

A close-up photograph of two individuals in dark business suits shaking hands. The background is blurred, suggesting an office or professional setting. The image is partially obscured by a semi-transparent dark blue horizontal band.

NOCIONES DE ECONOMETRIA INTERMEDIA

AUTOR: Jorge Dresdner Cid
Felipe Vásquez Lavin



2007 NOCIONES DE ECONOMETRIA INTERMEDIA.
Universidad de Concepción

Registro Propiedad Intelectual N° 136.998

I.S.B.N. 956-8029-48-6

Segunda Edición Noviembre 2007

Impresión:
Talleres Dirección de Docencia
Edmundo Larenas 64-A
Barrio Universitario
Concepción

IMPRESO EN CHILE / PRINTED IN CHILE

NOCIONES DE ECONOMETRIA INTERMEDIA

JORGE DRESDNER CID Y FELIPE VASQUEZ LAVIN

Junio 2003

Índice general

1. ALGEBRA MATRICIAL	1
1.1. Introducción	1
1.2. Matrices	2
1.2.1. Tipos de Matrices	3
1.2.2. Transposición	4
1.3. Operaciones de Matrices	5
1.3.1. Igualdad de Matrices	5
1.3.2. Adición - Sustracción	5
1.3.3. Multiplicación por Escalar	6
1.3.4. Multiplicación de Matrices	7
1.3.5. Producto Kronecker	10
1.3.6. Traza de una matriz	12
1.4. Determinante de una Matriz	13
1.4.1. Menor de una Matriz	14
1.4.2. Cofactor de una Matriz	15
1.4.3. Matriz de Cofactores	16
1.4.4. Matriz Adjunta	16
1.4.5. Método de Cofactores para el Determinante	16
1.4.6. Propiedades del Determinante.	17
1.5. Rango de una Matriz	18
1.6. Matriz Inversa	19
1.6.1. Propiedades de la Matriz Inversa	21
1.7. Diferenciación de Matrices	21
2. ESTADÍSTICA, PROBABILIDAD E INFERENCIA	25
2.1. Introducción: Definiciones	25
2.2. Momentos de la Distribución	31
2.3. Distribuciones de Probabilidad	34

2.3.1.	Distribución Normal	35
2.3.2.	Distribución Chi - Cuadrado	36
2.3.3.	Distribución t de Student	37
2.3.4.	Distribución F de Fisher:	38
2.4.	Inferencia y Estimación	38
2.4.1.	El problema de la estimación	40
2.4.2.	Métodos de Estimación	41
2.4.3.	Propiedades Deseadas de los Estimadores.	45
2.5.	Intervalos de Confianza y test de Hipótesis	54
3.	MODELO DE REGRESIÓN LÍNEAL GENERAL	61
3.1.	Introducción	61
3.2.	Modelo Clásico	62
3.3.	Supuestos del Modelo Clásico	64
3.4.	Mínimos Cuadrados Ordinarios	69
3.4.1.	Propiedades del Estimador MCO	76
3.4.2.	Estimador de la Varianza del Error	81
3.5.	Estimador Máximo Verosímil	85
3.6.	Estimación en Desviaciones de Media	87
3.7.	Criterios de Bondad del Ajuste	92
3.8.	Inferencia	98
3.8.1.	Pruebas de Hipótesis e Intervalos de Confianza.	98
3.8.2.	Test T Generalizado	104
3.8.3.	Prueba F General.	106
3.8.4.	Predicción.	114
4.	MINIMOS CUADRADOS GENERALIZADOS	121
4.1.	Introducción	121
4.2.	Mínimos Cuadrados Generalizados	122
4.3.	Estimador Máximo Verosímil	128
4.4.	Heterocedasticidad	130
4.4.1.	Detección de la Heterocedasticidad.	131
4.4.2.	Solución de la Heterocedasticidad.	139
4.5.	Autocorrelación	144
4.5.1.	Detección de Autocorrelación.	148
4.5.2.	Solución de la autocorrelación.	152

5. TOPICOS ADICIONALES	159
5.1. Multicolinealidad	160
5.1.1. Cómo Detectar la Multicolinealidad.	164
5.1.2. Cómo Solucionar la Multicolinealidad.	166
5.2. Prueba de Cambios Estructurales	169
5.3. Variables Dicótomicas	176
A. Ejercicios Complementarios	183
A.1. Estimación de Función de Precios y de Producción	183
A.1.1. Función de Precios	183
A.1.2. Función de Producción	190
A.2. Instrucciones para el programa E-Views	198
A.3. Tablas de Datos	200

Prólogo

El presente texto está dirigido a un público amplio, interesado en inferencia estadística en economía. La idea del texto ha nacido de una larga experiencia de enseñanza de econometría básica en la carrera de Ingeniería Comercial de la Universidad de Concepción. Los contenidos corresponden aproximadamente a lo que se cubre en un semestre de estudios parciales o cinco semanas de estudios intensivos en el primer nivel. En este sentido, el libro no presenta prerequisites en econometría. Eso sí, para un uso más óptimo del texto, es recomendable tener algunos conocimientos previos de economía básica y matemáticas.

La econometría se ha transformado en una herramienta indispensable para profesionales de distintas disciplinas que enfrentan la tarea de trabajar con datos económicos. El objetivo de este texto es presentar en forma comprimida, pero amigable, los rudimentos de econometría que debe tener cualquier profesional que se quiera desempeñar bien en un área de trabajo que contenga requerimientos de procesar y analizar datos económicos.

En el mercado existen distintos textos que cubren el material tratado en este libro, algunos con un nivel muy básico y otros de nivel muy avanzado. La necesidad de un texto de este tipo surge del convencimiento que falta un texto de nivel intermedio, que permita abordar en forma más general y con un grado de profundidad mayor, aunque sin llegar a un nivel de postgrado, los métodos econométricos básicos. Además, el presente texto está orientado a aplicar los conocimientos adquiridos en Econometría a la realidad nacional. Para ello, los ejemplos y aplicaciones se desarrollan con series de datos nacionales. Además, en el apéndice se presentan series de datos adicionales con los cuales se pueden realizar ejercicios no desarrollados en el texto. Ello no sólo permite al lector realizar ejercicios que relacionen el acontecer y la coyuntura nacional, con los métodos aprendidos, y eventualmente con la teoría económica, sino

que además introduce al lector en el mundo de las estadísticas disponibles en el país. El trabajo de Econometría no solo implica realizar regresiones y pruebas con los datos, sino que además implica recolectar, procesar y ordenar los datos. Esto último depende de la disponibilidad, accesibilidad y calidad de los datos disponibles. Al trabajar con series de datos nacionales, creemos que incentivamos al lector a completar las series, a buscar series para variables adicionales, a penetrar en el mundo de las instituciones que generan datos económicos en Chile. Todo esto es parte del trabajo de un economista.

Queremos reconocer la excelente y abnegada labor de asistencia realizada por Vanessa Maynou. Su contribución ha sido muy importante para poder completar el presente texto. También estamos en deuda con Romina Villaruel, cuyo trabajo y apoyo fue muy importante en las últimas etapas de edición de este libro. Finalmente, queremos reconocer el invaluable aporte a la generación de la idea de escribir este libro y a su gestación de nuestros alumnos, que a través de los años y sin saberlo, fueron diseñando la forma final que éste tomó.

JORGE DRESDNER C.

FELIPE VÁSQUEZ L.

Capítulo 1

ALGEBRA MATRICIAL

1.1. Introducción

El objetivo de este capítulo es entregar algunas nociones de álgebra matricial que son necesarias para la comprensión de los modelos y técnicas de estimación econométricas desarrollados a partir del capítulo 3. Se comienza con la definición de matrices y con una clasificación de éstas, para posteriormente estudiar las operaciones matriciales básicas, entre las cuales se mencionan la adición, la multiplicación, los determinantes y el Producto Kronecker. Otras herramientas que serán discutidas son la obtención de recíprocos (inversos multiplicativos) y sus respectivas propiedades, para finalizar con una breve introducción a la diferenciación matricial.

Las matrices son convenientes porque permiten expresar en forma ordenada y compacta una variedad de tópicos relevantes desde la perspectiva económica. Por ejemplo, se puede representar en forma matricial la información de múltiples series con numerosas observaciones de datos económicos, los cuales sirven para estimar diversos modelos teóricos y contrastarlos con la evidencia empírica. Así mismo, modelos teóricos con gran cantidad de relaciones se expresan en forma sencilla al trabajar con matrices. Por último, los resultados de los diversos problemas de estimación planteados son entregados en una forma ordenada y de fácil interpretación. Estas ventajas serán claras para el lector en la medida que profundice en el análisis de los datos económicos.

Los tópicos presentados en este capítulo no pretenden ser una revisión completa del álgebra matricial, sino sólo entregar los elementos necesarios

para comprender los procedimientos matriciales utilizados en estimaciones econométricas básicas. Para una profundización del álgebra matricial se recomienda literatura como Devaud, et al. (1991), Herstein y Winter (1989) y Jacob (1995)

1.2. Matrices

Las matrices se definen como un ordenamiento o arreglo rectangular de elementos dispuestos en filas y columnas. La cantidad de filas y columnas que posea una matriz determinará el orden de ésta. Así una matriz A que posea m filas y n columnas, tendrá orden $m \times n$ y la matriz se denotará como: $A_{m \times n}$. Su representación gráfica es

$$\mathbf{A}_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

donde a_{ij} denota al elemento de la matriz que está en la intersección de la i -ésima fila y la j -ésima columna. Veamos el caso de una matriz A de orden 2×3 , y de una matriz B de orden 2×2 :

$$\mathbf{A}_{2 \times 3} = \begin{bmatrix} 3 & -1 & 7 \\ 0 & 5 & 4 \end{bmatrix} \quad \mathbf{B}_{2 \times 2} = \begin{bmatrix} 2 & -5 \\ 6 & 4 \end{bmatrix},$$

donde -1 es el elemento a_{12} , y 4 es el elemento b_{22} .

En economía interesa muchas veces relacionar conjunto de datos. Por ejemplo, los datos anuales del **Producto Interno Bruto** (en millones de pesos de 1986) y el **Saldo de la Balanza Comercial** (al 31 de diciembre de cada año en millones de dólares) de la economía chilena entre los años 1991 y 1996 se presentan a continuación como la matriz \mathbf{X} :

$$\mathbf{X} = \begin{bmatrix} 4,841,447 & 1,485,1 \\ 5,435,881 & 722,0 \\ 5,815,646 & -989,8 \\ 6,147,610 & 732,0 \\ 6,800,952 & 1,369,1 \\ 7,305,141 & -1,095,0 \end{bmatrix}$$

El orden de esta matriz es de 6×2 . El elemento x_{21} indica que el Producto Interno Bruto fue superior a los 5 billones de pesos en 1992 y x_{62} que el Déficit en Balanza Comercial fue de casi 1.100 millones de dólares de 1996.

1.2.1. Tipos de Matrices

Casos especiales de matrices son el **vector columna** y el **vector fila**. El primero es una matriz que posee sólo una columna y n filas. Análogamente el vector fila posee sólo una fila y n columnas. Por ejemplo:

$$\mathbf{a} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} \quad \mathbf{a}' = [a_{11} \quad a_{12} \quad \cdots \quad a_{1n}],$$

En esta situación \mathbf{a} es un vector columna de orden $n \times 1$ y \mathbf{a}' es un vector fila de orden $1 \times n$. En general, esta es la forma de escribir los vectores usada en la literatura, por lo que de aquí en adelante se entenderá que un vector \mathbf{a} denota un vector columna, mientras que \mathbf{a}' un vector fila.

Otro caso particular de matrices que es de interés, son las matrices conocidas como **matrices cuadradas**, cuya característica es que posee igual número de columnas y filas. En este caso se dirá que la matriz es cuadrada de orden n con n columnas o filas. Un elemento importante de la matriz cuadrada es la **diagonal principal**, que es la diagonal formada por todos los elementos a_{ii} , para $i = 1, \dots, n$. Es decir, para esta matriz, la diagonal principal está formada por los elementos $a_{11}, a_{22}, \dots, a_{nn}$.

A partir de la diagonal principal se pueden definir geoméricamente las siguientes matrices:

1. **Matriz Diagonal:** Es aquella matriz que posee sólo los elementos de la diagonal principal distintos de cero.

2. **Matriz Triangular Superior:** Matriz en que todos los elementos bajo la diagonal principal son cero.

3. **Matriz Triangular Inferior:** Análogo a la matriz anterior, pero donde ahora los elementos sobre la diagonal principal son cero.

4. **Matriz Escalar:** Es una matriz diagonal, con la característica que todos los elementos de la diagonal principal son iguales.

5. **Matriz Identidad:** Es una matriz escalar cuyos elementos son 1. Esta siempre se denota por \mathbf{I} , o \mathbf{I}_n cuando se desea especificar su dimensión.

Por último, otra clase de matriz es la **matriz simétrica**, en la cual los elementos sobre la diagonal principal son el reflejo de los elementos bajo la diagonal principal. En otras palabras, el elemento a_{ij} es igual al elemento a_{ji} , para $i, j = 1 \dots n$. Por ejemplo, veamos la siguiente matriz simétrica:

$$\mathbf{A}_{3 \times 3} = \begin{bmatrix} 5 & 4 & -9 \\ 4 & 0 & 3 \\ -9 & 3 & 1 \end{bmatrix}$$

En esta matriz se puede observar que el elemento $a_{23}(3)$ es igual al elemento a_{32} , así como el elemento $a_{12}(4)$ es igual al elemento a_{21} . También se observa que los elementos de la diagonal principal son 5, 0 y 1, respectivamente.

1.2.2. Transposición

La **transpuesta** de una matriz \mathbf{A} , denotado por \mathbf{A}' , se obtiene mediante el intercambio de filas y columnas, es decir, las filas se convierten en columnas y viceversa, manteniendo su orden.

Si consideramos la siguiente matriz \mathbf{A} , al transponerla queda:

$$\mathbf{A}_{3 \times 3} = \begin{bmatrix} 5 & 4 & -9 \\ -11 & 0 & 3 \\ 2 & 8 & 1 \end{bmatrix} \quad \mathbf{A}'_{3 \times 3} = \begin{bmatrix} 5 & -11 & 2 \\ 4 & 0 & 8 \\ -9 & 3 & 1 \end{bmatrix}$$

En el caso particular de los vectores filas y vectores columnas, al transponerlos se convierten en vectores columnas y filas, respectivamente. Recuerde que escribimos \mathbf{a} como vector columna y \mathbf{a}' como vector fila. Por su parte, un escalar, que es una matriz de orden 1×1 , al transponerlo queda el mismo escalar.

Existen, además, matrices que son simétricas, las cuales al ser transpuestas quedan inalteradas. Es decir, si \mathbf{A} es una matriz simétrica, entonces:

$$\mathbf{A}' = \mathbf{A}$$

En la siguiente matriz, podemos observar que al transponer las filas y las columnas, obtenemos la misma matriz.

$$\mathbf{A}_4 = \begin{bmatrix} 3 & 5 & 23 & 4 \\ 5 & 1 & 9 & 7 \\ 23 & 9 & 0 & 5 \\ 4 & 7 & 5 & 1 \end{bmatrix} \Rightarrow \mathbf{A}'_4 = \begin{bmatrix} 3 & 5 & 23 & 4 \\ 5 & 1 & 9 & 7 \\ 23 & 9 & 0 & 5 \\ 4 & 7 & 5 & 1 \end{bmatrix}$$

Un caso especial de matrices simétricas, es la matriz identidad \mathbf{I}_n , la cual es de gran utilidad en econometría.

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \Rightarrow \mathbf{I}'_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Obviamente $(\mathbf{A}')' = \mathbf{A}$, es decir, al transponer una matriz transpuesta recobramos la matriz inicial.

1.3. Operaciones de Matrices

1.3.1. Igualdad de Matrices

Para que dos matrices, \mathbf{A} y \mathbf{B} , sean iguales, éstas deben tener el mismo orden o dimensión y sus elementos correspondientes deben ser idénticos, es decir $a_{ij} = b_{ij}, \forall i, j$.

1.3.2. Adición - Sustracción

La condición necesaria para sumar dos matrices \mathbf{A} y \mathbf{B} , es que ambas tengan el mismo orden. Así se puede definir la adición (sustracción) como:

$$\mathbf{A}_{m \times n} + (-) \mathbf{B}_{m \times n} = \mathbf{C}_{m \times n}$$

De tal manera que cada elemento de la matriz \mathbf{C} se obtiene de la siguiente forma:

$$c_{ij} = a_{ij} + (-) b_{ij}$$

En el siguiente ejemplo se suma la matriz \mathbf{A} de orden 2×3 con la matriz \mathbf{B} de la misma dimensión.

$$\begin{bmatrix} 5 & -8 & 1 \\ 9 & 0 & -13 \end{bmatrix} + \begin{bmatrix} -4 & 3 & 2 \\ 7 & -6 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -5 & 3 \\ 16 & -6 & -14 \end{bmatrix}$$

Propiedades de la Adición de Matrices

1. La adición de matrices es conmutativa, es decir:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

2. La suma (resta) de una matriz \mathbf{A} , con la matriz nula $\boldsymbol{\theta}$ (matriz que posee todos los elementos iguales a cero), es la matriz \mathbf{A} , es decir:

$$\mathbf{A}_{m \times n} + \boldsymbol{\theta}_{m \times n} = \mathbf{A}_{m \times n}$$

3. La adición (sustracción) de matrices es asociativa, es decir:

$$[\mathbf{A} + (-) \mathbf{B}] + (-) \mathbf{C} = \mathbf{A} + (-) [\mathbf{B} + (-) \mathbf{C}]$$

4. La transpuesta de la suma (resta) de dos matrices es igual a la suma (resta) de las matrices transpuestas, es decir:

$$(\mathbf{A} + (-) \mathbf{B})' = \mathbf{A}' + (-) \mathbf{B}'$$

1.3.3. Multiplicación por Escalar

Para multiplicar una matriz cualquiera de orden $m \times n$, por un escalar cualquiera λ , se multiplica cada uno de los elementos de la matriz por el escalar, de tal forma que queda:

$$\begin{aligned} \lambda \cdot \mathbf{A}_{m \times n} &= \lambda \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} \lambda \cdot a_{11} & \lambda \cdot a_{12} & \cdots & \lambda \cdot a_{1n} \\ \lambda \cdot a_{21} & \lambda \cdot a_{22} & \cdots & \lambda \cdot a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda \cdot a_{m1} & \lambda \cdot a_{m2} & \cdots & \lambda \cdot a_{mn} \end{bmatrix} \end{aligned}$$

1.3.4. Multiplicación de Matrices

Para multiplicar matrices se requiere que el número de columnas de la primera matriz sea igual al número de filas de la segunda matriz. Esto es porque se multiplican, en orden, los elementos de la fila i de la primera matriz por los elementos de la columna j de la segunda matriz. Por ello, la cantidad de elementos debe ser la misma. Sólo es posible multiplicar $\mathbf{A}_{m \times n}$ y $\mathbf{B}_{n \times q}$ sí y sólo si n es igual a n , y el resultado de la multiplicación denotado por la matriz \mathbf{C} será de orden $m \times q$. Por ejemplo,

$$\mathbf{A}_{m \times n} \cdot \mathbf{B}_{n \times q} = \mathbf{C}_{m \times q}$$

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & \cdots & b_{1q} \\ b_{21} & \cdots & b_{2q} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nq} \end{bmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1q} \\ c_{21} & \cdots & c_{2q} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mq} \end{bmatrix},$$

donde cada elemento de la matriz \mathbf{C} corresponde a lo siguiente:

$$c_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj}$$

Así, por ejemplo el primer elemento de la matriz \mathbf{C} , (c_{11}) corresponde a:

$$c_{11} = a_{11} \cdot b_{11} + a_{12} \cdot b_{21} + a_{13} \cdot b_{31} + \cdots + a_{1n} \cdot b_{n1}$$

Si el número de filas de la primera matriz \mathbf{A} , es igual al número de columnas de la segunda matriz \mathbf{B} , se dice que estas matrices son *conformables* para la multiplicación. Como ejemplo consideremos la siguiente matriz \mathbf{A} de orden 2×3 , que vamos a multiplicar por la matriz \mathbf{B} que necesariamente debe tener **3** filas, es decir, una posibilidad es que sea de orden 3×1 :

$$\mathbf{A} = \begin{bmatrix} 5 & 4 & -10 \\ -8 & 2 & 15 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix}$$

Sabiendo el orden de ambas matrices, \mathbf{A} y \mathbf{B} , conocemos el orden de la matriz resultante, ésta es de orden 2×1 . La matriz resultante es la siguiente:

$$\mathbf{A} \times \mathbf{B} = \begin{bmatrix} 5 \cdot 1 + 4 \cdot (-3) + (-10) \cdot (-2) \\ (-8) \cdot 1 + 2 \cdot (-3) + 15 \cdot (-2) \end{bmatrix} = \begin{bmatrix} 5 - 12 + 20 \\ (-8) - 6 - 30 \end{bmatrix} = \begin{bmatrix} 13 \\ -44 \end{bmatrix}$$

Matriz Idempotente

En el caso particular de que de la multiplicación de una matriz cuadrada \mathbf{A} por sí misma se obtenga como resultado la misma matriz \mathbf{A} , se habla de una matriz idempotente. Analíticamente se tiene que:

$$\mathbf{A}_{n \times n} \cdot \mathbf{A}_{n \times n} = \mathbf{A}_{n \times n}$$

Por ejemplo:

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 6 & 10 \\ -3 & -5 \end{bmatrix} \cdot \begin{bmatrix} 6 & 10 \\ -3 & -5 \end{bmatrix} \\ &= \begin{bmatrix} 6 \cdot 6 + 10 \cdot (-3) & 6 \cdot 10 + 10 \cdot (-5) \\ (-3) \cdot 6 + (-5) \cdot (-3) & (-3) \cdot 10 + (-5) \cdot (-5) \end{bmatrix} \\ &= \begin{bmatrix} 6 & 10 \\ -3 & -5 \end{bmatrix} \end{aligned}$$

Esto quiere decir que al multiplicar esta matriz por si misma un número de veces cualquiera siempre dará como resultado la matriz \mathbf{A} . La matriz \mathbf{I} es otro ejemplo de matriz idempotente.

Propiedades de la Multiplicación de Matrices

Algunas propiedades de la multiplicación de matrices son:

1. La multiplicación de matrices no es, en general, conmutativa, es decir:

$$\mathbf{A} \cdot \mathbf{B} \neq \mathbf{B} \cdot \mathbf{A}$$

Una conclusión que se puede sacar de esta propiedad, es que por lo general el orden que posea la matriz obtenida de multiplicar $\mathbf{A} \cdot \mathbf{B}$ es distinto al orden de la matriz obtenida de multiplicar $\mathbf{B} \cdot \mathbf{A}$.

Además, dado que esta operación no es conmutativa, el orden de los factores sí altera el producto. Es necesario definir dos operaciones de multiplicación de matrices. La multiplicación por la derecha o postmultiplicación y la multiplicación por la izquierda o premultiplicación. En el caso de $\mathbf{A} \cdot \mathbf{B}$, implica que \mathbf{A} esta premultiplicando a \mathbf{B} , o alternativamente, \mathbf{B} está postmultiplicando a \mathbf{A} .

2. Un vector fila ($\boldsymbol{\mu}'$), postmultiplicado por su traspuesta ($\boldsymbol{\mu}$) da como resultado un escalar. Este escalar va a ser la suma de los elementos del vector al cuadrado, es decir:

$$\boldsymbol{\mu}' \cdot \boldsymbol{\mu} = [\mu_1 \ \mu_2 \ \cdots \ \mu_n] \cdot \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \mu_1^2 + \mu_2^2 + \cdots + \mu_n^2 = \sum_{i=1}^n \mu_i^2$$

3. Un vector columna postmultiplicado por su traspuesta da como resultado una matriz cuadrada simétrica de orden n , donde n es el número de elementos del vector columna, es decir:

$$\boldsymbol{\mu} \cdot \boldsymbol{\mu}' = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \cdot [\mu_1 \ \mu_2 \ \cdots \ \mu_n] = \begin{bmatrix} \mu_1^2 & \mu_1\mu_2 & \cdots & \mu_1\mu_n \\ \mu_2\mu_1 & \mu_2^2 & \cdots & \mu_2\mu_n \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n\mu_1 & \mu_n\mu_2 & \cdots & \mu_n^2 \end{bmatrix}$$

4. La multiplicación de matrices es asociativa, es decir:

$$(\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C})$$

5. La multiplicación de matrices es distributiva con respecto a la suma (resta), es decir:

$$\mathbf{A} \cdot (\mathbf{B} + (-) \mathbf{C}) = (\mathbf{A} \cdot \mathbf{B}) + (-) (\mathbf{A} \cdot \mathbf{C})$$

6. La transposición del producto de dos matrices, es igual a la multiplicación de ambas matrices transpuestas en orden inverso, es decir:

$$(\mathbf{A} \cdot \mathbf{B})' = \mathbf{B}' \cdot \mathbf{A}'$$

Observe que en la suma (resta) de matrices transpuestas se mantiene el orden de los elementos, mientras que en la multiplicación el orden es inverso. Esta propiedad se puede generalizar en el caso que sean más de dos matrices las que se multiplican, esto es:

$$(\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C})' = \mathbf{C}' \cdot \mathbf{B}' \cdot \mathbf{A}'$$

7. La premultiplicación y postmultiplicación de una matriz A por la matriz identidad entrega como resultado la matriz A , es decir:

$$\mathbf{A}_{m \times n} \cdot \mathbf{I}_n = \mathbf{I}_m \cdot \mathbf{A}_{m \times n} = \mathbf{A}_{m \times n}$$

8. La transpuesta del producto de un escalar por una matriz es igual a la multiplicación del escalar por la transpuesta de la matriz, es decir:

$$(\lambda \cdot \mathbf{A})' = \mathbf{A}' \cdot \lambda' = \mathbf{A}' \cdot \lambda = \lambda \cdot \mathbf{A}'$$

1.3.5. Producto Kronecker

Esta operación entre matrices, implica la multiplicación entre dos matrices (\mathbf{A} y \mathbf{B}) pero en una forma particular, en la cual no se requiere igualdad entre las filas de la matriz \mathbf{A} y las columnas de la matriz \mathbf{B} , al contrario de los casos de adición y multiplicación presentados anteriormente. El producto de Kronecker implica multiplicar cada elemento de la matriz \mathbf{A} por toda la

matriz \mathbf{B} , como una simple multiplicación por un escalar. Si \mathbf{A} es de orden $m \times n$ y \mathbf{B} es de $p \times q$, entonces nos queda que la matriz \mathbf{C} es de orden $mp \times nq$, denotado por:

$$\mathbf{A}_{m \times n} \otimes \mathbf{B}_{p \times q} = \mathbf{C}_{mp \times nq}$$

donde \otimes denota el operador del producto de Kronecker. Para estas matrices la expresión más general de la matriz \mathbf{C} es:

$$\mathbf{C}_{mp \times nq} = \begin{bmatrix} a_{11} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix} & \cdots & a_{1n} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix} \\ a_{21} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix} & \cdots & a_{2n} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ a_{m1} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix} & \cdots & a_{mn} \cdot \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{bmatrix} \end{bmatrix}$$

Para clarificar la aplicación de este operador, veamos el ejemplo de dos matrices de 2×2 . El resultado se obtiene como se muestra a continuación.

$$\mathbf{A}_{2 \times 2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{B}_{2 \times 2} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

con estas matrices el producto de Kronecker queda expresado como:

$$\mathbf{C}_{4 \times 4} = \mathbf{A}_{2 \times 2} \otimes \mathbf{B}_{2 \times 2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$\mathbf{C}_{4 \times 4} = \begin{bmatrix} a_{11} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} & a_{12} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\ a_{21} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} & a_{22} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \end{bmatrix}$$

Cuyo resultado final es:

$$\mathbf{C}_{4 \times 4} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}$$

1.3.6. Traza de una matriz

La función traza de la matriz \mathbf{A} , denotado por $tr(\mathbf{A})$, es una función que entrega como resultado un escalar obtenido de la suma de los elementos de la diagonal principal de una matriz cuadrada, es decir:

$$tr(\mathbf{A}) = \sum_{i=1}^n a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$$

Si consideramos la siguiente matriz \mathbf{A} , su función traza es:

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & 3 \\ 5 & 9 & -1 \\ 2 & -8 & 6 \end{bmatrix}$$

$$tr(\mathbf{A}) = 4 + 9 + 6 = 19$$

Esta función tiene las siguientes propiedades:

1. La traza de la suma (resta) de dos matrices, \mathbf{A} y \mathbf{B} , del mismo orden es igual a la suma (resta) de las trazas, es decir:

$$tr(\mathbf{A} + (-) \mathbf{B}) = tr(\mathbf{A}) + (-) tr(\mathbf{B})$$

2. La traza del producto de la matriz \mathbf{A} y un escalar, es igual al escalar multiplicado por la traza de la matriz \mathbf{A} , es decir:

$$tr(\alpha \mathbf{A}) = \alpha tr(\mathbf{A})$$

3. La traza del producto de la matriz \mathbf{A} posmultiplicada por la matriz \mathbf{B} es igual a la traza del producto de la matriz \mathbf{A} premultiplicada por la matriz \mathbf{B} , es decir:

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}),$$

siempre que \mathbf{AB} y \mathbf{BA} estén definidos. Lo mismo es válido para los productos de las matrices \mathbf{A} , \mathbf{B} , \mathbf{C} siempre que éstos estén definidos:

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CBA})$$

4. La traza de una matriz \mathbf{B} , es igual a la traza de la matriz \mathbf{B} posmultiplicada por la matriz \mathbf{A} y premultiplicada por la inversa de la matriz \mathbf{A} , siempre y cuando la matriz \mathbf{A} tenga inversa y esté conforme para la multiplicación, es decir:

$$\text{tr}(\mathbf{B}) = \text{tr}(\mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{A})$$

1.4. Determinante de una Matriz

El **determinante** de una matriz es una función sobre una matriz cuadrada, que asocia a ésta un número real unívoco. Se denota por **det** \mathbf{A} o por el símbolo $|\mathbf{A}|$. El determinante de una matriz de orden n , se obtiene como la suma de $n!$ (factorial) de términos. Cada uno de estos términos se obtiene de la multiplicación de n elementos de la matriz, uno de cada fila y de cada columna, es decir:

$$|\mathbf{A}| = \sum^{n!} (-1)^\sigma \cdot a_{1i_1} \cdot a_{2i_2} \cdot \dots \cdot a_{ni_n}$$

Donde, σ se define como el número de permutaciones posibles entre los elementos de la matriz. Para una matriz de 2×2 el determinante está definido como:

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ |\mathbf{A}| &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (a_{11} \cdot a_{22}) - (a_{21} \cdot a_{12}) \end{aligned}$$

En este caso básico, la regla para obtener el determinante indica que se debe multiplicar los elementos de la diagonal principal (a_{11} y a_{22}), y a ellos restar el producto de los elementos de la otra diagonal (a_{12} y a_{21}).

Por su parte si la matriz es de dimensión de 3×3 , su determinante se define como:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$|\mathbf{A}| = (a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{32} \cdot a_{21}) - (a_{31} \cdot a_{22} \cdot a_{13} + a_{11} \cdot a_{32} \cdot a_{23} + a_{33} \cdot a_{21} \cdot a_{12})$$

En este segundo caso, el determinante de esta matriz se obtiene como la suma de los productos de los elementos que pertenecen a las diagonales que tienen sentido hacia abajo y a la derecha, menos la suma del producto de los elementos que pertenecen a las diagonales que tienen sentido hacia arriba y la derecha, tal como lo indica la fórmula anterior.

Este es un método que presenta cierta dificultad, especialmente si se trata de matrices de orden 5 o superiores, donde la suma va a estar formada por lo menos por 120 términos. Afortunadamente existen otros métodos que sirven para simplificar el cálculo del determinante. Para aplicar estos métodos es necesario discutir los conceptos de Menor, Cofactor, Matriz de Cofactores y Matriz Adjunta.

1.4.1. Menor de una Matriz

Supongamos que tenemos una matriz cuadrada \mathbf{A} de orden 3, es decir;

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

El menor es un determinante especial de la matriz. Para esta manera se pueden definir 9 menores, tantos como los elementos que contiene. Podemos definir el menor correspondiente a cada elemento, como el determinante de la matriz resultante de eliminar la fila y la columna de dicho elemento de la matriz original. Por ejemplo, para el elemento a_{21} de la matriz \mathbf{A} , el menor está dado por:

$$\mathbf{M}_{21} = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} = (a_{12} \cdot a_{33}) - (a_{32} \cdot a_{13})$$

y se denota \mathbf{M}_{21} , donde se ha construido una matriz nueva eliminando la segunda fila y la primera columna de la matriz \mathbf{A} . A esta nueva matriz se le calcula su determinante. Así, se pueden encontrar todos los menores de una determinada matriz.

Ejemplo: Consideremos la siguiente matriz cuadrada \mathbf{A} , de orden 3×3 .

$$\mathbf{A}_{3 \times 3} = \begin{bmatrix} 5 & 4 & -9 \\ -11 & 0 & 3 \\ 2 & 8 & 1 \end{bmatrix}$$

Se puede obtener el menor del elemento a_{22} , y éste queda de la siguiente manera:

$$\mathbf{M}_{22} = \begin{vmatrix} 5 & -9 \\ 2 & 1 \end{vmatrix} = 5 \cdot 1 - 2 \cdot (-9) = 23$$

1.4.2. Cofactor de una Matriz

El cofactor es otra magnitud importante de las matrices cuadradas. Se denota por c_{ij} , y se define de la siguiente manera:

$$c_{ij} = (-1)^{i+j} \mathbf{M}_{ij}$$

Donde M_{ij} es el menor discutido anteriormente. En otras palabras, el cofactor es un menor con un signo determinado. Si consideramos el ejemplo anterior, se puede determinar el cofactor del elemento a_{22} de la siguiente forma:

$$c_{22} = (-1)^{2+2} \cdot \mathbf{M}_{22} = 23$$

En este caso el cofactor corresponde con el menor antes calculado, aunque ello no siempre es así. Para los cofactores cuyos subíndices \mathbf{i}, \mathbf{j} sumen un valor impar entonces el signo que precederá al menor será negativo. Este será el caso, por ejemplo, del cofactor c_{21} .

1.4.3. Matriz de Cofactores

Una vez definidos los cofactores, se puede plantear la matriz de cofactores, que es la matriz resultante de reemplazar todos los elementos a_{ij} de la matriz original por sus cofactores. Generalmente se le denotará como **cof A**. Así la matriz de cofactores de **A**, definida en el ejemplo anterior, será:

$$\text{cof } \mathbf{A}_{3 \times 3} = \begin{bmatrix} -24 & 17 & -88 \\ -76 & 23 & -32 \\ 12 & 84 & 44 \end{bmatrix}$$

El lector debería asegurarse que entendió como se obtuvo esta matriz de cofactores.

1.4.4. Matriz Adjunta

La matriz adjunta, que es la transpuesta de la matriz de cofactores, se denota por $\text{Adj}(\mathbf{A})$. Analíticamente se expresa como:

$$\text{Adj}(\mathbf{A}) = (\text{cof } \mathbf{A})'$$

Retomando el ejemplo anterior la matriz adjunta de la matriz **A** queda de la siguiente manera:

$$\text{Adj}(\mathbf{A}_{3 \times 3}) = \begin{bmatrix} -24 & -76 & 12 \\ 17 & 23 & 84 \\ -88 & -32 & 44 \end{bmatrix}$$

Esta matriz, será de utilidad cuando se discuta la obtención de la matriz inversa.

1.4.5. Método de Cofactores para el Determinante

Ahora estamos en condiciones de retomar el cálculo de determinante a través de un método alternativo conocido como el *Método de los Cofactores*. Con este procedimiento el determinante se obtiene de sumar los productos de cada elemento y su correspondiente cofactor a lo largo de una fila o columna cualquiera de la matriz. Por ejemplo si usamos la matriz cuadrada de orden 2×2 tenemos:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Si sumamos los elementos de la primera columna, el determinante de esta matriz se obtiene como:

$$|\mathbf{A}| = a_{11} \cdot [(-1)^{1+1} \cdot a_{22}] + a_{21} [(-1)^{2+1} \cdot a_{12}] = a_{11} \cdot a_{22} - a_{21} \cdot a_{12}$$

Alternativamente, si sumamos por la segunda columna, obtenemos el mismo resultado.

$$|\mathbf{A}| = a_{12} [(-1)^{1+2} \cdot a_{21}] + a_{22} \cdot [(-1)^{2+2} \cdot a_{11}] = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

Consideremos la matriz del ejemplo anterior:

$$\mathbf{A}_{3 \times 3} = \begin{bmatrix} 5 & 4 & -9 \\ -11 & 0 & 3 \\ 2 & 8 & 1 \end{bmatrix}$$

El determinante de esta matriz queda expresado de la siguiente manera, si consideramos aleatoriamente la segunda columna:

$$\begin{aligned} |\mathbf{A}| &= a_{12} \cdot c_{12} + a_{22} \cdot c_{22} + a_{32} \cdot c_{32} \\ &= 4 \cdot 17 + 0 \cdot 23 + 8 \cdot 84 = 68 + 672 \\ &= 740 \end{aligned}$$

Este ejemplo ilustra un criterio de eficiencia al aplicar el uso de este método. Como es posible ver el elemento a_{22} es cero, por lo cual el producto correspondiente desaparece de la sumatoria. Ello simplifica el cálculo del determinante. Por ello uno debería elegir una fila o columna que tenga la mayor cantidad de ceros para reducir los cálculos lo más que se pueda.

1.4.6. Propiedades del Determinante.

1. Los determinantes de una matriz \mathbf{A} y de su transpuesta (\mathbf{A}) son iguales, es decir, $|\mathbf{A}| = |\mathbf{A}'|$.

2. Al intercambiar de posición dos filas o columnas cualesquiera de una matriz \mathbf{A} cambia el signo del determinante. Es decir, si \mathbf{B} resulta de intercambiar de posición dos filas (columnas) de \mathbf{A} , entonces

$$|\mathbf{B}| = -|\mathbf{A}|$$

3. El determinante del producto de dos matrices es igual al producto de los determinantes de las matrices.

$$|\mathbf{A} \cdot \mathbf{B}| = |\mathbf{A}| \cdot |\mathbf{B}|$$

4. El determinante del producto de una matriz de orden n , multiplicada por un escalar, es igual al escalar con potencia n multiplicado por el determinante de la matriz original, es decir:

$$|\alpha\mathbf{A}| = \alpha^n \cdot |\mathbf{A}|$$

5. El determinante de una matriz será cero si:
- la matriz posee dos filas (columnas) iguales;
 - Una de las filas (columnas) es una combinación lineal de otras filas (columnas);
 - Todos los elementos de una fila (columna) de la matriz son cero.

1.5. Rango de una Matriz

Una matriz se puede definir como un conjunto de vectores filas o vectores columnas. Esto permite aplicar la teoría de Espacios Vectoriales a las filas o columnas de una matriz, y hablar de filas (columnas) linealmente independientes o dependientes. Entonces el **rango** de una matriz se define como el máximo número de filas (columnas) linealmente independientes que tiene la matriz. Por ejemplo, si la matriz \mathbf{A} es cuadrada con orden n , y todas las filas son linealmente independientes, entonces el rango de la matriz será n . Se puede definir la **matriz singular**, que es aquella que posee filas (columnas)

Escrito matricialmente quedaría de la siguiente forma:

$$\mathbf{A}_{n \times n} \cdot \mathbf{x}_{n \times 1} = \mathbf{b}_{n \times 1}$$

Luego resolviendo el valor de \mathbf{x} se obtendría la solución completa del sistema de ecuaciones. Esto se soluciona premultiplicando ambos lados de la ecuación por la inversa de la matriz \mathbf{A} , es decir por \mathbf{A}^{-1} . De esta forma tenemos:

$$\begin{aligned} \mathbf{A}^{-1} \cdot \mathbf{A} \cdot \mathbf{x} &= \mathbf{A}^{-1} \cdot \mathbf{b} \\ \mathbf{I} \cdot \mathbf{x} &= \mathbf{A}^{-1} \cdot \mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1} \cdot \mathbf{b}, \end{aligned}$$

Lo cual es la solución del sistema de ecuaciones.

Una forma bastante común de encontrar esta matriz inversa es:

$$\mathbf{A}^{-1} = \frac{Adj(\mathbf{A})}{|\mathbf{A}|}$$

Consideremos como ejemplo la matriz de los ejemplos anteriores:

$$\mathbf{A}_{3 \times 3} = \begin{bmatrix} 5 & 4 & -9 \\ -11 & 0 & 3 \\ 2 & 8 & 1 \end{bmatrix}$$

Si aplicamos la definición para obtener la matriz inversa, debemos usar la matriz adjunta y el determinante obtenidos anteriormente¹:

$$\mathbf{A}^{-1} = \frac{\begin{bmatrix} -24 & -76 & 12 \\ 17 & 23 & 84 \\ -88 & -32 & 44 \end{bmatrix}}{740} = \begin{bmatrix} -24/740 & -76/740 & 12/740 \\ 17/740 & 23/740 & 84/740 \\ -88/740 & -32/740 & 44/740 \end{bmatrix}$$

En otras palabras, si a_{ij} es el elemento ij -ésimo de la matriz \mathbf{A}^{-1} , entonces:

$$a_{ij} = \frac{|C_{ji}|}{|\mathbf{A}|}$$

¹Ver Sección 1.4.

1.6.1. Propiedades de la Matriz Inversa

Algunas de las propiedades que presenta la matriz inversa son:

1. La inversa de una matriz invertida es la matriz original, es decir:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

2. La transposición e inversión son operaciones que pueden intercambiarse sin alterar el resultado, es decir:

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

3. La inversa del producto de dos matrices es equivalente al producto de las inversas de las matrices, pero con el orden intercambiado, siempre y cuando existan las inversas de cada una de ellas, es decir:

$$(\mathbf{A} \cdot \mathbf{B})^{-1} = \mathbf{B}^{-1} \cdot \mathbf{A}^{-1}$$

4. La inversa de una matriz triangular o diagonal es también una matriz triangular o diagonal, con la característica en las matrices diagonales, que los elementos son los recíprocos de los elementos originales.
5. Los determinantes de una matriz y su inversa, son recíprocos entre ellos, es decir:

$$|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$$

1.7. Diferenciación de Matrices

La diferenciación de matrices es una herramienta importante en econometría, especialmente cuando se requiere maximizar o minimizar funciones objetivo escritas en forma de matrices. Un ejemplo de ello es el análisis de regresión múltiple.

- i) Asuma que existe un vector columna \mathbf{m} , cuyos elementos son números y otro vector columna de variables \mathbf{x} tales que:

$$\mathbf{m}_{n \times 1} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} \quad \mathbf{x}_{n \times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Para multiplicar estos vectores es necesario trasponer uno de ellos, por ejemplo, se puede trasponer el vector \mathbf{m} , obteniéndose un escalar al multiplicar ambos vectores (es un polinomio). Es decir:

$$\mathbf{m}'\mathbf{x} = (m_1 \ m_2 \ \cdots \ m_n) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = m_1x_1 + m_2x_2 + \cdots + m_nx_n$$

Note que $\mathbf{m}'\mathbf{x}$ es igual a $\mathbf{x}'\mathbf{m}$. Luego si queremos derivar con respecto al vector columna \mathbf{x} , lo que se hace es derivar el polinomio resultante de la multiplicación por cada variable del vector columna, de la siguiente forma:

$$\frac{\partial(\mathbf{m}'\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial(\mathbf{m}'\mathbf{x})}{\partial x_1} \\ \frac{\partial(\mathbf{m}'\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial(\mathbf{m}'\mathbf{x})}{\partial x_t} \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_t \end{pmatrix} = \mathbf{m}$$

De esta manera conseguimos el vector \mathbf{m} .

- ii) Un segundo caso de relevancia, cuando se trabaja con formas cuadráticas, considera el vector columna \mathbf{x} del caso anterior más una matriz cuadrada \mathbf{M} simétrica de coeficientes de orden \mathbf{n} , del siguiente tipo:

$$\mathbf{M}_{n \times n} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix} \quad \mathbf{x}_{n \times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

La forma cuadrática se expresa de la siguiente forma:

$$\mathbf{x}'_{1 \times n} \cdot \mathbf{M}_{(n \times n)} \cdot \mathbf{x}_{(n \times 1)} = \sum_{i=1}^n \sum_{j=1}^n m_{ij} \cdot x_i \cdot x_j$$

Si se deriva la expresión anterior con respecto a cada variable del vector columna \mathbf{x} , se tiene, entonces, que:

$$\frac{\partial (\mathbf{x}'\mathbf{M}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{M}\mathbf{x} = 2\mathbf{M}'\mathbf{x}$$

Ejemplo: Supongamos que se tiene una matriz cuadrada simétrica de coeficientes \mathbf{M} de orden 2, y un vector \mathbf{x} de variables de orden 2×1 :

$$\mathbf{M}_{2 \times 2} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \quad \mathbf{x}_{2 \times 1} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Al establecer la forma cuadrática se tiene que:

$$\begin{aligned} \mathbf{x}'\mathbf{M}\mathbf{x} &= [x_1 \quad x_2] \cdot \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= [x_1 m_{11} + x_2 m_{21} \quad x_1 m_{12} + x_2 m_{22}] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= [x_1^2 m_{11} + 2x_1 x_2 m_{12} + x_2^2 m_{22}] \end{aligned}$$

donde se asumió que la matriz \mathbf{M} es simétrica, es decir $m_{12} = m_{21}$. Luego derivamos por el vector \mathbf{x} , es decir, derivamos con respecto a cada variable del vector. Esto queda:

$$\begin{aligned} \frac{\partial (\mathbf{x}'\mathbf{M}\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial (x_1^2 m_{11} + 2x_1 x_2 m_{12} + x_2^2 m_{22})}{\partial \mathbf{x}} = \begin{bmatrix} 2x_1 m_{11} + 2x_2 m_{12} \\ 2x_1 m_{12} + 2x_2 m_{22} \end{bmatrix} \\ &= \begin{bmatrix} 2m_{11} & 2m_{12} \\ 2m_{12} & 2m_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2 \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2\mathbf{M}\mathbf{x} = 2\mathbf{M}'\mathbf{x} \end{aligned}$$

El conocimiento de los tópicos tratados hasta este momento facilitará la comprensión de los siguientes capítulos, en los cuales se presentarán muchas veces ejemplos de problemas en forma matricial y donde se usarán las propiedades de las matrices para obtener resultados útiles desde la perspectiva del econométrista.

Capítulo 2

ESTADÍSTICA, PROBABILIDAD E INFERENCIA

2.1. Introducción: Definiciones

La economía está interesada en ciertos fenómenos que desde el punto de vista estadístico reciben el nombre genérico de **variables**. Las variables son magnitudes que pueden adoptar diferentes valores en distintos puntos de observación. En el plano macroeconómico algunos ejemplos de estas variables son el Producto Interno Bruto, las exportaciones y el tipo de cambio real, mientras que en el plano microeconómico, tenemos variables como, el precio de los bienes, el ingreso y el consumo de las familias.

Antes de profundizar en los tópicos econométricos se debe discutir algunos conceptos estadísticos de utilidad para comprender los procedimientos econométricos. Primero definamos un **suceso** o **punto muestral** como uno de los posibles resultados que pueden observarse en el comportamiento de una determinada variable. Cada variable puede, en principio, adoptar distintos valores (resultados). Sin embargo, en cada observación vemos sólo uno de los valores posibles, el cual corresponde a un suceso. Distinguiremos variables **discretas** y **continuas**. La variable será **discreta** si su recorrido es un conjunto finito de valores reales. Ejemplo de este tipo de variable es el número de viajes a un sitio recreacional tal como una playa, realizados por una familia el cual claramente sólo puede tomar números enteros. Por el contrario,

una variable será **continua** si el recorrido es un intervalo de la recta real o la recta real completa. En este caso, podemos mencionar como ejemplos el PIB, el ingreso familiar. Estas variables son continuas en el intervalo no negativo de la recta real, porque en principio pueden tomar cualquier valor mayor o igual a cero.

Una **muestra** es un conjunto de sucesos (medidas) seleccionadas de un **universo** o **espacio muestral**. Este último abarca la totalidad de resultados posibles de un experimento. A este universo se le da el nombre de **población**. Por ejemplo, el espacio muestral asociado al PIB, está limitado a ser un valor no negativo. En principio, el PIB puede tomar innumerables valores diferentes, pero si tomamos el PIB efectivo entre los años 1980 y 1990 tendremos una muestra determinada la cual fue escogida del universo muestral.

La información contenida en la muestra se utiliza de dos formas. La primera de éstas da origen a la *Estadística Descriptiva*. Ésta se preocupa de la recolección, ordenamiento y presentación de datos. Por ejemplo, cuál es el promedio, la desviación estándar, y la distribución de los datos de ingreso nacional en una determinada muestra de la población chilena para un determinado año. La segunda es la *Inferencia Estadística*. Ésta utiliza la información de la muestra para proyectar sus resultados al conjunto de la población. Un ejemplo es intentar, sobre la base del promedio de ingreso de la muestra, inferir cuál es el ingreso promedio de la población.

