881201

2

2a

UNIVERSIDAD ANAHUAC

ESCUELA DE ACTUARIA

Con Estudios Incorporados a la

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIVERSIDAD ANAHUAC

VINCE IN BONO MALUM

"COMBINACION DE SERIES DE TIEMPO Y CORTE TRANSVERSAL"

TESIS
que para obtener el título de
A C T U A R I O
p r e s e n t a

JAVIER VARGAS DIEZ BARROSO

Mexico, D.F. 1985

INDICE

NOTACION

Esta tesis esta elaborada con un procesador de palabras americano que no tiene caracteres griegos. Estos caracteres son los usados comunmente en los libros de Econometria por lo que nos vimos en la necesidad de usar otros simbolos disponibles en sustitucion de aquellos caracteres griegos que mas se repiten (alfa y beta), mientras que el resto, menos frecuente, los escribimos directamente a maquina.

Por la misma razon la tesis carece de acentos, y escribimos sólo aquellos que sirven para distinguir palabras ambivalentes (por ejemplo, para distinguir el "sí" afirmativo del "si" condicional).

A continuacion presentamos una lista de los simbolos utilizados en la tesis y su equivalencia en el alfabeto griego o en notacion matematica:

Simbolo	Equivalencia
SUM	Sumatoria
SQ	Raiz cuadrada de una expresion
a,A	Alfa (coeficiente de regresion)
b,B	Beta (coeficiente de regresion)
u	miu (en notacion matricial)
*	Producto de Kronecker

CAPITULO 1

INTRODUCCION.

Comenzar una tesis no es cosa facil, como no lo es tomar decisiones en la vida. Desde luego hay cosas mas importantes que una tesis, pero toda decision, toda eleccion o renuncia, es un paso importante en la vida y sólo dandoles el valor que merecen aprenderemos a decidir lo que realmente trascienda.

Emprender el camino significa tener una meta que alcanzar, unos objetivos que lograr. Y del aprecio y valoracion de esta meta dependera el esfuerzo que se ponga por alcanzarla. Muchas veces la ausencia de metas nobles, de objetivos claros y precisos, es causa del desaliento y abandono de los hombres en su afan de realizarse, cuando no de un eterno empezar que nunca llega.

Para algunos la tesis es un mero tramite para recibirse y constituye su unico sentido. Aunque este sea un objetivo, en mi caso no es, ni mucho menos, el principal. Al hacer este trabajo he procurado alcanzar dos objetivos, a mi modo de ver, mas importantes.

Mi primer objetivo es conocer mejor y profundizar en algun campo de mi carrera. Si el objetivo unico fuera cumplir con un requisito, un trabajo mediocre y superficial quedaria justificado. Por el contrario, considero que el tiempo que lleva elaborar la tesis, que de cualquier manera seria una inversion grande en horas de estudio, debe ser aprovechado con verdadera avaricia para hacerlo rendir en beneficio de mi propia formacion. Por ello es importante elegir un tema adecuado, que me guste, porque ayudara a lograr un trabajo serio sin escatimar esfuerzos.

Mi segundo objetivo, como consecuencia del anterior, es realizar un trabajo serio, pero sencillo y que pueda ser de utilidad, ya sea para quien gusta de lo teorico, o para aquellos que busquen sus aplicaciones. Por tanto no es finalidad de esta tesis "aportar" algo nuevo a las teorias matematicas, sino expresar de una manera mas sencilla lo que varios autores han expuesto.

Con estos dos objetivos en mente, busque un tema adecuado dentro del campo de la Econometria, cuyo estudio me habia gustado mucho en la carrera. Despues de considerar varias posibilidades, opte por el tema "COMBINACION DE SERIES DE TIEMPO Y CORTE TRANSVERSAL", con la idea de estudiar los modelos de regresion cuyos coeficientes fueran variables.

El trabajo, de manera muy sintetica, recoje los diversos modelos que existen de coeficientes variables cuya estimacion toma en cuenta precisamente la combinacion de series de tiempo y corte transversal. Despues de estudiarlos con detenimiento, presento un ejemplo del uso de los modelos aplicandolos a un modelo economico que estima la "DEMANDA DE AUTOMOVILES POR CATEGORIAS".

En cuanto a su estructura, la tesis esta dividida en seis capitulos y consta de dos apendices. El segundo capitulo resume algunas importantes propiedades del Algebra Lineal que seran de utilidad en el manejo posterior de los modelos. El capitulo 3 es una sintesis de los modelos de regresion, a manera de repaso general, y constituye la base sobre la cual se han de construir los modelos de coeficientes variables. En el capitulo 4 se estudian los modelos de coeficientes variables cuyas caracteristicas dependen de los supuestos que se hagan. El capitulo 5 muestra la aplicacion de los modelos a la demanda de automoviles por categorias, depurando las estimaciones en base a pruebas de hipotesis e interpreta los resultados obtenidos. Finalmente, en el capitulo 6 doy la conclusion del trabajo.

En el Apendice A se encuentran las tablas y las graficas de los datos empleados en el modelo del capitulo 5. En el apendice B estan los programas de computacion utilizados en las estimaciones. Al final, se expone la Bibliografia consultada en la elaboracion de la tesis.

CAPITULO 2

ELEMENTOS DE ALGEBRA LINEAL.

Cualquier estimacion al hacer una regresion involucra un gran numero de operaciones sencillas pero que en conjunto resultan sumamente tediosas. El algebra lineal proporciona una poderosa herramienta para manejar con sencillez la informacion, sin importar el volumen de la misma.

En este capitulo pretendemos recordar sinteticamente algunos elementos de algebra lineal que nos ayudaran al desarrollo de los capitulos siguientes. Su demostracion sale de nuestros objetivos y puede ser facilmente hallada en muchos libros de Algebra Lineal o de Econometria, algunos de los cuales se mencionan al final en la bibliografia.

2.1 DETERMINANTES.

El determinante de una matriz A se define asi:

$$\text{Det } A = |A| = \underset{j}{\text{SUM}} (-1)^t a_{1j_1} a_{2j_2} \dots a_{nj_n}$$

donde las j_i son valores de $j = \{1, 2, \dots, n\}$ y t es el numero de trasposiciones necesarias para reordenar las j_i's.

MENOR Aij es el determinante de la submatriz Aij, obtenida al suprimir el renglon i y la columna j de la matriz A.

$$\text{COFACTOR } A_{ij} = C_{ij} = (-1)^{i+j} \text{ Menor } A_{ij}$$

PROPIEDADES DE LOS DETERMINANTES.

2.1.1) $$|A| = \underset{i=1}{\overset{n}{\text{SUM}}} a_{ij} C_{ij} \qquad \text{para} \quad j = 1, 2, \dots, n$$
$$= \underset{j=1}{\overset{n}{\text{SUM}}} a_{ij} C_{ij} \qquad \text{para} \quad i = 1, 2, \dots, n$$

2.1.2) La suma de los elementos de un renglon por los cofactores de otro renglon es cero, es decir,

$$\underset{j}{\text{SUM}} a_{ij} C_{kj} = 0$$

2.1.3) $|A| = |A'|$ donde A' es la traspuesta de la matriz A.

2.1.4) Si A tiene una fila nula, entonces $|A| = 0$

2.1.5) Si A es triangular, su determinante es el producto de los elementos de la diagonal principal.

2.1.6) $|AB| = |A||B|$

2.2 LA MATRIZ INVERSA.

La matriz inversa de A es A^{-1}, tal que $A A^{-1} = A^{-1} A = I$

PROPIEDADES DE LA INVERSA.

2.2.1) A es invertible $\longleftrightarrow$ $|A| \neq 0$, es decir, si es no singular

2.2.2) $A^{-1} = (1 / |A|) \, \text{Adj } A$

donde $\text{Adj } A = (\text{Cof } A)'$ es la traspuesta del Cofactor de A, y

$\text{Cof } A = [C_{ij}]$ es la matriz de los cofactores de cada a_{ij}

2.2.3) La inversa de la inversa es la propia matriz, esto es,

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

2.2.4) El producto de dos matrices no singulares es una matriz no singular, y ademas

$(AB)^{-1} = B^{-1} A^{-1}$

2.2.5) La inversa de una traspuesta es la traspuesta de la inversa

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

2.2.6) El producto de dos matrices no singulares no puede ser una matriz nula, es decir,

A o B no singular y $AB = 0 \longrightarrow A = 0$ o $B = 0$

2.3 PRODUCTO DE KRONECKER.

Sea A una matriz de orden (M,N) y B otra de orden (K,L). Entonces el producto Kronecker de A y B, denotado $A \otimes B$, es una matriz de orden (MK,NL) dada por

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1N}B \\ a_{21}B & a_{22}B & \cdots & a_{2N}B \\ \vdots & \vdots & & \vdots \\ a_{M1}B & a_{M2}B & & a_{MN}B \end{bmatrix}$$

PROPIEDADES DEL PRODUCTO DE KRONECKER.

2.3.1) $(A \otimes B)(C \otimes D) = (AC \otimes BD)$

2.3.2) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

2.3.3) $(A \otimes B)' = A' \otimes B'$

2.3.4) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$

2.3.5) $(A \otimes B)\ X = AX \otimes B$

2.3.6) Generalmente $A \otimes B \neq B \otimes A$

2.4 VALORES Y VECTORES CARACTERISTICOS.

2.4.1) Sea A de orden (n,n) y $x \neq 0$. Las raices caracteristicas λ (lambda) y el vector caracteristico x son aquellos que satisfacen

$A\,x = \lambda x$ o bien,

$(A\,x - \lambda\,x) = 0$ lo cual impica que

$|\,A - \lambda I_n\,| = 0$

De esta igualdad se obtienen los n valores de λ y con ellos se forman los n vectores de x que satisfacen dichas condiciones.

2.4.2) MATRIZ ORTOGONAL es aquella cuya traspuesta es igual a su inversa, es decir,

$$X' = X^{-1} \quad \text{--->} \quad X' X = I \quad y \quad x_i' x_j = \begin{cases} 1 & si \quad i = j \\ 0 & si \quad i \neq j \end{cases}$$

2.4.3) Si una matriz es simetrica, sus raices caracteristicas son siempre no negativas y sus vectores caracteristicos son siempre ortogonales, es decir,

$$x_i' x_j = 0 \qquad \text{para } i \neq j$$

2.4.4) Si A es una matriz simetrica y X es ortogonal, entonces

$X' A X = \lambda In$ donde λ es la matriz diagonal de valores caracteristicos.

X puede ser, pues, la matriz de vectores caracteristicos de A.

2.5 FORMAS CUADRATICAS Y MATRICES POSITIVAS DEFINIDAS.

2.5.1) Sea A una matriz de orden (n,n) y x un vector columna de n elementos. La forma cuadratica de A esta definida por:

$$\begin{aligned} x' A x = a_{11} x_1^2 &+ 2 a_{12} x_1 x_2 + \ldots + 2 a_{1n} x_1 x_n \\ &+ a_{22} x_2^2 + \ldots + 2 a_{2n} x_2 x_n \\ &\ldots \\ &+ a_{nn} x_n^2 \end{aligned}$$

es decir, el escalar $x'A x$ es la suma ponderada de los productos de los elementos del vector x.

2.5.2) Si A es una matriz diagonal, su forma cuadratica es la suma de cuadrados de sus elementos,

$$x' A x = \text{SUM}_i \; a_{ii} x_i^2$$

2.5.3) Si A es POSITIVA DEFINIDA ---> $x' A x > 0$ para $x \neq 0$
y de ella se puede afirmar:
- Que es no singular, o sea, $|A| \neq 0$
- Que sus raices caracteristicas son positivas, y $|A| > 0$
- Que todos sus menores principales son positivos, siendo los menores principales los determinantes de las submatrices obtenidas al eliminar todos los renglones y columnas excepto las deseadas.

2.5.4) Sea X la matriz que diagonaliza a la matriz simetrica positiva definida A, es decir, $X'A\ X = \lambda I_n$. Como λ_i son positivos, podemos definir la matriz

$$D = \begin{bmatrix} d1 & & & & \\ & d2 & & & \\ & & . & & \\ & & & . & \\ & & & & dn \end{bmatrix} \quad \text{donde} \quad di = \frac{1}{SQ(\lambda_i)}$$

y multiplicarla por X, obteniendo $Q = XD$
de manera que

$$Q'A\ Q = I_n$$

Entonces

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

o bien,

$$A = P'P \qquad \text{donde} \quad P = Q^{-1}$$

Esta es la importante conclusion que nos recoge el siguiente teorema:

"Toda matriz simetrica positiva definida puede ser factorizada en el producto de una matriz ortogonal por su traspuesta".

CAPITULO 3

MODELOS DE REGRESION LINEAL

Un modelo, por definicion, es una representacion de un fenomeno, sea un sistema o un proceso. Dicho fenomeno es representado por el modelo en orden a explicarlo, a predecirlo y a controlarlo si es posible. En muchos casos los fenomenos de la vida real son muy complejos y requieren ser tratados mediante representaciones mas sencillas. Sin embargo cualquier modelo representa un compromiso entre la realidad y la facilidad de manejo; en ocasiones para lograr esto ultimo se requiere idealizar las condiciones del fenomeno, eliminando influencias extrañas y simplificando el proceso, aunque esto signifique sacrificar algo de realismo.

