significación conjunta de los coeficientes indica que la combinación de todas las variables explica el modelo planteado. Esta significación está dada por el estadístico  $F$  y a un nivel de significación de un 5%, para valores inferiores a 0.05, no hay evidencia para rechazar la hipótesis nula de no significación.

### **Comparación entre modelos**

Si se han obtenido más de un modelo para representar el problema, se puede escoger entre ellos usando los siguientes criterios:

- Criterio de expertos. El modelo que mejor se ajusta a la realidad del problema.
- El valor de  $R^2$  y  $R^2$  ajustada, los cuales indican el nivel de precisión con que la combinación de variables describe el modelo.
- Criterio de expertos calculado por EViews como resultados de la estimación. La salida muestra tres valores calculados para los criterios de Akaike, Schwarz y Hannan-Quinn. Valores pequeños de estos criterios indican un buen modelo. Los criterios segundo y tercero aplican una función de penalización para coeficientes innecesarios, ayudando a verificar si se tiene una especificación parsimoniosa.

### **CONCLUSIONES**

La metodología planteada permitió comprobar un grupo de requisitos necesarios para aplicar correctamente el estimador de MCO con el objetivo de realizar inferencia. Sin la comprobación de estos requisitos, la salida del estimador de MCO puede ser errada en la descripción de las relaciones entre variables de la ecuación y por ende no ser apropiado para la inferencia. Por otra parte, se explicó cómo validar el modelo obtenido una vez que se aplique la estimación de forma correcta. Finalmente, se concluyó que la aplicación de este conjunto de pasos permite aprobar el modelo resultante de la estimación con MCO como válido para realizar inferencia con datos de series de tiempo usando modelos lineales.

### **REFERENCIAS BIBLIOGRÁFICAS**

- ARMESTO, M.T., ENGEMANN, K.M. y OWYANG, M.T., 2010. "Forecasting with Mixed Frequencies". *Federal Reserve Bank of St. Louis Review*, vol. 92, no. 6, pp. 521-536.
- BARRETO RODRÍGUEZ, C.R., 2018. "MODELO DE REGRESIÓN DEL CONSUMO

- FINAL PRIVADO Y EL PRODUCTO BRUTO INTERNO NACIONAL DEL PERÚ DURANTE EL PERIODO 2000 - 2013". *In Crescendo*, vol. 9, no. 1, pp. 13-32.
- BUS, A.G. y NICOLINI-LLOSA, J.L., 2007. "Importaciones de Argentina, una estimación econométrica". Buenos Aires, Argentina.
- CARRILLO, P., 2004. "Las diferencias salariales entre el sector público y privado en el Ecuador". *Cuestiones Económicas*, vol. 20, no. 2, pp. 165-174.
- CORREA MORALES, J.C. y PATRICIA CARMONA, G., 2015. "Comparación de la regresión GINI con la regresión de Mínimos cuadrados ordinarios y otros modelos de regresión lineal robustos". *Comunicaciones en Estadística*, vol. 8, no. 2, pp. 129-161.
- GALEANO, P., 2011. "Aspectos generales sobre la presencia de datos atípicos en series temporales". *CAEPIA 2011 - I Workshop on Time Series*.
- GHYSELS, E., SANTA-CLARA, P. y VALKANOV, R., 2004. "The MIDAS Touch: Mixed Data Sampling Regression Models". [en línea]. S.l.: Disponible en: <https://escholarship.org/uc/item/9mf223rs>.
- GHYSELS, E., SINKO, A. y VALKANOV, R., 2007. "MIDAS Regressions: Further Results and New Directions". *Econometric Reviews*, vol. 26, no. 1, pp. 53-90.
- GUJARATI, D.N. y PORTER, D.C., 2010. *Econometría Quinta Edición*.
- IHS MARKIT, 2018a. *Eviews 10*. 2018.
- IHS MARKIT, 2018b. *Eviews 10 User's Guide II* [en línea]. Irvine, CA. ISBN 978-1-880411-44-5 (2nd Edition). Disponible en: <http://www.eviews.com>.
- MADRIGAL ESPINOZA, S.D., 2011. *Pronóstico de series temporales con estacionalidad*. Universidad Autónoma de Nuevo León.
- MOLERO OLIVA, L.E., 2012. "Estimación de la Ley de Okun para la economía venezolana. Período 1999-2009". *Revista de Ciencias Sociales*, vol. XVIII, no. 2, pp. 311-324.
- NOVOA HOYOS, A., SABOGAL SALAMANCA, M. y VARGAS WALTEROS, C., 2016. "Estimación de las relaciones entre la inversión en medios digitales y las variables financieras de la empresa: una aproximación para Colombia". *Revista EAN*, no. 80, pp. 12-25. ISSN 0120-8160. DOI 10.21158/01208160.n80.2016.1313.
- OLIVERA-CHÁVEZ, R.I., CERMEÑO-BAZÁN, R., SÁENZ DE MIERA-JUÁREZ, B., JIMÉNEZ-RUIZ, J.A. y REYNALES-SHIGEMATSU, L.M., 2010. "El efecto del

precio del tabaco sobre el consumo: Un análisis de datos agregados para México". *Salud Publica de Mexico*, vol. 52, no. 2, pp. 197-205. ISSN 00363634. DOI 10.1590/s0036-36342010000800015.

- SALMERÓN GÓMEZ, R. y RODRÍGUEZ MARTÍNEZ, E., 2017. "Métodos cuantitativos para un modelo de regresión lineal con multicolinealidad. Aplicación a rendimientos de letras del tesoro". *REVISTA DE MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA* [en línea], vol. 24, pp. 169-189. Disponible en:  
[www.upo.es/revistas/index.php/RevMetCuant/article/view/2886](http://www.upo.es/revistas/index.php/RevMetCuant/article/view/2886).
- VELARDE MORENO, O.E., ALARCÓN OSUNA, M.A. y BLANCO JIMÉNEZ, M., 2019. "Análisis del contenido nacional de las exportaciones de manufacturas mexicanas: el caso de la Alianza del Pacífico 2011-2016". *México y la Cuenca del Pacífico*, vol. 8, no. 24, pp. 57-77. DOI 10.32870/mycp.v8i24.595.
- VELOZ JARAMILLO, M. y CÁRDENAS-PÉREZ, A., 2018. "MODELIZACIÓN ECONOMETRICA BAJO LA METODOLOGÍA DE BOX-JENKINS. ESTUDIO EMPÍRICO A LA LIQUIDEZ DEL SISTEMA FINANCIERO ECUATORIANO". *REVISTA INVESTIGACION OPERACIONAL*, vol. 39, no. 4, pp. 592-606.
- VERGARA, M., 2006. "Nota técnica para estimar fronteras estocásticas: una aplicación a la banca chilena". Chile.
- WOOLDRIDGE, J.M., 2016. *Introductory Econometrics: A Modern Approach. Sixth Edition*. 6th. Michigan: Cengage Learning.
- ZAMBRANO-MOHUAD, G.A., RAMÍREZ-ALFONSO, K.R., MENDOZA-VÉLEZ, M.E., PAMBABAY-CALERO, J.J., BAUZ-OLVERA, S.A. y NIETO-WIGBY, J.R., 2019. "Modelización Econométrica mediante mínimos cuadrados ordinarios para la detección de factores determinantes del salario en la ciudad de Cuenca-Ecuador". *Matemática ESPOL-FCNM*, vol. 19, no. 1, pp. 1-16.