En econometría generalmente se cuenta con datos que contienen información de un determinado período o de un determinado segmento de los agentes económicos. Un objetivo es, comúnmente, inferir cómo se comportarán estas variables en otro período para el cual no tenemos información o cómo reaccionarán otros agentes económicos no considerados en la muestra. Es decir, generalmente, estamos interesados en el comportamiento de la población. Desde esta perspectiva, la econometría está interesada principalmente en la inferencia estadística.

Note además, que en Estadística se trabaja generalmente con experimentos muestrales en los cuales se selecciona con algún método (aleatorio generalmente) las observaciones de la población que pertenecerán a la muestra. Sin embargo, en el campo de la economía la obtención de la muestra no siempre responde al diseño de un experimento muestral y lo que ocurre generalmente es que los datos ya están predefinidos por la disponibilidad de información económica (es el caso de variables como el Producto Interno Bruto, el nivel de ahorro e inversión de la economía, etc.).

Las variables pueden ser clasificadas en función de su naturaleza deter-

minística o aleatoria. Se definen como variables **determinísticas** o **controlables** aquellas cuyo valor se conoce con certeza. En cambio se denomina variables **aleatorias** o **no controlables**, las que asumen un valor con alguna probabilidad asociada. De esta forma, el valor exacto que tomará esta variable es desconocido antes que un determinado experimento se desarrolle, pero puede asumir un conjunto de valores dentro de un determinado rango, cada uno de estos valores con una probabilidad distinta. Desde la perspectiva econométrica son de particular interés las variables aleatorias. Ello porque la medición de los fenómenos económicos es de carácter probabilístico. Por una parte, la ciencia económica no posee conocimiento sobre todos los acontecimientos que condicionan un evento y/o no puede medir todos los acontecimientos relevantes. Ello hace que en los intentos de aplicar las teorías y medir los fenómenos existan muchos elementos “aleatorios”, que reflejan la dimensión del desconocimiento del investigador de variables relevantes. Por otra parte, el interés de proyectar y predecir los valores de las variables económicas es siempre incierto y por ello probabilístico.

Existen distintas formas de concebir la noción de probabilidad. Por una parte, la idea de **probabilidad** asociada a una variable aleatoria se relaciona con la creencia en la ocurrencia de un evento incierto, y se basa en la noción de experimentos mutuamente excluyentes e igualmente probables. Es decir, si \mathbf{n} es el espacio muestral, la probabilidad de un resultado es $\mathbf{1/n}$ para un experimento con resultados excluyentes e igualmente probables. Por ejemplo, si hay \mathbf{n} bolitas de distintos colores en una caja, iguales en todos los otros sentidos, y se extrae una en forma aleatoria sin conocer el color, entonces la probabilidad de tomar un color determinado es $\mathbf{1/n}$.

Otra forma de entender la probabilidad es como una frecuencia relativa. En este caso, si el experimento se repite \mathbf{N} veces y se observa que un evento \mathbf{A} ocurre $\mathbf{n_i}$ veces, entonces $\mathbf{n_i/N}$ es la frecuencia relativa de \mathbf{A} en \mathbf{N} repeticiones del experimento. Al extender esta idea se puede entender la probabilidad como el límite de la frecuencia relativa cuando el número de veces que se repite el experimento tiende a infinito, siempre y cuando este límite exista. En este caso se asume que el experimento puede ser repetido en las mismas condiciones aunque sea conceptualmente.

Para que una medida sea probabilística debe cumplir con algunas condiciones. La probabilidad es una medida, cuyo valor no puede adoptar números negativos. Es más, debe tomar un valor entre cero y uno. Es decir, su valor máximo es uno y su valor mínimo es cero. Además, tiene como propiedad que la suma de todos los sucesos posibles, mutuamente excluyentes, es igual

a la unidad. Cumplidas estas condiciones podemos construir una *función de distribución de probabilidad* o simplemente *función de probabilidad* que nos mostrará cómo la masa de la probabilidad se distribuye entre los valores que puede tomar la variable aleatoria.

Para una variable aleatoria discreta X la *función de probabilidad* se define como una función $f(x)$ tal que para todo número real x , donde $x \in X$, $f(x) = P(X = x)$, lo cual representa la probabilidad que la variable aleatoria X tome el valor x^1 y se cumple que:

1. $0 \leq f(x) \leq 1$ si $x \in X$ y
2. $f(x) = 0$ si $x \notin X$

Además, si x_1, x_2, \dots, x_n son todos los valores de X entonces

3. $\sum_{i=1}^n f(x_i) = 1$

Ejemplo 1: Suponga que la variable **aleatoria** (X) es el número de caras que resulta del lanzamiento de una moneda que se arroja dos veces al aire. Para cada lanzamiento el resultado (o suceso) puede ser cara (C) o sello (S). Por ello los posibles resultados para ambos lanzamientos son

$$\begin{array}{ll} x_1 & SS \\ x_2 & SC \\ x_3 & CS \\ x_4 & CC \end{array}$$

Si asumimos que la moneda y el lanzamiento no tiene sesgo, la probabilidad de cualquier evento particular para un lanzamiento será $\frac{1}{2}$. Por ello para dos lanzamientos consecutivos e independientes será $\frac{1}{4}$ ($\frac{1}{2} \times \frac{1}{2}$). De tal forma podemos ordenar la información en la siguiente tabla:

Sucesos	N° de caras (x)	Función de probabilidad $f(x)$
SS	0	1/4
SC	1	1/4
CS	1	1/4
CC	2	1/4
Total		1

¹En general, denotamos la variable que puede tomar diversos valores con mayúsculas, y una realización específica (un número determinado) de la variable con minúsculas.

O si el orden en que salen los resultados en cada lanzamiento no nos interesa, podemos escribir

N° de caras (x)	Función de probabilidad $f(x)$
0	1/4
1	1/2
2	1/4
Total	1

La distribución de las probabilidades de ocurrencia de los eventos es lo que se conoce por función de distribución de probabilidades.

En el caso de una variable **continua** en que X puede tomar cualquier valor en al menos un intervalo sobre la línea de números reales, la probabilidad que X tome cualquier valor particular es cero. Esto porque la cantidad de valores posibles es infinito en cualquier intervalo de la recta real. Por ello, la probabilidad de que un valor específico ocurra es cero. De tal manera que la función de probabilidad, que para el caso de una variable continua se denomina *función de densidad de probabilidad*, tiene sentido sólo en un intervalo. Una función definida sobre los números reales es una función de densidad de probabilidad para una variable aleatoria continua si:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. $\forall a, b$ con $-\infty \leq a \leq b \leq \infty$, se cumple $P(a \leq x \leq b) = P(a < x < b) = \int_a^b f(x) dx$

En este caso es más claro el hecho que la $P(X = x) = 0$, ya que en este caso el límite superior e inferior de la integral coinciden, y esto implica que el área asociada a esta integral es cero.

Toda variable aleatoria tiene una *función de distribución acumulada* que se define como la suma de probabilidades para los valores de la variable aleatoria X , menores o iguales que un valor dado x . Es decir, la función de distribución acumulada está dada por, $F(x) = \mathbf{P}(X \leq x)$ para $-\infty < x < \infty$,

donde $F(x)$ es la probabilidad que la función tome valores menores o iguales a x . Para una variable discreta será:

$$F(x_0) = \sum_{-\infty}^{x_0} f(t),$$

mientras que para una variable continua su función de densidad acumulada es:

$$F(x_0) = \int_{-\infty}^{x_0} f(t) dt$$

En el ejemplo de la moneda que se arroja dos veces (ejemplo 1), la distribución acumulada (discreta) de que al menos una de las monedas sea sello es

$$P(X \leq 1) = F(1) = \sum_{x=0}^1 f(x) = f(0) + f(1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

Las propiedades que debe cumplir esta función de densidad acumulada son las siguientes:

1. $F(-\infty) = 0$
2. $F(\infty) = 1$
3. $P(a \leq x \leq b) = F(b) - F(a)$
4. $f(x) = \frac{\partial F(x)}{\partial x} = F'(x)$ para el caso continuo.

Por último, es necesario mencionar la función de probabilidad conjunta. Sea X e Y dos variables aleatorias. El correspondiente espacio de X e Y , corresponde a un espacio bidimensional, donde la probabilidad que $X = x$ e $Y = y$ simultáneamente, es denotado por $f(x, y) = P(X = x, Y = y)$. La función $f(x, y)$ es llamada la función de densidad conjunta de X e Y y tiene las siguientes propiedades:

1. $0 \leq f(x, y) \leq 1$
2. $\sum_x \sum_y f(x, y) = 1$ para el caso discreto. y
3. $\iint_{xy} f(x, y) = 1$ para el caso continuo.

2.2. Momentos de la Distribución

Para describir una distribución se utilizan parámetros denominados *momentos*. Se pueden definir los momentos de una distribución respecto de un valor específico o respecto al origen. Se llama r-ésimo momento de una distribución respecto de \mathbf{b} , al valor de la esperanza matemática, si es que ésta existe, de la siguiente expresión:

$$E(x - b)^r = \sum_{i=1}^n (x_i - b)^r f(x_i),$$

en su versión discreta, y

$$E(x - b)^r = \int_{-\infty}^{\infty} (x - b)^r f(x) dx,$$

en su versión continua.

Por ejemplo, el r-ésimo momento de la variable aleatoria X respecto de su origen ($b = 0$), denotado por μ_r , es :

$$\mu_r = E(x^r) = \sum_{i=1}^n x_i^r f(x_i) \quad \text{si es una distribución discreta}$$

$$\mu_r = E(x^r) = \int_{-\infty}^{\infty} x^r f(x) dx \quad \text{si es una distribución continua}$$

El primer momento respecto al origen, es decir del cero, recibe el nombre de **esperanza matemática** de la variable aleatoria, y representa el valor promedio de esta variable. Se denota simplemente como:

$$\mu = E(x) = \sum_{i=1}^n x_i f(x_i) \quad \text{si es una distribución discreta}$$

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{si es una distribución continua}$$

En el ejemplo 1, la esperanza matemática de la distribución es

$$\mu = E(X) = \sum_{i=0}^2 x_i f(x_i) = 0 \cdot \left(\frac{1}{4}\right) + 1 \cdot \left(\frac{1}{2}\right) + 2 \cdot \left(\frac{1}{4}\right) = 1$$

Otros momentos de uso común son los momentos respecto de la media μ . Si existe la esperanza matemática de una variable aleatoria, se pueden definir momentos con respecto a la media de la forma:

$$m_r = \sum_{i=1}^n (x_i - \mu)^r f(x_i) \quad \text{si es una distribución discreta}$$

$$m_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \quad \text{si es una distribución continua}$$

El primer momento respecto de la media es cero para cualquier variable aleatoria. Pero mucho más importante es el segundo momento con respecto a la media, llamado **varianza**, denotado por

$$Var(x) = \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 f(x_i)$$

en forma discreta, o

$$Var(x) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

en forma continua.

La raíz cuadrada de la varianza se conoce como **desviación estándar o típica**.

$$\sigma = \sqrt{Var(x)}$$

La varianza de X en el ejemplo 1 es

$$\begin{aligned} Var(X) &= \sum_{i=0}^2 (x_i - \mu)^2 f(x_i) \\ &= (0 - 1)^2 \cdot \left(\frac{1}{4}\right) + (1 - 1)^2 \cdot \left(\frac{1}{2}\right) + (2 - 1)^2 \cdot \left(\frac{1}{4}\right) = \frac{1}{2} \end{aligned}$$

Por ende la desviación estándar o típica es

$$\sqrt{Var(X)} = \sqrt{\frac{1}{2}} \approx 0,7071$$

Además, existen los denominados productos de momentos de variables aleatorias. En este sentido puede definirse el siguiente momento:

$$m_{rs} = E[(x - E(x))^r (y - E(y))^s]$$

Un momento especial es el producto de los primeros momentos respecto de la media de cada variable aleatoria definido como:

$$\mu_{11} = E [(x - E(x))(y - E(y))] = E(xy) - E(x)E(y)$$

Por su importancia, este momento acostumbra a denotarse en forma especial como σ_{xy} y es conocido como la **covarianza** entre variables aleatorias. De tal forma

$$Cov(x, y) = \sigma_{xy} = E(xy) - E(x)E(y)$$

La covarianza mide el grado de asociación entre las variables, asociación que puede ser positiva o negativa. Para comprender este argumento considere otra forma de expresar la covarianza en el caso de variables discretas:

$$Cov(x, y) = \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) f(x_i, y_i)$$

en que μ_x , μ_y y $f(x_i, y_i)$ representan la media de la variable aleatoria X , la media de la variable aleatoria Y , y la función de probabilidad conjunta, respectivamente. Se puede constatar que si los valores de ambas variables están por sobre la media o ambos valores están bajo la media simultáneamente, el valor de la covarianza será positivo. Por el contrario si un valor está por sobre la media y el otro bajo la media, el valor de la covarianza será negativo.

A partir de los momentos mencionados anteriormente se puede definir un parámetro conocido como coeficiente de correlación entre X e Y :

$$\rho_{xy} = \frac{Cov(x, y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$$

que refleja el grado de variación conjunta (asociación lineal) entre dos variables en relación a una medida de la variación total de ambos. Si no existe correlación entre las variables aleatorias ρ debe tener un valor igual a cero.

Existen una serie de propiedades útiles de los momentos ². Considere las variables aleatorias x_i y las constantes a_j , donde $i, j = 1, \dots, n$, entonces:

1. $E(x_i + a_j) = E(x_i) + E(a_j) = \mu_i + a_j$

²Una descripción detallada de estas propiedades puede encontrarse en Mora, *et al.* (1996).

2. $E(a_j x_i) = a_j E(x_i) = a_j \mu_i$
3. $E(a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n) = a_0 + a_1 E(x_1) + a_2 E(x_2) + \dots + a_n E(x_n)$
4. $E(x_1 \times x_2 \cdots \times x_n) = E(x_1) \times E(x_2) \cdots \times E(x_n)$ si y solo si los x_i son independientes entre si.
5. $Var(x_i + a_j) = Var(x_i)$
6. $Var(a_j x_i) = a_j^2 Var(x_i)$
7. $Var(x_i) = E(x_i^2) - \mu_i^2$
8. $Var(a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n) = \sum_{i=1}^n a_i^2 Var(x_i) + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^i a_i a_j cov(x_i, x_j)$

También se pueden obtener momentos a partir de una muestra de una determinada variable aleatoria. Los momentos muestrales más importantes son la media muestral (\bar{X}) y la varianza muestral (s^2), definidos:

$$\bar{X} = \frac{\sum X_i}{n}$$

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

Puede definirse además, la desviación típica o estándar muestral, s , como la raíz cuadrada de la varianza muestral, que puede interpretarse como una medida de la distancia promedio entre la media y las observaciones.

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

2.3. Distribuciones de Probabilidad

Para poder aplicar los conceptos estadísticos revisados hasta aquí a la medición de procesos económicos específicos, es necesario determinar la función de distribución de probabilidad, $f(x)$, que se utilizará. Existen múltiples distribuciones de probabilidad disponibles, pero por diversas razones algunas se utilizan en forma más profusa que otras. Aquí revisaremos las relevantes para nuestros propósitos.

2.3.1. Distribución Normal

La distribución normal es una de las distribuciones más utilizadas en estadística y en la modelación de fenómenos económicos. Una razón fundamental para esto es la existencia del teorema del límite central, el cual plantea que cuando el tamaño de la muestra se incrementa la distribución de una variable aleatoria cualquiera tiende a comportarse de acuerdo a una distribución normal. Por lo tanto, basta con tener una muestra suficientemente grande como para sustentar el uso de una distribución normal.³ Además, muchas de las distribuciones en la naturaleza y en la sociedad son normales. Finalmente, es una distribución simple de caracterizar.

Una variable aleatoria continua que puede tomar cualquier valor en la recta de los números reales, tendrá distribución normal si su función de densidad de probabilidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(x_i - \mu)^2}{-2\sigma^2}\right],$$

donde μ es la esperanza de X que es un valor real, y σ^2 es la varianza. La notación que se utiliza en este caso para indicar que X se distribuye en forma normal con media μ y varianza σ^2 es $X \sim N(\mu, \sigma^2)$.

El gráfico de la función de densidad normal es en forma de campana, y es simétrico con respecto a μ (media), donde alcanza su punto máximo. Además, posee dos puntos de inflexión en los valores $\mu + \sigma$ y $\mu - \sigma$.

Un caso particular de esta función es la **función normal estándar** (figura 2.1), cuya media es **cero** y su varianza es **uno**. Esto se indica $X \sim N(0, 1)$. La función de densidad normal estándar es:

$$f(x/0, 1) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right],$$

y su función de distribución acumulada es:

$$F(x/0, 1) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] dz$$

³Una versión formal del teorema del límite central es como sigue: Si \bar{X} es la media de una muestra aleatoria X_1, X_2, \dots, X_n de tamaño n de una distribución de media finita μ y varianza positiva finita σ^2 , entonces la distribución de $W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ tiende a distribuirse asintóticamente $N(0, 1)$. Ver Hogg y Tanis (1983).

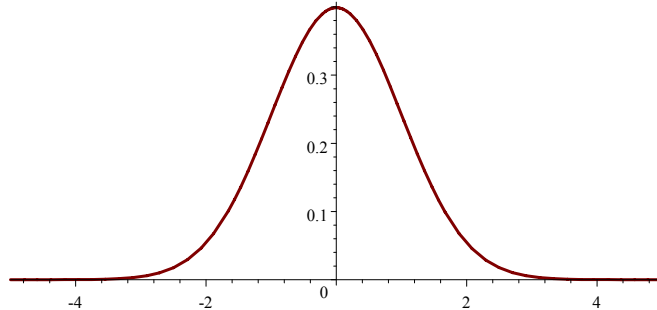


Figura 2.1: **Gráfica de la Distribución Normal Estándar**

Esta es una integral que sólo se puede calcular a través de métodos de aproximación y generalmente se encuentra tabulada en todos los libros de estadística.

En los casos en que la variable aleatoria se distribuye normal pero con media μ y varianza σ^2 , se puede estandarizar la distribución restando a cada observación la media de la distribución y dividir el resultado por la desviación estándar. Es decir:

$$z = \frac{x_i - \mu}{\sigma}$$

Esto permite el uso de la función de distribución y densidad de probabilidad normal estándar, sin pérdida de generalidad. En otras palabras, a partir de cualquier distribución normal se puede llegar a una distribución normal estándar.

Existen tres distribuciones de probabilidad específicas asociadas a la distribución normal y que serán utilizadas para inferencia estadística. Estas son conocidas con el nombre de Chi-Cuadrado (χ^2), **t** de Student (**t**) y **F** de Fisher (**F**).

2.3.2. Distribución Chi - Cuadrado

Considere una variable aleatoria X tal que $X \sim N(\mu, \sigma^2)$. Si se define $z = \frac{x - \mu}{\sigma}$, entonces $z \sim N(0, 1)$. Definamos una nueva variable aleatoria como $\chi_{(1)}^2 = z^2 = \left(\frac{x - \mu}{\sigma}\right)^2$ entonces decimos que χ^2 se distribuye Chi-

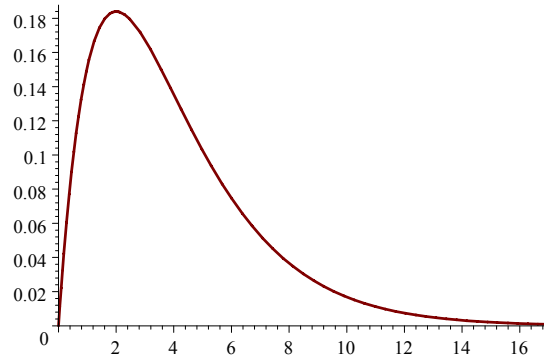


Figura 2.2: Gráfica de una Distribución Chi-Cuadrado

Cuadrado con 1 grado de libertad.

Si X_1, \dots, X_n son n variables aleatorias independientes e idénticamente distribuidas (i.i.d.) en forma normal estándar, y sea $u = \sum_{i=1}^n X_i^2$. Entonces u se distribuye Chi-cuadrado con n grados de libertad.

Lo que se quiere expresar, es que la suma de variables Chi-cuadrado tiene también una distribución Chi-cuadrado, y sus grados de libertad son igual al número de variables sumadas.

En general, si X_1, \dots, X_n son variables independientes distribuidas Chi-cuadrado con v_1, v_2, \dots, v_n grados de libertad respectivamente, entonces si $y = \sum_{i=1}^n X_i$, se dirá que y se distribuye Chi-cuadrado con $\sum_{i=1}^n v_i$ grados de libertad. El gráfico de la función de distribución Chi-cuadrado, es como el que se muestra en la figura 2.2. Como puede constatarse, la distribución χ^2 no es simétrica, como lo es la distribución normal. Además, la distribución χ^2 concentra su masa a la izquierda de su mediana.

2.3.3. Distribución t de Student

Sea X_1 una variable aleatoria normal estándar y X_2 una variable aleatoria distribuida Chi-cuadrado con v grados de libertad. Si ambas variables son independientes, entonces la variable aleatoria T :

$$T = \frac{X_1}{\sqrt{X_2/v}},$$

tiene una distribución **t de Student** con v grados de libertad y se denota por $(T \sim t(v))$. Su uso corresponde a casos donde es posible aplicar la distribución normal, pero bajo la restricción que se desconoce el valor de la verdadera varianza poblacional (σ^2).

2.3.4. Distribución F de Fisher:

Sean X_1 y X_2 variables aleatorias independientes distribuidas Chi - Cuadrado con v_1 y v_2 grados de libertad respectivamente, entonces la variable aleatoria Y :

$$Y = \frac{X_1/v_1}{X_2/v_2},$$

tiene distribución **F de Fisher** con v_1 grados de libertad en el numerador y v_2 grados de libertad en el denominador, es decir, $Y \sim F(v_1, v_2)$. Es posible demostrar que una distribución F con $(1, n)$ grados de libertad es equivalente al cuadrado de una distribución t con n grados de libertad.

Tanto para el caso de la distribución normal, como para los otros tres casos (chi-cuadrado, t-student, F-Fisher), existen tablas que muestran sus resultados. Estas tablas permiten realizar diversos tipos de juicios sobre el valor de los parámetros, lo que se discutirá en más detalle en la sección de pruebas de hipótesis.

2.4. Inferencia y Estimación

El propósito de la inferencia es utilizar la información contenida en una muestra para obtener conclusiones sobre los parámetros que representan a la población. Debido a que los parámetros originales, que caracterizan a la población son desconocidos, se utiliza una muestra con el fin de obtener una idea de cual es el valor y comportamiento del parámetro poblacional.

Dos tipos de inferencia son importantes desde el punto de vista econométrico: la **Estimación** y las **Pruebas de Hipótesis**. Para el primer caso, lo que interesa es calcular valores lo más cercanos posibles a los verdaderos parámetros poblacionales que permitan explicar el comportamiento de los agentes económicos. La economía trabaja con modelos que representan relaciones de comportamiento en los cuales se observan variables endógenas y variables exógenas que se relacionan a través de parámetros. Por ejemplo, considere la

relación entre el consumo y el ingreso dada por:

$$C = a + bY$$

donde C es el consumo, Y es el ingreso nacional y a y b son los parámetros del modelo que se interpretan como el consumo autónomo y la propensión marginal a consumir respectivamente. El objetivo de la econometría es estimar el valor de estos parámetros (a y b) a partir de información contenida en una muestra en la que recolectamos diversas observaciones de consumo e ingreso de distintos países o individuos, o simplemente en distintos momentos del tiempo.

En el segundo caso nos interesa realizar pruebas de hipótesis, es decir, analizar el rango de valores posibles que pueden tomar los parámetros poblacionales. En nuestro ejemplo es muy importante que la propensión marginal al consumo esté entre los valores $0 < b < 1$, ya que de otra forma carecería de interpretación económica. Note además, que si el valor es cercano a uno implica que los individuos gastan casi todo su ingreso marginal, dejando muy poco para el ahorro.

Dentro de las pruebas más importantes se encuentran las pruebas de significancia estadística, donde se verifica si con la información que se dispone se puede concluir que el valor obtenido es relevante desde el punto de vista estadístico. Como veremos más adelante, los valores obtenidos para nuestros parámetros (a y b en nuestro ejemplo) son variables aleatorias. Por lo tanto, no se puede descartar a priori que su valor sea cero. Por esta razón una de las pruebas más importantes es probar si el valor del parámetro obtenido es distinto de cero en términos estadísticos. Si no podemos rechazar esta hipótesis, ello implica que la variable que acompaña a este parámetro no ayuda a explicar el comportamiento de nuestro modelo.

Otros ejemplos de pruebas de hipótesis son la verificación de la elasticidad unitaria de la función de demanda por un determinado producto, o bien hipótesis de rendimientos constantes a escala para funciones de producción del tipo Cobb-Douglas. Veamos en detalle este último ejemplo. Considere la función de producción

$$Y = AK^\alpha L^\beta$$

Donde Y es el nivel de producto, A es el componente de tecnología, K representa el capital y L el trabajo. Para estimar los valores de α y β , se usa

una muestra de observaciones de producto, capital y trabajo de distintas empresas, o bien de una misma empresa en distintos periodos del tiempo. Para proceder con la estimación se utiliza la expresión de la función de producción en logaritmos:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L$$

Suponga que se desea verificar la existencia de rendimientos constantes a escala en la función de producción anterior. Para tal efecto, la prueba de hipótesis de rendimientos constantes a escala implica probar que $\alpha + \beta = 1$. El procedimiento de prueba de hipótesis debe analizar si la suma de los parámetros es estadísticamente distinto de uno o no.

La necesidad de realizar pruebas de hipótesis se justifica en el hecho que los estimadores de los parámetros poblacionales son aleatorios en el sentido que dependen de la información contenida en la muestra. Si la muestra cambia, entonces el valor estimado también se modifica. En otras palabras, el valor obtenido asume un valor específico con alguna probabilidad asociada, y no existe razón para descartar que tome otro valor cualquiera. Entonces, es relevante verificar si los parámetros obtenidos son cercanos estadísticamente a valores esperados por la teoría económica.

A continuación se explicará las formas típicas de obtención de los estimadores de los parámetros poblacionales. En la siguiente sección se analizará la construcción de Pruebas de Hipótesis.

2.4.1. El problema de la estimación

Previo a la discusión de los métodos de estimación es necesario precisar qué se entiende por un estimador. Sea X una variable aleatoria con función de densidad de probabilidad $f(X, \theta)$ donde θ es un parámetro desconocido de la población que se desea estimar. Si X_1, \dots, X_n es una muestra aleatoria de la población podemos definir un estimador como una función o regla de la forma:

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$$

donde el gorro sobre la variable denota que se trata de un estimador del parámetro poblacional. El estimador es una variable aleatoria que depende de la muestra observada. Un valor específico se le denomina **valor estimado**. En el ejemplo de la función de consumo el estimador del parámetro b , la

propensión marginal del consumo, va a depender de las observaciones del consumo y el ingreso. Es decir,

$$\hat{b} = \hat{b}(C_1, C_2, \dots, C_n, Y_1, Y_2, \dots, Y_n)$$

2.4.2. Métodos de Estimación

Para la estimación de los parámetros existen varios métodos. El texto se concentrará básicamente en el **Método de Máxima Verosimilitud** y el **Método de Mínimos Cuadrados**.

Método de Máxima Verosimilitud.

Este método elige los valores para los parámetros de tal forma que éstos maximicen la probabilidad de seleccionar aleatoriamente la muestra que se tiene actualmente. Alternativamente podemos definir un estimador Máximo Verosímil (**MV**) como aquél valor del parámetro para los cuales la muestra observada es la más probable de haber ocurrido⁴.

Asumiendo que las observaciones en la muestra son independientes y aleatoriamente escogidas, la función de verosimilitud será la función de **probabilidad conjunta** de la muestra. Entonces, para valores observados X_1, \dots, X_n de una variable aleatoria X , con una determinada función de $f(X, \theta)$, que depende de un vector de parámetros desconocido θ , la función de verosimilitud está definida por:

$$L(\theta) = \prod_{i=1}^n f(X, \theta)$$

El valor del vector $\hat{\theta}$ se obtiene por los procedimientos convencionales de maximización con:

$$\left. \frac{\partial L(\theta)}{\partial \hat{\theta}} \right|_{\theta=\hat{\theta}} = 0$$

Para la función de verosimilitud los valores que la maximizan son los mismos que maximizan la función expresada en logaritmos, es decir

$$\ln L(\theta) = \sum_{i=1}^n \ln f(X, \theta)$$

⁴O bien un estimador **MV** es aquel valor de θ que tenga la mayor probabilidad de generar la muestra observada.

y el problema de optimización queda dado por:

$$\underset{\hat{\theta}}{Max} \ln L(\theta) = \underset{\hat{\theta}}{Max} \sum_{i=1}^n \ln f(X, \theta),$$

cuya condición de primer orden⁵ es:

$$\left. \frac{\partial \ln L(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0$$

Ejemplo 1

Supongamos que se tiene una muestra aleatoria de una variable X distribuida normal con media μ y varianza σ^2 , y queremos encontrar el vector de parámetros estimados para θ que, en este ejemplo, corresponde a los estimadores de μ y σ^2 . La función de verosimilitud queda definida por:

$$L(\mu, \sigma^2) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[\frac{(x_i - \mu)^2}{-2\sigma^2} \right]$$

Aplicando logaritmo natural y simplificando se tiene que:

$$\begin{aligned} \ln L(\mu, \sigma^2) &= \sum_{i=1}^n \ln \left\{ \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[\frac{(x_i - \mu)^2}{-2\sigma^2} \right] \right\} \\ \ln L(\mu, \sigma^2) &= \sum_{i=1}^n \left\{ \ln 1 - \frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2} \right\} \\ \ln L(\mu, \sigma^2) &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

Luego obtenemos las primeras derivadas con respecto a μ y σ^2 .

$$\begin{aligned} \frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} &= -\frac{1}{2\sigma^2} (2) \sum_{i=1}^n (x_i - \mu) (-1) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi - \sum_{i=1}^n (x_i - \mu)^2 \left(\frac{1}{2} \right) (-1) (\sigma^2)^{-2} = 0 \\ &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2} = 0 \end{aligned}$$

⁵Naturalmente se requiere que las condiciones de segundo orden se cumplan para que la solución efectivamente sea un máximo. En los ejemplos discutidos aquí, esto es así.

Resolviendo estas ecuaciones se obtiene que:

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Vemos que, en este caso el estimador de la media es la media muestral y de la varianza es la varianza muestral.

Ejemplo 2

Consideremos la variable discreta y_i , cuya distribución es Bernoulli⁶, con probabilidad de éxito p . Se busca un estimador para este parámetro. En este caso la función de Verosimilitud está dada por la siguiente expresión:

$$L(p) = \prod_{i=1}^n (p)^{y_i} (1-p)^{1-y_i}$$

$$L(p) = (p)^{\sum_{i=1}^n y_i} (1-p)^{\sum_{i=1}^n (1-y_i)}$$

Aplicando logaritmo natural, y sus propiedades, se obtiene:

$$\ln L(p) = \sum_{i=1}^n y_i \ln(p) + \sum_{i=1}^n (1-y_i) \ln(1-p)$$

A partir de la ecuación anterior, se obtiene la primera derivada de la función **MV**:

$$\frac{\partial \ln L(p)}{\partial p} = \sum_{i=1}^n y_i \frac{1}{p} - \sum_{i=1}^n (1-y_i) \frac{1}{1-p}$$

⁶Un experimento del tipo Bernoulli es un experimento aleatorio, en que el resultado puede ser clasificado en sólo dos eventos mutuamente excluyentes, comúnmente denominados éxito y fracaso, con $y_i = 1$ para el éxito y $y_i = 0$ para fracaso. La probabilidad de éxito es definida por p y la de fracaso $1-p$. Luego la función de probabilidad puede escribirse como: $f(y_i) = p^{y_i} (1-p)^{1-y_i}$.

Simplificando e igualando a cero, para poder despejar p :

$$\begin{aligned} \frac{\partial \ln L}{\partial p} &= \sum_{i=1}^n y_i \frac{1}{\hat{p}} - \sum_{i=1}^n (1 - y_i) \frac{1}{1 - \hat{p}} = 0 \\ \frac{\partial \ln L}{\partial p} &= \frac{\sum_{i=1}^n y_i}{\hat{p}} - \sum_{i=1}^n \left(\frac{1}{1 - \hat{p}} \right) + \frac{\sum_{i=1}^n y_i}{1 - \hat{p}} = 0 \\ \frac{\partial \ln L}{\partial p} &= \sum_{i=1}^n y_i - \frac{n\hat{p}}{1 - \hat{p}} + \frac{\hat{p} \sum_{i=1}^n y_i}{1 - \hat{p}} = 0 \\ \frac{\partial \ln L}{\partial p} &= (1 - \hat{p}) \sum_{i=1}^n y_i - n\hat{p} + \hat{p} \sum_{i=1}^n y_i = 0 \\ \frac{\partial \ln L}{\partial p} &= \sum_{i=1}^n y_i - n\hat{p} = 0 \end{aligned}$$

Luego p está dado por:

$$\hat{p} = \frac{\sum_{i=1}^n y_i}{n} \quad (2.2)$$

donde el estimador de la probabilidad de éxito p está dado por la frecuencia de aciertos.

En este caso se tiene que cumplir la condición de segundo orden para que la solución corresponda efectivamente a un máximo. Es decir, la segunda derivada de la función verosimilitud debe ser menor que cero:

$$\frac{\partial^2 \ln L}{\partial p^2} = -n < 0$$

Método de Mínimos Cuadrados.

La idea de este método consiste en minimizar la suma de los residuos al cuadrado. Los residuos se definen como la diferencia entre el valor observado de una determinada variable y su valor esperado. El estimador Mínimo Cuadrático (**MC**) es aquel estimador $\hat{\theta}$ que minimiza la diferencia al cuadrado entre el valor observado y el valor esperado.

$$\min_{\hat{\theta}} Q = \min \sum \left[X_i - E \left(X_i \mid \hat{\theta} \right) \right]^2$$

La idea es que los parámetros θ se elijan de tal forma que esta sumatoria sea el mínimo valor posible.

Ejemplo 3

Asumamos una variable aleatoria Y , para la cual se desea estimar la media poblacional (μ). Para una muestra dada de observaciones de la variable aleatoria, el método de Mínimos Cuadrados opera minimizando la sumatoria de las diferencias al cuadrado entre el valor observado y la media. Esto se escribe como:

$$\begin{aligned}\text{mín } Q &= \text{mín } \sum (Y_i - \mu)^2 \\ \frac{\partial \text{mín } Q}{\partial \hat{\mu}} &= 2 \sum (Y_i - \hat{\mu}) (-1) = 0\end{aligned}$$

El resultado es⁷:

$$\hat{\mu} = \frac{\sum Y_i}{n}$$

Este nos dice que se utiliza la media muestral como estimador de la media poblacional.

Dos aspectos son importantes de considerar. Primero, los estimadores llevan un “gorro” con el fin de distinguirlos de los verdaderos parámetros poblacionales. Esto además, debe indicarnos que los estimadores son una variable aleatoria, que depende de la muestra con la que se está trabajando. Segundo, a diferencia del método de Máxima Verosimilitud, el método de Mínimos Cuadrados no requiere supuestos sobre la distribución de probabilidad de la variable aleatoria, para obtener el parámetro estimado. Esto sólo se requerirá en el momento de realizar pruebas de hipótesis.

2.4.3. Propiedades Deseadas de los Estimadores.

Los métodos revisados entregan estimadores para los parámetros poblacionales. Sin embargo, no sabemos que tan adecuados son estos estimadores. Es decir, cómo se comportan en relación a los *verdaderos* parámetros de la población.

⁷Es fácil observar en este caso que $\partial^2 \text{mín } Q / \partial \hat{\mu}^2$ es positivo, lo cual se requiere para que efectivamente se trate de una solución mínima.

Existe una serie de propiedades que son deseables en los estimadores, independiente del método por el cual se obtuvieron, ya que estas propiedades reducen los niveles de error entre el verdadero parámetro poblacional y el estimador. En la literatura se distingue entre las propiedades asociadas a muestras pequeñas y las propiedades de muestra grande o asintóticas. Las primeras se cumplen independientemente del tamaño de la muestra seleccionada, mientras que las propiedades asintóticas se cumplen en el límite, es decir, cuando el tamaño de la muestra tiende a infinito.

Las propiedades de los estimadores ayudan a seleccionar un estimador, cuando se cuenta con un grupo de estimadores. En otras palabras, son criterios que permiten decidir cual estimador es mejor como representante del parámetro poblacional desconocido.

Propiedades de Muestra Pequeña.

Las principales propiedades que vamos a buscar en un estimador obtenido de una muestra pequeña son las de **insesgamiento** y **eficiencia**. Por **insesgamiento** entenderemos que la media o esperanza del estimador sea igual al valor del parámetro poblacional. Matemáticamente se puede expresar como:

$$E(\hat{\theta}) = \theta$$

Esto significa que en el promedio el estimador es igual al verdadero parámetro poblacional. Es decir, si el experimento se repite infinitas veces, obteniendo muestras de tamaño n , el valor promedio del estimador para todas las muestras será θ . Por el contrario, si la esperanza del estimador es $E(\hat{\theta}) = \theta + \delta$, donde $\delta \neq 0$, se dice que el estimador es sesgado, con sesgo igual a δ .

Veamos un ejemplo de insesgamiento para el caso discreto de la variable distribuida Bernoulli, con probabilidad de éxito \mathbf{p} presentada anteriormente (ver Ejemplo 2). En este caso se aplica el operador de esperanza (E) al estimador del parámetro, es decir:

$$E(\hat{p}) = E\left(\frac{\sum_{i=1}^n y_i}{n}\right) = \sum_{i=1}^n \frac{E(y_i)}{n}$$

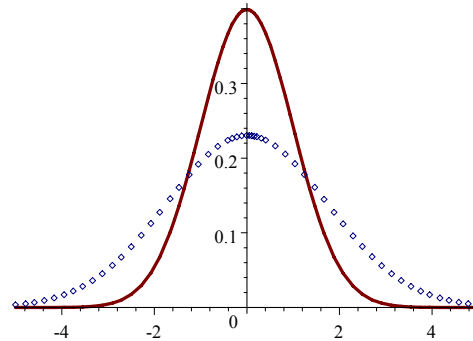


Figura 2.3: **Varianzas de la Distribución:** la curva de línea punteada está más dispersa alrededor de la media que la distribución de línea continua. Es decir, la varianza de la primera es mayor que la de la última.

Por la definición de esperanza sabemos que $E(y_i) = 1 \cdot p + 0 \cdot (1 - p) = p$. Por lo tanto, por la sumatoria obtenemos:

$$E(\hat{p}) = \frac{np}{n} = p$$

Es decir, el valor esperado del estimador \hat{p} corresponde al valor poblacional. De esta forma se puede observar que este estimador de \mathbf{p} es insesgado.

La **eficiencia** es un concepto que se aplica a estimadores insesgados y se refiere a que el estimador tenga la mínima varianza posible. Esto quiere decir que en repetidas muestras los valores de los estimadores tenderán a concentrarse en torno al valor del parámetro poblacional. Entre muchos estimadores se preferirá aquel que tenga la varianza tan pequeña como sea posible. Como observamos en la figura 2.3 dados dos estimadores diferentes (ambos insesgados) θ_1 y θ_2 , es preferible aquella distribución que es más estrecha $f(\theta_2)$ en torno al valor de la media, puesto que mientras menor sea la varianza, menor será la amplitud en torno a la media y por lo tanto mayor será la precisión del estimador.

Para probar que un estimador posee mínima varianza existe un teorema conocido como **Teorema de la Cota Inferior de Cramer Rao**⁸, el cual establece que la varianza de un estimador $\hat{\theta}$ será al menos tan grande como

⁸Cramer Rao Lower Bound.

el inverso negativo del valor esperado de la segunda derivada de la función de verosimilitud respecto al estimador. Por lo tanto, la mínima varianza se obtiene cuando se cumple la igualdad. Si algún estimador cumple con esta condición es eficiente. Este teorema se expresa de la siguiente forma:

$$\text{var}(\hat{\theta}) \geq -\frac{1}{E\left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}\right)}$$

Donde L es la función de verosimilitud. Este teorema proporciona una cota inferior para la varianza de cualquier estimador del parámetro.

Para el caso de muchos parámetros con función de verosimilitud dada por:

$$f(X|\theta) = f(X_1, \dots, X_n|\theta) = \prod_{i=1}^n f_i(X|\theta)$$

La matriz de esperanzas negativas de segundas derivadas de la función de verosimilitud queda definida por⁹:

$$\begin{aligned} I(\hat{\theta}) &= -E\left(\frac{\partial^2 L^*}{\partial \theta \partial \theta'}\right) \\ &= \begin{bmatrix} -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_1^2}\right) & -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_1 \partial \hat{\theta}_2}\right) & \dots & -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_1 \partial \hat{\theta}_n}\right) \\ -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_2 \partial \hat{\theta}_1}\right) & -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_2^2}\right) & \dots & -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_2 \partial \hat{\theta}_n}\right) \\ \vdots & \vdots & & \vdots \\ -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_n \partial \hat{\theta}_1}\right) & -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_n \partial \hat{\theta}_2}\right) & \dots & -E\left(\frac{\partial^2 \ln L^*}{\partial \hat{\theta}_n^2}\right) \end{bmatrix} \end{aligned}$$

Esta matriz nos entrega las mínimas varianzas de los estimadores máximo verosímiles ubicados sobre la diagonal principal, y fuera de ella las covarianzas de los mismos.

Ejemplo 4:

⁹Esta matriz se denomina comúnmente **matriz información**.

Para entender el concepto considere el ejemplo 1 dado anteriormente, donde la variable aleatoria X se distribuye normal con media μ y varianza σ^2 . En este caso, tenemos que considerar las varianzas de dos estimadores, $\hat{\mu}$ y $\hat{\sigma}^2$. Las primeras derivadas de la función de verosimilitud eran:

$$\begin{aligned}\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ \frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2}\end{aligned}$$

Las segundas derivadas de la función de verosimilitud son:

$$\begin{aligned}\frac{\partial^2 \ln L(\mu, \sigma^2)}{\partial \mu^2} &= \frac{1}{\sigma^2} \sum_{i=1}^n (-1) = -\frac{n}{\sigma^2} \\ \frac{\partial^2 \ln L(\mu, \sigma^2)}{\partial (\sigma^2)^2} &= \frac{n}{2(\sigma^2)^2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{(\sigma^2)^3} \\ \frac{\partial^2 \ln L(\mu, \sigma^2)}{\partial \mu \partial (\sigma^2)} &= \frac{\partial^2 \ln L(\mu, \sigma^2)}{\partial (\sigma^2) \partial \mu} = -\frac{\sum_{i=1}^n (x_i - \mu)}{(\sigma^2)^2}\end{aligned}$$

Que en términos matriciales se transforma en:

$$\begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{\sum_{i=1}^n (x_i - \mu)}{(\sigma^2)^2} \\ -\frac{\sum_{i=1}^n (x_i - \mu)}{(\sigma^2)^2} & \frac{n}{2(\sigma^2)^2} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{(\sigma^2)^3} \end{bmatrix}$$

Note que

$$\sum_{i=1}^n (x_i - \mu) = 0$$

y

$$\begin{aligned} \frac{\sum_{i=1}^n (x_i - \mu)^2}{(\sigma^2)^3} &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{\frac{n}{(\sigma^2)^3}} \\ &= \frac{n}{(\sigma^2)^2} \end{aligned}$$

El paso siguiente es calcular el negativo de la esperanza de esta matriz, lo que implica aplicar el operador esperanza a cada elemento de la matriz y cambiar el signo. Recordando las propiedades de los momentos la matriz se transforma en:

$$I(\hat{\theta}) = -E\left(\frac{\partial^2 L^*}{\partial\theta\partial\theta'}\right) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix}$$

Por último, se debe calcular la inversa de esta matriz, que es:

$$\begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix}^{-1} = \frac{\begin{bmatrix} \frac{n}{2(\sigma^2)^2} & 0 \\ 0 & \frac{n}{\sigma^2} \end{bmatrix}}{\frac{n^2}{2(\sigma^2)^3}} = \begin{bmatrix} \frac{1}{n}\sigma^2 & 0 \\ 0 & \frac{2(\sigma^2)^2}{n} \end{bmatrix}$$

En resumen, la mínima varianza que puede obtener un estimador de μ es $\frac{\sigma^2}{n}$, mientras que para el estimador de σ^2 es $\frac{2(\sigma^2)^2}{n}$. Con respecto a la covarianza entre los estimadores de μ y σ^2 se puede decir que el mínimo valor que puede tomar es cero.

Ejemplo 5

Veamos ahora el caso de la cota inferior para la variable discreta distribuida Bernoulli (ver Ejemplo 2) con probabilidad de éxito \mathbf{p} . Calculando

la varianza del estimador de \mathbf{p} , se tiene que:

$$\text{var}(\hat{p}) = \text{var}\left(\frac{\sum_{i=1}^n y_i}{n}\right)$$

$$\text{donde la } \text{var}\left(\sum_{i=1}^n y_i\right) = \sum_{i=1}^n \text{var}(y_i).$$

$$\text{Entonces, } \text{var}(\hat{p}) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(y_i)$$

$$\text{var}(\hat{p}) = \frac{1}{n^2} \cdot n \cdot p \cdot (1-p)$$

$$\text{var}(\hat{p}) = \frac{p(1-p)}{n}$$

El resultado de la segunda línea se deriva de: $\text{var}(y_i) = E(y_i^2) - [E(y_i)]^2$, además la $E(y_i) = 1 \cdot p + 0 \cdot (1-p) = p$ y $E(y_i^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$, por lo tanto la $\text{var}(y_i) = p - p^2 = p \cdot (1-p)$.

Recordemos que la función *MV*, para este caso es:

$$L(y_i|p) = \prod_{i=1}^n (p)^{y_i} (1-p)^{1-y_i}$$

Al aplicar logaritmo natural, con sus propiedades, se obtenía:

$$\ln L(y_i|p) = \sum_{i=1}^n y_i \ln(p) + \sum_{i=1}^n (1-y_i) \times \ln(1-p)$$

Luego, la primera derivada de la función de verosimilitud, con respecto a \mathbf{p} , es:

$$\frac{\partial \ln L}{\partial p} = \sum_{i=1}^n \frac{y_i}{p} - \sum_{i=1}^n \frac{(1-y_i)}{1-p}$$

A partir de esto obtenemos la segunda derivada con respecto a \mathbf{p} :

$$\frac{\partial^2 \ln L}{\partial p^2} = -\frac{1}{p^2} \sum_{i=1}^n y_i - \frac{1}{(1-p)^2} \sum_{i=1}^n (1-y_i)$$

Aplicando esperanza, multiplicado por -1 y simplificando se tiene:

$$\begin{aligned}
 -E\left(\frac{\partial^2 \ln L}{\partial p^2}\right) &= -E\left[-\frac{1}{p^2} \sum_{i=1}^n y_i - \frac{1}{(1-p)^2} \sum_{i=1}^n (1-y_i)\right] \\
 -E\left(\frac{\partial^2 \ln L}{\partial p^2}\right) &= \frac{\sum_{i=1}^n E(y_i)}{p^2} + \frac{\sum_{i=1}^n E(1-y_i)}{(1-p)^2} \\
 &= \frac{pn}{p^2} + \frac{\sum_{i=1}^n 1 - \sum_{i=1}^n E(y_i)}{(1-p)^2} \\
 &= \frac{n}{p} + \frac{n-pn}{(1-p)^2} = \frac{n}{p} + \frac{n(1-p)}{(1-p)^2} \\
 &= \frac{n}{p} + \frac{n}{1-p} = \frac{n(1-p) + pn}{p(1-p)} \\
 -E\left(\frac{\partial^2 \ln L}{\partial p^2}\right) &= \frac{n}{p(1-p)}
 \end{aligned}$$

Luego, podemos concluir que la cota inferior para la varianza de \mathbf{p} es:

$$-\frac{1}{E\left(\frac{\partial^2 \ln L}{\partial p^2}\right)} = \frac{p(1-p)}{n}$$

Hasta el momento hemos considerado la situación en que se debe seleccionar entre dos estimadores insesgados, para ello se sugiere utilizar el criterio de mínima varianza. Sin embargo, un problema especial surge cuando se quiere seleccionar entre estimadores que no son insesgados. Existen situaciones en que no se pueden obtener estimadores insesgados, por lo que el criterio de eficiencia no es aplicable para la selección entre éstos. Además pueden existir estimadores sesgados con varianza menor que los insesgados. Existe un trade-off entre insesgamiento y precisión: ¿Es preferible un estimador cuyo valor esperado sea igual al parámetro poblacional, aunque tenga una gran varianza, o un estimador que fluctúe poco alrededor de su valor esperado, aunque éste último difiera del parámetro poblacional?. En este caso existe un criterio que consiste en elegir aquel estimador que posea un menor **Error Cuadrático Medio** (*ECM*), el cual se define como:

$$ECM = E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'$$

donde $\hat{\boldsymbol{\theta}}$ y $\boldsymbol{\theta}$ son el vector de parámetros estimados y poblacionales respectivamente.