Un modelo econometrico representa matematicamente la relacion teorica existente entre algunas variables economicas . La econometria pretende emplear metodos estadisticos de analisis para probar hipotesis sobre estas relaciones, estimar sus magnitudes y hacer predicciones.

Dentro de los modelos econometricos estudiaremos aquellos que establecen una relacion lineal en los coeficientes de las variables, y no en las variables mismas, es decir, que establece una combinacion lineal entre la variable independiente y las otras variables del modelo, sin importar que ellas mismas puedan ser exponenciales o de otra forma. Por ejemplo, la relacion

$$Y = a + b X + c X^2$$

es una relacion lineal en sus coeficientes, a pesar de que la variable X es no lineal por estar elevada al cuadrado.

Por otra parte, cuando teoricamente se maneja una relacion no lineal en los coeficientes, muchas veces es posible hacer una transformacion que la convierta en lineal. Este es el caso donde la relacion

$$Y = a X^b$$

puede ser transformada a una relacion lineal al aplicar logaritmos:

$$\log Y = \log a + b \log X$$

A continuacion presentaremos un breve resumen de los modelos de regresion lineal estimados por los metodos de Minimos Cuadrados Ordinarios y Minimos Cuadrados Generalizados.

3.1 MODELO DE REGRESION DE DOS VARIABLES

Supongamos que queremos estimar la relacion lineal entre dos variables, donde la variable dependiente Y es funcion de la variable independiente X, expresada de la siguiente manera:

$$Y_t = A + B X_t + u_t \qquad t = 1, 2, \ldots, n \qquad (3.1.1)$$

donde

Y_t es la t-esima observacion de la variable dependiente

X_t es la t-esima observacion de la variable independiente, y

u_t es el t-esimo valor del termino de error

El conjunto de las n observaciones son tan solo una muestra del fenomeno real. Para cada observacion, Y es una variable aleatoria, X es una variable fija o no estocastica (es decir, conocida por el observador), y u es el termino aleatorio del error.

La presencia del error se debe a diversas causas. Primero, el error aparece porque el modelo es una simplificacion de la realidad; al omitir el modelo otras fuentes de variacion (sólo está en funcion de una) se incurre en error. Una segunda causa está asociada con la obtencion y medicion de los datos; la informacion de variables economicas muchas veces resulta dificil de obtener. Dadas estas razones resulta logico definir la relacion (3.1.1) como estocástica. Para cada valor de X, existe una distribucion de probabilidad de u y por lo mismo una distribucion de probabilidad de las Y's.

De acuerdo a lo anterior, el modelo de regresion lineal para dos variables queda especificado por los siguientes supuestos:

1) La relacion entre Y y X es lineal, como lo muestra (3.1.1)

2) Las variables X_t's son no estocasticas y sus valores son fijos.

3) a. El termino de error tiene un valor esperado de cero y una varianza constante para todas las observaciones, es decir,
   $E[u_t] = 0$ y $E[u_t^2] = \sigma^2$
   b. Las variables aleatorias del error no estan correlacionadas, es decir, errores correspondientes a diferentes observaciones tienen correlacion igual a cero.
   c. El termino de error tiene una distribucion normal.

Si el termino de error tiene varianza constante (como asumimos en la proposicion 3a), se dice que son "homocedásticos"; en cambio si la varianza cambia con las observaciones se dice que los errores son "heterocedásticos".

Si el supuesto 3b no se cumple, es decir, cuando los errores de diferentes observaciones sí están correlacionados, se dice que el proceso esta "autocorrelacionado". Si la correlacion existiese y fuese positiva, el error de un periodo tendera a asociarse a un error positivo en la siguiente observacion. Si fuese negativa, un error negativo está asociado con un error positivo para el siguiente periodo.

Como consecuencia de las proposiciones 2 y 3a podemos afirmar que la variable X y el termino de error u no están correlacionados. Finalmente la proposicion 3c sobre la normalidad de los errores implica que la variable dependiente Y tambien es normal, ya que es funcion lineal de ellos.

Cabe hacer notar que cada termino de error esta descrito en terminos de su varianza σ^2 que es desconocida. Por tanto nuestro modelo de regresion tiene tres parametros desconocidos (A, B y σ^2). Hemos descrito los supuestos del modelo en terminos del error pero se pueden definir tambien en terminos de la distribucion de probabilidad de la variable Y, diciendo

- La variable Y tiene valor esperado A + B X y varianza σ^2.

- Las variables aleatorias Y_i no estan correlacionadas.

ESTIMACION DE LOS PARAMETROS

Dado que nuestro modelo es lineal, nuestro objetivo es especificar una regla que nos permita definir la "mejor" linea recta que ajuste a los valores de las variables Y y X. Entre los varios metodos que se pueden sugerir, el de MINIMOS CUADRADOS resulta ser el más conveniente porque penaliza relativamente mas los errores grandes (observaciones lejos de la recta obtenida) que los pequeños. El criterio del metodo de minimos cuadrados es el siguiente: "la linea de 'mejor ajuste' es aquella que minimiza la suma de cuadrados de las desviaciones de los puntos observados respecto a la recta obtenida" (midiendo la distancia verticalmente).

De acuerdo con el metodo de Minimos Cuadrados, los estimadores de los parametros serian los siguientes:

$$b = \frac{\text{SUM } x_i\, y_i}{\text{SUM } x_i^2}$$

donde

b es el estimador del parametro B

$x_i = X_i - \bar{X}$ es la desviacion de X_i respecto a su media

$y_i = Y_i - \bar{Y}$ es la desviacion de Y_i respecto a su media

y

$$a = \bar{Y} - b\,\bar{X}$$

donde

a es el estimador del parametro A

Nos queda por estimar el ultimo parametro, la varianza del error que ademas es utilizada en los anteriores estimadores. El mismo metodo nos proporciona su estimador:

$$s^2 = \frac{\text{SUM } e_i^2}{N-2} = \frac{\text{SUM } (Y_i - a - b\,X_i)^2}{N-2}$$

PROPIEDADES DE LOS ESTIMADORES.

Tenemos ya una tecnica para estimar A y B. Cabe preguntarse si esta tecnica es buena o no. Nos interesa saber si a y b son "buenos" estimadores en el sentido de que posean ciertas propiedades estadisticas.

En particular interesa que un estimador:

1) Sea INSESGADO, es decir, que su media o valor esperado sea el mismo parametro. Por tanto esperamos que el valor esperado de a y b sean A y B respectivamente.

2) Sea EFICIENTE, es decir, que para una muestra dada, se espera que el estimador tenga una varianza menor a cualquier otro estimador insesgado. Si se diera que la varianza fuera cero, estariamos en el caso teorico de que el estimador es el mismo parametro.

3) Sea CONSISTENTE lo que quiere decir que al aumentar el tamaño de la muestra el estimador se aproxima a su parametro.

Dentro del conjunto de estimadores lineales insesgados, es decir, aquellos que son lineales en la variable independiente y son insesgados, los que tienen la menor varianza se llaman los "mejores estimadores lineales insesgados" y les nombraremos BLUE (del ingles "best linear unbiased estimator").

Bajo el metodo de minimos cuadrados, los estimadores a y b resultan BLUE. Y dado que Y_i es una variable aleatoria, entonces a y b tambien lo son y su distribucion estara en funcion de sus medias y varianzas:

$$E[b] = B \qquad\qquad E[a] = A$$

$$Var[b] = \frac{\sigma^2}{SUM\ x_i^2} \qquad\qquad Var[a] = \sigma^2 \frac{SUM\ X_i^2}{N\ SUM\ x_i^2}$$

y

$$Cov[a,b] = \frac{-\bar{X}\,\sigma^2}{SUM\ x_i^2}$$

El signo negativo de la covarianza nos indica que, cuando la media de X es positiva, entonces muy probablemente una sobrestimacion de a estara asociada a una subestimacion de b, y viceversa.

Como a y b dependen de la variable Y en forma lineal, y esta tiene una distribucion normal, podemos afirmar que a y b se distribuyen normalmente.

PRUEBAS DE HIPOTESIS E INTERVALOS DE CONFIANZA.

Conocer las distribuciones de los parametros a y b nos permite hacer pruebas de hipotesis sobre los parametros de la regresion. Los intervalos de confianza nos proporcionan un rango de valores que probablemente contiene el valor real de los parametros de la regresion. Con cada intervalo de confianza se asocia un "nivel se significancia"; determinado este ultimo, el intervalo de confianza es construido de manera que la probabilidad de que el intervalo contenga el parametro real de la regresion sea 1 menos el nivel de significancia.

Tanto las pruebas de hipotesis como los intervalos de confianza estan basados en las distribuciones de probabilidad por lo que conviene aqui hacer un pequeño parentesis para recordar algunas conclusiones sobre las mismas.

1) Distribucion Chi-cuadrada (Chi²). La suma de cuadrados de N variables aleatorias normales independientemente distribuidas (con media 0 y varianza 1) se distribuyen como una Chi² con N grados de libertad. Se puede demostrar que si se estima la varianza σ^2 de una muestra con distribucion normal con el estimador s^2, entonces $(N-1)s^2/\sigma^2$ se distribuye como una Chi cuadrada con (N-1) grados de libertad.

2) Distribucion t de student (t). Si X es una variable distribuida como una normal de media 0 y varianza 1, y Z esta distribuida como una Chi² con N grados de libertad, y ambas son independendientes, entonces X/SQ(Z/N) tiene una distribucion t con N grados de libertad.

3) Distribucion F. Si X y Z son independientes y se distribuyen como Chi² con N y M grados de libertad respectivamente, entonces

(X/N)/(Z/M) se distribuye como una F con (N,M) grados de libertad.

Nos interesa ahora establecer las pruebas de hipotesis. La mas comun es la llamada "hipotesis nula" que consiste en una proposicion o hipotesis que asume que cierto efecto no esta presente en el modelo.Como se espera aceptar el modelo, la prueba se diseña para ser rechazada. Por ejemplo, nuestra hipotesis nula puede ser que $B = 0$. Debemos usar nuestro estimador b para rechazar la hipotesis nula en favor de que la pendiente de la regresion no es cero.

La prueba estadistica para rechazar la hipotesis nula asociada a uno de los coeficientes de la regresion esta basada en la prueba t. Para usarla primero se debe estandarizar el parametro estimado restandole el valor hipoteticamente verdadero (la hipotesis nula en este caso sostiene que $B = Bo$) y dividiendolo entre su desviacion estandar estimada:

$$t_{N-2} = \frac{b - Bo}{s_B} = \frac{b - Bo}{s/SQ(SUM\ x_i^2)}$$

sigue una distribucion t con (N-2) grados de libertad. Ahora se selecciona un "valor critico" (tc) que asegura que la mitad del nivel de significancia de la distribucion t cae en ambas colas, es decir,

$$Prob\ (-tc < t_{N-2} < tc\ ) = 1 - \text{Nivel de significancia.}$$

Por tanto,

$$Prob\ (-tc < \frac{b - Bo}{s_B} < tc) = 1 - \text{Nivel de significancia.}$$

que resolviendo, obtenemos el siguiente intervalo de confianza para B:

$$b \pm tc\ s_B = b \pm tc\ s/\ (\ SQ(SUM\ x_i^2)\ )$$

De manera similar, el nivel de confianza para el parametro A es:

$$a \pm tc\ S_A = a \pm tc\ s\ (\ SQ(SUM\ Xi^2)\ /\ SQ(N\ SUM\ x_i^2)\ )$$

ANALISIS DE VARIANZA Y CORRELACION.

Hemos visto las propiedades de los estimadores de la regresion. Aunque hayan sido "buenos", queda aun la duda de si el diseño del modelo es el correcto o no; sin modelo no podemos explicar las variaciones que hay en la variable Y e interesa saber si nuestro modelo realmente "explica" las variaciones de Y, es decir, si la recta de la regresion ajusta con exactitud o no a las observaciones. El analisis de varianza y correlacion nos permite conocer esto precisamente.

El analisis esta basado en las observaciones, a partir de los residuales, es decir, de las diferencias que hay entre la observacion real y la estimacion hecha por el modelo. Despues de algo de algebra, se obtiene la siguiente relacion:

$$\mathrm{SUM}\ (Y_i - \bar{Y})^2 \quad = \quad \mathrm{SUM}\ (Y_i - \hat{Y}_i)^2 \quad + \quad \mathrm{SUM}\ (\hat{Y}_i - \bar{Y})^2$$

Variacion total de Y (o suma de cuadrados total)	Var. residual de Y (suma de cuadrados del error)	Var. explicada de Y (suma de cuadrados de la regresion)

$$TSS \quad = \quad ESS \quad + \quad RSS$$

Normalizando,

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

Y de aqui,

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}$$

donde

R^2 ("R cuadrada") es la proporcion de la variacion de Y explicada por la regresion de Y en X. Obviamente el rango de R^2 es de 0 a 1, siendo 1 cuando todos los puntos de la muestra ajustan con la linea, es decir, cuando el ajuste es perfecto. Por ello se usa mucho esta relacion para ver si el ajuste del modelo es bueno o no, esperando que este estadistico sea lo mas cercano posible a 1.

La division de la variacion de Y en sus dos componentes nos permite hacer una prueba sobre la correlacion de las variables Y y X. Bajo la hipotesis nula de que no hay correlacion entre ambas variables,

$$F_{1,N-2} = \frac{\text{varianza explicada}}{\text{varianza no explicada}} = \frac{RSS/1}{ESS/(N-2)}$$

se distribuye como una F con (1,N-2) grados de libertad. De esta manera se asocia un valor alto a la prueba si la relacion entre Y y X es fuerte, y un valor bajo a la prueba si la relacion entre dichas variables mas bien es debil.

3.2 MODELO DE REGRESION MULTIPLE.

Extendiendo el modelo de regresion de dos variables al caso en que la variable dependiente Y es funcion de una serie de variables independientes X1, X2, ... , XK podemos escribir

$$Y = B1 + B2\,X2 + \ldots + BK\,XK + e \qquad (3.2.1)$$

donde hemos asumido que X1 = 1 en todas las observaciones.

El estudio de este modelo es paralelo al anterior y diremos en primer lugar que los supuestos son los mismos, aunque aqui debemos incluir que no hay ninguna relacion lineal entre dos o mas de las variables independientes. Si se rompe este supuesto, es decir, si existe una relacion lineal o una alta correlacion entre las variables independientes, se dice que hay MULTICOLINEALIDAD y la interpretacion de los coeficientes estimados resulta complicada. Una manera de detectar si existe o no este problema es analizando la matriz de correlacion de los coeficientes; otro es conocer los valores caracteristicos de la matriz de informacion usando el metodo de "componentes principales"; si efectivamente hay multicolinealidad, alguno(s) de los valores sera(n) cero.

Dado que tenemos n ecuaciones del tipo (3.2.1), correspondientes a n observaciones, conviene escribir nuestro modelo en forma matricial:

$$Y = XB + e \qquad (3.2.2)$$

donde

Y es un vector de orden (N,1) de la variable dependiente

X es una matriz de orden (N,K) de las variables independientes

B es el vector de orden (K,1) de los parametros a estimar

e es el vector de orden (N,1) de los errores. El supuesto 3a del modelo anterior puede expresarse ahora $E\,[e\,e'] = \sigma^2\, I_N$

Aplicando de nuevo el criterio de Minimos Cuadrados, obtenemos:

$$b = (X'\,X)^{-1}\,X'\,Y \qquad (3.2.3)$$

La matriz (X' X) es llamada "matriz de productos cruzados" y es seguro que su inversa existe dado el supuesto de no correlacion entre las variables independientes.

El estimador b es BLU y su varianza es $\sigma^2\,(X'\,X)^{-1}$

La estimacion de la varianza de los errores puede obtenerse asi:

$$s^2 = \frac{\hat{e}'\,\hat{e}}{N - K} \qquad \text{donde } \hat{e} = Y - X\,b$$

Las pruebas sobre los coeficientes son tambien las mismas:

$$t_{N-K} = \frac{bi - Bi}{s/SQ(Var\ i)}$$

se distribuye como una t con (N-K) grados de libertad. Los intervalos de confianza se calculan de la misma manera que en el modelo anterior. Para las pruebas de correlacion, tenemos

$$Y'\,Y = b'\,X'\,X\,b + \hat{e}'\,\hat{e} \qquad \text{que equivale a}$$

$$TSS = RSS + ESS$$

Por tanto,

$$R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\hat{e}'\,\hat{e}}{Y'\,Y} = \frac{b'\,X'\,X\,b}{Y'\,Y} \qquad (3.2.4)$$

Entonces,

$$F = \frac{R^2}{1 - R^2}\,\frac{N-K}{K-1} = \frac{(b - B)'\,X'\,X\,(b - B)}{\hat{e}'\,\hat{e}}\,\frac{N-K}{K-1}$$

se distribuye como una F con (K-1,N-K) grados de libertad.

3.3 MINIMOS CUADRADOS GENERALIZADOS.

Hay ocasiones en que al diseñar un modelo economico el supuesto de que la varianza del error es constante resulta poco razonable. Por ejemplo, si la observaciones se hacen en corte transversal, esto es, a traves de diversas clases, firmas, individuos, etc., se espera que la dimension de las observaciones varie con el tamaño de la clase. Por tanto se espera que la varianza del error varíe con cada observacion, y tenemos un modelo "heterocedastico".

Cuando hay heterocedasticidad, el metodo de Minimos Cuadrados Ordinarios (M.C.O.) hasta ahora visto da mayor peso a las observaciones que tienen mayor varianza en los errores que a las de menor varianza. Por su ponderacion implicita, los estimadores de M.C.O son insesgados y consistentes, pero no son eficientes, es decir, su varianza no es la minima varianza; ademas la estimacion de sus varianzas sera sesgada.

El metodo de Minimos Cuadrados Generalizados (M.C.G) nos proporciona una manera de resolver el problema de heterocedasticidad. Hasta ahora el supuesto de homocedasticidad, en notacion matricial, decia

$$E[e e'] = \sigma^2 I$$

donde I es la matriz identidad de orden (N,N)

El supuesto de heterocedasticidad nos hace cambiar por el supuesto

$$E[e e'] = \sigma^2 \Omega \qquad (3.3.1)$$

donde Ω (Omega) es una matriz conocida de orden (N,N), y se asume que es positiva definida. Como vimos en el teorema 2.5.4, el supuesto de que Ω es una matriz positiva definida es suficiente para garantizarnos que existe una matriz P no singular de orden (N,N) tal que

$$P \Omega P' = I \qquad (3.3.2)$$

de manera que

$$\Omega = P^{-1} (P')^{-1} = (P'P)^{-1}$$

de donde

$$P' P = \Omega^{-1}$$

Podemos usar entonces la matriz P para transformar el modelo original:

$$P Y = P X B + P e \qquad (3.3.3)$$

o bien

$$Y^* = X^* B + e^*$$

donde

$$Y^* = P Y \qquad X^* = P X \qquad e^* = P e$$

El nuevo termino de error es ahora consistente con el modelo lineal clasico, dado que

$$E[e^* e^{*\prime}] = E[P e e' P'] = \sigma^2 P \Omega P' = \sigma^2 I \qquad (3.3.4)$$

Por tanto nuestro nuevo estimador sera

$$b = (X^{*\prime} X^*)^{-1} X^{*\prime} Y^*$$

que en terminos de los datos originales sera:

$$b = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y \qquad (3.3.5)$$

y su matriz de varianza-covarianza sera $\sigma^2 (X' \Omega^{-1} X)^{-1}$

En orden a poder aplicar M.C.G. es necesario estimar la matriz Ω ya que generalmente es desconocida, lo cual se puede hacer con la informacion que nos proporcionan las mismas observaciones y la especificaciones del modelo, como veremos en el capitulo siguiente.

CAPITULO 4

MODELOS DE COEFICIENTES VARIABLES

Los datos obtenidos a partir de la medicion de variables pueden provenir de diversas fuentes y ser de diversas formas. Los datos que describen el movimiento de una variable en el tiempo son llamados "series de tiempo" y pueden ser diarios, mensuales, anuales, etc. Los datos que describen las actividades de personas individuales, empresas u otras unidades en un momento determinado son llamados "datos en corte transversal".

El estudio de modelos que combinan datos en corte transversal y series de tiempo resulta de gran interes, pues nos permite aprovechar la informacion que se tiene de diversos individuos a traves del tiempo. Tendriamos entonces datos de este tipo:

$$Y_{it},\ X2_{it},\ X3_{it},\ \ldots,\ XK_{it} \qquad \begin{array}{l} i = 1, 2, \ldots, N \\ t = 1, 2, \ldots, T \end{array}$$

donde N denota el numero de clases o individuos y T el numero de observaciones en el tiempo. Y_{it} podria ser, por ejemplo, la produccion de maiz en la region i durante el periodo t. Esta variable estara en funcion de las (K-1) variables independientes.

Podemos entonces expresar nuestro modelo general de la manera siguiente:

$$Y_{it} = B1_{it} + \sum_{k=2}^{K} Bk_{it}\, Xk_{it} + e_{it} \qquad \begin{array}{l} i = 1, 2, \ldots, N \\ t = 1, 2, \ldots, T \end{array}$$

donde Bk_{it} es el k-esimo parametro a estimar
e_{it} es el termino de error correspondiente

De este modelo general podemos distinguir cinco casos:

1) $$Y_{it} = B1 + \sum_{k=2}^{K} Bk\, Xk_{it} + e_{it}$$

2) $$Y_{it} = B1_{i} + \sum_{k=2}^{K} Bk\, Xk_{it} + e_{it}$$

3) $$Y_{it} = B1_{it} + \sum_{k=2}^{K} Bk\, Xk_{it} + e_{it}$$

4) $$Y_{it} = B1_{i} + \sum_{k=2}^{K} Bk_{i}\, Xk_{it} + e_{it}$$

5) $$Y_{it} = B1_{it} + \sum_{k=2}^{K} Bk_{it}\, Xk_{it} + e_{it}$$

es decir, respectivamente,

1) Todos los coeficientes son constantes.

2) La interseccion varia entre los individuos pero es constante en cada individuo respecto al tiempo.

3) La interseccion varia entre los individuos y en cada uno.

4) La interseccion y la pendiente varian entre los individuos, pero permanece constante en cada uno respecto al tiempo.

5) La interseccion y la pendiente varian entre los individuos y en cada uno de ellos respecto al tiempo.

La importancia de este estudio radica en la interpretacion que de él se haga. No cabe duda que podemos establecer una relacion entre la variable Y y las otras variables mediante una simple regresion que no tome en cuenta la proveniencia de los datos. Sin embargo tambien es verdad que existen otras variables que no estamos en posibilidad de controlar o estimar y que afectan de manera distinta a cada clase o individuo. Pensemos, por ejemplo, que las diversas regiones donde se ha plantado el maiz tienen diferente precipitacion pluvial. En estos casos la aportacion de una determinada region a la variable dependiente puede ser sobrestimada por el modelo tradicional.

Podria pensarse, como de hecho lo considera el primer caso, que las variables X tienen el mismo efecto en Y (es decir, sus coeficientes son constantes en la regresion) en las diversas clases. A pesar de ello suele ser interesante probar este supuesto en la vida real. Por ejemplo, ¿la funcion de produccion de una industria tiene los mismos coeficientres en distintas regiones, o las funciones econometricas permanecen sin cambiar a traves de los años?

Este tipo de cuestiones se encuentran detras de nuestro estudio que en particular se orienta al uso del analisis de covarianza para:

1) Probar diferencias en las intersecciones, con el supuesto de que la pendiente permanece constante.

2) Probar las diferencias en las pendientes.

3) Probar diferencias en la relacion absoluta entre los individuos o clases y en cada individuo, considerando la informacion como un todo.

Cabe hacer una importante observacion sobre nuestro estudio. En el modelo tradicional se establece una combinacion lineal de las variables independientes (las X's) para determinar el valor de la variable dependiente (Y). Los parametros obtenidos en la regresion son los coeficientes de la combinacion lineal, como observamos en esta ecuacion:

$$\hat{Y}_i = \hat{B}1 + \hat{B}2_i \; X2_i + \ldots + \hat{B}K_i \; XK_i$$

Se ve claramente como estos coeficientes son los mismos para toda i, y la combinacion lineal nos define un espacio solucion de orden (K-1). Pensemos, por ejemplo, en el caso mas sencillo donde Y solo depende de una variable, digamos X. Entonces:

$$\hat{Y} = \hat{B}1 + \hat{B}2\ X$$

y esta es la ecuacion de una recta. La regresion ajusta, pues, las observaciones a una recta donde sus parametros estan fijos.

En nuestro modelo el primer caso pretende lo mismo, es decir, establecer un plano de soluciones de orden (K-1) en el que los parametros estan fijos.

Para los siguientes casos el modelo cobra un "dinamismo" singular ya que al ir variando sus coeficientes con cada individuo y en cada tiempo, procura ir "describiendo" con mayor realismo el comportamiento de las observaciones. En el ejemplo bidimensional estariamos buscando una recta cuya pendiente cambiaria en cada observacion, siendo no ya una recta rigida, sino una "recta" que sigue los lineamientos que las observaciones le indican.

Antes de pasar al analisis de cada uno de los casos debemos percatarnos de que en la practica al afrontar un problema de esta naturaleza surgira de inmediato la duda fundamental de cuál modelo debemos emplear. ¿Es correcto estimar con este modelo? ¿Cuál otro debe ser el correcto?