Desarrollando se tiene:

$$\begin{aligned} ECM &= E \left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \right] \\ &= E \left[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}) + E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}) + E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})' \right] \\ &= E \left[\left\{ (\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})) + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}) \right\} \left\{ (\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})) + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}) \right\}' \right] \\ &= E \left[\left\{ (\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})) + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}) \right\} \left\{ (\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))' + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})' \right\} \right] \\ &= E \left[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))' \right] + E \left[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})' \right] \\ &\quad + E \left[(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))' \right] + E \left[(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})' \right] \\ &= E \left[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))' \right] + E \left[(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})' \right] \\ &= \text{var}(\hat{\boldsymbol{\theta}}) + (\text{sesgo}(\hat{\boldsymbol{\theta}}))(\text{sesgo}(\hat{\boldsymbol{\theta}}))' \end{aligned}$$

Y para el caso particular de un solo parámetro el Error Cuadrático Medio es¹⁰:

$$ECM = \text{var}(\hat{\theta}) + (\text{sesgo}(\hat{\theta}))^2$$

Del desarrollo se puede concluir que el mínimo error cuadrático medio es un criterio que toma en consideración tanto el sesgo como la precisión para el estimador y selecciona aquel que tenga el menor valor. Note que el sesgo y la precisión, medida por la varianza, están ponderadas de igual forma en el cálculo del ECM.

¹⁰Donde se ha hecho uso de las propiedades del operador esperanza. Además, el segundo y el tercer término de ésta expresión son nulos, como puede comprobarse al expandir y sacar la esperanza de cada uno de éstos términos.

Propiedades de Muestra Grande

Para el propósito del texto nos concentraremos en la propiedad de **consistencia** para estimadores obtenidos de muestras grandes¹¹. Esta es una propiedad asintótica, ya que describe una condición que se da en el límite de la distribución de probabilidades del estimador, cuando el tamaño de la muestra tiende a infinito.

Un estimador se considerará **consistente** si:

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P \left[\left| \hat{\theta}_n - \theta \right| < \varepsilon \right] = 1$$

Esto quiere decir, que en la medida que aumenta el tamaño de la muestra n , la probabilidad de la diferencia absoluta entre el estimador y el verdadero parámetro sea menor que un número ε , será uno, donde ε es un número arbitrario que puede ser infinitamente pequeño. Esto quiere decir que si la muestra es suficientemente grande el valor del estimador tenderá a ser igual al valor del parámetro poblacional con certeza. Existen dos condiciones suficientes pero no necesarias, para que un estimador sea **consistente**, estas son:

1. $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$
2. $\lim_{n \rightarrow \infty} var(\hat{\theta}) = 0$

En otras palabras, si estas 2 condiciones se cumplen el estimador será consistente.

2.5. Intervalos de Confianza y test de Hipótesis

Si se considera que los estimadores son variables aleatorias, entonces es lógico preguntarse por la utilidad de los valores obtenidos en el proceso de estimación. En otras palabras, el estimador depende de la muestra disponible, por lo tanto si nosotros cambiamos la muestra, entonces también debería cambiar nuestro estimador. Si esto es cierto, existe una variedad de valores

¹¹Otras propiedades asintóticas son el insesgamiento asintótico y la eficiencia asintótica. Ver mas detalle en Greene (1998).

que el estimador puede tomar considerando distintas muestras y por lo tanto no sabemos cuanto se puede confiar en estos valores. Esto es muy importante cuando, por ejemplo, los valores obtenidos se pretenden usar con fines de diseño de política económica.

Afortunadamente, en las estimaciones realizadas (es decir en el valor del estimador y de su respectiva varianza) subyace mucha información que puede ser útil para afrontar estos problemas. Una alternativa para enfrentar este problema de información es el enfoque denominado **Estimación por Intervalos**, que consiste en construir un **Intervalo de Confianza** para el parámetro de interés.

El intervalo de confianza refleja un rango dentro del cual se espera que fluctue el valor del parámetro. Comúnmente se usa un 95 % de probabilidad. Esto implica que si se construye infinitas veces un intervalo de confianza, en el 95 % de los casos el intervalo contendrá al parámetro poblacional.

Supongamos que se desea encontrar un intervalo de confianza para la media de una variable aleatoria distribuida normal con media μ y varianza σ^2 . De una muestra se obtiene que la esperanza es \bar{x} y su varianza es s^2 . El intervalo de confianza se define de la siguiente manera:

$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

donde Z corresponde a una variable distribuida normal estándar y α es el nivel de significancia y $-z_{1-\frac{\alpha}{2}}$ y $z_{1-\frac{\alpha}{2}}$ son el límite inferior y superior del intervalo, respectivamente. Por ejemplo para un 95 % de confianza, $\alpha = 0,05$ ($1 - 0,05 = 0,95$).

Luego, recordemos que por hipótesis:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Estandarizando la variable aleatoria se tiene:

$$z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

Luego, reemplazando la variable normal estándar, en la definición del intervalo de confianza, y desarrollando

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(-z_{1-\frac{\alpha}{2}} \cdot \sqrt{\sigma^2/n} < \bar{x} - \mu < z_{1-\frac{\alpha}{2}} \cdot \sqrt{\sigma^2/n}\right) = 1 - \alpha$$

Por último, se tiene que el intervalo de confianza es:

$$P\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\sigma^2/n} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\sigma^2/n}\right) = 1 - \alpha \quad (2.3)$$

Note que éste es el caso para la media de una distribución normal. De tal manera que para identificar los límites del intervalo se requiere leerlos de una tabla de distribución normal. Para cada caso se requiere conocer la distribución de probabilidad del estimador, y usar por lo tanto una distribución adecuada para construir el intervalo de confianza deseado.

Lo que la ecuación 2.3 refleja es un rango dentro del cual varía la media, considerando un nivel de probabilidad de $1 - \alpha$. Por ejemplo, tomemos la estimación de la propensión marginal a consumir (PMC) de una función de consumo y asumamos que el estimador del parámetro poblacional es $\hat{b} = 0,84$ y su varianza igual a $\sigma^2/n = 0,0064$. El intervalo de confianza estaría dado por:

$$P\left(0,84 - 1,96 \cdot \sqrt{0,0064} < \mu < 0,84 + 1,96 \cdot \sqrt{0,0064}\right) = 0,95$$

$$P(0,6832 < \mu < 0,9968) = 0,95$$

donde el valor 1.96 se ha obtenido de una tabla de distribución normal estándar para un valor de 0.975 con n infinito.

En palabras, se puede decir que la propensión marginal al consumo debería fluctuar entre un rango de 0.683 y 0.997 con un 95 % de probabilidad.

Si bien es cierto, esta información es útil, existen ocasiones en que puede resultar muy vaga o riesgosa. Note que sin necesidad de realizar estimaciones podemos decir con un 100 % de confianza que la propensión marginal a consumir estará entre cero y uno. Obviamente, esta información no es sorprendente, y será de muy poca utilidad al momento de diseñar políticas económicas (por ejemplo, para activar la economía). Es decir, si el intervalo de confianza es muy amplio, no entrega mucha información.

Otra forma de abordar el problema es realizar lo que se conoce como *prueba de hipótesis puntual*. Existen ocasiones en que por alguna razón se cree que el estimador debe tener un valor específico (por ejemplo, se puede creer que la PMC es igual a 0.5). En las pruebas de hipótesis, se contrasta alguna determinada creencia respecto del valor del parámetro poblacional, a lo cual se le denomina **hipótesis**, con una creencia alternativa.

Cualquier prueba de hipótesis que se realice debe contener algunos elementos básicos. Estos son:

1. **Hipótesis Nula:** es la hipótesis que se pretende probar. Esta puede ser **simple** o **compuesta**. En el primer caso se desea verificar sólo una condición. Generalmente toma la forma de un enunciado sobre el valor específico que toma el parámetro relevante. Por ejemplo $\mu = \mu_0$, donde μ_0 representa un valor determinado. En caso de pruebas compuestas se quiere probar más de una condición. Cualquiera sea el caso se denotará como H_0 .
2. **Hipótesis Alternativa:** Es la hipótesis que sirve para contraponer y que complementa la hipótesis nula. La hipótesis alternativa debe cubrir todas las otras posibilidades de valores que pueda adoptar el parámetro relevante, distintos a la enunciada en la hipótesis nula. En nuestro ejemplo anterior, la hipótesis alternativa sería $\mu \neq \mu_0$. Generalmente se denota por H_1 .
3. **Estadístico de Prueba:** es una función de la muestra y que se contrastará con la hipótesis nula. Sobre la información que entregue tomaremos una decisión consistente sobre la veracidad o falsedad de la hipótesis nula.
4. **Zonas de Rechazo y de Aceptación:** son las zonas de la función de distribución de probabilidades donde se rechaza o acepta la hipótesis nula a favor de la hipótesis alternativa. Si el estadístico de prueba cae fuera de la zona de rechazo, entonces aceptaremos la hipótesis nula.

Existen dos errores que se pueden cometer al realizar una prueba de hipótesis. El primero que se denomina error de **tipo I** corresponde al error de rechazar la hipótesis nula cuando ésta es verdadera. El segundo es el error de **tipo II** que corresponde al error de aceptar la hipótesis nula, siendo esta falsa. La probabilidad de cometer el error de tipo I se denota por α y se conoce con el nombre de **nivel de significación**, mientras que $1 - \alpha$ se conoce como el **nivel de confianza**. Por otra parte la probabilidad de cometer el error tipo II, se denota por β , y $1 - \beta$ se conoce con el nombre de **potencia**.

El investigador requerirá contrastes donde las probabilidades de cometer alguno de los errores sean pequeñas, pero lamentablemente solo puede reducirse uno de ellos a costa de aumentar el otro. Los valores más comúnmente usados para α son 1 % y 5 %. Con estos valores queremos indicar que el error tipo I es generalmente considerado más importante y que lo queremos evitar.

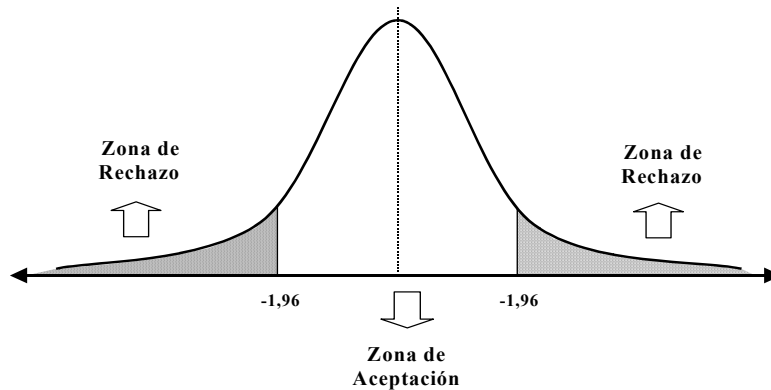


Figura 2.4: **Zonas de Aceptación y Rechazo de Hipótesis**

Para realizar la prueba, podemos usar el intervalo de confianza construido para el parámetro. Supongamos que se desea contrastar una hipótesis con respecto al valor de μ , lo cual se escribe como la hipótesis nula $H_0 : \mu = \mu_0$, y compararla con la hipótesis alternativa $H_1 : \mu \neq \mu_0$. Para verificar esta hipótesis debemos observar si μ_0 se encuentra dentro del intervalo de confianza. Si es así entonces podemos aceptar la hipótesis nula. En caso contrario, si μ_0 se encuentra fuera del intervalo, entonces aceptaríamos la hipótesis alterna.

En el caso particular de la PMC asumamos $H_0 : b = 0,5$. Este valor está fuera del intervalo de confianza construido, por lo que podemos decir que se rechaza la hipótesis nula de que el verdadero valor del parámetro es 0,5, considerando un 95% de probabilidad. En resumen, si se conoce el estimador de θ , se construye un intervalo al nivel de confianza deseado, y se acepta la hipótesis nula si θ_0 cae dentro del intervalo. El análisis es distinto si la hipótesis nula se refiere a una igualdad ($H_0 : \theta = \theta_0$), o si se refiere a una desigualdad ($H_0 : \theta \geq \theta_0$). La diferencia principal está en la zona de rechazo de la prueba que en el caso de igualdad tendrá dos colas, mientras en una desigualdad va a existir una sola cola. En el caso de una desigualdad mayor la zona de rechazo será la cola de la izquierda, mientras que en el caso de una desigualdad menor, la zona de rechazo será la cola de la derecha. Para entender mejor el resultado se presenta la figura 2.4 en la cual se muestran las zonas de aceptación y de rechazo de esta prueba de hipótesis, para un $\alpha = 0,05$.

Alternativamente, la prueba de hipótesis se puede realizar de forma puntual. No olvide que para realizar cualquier prueba es necesario definir la distribución de probabilidad que tiene el estimador. Para ejemplificar, supongamos que tenemos un estimador $\hat{\theta}$ que tiene una distribución normal con media θ y varianza σ^2/n . Y planteamos la siguiente hipótesis nula:

$$H_0 : \theta = \theta^*$$

y su hipótesis alterna:

$$H_1 : \theta \neq \theta^*$$

Para realizar la prueba se calcula el valor del estadístico z_c como una variable normalizada estándar:

$$z_c = \frac{\hat{\theta} - \theta^*}{\sqrt{\frac{\sigma^2}{n}}}$$

Donde $\hat{\theta}$ es el valor del estimador, θ^* es el valor que queremos asignar mediante la prueba de hipótesis y σ es la desviación estándar.¹² Luego una vez que conocemos el valor de z_c lo comparamos con los valores obtenidos de la distribución de probabilidades normal estándar tabulada. Si el nivel de confianza es del 95 %, el valor de tabla es de 1,96. Luego si z_c es mayor que el valor de tabla en valor absoluto rechazamos la hipótesis nula. Esto es válido en el caso de distribuciones simétricas, ya que en los otros casos se tendrá dos valores, uno para la cola superior (-1.96 en nuestro caso) y otro para la cola superior (1.96). Si el valor de z_c cae dentro de este intervalo, entonces aceptamos la hipótesis, en caso contrario la rechazamos.

Retomemos el ejemplo anterior en que la estimación de la propensión marginal a consumir de la función de consumo es $\hat{b} = 0,84$ y su varianza igual a $\sigma^2/n = 0,0064$. Planteamos la hipótesis que la propensión marginal a consumir es igual a 0.5, entonces nuestra prueba de hipótesis será:

- Hipótesis nula $H_0 : b = 0,5$

¹²Note la diferencia entre θ y θ^* . El primer operador se utiliza para identificar el verdadero valor del parámetro, el cual es desconocido. Mientras que θ^* denota cualquier creencia sobre el valor que puede tomar el parámetro.

- Hipótesis alterna $H_1 : b \neq 0,5$

Como se puede ver esta es una prueba de dos colas. Calculando nuestro estadístico de prueba:

$$z_c = \frac{\hat{b} - b}{\sqrt{s^2/n}} = \frac{0,84 - 0,5}{\sqrt{0,00064}} = 13,44$$

Si comparamos este valor con la tabla normal con un 5% de confianza, encontramos que este valor cae en la zona de rechazo. Por lo tanto podemos rechazar con un 5 por ciento de significancia que la propensión marginal al consumo sea igual a 0.5. Note que este caso es análogo a la prueba usando intervalos de confianza.

Un caso especial de este procedimiento es el **Test de Significancia**¹³, en el cual se determina si el valor del parámetro es estadísticamente distinto de cero. De esta forma la hipótesis nula es $H_0 : \theta = 0$, y por lo tanto la hipótesis alterna es $H_1 : \theta \neq 0$. Retomando el ejemplo de la función de consumo, si la hipótesis nula plantea que $b = 0$, y ello resulta ser verdadero, entonces, esto quiere decir que el nivel de ingreso no afecta el nivel de consumo. En otras palabras, lo que se busca probar es si la variable que acompaña a este parámetro es significativa para explicar el comportamiento de la variable dependiente. Si el valor del parámetro fuese cero o si aceptáramos esta hipótesis, implica que la variable explicativa no es relevante en nuestro modelo. Por ejemplo, si $b = 0$ tenemos:

$$z_c = \frac{0,84 - 0}{\sqrt{0,00064}} = 33,204$$

Lo que cae claramente en el área de rechazo. Obviamente, si nuestro modelo está correctamente especificado deberíamos rechazar la hipótesis nula. Si rechazamos la hipótesis nula de que el parámetro es cero, quiere decir que la variable usada en el modelo es significativa.

¹³Llamaremos Test de Significancia a la prueba de hipótesis mediante la cual se intenta probar si el valor del parámetro es igual a cero.

Capítulo 3

MODELO DE REGRESIÓN LÍNEAL GENERAL

3.1. Introducción

En este capítulo se presenta uno de los modelos más utilizados para la estimación econométrica, conocido como el modelo de regresión lineal general. En su versión sencilla, este modelo sólo tiene dos variables, una explicada y otra explicativa. En su forma general, este modelo puede incluir tantas variables explicativas como sea necesario. Debido a que la mayoría de los problemas prácticos en Econometría incluyen más de una variable explicativa, se adoptará el enfoque general.

En la exposición del tema se trabajará con un enfoque matricial, ya que con este instrumental es posible tratar con mayor facilidad el problema de estimación de un modelo lineal con más de dos variables. En la exposición del capítulo, se recurrirá al modelo simple cuando sea útil para ejemplificar o explicar algún punto de relevancia.

El modelo lineal general, también conocido como **modelo clásico**, es el modelo básico en econometría, por lo que todo el desarrollo de la teoría econométrica subsiguiente requiere un conocimiento cabal de éste. Su característica fundamental es su simplicidad, como consecuencia de los supuestos utilizados. Estos supuestos son un tanto restrictivos. Sin embargo, una vez que se domina el modelo básico, es posible levantar algunos supuestos y estudiar su efecto sobre los estimadores. Los capítulos posteriores discutirán estos aspectos. Por el momento, nos concentraremos en el modelo clásico.

3.2. Modelo Clásico

El Modelo lineal general está formado por una variable dependiente o explicada que denotaremos por Y_i , cuya variabilidad es explicada por k variables independientes o explicativas X_{ij} donde $j = 1, \dots, k$, y una variable aleatoria no observable comúnmente expresada como μ_i , y conocida como término de error. El subíndice i expresa la observación correspondiente de las variables, donde suponemos que $i = 1, \dots, n$. Por ejemplo, si las observaciones son distintos individuos, entonces i refleja el valor que tienen las variables Y y X_j para el individuo i . Esto se puede escribir como:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \mu_i \quad (3.1)$$

El sentido de esta relación es medir la contribución que cada variable X_j realiza a la “explicación” de la variable Y . Además se incluye un término constante, con valor igual a 1 para cada i (que corresponde al coeficiente β_0), para captar el efecto de aquellos factores que no varían entre las observaciones i . Finalmente, el término de error debería reflejar aquellas variables explicativas no incluidas en la ecuación, pero que sí varían entre las observaciones.

Como tenemos n individuos, existe una ecuación como 3.1 para cada individuo. Con el fin de simplificar la escritura y utilizar las ventajas de la notación matricial, el modelo puede resumirse en:

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times k} \boldsymbol{\beta}_{k \times 1} + \boldsymbol{\mu}_{n \times 1} \quad (3.2)$$

Donde \mathbf{Y} es un vector columna con n observaciones de la variable dependiente, \mathbf{X} es una matriz de orden $n \times k$, con unos (1) en su primera columna y con las observaciones de las variables explicativas en las restantes columnas, y $\boldsymbol{\beta}$ es un vector columna de coeficientes. Por último, $\boldsymbol{\mu}$ es un vector columna que contiene los errores para cada observación de la muestra.

Es decir:

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times k} \boldsymbol{\beta}_{k \times 1} + \boldsymbol{\mu}_{n \times 1} \quad (3.3)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

Pero, ¿qué refleja esta ecuación?. Primero que nada esta ecuación refleja lo que la teoría económica sugiere sobre algún modelo. En términos generales dice que los cambios en la variable Y se pueden explicar por un conjunto de variables explicativas denotadas por $\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$, lo que es conocido como la parte determinística del modelo, y además, existe una porción de los cambios en Y que no podemos explicar con este modelo o bien que no sabemos cómo explicar. Este último componente se conoce como componente estocástico del modelo y se incluye en el vector μ .

Desde el punto de vista de la estimación, uno generalmente cuenta con información tanto de la variable dependiente como de las independientes, y el problema que se enfrenta es determinar el valor del vector β , es decir estimar el valor de los coeficientes de las variables independientes. Estos coeficientes reflejan la contribución individual de cada variable explicativa a la determinación del nivel de la variable explicada.

La ecuación 3.2 puede representar cualquier modelo económico, como una función de consumo o una función de producción, etc. Lo que interesa desde la perspectiva econométrica es estimar los parámetros (los β 's) usando una muestra de observaciones tanto de las variables explicativas como de la variable explicada.

Por ejemplo, si se quiere conocer la Productividad Marginal del Trabajo, es necesario estimar una función de producción. Una de las funciones típicas usada en economía es la función Cobb-Douglas del tipo:

$$Y = AX_1^\alpha X_2^\beta$$

donde X_1 representa el capital, X_2 es el trabajo y A es una variable de posición. Los parámetros de interés asociados a los insumos son α y β . Aplicamos logaritmo para linealizar la ecuación. A partir de una serie de datos obtenidos del Boletín mensual del Banco Central de Chile¹ respecto del producto interno bruto, del capital y del trabajo para el período comprendido por el primer trimestre del año 1987 y el cuarto trimestre de 1993 se obtienen los siguientes resultados:

$$\ln(Y) = -1,7945 + 0,0826 \ln(X_1) + 1,5259 \ln(X_2) + \varepsilon_t$$

Así, la elasticidad parcial del capital está representada por $\hat{\beta}_1$ y tiene un valor igual a 0,0826 y la elasticidad parcial del trabajo está representada por

¹Página web del Banco Central y Colección de Estudios CIEPLAN N° 39

$\hat{\beta}_2$ e igual a 1,5259. Con los valores medios de todas las variables es posible obtener la Productividad Marginal para ambos factores productivos.

Retomando el modelo expresado en 3.2 el investigador está interesado en obtener una solución para el vector de parámetros β que tenga las propiedades deseadas de un estimador, como son insesgamiento y eficiencia². Para lograr esta solución, el modelo clásico asume una serie de supuestos que tienen la característica de simplificar apreciablemente la obtención del vector de estimadores y que además aseguran que estos estimadores tengan las propiedades deseadas.

3.3. Supuestos del Modelo Clásico

Una serie de supuestos sobre las características de los datos y la forma como se relacionan son parte del modelo lineal general. Estos supuestos son fundamentales para obtener las propiedades de los estimadores y son el punto de partida de los posteriores modelos de estimación. Cambios o levantamiento de estos supuestos imponen nuevos requerimientos en los métodos de estimación. Los supuestos del modelo son los siguientes:

Supuesto 1. La regresión es lineal en los parámetros.

Esto quiere decir que el vector β contiene expresiones lineales para cada uno de los parámetros. La linealidad debe cumplirse sólo en los parámetros, no es necesario que las variables explicativas sean lineales. Por ejemplo, el modelo $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \mu_i$ cumple con la propiedad de linealidad en los parámetros aunque X_1 esté al cuadrado.

Supuesto 2. El valor medio o esperanza del error μ_i es igual a cero, lo cual se escribe como:

$$E(\boldsymbol{\mu}) = \mathbf{0}$$

$$E \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E(\mu_1) \\ E(\mu_2) \\ \vdots \\ E(\mu_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

²Ver Capítulo 2

Esto quiere decir que el valor promedio de μ_i dado los valores de las variables explicativas (\mathbf{X}_{ki}) es cero. Lo cual implica que los valores positivos y negativos de μ_i se cancelan de tal manera que su efecto promedio sobre Y_i es cero.

Note que para cada observación tenemos un conjunto de variables de la forma:

$$(Y_i, X_{i1}, \dots, X_{ik}, \mu_i)$$

donde μ_i puede tomar cualquier valor dentro de un intervalo con una probabilidad asociada. Por lo tanto, para cada observación “ i ”, el supuesto (2) nos dice que $E(\mu_i)$ será igual a cero.³

Con la intención de aclarar más este concepto considere el caso del modelo lineal general en que sólo existe una variable explicativa, es decir:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \mu_i$$

Asuma, para efectos de este ejemplo, que $\beta_0 > 0$ y $\beta_1 > 0$. En el gráfico correspondiente a esta ecuación (figura 3.1) se observa que para un valor específico de X , llamémosle X_0 , se tiene una serie de posibles valores de Y_0 . Esto se refleja en la distribución de probabilidad dibujada sobre la línea vertical trazada desde X_0 . Esta gama de valores depende del error μ_i . Si $E(\mu_i) = 0$, entonces se espera que el punto de intersección entre X_0 y Y_0 esté sobre la recta $Y_i = \beta_0 + \beta_1 X_{1i}$.

Adicionalmente note que para valores dados de β_0 y β_1 , el valor que asuma μ_i dependerá de la diferencia $Y_i - \beta_0 - \beta_1 X_{1i} = \mu_i$.

Supuesto 3. Homocedasticidad y no Autocorrelación.

Este tercer supuesto sintetiza dos propiedades fundamentales del modelo lineal general; la *homocedasticidad* o igual varianza de μ_i y la inexistencia de *autocorrelación* entre los errores⁴. Estos conceptos están relacionados con la varianza y covarianza de los errores, por lo tanto necesitamos encontrar

³Lo que no significa que para una muestra particular la $\sum_i^n \frac{\mu_i}{n} = 0$, ya que en este caso cada error tomará un valor definido cuya suma no necesariamente es cero.

⁴Cabe mencionar que la heterocedasticidad es un fenómeno común en muestras de corte transversal mientras que la autocorrelación lo es en series de tiempo.

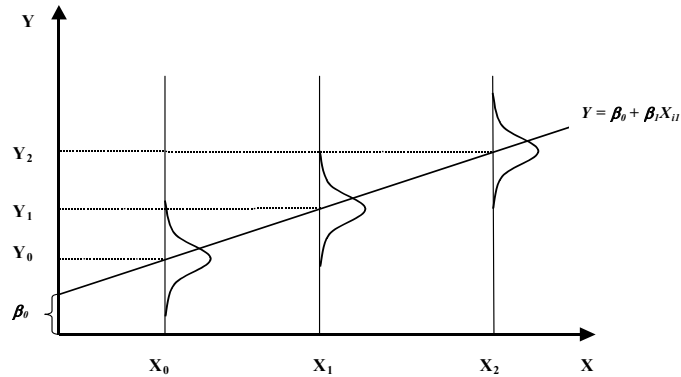


Figura 3.1: Distribución de los errores con homocedasticidad

esta matriz. Para ello aplicamos la definición de varianza en forma matricial y obtenemos:⁵

$$\begin{aligned}
 E(\boldsymbol{\mu}\boldsymbol{\mu}') &= E\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \mu_n \end{bmatrix} [\mu_1 \ \mu_2 \ \cdots \ \mu_n]\right) = E\left(\begin{matrix} \mu_1^2 & \mu_1\mu_2 & \cdot & \cdot & \cdot & \mu_1\mu_n \\ \mu_2\mu_1 & \mu_2^2 & \cdot & \cdot & \cdot & \mu_2\mu_n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mu_n\mu_1 & \mu_n\mu_2 & \cdot & \cdot & \cdot & \mu_n^2 \end{matrix}\right) \\
 &= \begin{pmatrix} E(\mu_1^2) & E(\mu_1\mu_2) & \cdot & \cdot & \cdot & E(\mu_1\mu_n) \\ E(\mu_2\mu_1) & E(\mu_2^2) & \cdot & \cdot & \cdot & E(\mu_2\mu_n) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ E(\mu_n\mu_1) & E(\mu_n\mu_2) & \cdot & \cdot & \cdot & E(\mu_n^2) \end{pmatrix}
 \end{aligned}$$

El supuesto de homocedasticidad significa que la varianza de los μ_i es la misma para toda observación “ i ”, es decir:

$$E(\mu_i^2) = \sigma_\mu^2 \quad \forall i = 1, \dots, n$$

donde σ_μ^2 es un escalar constante para todo i .

⁵Recuerde que $Var(\boldsymbol{\mu}) = E[(\boldsymbol{\mu} - E(\boldsymbol{\mu}))(\boldsymbol{\mu} - E(\boldsymbol{\mu}))']$. Pero como $E(\boldsymbol{\mu}) = \mathbf{0}$, obtenemos $E(\boldsymbol{\mu}\boldsymbol{\mu}')$

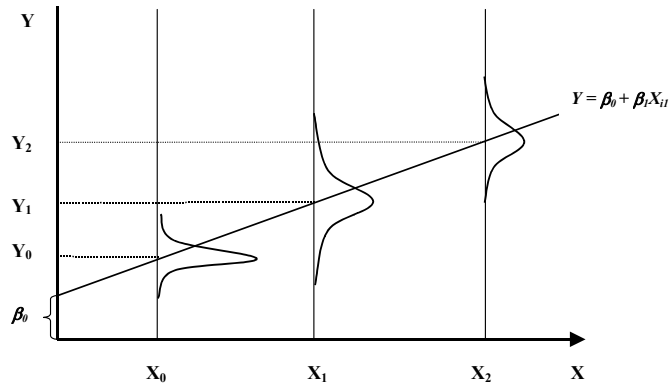


Figura 3.2: **Distribución de los errores con heterocedasticidad**

Esto se representa gráficamente (figura 3.1) con distribuciones de igual amplitud para los posibles resultados de Y_i , en cada valor de X_i . En el caso contrario, es decir, que las varianzas de los distintos μ_i fueran diferentes (heterocedasticidad), entonces enfrentaríamos distribuciones de probabilidad de distinta amplitud para cada valor X (ver figura 3.2)

Por su parte, ausencia de autocorrelación implica que la covarianza entre μ_i y μ_j es cero, denotado por

$$E(\mu_i \mu_j) = 0 \quad \forall i \neq j$$

En términos conceptuales, ausencia de autocorrelación nos indica que el valor del error en una observación i no es afectado por el valor que muestre el error de otra observación j . Por tanto, no existirá autocorrelación cuando, por ejemplo, el comportamiento en un período de un determinado elemento no afecte el comportamiento que presente dicho elemento en el período subsiguiente. Es decir, el nivel de consumo de un grupo familiar en un mes i no afectará el nivel de consumo de este mismo grupo familiar en el mes $i + 1$.

La homocedasticidad y no autocorrelación se sintetizan en la matriz de varianzas y covarianzas la cual se expresa como:

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \begin{pmatrix} \sigma_\mu^2 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \sigma_\mu^2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \sigma_\mu^2 \end{pmatrix} = \sigma_\mu^2 \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix} = \sigma_\mu^2 \mathbf{I}_n$$

Donde \mathbf{I}_n es la matriz identidad de orden n .

Supuesto 4. Los valores de \mathbf{X} son fijos en muestreos repetidos.

En otras palabras, las variables explicativas son determinísticas o predefinidas. Dado que \mathbf{X} contendrá solamente valores conocidos, no existirá correlación entre el término de error μ_i y ninguno de los X_{ij} , que pertenecen a \mathbf{X} donde $j = 1, \dots, k$, es decir:

$$\text{cov}(X_{ij}\mu_i) = 0 \quad \forall j$$

El aporte de este supuesto es que permite separar el efecto de las variables explicativas del efecto de los errores sobre la variable dependiente. En el caso de que este supuesto no se cumpla, no habría forma de determinar qué proporción de la variabilidad de Y_i se explica por las variables X_{ij} y qué parte se explica por el término de error.

Los cuatro supuestos mencionados anteriormente son los supuestos más importantes del modelo clásico y determinan las propiedades de los estimadores. Existen no obstante, otros supuestos adicionales mencionados comúnmente en la literatura, por ejemplo:

Supuesto 5. No existe multicolinealidad perfecta ($\text{Rango}(\mathbf{X}) = k < n$). No hay filas (columnas) linealmente dependientes. Esto tiene una implicancia directa sobre la posibilidad de identificar los parámetros deseados. Si existiera dependencia lineal entre las observaciones de la matriz \mathbf{X} , entonces no se podría estimar el modelo. Este enunciado será clarificado mas adelante.

Por último, existe un supuesto de tipo teórico, es decir, este supuesto no está relacionado con los datos. El supuesto es que *el modelo de regresión está correctamente especificado*. Obviamente si el modelo no está bien

especificado, entonces al aplicar cualquier método de estimación econométrica no entregará información útil al economista. Por esta razón, se sugiere que en primer término se debe tener un *Buen Modelo* y luego una *Buena Econometría*, poniendo el énfasis en la labor del economista con el fin de que construya un modelo que esté correctamente especificado. Además, el cometer errores en la especificación del modelo, como por ejemplo no considerar variables que son relevantes, podría sesgar los estimadores.

3.4. Mínimos Cuadrados Ordinarios

Para introducir el proceso de estimación llamemos a la ecuación $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ (ecuación 3.2, donde se han omitido los subíndices) *Función de Regresión Poblacional (FRP)*. Supondremos que esta ecuación representa el modelo teórico que pretendemos estimar y que refleja fielmente el comportamiento poblacional. El valor esperado de Y para la Función de Regresión Poblacional es

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta}) + E(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$$

gracias a los supuestos (2) y (4) del modelo.

Sin embargo, esta función no es conocida por el investigador, y por esta razón se desea estimar. Para hacer esto se acude a una muestra de observaciones que permita aproximarnos de la mejor forma posible al modelo teórico. Así, definimos una *Función de Regresión Muestral (FRM)*

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon} \quad (3.4)$$

Donde $\hat{\boldsymbol{\beta}}$ es el vector de estimadores de los parámetros poblacionales y $\boldsymbol{\varepsilon}$ es el término de error muestral (análogo al término de error $\boldsymbol{\mu}$). Usando una muestra del Universo se puede estimar la ecuación 3.4.

Para el caso de una sola variable explicativa, la Función de Regresión Poblacional y la Función de Regresión Muestral (ambas hipotéticas) se presentan en la figura 3.3. Al dibujar esta figura se asume que los coeficientes estimados ($\hat{\beta}$) difieren de los poblacionales (β). Se puede apreciar que la Función de Regresión Muestral subestima los verdaderos valores de la Función de Regresión Poblacional o de la variable dependiente a la izquierda de X_0 , mientras que a la derecha de este mismo punto los valores de la Función de Regresión Poblacional son sobrestimados. Para un valor determinado de

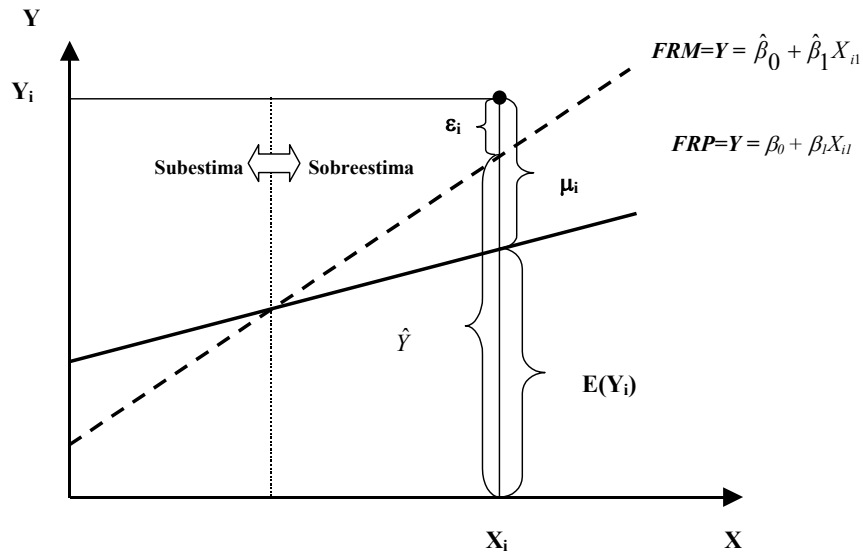


Figura 3.3: Función de Regresión Poblacional y Muestral

X , digamos X_i observamos que el valor efectivo de Y_i es superior al valor esperado $E(Y_i)$. La diferencia está dada por el error μ_i . Sin embargo, en la Función de Regresión Muestral el valor de $E(Y_i)$ se sobreestima y por ello el error estimado (e_i) es menor que el error efectivo (μ_i).

Note que no conocemos la función de regresión poblacional, por lo tanto debemos contar con algún criterio de estimación que nos acerque lo más posible a esta función. Para hacer esto tenemos los métodos de estimación econométrica, de los cuales se discutirán el de Mínimos Cuadrados Ordinarios (MCO) y el de Máxima Verosimilitud (MV).

El Método de los Mínimos Cuadrados está basado en la idea de buscar un estimador que minimice la dispersión de los errores de la regresión en torno a un estadístico relevante de la distribución (un momento). Es uno de los métodos de estimación más utilizados en la literatura.

Recordemos que la Función de Regresión Poblacional es:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \quad (3.5)$$

La esperanza condicional está dada por

$$\mathbf{E}(\mathbf{Y}/\mathbf{X}) = \mathbf{X}\boldsymbol{\beta} \quad (3.6)$$

Las razones de este resultado son el hecho que los \mathbf{X} son fijos (supuesto 4) y que β es el vector de parámetros poblacionales y como tal no es una variable aleatoria, además la $E(\mu) = 0$ por el segundo supuesto del modelo clásico.

Por su parte la función de regresión muestral es

$$\mathbf{Y} = \mathbf{X}\hat{\beta} + \varepsilon \quad (3.7)$$

con

$$\mathbf{E}(\mathbf{Y}/\mathbf{X}) = \hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} \quad (3.8)$$

Como se planteó en el capítulo 2, el objetivo de *MCO* es obtener el vector $\hat{\beta}$ que minimice la suma de los errores al cuadrado $\sum_1^n \varepsilon_i^2$, lo que en forma matricial se traduce en

$$MIN_{\hat{\beta}} \sum_1^n \varepsilon_i^2 = MIN_{\hat{\beta}} \varepsilon' \varepsilon = MIN_{\hat{\beta}} \left[\begin{array}{cccc} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{array} \right] \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix} \quad (3.9)$$

Despejando ε de la ecuación 3.7 tenemos

$$\varepsilon = \mathbf{Y} - \mathbf{X}\hat{\beta}$$

reemplazando la ecuación anterior en 3.9, se tiene

$$MIN_{\hat{\beta}} (\mathbf{Y} - \mathbf{X}\hat{\beta})' (\mathbf{Y} - \mathbf{X}\hat{\beta}) \quad (3.10)$$

utilizando algunas propiedades de la transposición de matrices analizadas en capitulos precedentes queda

$$MIN_{\hat{\beta}} (\mathbf{Y}' - \hat{\beta}' \mathbf{X}') (\mathbf{Y} - \mathbf{X}\hat{\beta}) = MIN_{\hat{\beta}} (\mathbf{Y}' \mathbf{Y} - \mathbf{Y}' \mathbf{X} \hat{\beta} - \hat{\beta}' \mathbf{X}' \mathbf{Y} + \hat{\beta}' \mathbf{X}' \mathbf{X} \hat{\beta}) \quad (3.11)$$

tanto $\mathbf{Y}' \mathbf{X} \hat{\beta}$ como $\hat{\beta}' \mathbf{X}' \mathbf{Y}$ tienen dimensión 1×1 , por ello su valor traspuesto no varía. En otras palabras $\mathbf{Y}' \mathbf{X} \hat{\beta} = \hat{\beta}' \mathbf{X}' \mathbf{Y}$, quedando la expresión 3.11 de la siguiente forma

$$MIN_{\hat{\beta}} (\mathbf{Y}' \mathbf{Y} - 2\hat{\beta}' \mathbf{X}' \mathbf{Y} + \hat{\beta}' \mathbf{X}' \mathbf{X} \hat{\beta}) \quad (3.12)$$

Así la ecuación anterior se convierte ahora en nuestro problema de optimización. Como sabemos de la Teoría de Optimización, para encontrar un mínimo debemos derivar esta expresión con respecto al vector de parámetros e igualar a cero, es decir

$$\frac{\partial(\mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})}{\partial\hat{\boldsymbol{\beta}}} = 0$$

Derivando y usando las propiedades de derivación de matrices presentadas en el capítulo 1 se tiene:

$$\begin{aligned} -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= 0 \\ \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{Y} \end{aligned}$$

Despejando el vector $\hat{\boldsymbol{\beta}}$ obtenemos

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) \quad (3.13)$$

Esta última ecuación es la matriz de estimadores de los parámetros poblacionales. Es claro que el vector de estimadores depende de las observaciones de la variable dependiente y de las variables explicativas, cuya información está resumida en las matrices $(\mathbf{X}'\mathbf{X})^{-1}$ y $(\mathbf{X}'\mathbf{Y})$ ya que:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & \sum X_{i1} & \sum X_{i2} & \dots & \sum X_{ik} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \dots & \sum X_{i1}X_{ik} \\ \sum X_{i2} & \sum X_{i2}X_{i1} & \sum X_{i2}^2 & \dots & \sum X_{i2}X_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{ik} & \sum X_{ik}X_{i1} & \sum X_{ik}X_{i2} & \dots & \sum X_{ik}^2 \end{bmatrix}$$

y

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & X_{31} & \dots & X_{n1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1k} & X_{2k} & X_{3k} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{i1} \\ \sum Y_i X_{i2} \\ \vdots \\ \sum Y_i X_{ik} \end{bmatrix}$$

El $\hat{\beta}$, no es otra cosa que un estimador de los verdaderos parámetros poblacionales y por lo tanto es una regla o función que depende de los datos observados. Ya que $\hat{\beta}$ depende de una variable aleatoria Y , debe a su vez también ser una variable aleatoria. Para el caso de una regresión simple (una variable independiente) se puede mostrar que la ecuación 3.13 queda

$$\hat{\beta} = \frac{\Sigma (X_i - \bar{X}) (Y_i - \bar{Y})}{\Sigma (X_i - \bar{X})^2} \quad (3.14)$$

donde una barra sobre la variable refleja su valor promedio.

La ecuación 3.14 permite tener una visión más intuitiva de lo que determina el valor de $\hat{\beta}$. Vemos que el denominador será siempre positivo, por lo tanto el signo que adopte $\hat{\beta}$ dependerá del valor del numerador. Si cuando los valores de X_i se encuentra bajo el promedio (\bar{X}), los valores de Y_i también se encuentran bajo el promedio (\bar{Y}) (y viceversa) entonces el numerador será positivo y $\hat{\beta} > 0$. En este caso, existe correlación positiva entre X_i e Y_i . Si por el contrario, cuando X_i esté debajo del promedio, Y_i esté sobre el promedio (y viceversa), entonces el numerador será negativo y $\hat{\beta} < 0$ (correlación negativa entre X_i e Y_i). Además, se puede observar que el denominador es una medida de la variación de X_i en torno a su promedio. Si la covariación (positiva o negativa) de X_i e Y_i es muy grande en relación a la variación de X_i , entonces $\hat{\beta}$ adoptará un valor grande. En caso que la covariación de X_i e Y_i sea menor que la variación de X_i , entonces el valor de $\hat{\beta}$ será pequeño (menor que uno).

Analícemos en más detalle el proceso de estimación. Para ello consideremos la expresión extendida del modelo que se desea estimar:

$$FRP : Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \mu_i$$

La estimación del modelo correspondiente a la función muestral

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik} + \varepsilon_i$$

el vector $\hat{\beta}$ está definido como:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix},$$

el cual se obtiene usando la formulación presentada en la ecuación 3.13. La interpretación de los coeficientes es la siguiente:

- $\hat{\beta}_0$ es el efecto promedio de todas las variables no aleatorias que no fueron incluidas en el modelo pero que afectan a la variable dependiente. Este es el efecto promedio en la regresión general y es un parámetro de posición de la recta.
- $\hat{\beta}_j = \frac{\partial Y_i}{\partial X_j}$ es la pendiente de la regresión con respecto a la variable X_j , es decir, mide el cambio en el valor de Y ante cambios en X_j .

Ejemplo 1: Estimación de una Función de Consumo

Como ejemplo consideremos una función de consumo Keynesiana básica

$$C = a + bY,$$

la cual señala que el consumo depende del ingreso disponible Y , pero que no se gasta todo el ingreso en consumo sino una porción b de éste, llamada *proporción marginal a consumir*, por otra parte existe un consumo autónomo, que no depende del ingreso y que está representado por la constante a en la ecuación anterior. Debería ser claro para el lector que se está describiendo el modelo teórico que está dado por la teoría económica.

Lo que interesa es estimar el valor de a y b usando información de consumo e ingreso de una economía (puede ser también de las familias), para lo cual debemos transformar nuestro modelo teórico determinístico en uno estocástico, es decir:

$$C = a + bY + \mu_i$$

que representa la función de regresión poblacional y

$$C = \hat{a} + \hat{b}Y + \varepsilon_i$$

que es la función de regresión muestral

Usando una muestra de datos⁶ de las Cuentas Nacionales sobre Consumo e Ingreso Disponible en Chile para el periodo comprendido entre los años 1960 a 1997 presentada en el apéndice A, podemos obtener estos estimadores.

⁶Los datos originales están expresados en millones de pesos de 1986. Para la estimación que se hace en este apartado, y en la subsección siguiente referida al estimador de la varianza del error; los datos han sido divididos por un millón para evitar resultados muy grandes. Se sugiere al lector verificar estos resultados para asegurar una buena comprensión del problema.

Primero considere la ecuación para el vector de estimadores:

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

En este caso particular la matriz $\mathbf{X}'\mathbf{X}$ esta dada por:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & \sum X_{i1} \\ \sum X_{i1} & \sum X_{i1}^2 \end{bmatrix}$$

Usando los datos de la muestra obtenemos:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 38 & 132,996 \\ 132,996 & 559,776 \end{bmatrix}$$

y su inversa esta dada por,

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0,1562 & -0,0371 \\ -0,0371 & 0,0106 \end{bmatrix} \quad (3.15)$$

a su vez

$$(\mathbf{X}'\mathbf{Y}) = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{i1} \end{bmatrix} = \begin{bmatrix} 93,4357 \\ 384,8900 \end{bmatrix} \quad (3.16)$$

Por último,

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) = \begin{bmatrix} 0,1562 & -0,0371 \\ -0,0371 & 0,0106 \end{bmatrix} \begin{bmatrix} 93,4357 \\ 384,8900 \end{bmatrix} \\ \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) = \begin{bmatrix} 0,3152 \\ 0,6134 \end{bmatrix} \end{aligned} \quad (3.17)$$

reemplazando en la función de regresión muestral, tenemos⁷:

$$\begin{aligned} C_t &= 0,3152 + 0,6134 \times Y_t + \varepsilon_t \\ n &= 38 \end{aligned}$$

donde C_t representa al consumo privado en el período t , Y_t es el ingreso real disponible en el período t .

Como podemos observar, la estimación obtenida cumple con la teoría económica, en el sentido que el coeficiente del ingreso es un número positivo

⁷Note que el parámetro de posición está expresado en unidad de millones de pesos de 1986.

menor que uno. El Consumo autónomo es de 0,3152 y la propensión marginal a consumir es de 0,6134.

Debería ser claro para el lector que este resultado está incompleto ya que como hemos discutido en el capítulo 2, los estimadores son variables aleatorias, lo cual significa que el coeficiente calculado es uno entre tantos valores posibles. Para completar el ejercicio de estimación econométrica es necesario calcular la varianza de los coeficientes estimados. Ello nos permitirá realizar pruebas de hipótesis respecto de los valores posibles de los coeficientes y construir intervalos de confianza para ellos.

El apartado siguiente discutirá las propiedades de los estimadores Mínimos Cuadráticos Ordinarios, centrándose en las propiedades de insesgamiento y mínima varianza. En el desarrollo de las demostraciones se analiza la forma de obtener la varianza de los estimadores.

3.4.1. Propiedades del Estimador MCO

En capítulos precedentes definimos un estimador como una función o regla de la forma $\hat{\beta} = \hat{\beta}(Y_1, \dots, Y_n)$ que permite calcular u obtener un valor para el parámetro poblacional. Además, se discutieron las propiedades deseadas de los estimadores. Es conveniente entonces preguntarse por las propiedades de los estimadores mínimos cuadráticos ordinarios del modelo lineal. Para obtener estas propiedades se hace uso de los supuestos presentados al inicio del capítulo.