Discutir este punto constituye el fondo del problema, porque una vez tomada la decision, resolverlo sera solo cuestion de aplicar tal o cual metodo.

La decision dependera de los supuestos que cumplan las variables en cuestion, pues cada modelo esta especificado por distintos supuestos.

La tabla 4.1 que presentamos en la pagina siguiente nos muestra los diversos modelos de coeficientes variables como resultado de una combinacion entre series de tiempo y cortes transversales y que seran objeto de nuestro analisis, comenzando por el mas sencillo.

NOTAS:

1) El termino "series de tiempo" se refiere a observaciones estadisticas en el tiempo, no al analisis matematico que suele hacerse de una variable en el tiempo.

2) Al hablar de "combinacion lineal" es necesario recordar que la linealidad se refiere a los parametros B's a estimar, no a las variables independientes. Estas ultimas pueden ser no lineales (exponenciales, cuadraticas, etc) pero mediante una transformacion conveniente (por ejemplo logaritmica) pueden ajustarse tambien a un comportamiento lineal.

TABLA 4.1

CUADRO DE MODELOS DE COEFICIENTES VARIABLES

El modelo lineal

$$Y_{it} = B1_{it} + \sum_{k=2}^{K} BK_{it}\, XK_{it} + e_{it}$$

- Coeficientes constantes
  $BK_{it} = BK$
  eit heterocedasticos
  y autocorrelacionados
  (COCO)
- Pendiente constante e interseccion variable
  - solo sobre individuos
    $B1_{it} = B1_i$
    $Bk_{it} = Bk$, $k>1$
    - Coef. fijos
      Mod. Variables artificiales.
      (MOVAR)
    - Coef. aleatorios
      Modelo de componentes del error
      (MOCER)
  - sobre individuos y tiempo
    $B1_{it} = B1 + u_i + L_t$
    $Bk_{it} = Bk$, $k>1$
    - Coef. fijos
      Mod. Variables artificiales
      (MOVAR)
    - Coef. aleatorios
      Modelo de componentes del error
      (MOCER)
- Interseccion y pendiente variables
  - solo sobre individuos
    $Bk_{it} = Bk_i$
    - Coefic. fijos
      Mod regresiones no relacionadas aparentemente
      (MORAN)
    - Coef. aleatorios
      Modelo Swamy de coeficientes aleatorios
      (MOSCA)
  - sobre indiv. y tiempo
    $Bk_{it} = Bk + uk_i + Lk_t$
    - uk_i y Lk_t aleatorias
      Modelo Hsiao de coeficientes aleatorios
      (MOHCA)

4.1 MODELO DE COEFICIENTES CONSTANTES (COCO)

El modelo mas sencillo que podemos aplicar a nuestros datos es:

$$Y = X B + e \qquad (4.1.1)$$

que expresado en terminos de un individuo cualquiera queda

$$Y_i = X_i B + e_i \qquad (4.1.2)$$

donde

$Y_i = (Y_{i1}, Y_{i2}, \dots , Y_{iT})'$

X_i es una matriz de orden (T,K)

$B = (B1, B2, \dots , BK)'$

$e_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$ es el termino del error y es tal que

$E [e\, e'] = \sigma^2 I_T$

Nuestro modelo asume que todos los coeficientes a traves de los individuos son constantes y la estimacion de los K parametros del vector B pueden ser facilmente obtenidos al aplicar Minimos Cuadrados Ordinarios al sistema (4.1.1). Llamando b al estimador que resulta de aplicar este metodo, tenemos que

$$b = (X' X)^{-1} X' Y \qquad (4.1.3)$$

Alternativamente, podriamos agrupar los datos para los diversos individuos y estimar el vector B aplicando M.C.O. a cualquier individuo ya que B es constante y es independiente de los individuos, como podemos apreciar en la ecuacion (4.1.2). Por tanto la expresion (4.1.3) sera equivalente, bajo nuestros supuestos, a

$$b = (X_i' X_i)^{-1} X_i Y_i \qquad \text{para cualquier } i$$

El conjunto de nuestros supuestos resulta altamente restrictiva. Por otra parte debemos recordar que generalmente las regresiones hechas sobre datos en corte transversal pueden y suelen originar problemas de heterocedasticidad, es decir, que la varianza no permanece constante entre los individuos. Esto suena logico si pensamos que la varianza en el volumen de produccion de maiz debe ser muy distinta entre las diversas regiones de un pais.

Igualmente las observaciones en el tiempo conllevan el problema de la autocorrelacion, es decir, los errores en un periodo determinan en parte el error del siguiente periodo en un individuo cualquiera.

Por todo ello se propone un cambio en la suposicion sobre el termino de error, asumiendo ahora que

$$E[e_i e'_j] = \frac{\sigma_{ij}}{1-\rho_i\rho_j}\begin{bmatrix} 1 & \rho_j & \dots & \rho_j^{T-1} \\ \rho_i & 1 & \dots & \rho_j^{T-2} \\ \vdots & \vdots & & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \dots & 1 \end{bmatrix}$$

$$= \Omega_{ij} \tag{4.1.4}$$

De hecho esta matriz de covarianza resulta de un proceso autorregresivo de primer orden:

$$e_{it} = \rho_i e_{i,t-1} + v_{it} \qquad i = 1, 2, \dots, N$$

donde

$$E[v_{it}] = 0, \qquad E[v_{it} v_{jt}] = \sigma_{it}, \qquad E[v_{it} v_{js}] = 0 \text{ para } j \neq s$$

El modelo asume, entonces:

1) Que los coeficientes son los mismos para todos los individuos, lo que equivale a decir que los datos provenientes de diversos individuos es poco significativo o no tiene relevancia, y que la variacion en Y es efectivamente explicada por la variacion en las variables X's.

2) Que el vector de error para un individuo determinado sigue un proceso autorregresivo de primer orden.

3) Que la varianza del error puede ser diferente para distintos individuos.

4) Los errores de distintos individuos estan contemporaneamente correlacionados como se puede apreciar en la matriz de covarianza (4.1.4), pero dicha correlacion declina en el tiempo y es distinta para distintos individuos.

Para algunos casos estos supuestos parecen ser razonables. Quiza el supuesto mas restrictivo es el que afirma que los coeficientes sean constantes para todos los individuos.

4.1.1 ESTIMACION.

La estimacion de B puede hacerse por el metodo de Minimos Cuadrados Generalizados (M.C.G.) a partir de las NT observaciones, donde la matriz de pesos esta dada por la matriz de covarianza.

La matriz de covarianza (4.1.4) puede escribirse en terminos del modelo completo asi:

$$E[e e'] = \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \dots & \Omega_{1N} \\ \vdots & \vdots & & \vdots \\ \Omega_{N1} & \Omega_{N2} & & \Omega_{NN} \end{bmatrix}$$

Si conocieramos todos los ρ_i y los σ_{ij}, entonces B podria estimarse mediante M.C.G.

$$\hat{B} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y \qquad (4.1.5)$$

donde

X es una matriz de orden (NT,K)

que es de minima varianza en cada individuo, es insesgado y tiene matriz de covarianza $(X' \Omega^{-1} X)^{-1}$.

Sin embargo ρ_i y σ_{ij} son parametros desconocidos y debemos sustituirlos por sus estimadores, obteniendo entonces el "estimador de minimos cuadrados generalizados estimado" (E.M.C.G.) que llamaremos B^.

Para obtener los estimadores $\hat{\rho}_i$ y $\hat{\sigma}_{ij}$ aplicamos Minimos Cuadrados Ordinarios (M.C.O.) separadamente a cada individuo y los estimamos de los residuales correspondientes:

$$\hat{\rho}_i = \frac{\sum_{t=2}^{T} \hat{e}_i' \hat{e}_{i,t-1}}{\sum_{t=2}^{T} (\hat{e}_{i,t-1})^2} \qquad i = 1, 2, \ldots, N$$

donde

$$\hat{e}_i = (\hat{e}_{i1}, \hat{e}_{i2}, \ldots, \hat{e}_{iT})' = Y_i - X_i b_i$$

e igualmente,

$$\hat{\sigma}_{ij} = \frac{\hat{e}_i' \hat{e}_j}{T} \quad \text{siendo N (N+1) / 2 estimaciones} \qquad (4.1.6)$$

4.1.2 PRUEBAS DE HIPOTESIS.

Nos interesa probar, en primer termino, si la ganancia en eficiencia al estimar los parametros por medio del supuesto de autocorrelacion es significativa o no. Si no es significativo, entonces aplicar el estimador (4.1.3) sera equivalente a usar el estimador (4.1.5).

NOTA:

3) Si NT es demasiado grande, la inversion de la matriz Ω puede complicarse aun cuando se cuente con poderosas computadoras. En estos casos se podria hacer una transformacion buscando una matriz P tal que $P \Omega P' = S \otimes I_T$ cuya inversion sea mas facil y donde S es de esta forma:

$$S = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix}$$

Una manera de saber si es significativa sera comparar la suma de cuadrados de los residuales de ambos estimadores. Bajo la hipotesis de que no existe autocorrelacion (y por tanto la ganancia NO es significativa), proponemos un estadistico (Cfr. Pindyck,p.255):

$$F = \frac{((Y'Y - b\,X'Y) - (Y-XB)'\,\Omega^{-1}(Y-XB)) \;/\; (N+N(N+1)/2)}{(Y-XB)'\,\Omega^{-1}\,(Y-XB) \;/\; (NT-K-N-N(N+1)/2)} \qquad (4.1.7)$$

y se distribuye como una F con (N+N(N+1)/2,NT-K-N-N(N+1)/2) grados de libertad (se obtienen al restar del numero de observaciones el numero de elementos a estimar, sean coeficientes o varianzas; aqui tenemos NT observaciones y debemos restar K parametros, N coeficientes de autorregresion y N(N+1)/2 covarianzas).

Por otra parte resultara interesante probar la hipotesis de que los coeficientes son constantes. En la seccion 4.4 veremos un estadistico que nos permita hacerlo.

4.2 MODELOS DE PENDIENTE CONSTANTE E INTERSECCION VARIABLE SOBRE LOS INDIVIDUOS.

El segundo modelo es aquel que capta las diferencias entre los individuos en la interseccion mientras que asume que la pendiente permanece constante. Este modelo puede escribirse asi:

$$Y_{it} = \bar{B1} + u_i + \mathrm{SUM}_{k=2}\, Bk\, Xk_{it} + e_{it} \qquad i = 1, 2, \ldots, N \quad j = 1, 2, \ldots, T \qquad (4.2.1)$$

donde

$B1_i = \bar{B1} + u_i$ es la interseccion del i-esimo individuo, $\bar{B1}$ es la "interseccion media" y u_i la desviacion de esta media para el mismo individuo.

La estimacion de este modelo dependera de si la variable u_i es considerada fija o aleatoria. Si es fija, usaremos entonces el modelo de variables artificiales (MOVAR), mientras que si es aleatoria usaremos el modelo de componentes del error (MOCER). Ambos seran explicados a continuacion.

4.2.1 MODELO DE VARIABLES ARTIFICIALES. (MOVAR)

Para captar el efecto de los individuos en nuestro modelo usaremos variables aritificiales (dummy variables), que es un intento de justificar la ausencia de otras variables independientes explicativas del modelo. La ecuacion (4.2.1) para el i-esimo individuo puede escribirse asi:

$$Y_i = (\overline{B1} + u_i)\, J_T + Xs_i\, Bs + e_i \qquad (4.2.2)$$

donde

$Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})'$

$e_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$

$J_T = (1, 1, \dots, 1)'$ y es de orden $(T,1)$

Xs_i es la matriz X_i quitando la primera columna del termino de interseccion, y es de orden $(T,K-1)$

$Bs = (B2, B3, \dots, BK)'$ son $(K-1)$ parametros fijos a estimar

$B1 + u_i$ son los N parametros fijos de las intersecciones.

Asumiremos que $E[e_i]=0$, $E[e_i' e_i]=\sigma_e^2 I_T$ y $E[e_i' e_j]=0$ para $i \neq j$. Bajo estos supuestos el estimador de M.C.O. es el mejor estimador lineal insesgado (en ingles "BLUE"). Para conocer su estructura, escribiremos (4.1.2) para las NT observaciones:

$$Y = (I_N \otimes J_T \;\; Xs) \begin{bmatrix} B1 \\ Bs \end{bmatrix} + E \qquad (4.2.3)$$

es decir,

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1T} \\ Y_{21} \\ \vdots \\ \vdots \\ Y_{NT} \end{bmatrix} = \begin{bmatrix} 1 & & & & X2_{11} & \dots & XK_{11} \\ 1 & & & & & & \\ \vdots & & & & \vdots & & \vdots \\ \vdots & & & & X2_{1T} & \dots & XK_{1T} \\ 1 & & & & & & \\ & 1 & & & X2_{21} & \dots & XK_{21} \\ & \vdots & & & & & \\ & & \ddots & & \vdots & & \vdots \\ & & & 1 & X2_{NT} & \dots & XK_{NT} \end{bmatrix} \begin{bmatrix} B1_1 \\ B1_2 \\ \vdots \\ B1_N \\ B2 \\ B3 \\ \vdots \\ BK \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1T} \\ e_{21} \\ \vdots \\ \vdots \\ e_{NT} \end{bmatrix}$$

De aqui, aplicando M.C.O., obtendremos:

$$\begin{bmatrix} \hat{B1} \\ \hat{Bs} \end{bmatrix} = \left[(I_N \otimes J_T \;\; Xs)' \, (I_N \otimes J_T \;\; Xs) \right]^{-1} (I_N \otimes J_T \;\; Xs)'$$

$$\begin{bmatrix} \hat{B1} \\ \hat{Bs} \end{bmatrix} = \left[\begin{bmatrix} I_N \otimes J_T' \\ Xs' \end{bmatrix} (I_N \otimes J_T \;\; Xs) \right]^{-1} \begin{bmatrix} I_N \otimes J_T' \\ Xs' \end{bmatrix} Y \qquad (4.2.4)$$

Observemos que el vector $(\hat{B1}, \hat{Bs})'$ es de orden $(N+K-1,1)$ que nos indica que los primeros N valores corresponden al subvector $(B1_1, B1_2, \dots, B1_N)$ mientras que los restantes $(K-1)$ valores son de los parametros $(B2, B3, \dots, BK)$.

El proceso se hace mas sencillo si a la ecuacion (4.2.1) la corregimos por su media respecto al tiempo, de esta manera:

$$(Y_{it} - \bar{Y}_i) = \sum_{k=2}^{K} Bk\ (Xk_{it} - \overline{Xk}_i) + e_{it} - \bar{e}_i \qquad (4.2.5)$$

donde

$$\bar{Y}_i = \sum_{t=1}^{T} Y_{it} \,/\, T\,, \qquad \overline{Xk}_{it} = \sum_{t=2}^{T} Xk_{it} \,/\, T \quad y \quad \bar{e}_{it} = \sum_{t=2}^{T} e_{it} \,/\, T$$

A partir de aqui podemos obtener Bs aplicando M.C.O. a (4.2.5), requiriendo la inversion de una matriz de orden (K-1) ya que logramos eliminar el termino de la interseccion que antes nos obligaba a invertir una matriz de orden (N+K-1).

La correccion, pues, en Xs en el t-esimo momento nos queda:

$$(\overline{X2}_{it} - \overline{X2}_{i}\ , \ldots ,\ \overline{XK}_{it} - \overline{XK}_{i}\ )$$

La transformacion completa de Xs para el individuo i sera entonces:

$$D_T\ Xs_i$$

donde

$$D_T = I_T - J_T\, J'_T \,/T = \begin{bmatrix} 1-1/T & -1/T & \ldots & -1/T \\ -1/T & 1-1/T & \ldots & -1/T \\ \vdots & \vdots & & \vdots \\ -1/T & -1/T & \ldots & 1-1/T \end{bmatrix}$$

La transformacion completa de Xs sera, finalmente:

$$(I_N \otimes D_T)\ Xs$$

y el nuevo sistema de ecuaciones quedara asi:

$$(I_N \otimes D_T)\ Y = (I_N \otimes D_T)\ Xs\ Bs + (I_N \otimes D_T)\ E \qquad (4.2.6)$$

Aplicando minimos cuadrados, obtenemos:

$$\hat{Bs} = \left[\ ((I_N \otimes D_T)Xs)'\ (I_N \otimes D_T)Xs\ \right]^{-1}\ ((I_N \otimes D_T)Xs)'\ (I_N \otimes D_T)\ Y$$

$$= \left[\ Xs'(I_N \otimes D_T)'(I_N \otimes D_T)Xs\ \right]^{-1}\ Xs'(I_N \otimes D_T)'(I_N \otimes D_T)\ Y$$

$$= \left[\ Xs'(I_N I_N \otimes D_T D_T)Xs\ \right]^{-1}\ Xs'(I_N I_N \otimes D_T D_T)\ Y$$

$$= \left[\ Xs'(I_N \otimes D_T)\ Xs\ \right]^{-1}\ Xs'(I_N \otimes D_T)\ Y \qquad (4.2.7)$$

Desarrollando las matrices y simplificando,

$$\hat{Bs} = \left[\ \sum_{i=1}^{N} Xs'_i D_T Xs'_i\ \right]^{-1} \sum_{i=1}^{N} Xs'_i D_T Y_i \qquad (4.2.8)$$

Tambien en este caso se requiere invertir una matriz de orden (K-1) y la matriz de varianza estara dada por

$$\sigma_e^2\ (\ Xs'\ (\ I_N \otimes D_T\ )\ Xs\ )^{-1} \qquad (4.2.9)$$

El estimador $\hat{Bs}$ esta obtenido a partir de las variaciones en el tiempo dentro de cada individuo (corrigiendolo por su media respecto al tiempo) y por esta razon nos referiremos a el como el "estimador interno" y lo escribiremos 'bs' por razones que mas adelante se expondran.

Hemos estimado Bs y nos falta estimar el vector de intersecciones B1, lo que podemos hacer sencillamente por este metodo:

$$\hat{B1}_i = \bar{Y}_i - \sum_{k=2}^{K} \hat{Bk}\, \bar{Xk}_i \qquad (4.2.10)$$

y promediando los resultados que se obtengan,

$$u_i = \hat{B1}_i - \bar{\bar{B1}}$$

Una solucion alternativa a la ecuacion (4.2.3) se presenta para el caso en que la computadora introduce automaticamente el termino constante. En este caso una reparametrizacion para evitar la singularidad de la matriz X es necesaria. Podemos escribir:

$$Y = \left[J_{NT} \begin{bmatrix} I_{N-1} \\ 0 \end{bmatrix} \otimes J_T \quad Xs \right] \begin{bmatrix} d \\ Bs \end{bmatrix} + E \qquad (4.2.11)$$

donde

$$d = (d_1, d_2, \ldots, d_N)'$$

$$B1_N = d_1 \quad y \quad B1_i = d_{i+1} + d_1 \qquad i = 1, 2, \ldots, N$$

Las variables artificiales, para N=3 como ejemplo, serian:

$$\left[J_{3T} \begin{bmatrix} I_2 \\ 0 \end{bmatrix} \otimes J_T Xs \right] = \begin{bmatrix} 1 & 1 & 1 & \ldots & \ldots & \ldots & 1 & 1 \\ 1 & 1 & \ldots & 1 & & & & \\ & & & & 1 & 1 & \ldots & 1 \end{bmatrix}'$$

la estimacion con las nuevas variables es exactamente la misma.

Como pruebas de hipotesis, las pruebas conocidas de M.C.O. son validas y resulta de especial interes la prueba F sobre los residuales bajo la hipotesis de que

$$B1_1 = B1_2 = \ldots = B1_N = \bar{\bar{B1}}$$

4.2.2 MODELO DE COMPONENTES DEL ERROR.

(MOCER)

Estudiaremos ahora el mismo modelo de la ecuacion (4.2.1) pero con la suposicion de que las u_i son variables aleatorias con media cero, varianza σu^2 y covarianza cero entre diferentes individuos. Se asume tambien que u_i y e_{it} no estan correlacionados. Para el individuo i, tenemos:

$$Y_i = X_i B + u_i J_T + e_i \qquad (4.2.12)$$

donde

X_i incluye el termino constante y es de orden (T,K)

$\bar{B} = (B1, B2, \ldots, BK)$

y la matriz de covarianza de la composicion del error es:

$$\Phi_i = E[(u_i J_T + e_i)(u_i J_T + e_i)'] = \sigma u^2 J_T J_T' + \sigma e^2 \qquad (4.2.13)$$

El supuesto de que las u_i son variables aleatorias implica:

1) Que los N individuos del problema pueden ser considerados como una muestra de una poblacion mayor.

2) Que las u_i no estan correlacionadas con las Xi, pues de otro modo su media estaria en funcion de las segundas, es decir $E[u_i] = f(Xi)$ y esto seria distinto de cero, violando el supuesto del modelo.

3) Que este modelo puede ser visto como uno donde todos los coeficientes son constantes y la matriz de covarianza del error es de la forma (4.2.13). Es interesante comparar esta matriz de covarianza con la vista para el modelo COCO (4.1.4). En la presente la matriz de covarianza es identica para todos los individuos; los errores en un individuo dado estan correlacionados en diferentes tiempos, pero dicha correlacion es constante y es identica para todos los individuos. En cambio en (4.1.4) vemos que la correlacion declina conforme haya mas tiempo de distancia y puede ser diferente para cada individuo.

Para la estimacion, incluyamos a todos los individuos:

$$Y = X B + u \otimes J_T + E \qquad (4.2.14)$$

donde

$$u = (u_1, u_2, \ldots, u_N)'$$

y la matriz de covarianza sera diagonal en bloques como sigue

$$\Phi = E[(u \otimes J_T + E)(u \otimes J_T + E)'] = I_N \otimes \Phi_i \qquad (4.2.15)$$

Si conocieramos σu^2 y σe^2, entonces estimando con M.C.G.:

$$\hat{B} = (X' \Phi^{-1} X)^{-1} X' \Phi^{-1} Y \qquad (4.2.16)$$

Haciendo una particion, $\hat{B} = (\hat{\bar{B}}1, \hat{B}s)'$, es posible demostrar que el segundo componente es una matriz ponderada del "estimador interno" (4.2.8) y otro estimador conocido como "estimador externo" por tomar en cuenta las variaciones entre los individuos.

Para demostrarlo, notemos primero que

$$\Phi^{-1} = I_N \otimes \varphi_i^{-1} = I_N \otimes \left[\frac{J_T J_T'}{T \sigma 1^2} + \frac{D_T}{\sigma e^2} \right] \qquad (4.2.17)$$

donde

$$\sigma 1^2 = T \sigma u^2 + \sigma e^2$$

Hagamos tambien las particiones $X = (J_{NT}, Xs)$ y $X_i = (J_T, Xs_i)$
Entonces, si las sustituimos junto con (4.2.17) en (4.2.16), obtenemos con algo de algebra

$$\hat{Bs} = \left[\frac{Xs'\ Q1\ Xs}{\sigma 1^2} + \frac{\sum_{i=1}^{N} Xs_i'\ D_T\ Xs_i}{\sigma e^2} \right]^{-1} \left[\frac{Xs'\ Q1\ Y}{\sigma 1^2} + \frac{\sum_{i=1}^{N} Xs_i'\ D_T\ Y_i}{\sigma e^2} \right]$$

donde
$Q1 = I_N * J_T J_T' - J_{NT} J_{NT}' / NT$ es una matriz idempotente tal que

$$Q1\ Xs = \begin{bmatrix} \overline{X21.} - \overline{X2..}, \ldots, \overline{XK1.} - \overline{XK..} \\ \overline{X22.} - \overline{X2..}, \ldots, \overline{XK2.} - \overline{XK..} \\ \vdots \qquad\qquad \vdots \\ \overline{X2N.} - \overline{X2..}, \ldots, \overline{XKN.} - \overline{XK..} \end{bmatrix} * J_T$$

y

$$\overline{Xk..} = \sum_{i=1}^{N} \sum_{t=1}^{T} Xk_{it} / NT$$

es decir, Q1 nos corrige por la media total las matrices Y y X.

El "estimador externo" sera entonces:

$$Bs^* = (Xs'\ Q1\ Xs)^{-1}\ Xs'\ Y \tag{4.2.18}$$

que es obtenida al aplicar M.C.O. a

$$Y_{i.} = \overline{B1} + u_i + \sum_{k=1}^{K} Bk\ \overline{Xk}_{i.} + \sum_{t=1}^{T} e_{it} \tag{4.2.19}$$

que es la ecuacion (4.2.1) promediada en el tiempo.
Podemos escribir ahora $\ddot{Bs}$ de esta manera: (4.2.20)

$$\hat{Bs} = \left[\frac{Xs'Q1\ Xs}{\sigma 1^2} + \frac{\sum_{i=1}^{N} Xs_i'D_T\ Xs_i}{\sigma e^2} \right]^{-1} \left[\frac{Xs'Q1\ Xs}{\sigma 1^2} Bs^* + \frac{\sum_{i=1}^{N} Xs_i'D_T\ Xs_i}{\sigma e^2} bs \right]$$

que es una matriz ponderada de Bs* y bs. Por tanto el estimador de M.C.G. puede verse como una combinacion eficiente del estimador de variables artificiales que usa la variacion dentro de cada individuo y el estimador Bs* que utiliza la variacion entre los individuos.

Para obtener el termino de interseccion, resolvemos:

$$\hat{\overline{B1}} = \overline{Y} - \sum^{K} \hat{Bk}\ \overline{Xk} \tag{4.2.21}$$

Hemos visto en la ecuacion (4.2.20) la relacion entre los estimadores, pero computacionalmente puede resultar complicada su aplicacion.