Existe un teorema denominado *TEOREMA de GAUSS-MARKOV*, el cual plantea que si se cumplen los supuestos clásicos, es decir:

1. $E(\mu_i) = 0$
2. Homocedasticidad: $E(\mu_i - E(\mu_i))^2 = E(\mu_i^2) = \sigma_\mu^2$
3. No autocorrelación: $E(\mu_i - E(\mu_i))(\mu_j - E(\mu_j)) = E(\mu_i \mu_j) = 0$, con $i \neq j$.⁸
4. $Cov(x_{ij}, \mu_i) = 0 \quad \forall j$

Entonces el estimador de *MCO* es el *Mejor Estimador Lineal Insesgado (MELI)*. Es decir, los estimadores *MCO* son LINEALES, INSESGADOS y tienen MINIMA VARIANZA. Revisemos cada una de estas propiedades por separado.

⁸Los supuestos de homocedasticidad y no autocorrelación se resumen en $E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_\mu^2 \mathbf{I}_n$

(a) **Linealidad**

Sabemos que $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) = \mathbf{A}\mathbf{Y}$
 donde $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ es fijo por el supuesto de que los valores de \mathbf{X} son fijos, lo que muestra que el estimador es lineal.

(b) **Insesgamiento**

Para demostrar que el estimador de *MCO* es insesgado, tomamos como punto de partida el valor del estimador obtenido por *MCO* y la Función de Regresión Poblacional:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) \quad (3.18)$$

$$\mathbf{Y} = \mathbf{X}\beta + \mu \quad (3.19)$$

reemplazando la ecuación 3.19 en 3.18 se obtiene

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mu) \\ \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu \\ \hat{\beta} &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu \end{aligned} \quad (3.20)$$

aplicando esperanza a 3.20

$$\begin{aligned} \mathbf{E}(\hat{\beta}) &= \mathbf{E}[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu] \\ \mathbf{E}(\hat{\beta}) &= \mathbf{E}(\beta) + \mathbf{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu] \\ \mathbf{E}(\hat{\beta}) &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\mu) \end{aligned}$$

Donde se ha utilizado el hecho que $\mathbf{E}(\beta) = \beta$. Por último, es necesario recordar el primer supuesto del modelo clásico, el cual muestra que $\mathbf{E}(\mu) = \mathbf{0}$, entonces:

$$\mathbf{E}(\hat{\beta}) = \beta$$

En palabras, esto nos dice que la esperanza del estimador de la función de regresión muestral, es igual al vector de parámetros β de la función de regresión poblacional. Es decir, el estimador es insesgado.

(c) Varianza Mínima

Antes de mostrar que la varianza del estimador MCO es la mínima, requerimos de una expresión para esta varianza. En el caso matricial deseamos encontrar la matriz de varianzas y covarianzas de los estimadores. Para encontrar el estimador de la varianza de $\hat{\beta}$ reordenemos la ecuación 3.20 tal que:

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu \quad (3.21)$$

la matriz de varianzas y covarianzas esta definida como $E(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'$, pero sabemos que $E(\hat{\beta}) = \beta$, por lo tanto

$$\begin{aligned} Var(\hat{\beta}) &= E(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))' \\ Var(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \end{aligned}$$

y reemplazando 3.21 se obtiene

$$\begin{aligned} Var(\hat{\beta}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu)'] \\ Var(\hat{\beta}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu\mu'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \end{aligned}$$

Donde se aplicó las propiedades de la trasposición de matrices. Además se puede verificar que $[(\mathbf{X}'\mathbf{X})^{-1}]' = [(\mathbf{X}'\mathbf{X})']^{-1} = (\mathbf{X}'\mathbf{X})^{-1}$. Por lo tanto, la expresión anterior queda como

$$Var(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mu\mu')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Además, asumiendo el supuesto de homocedasticidad y no autocorrelación, implica que

$$\begin{aligned} Var(\hat{\beta}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_{\mu}^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ Var(\hat{\beta}) &= \sigma_{\mu}^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (3.22)$$

Esta es la matriz de varianzas y covarianzas del vector de estimadores. Esta matriz, en su forma extendida es:

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{\beta}}) &= \sigma_{\mu}^2 \begin{bmatrix} \Sigma X_{i1}^2 & \Sigma X_{i1}X_{i2} & \cdots & \Sigma X_{i1}X_{ik} \\ \Sigma X_{i2}X_{i1} & \Sigma X_{i2}^2 & \cdots & \Sigma X_{i2}X_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma X_{ik}X_{i1} & \Sigma X_{ik}X_{i2} & \cdots & \Sigma X_{ik}^2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \cdots & \text{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_2) & \cdots & \text{Var}(\hat{\beta}_k) \end{bmatrix}
\end{aligned}$$

Con este resultado es posible demostrar que ésta es la mínima varianza para este vector. El procedimiento consiste en utilizar otro estimador lineal e insesgado y demostrar que para que su varianza sea mínima ésta debe ser igual a la varianza del estimador *MCO*. Tomemos entonces un $\boldsymbol{\beta}^*$ cualquiera tal que sea un estimador lineal insesgado de $\boldsymbol{\beta}$ donde

$$\boldsymbol{\beta}^* = \mathbf{A}\mathbf{Y}$$

donde

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}'$$

es decir el estimador es lineal, y \mathbf{C} es una matriz ($n \times k$) de constantes cualquiera. Entonces:

$$\boldsymbol{\beta}^* = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}'] \mathbf{Y}$$

$$\boldsymbol{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} + \mathbf{C}'\mathbf{Y}$$

reemplazando el modelo poblacional $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ en en la ecuación anterior

$$\begin{aligned}
\boldsymbol{\beta}^* &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' [\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}] + \mathbf{C}' [\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}] \\
\boldsymbol{\beta}^* &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu} + \mathbf{C}'\mathbf{X}\boldsymbol{\beta} + \mathbf{C}'\boldsymbol{\mu} \\
\boldsymbol{\beta}^* &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu} + \mathbf{C}'\mathbf{X}\boldsymbol{\beta} + \mathbf{C}'\boldsymbol{\mu}
\end{aligned} \tag{3.23}$$

para que la comparación a realizar tenga sentido el nuevo estimador debe ser insesgado, $\mathbf{E}(\boldsymbol{\beta}^*) = \boldsymbol{\beta}$. Aplicando esperanza a 3.23 se tiene

$$\begin{aligned}
\mathbf{E}(\boldsymbol{\beta}^*) &= \mathbf{E}(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\boldsymbol{\mu}) + \mathbf{C}'\mathbf{X}\boldsymbol{\beta} + \mathbf{C}'\mathbf{E}(\boldsymbol{\mu}) \\
\mathbf{E}(\boldsymbol{\beta}^*) &= \boldsymbol{\beta} + \mathbf{C}'\mathbf{X}\boldsymbol{\beta}
\end{aligned} \tag{3.24}$$

para que cumpla con la propiedad de insesgamiento debe ocurrir que $\mathbf{C}'\mathbf{X} = \mathbf{0}$.

Vale la pena resumir lo que hasta el momento se ha realizado. Sabemos que existe otro estimador denominado β^* , el cual es insesgado. De esta forma no podemos decir cual de los dos estimadores β^* ó $\hat{\beta}$ es mas recomendable utilizar. El criterio de decisión debería ser el de mínima varianza, nosotros conocemos la varianza de $\hat{\beta}$, pero ¿cuál es la varianza de β^* ?

Reordenando la ecuación 3.23 se obtiene una expresión para la matriz de varianzas y covarianzas de β^*

$$\beta^* - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu} + \mathbf{C}'\boldsymbol{\mu}$$

es decir,

$$\begin{aligned} \text{Var}(\beta^*) &= \mathbf{E} [(\beta^* - \mathbf{E}(\beta^*))(\beta^* - \mathbf{E}(\beta^*))'] \\ &= \mathbf{E} [(\beta^* - \beta)(\beta^* - \beta)'] \\ &= \mathbf{E} [((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu} + \mathbf{C}'\boldsymbol{\mu})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu} + \mathbf{C}'\boldsymbol{\mu})'] \\ &= \mathbf{E} [((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu} + \mathbf{C}'\boldsymbol{\mu})(\boldsymbol{\mu}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \boldsymbol{\mu}'\mathbf{C})] \\ &= \mathbf{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{C}'\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\quad + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{C} + \mathbf{C}'\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{C}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{C}'\mathbf{E}(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\quad + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{C} + \mathbf{C}'\mathbf{E}(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{C} \\ &= \sigma_{\mu}^2 (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \sigma_{\mu}^2 \mathbf{C}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\quad + \sigma_{\mu}^2 (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C} + \sigma_{\mu}^2 \mathbf{C}'\mathbf{C} \end{aligned}$$

de 3.24 se sabe que $\mathbf{C}'\mathbf{X} = \mathbf{X}'\mathbf{C} = \mathbf{0}$. Luego

$$\text{Var}(\beta^*) = \sigma_{\mu}^2 ((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{C}\mathbf{C}')$$

El primer componente de la ecuación es igual a la varianza del estimador *MCO* dado en 3.22. El segundo componente es una matriz ($n \times n$) no negativa definida. Por tanto la $\text{Var}(\beta^*) \geq \text{Var}(\hat{\beta})$. Como β^* fue definida en forma arbitraria, este resultado es válido para cualquier otro estimador lineal insesgado de β .

De esta forma hemos probado que el estimador *MCO* es el Mejor Estimador Lineal Insesgado (*MELI*), es decir, tiene la menor varianza dentro del conjunto de estimadores lineales insesgados.

Con la información obtenida hasta aquí existe sólo un aspecto que dificulta la estimación de la matriz de varianzas y covarianzas de los estimadores. Esto es, no conocemos el valor de σ_μ^2 . En la siguiente sección construiremos un estimador insesgado para la varianza del error.

3.4.2. Estimador de la Varianza del Error

Otra información que nos interesa es estimar la varianza del error, denotada por σ_μ^2 . Encontrar un estimador para esta expresión es importante especialmente para poder calcular la matriz de varianzas y covarianzas y poder realizar pruebas de hipótesis sobre los parámetros estimados. Usando la expresión para la función de regresión poblacional y la función de regresión muestral:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \quad (3.25)$$

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon} \quad (3.26)$$

es posible despejar $\boldsymbol{\varepsilon}$ de la función de regresión muestral,

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \quad (3.27)$$

y reemplazar el modelo poblacional de la ecuación 3.25 , luego se tiene:

$$\boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

reemplazando la expresión de la ecuación 3.20 obtenemos

$$\begin{aligned} \boldsymbol{\varepsilon} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} - \mathbf{X} \left[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\mu} \right] \\ \boldsymbol{\varepsilon} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\mu} \\ \boldsymbol{\varepsilon} &= \left[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \boldsymbol{\mu} \end{aligned}$$

Se define $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$, entonces

$$\boldsymbol{\varepsilon} = \mathbf{M}\boldsymbol{\mu} \quad (3.28)$$

\mathbf{M} posee propiedades útiles e interesantes, a saber: es una matriz idempotente, es decir se cumple que $\mathbf{M}' = \mathbf{M}$ y que $\mathbf{M}'\mathbf{M} = \mathbf{M}$. Probemos cada una de estas propiedades:

1. $\mathbf{M}' = \mathbf{M}$

$$(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' = \mathbf{I}_n' - (\mathbf{X}')' [(\mathbf{X}'\mathbf{X})^{-1}]' \mathbf{X}' = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

2. $\mathbf{M}'\mathbf{M} = \mathbf{M}$

$\mathbf{M}'\mathbf{M} = \mathbf{M}\mathbf{M}$

$$\begin{aligned} &= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \end{aligned}$$

Esto prueba que \mathbf{M} es una matriz idempotente⁹.

Retomando nuestro problema central, la ecuación 3.28 expresada en forma cuadrática se puede escribir como:

$$\begin{aligned} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{M}\boldsymbol{\mu})'(\mathbf{M}\boldsymbol{\mu}) \\ \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= \boldsymbol{\mu}'\mathbf{M}'\mathbf{M}\boldsymbol{\mu} \\ \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= \boldsymbol{\mu}'\mathbf{M}\boldsymbol{\mu} \\ \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= \sum_{i=1}^n \sum_{j=1}^n m_{ij} \mu_i \mu_j \end{aligned}$$

donde m_{ij} es el elemento de la fila i y columna j de la matriz \mathbf{M} .

Por lo tanto,

$$\begin{aligned} \mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \mathbf{E}(\boldsymbol{\mu}'\mathbf{M}\boldsymbol{\mu}) \\ \mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \sum_{i=1}^n \sum_{j=1}^n m_{ij} \mathbf{E}(\mu_i \mu_j) \end{aligned}$$

Por los supuestos (2) y (3) se obtiene:

$$\begin{aligned} \mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \sum_{i=1}^n m_{ii} \sigma_\mu^2 \\ \mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \sigma_\mu^2 \sum_{i=1}^n m_{ii} \\ \mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \sigma_\mu^2 \text{tr}(\mathbf{M}) \end{aligned}$$

⁹Donde $\mathbf{I}_n \mathbf{I}_n' = \mathbf{I}_n$.

donde, como se recordará, $tr(\mathbf{M})$ es la traza de la matriz \mathbf{M} .

Recordando que $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ y usando las propiedades de la traza tenemos

$$\begin{aligned}
 tr(\mathbf{M}) &= tr(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\
 &= tr(\mathbf{I}_n) - tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\
 &= \mathbf{n} - tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\
 &= \mathbf{n} - tr\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\right) \\
 &= \mathbf{n} - tr(\mathbf{I}_k) \\
 &= \mathbf{n} - \mathbf{k}
 \end{aligned}$$

entonces

$$\begin{aligned}
 \mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \sigma_\mu^2 tr(\mathbf{M}) \\
 \mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \sigma_\mu^2(\mathbf{n} - \mathbf{k})
 \end{aligned} \tag{3.29}$$

Como puede observarse, el valor esperado de los errores al cuadrado obtenidos de la estimación MCO, no es un estimador insesgado de σ_μ^2 . Sin embargo, este resultado también nos da la clave para generar un estimador insesgado.

De esta forma, se puede usar como estimador insesgado de σ_μ^2 la siguiente expresión:

$$\hat{\sigma}_\mu^2 = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\mathbf{n} - \mathbf{k}} \tag{3.30}$$

ya que

$$\mathbf{E}(\hat{\sigma}_\mu^2) = \mathbf{E}\left(\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\mathbf{n} - \mathbf{k}}\right) = \frac{\mathbf{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\mathbf{n} - \mathbf{k}} = \frac{\sigma_\mu^2(\mathbf{n} - \mathbf{k})}{\mathbf{n} - \mathbf{k}} = \sigma_\mu^2$$

Esta expresión de la varianza del error puede estimarse usando la información muestral contenida en la matriz de variables endógenas y exógenas.

$$\begin{aligned}
 \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
 \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{Y}' - \hat{\boldsymbol{\beta}}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
 \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}) \\
 \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})
 \end{aligned}$$

Usando el hecho que $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

$$\begin{aligned}\varepsilon'\varepsilon &= \mathbf{Y}'\mathbf{Y} - 2\hat{\beta}'\mathbf{X}'\mathbf{Y} + \hat{\beta}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \varepsilon'\varepsilon &= \mathbf{Y}'\mathbf{Y} - 2\hat{\beta}'\mathbf{X}'\mathbf{Y} + \hat{\beta}'\mathbf{X}'\mathbf{Y} \\ \varepsilon'\varepsilon &= \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}\end{aligned}$$

Luego,

$$\hat{\sigma}_\mu^2 = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}}{n - k}$$

Ejemplo 1: Estimación de Varianzas para la Función de Consumo

Estamos en condiciones de retomar el ejemplo de la función de consumo y estimar las varianzas y covarianzas de los estimadores, para ello utilizaremos las matrices 3.15, 3.16 y 3.17.

La suma de la variable dependiente al cuadrado es:

$$\mathbf{Y}'\mathbf{Y} = 266,010$$

De esta forma tenemos:

$$\begin{aligned}\hat{\beta}'(\mathbf{X}'\mathbf{Y}) &= [0.3152 \quad 0,6134] \begin{bmatrix} 93,4357 \\ 384,8900 \end{bmatrix} = 265,542 \\ n - k &= 38 - 2 = 36\end{aligned}$$

Finalmente

$$\hat{\sigma}_\mu^2 = \frac{266,010 - 265,542}{36} = 0,013$$

con este estimador de la varianza del error podemos calcular

$$\begin{aligned}Var(\hat{\beta}) &= \hat{\sigma}_\mu^2 \mathbf{I}_n (\mathbf{X}'\mathbf{X})^{-1} \\ Var(\hat{\beta}) &= 0,013 \begin{bmatrix} 0,1562 & -0,0371 \\ -0,0371 & 0,0106 \end{bmatrix} = \begin{bmatrix} 0,00203 & -0,00049 \\ -0,00049 & 0,00014 \end{bmatrix}\end{aligned}$$

así, la ecuación de consumo queda dada por¹⁰:

¹⁰Tanto $\hat{\beta}_0$ como su desviación estándar están expresados en unidad de millón.

$$C_t = 0,3152 + 0,6134 \times Y_t + \varepsilon_t$$

$$d.e. \quad (0.04506) \quad (0.0118)$$

en donde bajo el coeficiente estimado entre paréntesis se ha incorporado la respectiva desviación estándar de los estimadores.

3.5. Estimador Máximo Verosímil

Alternativamente al método de MCO, podemos usar el método de Máxima Verosimilitud para estimar los parámetros poblacionales. La idea básica de este último método de estimación, es que distintas poblaciones deberían generar muestras diferentes. Por lo tanto, una determinada muestra tiene mayor probabilidad de pertenecer a una población en vez de a otras. El estimador se debería escoger de tal forma que la muestra tenga la más alta probabilidad de pertenecer a la población en cuestión.

El Método de Máxima Verosimilitud, a diferencia del método anterior, requiere un supuesto sobre la distribución de probabilidad de la variable dependiente que permita construir la función de verosimilitud a maximizar. Manteniendo los supuestos clásicos vistos anteriormente, asumiremos que la distribución de probabilidad del error es normal con media cero y varianza constante e igual a σ_μ^2 . Esto se puede expresar como:

$$\mu_i \sim N(0, \sigma_\mu^2)$$

Además, asumiremos que los errores individuales se distribuyen en forma idéntica e independiente (iid). Si usamos el modelo poblacional $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ es posible obtener una función de distribución de probabilidad para \mathbf{Y} , dado que $\mathbf{X}\boldsymbol{\beta}$ se comporta como constante. Entonces, por las propiedades de la esperanza y la varianza:

$$\mathbf{Y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma_\mu^2)$$

Utilizando la función de distribución normal para estimar los parámetros $\boldsymbol{\beta}$ y σ_μ^2 , la función de máximo verosimilitud es:

$$L(Y_i/X_i, \boldsymbol{\beta}, \sigma_\mu^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp \left\{ -\frac{(y_i - x_i\boldsymbol{\beta})^2}{2\sigma_\mu^2} \right\}$$

Utilizando algunas propiedades del logaritmo, podemos simplificar la expresión, de tal manera de dejarla en términos de sólo sumas y multiplicaciones:

$$\begin{aligned} \ln L(Y_i/X_i, \boldsymbol{\beta}, \sigma_\mu^2) &= \sum_{i=1}^n \left[\ln(1) - \frac{1}{2} \ln(2\pi\sigma_\mu^2) + \ln\left(\exp\left\{-\frac{(y_i - x_i\boldsymbol{\beta})^2}{2\sigma_\mu^2}\right\}\right) \right] \\ \ln L(Y_i/X_i, \boldsymbol{\beta}, \sigma_\mu^2) &= -\sum_{i=1}^n \frac{1}{2} \ln(2\pi\sigma_\mu^2) - \sum_{i=1}^n \frac{(y_i - x_i\boldsymbol{\beta})^2}{2\sigma_\mu^2} \end{aligned}$$

Como podemos observar, el primer término es constante, la sumatoria de n veces esa constante es simplemente la multiplicación del término por n . Expresando en forma matricial:

$$\ln L(\boldsymbol{\beta}, \sigma_\mu^2/\mathbf{Y}) = -\frac{n}{2} \ln(2\pi\sigma_\mu^2) - \frac{1}{2\sigma_\mu^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Lo que se busca es maximizar la función de Verosimilitud, para lo cual se debe derivar con respecto a los parámetros $\boldsymbol{\beta}$ y σ^2 e igualar estas derivadas a cero para obtener las condiciones de primer orden. Las derivadas son las siguientes:

$$\frac{\partial \ln L(\boldsymbol{\beta}, \sigma_\mu^2/\mathbf{Y})}{\partial \hat{\boldsymbol{\beta}}} = -\frac{1}{2\sigma_\mu^2} \frac{\partial (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = 0$$

El resultado de esta derivación es formalmente idéntica a la obtenida con MCO (ver ecuación 3.13). Obtenemos:

$$-\frac{1}{2\sigma^2} \left(-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \right) = 0$$

Despejando $\hat{\boldsymbol{\beta}}$ se obtiene una expresión análoga a la obtenida por MCO. Esto quiere decir que los mismos estimadores se obtienen por el método de máxima verosimilitud.

$$\hat{\boldsymbol{\beta}}_{MCO} = \hat{\boldsymbol{\beta}}_{MV} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$$

En el caso de la varianza del error (σ_μ^2) la función de verosimilitud se deriva con respecto de σ_μ^2 y obtenemos:

$$\frac{\partial \ln L(\boldsymbol{\beta}, \sigma_\mu^2/\mathbf{Y})}{\partial \hat{\sigma}_\mu^2} = -\frac{n}{2} \frac{1}{2\pi\sigma_\mu^2} 2\pi + \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{2(\hat{\sigma}_\mu^2)^2} = 0$$

Simplificando se llega al siguiente resultado:

$$\hat{\sigma}_\mu^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n} = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n}$$

que es un estimador sesgado y distinto al de *MCO*.

3.6. Estimación en Desviaciones de Media

Es común en econometría expresar los datos en desviaciones respecto de la media. La estimación por desviaciones de media es interesante, ya que, presenta en forma más compacta las matrices que determinan el vector de parámetros y además, tiene ventajas para la interpretación de los resultados.

Para calcular las desviaciones de media es necesario restar a cada observación, tanto de la variable dependiente como de las variables explicativas, su respectiva media, es decir:

$$y_i = Y_i - \bar{Y}$$

$$x_{ij} = X_{ij} - \bar{X}_j$$

Esta forma de presentar los datos tiene ciertas ventajas entre las cuales se destacan:

- a) Es más fácil el manejo de la información muestral ya que los cálculos con desviaciones de media son menos demandantes que usar los datos originales, en especial si no se cuenta con recursos computacionales.
- b) La interpretación de los datos en desviaciones de media está directamente relacionada con la naturaleza del problema econométrico. En el modelo de regresión lineal se intenta explicar la variabilidad de la variable endógena. Esta variabilidad se expresa en términos de qué tan alejada está la observación con respecto a su media. La ecuación en desviaciones de media expresa en forma precisa este concepto.

Por lo tanto, un modelo expresado en desviaciones de media permitirá distinguir en forma directa qué porción del cambio en la variable dependiente es explicada por el modelo econométrico sugerido por la teoría.

Para expresar los datos en desviaciones respecto de la media considere la matriz

$$\mathbf{X} = \begin{pmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ 1 & X_{13} & X_{23} & \dots & X_{k3} \\ \vdots & & & & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{pmatrix}$$

Se puede eliminar la columna de unos, ya que el promedio de columnas de unos es uno (en la práctica implica eliminar el parámetro de posición) y se resta a todos los elementos de cada columna su respectiva media, quedando la nueva matriz de la siguiente forma:

$$\mathbf{x} = \begin{pmatrix} X_{11} - \bar{X}_1 & X_{21} - \bar{X}_2 & \dots & X_{k1} - \bar{X}_k \\ X_{12} - \bar{X}_1 & X_{22} - \bar{X}_2 & \dots & X_{k2} - \bar{X}_k \\ X_{13} - \bar{X}_1 & X_{23} - \bar{X}_2 & \dots & X_{k3} - \bar{X}_k \\ \vdots & \vdots & \vdots & \vdots \\ X_{1n} - \bar{X}_1 & X_{2n} - \bar{X}_2 & \dots & X_{kn} - \bar{X}_k \end{pmatrix} = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{k1} \\ x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}$$

donde $\bar{X}_i = \frac{\sum_{j=1}^n \bar{X}_{ij}}{n}$.

Hacemos las mismas operaciones con las matrices $\boldsymbol{\mu}$ e \mathbf{Y}

$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 - \bar{\mu} \\ \mu_2 - \bar{\mu} \\ \vdots \\ \mu_n - \bar{\mu} \end{pmatrix} = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \\ \vdots \\ \tilde{\mu}_n \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Note que para un variable particular X_i podemos expresar la suma de todos sus elementos como:

$$\sum X_i = \mathbf{i}'\mathbf{X}_i$$

Donde \mathbf{i} es un vector de unos.

Adicionalmente se tiene que

$$\sum aX_i = a\mathbf{i}'\mathbf{X}_i$$

si $a = 1/n$ entonces

$$\sum aX_i = \frac{1}{n}\mathbf{i}'\mathbf{X}_i$$

y

$$\begin{aligned}\bar{X}_i &= \frac{\sum X_i}{n} = \frac{1}{n} \mathbf{i}' \mathbf{X}_i \\ i\bar{X}_i &= i \frac{\sum X_i}{n} = \mathbf{i} \frac{1}{n} \mathbf{i}' \mathbf{X}_i = \frac{1}{n} \mathbf{ii}' \mathbf{X}_i\end{aligned}$$

Esta expresión es útil para expresar las matrices en desviaciones de medias. Es decir:

$$\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_n \end{bmatrix} = \frac{1}{n} \mathbf{ii}' \mathbf{X}$$

y

$$\begin{aligned}\mathbf{x} &= \begin{bmatrix} X_1 - \bar{X}_1 \\ X_2 - \bar{X}_2 \\ \vdots \\ X_n - \bar{X}_n \end{bmatrix} = \mathbf{X} - \mathbf{i}\bar{\mathbf{X}} = \mathbf{X} - \frac{1}{n} \mathbf{ii}' \mathbf{X} \\ \mathbf{x} &= \left[I - \frac{1}{n} \mathbf{ii}' \right] \mathbf{X} = \mathbf{M}_0 \mathbf{X}\end{aligned}\tag{3.31}$$

Donde $\mathbf{M}_0 = I - \frac{1}{n} \mathbf{ii}'$.¹¹

Si retomamos el problema de estimación tenemos una ecuación para la Función de Regresión Poblacional expresada en desviaciones de media. Sabemos por la ecuación 3.1 que la Función de Regresión Poblacional puede escribirse

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \mu_i$$

Si aplicamos el operador de la esperanza, obtenemos

$$\mathbf{E}(Y_i) = \beta_0 + \beta_1 \mathbf{E}(X_{i1}) + \dots + \beta_k \mathbf{E}(X_{ik}) + \mathbf{E}(\mu_i)\tag{3.32}$$

Entonces (1) – (2) da

¹¹Nótese que M_0 es una matriz simétrica e idempotente.

$$Y_i - \mathbf{E}(Y_i) = \beta_1 (X_{i1} - \mathbf{E}(X_{i1})) + \dots + \beta_k (X_{ik} - \mathbf{E}(X_{ik})) + (\mu_i - \mathbf{E}(\mu_i))$$

o lo que es lo mismo en forma compacta

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \tilde{\boldsymbol{\mu}}$$

y una Función de Regresión Muestral

$$\mathbf{y} = \mathbf{x}\hat{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}$$

donde $\tilde{\boldsymbol{\varepsilon}}$ es la diferencia entre ε_i y $\bar{\varepsilon}$.

Se puede mostrar que $\bar{\varepsilon} = 0$. Para ello considere las ecuaciones normales del proceso de minimización de MCO del modelo lineal general :

$$\begin{aligned} \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}'\mathbf{Y} &= \mathbf{0} \\ -\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= \mathbf{0} \\ -\mathbf{X}'\boldsymbol{\varepsilon} &= \mathbf{0} \end{aligned}$$

De donde se puede deducir que:

$$\sum \varepsilon_i = 0 \tag{3.33}$$

$$\sum \varepsilon_i X_{1i} = 0$$

$$\sum \varepsilon_i X_{2i} = 0$$

$$\vdots$$

$$\sum \varepsilon_i X_{ki} = 0 \tag{3.34}$$

Por lo tanto, el nuevo vector de errores queda de la siguiente manera:

$$\tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 - \bar{\varepsilon} \\ \varepsilon_2 - \bar{\varepsilon} \\ \vdots \\ \varepsilon_n - \bar{\varepsilon} \end{pmatrix}$$

Donde $\bar{\varepsilon} = \frac{\sum \varepsilon_i}{n}$, pero de (3.33) se sabe que $\sum \varepsilon_i = 0$, entonces el nuevo vector $\tilde{\varepsilon}$ queda

$$\tilde{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \varepsilon$$

Desde luego, lo que se busca es minimizar la sumatoria de residuos al cuadrado:

$$\begin{aligned} \text{MIN } \sum \varepsilon_i^2 &= \text{MIN } (\varepsilon' \varepsilon) \\ \varepsilon' \varepsilon &= (\mathbf{y} - \mathbf{x}\hat{\beta})'(\mathbf{y} - \mathbf{x}\hat{\beta}) \\ \varepsilon' \varepsilon &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{x}\hat{\beta} - \hat{\beta}'\mathbf{x}'\mathbf{y} + \hat{\beta}'\mathbf{x}'\mathbf{x}\hat{\beta} \\ \varepsilon' \varepsilon &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{x}'\mathbf{y} + \hat{\beta}'\mathbf{x}'\mathbf{x}\hat{\beta} \end{aligned}$$

derivando respecto del vector de parámetros e igualando el resultado a cero se tiene

$$\begin{aligned} \frac{\partial \varepsilon' \varepsilon}{\partial \hat{\beta}} &= -2\mathbf{x}'\mathbf{y} + 2\mathbf{x}'\mathbf{x}\hat{\beta} = 0 \\ \hat{\beta} &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \end{aligned}$$

que es el nuevo estimador *MCO* usando los datos en desviaciones de media. Aunque la formulación parece idéntica a la ecuación 3.13, en este caso no se obtiene una estimación para el parámetro de posición, $\hat{\beta}_0$, ya que este ha sido eliminado en el proceso de transformación de los datos. Como puede observarse de la ecuación 3.32, para obtener este estimador basta con calcular:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}\bar{X}$$

Esta forma de expresar los datos en desviaciones de medias permite calcular los estimadores con una matriz mas pequeña que en el caso general, pero su utilidad es mas clara al momento de estudiar los criterios de bondad de ajuste del modelo lineal general.

3.7. Criterios de Bondad del Ajuste

Se denomina *Criterios de Bondad del Ajuste* a estadísticos que permiten evaluar la capacidad que tiene el modelo de explicar la variación de la variable dependiente. En modelos lineales, el criterio más utilizado es el denominado R^2 (“R cuadrado”).

Para entender qué es el R^2 retomemos el modelo en desviaciones de medias:

$$\mathbf{y} = \mathbf{x}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$$

Si consideramos una sola variable explicativa el modelo se reduce a

$$y_i = \hat{\beta}x_i + \varepsilon_i$$

es decir,

$$Y_i - \bar{Y} = \hat{\beta}(X_i - \bar{X}) + \varepsilon_i$$

Elevando al cuadrado y aplicando sumatoria tenemos

$$\begin{aligned} \sum (Y_i - \bar{Y})^2 &= \sum \hat{\beta}^2 (X_i - \bar{X})^2 + \sum \varepsilon_i^2 \\ STC &= SEC + SRC \end{aligned} \quad (3.35)$$

Note que el producto de $\sum x_i \varepsilon_i = 0$ (ver ec.3.33).

El primer término se denomina *Sumatoria Total de Cuadrados (STC)* y es una medida de la variabilidad de la variable dependiente respecto de su media. Esto es lo que deseamos explicar. Esta variación se descompone en:

- (1) Un componente explicado por el modelo ($\sum (X_i - \bar{X})^2 \hat{\beta}^2$), que conocemos como *Sumatoria Explicada de Cuadrados (SEC)* y
- (2) Un componente no explicado ($\sum \varepsilon_i^2$) llamado *Sumatoria Residual de Cuadrados (SRC)*.

El aporte del modelo teórico sugerido es solamente la *SEC*, ya que la *SRC* refleja nuestra ignorancia respecto de otras variables o factores que explican los cambios en nuestra variable dependiente. De esta forma, el R^2 se define como:

$$R^2 = \frac{SEC}{STC} \quad (3.36)$$

Mientras más cercano a uno sea el valor de R^2 , entonces mayor será el poder explicativo del modelo. Este criterio indica el porcentaje de la varianza de la variable dependiente que es “explicada” (en un sentido estadístico) por el modelo y puede estar en el rango de 0 a 1. Es decir, se mueve entre un cero o cien por ciento de la “explicación” de la variable dependiente.

Note que dicha definición se mantiene en el caso de más variables explicativas, es decir en el caso del modelo lineal general,

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$$

y usando la ecuación 3.31 tenemos que

$$\mathbf{M}_0\mathbf{Y} = \mathbf{M}_0\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{M}_0\boldsymbol{\varepsilon}$$

donde $\mathbf{M}_0\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}$. Expresado en forma cuadrática

$$\begin{aligned} (\mathbf{M}_0\mathbf{Y})'\mathbf{M}_0\mathbf{Y} &= (\mathbf{M}_0\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{M}_0\boldsymbol{\varepsilon})' (\mathbf{M}_0\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{M}_0\boldsymbol{\varepsilon}) \\ \mathbf{Y}'\mathbf{M}'_0\mathbf{M}_0\mathbf{Y} &= (\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{M}'_0 + \boldsymbol{\varepsilon}'\mathbf{M}'_0) (\mathbf{M}_0\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{M}_0\boldsymbol{\varepsilon}) \\ \mathbf{Y}'\mathbf{M}_0\mathbf{Y} &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{M}'_0\mathbf{M}_0\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{M}'_0\mathbf{M}_0\boldsymbol{\varepsilon} + \\ &\quad \boldsymbol{\varepsilon}'\mathbf{M}'_0\mathbf{M}_0\mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}'\mathbf{M}'_0\mathbf{M}_0\boldsymbol{\varepsilon} \\ \mathbf{Y}'\mathbf{M}_0\mathbf{Y} &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{M}'_0\mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \end{aligned}$$

Donde se ha hecho uso de $\mathbf{M}'_0\mathbf{M}_0 = \mathbf{M}_0$ y $\boldsymbol{\varepsilon}'\mathbf{M}_0\mathbf{X} = \boldsymbol{\varepsilon}'\mathbf{X} = \mathbf{0}$.

Luego la $STC = \mathbf{Y}'\mathbf{M}_0\mathbf{Y}$, la $SEC = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{M}'_0\mathbf{X}\hat{\boldsymbol{\beta}}$ y la $SRC = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$.

Y el R^2 queda definido en forma matricial como:

$$\begin{aligned} R^2 &= \frac{SEC}{STC} = \frac{STC - SRC}{STC} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{M}'_0\mathbf{X}\hat{\boldsymbol{\beta}}}{\mathbf{Y}'\mathbf{M}_0\mathbf{Y}} \\ &= 1 - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\mathbf{Y}'\mathbf{M}_0\mathbf{Y}} \end{aligned}$$

Expresada en forma más sencilla se tiene:

$$\begin{aligned} STC &= \sum_{i=1}^n y_i^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = \mathbf{Y}'\mathbf{Y} - n\bar{Y}^2 \\ SRC &= \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} \end{aligned}$$

entonces

$$R^2 = \frac{STC - SRC}{STC} = \frac{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2 - (\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y})}{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2} \quad (3.37)$$

$$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2}{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2}$$

También es posible expresar su valor en términos de desviaciones de medias, para lo cual usamos una forma conveniente de expresar la sumatoria de errores al cuadrado tal como:

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}} &= (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}) = (\mathbf{y}' - \hat{\boldsymbol{\beta}}'\mathbf{x}')(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} - \mathbf{y}'\mathbf{x}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{x}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{x}\hat{\boldsymbol{\beta}} \end{aligned}$$

$$R^2 = \frac{\sum y_i^2 - \sum \varepsilon_i^2}{\sum y_i^2} = \frac{\mathbf{y}'\mathbf{y} - \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{y} + \mathbf{y}'\mathbf{x}\hat{\boldsymbol{\beta}}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{y}'\mathbf{x}\hat{\boldsymbol{\beta}}}{\mathbf{y}'\mathbf{y}} \quad (3.38)$$

$$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} \quad (3.39)$$

Uno de los problemas que presenta el R^2 es que su valor aumenta si se incrementa la cantidad de variables explicativas. Esto se deriva del hecho que $E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = \sigma^2(n - k)$ (ver ecuación 3.29). Entonces al aumentar k , para un número fijo de observaciones n , $E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})$ se reducirá y el R^2 aumentará. Ello sucederá aunque las nuevas variables explicativas agregadas no aporten a la “explicación” de la variabilidad de la variable dependiente. Esto hace poco confiable este criterio de bondad de ajuste. Para evitar esto se ha propuesto un R^2 corregido, el cual considera los grados de libertad que tienen y pierden las variables. Retomando la definición inicial

$$R^2 = 1 - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\mathbf{Y}'\mathbf{M}_0\mathbf{Y}} = 1 - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\mathbf{y}'\mathbf{y}} = 1 - \frac{\sum \varepsilon_i^2}{\sum y_i^2}$$

se puede dividir la SRC y la STC por sus respectivos grados de libertad dando origen a un R^2 corregido de la forma

$$\bar{R}^2 = 1 - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/(n-k)}{\mathbf{y}'\mathbf{y}/(n-1)} = 1 - \frac{\sum \varepsilon_i^2/(n-k)}{\sum y_i^2/(n-1)}$$

Esta medida considera el efecto de agregar más variables o más observaciones al modelo. Como es posible observar, al aumentar el número de variables explicativas (k), sin reducir la sumatoria residual de cuadrados, aumenta el término $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/(n-k)$ y por lo tanto se ve reducido el \bar{R}^2 . Este indicador permite la comparación del poder explicativo de modelos donde se agregan o quitan variables.

La relación entre R^2 y \bar{R}^2 se puede obtener de la siguiente forma:

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{\sum \varepsilon_i^2/(n-k)}{\sum y_i^2/(n-1)} \\ &= 1 - \frac{(n-1) \sum \varepsilon_i^2}{(n-k) \sum y_i^2} \\ &= 1 - \frac{(n-1)}{(n-k)} (1 - R^2) \end{aligned}$$

de donde es posible observar que siempre $\bar{R}^2 \leq R^2$.

Además, cabe mencionar que \bar{R}^2 no está delimitado a valores iguales o mayores que cero. Cuando el modelo incluye muchas variables con muy bajo grado explicativo \bar{R}^2 puede ser negativo.

Ejemplo 1: La Función de Consumo Keynesiana

En este punto estamos en condiciones de presentar el ejemplo 1 completo. Usando el software *E-views* y las cifras de consumo e ingreso adjuntadas en el archivo consumo.xls se obtienen los siguientes resultados:

$$\begin{aligned} C_t &= 0,3152 + 0,6134 \times Y_t + \varepsilon_t \\ &\quad (0,04506) \quad (0,0118) \\ n &= 38 \\ R^2 &= 0,9869 \\ \bar{R}^2 &= 0,9866 \end{aligned}$$

donde C_t representa al consumo privado en el período t , Y_t es el ingreso real disponible en el período t . Las cifras entre paréntesis corresponden a la desviación estándar del coeficiente estimado (la raíz cuadrada de la varianza).

El primer parámetro, $\hat{\beta}_1$, corresponde al consumo autónomo. Es decir, si el ingreso fuese igual a cero, de todos modos existiría un consumo igual a 3152 millones de pesos¹². El valor de $\hat{\beta}_2$, igual a 0.6134, se interpreta de la siguiente forma: si el ingreso aumenta en una unidad el consumo se incrementa en 0.6134 unidades. O sea si el ingreso real disponible se incrementara en 100 millones de pesos, el modelo predice que el consumo privado aumentaría en 61.34 millones de pesos. Esto desde el punto de vista económico se conoce como la *Propensión Marginal al Consumo*.

Al evaluar los coeficientes de determinación (R^2 normal y ajustado), podemos concluir que la estimación obtenida se trata de una buena aproximación, explicando el 98,66 por ciento de la variabilidad del consumo. Además, como es posible comprobar, el valor del \bar{R}^2 es (aunque marginalmente) menor que el del R^2 . Esto es consecuencia que $k = 2$ en este caso.

Ejemplo 2: Estimación de la Función de Importaciones

Un caso interesante de analizar es la Función de Importaciones. Recordemos que, en términos generales, el nivel de importaciones depende del PIB y del Tipo de Cambio Real (TCR). Esto se debe a que:

- a) Un mayor nivel de producción genera mayores ingresos a nivel nacional. Estos ingresos permiten aumentar nuestras compras en el exterior, es decir, elevar el nivel de importaciones. De esta forma, existe una relación directa entre PIB e importaciones.
- b) El Tipo de Cambio Real refleja la cantidad de bienes nacionales que deberíamos sacrificar para obtener una unidad de bienes importados. Si se tiene un TCR elevado, nuestra capacidad de compra del exterior se deteriora, ya que cada unidad importada resulta más cara. Este hecho nos demuestra, entonces, que a medida que el tipo de cambio real crece el nivel de importaciones disminuye, existiendo así una relación inversa entre ellos. Sin embargo, no es obvio que el efecto de un cambio en el TCR sobre las importaciones se deje sentir completamente en el mismo período. Por ello es importante en la especificación del modelo permitir los efectos rezagados del TCR sobre las importaciones.

¹²Expresado en moneda constante de 1986.

Aclarada la relación existente entre las variables, procedemos a estimar la Función de Importaciones a partir de los datos¹³ tomados de las Cuentas Nacionales para el periodo comprendido entre el segundo trimestre del año 1990 y el tercer trimestre del año 2000 los cuáles puede encontrar en el archivo importaciones.xls . Los resultados se muestran a continuación:

$$\begin{aligned} \ln(M_t) &= 0,47078 + 0,99234 \ln(PIB_t) - 0,01299 TCR_{t-1} + \varepsilon_t \\ &\quad (2,6156) \quad (0,1638) \quad (0,0031) \\ n &= 42 \\ R^2 &= 0,9617 \\ \bar{R}^2 &= 0,9598 \end{aligned}$$

donde M_t es el nivel de importaciones en el periodo t , PIB_t es el producto interno bruto en el periodo t , TCR_{t-1} es el tipo de cambio real para el periodo $t - 1$, y \ln indica el logaritmo natural. Tal como en la función de consumo, las cifras entre paréntesis corresponden a la desviación estándar de cada estimador, es decir, a la raíz cuadrada de la varianza.

Considerando el R^2 la regresión explica un 96.17% el comportamiento de las importaciones en relación a los cambios que se produzcan en el producto y en el tipo de cambio real en los distintos periodos considerados.

Para poder ver el efecto que tiene la inclusión de más variables sobre el R^2 se estimará la función de importaciones sin considerar el efecto que puede tener sobre ellas el tipo de cambio real (TCR). Los resultados se muestran a continuación:

$$\begin{aligned} \ln(M_t) &= -9,97143 + 1,63785 \ln(PIB_t) + \varepsilon_t \\ &\quad (0,89613) \quad (0,06263) \\ n &= 42 \\ R^2 &= 0,9447 \\ \bar{R}^2 &= 0,9434 \end{aligned}$$

Como se esperaba, el coeficiente de determinación R^2 disminuye al reducir el número de variables explicativas, o dicho de otra manera, el valor de R^2 se incrementa al aumentar el número de variables explicativas. Estos

¹³Fuente: Boletín Mensual del Banco Central de Chile, su página web, Indicadores Económicos. Informe Económico y Financiero.

resultados apoyan el hecho que esta medida puede no ser muy confiable y se hace necesario considerar como mejor medida el \bar{R}^2 . Puede observar que si bien este último también cae al omitir una variable, lo hace en menor medida que el R^2 .

3.8. Inferencia

En esta sección se discutirán algunos conceptos relacionados con las pruebas de significancia individual, pruebas de significancia global, y predicción del modelo lineal general.

Si bien en las secciones anteriores hemos visto como se pueden estimar los coeficientes de modelos teóricos, no debemos perder de vista la naturaleza estadística de estos resultados.

Considere el ejemplo 1, donde se obtuvo un coeficiente estimado de 0.6137 para la Propensión Marginal al Consumo. Sin embargo el parámetro estimado es a su vez una variable aleatoria que puede asumir un rango posible de valores con mayor o menor probabilidad. Por tanto es posible que el valor del parámetro poblacional no sea 0.6137, sino 0.5 ó 0.7. Es más, a menos que tengamos alguna forma de comprobarlo, no sabemos si el parámetro podría ser cero. Para poder determinar la probabilidad de un evento de esta naturaleza, debemos desarrollar la inferencia estadística que, entre otras cosas, nos permite establecer Pruebas de Hipótesis para dilucidar este tipo de problemas.

3.8.1. Pruebas de Hipótesis e Intervalos de Confianza.

Recordemos que la matriz de varianzas y covarianzas de los parámetros estimados se calcula como (ver ecuación 3.22):

$$Var(\hat{\beta}) = \sigma_{\mu}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

donde la varianza del primer parámetro estimado viene dado por:

$$Var(\hat{\beta}_1) = \sigma_{\mu}^2 a_{11}$$

siendo a_{11} el primer elemento de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. En el caso de un parámetro cualquiera, la varianza es:

$$Var(\hat{\beta}_i) = \sigma_{\mu}^2 a_{ii}$$

A diferencia de la estimación por Máxima Verosimilitud, en el método de Mínimos Cuadrados ordinarios no es necesario realizar supuestos sobre que distribución tienen los errores del modelo. Basta con asumir que esta distribución tiene media cero y varianza constante. Sin embargo, cuando se pretende realizar inferencia estadística se debe incorporar un supuesto adicional al modelo lineal general. Este supuesto es conocido como el *supuesto de normalidad* e implica asumir que el error se distribuye en forma normal. La idea básica es que si queremos conocer con que probabilidad es posible observar un determinado valor del parámetro, debemos conocer la forma de su distribución. Lo anterior se resume en que

$$\mu \sim N(0, \sigma_\mu^2)$$

Dado este supuesto, es posible concluir que los parámetros estimados se distribuyen en forma normal con media β y varianza $\sigma_\mu^2(\mathbf{X}'\mathbf{X})^{-1}$, es decir:

$$\hat{\beta} \sim \mathbf{N}(\beta, \sigma_\mu^2(\mathbf{X}'\mathbf{X})^{-1})$$

Anteriormente demostramos que $\hat{\beta}$ es un estimador insesgado, la varianza de $\hat{\beta}$ también ha sido obtenida previamente. Finalmente, sabemos que la regresión es lineal en los parámetros. Ello significa que $\hat{\beta}$ es una combinación lineal de los valores de Y_i . De acuerdo a un teorema básico de estadística, si los Y_i se distribuyen en forma normal, entonces cualquier combinación lineal de estos también se distribuirá en forma normal.

Si se conociera el valor de σ_μ^2 , se podrían realizar pruebas de hipótesis directamente usando la distribución normal. Sin embargo, este valor no es conocido y solo tenemos un estimador para éste. La solución a este problema consiste en crear una variable adicional Chi-Cuadrado, con la cual transformamos nuestra distribución normal original en una distribución t de student, en la cual podemos usar el estimador $\hat{\sigma}_\mu^2$ en reemplazo de σ_μ^2 .

Veamos primero la creación de la variable Chi-Cuadrado. Para hacer esto considere el supuesto inicial $\mu_i \sim N(0, \sigma_\mu^2)$. y recordando la información discutida en el capítulo 2, se puede estandarizar esta variable dividiendo por su desviación estándar

$$\frac{\mu_i}{\sqrt{\sigma_\mu^2}} \sim N(0, 1)$$

Si elevamos esto al cuadrado construimos una variable distribuida Chi-

Cuadrado con 1 grado de libertad.

$$\left(\frac{\mu_i}{\sqrt{\sigma_\mu^2}}\right)^2 = \frac{\mu_i^2}{\sigma_\mu^2} \sim \chi^2(1)$$

La suma de los n términos de error al cuadrado (observaciones), tendrá también comportamiento Chi-Cuadrado, pero con n grados de libertad.

$$\sum_i^n \frac{\mu_i^2}{\sigma_\mu^2} \sim \chi^2(n)$$

En este punto podemos realizar el siguiente reemplazo

$$\sum_i^n \mu_i^2 \approx \sum_i^n \varepsilon_i^2$$

La ecuación anterior nos queda ¹⁴

$$\frac{\sum_i^n \varepsilon_i^2}{\sigma_\mu^2} \sim \chi^2(n - k)$$

y recordando que

$$\hat{\sigma}_\mu^2 = \frac{\sum_i^n \varepsilon_i^2}{n - k}$$

se puede despejar $\sum_i^n \varepsilon_i^2 = \hat{\sigma}_\mu^2(n - k)$, luego podemos concluir que

$$\frac{\sum_i \varepsilon_i^2}{\sigma_\mu^2} = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\sigma_\mu^2} = \frac{(n - k)\hat{\sigma}_\mu^2}{\sigma_\mu^2} \sim \chi^2(n - k) \quad (3.40)$$

Esta variable nos permite solucionar el problema mencionado en el párrafo anterior, ya que ahora se puede construir una variable con distribución t de Student. Los pasos son los siguientes:

¹⁴Perdemos k grados de libertad ya que para estimar la sumatoria de cuadrados residuales fue necesario calcular los k coeficientes de la regresión.