Para resolver en la practica nuestro modelo, pensamos en una transformacion a partir de una matriz P, tal que $P'P = c\Phi^{-1}$ donde c es cualquier escalar.

Buscaremos entonces una P_i y Φ_i derivadas de la matriz ortogonal que diagonalice φ_i. Las raices caracteristicas de Φ_i son $\sigma 1^2$ y σ_e^2 con multiplicidad 1 y (T-1) respectivamente. De acuerdo con varios autores, dicha matriz seria $C = (1/SQ(T)\ J_T,\ C1)'$ donde C1 es cualquier matriz de orden (T-1,T), tal que

$$C1\ J_T = 0\ , \qquad C1\ C1' = I_{T-1} \quad y \quad C1'C1 = D_T$$

es decir,

$$C'\varphi_i C = \begin{bmatrix} \sigma 1^2 & & & \\ & \sigma e^2 & & \\ & & \ddots & \\ & & & \sigma e^2 \end{bmatrix}$$

Una matriz de transformacion seria $P = I_N \otimes P_i$ donde

$$P_i = I_T - \left(1 - \frac{\sigma e}{\sigma 1}\right) \frac{J_T\ J_T'}{T} \tag{4.2.22}$$

Multiplicando por P ambos lados de la ecuacion (4.2.14) resulta

$$Y_{it} - a\ \bar{Y}_{i.} = (1-a)\ \overline{B1} + \sum_{k=2}^{K} Bk\ (Xk_{it} - a\ \overline{Xk}_{i.}) + v_{it} \tag{4.2.23}$$

donde $a = 1 - \dfrac{\sigma e}{\sigma_1}$

y v_{it} son errores homocedasticos y no correlacionados.

PREDICCION DE u_i

Siendo u_i una variable aleatoria, podemos establecer un predictor que nos permita conocer como varia con los individuos. El estimador BLU seria:

$$\hat{u}_i = \frac{\sigma_u^2}{\sigma_1^2} J_T'\ (Y_i - X_i \hat{B}) \tag{4.2.24}$$

ESTIMACION DE LOS COMPONENTES DE VARIANZA.

Tanto el estimador $\hat{B}$ como el predictor $\hat{u}_i$ dependen de las varianzas de e y de u que generalmente son desconocidas. Existen varios metodos para estimar estos componentes de varianza, pero solo estudiaremos los que usan los residuales de Bs^* y de bs. Entonces,

$$\sigma 1^2 = \frac{e^{*\prime}\ e^*}{N-K} \tag{4.2.25}$$

donde

$e^* = Q1\ Y - Q1\ Xs\ Bs^*$ son los residuales de B^* y

$$\sigma e^2 = \frac{\hat{e}'\hat{e}}{N(T-1)-K} \tag{4.2.26}$$

donde

$\hat{e} = (I_N \otimes D_T)\ Y - (I_N \otimes D_T)\ Xs\ bs$ son los residuales de bs.

De aqui obtenemos facilmente

$$\sigma_u^2 = \frac{\sigma 1^2 - \sigma e^2}{T} \tag{4.2.27}$$

que es el estimador de la varianza de u_i y que tiene la clara desventaja de que puede ser negativa. Si esto ocurre, puede ser indicacion de que el efecto del tiempo ha sido equivocadamente omitido. Una solucion alternativa seria considerarla cero, volviendo entonces al modelo de Minimos Cuadrados Generalizados ya que no existe efecto diferente en los indiviudos (σu^2).

4.2.3 ELECCION DE SUPUESTOS.

Cuando la interseccion varia con los individuos, ¿cuál de los dos modelos, MOVAR o MOCER debemos usar? ¿Los efectos de la variacion son fijos o son aleatorios? La eleccion no es sencilla.

Yair Mundlak sugiere que en ambos casos se considere u_i como aleatoria, pero que al usar MOVAR nuestra inferencia está condicionada al valor de u_i en la muestra, mientras que en MOCER la inferencia es incondicional ya que se hacen suposiciones especificas de su distribucion; si ademas dichas suposiciones son correctas, el estimador de MOCER sera mas eficiente al aprovechar mayor informacion.

La cuestion relevante estara, pues, en saber si las u_i tienen una distribucion i.i.d. $(0, \sigma u^2)$ o cuándo siguen algun otro proceso aleatorio. Mundlak sostiene que lo mas razonable es que dicho proceso estabelzca una correlacion entre las u_i y las X_i, en cuyo caso $E[u_i] = f(X_i)$, violando el supuesto de MOCER y dando un estimador sesgado por lo que en este caso convendra usa MOVAR.

En resumen, si las u_i pueden ser consideradas como extracciones ALEATORIAS y su distribucion NO esta correlacionada con X_i entonces el estimador de MOCER sera BLU. Si dicha distribucion no se puede explicar o esta correlacionada, entonces el estimador BLU sera el de MOVAR.

Una prueba de hipotesis para ver si estan correlacionadas resultara conveniente. Bajo la hipotesis nula de que MOCER sera el modelo conveniente (es decir, que no hay correlacion),

$$m = (bs - \hat{B}s)'\ (M1 - M0)^{-1}\ (bs - \hat{B}s)$$

se distribuye asintoticamente como una Chi^2 (K-1) donde

$M1 = \hat{\sigma}_e^2 \, (Xs' \, (I_N \otimes D_T) \, Xs)^{-1}$ es la matriz de covarianza de bs

$M0 = ((1/\sigma_1^2) \, Xs' Q1 \, Xs \; + (1/\sigma_e^2)(Xs' (I_N \otimes D_T) \, Xs))^{-1}$ es la de $\hat{B}s$

Si se rechaza la hipotesis nula, MOVAR sera mas conveniente ya que u_i y X_i estan correlacionadas provocando mucha diferencias entre los estimadores de ambos modelos.

4.3 MODELOS DE PENDIENTE CONSTANTE E INTERSECCION VARIABLE SOBRE TIEMPO E INDIVIDUOS

El modelo estudiado en la seccion anterior puede extenderse al incluir en la interseccion un componente variable sobre el tiempo, pero constante sobre los individuos. Nuestro modelo seria

$$Y_{it} = \overline{B1} + u_i + 1_t + \sum_{k=1}^{K} Bk \, Xk_{it} + e_{it} \qquad i = 1, \ldots, N \quad j = 1, \ldots, T \quad (4.3.1)$$

donde

$\overline{B1} + u_i + 1_t$ es el termino de interseccion, siendo 1_t el termino que en un periodo dado afecta de igual manera a todos los individuos.

El analisis de este modelo sera paralelo al modelo anterior, dependiendo de cuando u_i y 1_t sean fijas o aleatorias.

4.3.1 MODELO DE VARIABLES ARTIFICIALES (MOVAR)

Cuando son consideradas fijas, podemos imponer la restriccion de que SUM $u_i = 0$ y SUM $1_t = 0$ ya que una de ellas resulta redundante.

Para resolver (4.3.1) por M.C.O. necesitamos reparametrizarla incorporando las restricciones. Una reparametrizacion es definir

$$B1_i = \overline{B1} + u_i \qquad i = 1, 2, \ldots, N$$

$$1\%_t = 1_t - 1_T \qquad t = 1, 2, \ldots, T$$

quedando el modelo para el individuo i

$$Y_i = B1_i \, J_T + \begin{bmatrix} I_{T-1} \\ 0' \end{bmatrix} 1\% + Xs_i Bs + e_i \qquad (4.3.2)$$

donde

$$1\% = (11\%,\ 12\%,\ \ldots,\ 1T\%)'$$

Para todos las observaciones, el modelo quedaria

$$Y = \begin{bmatrix} I_N \otimes J_T & J_N \otimes \begin{bmatrix} I_{T-1} \\ 0' \end{bmatrix} & Xs \end{bmatrix} \begin{bmatrix} B1 \\ 1\% \\ Bs \end{bmatrix} + E \qquad (4.3.3)$$

donde

$$B1 = (B1_1,\ B1_2,\ \ldots,\ B1_N)'$$
$$Y = (Y_1,\ Y_2,\ \ldots,\ Y_N)$$

Asuminos que $E[E] = 0$ y que $E[E E'] = \sigma e^2 I_{NT}$ y por tanto el vector estimador $(B1, 1\%, Bs)'$ de orden $(N+T+K-2,1)$ es BLUE. Como ejemplo para N=3

$$\begin{bmatrix} I_3 \otimes J_T & J_3 \otimes \begin{bmatrix} I_{T-1} \\ 0' \end{bmatrix} \end{bmatrix} = \begin{bmatrix}
1 & & & 1 & & & \\
1 & & & & 1 & & \\
\vdots & & & & & \ddots & \\
1 & & & & & & 0 \\
 & 1 & & 1 & & & \\
 & 1 & & & 1 & & \\
 & \vdots & & & & \ddots & \\
 & 1 & & & & & 0 \\
 & & 1 & 1 & & & \\
 & & 1 & & 1 & & \\
 & & \vdots & & & \ddots & \\
 & & 1 & & & & 0
\end{bmatrix}$$

El estimador de M.C.O. puede obtenerse de (4.3.3), pero si el vector es demasiado grande, podemos entonces particionarlo y obtener el vector de coeficientes para la pendiente:

$$bs = (Xs' \, Q \, Xs)^{-1} \, Xs \, Q \, Y \qquad (4.3.4)$$

donde

$$Q = I_{NT} - I_N \otimes \frac{J_T J_T'}{T} - \frac{J_N J_N'}{N} \otimes I_T + \frac{J_{NT} J_{NT}'}{NT}$$

es una matriz idempotente que transforma las observaciones de X e Y corrigiendolas por su media, de forma que, por ejemplo, QXs es una matriz de orden (NT,K-1) con elemento tipico dado por

$$Xk_{it} - \overline{Xk}_{i.} - \overline{Xk}_{.t} + \overline{Kk}_{..} \qquad \text{definidos como los promedios}$$

Si la ecuacion (4.3.1) es promediada sobre i, sobre t y sobre ambos, obtendremos respectivamente

$$\bar{Y}_{i.} = \overline{B1} + u_i + \sum_{k=2}^{K} \overline{Xk}_{i.} \, Bk + \sum_{t=1}^{T} e_{it} / T \qquad (4.3.5)$$

$$\bar{Y}_{.t} = \overline{B1} + 1_t + \sum_{k=2}^{K} \overline{Xk}_{.t} \, Bk + \sum_{i=1}^{N} e_{it} / N \qquad (4.3.6)$$

$$\bar{Y}_{..} = \bar{\bar{B}}1 + \sum_{k=2}^{K} \bar{\bar{X}}k_{..}\ Bk + \sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}/NT \tag{4.3.7}$$

Restando (4.3.5) y (4.3.6) a (4.3.1) y (4.3.7), tenemos

$$Y_{it} - \bar{Y}_{i.} - \bar{Y}_{.t} + \bar{Y}_{..} = \sum_{k=2}^{K} (Xk_{it} - \bar{\bar{X}}k_{i.} - \bar{\bar{X}}k_{.t} + \bar{\bar{X}}k_{..})\ Bk + v_{it} \tag{4.3.8}$$

donde v_{it} se define de manera obvia.

El estimador de (4.3.4) puede considerarse como el estimador de M.C.O. aplicado a (4.3.8). Tambien puede ser visto como el estimador de M.C.G. donde Q O = O es la matriz de covarianza de los errores y Q es la inversa generalizada de si misma. La matriz de covarianza de bs es $\sigma e^2 (Xs'\ Q\ Xs)^{-1}$.

Los coeficientes restantes pueden obtenerse de:

$$\hat{u}_i = (\bar{Y}_{i.} - \bar{Y}_{..}) - \sum_{k=2}^{K} (\bar{\bar{X}}k_{i.} - \bar{\bar{X}}k_{..})\ bk \tag{4.3.9}$$

$$\hat{l}_t = (\hat{l}_{.t} - \bar{Y}_{..}) - \sum_{k=2}^{K} (\bar{\bar{X}}k_{.t} - \bar{\bar{X}}k_{..})\ bk \tag{4.3.10}$$

y

$$\hat{b}1 = \bar{Y}_{..} - \sum_{k=2}^{K} \bar{\bar{X}}k_{..}\ bk \tag{4.3.12}$$

Para probar la hipotesis sobre la inclusion o no de variables artificiales podemos usar tambien la prueba F que compara las sumas de cuadrados restringidos y no restringidos.

4.3.2 MODELO DE COMPONENTES DEL ERROR (MOCER)

En esta seccion asumimos que los efectos del tiempo y de los individuos no son fijos sino aleatorios, de forma que

$E[u_i] = 0 \qquad E[u_i^2] = \sigma_u^2 \qquad E[u_i u_j] = 0$ para $i \neq j$

$E[l_t] = 0 \qquad E[l_t^2] = \sigma_l^2 \qquad E[l_t l_s] = 0$ para $t \neq s$

y entre ellos no estan correlacionados.