1. Dado que $\hat{\beta} \sim \mathbf{N}(\beta, \sigma_\mu^2(\mathbf{X}'\mathbf{X})^{-1})$ se construye una distribución normal estándar de la forma

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma_\mu^2(\mathbf{X}'\mathbf{X})^{-1}}}$$

que para el caso particular de un solo coeficiente es:

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma_\mu^2 a_{ii}}} \quad (3.41)$$

2. Se sabe del capítulo 2 que una distribución normal estándar, dividida por la raíz de una Chi-cuadrado a su vez dividida por sus grados de libertad, es una distribución t-student. Por tanto si utilizamos 3.41 y 3.40, obtenemos

$$\frac{\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma_\mu^2 a_{ii}}}}{\sqrt{\frac{(n-k)\hat{\sigma}_\mu^2/\sigma_\mu^2}{n-k}}} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}_\mu^2 a_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}}} \sim t_{(n-k)}$$

Luego, con esta nueva variable aleatoria construimos un intervalo de confianza para realizar las pruebas de hipótesis. El intervalo de confianza con un 95% de confianza viene dado por

$$\Pr\left(-1,96 < \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}}} < 1,96\right) = 0,95$$

donde los valores se obtienen de una tabla estadística para el valor de t . En este caso, se escogió un test con dos colas.

Sin embargo, regularmente se usa la prueba puntual y no el intervalo de confianza. Para el caso de las pruebas puntuales existe una regla práctica que consiste en comparar el valor t-calculado igual a

$$t_c = \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}}}$$

con el valor asociado al t de tabla (t_t). En el caso de que el número de observaciones sea mayor que 20 se utiliza como regla aproximada un valor

de tabla o valor crítico de 2. Esto porque como puede verse en la tabla de la distribución t , para una prueba de dos colas con un 95 % de confianza el valor de tabla con 20 grados de libertad es 2.086. El criterio de decisión implica que si $t_c > t_t$, entonces se rechaza la hipótesis nula cualquiera que esta sea. Como ésta es una distribución simétrica importa sólo el valor absoluto.

Ejemplo 1: Hipótesis en la Función de Consumo

- Recuerde la Función de Consumo dada en Ejemplo 1 Sección 3.4. Para verificar la significancia de los parámetros se analiza la Prueba de Significancia Individual medida a través de la Prueba t . En el caso que estemos interesados en probar si el parámetro estimado $\hat{\beta}_2$ es distinto de cero con un 95 % de confianza, podemos establecer la hipótesis nula, H_0 , de la siguiente forma:

$$\begin{aligned} H_0 & : \beta_2 = 0 \\ H_1 & : \beta_2 \neq 0 \end{aligned}$$

donde H_1 es la hipótesis alterna.

El t calculado se encuentra de la siguiente forma:

$$t^c = \frac{\hat{\beta}_2 - \beta_2}{S_{\hat{\beta}_2}} = \frac{0,6137 - 0}{0,0118} = 52,0085$$

Es posible observar que el t calculado es mayor que el t de tabla¹⁵. El estadístico t rechaza la hipótesis nula. Es decir, el incluir la variable ingreso en la regresión, resulta ser significativa. O en otras palabras, con una muy alta probabilidad se rechaza la posibilidad que el valor (“verdadero”) del parámetro poblacional sea cero. De esta forma, se valida el signo que lo acompaña (en este caso positivo). Este hecho es importante ya que, como se deduce de lo dicho anteriormente, es acorde con la teoría económica en el sentido que existe una relación directa entre ingreso y consumo. Así, cuando el ingreso real disponible aumenta en una unidad en el periodo t , el consumo en igual periodo también se incrementará.

¹⁵Para una muestra de 38 observaciones y un 95 % de confianza el t de tabla es de 2,021.

También es posible verificar si un parámetro es o no significativo a través de un intervalo de confianza. Cualquiera de los dos métodos, la prueba t individual o el intervalo de confianza, nos conducen a la misma conclusión.

El intervalo de confianza se encuentra de la siguiente forma:

$$\left(\hat{\beta}_2 - t_{\frac{\alpha}{2}} S_{\hat{\beta}_2} \leq \beta_2 \leq \hat{\beta}_2 + t_{\frac{\alpha}{2}} S_{\hat{\beta}_2} \right)$$

donde $S_{\hat{\beta}_2}$ es la desviación estándar del estimador $\hat{\beta}_2$, reemplazando los correspondientes valores tenemos:

$$(0,6137 - 2,021 \times 0,0118 \leq \beta_2 \leq 0,6137 + 2,021 \times 0,0118)$$

$$(0,5899 \leq \beta_2 \leq 0,6375)$$

Podrá observar que dentro del intervalo no se encuentra la posibilidad que β_2 tome el valor cero. Este resultado coincide con lo interpretado en la prueba t individual ya desarrollada. Por lo tanto, podrá hacer uso del método que usted desee para probar la significancia de los parámetros. A veces el intervalo de confianza puede ser preferido desde un punto de vista visual. En el ejemplo anterior, la prueba no sólo dice que $\beta_2 \neq 0$, sino además aproximadamente en qué rango se encuentra el valor verdadero con un 95% de confianza. Se observa que este valor debería estar entre 0.59 y 0.64, información que no se obtiene directamente con la prueba puntual. En otras ocasiones, cuando lo interesante es probar sólo si la H_0 es válida o no, puede ser preferible la prueba puntual.

Ejemplo 2: Hipótesis en la Función de Importaciones.

Llevaremos a cabo ésta prueba sobre los dos parámetros más relevantes de la Función de Importaciones¹⁶, en éste caso respecto a $\hat{\beta}_1$ (PIB) y $\hat{\beta}_2$ (TCR_{t-1}).

Las pruebas de hipótesis para β_1 quedan planteadas como:

$$\begin{aligned} H_0 & : \beta_1 = 0 \\ H_1 & : \beta_1 \neq 0 \end{aligned}$$

El estadístico t lo hallamos como se muestra a continuación:

¹⁶Ver ejemplo 2.

$$t^c = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} = \frac{0,99234 - 0}{0,1638} = 6.0582$$

:

Las pruebas de hipótesis para β_2 son:

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

El estadístico t será entonces:

$$t^c = \frac{\hat{\beta}_3 - \beta_3}{S_{\hat{\beta}_3}} = \frac{-0,01299 - 0}{0,0031} = -4.1903$$

Si tomamos el valor absoluto los valores del estadístico t encontrado para cada uno de los parámetros y lo comparamos con el de tabla podremos comprobar que en ambos casos el t calculado es más grande que el t de tabla, ubicándose en la zona de rechazo. Es decir, se rechaza la H_0 y se acepta la H_1 , lo cual significa que con un 95% de confianza se puede decir que el valor de los parámetros poblacionales es distinto de cero. Interpretando este resultado se valida el signo que posee el coeficiente estimado de β_1 y β_2 .

3.8.2. Test T Generalizado

Existen también pruebas de hipótesis de tipo compuestas, en las cuales se desea realizar una prueba de hipótesis sobre varios parámetros a la vez. Por ejemplo, cuando se estima una función de producción del tipo Cobb Douglas, generalmente se está interesado en comprobar la hipótesis de rendimientos constantes a escala. Como recordará, para que exista rendimientos constantes a escala en esta función, se requiere que la suma de los parámetros asociados a los factores productivos sea igual a 1. Entonces si la función de producción es:

$$Y = AX_1^\alpha X_2^\beta X_3^\gamma$$

donde Y es el nivel de producción y X_1 , X_2 y X_3 son los factores productivos, la prueba de hipótesis es que la suma de los coeficientes sea igual a 1.

$$\alpha + \beta + \gamma = 1$$

Para este tipo de hipótesis utilizamos la llamada *Prueba T Generalizada*. En esta prueba se utiliza un vector columna \mathbf{t} de constantes, de tal forma que la hipótesis se plantea en términos de

$$\mathbf{t}'\hat{\boldsymbol{\beta}} = t_1\hat{\beta}_1 + t_2\hat{\beta}_2 + \dots + t_k\hat{\beta}_k$$

donde, los t_i pueden tomar cualquier valor.

Para poder realizar la prueba, primero necesitamos encontrar la distribución de $\mathbf{t}'\hat{\boldsymbol{\beta}}$. Ya que, el vector de estimadores tiene una distribución normal y t es un vector de constantes, entonces la esperanza y varianza para $\mathbf{t}'\hat{\boldsymbol{\beta}}$ están dadas por¹⁷:

$$E(\mathbf{t}'\hat{\boldsymbol{\beta}}) = \mathbf{t}'E(\hat{\boldsymbol{\beta}}) = \mathbf{t}'\boldsymbol{\beta}$$

$$Var(\mathbf{t}'\hat{\boldsymbol{\beta}}) = \mathbf{t}'var(\hat{\boldsymbol{\beta}})\mathbf{t} = \mathbf{t}'\sigma_\mu^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t} = \sigma_\mu^2\mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}$$

Una vez que tenemos estos datos, podemos observar que:

$$\mathbf{t}'\hat{\boldsymbol{\beta}} \sim N(\mathbf{t}'\boldsymbol{\beta}, \sigma_\mu^2 \mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t})$$

y

$$\frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\sigma_\mu^2\mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}}} \sim N(0, 1)$$

Nuevamente, debido a que no conocemos el valor de σ^2 , construimos una variable aleatoria t , de la siguiente manera:

$$\frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\sigma_\mu^2\mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}}} = \frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\hat{\sigma}_\mu^2\mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}}} \sim t_{(n-k)}$$

$$\sqrt{\frac{(n-k)\frac{\hat{\sigma}_\mu^2}{\sigma_\mu^2}}{n-k}}$$

Observe que la expresión $\mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}$ es una matriz de orden 1×1 , por lo tanto la operación de raíz cuadrada no presenta ningún problema.

¹⁷Para un vector X de variables aleatorias de orden $k \times 1$, la forma $AX + b$, con A y b matrices de constantes de orden $n \times k$ y $n \times 1$ respectivamente, se tiene

1. $E(AX + b) = AE(X) + b$
2. $Var(AX + b) = Avar(X)A'$

Volviendo al ejemplo con que comenzamos este punto, nuestro vector de constantes \mathbf{t} , será:

$$\mathbf{t} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

De esta forma tenemos que:

$$\mathbf{t}'\boldsymbol{\beta} = [1 \quad 1 \quad 1] \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \alpha + \beta + \gamma$$

Así el problema se resume en las siguientes pruebas de hipótesis:

$$\begin{aligned} H_0 &: \mathbf{t}'\boldsymbol{\beta} = 1 \\ H_1 &: \mathbf{t}'\boldsymbol{\beta} \neq 1 \end{aligned}$$

para la cual operan los mismos criterios establecidos anteriormente en el caso de pruebas simples.

3.8.3. Prueba F General.

La característica principal de esta prueba consiste en que permite el contraste de varias hipótesis en forma simultánea. Es posible por ejemplo una hipótesis de la forma

$$\begin{aligned} H_0 &: \beta_1 = -1 \\ &\beta_2 = 1 \\ &2\beta_1 + 3\beta_2 = 1 \end{aligned}$$

En este caso la hipótesis alternativa, es el **NO** cumplimiento de la hipótesis nula. Es decir, bastaría que no se cumpla cualquiera de las hipótesis propuestas, para que la hipótesis sea falsa.

Existen dos alternativas para realizar esta prueba.

Alternativa 1.

La primera es vía matricial, en que se consideran dos matrices $\mathbf{R}_{q \times k}$ y $\mathbf{r}_{q \times 1}$, donde q se refiere a la cantidad de restricciones (conocidas) que se le están imponiendo al modelo. En este caso, la hipótesis nula se plantea de la siguiente forma:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

Si consideramos el ejemplo anterior, suponiendo que existen sólo dos parámetros, las matrices relevantes serían:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Existen algunas condiciones que se deben mantener al realizar esta prueba. Estas son:

1. El número de restricciones (q) debe ser inferior o igual al número de parámetros (en nuestro ejemplo, no se cumple esta condición y la tercera restricción se puede escribir como una combinación lineal de las dos primeras). En otras palabras, el rango de la matriz \mathbf{R} debe ser menor o igual al número de restricciones lineales.
2. Cualquier restricción será válida, y se podrá representar en forma matricial, siempre y cuando ésta sea lineal. En otros casos no existen las matrices \mathbf{r} y \mathbf{R} .

Para encontrar una distribución útil para esta prueba de hipótesis, recuerde que el vector de estimadores presenta la siguiente distribución:

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma_{\mu}^2 (\mathbf{X}'\mathbf{X})^{-1}\right)$$

Y además, utilizaremos el hecho que

$$\begin{aligned} \mathbf{E}(\mathbf{A}\mathbf{X}) &= \mathbf{A} \times \mathbf{E}(\mathbf{X}) \\ \mathbf{Var}(\mathbf{A}\mathbf{X}) &= \mathbf{A}\mathbf{Var}(\mathbf{X})\mathbf{A}' \end{aligned}$$

donde \mathbf{A} es constante (una matriz de escalares constantes). Entonces, al multiplicar la matriz de parámetros estimados de $\hat{\boldsymbol{\beta}}$, por la matriz de constantes \mathbf{R} , obtenemos la siguiente distribución de probabilidad:

$$\mathbf{R}\hat{\boldsymbol{\beta}} \sim N\left(\mathbf{R}\boldsymbol{\beta}, \sigma_{\mu}^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)$$

Al igual que en el caso de una sola variable aleatoria, cada una de las variables incluidas en el vector $\hat{\boldsymbol{\beta}}$ se pueden estandarizar. Sin embargo, en

este caso particular se debe considerar la distribución de una forma cuadrática en la cual puede existir correlación entre los elementos del vector $\hat{\beta}$ ¹⁸.

Así el vector

$$\left(\mathbf{R}\hat{\beta} - \mathbf{R}\beta\right) \sim N\left(\mathbf{0}, \sigma_{\mu}^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)$$

y la forma cuadrática de la expresión anterior puede transformarse en una variable con distribución Chi-Cuadrado tal que:

$$\frac{\left(\mathbf{R}\hat{\beta} - \mathbf{R}\beta\right)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{R}\beta\right)}{\sigma_{\mu}^2} \sim \chi_{(q)}^2 \quad (3.42)$$

Bajo la hipótesis nula

$$\frac{\left(\mathbf{R}\hat{\beta} - \mathbf{r}\right)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{r}\right)}{\sigma_{\mu}^2} \sim \chi_{(q)}^2 \quad (3.43)$$

Donde q es el número de restricciones lineales impuestas en $\mathbf{R}\beta = \mathbf{r}$.

En este momento nos volvemos a enfrentar al problema que σ_{μ}^2 no es conocida. Por lo tanto utilizamos el mismo procedimiento anterior, dividir la ecuación ?? por:

$$\frac{\hat{\sigma}_{\mu}^2 (n - k)}{\sigma_{\mu}^2} \sim \chi_{(n-k)}^2,$$

lo cual ajustado por los respectivos grados de libertad de cada variable Chi-Cuadrado se distribuye $F(v_1, v_2)$ ¹⁹.

En este caso la variable F queda expresada de la siguiente manera:

$$F = \frac{\left(\mathbf{R}\hat{\beta} - \mathbf{R}\beta\right)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{R}\beta\right)}{\frac{\sigma_{\mu}^2 * q}{\frac{\hat{\sigma}_{\mu}^2 (n - k)}{\sigma_{\mu}^2} / (n - k)}} \sim F_{(q, n-k)}$$

$$F = \frac{1}{q} \frac{\left(\mathbf{R}\hat{\beta} - \mathbf{R}\beta\right)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{R}\beta\right)}{\hat{\sigma}_{\mu}^2} \sim F_{(q, n-k)}$$

¹⁸Para una demostración ver Johnston (1992).

¹⁹Ver sección 2.3.4.

Como conocemos el valor de $\hat{\sigma}_\mu^2$ lo podemos reemplazar, quedando la siguiente expresión:

$$F = \frac{1}{q} \frac{\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}\right)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}\right)}{\frac{\mathbf{e}'\mathbf{e}}{n-k}} \sim F_{(q,n-k)}$$

$$F = \frac{n-k}{q} \frac{\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}\right)' \left(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}\right)}{\mathbf{e}'\mathbf{e}} \sim F_{(q,n-k)}$$

En este caso operan los mismos criterios que para las pruebas anteriores. Es decir, si el F calculado es mayor que el F de tabla se rechaza la hipótesis nula.

Alternativa 2.

Un método alternativo y más simple para realizar las pruebas de significancia global con distribución F es mediante la utilización de las Sumas Residuales del Modelo.

Estas sumas residuales se calculan para el caso de un modelo restringido a la hipótesis que se desea probar y un modelo no restringido a la hipótesis nula. De lo que se trata es de comparar si estas sumas residuales son o no estadísticamente diferentes. Si bajo la hipótesis nula las sumatorias no son distintas de las obtenidas en el modelo no restringido, entonces podemos aceptar la hipótesis nula.

Para clarificar el concepto, tomemos el modelo lineal general en desviaciones de medias y por fines de simplicidad asumamos que existe sólo una variable explicativa

$$y_i = \hat{\beta}x_i + \varepsilon_i$$

Elevando esta expresión al cuadrado y aplicando sumatoria se tiene:

$$\sum y_i^2 = \hat{\beta}^2 \sum x_i^2 + \sum \varepsilon_i^2 \quad (3.44)$$

que es la sumatoria total de cuadrados (STC) dividida en una sumatoria explicada (SEC) y una residual (SRC).

$$STC = SEC + SRC$$

Ahora bien, sabemos que $\hat{\beta} \sim \mathbf{N}(\beta, \sigma_\mu^2(\mathbf{X}'\mathbf{X})^{-1})$ y por lo tanto

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma_\mu^2(\mathbf{X}'\mathbf{X})^{-1}}} \sim \mathbf{N}(\mathbf{0}, \mathbf{1})$$

que para nuestro caso simple es

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma_\mu^2}{\sum x_i^2}}} = \frac{(\hat{\beta} - \beta)\sqrt{\sum x_i^2}}{\sigma_\mu} \sim \mathbf{N}(0,1)$$

Si este resultado lo elevamos al cuadrado, tendremos una distribución χ^2 con un grado de libertad.

$$\frac{(\hat{\beta} - \beta)^2 \sum x_i^2}{\sigma_\mu^2} \sim \chi^2(1)$$

Asumamos la siguiente hipótesis nula

$$H_0 : \beta = 0$$

Bajo esta hipótesis tenemos dos resultados importantes:

1. $\frac{(\hat{\beta})^2 \sum x_i^2}{\sigma_\mu^2} \sim \chi^2(1)$ y
2. $\sum \mathbf{y}^2 = \sum \boldsymbol{\varepsilon}^2$, es decir la *STC* es igual a la *SRC*. Llamaremos *Sumatoria Residual Restringida de Cuadrados* (SRRC) a la *STC* bajo la hipótesis nula H_0 .

Como es posible verificar de 3.44 y del resultado 1.

$$\frac{\hat{\beta}^2 \sum x_i^2}{\sigma_\mu^2} = \frac{SEC}{\sigma_\mu^2} \sim \chi^2(1)$$

Pero la $SEC = STC - SRC$, y bajo la Hipótesis nula $STC = SRRC$. Reemplazando en la ecuación anterior obtenemos

$$\frac{SRRC - SRC}{\sigma_\mu^2} \sim \chi^2(1)$$

Se puede construir una distribución F dividiendo dos distribuciones chi-cuadrado, divididos por sus respectivos grados de libertad. Por lo tanto,

$$\frac{\frac{SRRC - SRC}{1 \cdot \sigma_\mu^2}}{\frac{\hat{\sigma}_\mu^2}{(n-k) \frac{\sigma_\mu^2}{n-k}}} = \frac{SRRC - SRC}{\hat{\sigma}_\mu^2} \sim \mathbf{F}(1, n - k)$$

también sabemos que $\hat{\sigma}_\mu^2 = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n-k}$, entonces

$$\frac{SRRC - SRC}{\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n-k}} = \frac{(n-k)(SRRC - SRC)}{1 \cdot SRC} \sim \mathbf{F}(1, n - k)$$

En términos generales, la hipótesis se puede plantear para el caso de más coeficientes cambiando el número de restricciones lineales impuestas al modelo

$$\frac{1}{q} \cdot \frac{SRRC - SRC}{\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n-k}} = \frac{(n-k)(SRRC - SRC)}{q \cdot SRC} \sim \mathbf{F}(q, n - k)$$

Con este resultado es posible probar muchas hipótesis alternativas. Para cualquier hipótesis el procedimiento se realiza en tres etapas:

1. En una primera etapa se estima el modelo sin ninguna restricción.

$$\mathbf{Y} = \hat{\boldsymbol{\beta}}\mathbf{X} + \boldsymbol{\varepsilon}$$

De aquí se obtiene la sumatoria de cuadrados no restringida

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sum \varepsilon_i^2 = SRC$$

2. En una segunda etapa se estima el modelo sujeto a la hipótesis nula

$$H_0 : \boldsymbol{\beta} = \bar{\boldsymbol{\beta}} \quad \mathbf{Y} = \bar{\boldsymbol{\beta}}\mathbf{X} + \bar{\boldsymbol{\varepsilon}}$$

y se calcula la sumatoria de residuos al cuadrado restringida.

$$\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}} = \sum \bar{\varepsilon}_i^2 = SRRC.$$

3. En una tercera etapa se calcula el valor F y compara con el valor de tabla.

$$\frac{(n-k)(SRRC - SRC)}{q \quad SRC} \sim \mathbf{F}(q, n-k)$$

Regularmente, en los resultados econométricos que se incorporan en los programas estándar, se presenta un valor F calculado que esta sujeto a la hipótesis nula que todos los coeficientes (excepto el intercepto) son iguales a cero, es decir

$$H_0 : \beta = \mathbf{0}$$

Esto es lo que se conoce como una hipótesis de significancia global. Es decir, lo que se quiere probar es si el modelo tiene algún poder “explicativo”. Como el número de parámetros es k , el número de grados de libertad de la prueba es $k-1$, ya que se descuenta el parámetro de posición. La forma general en que se calcula esta prueba es

$$F = \frac{SRRC - SRC}{SRC} \left(\frac{n-k}{k-1} \right) \sim F_{(k-1, n-k)}$$

Esta expresión también puede ser calculada usando el estimador del R^2 . Si recordamos que $R^2 = \frac{STC - SRC}{STC}$, entonces bajo la hipótesis nula $SRRC = STC$, por lo tanto

$$F = \frac{STC - SRC}{SRC} \left(\frac{n-k}{k-1} \right) \sim F_{(k-1, n-k)}$$

y recordando que $STC - SRC = SEC$ se tiene

$$F = \frac{SEC}{STC - SEC} \left(\frac{n-k}{k-1} \right) \sim F_{(k-1, n-k)}$$

por último se divide numerador y denominador por STC

$$F = \frac{\frac{SEC}{STC}}{1 - \frac{SEC}{STC}} \left(\frac{n-k}{k-1} \right) \sim F_{(k-1, n-k)}$$

se obtiene

$$F = \frac{R^2}{1 - R^2} \left(\frac{n-k}{k-1} \right) \sim F_{(k-1, n-k)}$$

Lo cual implica que a través del R^2 se puede evaluar la significancia global del modelo. En otras palabras, se puede verificar si la ecuación de regresión en su conjunto es significativa.

Para ejemplificar como opera esta prueba considere la siguiente función de regresión muestral:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \varepsilon_i$$

y asumamos la siguiente hipótesis:

$$H_0 : \begin{matrix} \beta_1 = 0 \\ \beta_2 = 1 \end{matrix}$$

Usando las etapas anteriores, se debe estimar el modelo en su forma no restringida, es decir:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \varepsilon_i$$

Con este modelo original estimado obtenemos:

$$\sum_i \varepsilon_i^2 = SRC$$

posteriormente calculamos el modelo restringido:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \mu_i$$

$$Y_i = 0 + 1X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \mu_i$$

$$Y_i - X_{2i} = \tilde{\beta}_3 X_{3i} + \tilde{\beta}_4 X_{4i} + \mu_i$$

y a través de M.C.O. obtenemos la suma de residuos cuadrados restringidos

$$\sum \mu_i^2 = SRRC$$

Con estos datos calculamos la variable aleatoria Fischer:

$$F = \frac{(SRRC - SRC)(n - 4)}{SRC} \sim F_{(2, n-4)}$$

y lo comparamos con el valor F de tabla, los mismos criterios anteriores son usados para aceptar o rechazar la hipótesis nula.

3.8.4. Predicción.

En muchas ocasiones el investigador está interesado en pronosticar el comportamiento de las variables estudiadas. Por ejemplo, el Banco Central o la autoridad pertinente debe preguntarse cuánto variará la inversión total si cambia la tasa de interés, o bien cuánto será el consumo agregado si el PIB toma un valor determinado. Para saber este resultado basta, en este último caso, con reemplazar el valor hipotético del PIB dentro de la función de consumo estimada y obtener una predicción del consumo. En este caso, lo que se desea conocer es el valor que tomará la variable dependiente para distintos valores posibles que tome el vector de variables independientes. Para dicho efecto, el vector de parámetros (que constituía la incógnita en la estimación) se asume fijo.

La exactitud de esta predicción va a depender de varias cosas. Entre ellas, del tipo de *predictor* que se requiera y del grado de variabilidad que presenten las variables explicativas. Existen dos tipos de predicción, la *predicción individual* que, como su nombre lo indica, predice un valor *individual* de la variable dependiente y, existe además, la *predicción media*, en la cual se predice el valor esperado de Y . En ambos casos la predicción está sujeta al valor asignado a las variables explicativas y a los valores de los parámetros estimados.

La notación utilizada en este apartado es la siguiente: denominaremos como Y_0 al valor que se quiere predecir y que está asociado al verdadero modelo poblacional y \mathbf{X}_0 al vector fila de valores conocidos de las variables explicativas que usaremos para predecir. El predictor está dado por el valor $\hat{Y}_0 = \mathbf{X}_0\hat{\boldsymbol{\beta}}$.²⁰

Predicción Individual

En la predicción individual se desea conocer el valor de Y para una observación específica, es decir

$$Y_0 = \mathbf{X}_0\boldsymbol{\beta} + \mu_0$$

donde \mathbf{X}_0 refleja un vector de valores en los cuales se desea evaluar la ecuación de regresión y $\boldsymbol{\beta}$ es el vector (cierto) de parámetros poblacionales. El resultado de esa evaluación será un valor Y_0 . Sin embargo no conocemos

²⁰Se asume que éste es un predictor insesgado.

con exactitud ese valor, por lo que usamos el predictor $\hat{Y}_0 = \mathbf{X}_0\hat{\boldsymbol{\beta}}$. Se define entonces el error de predicción como la diferencia entre el valor predicho y el valor que debería tener la variable dependiente si hubiésemos usado la verdadera función poblacional, o sea:

$$ep_0^i = \hat{Y}_0 - Y_0$$

donde el supraíndice i nos indica que corresponde al error de predicción individual. Reemplazando las correspondientes ecuaciones, tenemos:

$$ep_0^i = \mathbf{X}_0 \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right) - \mu_0$$

Este error de predicción tiene los siguientes momentos:

La esperanza es:

$$\begin{aligned} \mathbf{E}(ep_0^i) &= \mathbf{E}[\hat{Y}_0 - Y_0] \\ \mathbf{E}(ep_0^i) &= \mathbf{E}[\mathbf{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mu_0] \\ \mathbf{E}(ep_0^i) &= \mathbf{X}_0\mathbf{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] - \mathbf{E}[\mu_0] \\ \mathbf{E}(ep_0^i) &= 0 \end{aligned}$$

dato que el predictor es insesgado y la esperanza del término de error es cero.

Y la varianza por su parte es:

$$\begin{aligned} \mathbf{Var}(ep_0^i) &= \mathbf{E}[(ep_0^i - E(ep_0^i))(ep_0^i - E(ep_0^i))'] \\ \mathbf{Var}(ep_0^i) &= \mathbf{E}[(ep_0^i)(ep_0^i)'] \end{aligned}$$

$$\mathbf{Var}(ep_0^i) = \mathbf{E} \left[\left(\mathbf{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mu_0 \right) \left(\mathbf{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mu_0 \right)' \right]$$

$$\mathbf{Var}(ep_0^i) = \mathbf{E} \left[\left(\mathbf{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mu_0 \right) \left((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}_0' - \mu_0' \right) \right]$$

$$\mathbf{Var}(ep_0^i) = \mathbf{E} \left[\begin{array}{c} (\mathbf{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mu_0) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}_0' \\ -\mathbf{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \mu_0' - \mu_0 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}_0' + \mu_0 \mu_0' \end{array} \right]$$

$$\begin{aligned} \mathbf{Var}(ep_0^i) &= \mathbf{X}_0 \mathbf{E} \left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right) \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right)' \right] \mathbf{X}_0' - \mathbf{X}_0 \mathbf{E} \left(\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right) \mu_0' \right) \\ &\quad - \mathbf{E} \left(\mu_0 \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right)' \right) \mathbf{X}_0' + \mathbf{E}(\mu_0 \mu_0') \end{aligned}$$

$$\mathbf{Var}(ep_0^i) = \mathbf{X}_0 \sigma_\mu^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0' + \sigma_\mu^2$$

$$\mathbf{Var}(ep_0^i) = \sigma_\mu^2 \left[\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0' + 1 \right]$$

Asumiendo que la variable poblacional y el predictor se distribuyen en forma normal

$$ep_0^i \sim N \left(0, \sigma_\mu^2 \left[\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0' + 1 \right] \right)$$

Y estandarizando

$$\frac{ep_0^i}{\sqrt{\sigma_\mu^2 \left[\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0' + 1 \right]}} \sim N(0, 1)$$

Luego realizamos el procedimiento usual para transformar nuestra distribución normal en una *t* de Student:

$$\frac{ep_0^i}{\sqrt{\hat{\sigma}_\mu^2 \left[\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0' + 1 \right]}} \sim t_{n-k}$$

Con la información obtenida construimos un intervalo de confianza para la predicción:

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 \left[\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0' + 1 \right]} \leq Y_0 \leq \hat{Y}_0 + t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 \left[\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0' + 1 \right]}$$

Ejemplo 1: Predicción Individual del Consumo

Intentaremos predecir el consumo cuando el ingreso alcanza los 8.200.000 millones de pesos²¹. La predicción individual se lleva a cabo a través del siguiente intervalo de confianza²²:

²¹Utilizando los resultados de la estimación de la sección 3.4 y recordando que se expresa en unidad de millón.

²²Los valores han sido aproximados. Usted puede obtener los verdaderos resultados a partir de los datos contenidos en el apéndice A.

$$\hat{C}_0 - t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0 + 1]} \leq C_0 \leq \hat{C}_0 + t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0 + 1]}$$

Se tiene que:

$$\hat{\beta} = \begin{bmatrix} 0,3152 \\ 0,6137 \end{bmatrix} \quad \mathbf{X}_0 = [1 \quad 8200000]$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0,15623 & -0,0371 \\ -0,0371 & 0,0106 \end{bmatrix}$$

$$\hat{C}_0 = \mathbf{X}_0 \hat{\beta} = 5343292 \quad \hat{\sigma}_\mu^2 = 0,013$$

Haciendo los cálculos necesarios y reemplazando en el intervalo de confianza con un t de tabla de 2.021 para un 5% de significancia y un tamaño muestral de 38 observaciones encontramos que:

$$5084310 \leq C_0 \leq 5602274$$

De esta forma, cuando el ingreso alcanza los 8200000 millones de pesos²³, el consumo se hallará entre 5084310 y 5602274 millones de pesos con una probabilidad del 95%.

Ejemplo 2: Predicción Individual de las Importaciones

Predeciremos el nivel de importaciones dado un valor determinado para cada variable independiente²⁴.

Los valores que se asumen que adoptan quedan expresadas en la matriz \mathbf{X}_0 :

$$\hat{\beta} = \begin{bmatrix} 0,47078 \\ 0,99234 \\ -0,01299 \end{bmatrix} \quad \mathbf{X}_0 = [1 \quad \ln 3000000 \quad 90]$$

luego,

²³En moneda constante de 1986.

²⁴Utilizando los valores obtenidos de la estimación de la función de importaciones de la sección 3.7.

$$\hat{M}_0 = \mathbf{X}_0 \hat{\boldsymbol{\beta}} = 14,101$$

Calculando $\hat{\sigma}_\mu^2$ para la función de importaciones como lo muestra la sección 3.4.2. y siguiendo el mismo procedimiento del ejemplo anterior, podemos reemplazar los valores obtenidos en el intervalo de confianza. De esta forma, el intervalo de confianza para la predicción individual queda como sigue:

$$M_0 \in \left(14,101 \mp 2,021 \sqrt{0,0051 \times (1 + 1,5156)} \right)$$

$$(13,8712 \leq \ln(M_0) \leq 14,3290)$$

Dado los valores antes planteados para los estimadores, se puede esperar que las importaciones se ubiquen entre²⁵ 1057269 y 1671112 millones de pesos de 1986.

Predicción Media

En el caso de la predicción media interesa conocer la esperanza de la variable dependiente, dado un nivel de las variables explicativas, es decir $E(Y/\mathbf{X}_0)$. El predictor sigue siendo $\hat{Y} = \mathbf{X}_0 \hat{\boldsymbol{\beta}}$. En este caso, se define el error de predicción de la siguiente manera:

$$\begin{aligned} ep_m &= \hat{Y}_0 - E(Y/\mathbf{X}_0) \\ ep_m &= \mathbf{X}_0 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{aligned}$$

Nuevamente, deseamos encontrar la esperanza y varianza de este error de predicción para construir un intervalo de confianza para la predicción media.

$$\begin{aligned} E(ep_m) &= E[\hat{Y}_0 - E(Y/\mathbf{X}_0)] \\ E(ep_m) &= E[\mathbf{X}_0 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ E(ep_m) &= \mathbf{X}_0 E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ E(ep_m) &= 0 \end{aligned}$$

²⁵Aplicando la exponencial a los valores extremos hallados en el intervalo de confianza obtenemos el rango probable para el nivel de importaciones.

y

$$\begin{aligned}
 \text{Var}(ep_m) &= E[(ep_m - E(ep_m))(ep_m - E(ep_m))'] \\
 \text{Var}(ep_m) &= E[(ep_m)(ep_m)'] \\
 \text{Var}(ep_m) &= E\left[\left(\mathbf{X}_0(\hat{\beta} - \beta)\right)\left(\mathbf{X}_0(\hat{\beta} - \beta)\right)'\right] \\
 \text{Var}(ep_m) &= E\left[\left(\mathbf{X}_0(\hat{\beta} - \beta)\right)\left((\hat{\beta} - \beta)' \mathbf{X}_0'\right)\right] \\
 \text{Var}(ep_m) &= \mathbf{X}_0 E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] \mathbf{X}_0' \\
 \text{Var}(ep_m) &= \sigma_\mu^2 \left[\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0'\right]
 \end{aligned}$$

Asumiendo una distribución normal

$$ep_m \sim N\left(0, \sigma_\mu^2 \left[\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0'\right]\right),$$

o bien

$$\frac{ep_m}{\sqrt{\sigma_\mu^2 \left[\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0'\right]}} \sim N(0, 1)$$

Lo cual se puede transformar en una distribución t de Student:

$$\frac{ep_m}{\sqrt{\hat{\sigma}_\mu^2 \left[\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0'\right]}} \sim t_{n-k}$$

El intervalo de confianza queda como

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 \left[\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0'\right]} \leq E(Y_0 | X_0) \leq \hat{Y}_0 + t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 \left[\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_0'\right]}$$

La diferencia de varianzas, que se observa, entre ambos errores de predicción, se debe a que la predicción individual es más exigente. Es decir, si construimos un intervalo de confianza, digamos al 95 %, en el caso de la predicción individual, éste intervalo debe ser más amplio, para asegurarnos de que el intervalo contenga al valor individual. No así en el caso de la predicción media, ya que como sólo interesa el valor medio, un intervalo más estrecho es suficiente para contener este valor.

Ejemplo 1: Predicción Media del Consumo

Para los mismos valores dados en la predicción individual ahora los desarrollaremos para la predicción media del consumo.

El intervalo de confianza relevante para esta predicción es:

$$\hat{C}_0 - t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0]} \leq C_0 \leq \hat{C}_0 + t_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0]}$$

luego, la predicción media para el consumo con un 95 % de confianza es:

$$5225561 \leq E(C_0/X_0) \leq 5461024$$

Para la predicción media, el consumo se encontrará entre 5225561 y 5461024 millones de pesos de 1986 cuando el ingreso alcanza los 8200000 millones de pesos. Este intervalo es más estrecho que el obtenido previamente con la predicción individual.

Ejemplo 2: Predicción Media de las Importaciones

Suponiendo los mismos valores para los parámetros de la Función de Importaciones en la predicción individual, ahora encontraremos la predicción media. Reemplazando los cálculos en el intervalo de confianza entregado en el ejemplo anterior, encontramos que el intervalo de confianza para las importaciones es:

$$(13,9224 \leq E(\ln M_0/X_0) \leq 14,2778)$$

Así, cuando el PIB alcanza un nivel de 3000000 y el tipo de cambio real del periodo $t-1$ es de 90, el nivel de importaciones podría hallarse²⁶ en promedio entre los 1112811 y 1587705 millones de pesos de 1986. Este intervalo también es más estrecho que en el caso de la predicción individual.

²⁶Aplicando exponenciales al rango anterior.

Capítulo 4

MINIMOS CUADRADOS GENERALIZADOS

4.1. Introducción

En el capítulo anterior revisamos el modelo clásico de regresión. Este modelo se basa en varios supuestos centrales. Entre estos supuestos se encuentran aquellos relativos a la estructura del término de error. Básicamente, se asume que los errores de la regresión poblacional son homocedásticos y que no presentan autocorrelación.

Asumir que los términos de error de las funciones de regresión poblacional cumplen con estos supuestos puede ser poco adecuado. En la práctica, la excepción es encontrarse con términos de error que cumplan con estos supuestos. En este capítulo nos concentraremos en los problemas de estimación que surgen cuando se levantan estos supuestos sobre la matriz de varianzas y covarianzas. Estudiaremos los dos principales problemas que se encuentran en una estimación: *Heterocedasticidad y Autocorrelación*.

El primero se refiere a que las varianzas de los errores poblacionales no poseen la misma varianza. Es decir, los elementos de la diagonal principal de la matriz de covarianzas no son iguales.

El segundo problema se refiere a que los errores poblacionales dejan de ser independientes entre observaciones. Esto queda representado cuando los elementos fuera de la diagonal principal dejan de ser ceros.

El principal obstáculo que presentan estos problemas se refleja en los estimadores y en su condición de estimadores **MELI** (mejores estimadores

lineales insesgados). Cuando no se cumple el supuesto de homocedasticidad, o no autocorrelación, los estimadores dejan de tener la mínima varianza. Esto implica que hacer inferencias a partir de estos estimadores, puede no ser lo más apropiado. Las pruebas de hipótesis pueden dejar de tener sentido. Para corregir el problema se utiliza un método denominado *Mínimos Cuadrados Generalizados* (M.C.G.).

A continuación, primero se presenta un modelo de regresión general que permite que la matriz de varianza y covarianza de los errores no cumpla los supuestos clásicos. Este modelo de regresión lo llamaremos *Mínimos Cuadrados Generalizados*. Luego, investigaremos específicamente los problemas de heterocedasticidad y autocorrelación, en ese orden.

4.2. Mínimos Cuadrados Generalizados

Recordemos la función de regresión poblacional en su forma matricial:

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times k} \boldsymbol{\beta}_{k \times 1} + \boldsymbol{\mu}_{n \times 1}$$

y los supuestos clásicos más relevantes,

$$E(\boldsymbol{\mu}) = \mathbf{0} \quad (4.1)$$

$$\text{rango}(\mathbf{X}) = k < n \quad (4.2)$$

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_{\mu}^2 \mathbf{I}_n \quad (4.3)$$

$$E(\boldsymbol{\mu}\mathbf{X}) = \mathbf{0} \quad (4.4)$$

Por el Teorema de Gauss-Markov sabemos que si se cumplen estos supuestos obtenemos estimadores que son MELI.

En casos prácticos es poco común encontrar situaciones en que la matriz de covarianzas de los errores esté dada por una matriz tal como la presentada en 4.3. Si esto no se cumple, la razón puede estar en dos tipos de problemas.

Uno de estos problemas es conocido como *Heterocedasticidad* y se presenta cuando la varianza de los errores es distinta para cada observación. Cuando los estimadores son MELI, la varianza de Y_i condicional a los valores dados de X_i , permanece constante independientemente de los valores que tome X . Cuando no se cumple este supuesto la varianza condicional de Y_i aumenta o disminuye a medida que X se incrementa.

El problema de heterocedasticidad es típico en datos de corte transversal, donde se trabaja con miembros de una población en un momento determinado, tales como familias o industrias, las cuales pueden ser de diferentes

tamaños: Por ejemplo, las firmas se clasifican como grandes, pequeñas o medianas, y los niveles de ingresos de las familias en altos, bajos o medios. Si tomamos el caso de los ingresos, el problema de la heterocedasticidad queda ejemplificado al estudiar los patrones de consumo de pan de familias pertenecientes a diferentes estratos sociales. Es de esperar que el nivel de consumo de pan dependa del nivel de ingreso de cada familia. No obstante, también puede depender de variables no observables, como son los patrones culturales. Recuerde que estas variables no observables caen en el término de error de la regresión. Si las familias de bajos ingresos tienen patrones culturales distintos a las de altos ingresos, entonces probablemente la dimensión y variabilidad del término de error será distinto para familias de distintos niveles de ingreso, representando un caso típico de heterocedasticidad. En este caso, el supuesto que la varianza del error es igual para todas las observaciones puede ser poco adecuado.

El segundo problema que puede explicar el no cumplimiento del supuesto dado en 4.3 es que los términos de error no sean independientes entre sí. Esto implica que los elementos fuera de la diagonal principal de la matriz de varianzas y covarianzas de los errores serán distintos de cero. Esto se conoce como *Autocorrelación*, o también correlación serial de errores. Cuando los estimadores sí cumplen con el supuesto de no autocorrelación se supone que el término de perturbación perteneciente a una observación no está influenciado por el término de perturbación perteneciente a otra. El problema de autocorrelación es común en series de observaciones ordenadas en el tiempo. Por ejemplo, es lógico pensar que si el consumo de una familia fue “excesivamente” alto en un período, indicando con el término excesivo que está sobre lo que la regresión predice que debería ser en promedio, dado el valor de las variables independientes (ingreso por ejemplo), también lo sea en el siguiente período. En este caso, lo que veremos en la estimación es que los errores de dos observaciones sucesivas tenderán a tener el mismo signo y tamaño. Es decir, estaríamos en presencia de autocorrelación positiva.

De esta forma entonces, si relajamos el supuesto de homocedasticidad y de no autocorrelación, la matriz de covarianzas queda expresada como:

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_{\mu}^2 \boldsymbol{\Psi}_n \quad (2)$$

donde $\boldsymbol{\Psi}$ no es necesariamente igual a la matriz identidad, sino que es una matriz simétrica ordinaria. Por lo tanto la matriz de varianzas y covarianzas

de los errores esta dada por:

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \begin{bmatrix} \text{var}(\mu_1) & \text{cov}(\mu_1, \mu_2) & \cdots & \text{cov}(\mu_1, \mu_n) \\ \text{cov}(\mu_2, \mu_1) & \text{var}(\mu_2) & \cdots & \text{cov}(\mu_2, \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\mu_n, \mu_1) & \text{cov}(\mu_n, \mu_2) & \cdots & \text{var}(\mu_n) \end{bmatrix}$$

Observando esta matriz se puede determinar si existe homocedasticidad o heterocedasticidad, como también si existe o no autocorrelación de errores. Recuerde que homocedasticidad implica que $\text{Var}(\mu_i) = \sigma_\mu^2, \forall i$ y ausencia de autocorrelación que $\text{Cov}(\mu_i, \mu_j) = 0, \forall i \neq j, i, j = 1, \dots, n$.

Resumiendo el dilema econométrico, nos enfrentamos al siguiente problema de estimación:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \quad \text{con} \quad E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_\mu^2 \boldsymbol{\Psi}_n$$

Si recordamos las propiedades que poseía el estimador de mínimos cuadrados ordinarios, sabemos que éste era insesgado y tenía mínima varianza. Si continuamos usando el método de mínimos cuadrados, obtendremos el mismo vector de parámetros, es decir $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$. Sin embargo, no sabemos qué propiedades tienen estos estimadores, dado que no podemos aplicar el teorema de Gauss Markov debido a que no se cumplen los supuestos clásicos. Entonces, ¿ Cuáles son las propiedades de este estimador?

Podemos descomponer el vector de estimadores y aplicar esperanza tal que¹:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MCO} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\mu} \\ E(\hat{\boldsymbol{\beta}}_{MCO}) &= E(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\mu}) \\ E(\hat{\boldsymbol{\beta}}_{MCO}) &= E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\boldsymbol{\mu}) \\ E(\hat{\boldsymbol{\beta}}_{MCO}) &= \boldsymbol{\beta} \end{aligned}$$

Lo cual implica que $\hat{\boldsymbol{\beta}}$ es insesgado. En otras palabras, a pesar de haber relajado el supuesto de homocedasticidad y no autocorrelación el estimador $\hat{\boldsymbol{\beta}}$ de M.C.O. sigue siendo insesgado.

¹Ver sección 3.4.1.

Su varianza es:

$$\begin{aligned}
 Var(\hat{\beta}_{MCO}) &= E\left[\left(\hat{\beta}_{MCO} - \beta\right)\left(\hat{\beta}_{MCO} - \beta\right)'\right] \\
 Var(\hat{\beta}_{MCO}) &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\mu}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\mu}\right]' \\
 Var(\hat{\beta}_{MCO}) &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\boldsymbol{\mu}\boldsymbol{\mu}'\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\
 Var(\hat{\beta}_{MCO}) &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left[\left(\boldsymbol{\mu}\boldsymbol{\mu}'\right)\right]\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\
 Var(\hat{\beta}_{MCO}) &= \sigma_{\mu}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Psi}_n\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}
 \end{aligned}$$

La cual es claramente distinta al caso en que se cumplen todos los supuestos clásicos. En resumen, hemos obtenido un estimador insesgado, pero que no posee mínima varianza. Por lo tanto, no es el mejor estimador insesgado y debemos encontrar uno que lo sea. Una manera de encontrar este estimador insesgado de mínima varianza es transformar $\boldsymbol{\Psi}_n$ de tal forma que cumpla los supuestos clásicos y se pueda recurrir al teorema de Gauss Markov.

Sabemos que la matriz de covarianzas es:

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_{\mu}^2\boldsymbol{\Psi}_n$$

Por lo tanto, debemos encontrar una matriz \mathbf{P} tal que se cumpla la siguiente característica:

$$\mathbf{P}\boldsymbol{\Psi}_n\mathbf{P}' = \mathbf{I}_n,$$

es decir, que premultiplicando esta matriz y postmultiplicando la transpuesta de esta matriz con la matriz de covarianzas, se obtiene una nueva matriz que asegura el cumplimiento de los supuestos clásicos.

Premultiplicando por \mathbf{P}^{-1} :

$$\begin{aligned}
 \mathbf{P}^{-1}\mathbf{P}\boldsymbol{\Psi}_n\mathbf{P}' &= \mathbf{P}^{-1}\mathbf{I}_n \\
 \boldsymbol{\Psi}_n\mathbf{P}' &= \mathbf{P}^{-1}\mathbf{I}_n
 \end{aligned}$$

Postmultiplicando por $(\mathbf{P}')^{-1}$:

$$\begin{aligned}
\Psi_n \mathbf{P}' (\mathbf{P}')^{-1} &= \mathbf{P}^{-1} (\mathbf{P}')^{-1} \\
\Psi_n &= \mathbf{P}^{-1} (\mathbf{P}')^{-1} \\
\Psi_n &= \mathbf{P}^{-1} (\mathbf{P}^{-1})' \\
\Psi_n &= (\mathbf{P}'\mathbf{P})^{-1} \\
\Psi_n^{-1} &= \mathbf{P}'\mathbf{P}
\end{aligned}$$

Este resultado será útil para obtener estimadores de mínima varianza.

En resumen, lo que se necesita para obtener estimadores eficientes a través de mínimos cuadrados ordinarios es que se cumpla $E(\boldsymbol{\mu}) = \mathbf{0}$ y $E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_\mu^2 \mathbf{I}_n$. Sin embargo lo que tenemos ahora ya no es lo mismo, sino que tenemos una matriz de covarianzas no escalar, que causa estimadores insesgados, pero no eficientes. Para resolver el problema aplicamos el método de **Mínimos Cuadrados Generalizados** que consiste en encontrar una matriz $\mathbf{P}_{n \times n}$ tal que:

$$\mathbf{P}\Psi\mathbf{P}' = \mathbf{I}_n \quad \Longleftrightarrow \quad \mathbf{P}'\mathbf{P} = \Psi_n^{-1}$$

Si suponemos que hemos encontrado la matriz \mathbf{P} , la utilizamos para transformar los datos originales. Para ello premultiplicamos la función de Regresión Muestral por esta matriz:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \\
\mathbf{P}_{n \times n} \mathbf{Y}_{n \times 1} &= \mathbf{P}_{n \times n} \mathbf{X}_{n \times k} \boldsymbol{\beta}_{k \times 1} + \mathbf{P}_{n \times n} \boldsymbol{\mu}_{n \times 1} \\
\mathbf{Y}^* &= \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\mu}^*
\end{aligned}$$

obteniendo nuevas variables \mathbf{Y}^* , \mathbf{X}^* y $\boldsymbol{\mu}^{*2}$, cuyo único cambio ha sido el amplificarse por constantes. Analizemos qué sucede con los supuestos clásicos con esta nueva regresión. Primero observemos qué ocurre con la esperanza de $\boldsymbol{\mu}^*$:

$$E(\boldsymbol{\mu}^*) = E(\mathbf{P}\boldsymbol{\mu}) = \mathbf{P}E(\boldsymbol{\mu}) = \mathbf{0}$$

con la varianza se tiene:

$$\begin{aligned}
E(\boldsymbol{\mu}^* (\boldsymbol{\mu}^*)') &= E[(\mathbf{P}\boldsymbol{\mu}) (\mathbf{P}\boldsymbol{\mu})'] = E[\mathbf{P}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{P}'] \\
E(\boldsymbol{\mu}^* (\boldsymbol{\mu}^*)') &= \mathbf{P}E[\boldsymbol{\mu}\boldsymbol{\mu}']\mathbf{P}' = \mathbf{P}\sigma_\mu^2 \Psi_n \mathbf{P}' \\
E(\boldsymbol{\mu}^* (\boldsymbol{\mu}^*)') &= \sigma_\mu^2 \mathbf{P}\Psi_n \mathbf{P}' \\
E(\boldsymbol{\mu}^* (\boldsymbol{\mu}^*)') &= \sigma_\mu^2 \mathbf{I}_n
\end{aligned}$$

²Note que las nuevas variables individuales son combinaciones lineales de **todas** las variables individuales originales.

con lo cual se cumplen los supuestos clásicos.

Así podemos aplicar el procedimiento de Mínimos Cuadrados a la regresión $\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\mu}^*$, para obtener estimadores insesgados y eficientes. Es decir:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{MCG} &= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^* \\ \hat{\boldsymbol{\beta}}_{MCG} &= ((\mathbf{PX})'(\mathbf{PX}))^{-1}(\mathbf{PX})'(\mathbf{PY}) \\ \hat{\boldsymbol{\beta}}_{MCG} &= (\mathbf{X}'\mathbf{P}'\mathbf{PX})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{PY} \\ \hat{\boldsymbol{\beta}}_{MCG} &= (\mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{Y}\end{aligned}$$

Estos estimadores tienen las mismas propiedades que los estimadores obtenidos por Mínimos Cuadrados Ordinarios, dado que los nuevos términos de errores, $\boldsymbol{\mu}^*$, cumplen con todos los supuestos del modelo clásico. Es decir, estos estimadores son los mejores estimadores lineales insesgados.

Recordemos que $\hat{\sigma}_\mu^2$, en Mínimos Cuadrados Ordinarios estaba determinado como:

$$\hat{\sigma}_\mu^2 = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n - k} \quad (4.5)$$

En el caso de MCG $\hat{\sigma}_\mu^2$ se obtiene de manera similar:

$$\hat{\sigma}_\mu^2 = \frac{(\boldsymbol{\varepsilon}^*)'(\boldsymbol{\varepsilon}^*)}{n - k}$$

donde $\boldsymbol{\varepsilon}^*$ se obtiene a partir de la función de regresión muestral.

$$\begin{aligned}\mathbf{Y}^* &= \mathbf{X}^*\hat{\boldsymbol{\beta}}_{MCG} + \boldsymbol{\varepsilon}^* \\ \boldsymbol{\varepsilon}^* &= \mathbf{Y}^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}_{MCG}\end{aligned}$$

podemos reemplazar esta última ecuación en 4.5, quedando la expresión co-

mo:

$$\begin{aligned}
\hat{\sigma}_\mu^2 &= \frac{(\mathbf{Y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}_{MCG})' (\mathbf{Y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}_{MCG})}{n - k} \\
\hat{\sigma}_\mu^2 &= \frac{(\mathbf{P}\mathbf{Y} - \mathbf{P}\mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})' (\mathbf{P}\mathbf{Y} - \mathbf{P}\mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})}{n - k} \\
\hat{\sigma}_\mu^2 &= \frac{[\mathbf{P} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})]' [\mathbf{P} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})]}{n - k} \\
\hat{\sigma}_\mu^2 &= \frac{(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})' \mathbf{P}' \mathbf{P} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})}{n - k} \\
\hat{\sigma}_\mu^2 &= \frac{(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})' \boldsymbol{\Psi}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MCG})}{n - k}
\end{aligned}$$

que es nuestro nuevo estimador de la varianza del error.

4.3. Estimador Máximo Verosímil

Si añadimos ahora, el supuesto que el vector de errores $\boldsymbol{\mu}$ se distribuye normal, manteniendo los supuestos que $E(\boldsymbol{\mu}) = \mathbf{0}$ y $E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_\mu^2 \boldsymbol{\Psi}_n$, donde, $\boldsymbol{\Psi}_n$ es una matriz simétrica, conocida de orden $n \times n$, podemos obtener el estimador máximo verosímil. Para el caso del modelo lineal general, teníamos que la función de verosimilitud estaba dada por:

$$\begin{aligned}
f(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{- (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{2\sigma^2} \right] \\
f(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} |\mathbf{I}_n|^{-\frac{1}{2}} \exp \left[\frac{- (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{I}_n^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{2\sigma^2} \right]
\end{aligned}$$

Sin embargo, dado el levantamiento de los supuestos clásicos, la función de verosimilitud cambia. La función de densidad multivariada del caso general

de covarianzas puede ser expresado analíticamente como:

$$f(\mathbf{Y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} |\boldsymbol{\Psi}_n|^{-\frac{1}{2}} \exp \left[\frac{-\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \boldsymbol{\Psi}_n^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{2\sigma^2} \right]$$

Recordando los principios del estimador de máximo verosimilitud introducidos en los capítulos anteriores, podemos escribir la función del logaritmo Máximo Verosimilitud como:

$$\ln L(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln |\boldsymbol{\Psi}_n| - \frac{\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \boldsymbol{\Psi}_n^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{2\sigma^2}$$

Luego, como lo que buscamos es encontrar los valores de $\boldsymbol{\beta}$ y σ^2 que maximizen la función de Verosimilitud, se deriva con respecto a dichos parámetros y luego se iguala la derivada a cero. Esto es:

$$\begin{aligned} \frac{\partial \ln L(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y})}{\partial \boldsymbol{\beta}} &= -\frac{1}{\hat{\sigma}^2} \left[-\mathbf{X}' \boldsymbol{\Psi}_n^{-1} \mathbf{Y} + (\mathbf{X}' \boldsymbol{\Psi}_n^{-1} \mathbf{X}) \hat{\boldsymbol{\beta}} \right] = 0 \\ \frac{\partial \ln L(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y})}{\partial \sigma^2} &= \frac{-n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \boldsymbol{\Psi}_n^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = 0 \end{aligned}$$

Simplificando las expresiones anteriores, lo que vamos a tener es:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MCG} &= (\mathbf{X}' \boldsymbol{\Psi}_n^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Psi}_n^{-1} \mathbf{Y} \\ \hat{\sigma}^2 &= \frac{\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \boldsymbol{\Psi}_n^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{n} \end{aligned}$$

Podemos observar que el estimador obtenido para el parámetro $\boldsymbol{\beta}$ es el mismo que se obtiene al estimar los parámetros por Mínimos Cuadrados Generalizados; es decir, se obtiene el mejor estimador lineal insesgado. No es el caso para el estimador de σ^2 , donde el estimador insesgado que se obtenía por Mínimos Cuadrados Generalizados, era:

$$\hat{\sigma}^2 = \frac{\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \boldsymbol{\Psi}_n^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{n - k}$$

Es decir, la diferencia entre ambos está determinada:

$$\hat{\sigma}_{MV}^2 = \frac{n - k}{n} \hat{\sigma}_{MCG}^2$$

En la medida que aumenta el tamaño muestral, disminuye el sesgo del estimador Máximo Verosímil, acercándose cada vez más al estimador MCG.

En los párrafos precedentes hemos discutido el caso general en que la matriz de varianzas y covarianzas de los errores no cumplen con los supuestos clásicos. A continuación analizaremos en forma separada cada uno de los problemas planteados, es decir, heterocedasticidad y autocorrelación. La discusión se centrará en la identificación y corrección de cada uno de estos problemas.

4.4. Heterocedasticidad

Como definimos anteriormente, la heterocedasticidad ocurre cuando la matriz de covarianza deja de tener una estructura escalar, es decir, cuando los elementos de la diagonal principal no son todos iguales. La notación de esto sería:

$$Var(\mu_i) = \sigma_i^2 \quad i = 1, \dots, n$$

En forma explícita, la matriz sería:

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \begin{bmatrix} \overbrace{Var(\mu_1)}^{\sigma_1^2} & Cov(\mu_1, \mu_2) & \cdots & Cov(\mu_1, \mu_n) \\ Cov(\mu_2, \mu_1) & \overbrace{Var(\mu_2)}^{\sigma_2^2} & \cdots & Cov(\mu_2, \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\mu_n, \mu_1) & Cov(\mu_n, \mu_2) & \cdots & \overbrace{Var(\mu_n)}^{\sigma_n^2} \end{bmatrix}$$

Para empezar vamos a mantener el supuesto que la covarianza entre los errores poblacionales es cero, es decir, $Cov(\mu_i, \mu_j) = 0$. En otras palabras, vamos a suponer que no existe autocorrelación de errores. Entonces, la matriz de varianzas covarianzas queda de esta forma:

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} = \sigma_\mu^2 \boldsymbol{\Psi}_n$$

La matriz \mathbf{P} que debemos encontrar para poder estimar por Mínimos Cuadrados Generalizados es:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix}$$

por lo tanto:

$$\mathbf{P}' \times \mathbf{P} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix}$$

$$\mathbf{P}' \times \mathbf{P} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix}$$

Recordemos que los estimadores obtenidos por Mínimos cuadrados Ordinarios en presencia de heterocedasticidad y/o autocorrelación no son **MELI**. Es decir, los estimadores son insesgados pero no poseen mínima varianza.

4.4.1. Detección de la Heterocedasticidad.

Cuando uno se enfrenta a una estimación no sabe si la regresión tiene o no heterocedasticidad. Por ello, lo primero que se debe hacer es probar si ésta existe o no.

Existen diversos métodos para probar la presencia de heterocedasticidad. Nosotros discutiremos sólo tres de ellos; la prueba de Goldfeld - Quandt, la

prueba de Breusch-Pagan y la prueba de White. Todos estos métodos parten de la base que primero se realizó una estimación con Mínimos Cuadrados Ordinarios. Las pruebas se aplican a los resultados obtenidos de esta estimación.

Prueba de Goldfeld - Quandt

Una de las pruebas más utilizada es la prueba de Goldfeld - Quandt (**GQ**). Para realizar esta prueba se divide la muestra en dos grupos ordenados según el valor de la variable explicativa. Es decir, un grupo para los valores altos de la variable y el otro con los valores bajos. Se ajusta la regresión original a los dos grupos por separado. Luego se ejecuta la prueba F de igualdad de varianzas, para probar si la varianza que se obtiene de las regresiones con una de las submuestras es significativamente distinta de la varianza de la regresión que se obtiene con la otra submuestra. En la medida que se cumpla la hipótesis nula, se acepta la hipótesis de homocedasticidad, es decir que la varianza entre las dos submuestras no difiere.

A continuación, analizaremos más detalladamente el desarrollo de la prueba de **GQ**. Supongamos que tenemos la siguiente función de regresión:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

Supongamos además, que la varianza de los errores está correlacionada con la variable exógena X_i en la forma:

$$\sigma_i^2 = \sigma^2 X_i^2$$

Vamos a dividir la muestra en dos grupos clasificados de acuerdo a la variable X_i , una muestra para valores pequeños de X_i y otra muestra para valores grandes.

Posteriormente ajustamos el modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ para cada una de las muestras por separado. Si se cumplen los supuestos clásicos se espera que la varianza de los errores de estas dos muestras no varíen significativamente. Por lo tanto, para probar homocedasticidad se compara el estimador de la varianza de los errores de la primera muestra (denotado por $\hat{\sigma}_1^2$) con el estimador de la varianza de la segunda muestra ($\hat{\sigma}_2^2$). La hipótesis nula es $H_0 : \hat{\sigma}_1^2 = \hat{\sigma}_2^2$.

Sin embargo debemos encontrar la distribución asociada a esta prueba de hipótesis. Para ello, recordemos que uno de los supuestos del modelo clásico es que el error tiene distribución normal, es decir:

$$\mu_i \sim N(0, \sigma^2)$$

estandarizando esta distribución se tiene

$$\frac{\mu_i}{\sigma} \sim N(0, 1)$$

Elevando al cuadrado y aplicando sumatoria obtenemos una distribución Chi cuadrado con n grados de libertad, es decir:

$$\frac{\sum \mu_i^2}{\sigma^2} \sim \chi^2(n)$$

note que

$$\sum \mu_i^2 \approx \sum \varepsilon_i^2$$

Además, el estimador de σ^2 se definió de la siguiente manera:

$$\hat{\sigma}^2 = \frac{\varepsilon' \varepsilon}{n - k} = \frac{\sum \varepsilon_i^2}{n - k}$$

Reemplazando en la variable Chi Cuadrado, obtenemos una nueva variable:

$$\frac{\hat{\sigma}^2 (n - k)}{\sigma^2} \sim \chi^2(n - k) \quad (4.6)$$

Con esta distribución es posible usar una distribución F para verificar la hipótesis de igualdad de varianzas de las dos regresiones.

Las etapas de la prueba son las siguientes:

1. Ordenar las observaciones de acuerdo a los valores de X_i que se sospecha generan la heterocedasticidad, comenzando por el valor más bajo.
2. Se omiten un número α de observaciones centrales, donde α se especifica a priori. El resto de observaciones se divide en los dos grupos restantes (según el orden establecido).
3. Se estiman dos regresiones usando Mínimos Cuadrados Ordinarios. De esta manera se obtienen las sumas residuales de ambos grupos (SCR1 y SCR2) las cuales permiten estimar $\hat{\sigma}_\mu^2$ de cada submuestra.

De la ecuación 4.6 sabemos que

$$\frac{\hat{\sigma}_1^2 \left(\frac{n}{2} - \frac{\alpha}{2} - k \right)}{\sigma_1^2} \sim \chi^2 \left(\frac{n}{2} - \frac{\alpha}{2} - k \right) \quad \text{para la muestra 1} \quad (4.7)$$

mientras que

$$\frac{\hat{\sigma}_2^2 \left(\frac{n}{2} - \frac{\alpha}{2} - k \right)}{\sigma_2^2} \sim \chi^2 \left(\frac{n}{2} - \frac{\alpha}{2} - k \right) \quad \text{para la muestra 2} \quad (4.8)$$

4. Se verifica la siguiente prueba de hipótesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_\mu^2$$

para ello se construye un estadístico F usando las ecuaciones 4.7 y 4.8. Expresándola en una forma más general, se tiene

$$\frac{\frac{\hat{\sigma}_1^2}{\sigma_1^2} \left(\frac{n}{2} - \frac{\alpha}{2} - k \right)}{\frac{\hat{\sigma}_2^2}{\sigma_2^2} \left(\frac{n}{2} - \frac{\alpha}{2} - k \right)} \sim F \left(\frac{n-\alpha}{2} - k, \frac{n-\alpha}{2} - k \right)$$

Bajo la hipótesis nula de homocedasticidad se tiene :

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F \left(\frac{n-\alpha}{2} - k, \frac{n-\alpha}{2} - k \right)$$

Este valor se compara con el F de tabla y si se acepta H_0 decimos que hay homocedasticidad, mientras que si se rechaza la hipótesis nula se sospecha de presencia de heterocedasticidad. Debe tenerse presente que en el numerador de la expresión anterior debe colocarse la SCR del grupo con mayor varianza.

Puede suceder en algunos casos que, luego de eliminar α observaciones centrales, se obtengan dos grupos de diferente tamaño, ocasión en la cual resulta más conveniente expresar lo anterior como

$$\frac{\hat{\sigma}_i^2}{\hat{\sigma}_j^2} \sim F(n_i - k, n_j - k)$$

donde i y j representan a cada grupo, siendo $\hat{\sigma}_i^2 > \hat{\sigma}_j^2$.

Una de las limitaciones de la prueba de **GQ** es que asume que la heterocedasticidad es “generada” por una sola variable explicativa.

Prueba de Breusch - Pagan

En algunos casos es interesante plantear la hipótesis de que la varianza de los errores es una función de más de una variable explicativa. Esta hipótesis no es posible probarla con la Prueba de **GQ**. En la prueba de Breusch-Pagan (**BP**) se asume que la varianza es una función (no necesariamente lineal) de un conjunto de variables:

$$\sigma_i^2 = h(\mathbf{Z}'_i, \boldsymbol{\alpha})$$

donde \mathbf{Z}'_i es un vector de variables y $\boldsymbol{\alpha}$ un vector de parámetros. Un caso especial de esta función sería:

$$\sigma_i^2 = \alpha_0 + \alpha_1 \times Z_{i1} + \alpha_2 \times Z_{i2} + \dots + \alpha_p \times Z_{ip}$$

Esta prueba también puede dividirse en etapas. Éstas son:

1. Estimar el modelo mediante Mínimos Cuadrados Ordinarios y obtener los residuos ε_i^2 y un estimador para la varianza de los errores $\hat{\sigma}_\mu^2 = \frac{\varepsilon/\varepsilon}{\mathbf{n}}$, que corresponde al estimador máximo verosímil.

2. Construir una variable

$$p_i = \frac{\varepsilon_i^2}{\hat{\sigma}_\mu^2}$$

3. Se estima una regresión de p sobre una constante y un conjunto de variables Z'_i .

$$p_i = \alpha_0 + \alpha_1 Z_{i2} + \alpha_3 Z_{i3} + \dots + \alpha_m Z_{im} + v_i \quad (4.9)$$

obteniendo la Suma Explicada de Cuadrados. Recordemos que $STC = SEC + SRC$, donde STC es la suma total de cuadrados, SEC es la suma explicada de cuadrados y SRC es la suma residual de cuadrados.

4. Bajo la hipótesis nula de homocedasticidad y distribución normal del error, el cociente :

$$\frac{SEC}{2}$$

tiene asintóticamente una distribución χ^2 con $m - 1$ grados de libertad, donde m es el número de parámetros de la ecuación 4.9 y SEC es la suma obtenida en el paso 3.

En la medida que los residuos fuesen homocedásticos, las variables Z 's no tendrían poder explicativo y el valor de la variable calculada en el punto 4, debería ser pequeño tendiendo a aceptar la hipótesis nula. Por el contrario, si el valor es mayor que el valor de la distribución χ^2 al correspondiente nivel de significancia, entonces se rechaza la hipótesis de homocedasticidad.

La matriz de variables \mathbf{Z}' , debe contener pocas variables que no se encuentren ya incluidas como variables explicativas en el modelo original. Los cuadrados de las variables explicativas son candidatos a ser considerados dentro de esta regresión.

Una limitación compartida por las dos pruebas de heterocedasticidad revisadas es que ambas asumen que el investigador conoce la forma funcional que toma la heterocedasticidad.

Prueba de White

Esta prueba de heterocedasticidad, a diferencia de las anteriores, es una prueba en la cual no se precisa la forma particular que adopta la heterocedasticidad. Las etapas para la detección son las siguientes:

1. Estimar el modelo por Mínimos Cuadrados Ordinarios ignorando la posible heterocedasticidad y obtener los residuos ε_i^2 .
2. Estimar una regresión del cuadrado de los residuos obtenidos en la estimación del modelo original, sobre una constante, las variables explicativas del modelo original, sus cuadrados y sus productos cruzados de segundo orden.
3. Construimos la siguiente variable aleatoria:

$$n \times R^{2*} \sim \chi^2(k - 1)$$

Donde R^2 es el coeficiente de determinación de la regresión estimada en el paso 2, k es el número de parámetros y n es el número de observaciones.

Considerando la hipótesis nula de homocedasticidad, esta variable se distribuye asintóticamente como χ^2 con $k - 1$ grados de libertad (se excluye la constante). Por tanto, se requiere comparar el valor obtenido con el valor crítico relevante al nivel de significancia deseado de la tabla de la distribución χ^2 .

Si aumenta el tamaño muestral, el coeficiente de determinación deberá tender a cero bajo el supuesto de homocedasticidad (hipótesis nula). Sólo si la varianza del error depende de las variables explicativas entonces el R^{2*} no tenderá a ser cero y la prueba tenderá a rechazar la hipótesis nula.

Ejemplo 3. Función de Salarios (Mincer)

Como se dijo anteriormente, el problema de heterocedasticidad se presenta a menudo en muestras de corte transversal y es por este motivo que se ha estimado una ecuación de salarios conocida en la literatura como Ecuación de Mincer. Esta ecuación sirve para determinar el nivel de ingresos de un individuo dadas sus características de capital humano, es decir, dados sus años de escolaridad y experiencia laboral. En la base de datos *mincer.xls*³ se presentan los datos utilizados para esta estimación. Los resultados de la ecuación de MCO es:

$$\begin{aligned} \ln y_h &= 3,9997 + 0,1450S + 0,0372X - 0,0004X^2 + \varepsilon \\ d.s. &: 0,1010 \quad 0,0065 \quad 0,0059 \quad 0,0001 \\ t &: 39,5894 \quad 22,3082 \quad 6,3037 \quad -3,1412 \\ R^2 &= 0,2920 \\ \overline{R^2} &= 0,2904 \\ n &= 1268 \end{aligned}$$

Donde y_h representa el ingreso por hora, S representa la escolaridad, X es una variable que mide la experiencia laboral y X^2 es la experiencia laboral al cuadrado, que permite captar el efecto de la depreciación del capital humano. Los resultados indican que los parámetros son significativos individualmente y si es llevada a cabo una Prueba de Significancia Global se verificará la significancia de la regresión en su conjunto. Ahora, procederemos a examinar la existencia de heterocedasticidad a través del Test de Goldfeld Quandt.

Supondremos que la variable que probablemente genera la heterocedasticidad es la experiencia (X). Por lo tanto se ordena la base de datos según esta variable y se eliminan α observaciones centrales (en este caso se eliminó el 20% de ellas equivalente a 254). Para los dos grupos restantes (con 507 observaciones cada uno) se estimaron los siguientes modelos:

³Fuente: Encuesta CASEN año 1994 para hombres de la VII Región

$$\begin{aligned}\ln y_h &= 3,9169 + 0,1532S + 0,0321X - 0,0003X^2 + \varepsilon \\ SCR_1 &= 237,5660 \\ n &= 507\end{aligned}$$

para el primer grupo, y

$$\begin{aligned}\ln y_h &= 4,2456 + 0,1453S + 0,0286X - 0,0003X^2 + \varepsilon \\ SCR_2 &= 355,2361 \\ n &= 507\end{aligned}$$

para el segundo grupo. A partir de estas dos ecuaciones se obtienen sus respectivas SCR.

Finalmente, se plantea la hipótesis nula como:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_\mu^2$$

El estadístico se obtiene con el cociente entre las SCR teniendo cuidado de que en el numerador se ubique aquélla que tenga el valor más alto, es decir⁴,

$$F^c = \frac{355,2361}{237,5660} = 1,4953$$

El F de tabla para este ejemplo corresponde al de 504 grados de libertad, tanto para el numerador como para el denominador. Desafortunadamente, al buscar en la tabla no encontramos el valor del estadístico para estos grados de libertad. Es por ello que comparamos el calculado con el $F_{500,\infty}$ de tabla, el cual corresponde a un valor de 1.11 para un 95 % de probabilidades. Dado que el F calculado es superior al F crítico concluiríamos que el modelo es heterocedástico, ya que se rechaza la hipótesis nula.

Sin embargo, el estadístico hallado se encuentra muy cerca de la zona de aceptación. En otras palabras, nuestra prueba de hipótesis no parece ser muy potente. Este escenario nos hace dudar de la existencia de heterocedasticidad en la muestra.

Una posible explicación para nuestro problema radica en que el Test de **GQ** es confiable siempre y cuando se tenga certeza de cuál es la variable que

⁴Nótese que el F calculado queda expresado solo en términos de las SCR ya que las dos ecuaciones tienen el mismo número de grados de libertad.

genera la heterocedasticidad. Obviamente, nosotros no sabemos con certeza cual es esta variable. Ante esta situación, se hará la detección a través del Test de White. Este test consiste en regresionar los residuos estimados al cuadrado en contra de todas las demás variables, las mismas elevadas al cuadrado y los productos cruzados como se muestra a continuación:

$$\begin{aligned}\varepsilon^2 &= 0,0563 - 0,0006X^2 + (3,25E * 10^{-7}) X^{2^2} - (1,85 * 10^{-5}) X^2X + \\ &\quad (2,86 * 10^{-5}) X^2S + 0,0501X - 0,0020XS + 0,0003S + 0,0018S^2 \\ R^2 &= 0,0087 \\ n &= 1268\end{aligned}$$

A partir de esta información, se obtiene la siguiente variable aleatoria:

$$nR^{2*} \sim \chi_{p-1}^2$$

Luego,

$$1268 \cdot 0,0087 = 10,9778$$

Finalmente, buscamos el estadístico χ^2 para (9-1) grados de libertad y un 5% de confianza, el cual es igual a 15.5073 y dado que este valor se encuentra en la zona de aceptación, la Función de Mincer estimada es homocedástica.

4.4.2. Solución de la Heterocedasticidad.

Una vez que se detecta la presencia de heterocedasticidad en los residuos, la pregunta siguiente es cómo se resuelve para obtener estimaciones eficientes.

Dos casos son posibles de analizar para corregir la heterocedasticidad. Primero, si conocemos las varianzas de los errores (σ_i^2) utilizamos un método particular de Mínimos Cuadrados Generalizados, conocido como *Mínimos Cuadrados Ponderados*. Supongamos que tenemos la Función de Regresión:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

Al estimar los parámetros por Mínimos Cuadrados Ordinarios, obtenemos la siguiente matriz:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Como se discutió anteriormente para poder utilizar Mínimos Cuadrados Generalizados, teníamos que encontrar una matriz P tal que

$$\mathbf{P}\boldsymbol{\Psi}\mathbf{P}' = \mathbf{I}_n$$

Si conocemos σ_i^2 la matriz \mathbf{P} puede expresarse como:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix}$$

$$\mathbf{P}' \times \mathbf{P} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix}$$

Como es sencillo de comprobar, si \mathbf{P} se define de esta manera, entonces $\mathbf{P}\Psi\mathbf{P}' = \mathbf{I}_n$.

El procedimiento de corrección implica simplemente multiplicar la función de regresión por esta matriz \mathbf{P} . En este caso, dada la forma de \mathbf{P} esto significa simplemente dividir cada una de las variables asociadas a un individuo por la desviación estándar relevante para ese individuo.

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\mu}$$

$$\frac{Y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{X_{1i}}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \cdots + \beta_K \frac{X_{Ki}}{\sigma_i} + \frac{\mu_i}{\sigma_i}$$

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\mu}^*$$

Entonces podemos comprobar que este modelo cumple con los supuestos clásicos ya que la varianza de $\boldsymbol{\mu}^*$ es constante:

$$Var(\mu_i^*) = Var\left(\frac{\mu_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2} Var(\mu_i) = 1$$

Por lo tanto, los estimadores que se obtienen son insesgados y de mínima varianza (dado el teorema de Gauss-Markov).

El segundo caso que se debe enfrentar, es la corrección de la heterocedasticidad cuando no se conoce los valores de σ_i^2 . En este caso para solucionar el

problema debemos asumir algún tipo de comportamiento de la varianza de los errores. Es decir, debemos plantearnos el supuesto que la varianza se correlaciona en alguna forma con alguna variable o algún parámetro del modelo o ambos.

Existen varios supuestos típicos, entre los que se destacan:

1. Supuesto de correlación con cuadrado de variable explicativa

Se asume que la varianza está correlacionada con alguna de las variables explicativas al cuadrado. Expresándolo matemáticamente:

$$\sigma_i^2 = \sigma_\mu^2 X_{ji}^2$$

En este caso la matriz \mathbf{P} es la siguiente:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{X_{j1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{X_{j2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{X_{jn}} \end{bmatrix}$$

Ahora, la función transformada a estimar es la siguiente:

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\mu}$$

$$\frac{Y_i}{X_{ji}} = \beta_0 \frac{1}{X_{ji}} + \dots + \beta_j \frac{X_{ji}}{X_{ji}} + \dots + \beta_K \frac{X_{Ki}}{X_{ji}} + \frac{\mu_i}{X_{ji}}$$

$$\frac{Y_i}{X_{ji}} = \beta_0 \frac{1}{X_{ji}} + \dots + \beta_j + \dots + \beta_K \frac{X_{Ki}}{X_{ji}} + \frac{\mu_i}{X_{ji}}$$

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\mu}^*$$

Calculando la varianza del error poblacional, nos encontramos con que cumple el supuesto de la homocedasticidad, por lo que se pueden obtener estimadores insesgados y de mínima varianza por Mínimos Cuadrados Ordinarios. La varianza queda:

$$Var(\mu_i^*) = E\left(\frac{\mu_i}{X_{ji}}\right)^2 = \frac{1}{X_{ji}^2} \sigma_\mu^2 X_{ji}^2 = \sigma_\mu^2$$

En este, y en los casos siguientes, es necesario ser cuidadoso con la interpretación de los parámetros estimados. Fíjese que en este caso β_j hace las veces de parámetro de posición (“constante”) en la regresión y que β_0 relaciona la variable dependiente con la variable explicativa X_j . La interpretación de los parámetros no es obvia.

2. Supuesto de correlación con nivel de variable explicativa

Este caso es similar al primero, con la diferencia de que la variable está expresada en primer grado, es decir la varianza del error es:

$$E(\mu_i^2) = \sigma_\mu^2 X_{ji}$$

En este nuevo caso, la matriz \mathbf{P} queda de la siguiente manera:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{\sqrt{X_{j1}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{X_{j2}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{X_{jn}}} \end{bmatrix}$$

La función transformada a estimar es la siguiente:

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\mu}$$

$$\frac{Y_i}{\sqrt{X_{ji}}} = \beta_0 \frac{1}{\sqrt{X_{ji}}} + \beta_1 \frac{X_{1i}}{\sqrt{X_{ji}}} + \dots + \beta_j \frac{X_{ji}}{\sqrt{X_{ji}}} + \dots + \beta_K \frac{X_{Ki}}{\sqrt{X_{ji}}} + \frac{\mu_i}{\sqrt{X_{ji}}}$$

$$\frac{Y_i}{\sqrt{X_{ji}}} = \beta_0 \frac{1}{\sqrt{X_{ji}}} + \beta_1 \frac{X_{1i}}{\sqrt{X_{ji}}} + \dots + \beta_j \sqrt{X_{ji}} + \dots + \beta_K \frac{X_{Ki}}{\sqrt{X_{ji}}} + \frac{\mu_i}{\sqrt{X_{ji}}}$$

$$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\mu}^*$$

Luego al calcular la varianza del nuevo error poblacional tenemos:

$$\text{Var}(\mu_i^*) = E\left(\frac{\mu_i}{\sqrt{X_{ji}}}\right)^2 = \frac{1}{X_{ji}} \sigma_\mu^2 X_{ji} = \sigma_\mu^2$$

3. Supuesto de correlación con cuadrado de la esperanza de variable dependiente

Consiste en suponer que la varianza está correlacionada con la esperanza de la variable dependiente al cuadrado, es decir:

$$\sigma_i^2 = \sigma_\mu^2 \times E[Y_i]^2$$

La matriz \mathbf{P} es la siguiente:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{E(\hat{Y}_1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{E(\hat{Y}_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{E(\hat{Y}_n)} \end{bmatrix}$$

Ahora, la función a estimar transformada al premultiplicarla por la matriz \mathbf{P} , es la siguiente:

$$\frac{Y_i}{E(\hat{Y}_i)} = \beta_0 \frac{1}{E(Y_i)} + \beta_1 \frac{X_{1i}}{E(Y_i)} + \beta_2 \frac{X_{2i}}{E(Y_i)} + \cdots + \beta_K \frac{X_{Ki}}{E(Y_i)} + \frac{\mu_i}{E(Y_i)}$$

La varianza del error en este caso queda:

$$Var(\mu^*) = E\left(\frac{\mu_i}{E(Y_i)}\right)^2 = \frac{1}{E(Y_i)} \sigma_\mu^2 E(Y_i) = \sigma_\mu^2$$

El problema que presenta este caso, es que la esperanza de la variable Y , depende de los valores de los parámetros estimados, por lo que no se podría utilizar en un solo paso. Sin embargo, lo que sí se puede conocer es \hat{Y}_i , que es un estimador insesgado de la esperanza. Para eso, se estima el modelo original sin considerar problemas de heterocedasticidad, obteniendo \hat{Y}_i . Luego se transforma la ecuación del siguiente modo:

$$\frac{Y_i}{\hat{Y}_i} = \beta_0 \frac{1}{\hat{Y}_i} + \beta_1 \frac{X_{1i}}{\hat{Y}_i} + \beta_2 \frac{X_{2i}}{\hat{Y}_i} + \cdots + \beta_K \frac{X_{Ki}}{\hat{Y}_i}$$

Este proceso se conoce como *Mínimos Cuadrados Ponderados en Dos Etapas*, y se puede continuar iterando un número ilimitado de veces hasta que eventualmente se llegue a valores estables.

4. Supuesto de relación logarítmica

Un cuarto método que se utiliza para eliminar la heterocedasticidad es transformar los datos en logaritmos. En este método no se realiza ningún supuesto sobre la varianza del error, de modo que la función a estimar es:

$$\ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \dots + \beta_K \ln X_{Ki}$$

Se espera que la transformación reduzca la posibilidad de presenciar heterocedasticidad.

En general, no es fácil decidir cuál de estos casos se debe utilizar. La decisión dependerá de la naturaleza del problema estudiado y de la severidad de la heterocedasticidad. Recuerde que en los tres primeros casos sólo se está especulando respecto de cómo podrá ser el comportamiento de la varianza, a menos que se cuente con información previa que indique la forma específica que toma la heterocedasticidad. Note además, que cuando se tienen más de dos variables explicativas, no se puede saber a priori cuál de ellas es la mejor para transformar los datos. Una buena medida práctica consiste en graficar los errores al cuadrado versus cada una de las variables explicativas (alternativamente, calcular los coeficientes de correlación), para determinar cual está más relacionada.

4.5. Autocorrelación

En este capítulo hemos estudiado las diferentes formas que puede tomar la matriz de varianzas y covarianzas de los errores. Primero presentamos el caso general, en que la matriz de covarianzas es distinta a la varianza del error multiplicada por la matriz identidad. Luego resolvimos el caso particular de heterocedasticidad, situación que es típica en datos de corte transversal. Ahora analizaremos el caso en que las covarianzas entre los errores son distintas de cero. Esta situación es común en datos provenientes de *Series de Tiempo*, donde la información tanto de la variable dependiente como de las explicativas ha sido obtenida en períodos sucesivos de tiempo.

Existen muchos casos en los que podemos esperar que una variable observada en el presente esté correlacionada o determinada por los valores de otras variables o de sí misma, pero de períodos anteriores al que se está observando. Esta situación puede explicarse por rezagos en las respuestas de los agentes económicos ante cambios de las condiciones del entorno. En otras

palabras, es común observar que ante cambios en las variables económicas, los agentes no internalicen inmediatamente estos cambios sino que, por el contrario, la nueva información es incorporada paulatinamente a las decisiones de los agentes. Desde el punto de vista de la regresión esto puede verse reflejado en el componente de error del modelo. Este fenómeno puede deberse a que la variable Y , depende de valores pasados de sí misma, a que la matriz de variables explicativas contiene valores corrientes y rezagados de algunas variables explicatorias, o bien a que el error dependa de valores de errores previos. La última situación es en la que estamos interesados y se conoce en la literatura como autocorrelación de errores.

Existen muchas formas de *autocorrelación* y cada una de ellas tiene implicancias sobre la estructura de la matriz de varianzas y covarianzas del error. Nosotros analizaremos un caso particular conocido como Autorregresión de Primer Orden.

Supongamos que tenemos el siguiente modelo:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \mu_t$$

donde t representa el período de tiempo correspondiente. La relación entre el error actual y el error del período anterior se define como:

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t \quad -1 < \rho < 1$$

Esto es lo que se conoce como **Proceso Autoregresivo De Primer Orden**. Lo que esta formulación nos indica es que el error en el período t está relacionado linealmente con el error del período $t - 1$. La forma que toma esta relación depende del signo y magnitud que adopte el parámetro ρ , el cual se conoce como coeficiente de correlación. Si ρ es positivo entonces dos errores consecutivos tenderán a tener el mismo signo. En este caso se dice que existe autocorrelación positiva de errores. Si ρ es negativo, entonces dos errores consecutivos tenderán a tener signo contrario, existiendo en este caso autocorrelación negativa. Finalmente, si ρ es cero o muy pequeño la autocorrelación desaparece.

En esta formulación se asume que ε_t cumple con todos los supuestos clásicos, es decir:

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ Var(\varepsilon_t) &= \sigma_\varepsilon^2 \\ E(\varepsilon_t, \varepsilon_{t-j}) &= Cov(\varepsilon_t, \varepsilon_{t-j}) = 0 \quad \forall j \neq 0 \end{aligned} \quad (4.10)$$

Es posible mostrar que los errores poblacionales μ_t , cumplen con algunos de los supuestos clásicos, pero no con todos, ya que la condición de que la covarianza entre los errores sea cero no se mantiene cuando estamos en presencia de autocorrelación.

Para demostrar este enunciado reemplazamos en forma sucesiva el error del período anterior en el error corriente tal que:

$$\begin{aligned}\mu_t &= \rho\mu_{t-1} + \varepsilon_t \\ \mu_{t-1} &= \rho\mu_{t-2} + \varepsilon_{t-1} \\ \mu_{t-2} &= \rho\mu_{t-3} + \varepsilon_{t-2} \\ &\vdots \\ \mu_{t-j} &= \rho\mu_{t-j-1} + \varepsilon_{t-j} \\ &\vdots\end{aligned}$$

reemplazando μ_{t-1} en la primera ecuación:

$$\mu_t = \rho(\rho\mu_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

y sucesivamente

$$\mu_t = \rho(\rho(\rho\mu_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t$$

simplificando:

$$\mu_t = \rho^3\mu_{t-3} + \rho^2\varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$$

por lo tanto, podemos generalizar la expresión como:

$$\mu_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j},$$

dado que el $\lim_{j \rightarrow \infty} \rho^j \mu_{t-j} = 0$.

Comprobemos qué sucede con los supuestos clásicos. Para ello calculemos la esperanza del error:

$$E(\mu_t) = E\left(\sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \rho^j E(\varepsilon_{t-j}) = 0$$

dado que $E(\varepsilon_{t-j}) = 0$ por el supuesto expresado en 4.10

En el caso de la varianza se tiene:

$$\begin{aligned} Var(\mu_t) &= E(\mu_t - E(\mu_t))^2 = E\left(\sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}\right)^2 \\ Var(\mu_t) &= E[(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \dots)(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \dots)] \\ Var(\mu_t) &= E[\varepsilon_t^2 + \rho\varepsilon_t\varepsilon_{t-1} + \rho^2\varepsilon_t\varepsilon_{t-2} + \dots + \rho\varepsilon_{t-1}\varepsilon_t + \rho^2\varepsilon_{t-1}^2 + \rho^3\varepsilon_{t-1}\varepsilon_{t-2} + \dots] \end{aligned}$$

Aplicando esperanza a cada uno de los términos obtenemos

$$\begin{aligned} Var(\mu_t) &= E[\varepsilon_t^2] + \rho^2 E[\varepsilon_{t-1}^2] + \rho^4 E[\varepsilon_{t-2}^2] + \rho^6 E[\varepsilon_{t-3}^2] + \dots \\ Var(\mu_t) &= \sigma_\varepsilon^2 + \rho^2 \sigma_\varepsilon^2 + \rho^4 \sigma_\varepsilon^2 + \rho^6 \sigma_\varepsilon^2 + \dots \end{aligned}$$

dado que $E(\varepsilon_t \varepsilon_{t-j}) = 0 \quad \forall j \neq 0$.

$$\begin{aligned} Var(\mu_t) &= \sigma_\varepsilon^2 (1 + \rho^2 + \rho^4 + \rho^6 + \dots) = \sigma_\varepsilon^2 (1 + \rho^2 + (\rho^2)^2 + (\rho^2)^3 + \dots) \\ Var(\mu_t) &= \sigma_\varepsilon^2 \left(\frac{1}{1 - \rho^2} \right) \end{aligned}$$

Con la covarianza entre los errores podemos observar:

$$\begin{aligned} Cov(\mu_t, \mu_{t-1}) &= E(\mu_t - E(\mu_t))(\mu_{t-1} - E(\mu_{t-1})) = E(\mu_t \cdot \mu_{t-1}) \\ Cov(\mu_t, \mu_{t-1}) &= E[(\rho\mu_{t-1} + \varepsilon_t)\mu_{t-1}] \\ Cov(\mu_t, \mu_{t-1}) &= E[\rho\mu_{t-1}^2 + \varepsilon_t\mu_{t-1}] \\ Cov(\mu_t, \mu_{t-1}) &= \rho E[\mu_{t-1}^2] + E[\varepsilon_t] \times E[\mu_{t-1}] \\ Cov(\mu_t, \mu_{t-1}) &= \rho\sigma_\mu^2 \\ Cov(\mu_t, \mu_{t-1}) &= \left(\frac{\rho}{1 - \rho^2} \right) \sigma_\varepsilon^2 \end{aligned}$$

Generalizando se puede obtener que:

$$\begin{aligned} Cov(\mu_t, \mu_{t-1}) &= \frac{\rho}{1 - \rho^2} \sigma_\varepsilon^2 \\ Cov(\mu_t, \mu_{t-2}) &= \frac{\rho^2}{1 - \rho^2} \sigma_\varepsilon^2 \\ Cov(\mu_t, \mu_{t-3}) &= \frac{\rho^3}{1 - \rho^2} \sigma_\varepsilon^2 \\ &\vdots \\ Cov(\mu_t, \mu_{t-k}) &= \frac{\rho^k}{1 - \rho^2} \sigma_\varepsilon^2 \end{aligned}$$

Así podemos construir la matriz de varianza - covarianza de los errores como:

$$E(\boldsymbol{\mu}\boldsymbol{\mu}') = \sigma_\varepsilon^2 \Psi_n = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^k \\ \rho & 1 & \rho & \dots & \rho^{k-1} \\ \rho^2 & \rho & 1 & \dots & \rho^{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^k & \rho^{k-1} & \rho^{k-2} & \dots & 1 \end{bmatrix}$$

Como se demostró en la sección 4.2 el estimador mínimo cuadrático seguirá siendo insesgado, pero no de mínima varianza, ya que existe otro estimador insesgado de mínima varianza que es el obtenido por Mínimos Cuadrados Generalizados.

4.5.1. Detección de Autocorrelación.

Al igual que en el caso de la heterocedasticidad, cuando se estima un modelo no se sabe si los resultados sufren de autocorrelación. Para ello existen distintas pruebas que pueden realizarse sobre los residuos de la regresión, que permiten detectar autocorrelación.

Prueba de Durbin-Watson

La prueba de Durbin Watson nos permite verificar la no existencia de autocorrelación de primer orden. El estadístico viene definido por:

$$dw = \frac{\sum_{t=2}^{\infty} (e_t - e_{t-1})^2}{\sum_{t=1}^{\infty} e_t^2} = \frac{\sum_{t=2}^{\infty} e_t^2 - 2 \sum_{t=2}^{\infty} e_t e_{t-1} + \sum_{t=2}^{\infty} e_{t-1}^2}{\sum_{t=1}^{\infty} e_t^2}$$

si asumimos que $\sum_{t=2}^{\infty} e_t^2 \approx \sum_{t=2}^{\infty} e_{t-1}^2$ entonces podemos escribir

$$\begin{aligned} dw &= \frac{2 \sum_{t=2}^{\infty} e_t^2 - 2 \sum_{t=2}^{\infty} e_t e_{t-1}}{\sum_{t=1}^{\infty} e_t^2} \\ &= 2 \frac{\sum_{t=2}^{\infty} e_t^2}{\sum_{t=1}^{\infty} e_t^2} - 2 \frac{\sum_{t=2}^{\infty} e_t e_{t-1}}{\sum_{t=1}^{\infty} e_t^2} \\ &= 2 - 2 \frac{\sum_{t=2}^{\infty} e_t e_{t-1}}{\sum_{t=1}^{\infty} e_t^2} \end{aligned}$$

Ahora, consideremos el modelo autorregresivo de primer orden

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t$$

Podemos estimar este modelo con MCO como

$$e_t = \hat{\rho} e_{t-1} + v_t$$

donde por las propiedades de los estimadores MCO y por la aproximación realizada anteriormente, sabemos que

$$\hat{\rho} = \frac{\sum_{t=2}^{\infty} e_t e_{t-1}}{\sum_{t=1}^{\infty} e_t^2} = \frac{\sum_{t=2}^{\infty} e_t e_{t-1}}{\sum_{t=1}^{\infty} e_{t-1}^2}$$

por lo tanto el estadístico dw queda

$$dw = 2(1 - \hat{\rho}) \tag{4.11}$$

Se puede demostrar que este estadístico va a estar acotado por los valores 0 y 4, donde los valores cercanos a 0 van a indicar autocorrelación positiva, cercanos a 4 autocorrelación negativa y cercanos a 2 indicarán la no existencia de autocorrelación.

Para entender mejor la relación entre este estadístico y el nivel de autocorrelación ρ considere la ecuación 4.11 para la cual tenemos los siguientes resultados

- Si $\hat{\rho} = 0$, lo cual implica ausencia de autocorrelación, entonces $dw = 2$.
- Si $\hat{\rho} = 1$, lo cual implica autocorrelación positiva total, entonces $dw = 0$.
- Si $\hat{\rho} = -1$, lo cual implica autocorrelación negativa total, entonces $dw = 4$.

En el caso de valores intermedios, es decir distintos de 0, 2, ó 4, se requiere una distribución para el estadígrafo. Sin embargo, el dw es calculado con base en e_i , que a su vez depende de los X dados. Por consiguiente, a diferencia de las pruebas t , F o χ^2 no hay un valor crítico único que lleve a rechazar o aceptar la hipótesis nula de que no hay correlación serial de primer orden en las perturbaciones μ_i .

Sin embargo, la Prueba Durbin - Watson cuenta con un límite inferior d_i y un límite superior d_s tales que si el dw calculado cae por fuera de éstos valores críticos puede tomarse una decisión sobre la posible presencia de correlación serial positiva o negativa. En el caso que el valor este dentro de este intervalo, entonces caemos en un área de indecisión. Estos límites dependen únicamente del número de observaciones y del número de variables explicativas. En la figura 4.1 se presentan todas las áreas posibles para el estadístico de Durbin - Watson. Si el dw es menor d_i estamos en presencia de autocorrelación positiva, por el contrario si el dw es mayor que $4 - d_s$ estamos en presencia de autocorrelación negativa. Si el dw está entre el límite d_s y $4 - d_i$ podemos aceptar la hipótesis de no autocorrelación, mientras que si el dw cae en las áreas sombreadas de la figura entonces, no se puede decidir con esta información si existe o no autocorrelación en los errores.