Las observaciones para el i-esimo individuo quedan:

$$Y_i = X_i B + u_i J_T + I_T l_- + e_i \qquad (4.3.12)$$

donde

$$\underline{l} = (l_1, l_2, \ldots, l_T)'$$

La matriz de covarianza para el error compuesto es:

$$\phi_{ii} = E\,[\,(u_i J_T + I_T l + e_i)(u_i J_T + I_T l + e_i)'\,]$$

$$= \sigma u^2 J_T J_T' + \sigma l^2 I_T + \sigma e^2 I_T \qquad (4.3.13)$$

y la covarianza entre los vectores de dos individuos es

$$\phi_{ij} = E\,[\,(u_i J_T + I_T l + e_i)(u_j J_T + I_T l + e_j)'\,]$$

$$= \sigma l^2 I_T \qquad (4.3.14)$$

Tomando en cuenta todas las NT observaciones, el modelo queda

$$Y = X B + \underline{u} \otimes J_T + (J_N \otimes I_T)\underline{l} + E \qquad (4.3.15)$$

donde

$$\underline{u} = (u_1, u_2, \ldots, u_N)'$$

y la matriz de covarianza completa es

$$\Phi = E\,[(u \otimes J_T + (J_N \otimes I_T)\underline{l} + E)(u \otimes J_T + (J_N \otimes I_T)\underline{l} + E)']$$

$$= \sigma u^2 (I_N \otimes J_T J_T') + \sigma l^2 (J_N J_N' \otimes I_T) + \sigma e^2 I_{NT} \qquad (4.3.16)$$

Aqui podemos ver que Φ no es diagonal en bloque, y por tanto los errores en diferentes individuos estan contemporaneamente correlacionados.

ESTIMACION POR M.C.G.

Para estimar B, el estimador de M.C.G.

$$\hat{B} = (X' \Phi^{-1} X)^{-1} X' \Phi^{-1} Y \qquad (4.3.17)$$

es BLUE con matriz de covarianza $(X' \Phi^{-1} X)^{-1}$

Si B es particionada en (B1,Bs), es posible demostrar que el estimador de los coeficientes de la pendiente, Bs, es una matriz ponderada de otros tres estimadores.

La multiplicacion directa muestra que:

$$\Phi = \frac{Q}{\sigma_e^2} + \frac{Q1}{\sigma_1^2} + \frac{Q2}{\sigma_2^2} + \frac{Q3}{\sigma_3^2} \qquad (4.3.18)$$

donde

$$Q2 = (1/N)\, J_N J_N' \otimes I_T - (1/NT)\, J_{NT} J_{NT}'$$

$$Q3 = (1/NT)\, J_{NT} J_{NT}'$$

$$\sigma 2^2 = \sigma e^2 + N \sigma 1^2$$

$$\sigma 3^2 = \sigma_e^2 + T \sigma u^2 + N \sigma 1^2$$

Particionando B y usando (4.3.18) tenemos, despues de algo de algebra,

$$Bs = [(1/\sigma 1^2)Xs'Q1\ Xs + (1/\sigma 2^2)Xs'Q2\ Xs + (1/\sigma e^2)Xs'Q\ Xs\ ]^{-1}$$

$$[(1/\sigma 1^2)Xs'Q1\ Xs\ Bs^* + (1/\sigma 2^2)Xs'Q2\ Xs\ Bs^{**} + (1/\sigma e^2)Xs'Q\ Xs\ bs] \qquad (4.3.19)$$

y $$\hat{\bar{B1}} = \bar{Y}_{..} - \sum_{k=2}^{K} \bar{X}k_{..}\, Bk \qquad (4.3.20)$$

donde

$Bs^* = (Xs'\ Q1\ Xs)^{-1}\ Xs'\ Q1\ Y$ utiliza la variacion ENTRE los individuos

$Bs^{**} = (Xs'\ Q2\ Xs)^{-1}\ Xs'\ Q2\ Y$ utiliza la variacion en el tiempo en cada individuo

bs es el estimador de var. artificiales y usa la variacion no explicada por las diferencias en individuos o periodos de tiempo

PREDICCION DE COMPONENTES ALEATORIOS.

Los predictores BLUE para los coeficientes aleatorios son:

$$\hat{u}_i = \frac{T \sigma u^2}{\sigma 1^2} (\bar{Y}_{i.} - \hat{\bar{B1}} - \sum_{k=2}^{K} Bk\ \bar{X}k_{i.}) \qquad (4.3.21)$$

y $$\hat{l} = \frac{N \sigma 1^2}{\sigma 2^2} (\bar{Y}_{.t} - \hat{\bar{B1}} - \sum_{k=2}^{K} Bk\ \bar{X}k_{.t}) \qquad (4.3.22)$$

El primer predictor da informacion sobre el futuro comportamiento de los individuos y parece mas interesante que la prediccion de l_t que informa sobre realizaciones pasadas.

ESTIMACION DE LOS COMPONENTES DE VARIANZA.

Los estimadores de B y los predictores de u y l dependen de las varianzas desconocidas σe^2, σu^2 y σl^2. Para reemplazarlas podemos usar los estimadores del apartado anterior adaptandolos para incluir la variacion del tiempo. Entonces,

$$\hat{\sigma}1^2 = \frac{e*' \, e*}{N-K} \tag{4.3.23}$$

$$\hat{\sigma}2^2 = \frac{eo' \, eo}{T-K} \tag{4.3.24}$$

$$\hat{\sigma}e^2 = \frac{\hat{e}' \, \hat{e}}{(N-1)(T-1)-(K-1)} \tag{4.3.25}$$

donde

$e* = Q1\, Y - Q1\, Xs\, Bs*$ son los residuales de $Bs*$

$eo = Q2\, Y - Q2\, Xs\, Bso$ son los residuales de Bso

$\hat{e} = Q\, Y - Q\, Xs\, bs$ son los residuales de bs

Para probar las hipotesis de $u = 0$, $l = 0$, o $u=l=0$ podemos usar el estimador de variables artificiales y la prueba F que compara las sumas de cuadrados restringidos y no restringidos.

Alternativamente una prueba asintotica para la prueba var(u) = var(l) = 0 viene dada por

$$g = \frac{N\,T}{2} \left\{ \left[\frac{1}{T-1} \left(\mathrm{SUM}_{i=1}^{N} \left(\mathrm{SUM}_{t=1}^{T} \hat{e}_{it} \right)^2 - 1 \right) \right]^2 + \frac{1}{N-1} \left[\mathrm{SUM}_{t=1}^{T} \left(\mathrm{SUM}_{i=1}^{N} \hat{e}_{it} \right)^2 - 1 \right]^2 \right\} \tag{4.3.26}$$

que se distribuye asintoticamente como una Chi^2 (2) donde $\hat{e}$ son los residuales de la regresion de Y en X.

Los supuestos de este modelo son muy restrictivos, pues implican:

1) Que la covarianza contemporanea entre dos individuos es la misma para cualquier par de individuos.

2) Que la covarianza entre dos observaciones de un individuo dado es constante en el tiempo y la misma para cada individuo.

4.3.3 ELECCION DE SUPUESTOS.

Como mencione en la seccion 4.2.3, el uso del modelo de variables artificiales puede considerarse como una inferencia condicional en las u_i y l_t en la muestra y por tanto sin importar las suposiciones que sobre ellas se hagan.

El modelo de componentes del error (MOCER) hace un conjunto especifico de supuestos acerca de los coeficientes aleatorios y, si son correctos, el estimador de M.C.G. sera mas eficiente. Pero si las especificaciones son incorrectas, este estimador resultara sesgado.

El punto decisivo es saber cuándo las u_i y las l_t están correlacionadas con las Xk_{it}. Si existe correlacion, entonces el estimador de MOVAR sera el mas apropiado. Si no existe correlacion, MOCER sera mas razonable.

Para decidir si existe o no correlacion, podemos probar la hipotesis nula de NO correlacion comparando los estimadores de MOVAR y MOCER.

Si la hipotesis nula es correcta, entonces

$$m = (bs - \hat{B}s)' \, (M1 - M0)^{-1} \, (bs - \hat{B}s)$$

se distribuye asintoticamente como una $Chi^2 (K-1)$

donde

$$M1 = \sigma e^2 \, (Xs' \, Q \, Xs)^{-1} \quad \text{es la matriz de covarianza para bs}$$

$$M0 = (\sigma 1^2 \, (Xs' \, D1 \, Xs) + \sigma 2^2 \, (Xs' \, D2 \, Xs) + \sigma e^2 \, (Xs' \, D \, Xs) \,)^{-1}$$

es la matriz de covarianza de $\hat{B}s$.

Las varianzas desconocidas en esta prueba pueden ser reemplazadas por sus estimadores sin cambiar la distribucion.

4.4 MODELOS DE PENDIENTE VARIABLE SOBRE LOS INDIVIDUOS.

Es posible que la diferencia en el comportamiento de los individuos se manifieste no solo en el termino de interseccion, sino tambien en los coeficientes de la pendiente. Si este es el caso, nuestro modelo podria ser:

$$Y_{it} = \sum_{k=1}^{K} Xk_{it}\, Bk_{i} + e_{it} \qquad \begin{matrix} i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, T \end{matrix} \qquad (4.4.1)$$

Este modelo nos hace suponer que el efecto en la variable dependiente (Y) al cambiar las variables independientes es distinto para cada individuo, pero en cada uno de ellos permanece constante en el tiempo.

Tambien aqui podremos distringuir dos modelos, correspondientes a cuando los Bk_i son fijos o son aleatorios.

4.4.1 MODELO DE "REGRESIONES APARENTEMENTE NO RELACIONADAS".

(MORAN)

Si los coeficientes son tratados como parametros fijos, la ecuacion (4.4.1) puede expresarse en general:

$$Y = Z R + E \qquad (4.4.2)$$

O bien,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 & & & & \\ & X_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & X_N \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ B_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_N \end{bmatrix} \qquad (4.4.3)$$

Asumiremos que

$$E[e_i] = 0 \quad y \quad E[e_i e_i'] = \sigma_{ii} I_T$$

De otra forma, si la covarianza fuera cero, podriamos resolver (4.4.3) por M.C.O. y el estimador seria BLUE. Tenemos entonces:

$$E[\,E\,E'\,] = \Theta = \Sigma \otimes I_T$$

donde

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix} \qquad (4.4.4)$$

Podemos ahora estimar R por M.C.G.

$$\hat{R} = (Z'\Theta^{-1}Z)^{-1} Z'\Theta^{-1}Y = (Z'(\Sigma^{-1} \otimes I_T)Z)^{-1} Z'(\Sigma^{-1} \otimes I_T)Y \qquad (4.4.5)$$

El estimador $\hat{R}$ es, pues, BLUE y tiene por matriz de covarianza

$$E[(\hat{R} - R)(\hat{R} - R)'] = (Z'\Theta^{-1}Z)^{-1} = (Z'(\Sigma^{-1} \otimes I_T)Z)^{-1}$$

Si aplicaramos M.C.O. a un individuo determinado, el estimador b_i de ese individuo sera tambien BLUE. Sin embargo podemos pensar en una clase de estimadores mas amplia, digamos la de estimadores insesgados. Dentro de esta clase, el estimador de B_i, el i-esimo subvector del estimador de R, es mejor que b_i porque toma en cuenta la correlacion entre e_i y otros vectores de error (recordemos que su covarianza es no nula) y tambien porque usa informacion de las variables explicatorias que esta oculta en el sistema y que no existe para la i-esima ecuacion individual. Esta es la razon por la cual Zellner en 1962 llamo a este modelo "Regresiones aparentemente no relacionadas". Solo cuando $_{ij}$=0 para toda i≠j, o cuando $X_1 = X_2 = \ldots = X_N$ los estimadores b_i y B_i seran identicos y no se ganara en eficiencia.

Generalmente desconocemos Σ por lo que debemos estimarla basandonos en los residuales de M.C.O.,

$$\hat{\sigma}_{ij} = (1/T)\ \hat{e}_i{}'\hat{e}_j \qquad i, j = 1, 2, \ldots, N \qquad (4.4.6)$$

donde

$$\hat{e}_i = Y_i - X_i b_i$$

Por tanto,

$$\hat{R} = (Z'(\hat{\Sigma}^{-1} \otimes I_T)Z)\ Z'(\hat{\Sigma}^{-1} \otimes I_T)\ Y \qquad (4.4.7)$$

es el estimador BLUE de R y se conoce como el estimador de MORAN.

PRUEBAS DE HIPOTESIS.

Como vimos en la seccion 4.1 interesa conocer si es o no significativa la informacion que nos proporciona la matriz de covarianza de los errores. La solucion alternativa sera considerar la restriccion de que los parametros son constantes para todos los individuos, incluso bajo el esquema de autocorrelacion visto en dicha seccion.

Nuevamente usaremos un estadistico que relaciona los residuales restringidos y no restringidos. Si probamos que la estimacion a partir de MORAN reduce la suma de los cuadrados de los residuales con respecto al modelo COCO, habremos probado que la informacion aportada por la matriz de covarianza es significativa, y por tanto rechazaremos la hipotesis de que los coeficientes son constantes.