Ejemplo 1. Función de Consumo

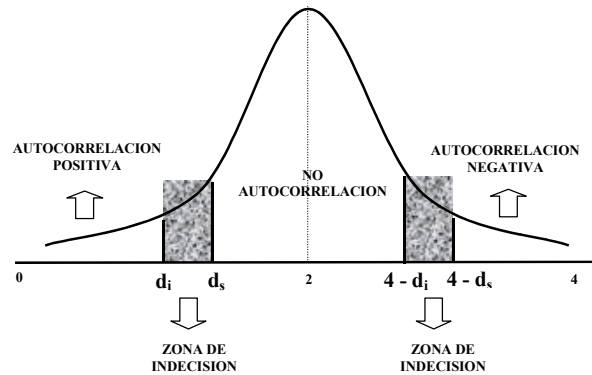


Figura 4.1: Regla de decisión para Autocorrelación

Retomemos la Función de Consumo estimada en el capítulo anterior. Incorporaremos a los resultados mostrados en esa oportunidad el estadístico dw asociado a la muestra. Este estadístico es generalmente entregado por los programas econométricos. Los resultados obtenidos son:

$$\begin{aligned}
 C_t &= 0,3152 + 0,6134Y_t + \varepsilon_t \\
 d.s. &= 0,04506 \quad 0,0118 \\
 R^2 &= 0,9869 \\
 \overline{R^2} &= 0,9866 \\
 dw &= 1,22 \\
 n &= 38
 \end{aligned}$$

donde C_t es el nivel de consumo en el período t e, Y_t es el ingreso real disponible en el mismo período, tal como se presentó en su oportunidad en el Capítulo 3. También se tiene que el estadístico dw es de 1.22 para una muestra de tamaño 38. Procederemos entonces a detectar la existencia de autocorrelación. Primero, buscamos en la tabla del estadístico Durbin - Watson los límites inferior y superior para un $n = 38$ y k variables explicativas, en este caso, $k = 1$. De esta manera se tienen los siguientes límites:

$$\begin{aligned}
 d_i &= 1,427 \\
 d_s &= 1,535
 \end{aligned}$$

Ubicando los límites en la recta de decisión, es posible verificar que el estadístico dw arrojado en los resultados es menor al límite inferior hallado en la tabla. Es decir, con un 95% de probabilidad, la muestra presenta autocorrelación positiva.

Ejemplo 2. Función de Importaciones

En forma análoga, recordemos la Función de Importaciones estimada en el capítulo 3. Para ella los resultados son como se muestra a continuación:

$$\begin{aligned} \ln(M_t) &= 0,47078 + 0,9923 \ln(PIB) - 0,01299TCR_{t-1} + \varepsilon_t \\ d.s. &= 2,6156 \quad 0,1638 \quad 0,0031 \\ R^2 &= 0,9617 \\ \overline{R^2} &= 0,9598 \\ dw &= 1,29 \\ n &= 42 \end{aligned}$$

Buscaremos entonces los límites en la tabla de Durbin-Watson para un nivel de significancia del 5%, en el cruce de n igual a 42 y k igual a 2. Llevando a cabo este procedimiento se obtiene:

$$\begin{aligned} d_i &= 1,391 \\ d_s &= 1,600 \end{aligned}$$

Al igual que en el ejemplo anterior, encontramos que el estadístico dw es menor al límite inferior, por lo que podemos concluir que existe correlación serial positiva en la muestra.

4.5.2. Solución de la autocorrelación.

Recordemos que el estimador del nivel de correlación está dado por:

$$\hat{\rho} = \frac{\sum_{t=2}^{\infty} e_t e_{t-1}}{\sum_{t=1}^{\infty} e_t^2}$$

Además, la relación entre el estadístico de Durbin-Watson y el coeficiente de correlación ρ , es la siguiente:

$$dw = 2(1 - \rho)$$

Luego, lo que hacemos para solucionar el problema de autocorrelación depende de si conocemos o no conocemos ρ .

a. Si conocemos ρ

En este caso Durbin y Watson sugieren un procedimiento en el que se estiman los parámetros en base a la siguiente transformación de la función a estimar:

$$Y_t = \beta X_t + \mu_t$$

Rezagando todas las variables un período obtenemos

$$Y_{t-1} = \beta X_{t-1} + \mu_{t-1}$$

Podemos multiplicar toda la expresión por el coeficiente de correlación ρ , de modo de obtener:

$$\rho Y_{t-1} = \beta \rho X_{t-1} + \rho \mu_{t-1}$$

Ahora, si restamos la última ecuación a la primera, tenemos:

$$Y_t - \rho Y_{t-1} = \beta (X_t - \rho X_{t-1}) + \mu_t - \rho \mu_{t-1}$$

Para evitar mayores confusiones, se pueden definir nuevas variables de la siguiente forma:

$$\begin{aligned} Y_t^* &= Y_t - \rho Y_{t-1} \\ X_t^* &= X_t - \rho X_{t-1} \\ \varepsilon_t &= \mu_t - \rho \mu_{t-1} \end{aligned}$$

La función de regresión queda

$$Y_t^* = \beta X_t^* + \varepsilon_t$$

Por la definición de proceso autorregresivo de primer orden y las propiedades de ε_t especificadas en la ecuación 4.10, podemos observar que los errores de esta regresión cumplen con los supuestos del modelo clásico. Por ende, los estimadores obtenidos por este proceso son insesgados y poseen mínima varianza.

b. Si no conocemos ρ

En este caso Cochrane-Orcutt proponen un procedimiento iterativo, bastante similar al anterior, con la diferencia de que se ocupa un estimador del nivel de autocorrelación ρ , que se obtiene a partir de la prueba Durbin-Watson.

El procedimiento es el siguiente:

1. Estimar la regresión original por Mínimos Cuadrados Ordinarios, ignorando la presencia de autocorrelación entre los errores.
2. Utilizar los residuos obtenidos en la etapa anterior para determinar el parámetro ρ por la regresión:

$$e_t = \hat{\rho}e_{t-1} + \varepsilon_t$$

o por medio del estadístico dw de la regresión original, es decir:

$$\hat{\rho} = \frac{\sum_{t=2}^{\infty} e_t e_{t-1}}{\sum_{t=1}^{\infty} e_t^2}$$

3. Se utiliza $\hat{\rho}$ para obtener las variables cuasidiferenciadas:

$$\begin{aligned} Y_t - \hat{\rho}Y_{t-1} &= \beta(X_t - \hat{\rho}X_{t-1}) + \varepsilon_t \\ Y_t^* &= \tilde{\beta}X_t^* + \varepsilon_t \end{aligned}$$

4. Estimar por Mínimos Cuadrados Ordinarios el modelo con variables transformadas para obtener el estimador de $\tilde{\beta}$.
5. Utilizar el vector $\tilde{\beta}$ para generar una nueva serie de residuos y estimar nuevamente el parámetro $\hat{\rho}$.

Se continúa con este proceso hasta alcanzar un nivel de convergencia fijado de antemano.

Ejemplo 1. Función de Consumo

Dado que se detectó la presencia de autocorrelación positiva en nuestra estimación de la Función de Consumo, ahora procederemos a solucionarla. De la información anterior se sabe que,

$$dw = 1,22$$

También sabemos que,

$$dw = 2(1 - \hat{\rho})$$

Luego,

$$\begin{aligned}\hat{\rho} &= 1 - \frac{dw}{2} = 1 - \frac{1,22}{2} \\ \hat{\rho} &= 0,39\end{aligned}$$

Ahora, se procede a generar una nueva base de datos a partir de la original de manera tal que:

$$C_t^* = \beta_0 + \beta_1 \times Y_t^* + \varepsilon_t$$

donde

$$\begin{aligned}C_t^* &= C_t - \hat{\rho}C_{t-1} \\ Y_t^* &= Y_t - \hat{\rho}Y_{t-1}\end{aligned}$$

siendo, como siempre, C el nivel de consumo, Y el ingreso real disponible y ε_t el error que cumplirá con los supuestos del modelo clásico. Llevando a cabo la estimación de los datos transformados se tiene:

$$\begin{aligned}C_t^* &= 191606,8 + 0,6133Y_t^* + \varepsilon_t \\ d.s. &= 41635 \quad 0,0169 \\ R^2 &= 0,9740 \\ \overline{R^2} &= 0,9733 \\ n &= 37 \\ dw &= 1,91\end{aligned}$$

Es posible observar que el estadístico dw , muy cercano a 2, nos indica la presencia de no autocorrelación. Por lo tanto, el problema ha sido solucionado. También se tiene que el valor estimado de la propensión marginal al consumo no varía significativamente con la corrección por autocorrelación, como debería esperarse. Lo mismo se observa de las desviaciones estándar, que hace que el estimador fluctúe dentro de un rango cercano al obtenido antes de la transformación. Note que este procedimiento implica la pérdida de un grado de libertad al calcularse las cuasi diferencias de las variables originales. Además, el nuevo término constante es $\tilde{\beta}_0 = \hat{\beta}_0 - \hat{\rho}\hat{\beta}_0$. Utilizando los parámetros estimados originales y la estimación del coeficiente de correlación, se puede observar la correspondencia de los resultados en ambos casos.

Ejemplo 2. Función de Importaciones

De la misma forma, procederemos a solucionar la autocorrelación que se halla presente en nuestra estimación de la Función de Importaciones. En primer lugar, encontraremos el valor de $\hat{\rho}$ relevante para nuestro problema:

$$\begin{aligned}\hat{\rho} &= 1 - \frac{dw}{2} = 1 - \frac{1,29}{2} \\ \hat{\rho} &= 0,355\end{aligned}$$

A continuación se procede a transformar las variables del siguiente modo:

$$\begin{aligned}M_t^* &= M_t - \hat{\rho}M_{t-1} \\ PIB_t^* &= PIB_t - \hat{\rho}PIB_{t-1} \\ TCR_t^* &= TCR_t - \hat{\rho}TCR_{t-1}\end{aligned}$$

Para terminar, se realiza la regresión por Mínimos Cuadrados Ordinarios de las variables transformadas de manera tal que:

$$\begin{aligned}\ln(M_t^*) &= \beta_0 + \beta_1 \ln(PIB_t^*) + \beta_2 TCR_{t-1}^* + \varepsilon_t \\ \ln(M_t^*) &= 1,6533 + 0,918 \ln(PIB_t^*) - 0,0229 TCR_{t-1}^* + \varepsilon_t \\ d.s. &= 2,5116 \quad 0,1624 \quad 0,0049 \\ R^2 &= 0,9179 \\ \overline{R^2} &= 0,9135 \\ n &= 41 \\ dw &= 2,03\end{aligned}$$

De la misma manera que en el ejemplo anterior, se observa que la estimación a partir de las variables transformadas ha solucionado el problema de autocorrelación. También, como se esperaba, los estimadores no han variado significativamente de magnitud y tampoco lo hacen sus desviaciones estándar.

Capítulo 5

TOPICOS ADICIONALES

En este capítulo se abordan tres tópicos básicos adicionales para el análisis econométrico. Estos tópicos son (i) la existencia de multicolinealidad entre las variables explicativas, (ii) el test de cambio estructural y (iii) el uso de las variables cualitativas dentro de la matriz de variables explicativas.

El primero de los temas a tratar, conocido como multicolinealidad, se presenta cuando las variables explicativas están “fuertemente” correlacionadas entre sí, lo cual tendrá implicancias sobre la estimación de los parámetros y de las varianzas de los estimadores.

Por su parte, el test de Cambio Estructural, generalmente conocido como test de Chow, es una técnica que intenta verificar la existencia de cambios en la estructura de la economía entre dos o más períodos de tiempo o dos muestras distintas de agentes económicos.

Por último, las variables cualitativas o dummies son utilizadas para incorporar en la regresión distintos elementos de “control” de diferencias poblacionales que no son continuos, tales como el género, analfabetismo, el estado civil, entre otros. Existen muchas variables de este tipo que se consideran relevantes en la explicación del comportamiento de los individuos, y que deben expresarse en términos de la presencia o ausencia de un determinado atributo.

A continuación desarrollaremos cada uno de estos puntos por separado.

5.1. Multicolinealidad

En los estudios empíricos es frecuente encontrar niveles de correlación entre las variables explicativas del modelo de regresión. La existencia de algún grado de asociación entre las variables explicativas del modelo tiene efectos sobre la estimación de los parámetros y de sus varianzas..

Retomemos el modelo lineal general, el cual se expresa como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

donde, la matrix \mathbf{X} de orden $n \times k$, contiene n observaciones y k variables explicativas. Desde la perspectiva del investigador, un buen resultado econométrico es obtener una alta correlación entre la variable dependiente Y y todas las variables explicativas que componen la matrix \mathbf{X} , lo cual se expresa en un alto valor para el estadístico R^2 . Sin embargo, es probable que las propias variables contenidas en la matrix de variables explicativas estén a su vez altamente correlacionadas entre ellas. Por esta razón, se hace necesario evaluar las implicancias que esta correlación tiene sobre los estimadores y sobre las propiedades de éstos.

Conceptualmente, al existir multicolinealidad no es factible separar, en forma nítida, los efectos sobre la variable dependiente de cada una de las variables explicativas. Por lo tanto, no podemos interpretar los coeficientes de la regresión adecuadamente. Un problema serio que impone la multicolinealidad es cuando las variables independientes están correlacionadas de tal forma que alguna de las columnas de la matrix de variables explicativas se puede escribir como una combinación lineal de las otras. Es decir, nos enfrentamos a una situación de colinealidad perfecta, y por lo tanto, no es posible obtener la matrix inversa de $\mathbf{X}'\mathbf{X}$. Como podemos recordar del capítulo tres, el vector de parámetros y la matrix de varianzas se estiman de la forma

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \\ \text{var}(\hat{\boldsymbol{\beta}}) &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Por lo tanto, es importante encontrar la matrix $(\mathbf{X}'\mathbf{X})^{-1}$. Nótese que para obtener esta matrix es necesario calcular la matrix de cofactores y el determinante de ella. El determinante de la matrix $\mathbf{X}'\mathbf{X}$ depende del grado de correlación existente entre las variables explicativas. Se sabe además, que

a mayor multicolinealidad más pequeño será el valor de $\mathbf{X}'\mathbf{X}$ y por tanto mayor el valor de su inversa. Si la multicolinealidad es perfecta (en otras palabras, el grado de correlación es igual a uno) la inversa de la matriz $\mathbf{X}'\mathbf{X}$ no existe, y por lo tanto no se pueden encontrar estimadores para los parámetros del modelo ni para las varianzas. Por otra parte, si la multicolinealidad no es perfecta, pero es alta (correlación distinta de uno), entonces se obtienen estimadores cuyas varianzas son muy grandes.

En la práctica, es poco probable encontrar muestras de datos económicos donde no exista multicolinealidad. Bajo estas circunstancias lo relevante es preguntarse ¿cuál es el grado de multicolinealidad que es tolerable en un determinado estudio?, o bien ¿a qué nivel de multicolinealidad se ven afectados seriamente los estimadores y sus varianzas?.

Para estructurar el análisis es útil dividir el problema considerando dos tipos de multicolinealidad: i) Perfecta e ii) Imperfecta.

Multicolinealidad Perfecta

El problema principal que enfrentamos en el caso de multicolinealidad perfecta es que la matriz $\mathbf{X}'\mathbf{X}$ es singular, es decir, su determinante es cero. Por tanto, no es posible estimar la matriz de parámetros β . Podemos describir el caso de multicolinealidad perfecta de la siguiente forma:

$$\mathbf{X}\mathbf{c} = \mathbf{X}_1c_1 + \mathbf{X}_2c_2 + \dots + \mathbf{X}_kc_k = \mathbf{0} \quad (5.1)$$

$$\mathbf{X}_1 = -\mathbf{X}_2 \frac{c_2}{c_1} - \dots - \mathbf{X}_k \frac{c_k}{c_1} = \mathbf{0} \quad (5.2)$$

donde \mathbf{c} es un vector de constantes no todas iguales a cero, $\mathbf{c} = (c_1, c_2, \dots, c_k)$. Vale decir, la ecuación 5.1 indica que una variable (X_1 por ejemplo) es una función lineal del resto de las variables explicativas para los n individuos. En este caso, el sistema $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ no puede resolver las k ecuaciones.

Multicolinealidad Imperfecta

Existe además, la llamada multicolinealidad imperfecta, que no es detectable a simple vista, puesto que la matriz $\mathbf{X}'\mathbf{X}$ es invertible y se obtendrá un estimador para β . En este caso podemos escribir la ecuación 5.1 como:

$$\begin{aligned} \mathbf{X}\mathbf{c} &= \mathbf{X}_1c_1 + \mathbf{X}_2c_2 + \dots + \mathbf{X}_kc_k \approx \mathbf{0} \\ \mathbf{X}_1 &= -\mathbf{X}_2 \frac{c_2}{c_1} - \dots - \mathbf{X}_k \frac{c_k}{c_1} + \mathbf{v}_1 \end{aligned} \quad (5.3)$$

Donde \approx significa “similar o cercano a”. Note que \mathbf{v}_1 puede ser tratado como un error (similar a ε del modelo lineal general), y la ecuación 5.3 se puede escribir como:

$$\mathbf{X}_1 = -\mathbf{X}_2 d_1 - \dots - \mathbf{X}_k d_k + \mathbf{v}_1 \quad (5.4)$$

que es igual a una ecuación de regresión entre X_1 y las demás variables explicativas. Nótese además que X_1 ha sido escogida arbitrariamente. Llamaremos a la ecuación 5.4 *ecuación auxiliar*. Esto refleja que cualquier variable del modelo puede escribirse como una combinación lineal perfecta o imperfecta del resto de las variables. Por ende, es posible estimar un modelo de regresión lineal considerando como variable dependiente cualquiera de las variables explicativas del modelo y como variables explicativas a todas las demás.

Para clarificar el impacto sobre los estimadores y las varianzas de la multicolinealidad, evaluemos la relación entre las variables explicativas y la estimación de los parámetros y de la varianza. Para ello tomemos el modelo lineal con desviaciones de media

$$y_t = \beta_2 x_{2t} + \beta_3 x_{3t} + \mu_t$$

donde $y_t = Y_t - \bar{Y}$ y $x_{jt} = X_{jt} - \bar{X}_j$. Asumamos que existe una relación entre las variables X_2 y X_3 del siguiente tipo:

$$X_2 = \lambda X_3, \quad \text{donde } \lambda \text{ es una constante}$$

Recordemos que el estimador a través de Mínimos Cuadrados Ordinarios es el siguiente (con este estimador es suficiente, puesto que no se ha violado ningún supuesto del modelo clásico):

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}$$

Luego, podemos expresar la matriz \mathbf{X} como una matriz de orden 1×2 , donde cada elemento de dicha matriz (X_2 y X_3) corresponde a un vector columna de orden $n \times 1$. Luego la matriz $\mathbf{X}'\mathbf{X}$ quedaría expresada como:

$$(\mathbf{x}'\mathbf{x}) = \begin{pmatrix} \mathbf{x}'_2 \\ \mathbf{x}'_3 \end{pmatrix} \begin{pmatrix} \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_2 \mathbf{x}_2 & \mathbf{x}'_2 \mathbf{x}_3 \\ \mathbf{x}'_3 \mathbf{x}_2 & \mathbf{x}'_3 \mathbf{x}_3 \end{pmatrix}$$

Lo que nos interesa conocer, es la matriz inversa de $\mathbf{x}'\mathbf{x}$, recordando el caso de una matriz cuadrada, donde la inversa se define como la matriz adjunta dividida por su determinante. Esto sería:

$$(\mathbf{x}'\mathbf{x})^{-1} = \begin{pmatrix} \mathbf{x}'_2\mathbf{x}_2 & \mathbf{x}'_2\mathbf{x}_3 \\ \mathbf{x}'_3\mathbf{x}_2 & \mathbf{x}'_3\mathbf{x}_3 \end{pmatrix}^{-1} = \frac{\begin{pmatrix} \mathbf{x}'_3\mathbf{x}_3 & -\mathbf{x}'_2\mathbf{x}_3 \\ -\mathbf{x}'_3\mathbf{x}_2 & \mathbf{x}'_2\mathbf{x}_2 \end{pmatrix}}{\begin{vmatrix} \mathbf{x}'_2\mathbf{x}_2 & \mathbf{x}'_2\mathbf{x}_3 \\ \mathbf{x}'_3\mathbf{x}_2 & \mathbf{x}'_3\mathbf{x}_3 \end{vmatrix}}$$

$$(\mathbf{x}'\mathbf{x})^{-1} = \frac{1}{\mathbf{x}'_2\mathbf{x}_2\mathbf{x}'_3\mathbf{x}_3 - (\mathbf{x}'_3\mathbf{x}_2)^2} \begin{pmatrix} \mathbf{x}'_3\mathbf{x}_3 & -\mathbf{x}'_2\mathbf{x}_3 \\ -\mathbf{x}'_3\mathbf{x}_2 & \mathbf{x}'_2\mathbf{x}_2 \end{pmatrix}$$

Luego, la matriz de varianzas, se define como:

$$Var(\boldsymbol{\beta}) = \sigma^2 (\mathbf{x}'\mathbf{x})^{-1}$$

Basta con tomar un elemento de la matriz para ejemplificar el problema. Si analizamos el caso de la varianza del estimador de β_2 , este sería:

$$Var(\beta_2) = \frac{\sigma^2 \mathbf{x}'_3\mathbf{x}_3}{\mathbf{x}'_2\mathbf{x}_2\mathbf{x}'_3\mathbf{x}_3 - (\mathbf{x}'_3\mathbf{x}_2)^2} = \frac{\sigma^2 \mathbf{x}'_3\mathbf{x}_3}{\mathbf{x}'_2\mathbf{x}_2\mathbf{x}'_3\mathbf{x}_3 \left(1 - \frac{(\mathbf{x}'_3\mathbf{x}_2)^2}{\mathbf{x}'_2\mathbf{x}_2\mathbf{x}'_3\mathbf{x}_3} \right)}$$

$$Var(\beta_2) = \frac{\sigma^2}{\mathbf{x}'_2\mathbf{x}_2 \left(1 - \frac{(\mathbf{x}'_3\mathbf{x}_2)^2}{\mathbf{x}'_2\mathbf{x}_2\mathbf{x}'_3\mathbf{x}_3} \right)}$$

Si analizamos la fracción del denominador de este último término:

$$\frac{(\mathbf{x}'_3\mathbf{x}_2)^2}{\mathbf{x}'_2\mathbf{x}_2\mathbf{x}'_3\mathbf{x}_3} = \frac{\sum (x_{2t} \times x_{3t})^2}{\sum x_{3t}^2 \sum x_{2t}^2} = \left[\frac{\sum (X_{2t} - \bar{X}_2)(X_{3t} - \bar{X}_3)}{\sqrt{\sum (X_{3t} - \bar{X}_3)^2 \sum (X_{2t} - \bar{X}_2)^2}} \right]^2 = r_{23}^2$$

donde r_{23}^2 es el coeficiente de correlación de la variable X_2 y X_3 . Recuerde que $0 \leq r_{23}^2 \leq 1$.

Por lo tanto se tiene que:

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\mathbf{x}'_2\mathbf{x}_2 (1 - r_{23}^2)}$$

Teniendo en cuenta las dos posibilidades de Multicolinealidad, se puede observar

- i Si $X_2 = \lambda X_3$, entonces $r^2 = 1$, por lo que la varianza de $\hat{\beta}$ se indetermina. El valor del parámetro también se indetermina, es decir, es imposible obtener un valor.
- ii Si $X_2 = \lambda X_3 + \nu_i$, entonces ν_i permite que la correlación no sea perfecta, pudiendo estimar los parámetros. Sin embargo, mientras mayor sea la correlación existente entre las variables X_2 y X_3 , menor será el denominador de la varianza, y por lo tanto mayor será el valor de ésta.

Cabe mencionar que dado el teorema de Gauss-Markov los estimadores obtenidos de un modelo con multicolinealidad imperfecta seguirán siendo *MELI*, ya que no se ha violado ningún supuesto clásico que sea fundamental para obtener las propiedades de insesgamiento y mínima varianza. No obstante, las varianzas de los estimadores serán muy grandes, por lo que en las pruebas de hipótesis se tenderá a aceptar la hipótesis nula de no significancia individual de los parámetros. Este resultado conducirá erróneamente al investigador a eliminar variables supuestamente no significativas. Además, las estimaciones puntuales de los parámetros serán poco informativas.

Como consecuencia de la multicolinealidad, los estimadores serán volátiles. Esto significa que pequeñas variaciones en el número de observaciones de la muestra, pueden generar cambios importantes en los valores de los estimadores.

En resumen, podemos decir que la multicolinealidad afecta la confiabilidad y la precisión de los estimadores.

5.1.1. Cómo Detectar la Multicolinealidad.

En la sección anterior discutimos los efectos de la multicolinealidad sobre los estimadores y sus varianzas. La siguiente discusión se centrará en la identificación o detección del fenómeno.

De la discusión se deriva que un síntoma clave de la existencia de multicolinealidad, es la presencia de un coeficiente de determinación R^2 alto, junto con pocas variables significativas individualmente. Dado que obtenemos un R^2 alto, la prueba F global rechazará la hipótesis de que los parámetros son todos iguales a cero. Lo cual es abiertamente contradictorio con el hecho que las variables no sean significativas individualmente. Nótese que ésta es una forma que nos permite sospechar la presencia de multicolinealidad, pero en ningún caso representa una prueba formal de ésta.

Una forma bastante corriente de verificar la presencia de multicolinealidad, es observar la matriz de correlación entre las variables explicativas. Para el caso de dos variables es factible evaluar el valor del determinante de esta matriz para evaluar la existencia de multicolinealidad. El determinante de la matriz de correlación está dada por

$$\left| \begin{bmatrix} 1 & r_{23}^2 \\ r_{23}^2 & 1 \end{bmatrix} \right| = R$$

Donde los elementos ubicados en la primera fila y en la primera columna corresponden al coeficiente de correlación de la variable X_2 en relación a las demás variables, y los ubicados en la segunda fila y segunda columna lo son para la variable X_3 . De esta forma, se tiene, en este caso, que los elementos a_{11} y a_{22} corresponden a los coeficientes de correlación de X_2 y X_3 en relación a sí mismos, respectivamente, y es por ello que su valor es 1. En cuanto a los elementos a_{12} y a_{21} representan el coeficiente de correlación primero entre X_2 y X_3 y el segundo entre X_3 y X_2 , dada la igualdad ambos se representan como r_{23}^2 .

Al estimar el determinante de la matriz, es decir $R = 1 - (r_{23}^2)^2$, se obtiene un valor que si tiende a cero, podemos concluir que la multicolinealidad es un problema serio. No obstante, cuando se tiene más de dos variables explicativas la utilización de este método se dificulta por la incapacidad de interpretar el valor del determinante en forma directa.

Alternativamente, es posible una regresión auxiliar donde se regresiona cada variable explicativa con respecto a las otras. Posteriormente se procede a testear mediante una prueba F la significancia global de la regresión auxiliar. Si el valor calculado es mayor que el valor F de tabla, entonces se dice que la variable X_i es colineal con las demás variables.

Aunque el modelo anterior permite revelar la existencia de correlación, debemos aceptar que en todas las muestras encontraremos que las variables están correlacionadas. Por lo tanto, debemos preguntarnos ¿cuándo la Multicolinealidad se transforma en un problema del cual debemos preocuparnos?. Una forma de responder esta pregunta es utilizar la **Regla o Criterio de Klein**, que sugiere que la multicolinealidad puede ser un problema serio sólo si el R^2 de la regresión auxiliar es mayor que el coeficiente de determinación global, es decir, el R^2 obtenido al regresionar la variable Y sobre todas las variables X .

5.1.2. Cómo Solucionar la Multicolinealidad.

Una vez que conocemos el grado de multicolinealidad y éste es importante, entonces debemos intentar solucionarlo. En ese sentido las posibilidades son diversas y se enuncian a continuación:

1. La primera alternativa consiste en incorporar mayor información muestral, es decir, aumentar el tamaño de la muestra. Debido a que la multicolinealidad es un problema muestral, es razonable pensar que mientras más grande sea la muestra, menor será la probabilidad de que la multicolinealidad sea severa. En la misma línea de razonamiento se puede combinar información de corte transversal con series de tiempo (datos de panel).
2. Una segunda posibilidad, es usar información a priori respecto a la relación entre las variables del modelo. Es decir, se entrega de antemano la relación que puede existir entre dos variables. Esta información puede provenir de trabajos empíricos anteriores, donde no existan grandes problemas de multicolinealidad, o de la teoría que sustenta el estudio. De tal forma, lo que se estima es un modelo restringido a la nueva información del modelo. Note que se requiere conocer la forma en que se interrelacionan las variables explicativas, lo cual puede ser un requerimiento bastante restrictivo.
3. Una tercera alternativa es la *transformación de variables*, donde lo que se hace es estimar el modelo con las variables expresadas de otra forma. Una manera común de expresar las variables es en primeras diferencias. Es decir, si tenemos el siguiente modelo, para un período t :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \mu_t$$

De la misma forma se puede obtener el modelo para el período $t - 1$:

$$Y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + \dots + \beta_k X_{kt-1} + \mu_{t-1}$$

Así, si restamos estas ecuaciones vamos a obtener:

$$(Y_t - Y_{t-1}) = \beta_2 (X_{2t} - X_{2t-1}) + \dots + \beta_k (X_{kt} - X_{kt-1}) + \nu_t$$

Donde el término $\nu_t = \mu_t - \mu_{t-1}$. Utilizando este modelo se reduce frecuentemente el grado de multicolinealidad, ya que si existen variables que

están altamente correlacionadas, no hay razón a priori para pensar que las primeras diferencias lo estén. Sin embargo, el problema que puede surgir en esta alternativa es el no cumplimiento de los supuestos del modelo clásico por parte del término de error ν_t . Adicionalmente, se pierde una observación y por consiguiente un grado de libertad, que puede ser muy perjudicial especialmente en el caso de muestras pequeñas. También, puede no ser un método adecuado para casos de datos de corte transversal, donde no hay un ordenamiento temporal o lógico de las observaciones.

4. Por último, una cuarta solución para enfrentar el problema de la multicolinealidad es la eliminación de variables. Aunque esta solución es sencilla, implica incurrir potencialmente en un *sesgo de especificación* o *error de especificación*. El sesgo de especificación surge de la especificación incorrecta del modelo utilizado en el análisis. Veamos esto con un pequeño ejemplo. Supongamos que el modelo definido en desviaciones es el siguiente:

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \mu_i \quad (5.5)$$

Pero al existir evidencia de multicolinealidad importante entre X_2 y X_3 , corregimos el modelo omitiendo la variable X_3 . De esta forma el modelo a estimar es:

$$y_i = \beta_2 x_{2i} + \nu_i$$

Aplicando la fórmula del estimador MCO obtenemos un estimador para β_2 como:

$$\tilde{\beta}_2 = \frac{\sum x_{2i} y_i}{\sum x_{2i}^2}$$

Sin embargo, la variable dependiente y_i se define según el modelo original dado en 5.5.

Reemplazando esta definición, en la ecuación anterior lo que tenemos es:

$$\tilde{\beta}_2 = \frac{\sum x_{2i} (\beta_2 x_{2i} + \beta_3 x_{3i} + \mu_i)}{\sum x_{2i}^2}$$

Simplificando logramos:

$$\tilde{\beta}_2 = \beta_2 + \beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2} + \frac{\sum x_{2i} \mu_i}{\sum x_{2i}^2}$$

Luego, si aplicamos el operador esperanza, lo que vamos a encontrar es lo siguiente:

$$E(\tilde{\beta}_2) = \beta_2 + \beta_3 b_{23}$$

Donde b_{23} representa el coeficiente de la pendiente de la regresión de X_3 sobre X_2 . Por lo tanto $\tilde{\beta}_2$ es una estimación sesgada de β_2 , mientras b_{23} sea distinta de cero. En este sentido, uno podría esperar que b_{23} sea distinto de cero, dado que existe una relación entre ambas que originó la eliminación de X_3 del modelo original.

Ejemplo 2. Función de Importaciones

El ejemplo 1, o de la Función de Consumo, no tiene posibilidades de presentar multicolinealidad debido a que posee sólo una variable explicativa, es por ello que para ejemplificar la detección de multicolinealidad se hará uso del ejemplo 2 de la Función de Importaciones la cual depende del PIB y del TCR.

Procediendo a la detección de este fenómeno regresionaremos la primera variable explicativa en función de la segunda. Es decir,

$$\ln(PIB_t) = f(TCR_{t-1})$$

Los resultados obtenidos de la Regresión Auxiliar son:

$$\begin{aligned} \ln(PIB_t) &= 15,9548 - 0,0180TCR_{t-1} + \varepsilon_t \\ t^c &= (178,5611) (-18,5841) \\ R^2 &= 0,8962 \\ n &= 42 \end{aligned}$$

La Prueba de Significancia Global se construye como:

$$R = \frac{\frac{0,8962}{3-2}}{\frac{1-0,8962}{42-3+1}} = 345,3565 \sim F(1, 40)$$

El valor de tabla es de 4.08 para un 5 % de significancia, resultando un modelo significativo, lo que implica un 95 % de probabilidades de que haya una alta correlación entre las variables explicativas, debido a que se rechaza la hipótesis nula. Según el Criterio de Klein podemos detectar la gravedad de esta situación al comparar el R^2 de la regresión auxiliar con el del modelo

general. Los resultados obtenidos para la Función de Importaciones en el capítulo 3 nos entrega el valor del R^2 , el cual es de 0.9617, que es mayor al R^2 de la Regresión Auxiliar de 0.8962. Por lo tanto, según este criterio la multicolinealidad presente en este modelo no es suficientemente grave. Sin embargo, si queremos obtener estimaciones más precisas de las varianzas de los parámetros estimados, deberíamos buscar una solución a este problema. Por ejemplo, alargando las series de datos.

5.2. Prueba de Cambios Estructurales

Un tema que ha interesado a los economistas es la posibilidad de identificar cambios en la estructura de la economía entre dos períodos, o cambios en la estructura de comportamiento de dos o más grupos de agentes económicos. En la figura 5.1 se presenta un caso que sirve para ilustrar la idea asociada a cambios estructurales.

En el eje Horizontal se mide el tiempo mientras que en el eje vertical se mide cualquier variable económica de interés, tal como el consumo. Si consideramos un momento en el tiempo, digamos el año 1982 señalado en la figura, podemos sospechar que por alguna razón (la crisis económica, por ejemplo) el período anterior a 1982 y el período posterior a 1982 tienen comportamientos distintos en términos de sus parámetros que caracterizan el período. Específicamente el parámetro de pendiente y/o posición de la función de consumo podrían diferir entre períodos.

Si el comportamiento de la economía fuese distinto en ambos períodos, lo correcto sería estimar una ecuación para cada intervalo (regresiones no restringidas del periodo 1 y 2). Si por el contrario, estimamos un solo modelo para todo el período de estudio, entonces implícitamente estamos asumiendo que el valor de los parámetros es el mismo para todo el período muestral. En otras palabras estamos imponiendo una restricción sobre los parámetros de ambos períodos.

El objetivo de la prueba de cambio estructural es determinar si existen diferencias de estructura o comportamiento entre dos muestras o dos períodos de una regresión.

Para verificar esta hipótesis podemos utilizar la prueba F descrita en capítulos anteriores. Recordemos que en esta prueba se comparan las sumatorias de cuadrados residuales de las estimaciones restringidas y no restringidas. Esta prueba, cuando se aplica a problemas de cambio estructural, también se

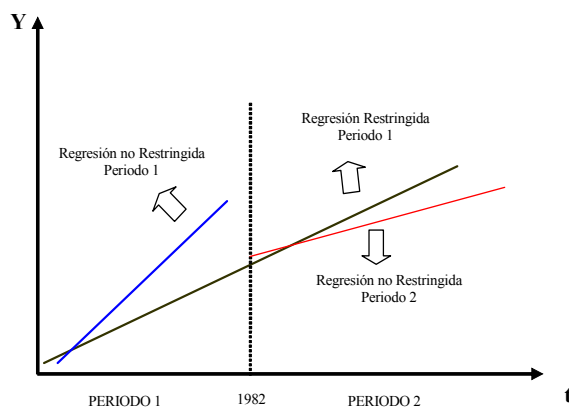


Figura 5.1: Cambio Estructural

denomina “Prueba de Chow”. Si pensamos, que el comportamiento del modelo es el mismo en los dos períodos, estimamos un solo modelo usando el total de observaciones disponibles (N). Implícitamente estamos asumiendo que:

$$\begin{aligned}\beta_0^1 &= \beta_0^2 \\ \beta_1^1 &= \beta_1^2 \\ &\vdots \\ \beta_k^1 &= \beta_k^2\end{aligned}$$

donde β_i^j es el i -ésimo parámetro en el período j ($j = 1, 2$ $i = 1, 2, \dots, k$).

Es decir, que los parámetros del periodo 1 son iguales a los parámetros del periodo 2. De esta estimación se obtiene una sumatoria de cuadrados residuales restringida a la hipótesis nula (SCR_R^*).

Por el contrario, si pensamos que los dos períodos difieren significativamente, podemos estimar dos modelos con distinto número de observaciones en cada uno de ellos (N_1 y N_2 , respectivamente). Para este caso tendremos 2 sumatorias de cuadrados residuales (SCR_1 y SCR_2), la sumatoria de ellas representa la sumatoria de cuadrados no restringida del modelo (SCR_n). Si no existiera diferencia en los parámetros de ambos períodos, no deberíamos encontrarnos con diferencias significativas entre la sumatoria de cuadrados residuales restringida y no restringida. De esta forma podemos construir el siguiente estadígrafo:

$$F = \frac{\frac{SCR R^* - SCR_n}{k}}{\frac{SCR_n}{N - sk}} \sim F_{(k, n-sk)}$$

donde:

$SCR R^*$: Suma de Cuadrados Residuales Restringidos (del total de la muestra).

SCR_n : Suma de Cuadrados Residuales no Restringida (suma de la SCR obtenidas en la estimación de cada grupo de la muestra).

N : Número de Observaciones.

k : Número de Parámetros.

s : Número de Sectores Agrupados o de Períodos Agrupados.

En resumen la prueba consiste en los siguientes pasos:

1. Estimar el modelo general

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t$$

por MCO para el número total de observaciones. Calcular $SCR R^*$.

2. Estimar el mismo modelo para cada muestra por separado.

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t \quad \text{para } N_1 \\ Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t \quad \text{para } N_2 \end{aligned}$$

De cada regresión se obtienen las sumas de cuadrados residuales (SCR_1 y SCR_2). Con estas se obtiene la SCR_n como la suma de las anteriores:

$$SCR_1 + SCR_2 = SCR_n$$

3. Se plantea la hipótesis nula de la siguiente manera:

$$H_0 : \begin{aligned} \beta_0^1 &= \beta_0^2 \\ \beta_1^1 &= \beta_1^2 \\ &\vdots \\ \beta_k^1 &= \beta_k^2 \end{aligned}$$

4. Luego se calcula la variable aleatoria F:

$$F = \frac{\frac{SCRR^* - SCR_n}{k}}{\frac{SCR_n}{N - sk}} = \frac{(SCRR^* - SCR_n) (N_1 + N_2 - sk)}{SCR_n k} \sim F_{(k, n-sk)}$$

5. Por último, se compara el valor anterior con el valor de la distribución F, de manera que si el valor calculado es menor que el valor de la distribución entregado por la tabla, entonces se acepta la hipótesis nula, es decir, se acepta la hipótesis de que no existen cambios estructurales, y los dos períodos o muestras se comportan de la misma manera. En caso contrario, si el F calculado es mayor que el valor F de tabla, se rechaza la hipótesis nula. Ello constituye indicio de que la estructura de determinación de la variable habría cambiado entre ambos períodos.

A continuación, presentamos la evaluación de este test para dos ejemplos distintos, el primero es una muestra de series de tiempo y el segundo es de corte transversal.

Ejemplo 1. Función de Consumo

Retomando la Función de Consumo para el caso chileno, que comprende el período desde 1960 hasta 1997, estimada en capítulos anteriores, verificaremos la posibilidad que exista un cambio estructural en los parámetros antes y después del año 1982, motivados por la crisis que experimentó nuestro país en aquel año.

Para ello seguiremos el procedimiento propuesto por la Prueba de Chow, recién presentado.

1. A partir de la estimación de la función de consumo realizado en el capítulo 3 se obtiene la Suma de Cuadrados Residuales Restringida, dado que en aquella se asumió que el comportamiento del modelo era el mismo en ambos períodos. De esta forma se tiene que,

$$\begin{aligned} C_t &= 0,3152 + 0,6134Y_t + \varepsilon_t \\ d.s. &= 0,04506 \quad 0,0118 \\ t^c &= (6,8947) \quad (52,2269) \\ SCRR^* &= 4,69 * 10^{11} \\ n &= 38 \end{aligned}$$

2. Se divide la muestra en dos períodos: el período 1 comprende los años desde 1960 hasta 1982 inclusive, mientras el segundo período comprende los años restantes, es decir, desde 1983 hasta 1997. Luego, procedemos a estimar el modelo para cada período:

$$\begin{aligned} C_t &= 163415,8 + 0,6700Y_t + \varepsilon_t \\ t^c &= 1,1103 \quad 12,1752 \\ SCR_1 &= 3,15 * 10^{11} \\ N_1 &= 23 \end{aligned}$$

$$\begin{aligned} C_t &= 334518,3 + 0,6085Y_t + \varepsilon_t \\ t^c &= 4,2974 \quad 40,1299 \\ SCR_2 &= 1,36 * 10^{11} \\ N_2 &= 15 \end{aligned}$$

De esta forma, la Suma de Cuadrados Residuales no Restringido es:

$$\begin{aligned} SCR_n &= SCR_1 + SCR_2 = (3,15 * 10^{11}) + (1,36 * 10^{11}) \\ SCR_n &= 4,51 * 10^{11} \end{aligned}$$

3. La hipótesis nula se plantea como:

$$H_0 : \begin{aligned} \beta_0^1 &= \beta_0^2 \\ \beta_1^1 &= \beta_1^2 \end{aligned}$$

4. Construimos a continuación el estadístico:

$$F = \frac{\frac{SCR_n^* - SCR_n}{k}}{\frac{SCR_n}{n - sk}} = \frac{\frac{(4,69 * 10^{11}) - (4,51 * 10^{11})}{2}}{\frac{4,51 * 10^{11}}{38 - 4}} = 0,6785$$

5. Comparando con el valor de tabla que es de 3,23 es posible verificar que el valor anterior es menor, por lo que no existe evidencia suficiente para rechazar la hipótesis nula. Por ello es posible decir con un 95 % de probabilidades que no existe cambio estructural en el modelo, es decir, los dos períodos considerados se comportan de la misma manera.

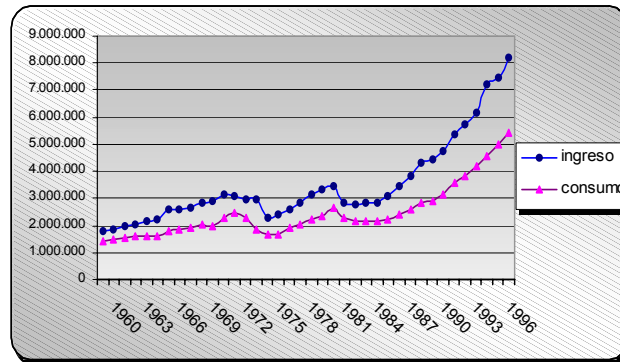


Figura 5.2: Relación Consumo-Ingreso (1960-1997)

Es posible ver en la Figura 5.2 que la relación entre el consumo y el ingreso a través de todo el período considerado no cambia significativamente, corroborando el resultado encontrado a través de la Prueba de Cambio Estructural. Por decirlo en forma simple, las dos series “se mueven” en forma relativamente sincronizada a través del tiempo. Por lo tanto, se asevera el hecho de que es posible concluir con un 95 % de probabilidad que no hay evidencia de cambio estructural o de comportamiento entre el consumo y el ingreso disponible entre los períodos 1960-1982 y 1983-1997 para el caso chileno.

Ejemplo 2. Función de Mincer

El presente modelo ha sido presentado en capítulos anteriores y consiste en una estimación de salarios para los trabajadores de sexo masculino cuyos datos fueron obtenidos a través de la encuesta CASEN para la VIII Región en el año 1994. Plantearemos esta Prueba para evaluar si existen diferencias entre trabajadores divididos en 3 grupos según los años de escolaridad. El primer grupo contendrá a trabajadores que posean menos de 10 años de escolaridad, el segundo grupo que posea entre 10 y menos de 14 años de estudios, y el último con 14 años de escolaridad inclusive y más.

1. El modelo estimado con la muestra total es:

$$\begin{aligned} \ln y_i &= 3,9997 - 0,0004X^2 + 0,0372X + 0,1450S + \varepsilon_t \\ d.s. &= 0,1010 \quad 0,0001 \quad 0,0059 \quad 0,0065 \\ t^c &= 39,5894 \quad -3,1412 \quad 6,3037 \quad 22,3082 \\ SCRR^* &= 779,7495 \\ n &= 1268 \end{aligned}$$

2. Procediendo a las estimaciones de cada sector se tienen sus SCR:

$$\begin{aligned} SCR_1 &= 314,3561 \\ N_1 &= 534 \\ SCR_2 &= 332,3691 \\ N_2 &= 554 \\ SCR_3 &= 101,8453 \\ N_3 &= 180 \end{aligned}$$

Por lo tanto la Suma de Cuadrados Residuales no Restringida es:

$$\begin{aligned} SCR_n &= SCR_1 + SCR_2 + SCR_3 \\ SCR_n &= 314,3561 + 332,3691 + 101,8453 = 748,5705 \end{aligned}$$

3. La hipótesis nula es:

$$H_0 : \begin{aligned} \beta_0^1 &= \beta_0^2 = \beta_0^3 \\ \beta_1^1 &= \beta_1^2 = \beta_1^3 \\ \beta_2^1 &= \beta_2^2 = \beta_2^3 \\ \beta_3^1 &= \beta_3^2 = \beta_3^3 \end{aligned}$$

4. Construyendo el estadístico se tiene:

$$F = \frac{\frac{779,7495 - 748,5705}{4}}{\frac{748,5705}{1268 - 12}} = 13,0785$$

5. Al comparar el estadístico calculado obtenido en el paso anterior con el valor de la distribución entregado por la tabla (3,26) se detecta que nos

hallamos en la zona de rechazo, por lo que es posible aseverar que con un 95 % de confianza existe un cambio estructural en los parámetros, dado que la muestra no se comporta de la misma forma según el nivel de estudios que tenga el trabajador.

En este caso, dado la forma en que se estableció el test, se sabe que los parámetros no deberían ser iguales entre los tres grupos. Lo que no se sabe, es si éstos sí podrían ser iguales entre dos grupos. ¿Cómo se podría probar esta hipótesis?.

5.3. Variables Dicótomicas

Existen muchas variables explicativas que son discontinuas o en la práctica no pueden ser medidas en forma continua. Estas variables generalmente se les conoce como variables cualitativas, dicotómicas o variables dummy. Por ejemplo, en la estimación de la demanda por un producto puede ser relevante si el consumidor es hombre o mujer. También puede ser de importancia el nivel educacional que esta persona obtuvo a través de su vida (educación primaria, secundaria o superior), etc. De esta forma podemos dividir la muestra en dos partes, una de ellas contiene a todas las observaciones que presentan el atributo de relevancia y la otra parte contiene a todos los que no lo presentan. Así podemos definir una variable dicotómica de la siguiente forma:

$$D = \begin{cases} 1 & \text{Si el individuo presenta el atributo} \\ 0 & \text{en otro caso} \end{cases}$$

El uso de las variables cualitativas puede afectar de tres maneras la estimación del modelo. Primero, puede afectar el intercepto, es decir, el hecho que una muestra presente un determinado atributo afecta la posición de la curva. Segundo, puede afectar la pendiente y por último, puede afectar tanto la pendiente como el parámetro de posición.

Consideremos el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \mu_t$$

y analicemos cada caso por separado:

1. Sólo cambia el parámetro de posición

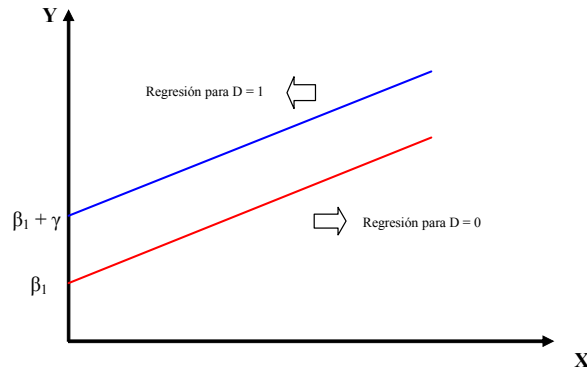


Figura 5.3: Cambio en parámetro de posición

En este caso se define una variable D que se comporta de la siguiente manera:

$$D = \begin{cases} 1 & \text{si posee el atributo} \\ 0 & \text{si no posee el atributo} \end{cases}$$

entonces la función de regresión queda definida de la siguiente manera:

$$Y_t = \beta_1 + \gamma D + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \mu_t$$

Ahora, según el comportamiento de D tenemos dos funciones de regresión, dependiendo del valor que tome la nueva variable dicotómica:

$$\begin{aligned} D = 1 &\Rightarrow Y_t = (\beta_1 + \gamma) + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \mu_t \\ D = 0 &\Rightarrow Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \mu_t \end{aligned}$$

La introducción de la variable muda hace que el intercepto de la regresión cambie para aquellas personas que poseen el atributo. Ahora el intercepto se compone de dos partes: β_1 y γ . En cambio, para las personas que no poseen el atributo, el intercepto sigue siendo sólo β_1 . De esta forma, la variable muda permite diferenciar el comportamiento entre los distintos grupos de observaciones (personas). En el caso que γ resulta ser positivo, la regresión puede graficarse como en la figura 5.3

2. Sólo cambia la pendiente

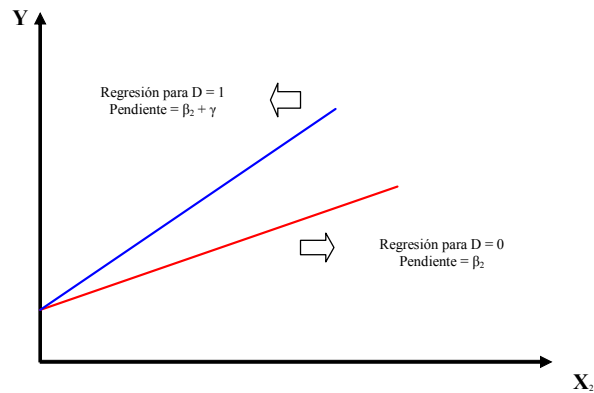


Figura 5.4: Cambio de Pendiente

Suponga ahora que lo que cambia es el impacto que tiene la variable explicativa X_2 sobre la variable dependiente entre distintos grupos de individuos.

En este caso también se define una variable D que se comporta de la siguiente manera:

$$D = \begin{cases} 1 & \text{si posee el atributo} \\ 0 & \text{si no posee el atributo} \end{cases}$$

entonces la función de regresión queda definida de la siguiente manera:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \gamma D X_{2t} + \dots + \beta_k X_{kt} + \mu_t$$

Ahora, según el comportamiento de D tenemos dos funciones de regresión:

$$\begin{aligned} D = 1 & \text{ entonces } Y_t = \beta_1 + (\beta_2 + \gamma) X_{2t} + \dots + \beta_k X_{kt} + \mu_t \\ D = 0 & \text{ entonces } Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \mu_t \end{aligned}$$

En este caso lo que cambia es el coeficiente de pendiente para X_2 entre distintos grupos de observaciones. Para un grupo el coeficiente es $\beta_2 + \gamma$, y para el otro sólo β_2 . Esto se muestra en la figura 5.4 cuando $\gamma > 0$.

3. Cambia tanto la posición como la pendiente

En este caso definimos una variable D que se comporta de la siguiente manera:

$$D = \begin{cases} 1 & \text{si posee el atributo} \\ 0 & \text{si no posee el atributo} \end{cases}$$

entonces, la función de regresión queda definida de la siguiente forma:

$$Y_t = \beta_1 + \gamma_1 D + \beta_2 X_{2t} + \gamma_2 D X_{2t} + \dots + \beta_k X_{kt} + \mu_t$$

Entonces, según el comportamiento de D tenemos dos funciones de regresión:

$$\begin{aligned} D = 1 & \text{ entonces } Y_t = (\beta_1 + \gamma_1) + (\beta_2 + \gamma_2) X_{2t} + \dots + \beta_k X_{kt} + \mu_t \\ D = 0 & \text{ entonces } Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t \end{aligned}$$

Como vemos, en este caso se diferencian los grupos tanto por el coeficiente de posición, como por el coeficiente de pendiente para la variable X_2 .

Como se ha hecho en casos anteriores, mostraremos dos ejemplos en los que se ha introducido variables dummies. Pero antes de pasar a ellos, es preciso tener presente que es posible la utilización de este tipo de variables para caracterizar atributos múltiples. Considere, por ejemplo, una situación en la que se tienen tres atributos: a , b y c (y no dos como en las referencias anteriores). Asumimos que estos atributos son excluyentes, pero que todas las observaciones deben tener uno de ellos. Por ejemplo, para una muestra de individuos se pide su estado civil, y se clasifican todos los individuos en tres categorías: soltero, casado, u otro (divorciado, viudo, separado, etc.). Dado que las variables dicotómicas implican sólo dos características, en este caso será necesario la introducción de dos variables dummies, siendo:

$$D_1 = \begin{cases} 1 & \text{si posee el atributo } a \\ 0 & \text{si no lo posee} \end{cases}$$

y

$$D_2 = \begin{cases} 1 & \text{si posee el atributo } b \\ 0 & \text{si no lo posee} \end{cases}$$

Luego, se introducen ambas variables según lo que se quiera evaluar a partir de los casos (1), (2) y (3) descritos recientemente. Fíjese que en este caso, si un individuo presenta la siguiente situación, $D_1 = 0$, $D_2 = 0$, entonces significa que posee el atributo c . Queda claro, entonces, que la regla de introducción de variables cualitativas es que si se tienen " n " atributos entonces se deben introducir $(n - 1)$ variables dummies, y que el n -ésimo atributo queda como "base" de comparación..

Ejemplo 1. Función de Consumo

En la sección anterior se evaluó si la Función de Consumo para el caso chileno presentaba evidencia suficiente como para esperar un cambio estructural en los años posteriores a 1982 respecto del período anterior al mismo. Las variables cualitativas permiten el uso de un método alternativo al Test de Chow. Este método implica introducir una dummy de la siguiente forma:

$$D_1 = \begin{cases} 1 & \text{si es el periodo 1960 - 1982} \\ 0 & \text{en otro periodo} \end{cases}$$

Luego, se procede a introducir la variable D_1 de manera tal que afecte la pendiente y el intercepto. Finalmente la ecuación a estimar para el total de la muestra es:

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 D_1 + \beta_3 D_1 \times Y_t + \mu_t$$

obteniéndose

$$\begin{aligned} C_t &= 334518,3 + 0,6085Y_t - 171102,5D_1 + 0,0615D_1 \times Y_t + \varepsilon_t \\ d.s. &= 87575,67 \quad 0,0171 \quad 163836,2 \quad 0,0545 \\ t &= 3,8198 \quad 35,6698 \quad -1,0444 \quad 1,1282 \\ R^2 &= 0,9875 \\ \bar{R}^2 &= 0,9863 \\ F &= 891,4355 \end{aligned}$$

Cabe señalar que los parámetros estimados para cada período son idénticos a los obtenidos con la prueba de Chow previamente en la sección 5.2. Si calculamos el intercepto para el período 1960-1982 de los resultados con variables mudas obtenemos $\hat{\beta}_0 + \hat{\beta}_2 = 163415,8$. El coeficiente de pendiente para el mismo período es $\hat{\beta}_1 + \hat{\beta}_3 = 0,6700$. Ambas estimaciones son las mismas que obtuvimos previamente para el primer período. Y los resultados obtenidos cuando $D_1 = 0$, corresponden exactamente a los obtenidos previamente para el segundo período (1983-1997). Cabe tener presente que la estimación anterior afecta tanto la pendiente como el intercepto, hecho que es justamente el que se quiere evaluar para encontrar si hay evidencia de cambio estructural. Si analizamos el estadístico t es sencillo darse cuenta que la variable D_1 no ha resultado significativa en ninguno de los casos. Por lo tanto, no es posible explicar los cambios experimentados por el consumo

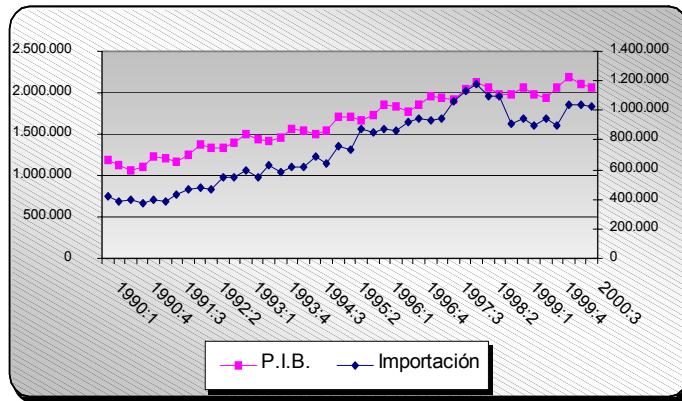


Figura 5.5: Series de Importación y P.I.B.

a través del comportamiento de la variable introducida, ya sea afectando el intercepto o en el caso en que interactúa con la variable ingreso lo cual afecta la pendiente. Estos resultados nos indican entonces, de manera análoga a los resultados hallados con el Test de Chow, que no se verifica un cambio de comportamiento antes y después del año considerado.

Ejemplo 2. Función de Importaciones

Los datos de la Función de Importaciones en relación con el P.I.B. utilizados para la estimación efectuada en el capítulo 3 se encuentran expresados en la figura 5.5. En ella es posible visualizar un quiebre en la relación Importación - P.I.B. en el período comprendido entre el primer trimestre de 1998 y el último trimestre del siguiente año. Este quiebre claramente está reflejando los efectos que la Crisis Asiática impuso sobre nuestra economía. Dada esta situación, haremos uso de una variable dummy que permita representar este período interactuando con el Producto Interno Bruto, medida económica que experimenta un cambio en relación a los niveles de importación.

Sea, entonces, la siguiente variable cualitativa

$$D_1 = \begin{cases} 1 & \text{si es el periodo } 1998 : 3 - 1999 : 4 \\ 0 & \text{en otro caso} \end{cases}$$

La ecuación a estimar se reduce a lo siguiente:

$$\ln(M_t) = \beta_0 + \beta_1 \ln(PIB_t) + \beta_2 TCR_{t-1} + \beta_3 D_1 \times \ln(PIB_t) + \mu_t$$

Los resultados de la estimación se presentan a continuación:

$$\begin{aligned}
 \ln(M_t) &= 1,4735 + 0,9404 \ln(PIB_t) - 0,0157TCR_{t-1} - 0,0082D_1 \ln(PIB_t) \\
 d.s. &= 2,2504 \quad 0,1407 \quad 0,0028 \quad 0,0021 \\
 t &= 0,6548 \quad 6,6850 \quad - 5,6841 \quad - 3,9209 \\
 R^2 &= 0,9728 \\
 \bar{R}^2 &= 0,9706 \\
 dw &= 1,8136 \\
 F &= 452,2442
 \end{aligned}$$

Los resultados nos indican que la introducción de la variable D_1 que permite ajustar el quiebre en la relación importación - P.I.B. es significativa pudiendo explicar, en el 95 % de los casos, el comportamiento de las importaciones. La relación de esta nueva variable con la explicada es negativa, resultado que indica que una de las consecuencias que tuvo la crisis asiática en nuestro país fue cambiar la relación entre producto e importaciones. Cambios en el producto tuvieron un efecto más reducido sobre las importaciones en este periodo. De esta forma, se tienen dos funciones de regresión dependiendo del comportamiento de la variable dummy. Como se explicó, la forma en que ha sido incorporada a la estimación la variable D_1 afecta a la pendiente, teniéndose para

$$\begin{aligned}
 D = 1 \quad \ln(M_t) &= 1,4735 + (0,9404 - 0,0082) \ln(PIB_t) - 0,0157TCR_{t-1} \\
 D = 0 \quad \ln(M_t) &= 1,4735 + 0,9404 \ln(PIB_t) - 0,0157TCR_{t-1}
 \end{aligned}$$

Para los años descritos por D_1 la pendiente de la Función de Regresión Muestral disminuye en el coeficiente estimado para esta variable. Si comparamos los valores del \bar{R}^2 y del dw obtenidos en la anterior estimación con los hallados en capítulos anteriores que no se les había introducido variables cualitativas, observaremos que ambas medidas estadísticas han mejorado.

Apéndice A

Ejercicios Complementarios

En este apéndice entregamos al lector ejercicios complementarios relacionados a los temas de estimación e inferencia discutidos en el Capítulo 3. El apéndice contiene los siguientes elementos:

1. Análisis de estimaciones y resultados para dos funciones adicionales a las presentadas en el Capítulo 3: la Función de Precios y la Función de Producción.
2. Todas las bases de datos usadas en el libro se entregan en un diskette. Por lo tanto, es factible utilizarlas para replicar los resultados presentados. Por esta razón entregamos una serie de Instrucciones del Programa E-Views.
3. Series de datos de los ejemplos desarrollados.

A.1. Estimación de Función de Precios y de Producción

A.1.1. Función de Precios

Nuestro modelo teórico asume que el incremento porcentual en los precios depende del crecimiento en el índice de remuneraciones y del incremento porcentual en el tipo de cambio nominal. Este se basa en un enfoque de

fijación de precios por costos. Es decir,

$$\frac{(P_t - P_{t-1})}{P_{t-1}} = \beta_0 + \beta_1 \frac{(W_t - W_{t-1})}{W_{t-1}} + \beta_2 \frac{(E_t - E_{t-1})}{E_{t-1}} + \mu_t$$

donde P es el índice de precios, W es el índice de remuneraciones y E es el valor del tipo de cambio nominal. Todas las variables se presentan en tasas de cambio porcentual.

Las razones que explican este modelo son las siguientes:

- El incremento en los salarios afecta los costos unitarios de la empresa, y la empresa en la medida que intenta mantener sus márgenes de utilidad, ajusta los precios al alza.
- El tipo de cambio nominal también afecta los costos unitarios de las empresas, ya sea por el efecto directo sobre los precios de los insumos importados, ya sea a través de un efecto indirecto por medio del impacto sobre insumos domésticos que utilizan insumos importados para su producción. El efecto sobre los precios es positivo.

Ahora que hemos descrito el marco conceptual de la función de precios podemos analizar la estimación realizada en base a los datos que se encuentran en el archivo *precios.xls* para el período comprendido entre el primer trimestre de 1980 y el mismo trimestre del año 2000. Los resultados obtenidos son los siguientes:

$$\begin{aligned} \frac{(P_t - P_{t-1})}{P_{t-1}} &= 0,0079 + 0,4916 \frac{(W_t - W_{t-1})}{W_{t-1}} + 0,2258 \frac{(E_t - E_{t-1})}{E_{t-1}} + \varepsilon_t \\ &\quad (0,0038) \quad (0,0745) \quad (0,0293) \\ t &= 2,0587 \quad 6,5956 \quad 7,6982 \\ R^2 &= 0,5607 \\ \overline{R^2} &= 0,5493 \\ dw &= 2,0117 \\ n &= 80 \end{aligned}$$

Cabe mencionar que en este caso el parámetro estimado tiene la interpretación de una elasticidad. Por ejemplo, para el caso de $\hat{\beta}_1$ un incremento de un 10% en la tasa de crecimiento del índice de remuneraciones aumentará el nivel de precios en un 4,9%.

A.1. ESTIMACIÓN DE FUNCIÓN DE PRECIOS Y DE PRODUCCIÓN 185

A partir del $\overline{R^2}$ podemos decir que la estimación de nuestra Función de Precios explica en un 55% al modelo poblacional. El estadístico de Durbin Watson nos indica presencia de no autocorrelación en los errores.

Parte 1. Prueba t Individual

Con esta prueba podremos verificar si los parámetros encontrados son compatibles con lo sugerido por la teoría económica al probar su significancia. Dentro de los resultados presentados en el esquema anterior, encontramos el valor del estadístico t resultante de aplicar esta Prueba de Hipótesis. Recuerde que este test nos permite validar el signo que acompaña a cada parámetro indicando si es posible explicar los cambios experimentados por la variable dependiente a través de las variaciones producidas en las variables explicativas.

Al llevar a cabo esta prueba, se plantea para el segundo parámetro, $\hat{\beta}_1$, la siguiente hipótesis nula:

$$H_0 : \beta_1 = 0$$

Como se explicó en el capítulo 3, se calcula el estadístico t como se muestra a continuación:

$$t^c = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}}$$

Desarrollando la ecuación anterior se obtiene un valor igual a 6.5956 para el t^c , que al ser comparado con el de tabla vemos que se encuentra en la zona de rechazo. Esto significa que no se acepta la hipótesis nula, lo cual conduce a probar la significancia de $\hat{\beta}_1$.

En forma análoga se obtuvieron los valores del estadístico t para los restantes parámetros. Dado que éste estadístico es mayor que 2, es decir que el t de tabla, se rechaza la hipótesis nula en todos los casos analizados. Con estos resultados sabemos entonces que los estimadores permiten explicar las variaciones que experimenta el índice de precios, lo cual resulta compatible con la teoría económica, ya que la regresión efectuada a través de MCO satisface la relación existente entre las variables explicativas y la variable dependiente.

Parte 2. Intervalo de Confianza

Un método alternativo de llevar a cabo las Pruebas de Hipótesis, consiste en construir un intervalo de confianza para el parámetro, aunque el resultado de la Prueba de Significancia es el mismo, los intervalos de confianza nos entregan más información que el procedimiento anterior, ya que nos indica entre qué rangos puede variar un parámetro. Para ejemplificar este caso buscaremos el intervalo de confianza para $\hat{\beta}_2$ y verificaremos si se llega a la misma conclusión que con la Prueba t Individual.

Planteando el intervalo para $\hat{\beta}_2$:

$$\Pr(\hat{\beta}_2 - t_{\alpha/2} \cdot S_{\hat{\beta}_2} \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2} \cdot S_{\hat{\beta}_2}) = 1 - \alpha$$

Si elegimos $\alpha = 0,05$ (95 % de confianza), obtenemos

$$(0,2258 - 1,96 \cdot 0,0293 \leq \beta_2 \leq 0,2258 + 1,96 \cdot 0,0293)$$

$$(0,1684 \leq \beta_2 \leq 0,2832)$$

Es posible apreciar de los resultados que existe un 95 % de posibilidades que β_2 tome algún valor entre 0.1684 y 0.2832. Tal como se planteara al comienzo de esta parte, el intervalo de confianza nos guía a la misma conclusión que la Prueba t Individual, ya que, se observa que $\beta_2 = 0$ cae fuera del intervalo. Además, el intervalo de confianza nos indica entre qué valores puede hallarse con mayor probabilidad el valor poblacional, información que no es posible obtener a través del Test de Significancia Individual.

Parte 3. Prueba T Generalizada

A continuación intentaremos probar la existencia de homogeneidad de grado 1 en salarios y tipo de cambio. Es decir, si incrementos en los costos se traspasan completamente a precios. La hipótesis se plantea de la siguiente manera:

$$H_0 : \beta_1 + \beta_2 = 1$$

Podemos establecer t como

$$\mathbf{t} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

A.1. ESTIMACIÓN DE FUNCIÓN DE PRECIOS Y DE PRODUCCIÓN 187

De esta forma, la prueba de hipótesis puede expresarse como:

$$H_0 : \mathbf{t}'\boldsymbol{\beta} = 1$$

$$H_1 : \mathbf{t}'\boldsymbol{\beta} \neq 1$$

El estadístico t se encuentra al aplicar la siguiente ecuación:

$$t^c = \frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\text{VAR}(\mathbf{t}'\boldsymbol{\beta})}} = \frac{\mathbf{t}'\hat{\boldsymbol{\beta}} - \mathbf{t}'\boldsymbol{\beta}}{\sqrt{\sigma_\mu^2 \mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}}}$$

Por lo tanto requerimos de la siguiente información:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0,0079 \\ 0,4916 \\ 0,2258 \end{bmatrix} \quad \text{VAR}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 1,46 \cdot 10^{-5} & -0,000238 & -3,39 \cdot 10^{-5} \\ -0,000238 & 0,005556 & 0,000101 \\ -3,39 \cdot 10^{-5} & 0,000101 & 0,000860 \end{bmatrix}$$

Así, podemos obtener:

$$\mathbf{t}'\hat{\boldsymbol{\beta}} = 0,7174$$

$$\text{var}(\mathbf{t}'\boldsymbol{\beta}) = \mathbf{t}'\text{var}(\hat{\boldsymbol{\beta}})\mathbf{t} = 0,006618$$

Reemplazando:

$$t^c = \frac{0,7174 - 1}{\sqrt{0,006618}} = -3,4738$$

Dado que el estadístico t se encuentra en la zona de rechazo al exceder al de tabla (1.99 para un 5% de significancia), rechazamos la hipótesis nula planteada.

Parte 4. Prueba de Significancia Global

En este apartado intentaremos probar si la regresión de la función de precios es significativa en conjunto. Para ello planteamos como hipótesis nula que todos los parámetros relevantes del punto de vista teórico son iguales a cero. Si aceptamos esta hipótesis, entonces estaremos en condiciones de decir que la regresión en conjunto no explica los cambios experimentados en el

nivel de precios. Por lo tanto, lo que esperamos hallar con esta prueba es el rechazo de la hipótesis nula.

Las pruebas de hipótesis son las siguientes:

$$H_0 : \beta_1 = \beta_2 = 0$$

Calculando el estadístico F :

$$F^c = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,5607/(3-1)}{(1-0,5607)/(80-3)} = 49,14$$

Es posible concluir al comparar con el F de tabla (3,11), que este último es menor al calculado, lo cual indica que nos encontramos en la zona de rechazo. Así, es posible decir que la regresión encontrada para la Función de Precios sí explica el comportamiento que experimenta nuestra variable dependiente.

Parte 5. Predicción Individual

Para esta parte del análisis buscaremos qué valor adopta la variable dependiente, es decir, el nivel de precios, cuando las variables independientes o exógenas toman un valor específico.

De esta manera, podemos plantear una situación en la que la tasa de crecimiento de las remuneraciones alcanza un nivel igual a 4% y el cambio porcentual en el tipo de cambio nominal es de 3% sobre base anual. Estas tasas son las que se esperan “normalmente” para estos indicadores.

El intervalo de confianza que define un valor específico para el nivel de precios es:

$$P_0 \in \left(\hat{P}_0 \pm t_{\alpha/2} \times \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_0 + 1]} \right)$$

Siendo:

$$\hat{P}_0 = \mathbf{X}_0\hat{\beta}$$

$$\hat{P}_0 = [1 \quad 4 \quad 3] \begin{bmatrix} 0,0079 \\ 0,4916 \\ 0,2258 \end{bmatrix} = 2,6517$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0,0529 & -0,8623 & -0,1228 \\ -0,8623 & 20,1295 & 0,3659 \\ -0,1228 & 0,3659 & 3,1158 \end{bmatrix} \quad \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_0 = 351,3135$$

$$\hat{\sigma}_\mu^2 = 2,75974 \cdot 10^{-4}$$

Resolviendo para un 95 % de confianza:

$$P_0 \in \left(2,6517 \pm 1,96 \times \sqrt{2,75974 \cdot 10^{-4} [351,3135 + 1]} \right)$$

$$(2,04054 \leq P_0 \leq 3,26286)$$

Es posible afirmar con un 95 % de confianza que, dado el escenario planteado en el enunciado, la variación en el nivel de precios se ubicará entre 2.04 % y 3.26 %, anual, aproximadamente.

Parte 6. Predicción Media

Llegando al final de nuestro desarrollo de la Función de Precios, llevaremos a cabo la predicción media del nivel de precios dada la misma situación anterior para los parámetros. Así, el predictor es \hat{P}_0 igual a 1.3298 y su intervalo de confianza para la Predicción Media está dado por:

$$E(P_0/X_0) \in \left(\hat{P}_0 \pm t_{\alpha/2} \times \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_0]} \right)$$

Reemplazando:

$$E(P_0/X_0) \in \left(2,6517 \pm 1,96 \times \sqrt{2,75974 \cdot 10^{-4} \times 351,3135} \right)$$

$$(2,04141 \leq E(P_0/X_0) \leq 3,26199)$$

De esta forma, el valor esperado en la tasa de crecimiento del nivel de precios se encuentra entre 2.04141 y 3.26199 con un 95 % de confianza.

A.1.2. Función de Producción

La Función de Producción que se utiliza asume la presencia de dos factores productivos, capital y trabajo. La forma funcional que adoptaremos será del tipo Cobb-Douglas. Para hacer posible la regresión con Mínimos Cuadrados Ordinarios hemos procedido a linealizarla con logaritmos, quedando con la forma que se observa más abajo.

A continuación se detallan los resultados obtenidos para esta función partiendo de las series contenidas en el archivo *producción.xls* para el período trimestral 1987:1 - 1993:4.

$$\begin{aligned}
 \ln(PIB_t) &= -1,7945 + 0,0826 \ln(capital_t) + 1,5259 \ln(trabajo_t) + \varepsilon_t \\
 &\quad (1,6639) \quad (0,1691) \quad (0,3575) \\
 t &= -1,0785 \quad 0,4882 \quad 4,2678 \\
 R^2 &= 0,9347 \\
 \overline{R^2} &= 0,9294 \\
 dw &= 0,7239 \\
 n &= 28
 \end{aligned}$$

El primer término, $\hat{\beta}_0$, representa el efecto medio que tienen otras variables sobre el producto, las cuales no han sido consideradas en forma explícita en el modelo.

Los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$ corresponden a la elasticidad parcial del capital (ξ_K) y a la elasticidad parcial del trabajo (ξ_L), respectivamente y son iguales a 0,0826 y 1,5259. Al observar los coeficientes de determinación verificamos que la estimación llevada a cabo explica en un 93% los cambios en el producto. El estadístico de Durbin y Watson indica la violación de uno de los supuestos del modelo clásico: la ausencia de autocorrelación en los residuos. En este caso, se tiene correlación serial positiva de primer orden en los residuos. Dada esta situación, no tiene validez realizar inferencia sobre los parámetros estimados por MCO, pero a pesar de ello, y teniendo presente que debe hacerse una transformación de las variables para eliminar la autocorrelación, continuaremos con el procedimiento efectuado en la función anteriormente estimada, esta vez para la Función de Producción.

Parte 1. Prueba t Individual

La Prueba de Significancia Individual arroja diferentes resultados para cada uno de los estimadores. Si se considera un t de tabla igual a 2.048¹, se rechaza H_0 al evaluar si el coeficiente que acompaña la variable trabajo es igual a cero, lo cual prueba su significancia. En el caso del capital, el estadístico t se encuentra en la zona de aceptación lo que nos indica que el capital no ayuda a explicar el comportamiento del producto.

Para ejemplificar, como lo hemos hecho para las demás funciones presentadas tanto en el capítulo 3 como en este apartado, desarrollaremos la Prueba t para $\hat{\beta}_1$.

Planteamos el estadístico t como:

$$t^c = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}}$$

Siendo la Prueba de Hipótesis como se presenta:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

Reemplazando se tiene:

$$t^c = \frac{0,0826 - 0}{0,1691} = 0,488$$

Efectivamente, al comparar el resultado obtenido en nuestra Prueba t con el de tabla, aceptamos H_0 al encontrarnos en la zona de aceptación.

Análogamente para $\hat{\beta}_2$ se tiene:

$$t^c = \frac{\hat{\beta}_2 - \beta_2}{S_{\hat{\beta}_2}}$$

La hipótesis nula es:

$$\begin{aligned} H_0 &: \beta_2 = 0 \\ H_1 &: \beta_2 \neq 0 \end{aligned}$$

¹Valor de tabla para un tamaño muestral de 28 observaciones y un nivel de significancia de 0.05.

Finalmente,

$$t^c = \frac{1,5259 - 0}{0,3575} = 4,268$$

Dado que $t^c > t^t$ se tiene que $\hat{\beta}_2$ es un estimador significativo.

Parte 2. Intervalo de Confianza

Ya que nos encontramos en un escenario en el que uno de los parámetros encontrados no es significativo buscaremos su intervalo de confianza.

Procedemos planteando el intervalo de confianza para $\hat{\beta}_1$:

$$\left(\hat{\beta}_1 - t_{\alpha/2} \cdot S_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2} \cdot S_{\hat{\beta}_1} \right)$$

Asumiendo un $\alpha = 0,05$ y reemplazando la información obtenemos:

$$(0,0826 - 2,048 \cdot 0,1691 \leq \beta_1 \leq 0,0826 + 2,048 \cdot 0,1691)$$

$$(-0,2637 \leq \beta_1 \leq 0,4289)$$

Efectivamente, es sencillo verificar que β_1 tiene una alta probabilidad de tomar el valor cero, lo que nos conduce a la misma conclusión que con la Prueba t Individual.

Llevando a cabo el mismo procedimiento para $\hat{\beta}_2$, el intervalo de confianza se plantea como:

$$\left(\hat{\beta}_2 - t_{\alpha/2} \cdot S_{\hat{\beta}_2} \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2} \cdot S_{\hat{\beta}_2} \right)$$

Reemplazando se tiene,

$$(1,5259 - 2,048 \cdot 0,3575 \leq \beta_2 \leq 1,5259 + 2,048 \cdot 0,3575)$$

$$0,7937 \leq \beta_2 \leq 2,2581$$

Se observa que el valor cero queda fuera del intervalo de confianza, teniendo $\hat{\beta}_2$ altas probabilidades de ser significativo, conclusión que también se deduce de la Prueba t Individual.

Parte 3. Prueba T Generalizada

Una prueba interesante desde la perspectiva económica es comprobar si la función posee rendimientos constantes a escala, es decir $\beta_1 + \beta_2 = 1$. Esta prueba es fácil de verificar usando la Prueba T Generalizada. Para poder llevarla a cabo, nuestra matriz \mathbf{t} será un vector de constantes iguales a 1 con excepción del primer elemento que tomará el valor cero dado que representa la constante en nuestra función, quedando de la siguiente manera:

$$\mathbf{t} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Recuerde que la forma funcional original para medir el nivel de producto es:

$$PIB = AK^{\beta_1}L^{\beta_2}$$

Por lo tanto, la verificación de rendimientos constantes a escala equivale a estudiar el grado de homogeneidad r de la función, el cual también se interpreta como el rendimiento que presenta la función de producción. En este caso, se tiene que para que la función presente este tipo de rendimientos debe suceder que la suma de sus exponentes sea igual a 1. De esta forma tendremos que:

$$\mathbf{t}'\boldsymbol{\beta} = [0 \quad 1 \quad 1] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = (\beta_1 + \beta_2) = 1$$

Para que la Función de Producción presente rendimientos constantes a escala, la suma de sus exponentes debe ser igual a uno.

Por consiguiente, el problema se resume en la siguiente prueba de hipótesis:

$$\begin{aligned} H_0 &: \mathbf{t}'\boldsymbol{\beta} = 1 \\ H_1 &: \mathbf{t}'\boldsymbol{\beta} \neq 1 \end{aligned}$$

Las demás matrices necesarias para resolver este problema son:

$$\hat{\beta} = \begin{bmatrix} -1,7945 \\ 0,0826 \\ 1,5259 \end{bmatrix} \quad \text{VAR}(\hat{\beta}) = \begin{bmatrix} 2,7685 & 0,2510 & -0,5811 \\ 0,2510 & 0,0286 & -0,0585 \\ -0,5811 & -0,0585 & 0,1278 \end{bmatrix}$$

El estadístico t relevante para esta prueba es:

$$t^c = \frac{\mathbf{t}'\hat{\beta} - \mathbf{t}'\beta}{\sqrt{\text{var}(\mathbf{t}'\beta)}}$$

Siendo:

$$\text{var}(\mathbf{t}'\beta) = \mathbf{t}'\text{var}(\hat{\beta})\mathbf{t} = 0,0394$$

Resolviendo:

$$t^c = \frac{1,6085 - 1}{\sqrt{0,0394}} = 3,066$$

En conclusión, se rechaza la hipótesis nula dado que el estadístico encontrado supera al de tabla, lo cual indica que nuestra función de producción no presenta rendimientos constantes a escala. Por el contrario, si sumamos los exponentes de nuestra función, el resultado será mayor a la unidad, lo que nos indica la presencia de rendimientos crecientes a escala.

Además, cabe agregar que los resultados indican que el factor capital no ayuda a explicar el producto. Estos resultados pueden deberse a varios factores:

- a) Que la teoría económica utilizada para explicar la producción no es la adecuada.
- b) Que la forma de la función de producción utilizada (Cobb-Douglas) no sea la más apropiada.
- c) Que la medición de las variables utilizadas es defectuosa.
- d) Que efectivamente el capital no ayuda a explicar la producción en Chile en el período muestral.

Antes de inclinarse por la última opción, es necesario investigar las alternativas anteriores. Realizar esto escapa, naturalmente, al ámbito de un libro de Econometría.

Parte 4. Prueba de Significancia Global

Esta prueba nos permitirá concluir si realmente la función encontrada explica en su conjunto los cambios que experimenta el PIB, a pesar que existan parámetros que resultaron no ser significativos como ya analizamos.

Procedemos, entonces, a plantear la hipótesis nula:

$$H_0 : \beta_1 = \beta_2 = 0$$

El estadístico F es:

$$F^c = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,9347/(3-1)}{(1-0,9347)/(28-3)} = 178,924$$

Efectivamente, ocurre que a pesar de que no todos los estimadores han resultado significativos al ser evaluados individualmente, la regresión en conjunto sí resulta serlo. Es posible llegar a esta conclusión al verificar que el F^c es mayor al F de tabla ($F^t = 3,4$), lo que ubica al estadístico en la zona de rechazo. Esto es justamente lo que se busca, ya que si esta hipótesis es rechazada nos indica con un 95% de posibilidades que los parámetros no serán iguales a cero simultáneamente, pudiendo de esta manera explicar el comportamiento del PIB a través de los cambios que ellos experimenten.

Parte 5. Predicción Individual

Para hallar el valor específico de la producción, dado un escenario determinado de las variables independientes, debemos hacer uso del intervalo de confianza para la predicción individual.

El escenario que nos impondremos para las variables exógenas son los que se presentan a continuación:

$$K = 6000$$

$$L = 5200$$

Es decir, conocer el rango en que se encontraría la producción nacional si se tuviera una acumulación de capital valorada en 6000 (expresado en unidades de miles de millones de pesos de 1986) y la fuerza de trabajo de 5200 personas (expresada en unidades de mil).

Recordando el intervalo de confianza:

$$\ln PIB_0 \in \left(\ln \widehat{PIB}_0 \pm t_{\alpha/2} \times \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0 + 1]} \right)$$

Siendo:

$$\ln \widehat{PIB}_0 = \mathbf{X}_0 \hat{\boldsymbol{\beta}}$$

$$\ln \widehat{PIB}_0 = \begin{bmatrix} 1 & \ln 6000 & \ln 5200 \end{bmatrix} \begin{bmatrix} -1,7945 \\ 0,0826 \\ 1,5259 \end{bmatrix} = 11,980$$

$$\hat{\sigma}_\mu^2 = 0,0014$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1977,5 & 179,286 & -415,071 \\ 179,286 & 20,429 & -41,786 \\ -415,071 & -41,786 & 91,286 \end{bmatrix}$$

Reemplazando y resolviendo para un 95 % de confianza se obtiene que:

$$\ln PIB_0 \in \left(11,980 \pm 2,048 \times \sqrt{0,0014 [2,39336 + 1]} \right)$$

$$(11,839 \leq \ln PIB_0 \leq 12,121)$$

Estos valores se encuentran en forma exponencial. Aplicando exponencial se tiene

$$138550 \leq PIB_0 \leq 183510$$

De esta forma, si las variables exógenas alcanzaran el nivel planteado, el PIB se ubicaría entre 138550 y 183510 millones de pesos.

Parte 6. Predicción Media

Finalizando el análisis de la Función de Producción, calcularemos el valor esperado del nivel de producto cuando se presenta la misma situación planteada para la predicción individual.

El intervalo de confianza relevante para resolver este problema es:

$$E(\ln PIB_0/X_0) \in \left(\ln \widehat{PIB}_0 \pm t_{\alpha/2} \times \sqrt{\hat{\sigma}_\mu^2 [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0]} \right)$$

*A.1. ESTIMACIÓN DE FUNCIÓN DE PRECIOS Y DE PRODUCCIÓN*197

Resolviendo se llega a que:

$$(11,861 \leq E(\ln PIB_0/X_0) \leq 12,099)$$

Finalmente,

$$141630 \leq E(PIB_0/X_0) \leq 179690$$

Se observa que, con un 95% de posibilidades y un escenario como el establecido al inicio a las variables, el valor medio del producto se encontrará en el intervalo encontrado.

A.2. Instrucciones para el programa E-Views

Las estimaciones de las distintas bases de datos que aparecen en este texto se han llevado a cabo con el software econométrico **E-Views**. Si usted mismo desea efectuar las regresiones deberá seguir los siguientes pasos:

- Al abrir E-Views deberá crear un nuevo **workfile**. Para ello tendrá que acceder al menú **File** en la barra de menú, luego elegir **New** y finalmente **Workfile**. Se desplegará un cuadro de diálogo en el que se debe especificar la frecuencia de las observaciones a estimar e ingresar tanto el inicio como término de dichas observaciones.
- Nuevamente se desplegará una ventana (Workfile) en la que aparecerán los parámetros para la constante (**c**) y para los residuos (**resid**).
- Para que los datos puedan ser leídos por E-Views es necesario importarlos desde los archivos adjuntados, que se encuentran en formato Excel. Para ello deberá presionar el botón **Procs** en la Barra de Título de la ventana de Workfile. Luego se abrirá un menú donde se elegirá el submenú **Import**, para concluir eligiendo la opción **Read Text-Lotus-Excel**.
- En este nuevo cuadro se debe especificar archivo, unidad y formato. Para nuestro caso la unidad será **A:**, el formato **Excel** y el archivo, el que contenga la información a estimar, por ejemplo *consumo.xls*.
- Se abrirá un nuevo menú en el que se debe tener cuidado que las series estén con la opción columnas. También se deberá especificar la primera celda que contiene datos. Para *consumo.xls* será **B2** y en el cuadro principal escribirá el nombre de las series separadas por un espacio (en el orden en que aparecen en la base de datos), es decir, para nuestro ejemplo, **consumo** seguido de un **espacio** seguido de **ingreso**. Una vez llevado a cabo, aparecerán en el Workfile junto al parámetro de la constante y de los residuos las series de consumo e ingreso ordenados en columnas.
- Solo ahora que tenemos los datos es posible llevar a cabo la estimación. Para ello debe elegir dentro del menú **Quick** de la barra principal la opción **Estimate Equation**.

- Para terminar, en el cuadro de diálogo debemos ingresar primero la variable dependiente seguido de la constante y finalmente las variables independientes, cada una separada de la otra por un espacio y sin comas. El método a elegir es el de Mínimos Cuadrados (**Least Squares** en inglés). Para guardar los resultados de la estimación, al cerrar la ventana en la que éstos aparecen deberá hacer clic en el botón **Name** y luego dar un nombre a la ecuación, finalmente termina haciendo clic en **ok**. Esta ecuación aparecerá junto con las demás series en el Workfile identificada por un signo igual.

Si desea obtener los resultados en forma gráfica se debe regresar a la ventana de Workfile. Luego se selecciona(n) las series que se desean graficar (con ayuda del comando CTRL en el teclado) y con el botón derecho del mouse elegimos la opción **as Group**. Una nueva ventana se abrirá donde se mostrarán las observaciones de cada serie. En esta nueva ventana haga clic en el menú **View**, luego elija la opción **Graph** y finalmente **Line**.

A.3. Tablas de Datos

*Tabla 1. Datos para estimar una Función de Consumo
(millones de pesos de 1986)*

Periodo	Consumo	Ingreso	Periodo	Consumo	Ingreso
1960	1403565	1780341	1979	2193805	3118665
1961	1484139	1851872	1980	2345662	3338607
1962	1544760	1975044	1981	2659718	3422684
1963	1609861	2058054	1982	2267392	2806587
1964	1600076	2160352	1983	2137419	2743290
1965	1598521	2200940	1984	2150843	2836445
1966	1778919	2567978	1985	2129737	2858357
1967	1839146	2573875	1986	2238746	3063564
1968	1909414	2675635	1987	2400565	3430872
1969	2009923	2853313	1988	2569303	3820065
1970	1998181	2902718	1989	2829978	4291993
1971	2261128	3159419	1990	2892007	4428160
1972	2435609	3072439	1991	3148534	4769029
1973	2275531	2968435	1992	3582720	5385467
1974	1859731	2980573	1993	3848849	5728410
1975	1647317	2268325	1994	4163544	6180585
1976	1651487	2400603	1995	4572265	7225132
1977	1915879	2616966	1996	5003503	7480542
1978	2060109	2826672	1997	5417874	8173955

Fuente:Boletín Estadístico Mensual. Banco Central

En la tabla para la estimación de la Función de Producción la serie del PIB ha sido encadenada con las tasas de crecimiento observadas a partir del año 1990 en moneda de 1986. Por su parte la serie de capital (K) se encuentra en miles de millones de pesos de 1986 y el trabajo (L) en miles de personas.

Tabla 2. Datos para estimar una Función de Producción

Periodo	PIB	K	L	Periodo	PIB	K	L
1987:1	99892	3487.4	3879.00	1990:3	108991	4519.9	4381.17
1987:2	101322	3532.1	3884.40	1990:4	114592	4611.6	4463.70
1987:3	96505	3582.6	3831.30	1991:1	125957	4674.1	4546.73
1987:4	100511	3625.5	3937.43	1991:2	124566	4741.4	4467.93
1988:1	105835	3671.6	4030.93	1991:3	119462	4814.1	4443.77
1988:2	107816	3705.0	4065.80	1991:4	129417	4922.1	4564.23
1988:3	104655	3750.7	4083.27	1992:1	142330	5006.6	4693.27
1988:4	109224	3811.5	4224.80	1992:2	137957	5097.4	4668.20
1989:1	116240	3889.5	4317.27	1992:3	137180	5221.8	4655.93
1989:2	120965	3980.5	4313.23	1992:4	143249	5376.0	4796.53
1989:3	115486	4081.3	4307.73	1993:1	154080	5541.1	4940.87
1989:4	117552	4192.2	4410.17	1993:2	149006	5697.6	4923.17
1990:1	122583	4310.3	4502.37	1993:3	146816	5822.7	4972.20
1990:2	116373	4412.0	4429.77	1993:4	149991	5945.1	5069.82

Fuente: Series de PIB y L del Boletín Estadístico Mensual del Banco Central.
Serie de capital de Lehmann, S. (1994)

*Tabla 3. Datos para estimar una Función de Importaciones
Datos de M y PIB en millones.de pesos de 1986*

Periodo	M	PIB	TCR	Periodo	M	PIB	TCR
1990:1	419740	1188378	115.86	1995:3	878448	1670564	85.95
1990:2	386714	1128171	110.36	1995:4	850732	1725135	88.28
1990:3	400124	1056616	107.29	1996:1	878918	1859798	86.91
1990:4	368772	1110906	107.78	1996:2	862678	1825502	84.26
1991:1	399807	1221084	113.77	1996:3	920462	1762695	83.92
1991:2	383750	1207608	104.91	1996:4	944680	1857145	83.91
1991:3	433689	1158126	103.31	1997:1	931708	1953561	80.48
1991:4	467629	1254629	103.73	1997:2	947008	1937209	79.07
1992:1	479432	1379825	99.32	1997:3	1064112	1916143	76.79
1992:2	467554	1337429	95.43	1997:4	1128409	2038219	76.29
1992:3	554085	1329893	100.58	1998:1	1179419	2125595	77.82
1992:4	551239	1388734	96.03	1998:2	1099344	2059100	77.58
1993:1	590818	1493715	96.79	1998:3	1102460	1980364	78.39
1993:2	550337	1444535	99.63	1998:4	910388	1987952	78.24
1993:3	624531	1423310	96.67	1999:1	943905	2066145	79
1993:4	577670	1454086	94.42	1999:2	895176	1982669	78.74
1994:1	621553	1567691	95.92	1999:3	944212	1944121	83.62
1994:2	621922	1536989	93.65	1999:4	896239	2066832	87.79
1994:3	693114	1498487	94.53	2000:1	1041253	2178427	83.12
1994:4	643892	1544443	92.90	2000:2	1036687	2102912	82.84
1995:1	760598	1707200	93.39	2000:3	1031309	2056244	88.24
1995:2	735334	1698053	88.61				

Fuente:Sitio Web Banco Central. Informe Económico y Financiero.

Tabla 4. Datos para estimar una Función de Precios

Periodo	P	W	T.C.N.	Periodo	P	W	T.C.N.
1980:1	6.04	7.54	39.00	1985:2	16.82	20.29	148.87
1980:2	6.49	8.33	39.00	1985:3	18.22	21.55	175.16
1980:3	6.91	8.79	39.00	1985:4	18.97	22.71	180.78
1980:4	7.44	9.97	39.00	1986:1	19.97	24.15	186.92
1981:1	7.76	10.38	39.00	1986:2	20.70	25.32	188.64
1981:2	7.97	10.85	39.00	1986:3	21.34	26.19	194.38
1981:3	8.15	11.72	39.00	1986:4	22.25	27.38	201.78
1981:4	8.29	12.19	39.00	1987:1	23.42	28.69	206.12
1982:1	8.35	12.28	39.00	1987:2	24.67	29.86	214.07
1982:2	8.33	12.37	40.34	1987:3	25.73	31.39	224.62
1982:3	8.83	12.22	55.01	1987:4	27.18	33.35	232.81
1982:4	9.86	12.64	69.28	1988:1	27.86	35.39	242.35
1983:1	10.30	13.14	74.97	1988:2	28.61	37.02	245.11
1983:2	10.89	13.69	75.27	1988:3	29.05	38.68	246.53
1983:3	11.56	14.42	79.68	1988:4	30.32	39.59	246.05
1983:4	12.26	15.03	85.23	1989:1	31.46	41.57	247.53
1984:1	12.47	16.11	88.05	1989:2	32.82	43.74	255.15
1984:2	13.02	16.65	89.96	1989:3	34.50	45.83	276.54
1984:3	13.45	17.01	95.17	1989:4	36.78	48.54	288.60
1984:4	15.02	17.76	120.73	1990:1	38.83	52.32	295.29
1985:1	16.05	19.61	135.50	1990:2	40.90	55.39	296.72

Periodo	P	W	T.C.N.	Periodo	P	W	T.C.N.
1990:3	43.67	58.80	302.82	1995:3	82.77	137.42	386.63
1990:4	47.46	64.11	324.78	1995:4	84.31	140.34	409.30
1991:1	48.19	69.38	338.33	1996:1	85.22	151.23	410.35
1991:2	50.62	71.74	341.71	1996:2	87.24	154.60	408.17
1991:3	53.23	73.98	351.74	1996:3	88.37	158.05	411.22
1991:4	56.04	79.45	365.08	1996:4	89.76	161.38	419.33
1992:1	57.19	84.42	355.32	1997:1	91.15	166.31	418.01
1992:2	58.64	87.42	349.29	1997:2	92.11	168.23	415.35
1992:3	60.81	90.02	368.72	1997:3	93.43	170.91	411.22
1992:4	63.54	93.55	376.98	1997:4	95.41	174.01	419.33
1993:1	64.26	98.60	389.69	1998:1	96.25	179.27	451.48
1993:2	66.25	101.14	403.16	1998:2	97.07	182.10	454.45
1993:3	68.67	104.16	406.88	1998:3	98.16	184.46	468.80
1993:4	71.56	109.77	416.94	1998:4	99.63	187.38	466.42
1994:1	72.85	116.48	429.86	1999:1	99.93	191.08	487.20
1994:2	74.67	119.02	423.28	1999:2	100.88	193.16	489.85
1994:3	76.42	122.26	418.26	1999:3	101.29	194.96	518.11
1994:4	77.79	126.02	409.30	1999:4	102.07	196.82	539.97
1995:1	79.01	131.15	409.46	2000:1	103.12	200.60	512.56
1995:2	80.47	134.62	381.70				

Fuente: Precios y Salarios del Instituto Nacional de Estadísticas.

Tipo de Cambio de Boletín Mensual del Banco Central.

Bibliografía

- [1] **Devaud, G. et. al. (1991)**. *Algebra Lineal*, Universidad de Concepción, Facultad de Ciencias Físicas y Matemáticas, Concepción.
- [2] **Gerber, H. (1992)**. *Algebra Lineal*, Grupo Editorial Iberoamericana, México D.F.
- [3] **Greene, William (1998)**. *Análisis Económico*. Prentice Hall, Madrid
- [4] **Grossman, S.I. (1996)**. *Algebra Lineal*, McGraw-Hill, México D.F.
- [5] **Gujarati, D.(1997)**. *Econometría Básica*. Mc Graw Hill, Santafé de Bogotá
- [6] **Herstein, I.N. y Winter, D. (1989)**. *Algebra Lineal y Teoría de Matrices*, Grupo Editorial Iberoamericana, México D.F.
- [7] **Hogg R., y E. Taniss (1983)**. *Probability and Statistical Inference* MacMillan second edition
- [8] **Jacob, B. (1995)**. *Linear Functions and Matrix Theory*, Springer-Verlag, New York.
- [9] **Johnston, J. (1975)**: *Métodos de Econometría*. Vicens Vives, Barcelona
- [10] **Judge G., Hill R., W. Griffiths, H. Lütkepohl and T.C.Lee. (1988)**. *Introduction to the Theory and Practice of Econometrics*, John Wiles and sons, New York.”
- [11] **Lipschutz, S. (1992)**. *Algebra Lineal*, McGraw-Hill, Madrid.

- [12] **Maddala, G. (1985).** *Econometría*, Mc Graw Hill, México, D.F.
- [13] **Maddala, G. (1996).** *Introducción a la Econometría*. Prentice Hall, México, D.F.
- [14] **Mora A., Cid L. y Valenzuela M. (1996)** *Probabilidades y Estadística* Departamento de Estadística Universidad de Concepción.