Para ello nos ayudaremos del estadistico

$$F = \frac{((Y-XB)'\,\Omega^{-1}\,(Y-XB) - (Y-ZR)'\Theta^{-1}\,(Y-ZR))\ /\ (NK-K-N)}{(Y-ZR)'\Theta^{-1}\,(Y-ZR)\ /\ (NT-NK-N(N+1)/2)} \qquad (4.4.8)$$

que se distribuye como una F con (NK-K-N, NT-NK-N(N+1)/2) grados de libertad, cuya deduccion es igual que en la prueba (4.1.7).

4.4.2 MODELO SWAMY DE COEFICIENTES ALEATORIOS.

(MOSCA)

Cuando el vector de coeficientes de cada individuo B_i es considerardo como un vector aleatorio extraido de una distribucion de probabilidad con media $\bar{B}$ y matriz de covarianza CV, podemos escribir el modelo para el i-esimo individuo asi

$$Y_i = X_i B_i + e_i \qquad i = 1, 2, \ldots , N \qquad (4.4.9)$$

donde

$$B_i = \bar{B} + u_i$$

$$E[u_i] = 0 , \quad E[u_i u_i] = CV \quad y \quad E[u_i u_j] = 0 \quad para\ i \neq j$$

Hemos asumido en los modelos anteriores diferentes suposiciones sobre el error "e" y cualquiera de ellas puede aplicarse al presente, pero consideremos ahora las mas sencillas, esto es, que

$$E[e_i] = 0 , \quad E[e_i e'_i] = \sigma^2 I \quad y \quad E[e_i e'_j] = 0 \quad para\ i \neq j$$

Esto implica:

1) Que los errores entre los individuos son heterocedasticos pero no correlacionados.

2) Que si los B_i fueran parametros fijos, entonces el estimador de M.C.O. b_i seria BLUE ya que no habria correlacion entre individuos, es decir, cada modelo seria independiente.

ESTIMACION POR MINIMOS CUADRADOS GENERALIZADOS.

En terminos generales podemos escribir:

$$Y = X\bar{B} + Z\underline{u} + E \qquad (4.4.10)$$

donde

$$X = (X_1, X_2, \ldots, X_N)'$$

$$\underline{u} = (u_1, u_2, \ldots, u_N)'$$

Z esta definida como en (4.4.2)

La matriz de covarianza para el error compuesto es:

$$\Phi = E[(Z\underline{u} + E)(Z\underline{u} + E)] \qquad (4.4.11)$$

que es una matriz diagonal en bloque cuyo elemento i es

$$\Phi_{ii} = X_i CV X_i' + \sigma_{ii} I_T \qquad (4.4.12)$$

El estimador de M.C.G. de $\bar{B}$, con sus propiedades usuales, seria

$$\hat{\bar{B}} = (X' \Phi^{-1} X)^{-1} X' \Phi^{-1} Y$$

$$= (\sum_{j=1}^{N} X_j' \Phi_{jj}^{-1} X_j) \sum_{i=1}^{N} X_i' \Phi_{ii}^{-1} Y_i$$

$$= \sum_{i=1}^{N} W_i b_i \qquad (4.4.13)$$

donde

$$W_i = (\sum_{j=1}^{N} (CV + \sigma_{jj}(X_j'X_j)^{-1})^{-1})^{-1} (CV + \sigma_{ii}(X_i'X_i)^{-1})^{-1}$$

y

$$b_i = (X_i' X_i)^{-1} X_i' Y_i$$

La ecuacion (4.4.13) nos muestra como el estimador $\hat{\bar{B}}$ es una matriz ponderada de los estimadores b_i con los pesos inversamente proporcionados a sus matrices de covarianza.

La matriz de covarianza de $\hat{\bar{B}}$ es:

$$(X'\Phi^{-1}X)^{-1} = (\mathrm{SUM}_{j=1}^{N} X_j'\ \phi_{jj}^{-1}\ X_j)^{-1}$$

$$= (\mathrm{SUM}_{j=1}^{N} (CV + \sigma_{jj}(X_j'X_j)^{-1})^{-1})^{-1}$$

ESTIMACION DE LAS VARIANZAS.

Como en los modelos anteriores, aqui tambien requerimos estimar las varianzas CV y σ_{ii} para hacer operativos los estimadores. Swamy demostro que podemos usar los residuales de M.C. para estimarlas insesgadamente:

$$\hat{\sigma}_{ii} = (1/\ T\text{-}K)\ \hat{e}_i\ \hat{e}_i'$$

y

$$\hat{CV} = Sb\ /\ (N-1)\ - (1/N)\ \mathrm{SUM}_{i=1}^{N} \hat{\sigma}_{ii}\ (X_i'X_i)^{-1} \qquad (4.4.14)$$

donde

$$Sb = \mathrm{SUM}_{i=1}^{N} b_i b_i' \ - (1/N)\ \mathrm{SUM}_{i=1}^{N} b_i\ \mathrm{SUM}_{i=1}^{N} b_i'$$

Una dificultad con el estimador de CV es que debe ser no-negativo definido. Swamy sugiere metodos para corregirlo, pero a costa de eliminar la propiedad de insesgo. La solucion mas sencilla parece ser entonces aceptar el estimador sesgado CV^ = Sb / (N-1) que es un estimador no-negativo definido y es consistente respecto a los incrementos en T.

De gran interes resulta ahora la prueba de la hipotesis nula:

$$Ho:\quad B_1 = B_2 = \ldots = B_N = \bar{B}$$

Si resulta cierta, el vector de coeficientes individuales no son aleatorios y son identicos a la media. Entonces el modelo conjunto (4.4.9) resulta ser de errores heterocedasticos, donde las varianzas son constantes dentro de los subgrupos de observaciones.

El estadistico

$$g = \mathrm{SUM}_{i=1}^{N} \frac{(b_i - \hat{B})'\ X_i'X_i\ (b_i - \hat{B})}{\hat{\sigma}_{ii}} \qquad (4.4.15)$$

donde

$$\hat{B} = (\mathrm{SUM}_{i=1}^{N} \sigma_{ii}^{-1} X_i'X_i)^{-1}\ \mathrm{SUM}_{i=1}^{N} \sigma_{ii}^{-1} X_i'X_i\ b_i$$

tiene una distribucion asintotica Chi²(K(N-1)) bajo Ho. Puede hacerse esta misma prueba cuando las B_i son consideradas fijas y diferentes.

4.4.3 ELECCION DE SUPUESTOS.

La discusion corre paralela al caso de pendiente constante e interseccion variable. Cuando existe correlacion entre los coeficientes y las variables explicatorias o independientes, parece mas razonable usar MORAN, en tanto que si no hay correlacion o no se puede especificar en el modelo, conviene usar MOSCA.

La prueba de hipotesis sobre la NO correlacion ha sido estudiada por Pudney, basandose en la covarianza entre el estimador de M.C.O. de cada individuo y las medias de las variables explicatorias de cada uno de los individuos:

$$\bar{X}_i' = J_T' X_i / T$$

La covarianza muestral esta dada por

$$SXb = 1/N \left(\sum_{i=1}^{N} b_i \bar{X}_i' - 1/N \sum_{i=1}^{N} b_i \sum_{i=1}^{N} \bar{X}_i' \right) \qquad (4.4.16)$$

y el estadistico para la prueba es:

$$z = N \eta' V^{-1} \eta$$

donde

η es el vector de (SXb) para todas las i.

$$V = 1/N \sum_{i=1}^{N} (\bar{X}_i \otimes I_K) \left(\hat{CV} + \sigma e^2 (X_i' X_i)^{-1} \right) (\bar{X}_i' \otimes I_K)$$

4.5 MODELOS DE PENDIENTE VARIABLE SOBRE EL TIEMPO Y LOS INDIVIDUOS.

En este modelo todos los coeficientes tienen una aportacion especifica a cada individuo y en cada periodo de tiempo. Podemos escribirlo de la siguiente manera

$$Y_{it} = \sum_{k=1}^{K} (\bar{B}k + u_i + l_t) Xk_{it} + e_{it} \qquad i = 1, 2, \ldots, N \quad t = 1, 2, \ldots, T \qquad (4.5.1)$$

donde
$$Bk_{it} = \bar{Bk} + u_i + l_t$$

En este modelo, como en los anteriores, podemos considerar los parametros como fijos o aleatorios. Generalmente se pueden cosiderar aleatorios pero algun caso requiere que sean fijos. O bien podemos considerar que uno de ellos es fijo y el otro aleatorio como puede ser el caso cuando se tienen pocas observaciones respecto al tiempo.

Si en la ecuacion (4.5.1) sustituimos el efecto del tiempo con variables artificiales, tenemos exactamente el modelo anterior (MOSCA).

4.5.1 MODELO HSIAO DE COEFICIENTES ALEATORIOS. (MOHCA)

La estimacion y pruebas de hipotesis del modelo (4.5.1) lo debemos a Hsiao, que para el i-esimo individuo escribe la ecuacion asi,

$$Y_i = X_i \bar{B} + X_i \underline{u}_i + \bar{Z}_i \underline{l} + e_i \qquad (4.5.2)$$

donde

Y_i es de dimension (t,1) y X_i de (T,K)

$\underline{u}_i = (u1_i, u2_i, \ldots, uK_i)'$

$\underline{l} = (l_1, l_2, \ldots, l_T)'$

$l_t = (l1_t, l2_t, \ldots, lK_t)'$

$e_i = (e_{i1}, e_{i2}, \ldots, e_{iT})'$

$X_{it}' = (X1_{it}, X2_{it}, \ldots, XK_{it})$

y

$$\bar{Z}_i = \begin{bmatrix} X_{i1}' & & & & \\ & X_{i2}' & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & X_{iT}' \end{bmatrix}$$

Los supuestos que propone Hsiao son:

$E[e_i] = 0$, $\quad E[e_i e_i'] = \sigma e^2 I \quad$ $E[e_i e_j'] = 0$ para $i \neq j$

$E[u_i] = 0$, $\quad E[u_i u_i'] = \delta k\, I = CD \quad$ $E[u_i u_j'] = 0$ para $i \neq j$

$E[l_t] = 0$, $\quad E[l_t l_t'] = \alpha k\, I = A \quad$ $E[l_t l_s'] = 0$ para $t \neq s$

y entre ellos no esta correlacionados.

Al incluir todas las observaciones, tenemos

$$Y = X\underline{B} + Z\underline{u} + \bar{Z}\,l + E \qquad (4.5.3)$$

con las definiciones conocidas.

Si $\underline{u}$ y $\underline{l}$ son considerados fijos, $\underline{B}$ $\underline{u}$ y $\underline{l}$ pueden estimarse aplicando M.C.O. a (4.5.3), aunque habiendo reparametrizado corrigiendo por su media y tomando en cuenta que NT es suficientemente grande.

La matriz $(X,Z,\bar{Z})$ es de orden (NT, (N+T+1)K) y de rango (N+T-1)K por lo que hay 2 K parametros redundantes. Para una conveniente estimacion convendria eliminar

$$(u1_N\,,\ u2_N\,,\ \ldots\ ,\ uK_N\,,\ l1_T\,,\ l2_T\,,\ \ldots\ ,\ lK_T)$$

eliminar tambien sus columnas correspondientes en Z y $\bar{Z}$ y redefinir los otros parametros. En todo caso debemos recordar que estamos asumiendo que $NT > (N+T-1)K$.

ESTIMACION DE VARIANZAS Y PREDICCION

Cuando u y l son aleatorias, la matriz de covarianza de los errores es

$$\Phi = E\,[(Z u + \bar{Z}\,l + E)\ (Z u + \bar{Z}\,l + E)']$$

$$= Z\,(\,I_N \otimes CD\,)\,Z' + \bar{Z}\,(I_T \otimes A\,)\,\bar{Z}\,' + \sigma_e^2\, I_{NT} \qquad (4.5.4)$$

y el estimador

$$\hat{\underline{B}} = (X'\,\Phi^{-1}\,X)^{-1}\,X'\,\Phi^{-1}\,Y \qquad (4.5.5)$$

es BLUE para estiamar $\underline{B}$ con matriz de covarianza $(X'\Phi^{-1}\,X)^{-1}$

Siendo Φ de orden (NT,NT) su inversion puede complicarse. Sin embargo, Hsiao muestra que Φ^{-1} puede escribirse asi

$$\Phi^{-1} = I - ZGZ' - \bar{Z}C\bar{Z}'(I-ZGZ')$$

$$+ (I + \bar{Z}C\bar{Z}')\ Z\ (Z'\bar{Z}C\bar{Z}'Z - G^{-1})^{-1}\ Z'\bar{Z}C\bar{Z}'\ (I - ZGZ') \qquad (4.5.6)$$

donde

$$C = (\bar{Z}'\bar{Z} + (I_T \otimes A^{-1}))^{-1}$$

$$G = (Z'Z + (I_N \otimes CD^{-1}))^{-1}$$

En esta expresion la inversion de mayor orden se reduce al MAX (NK,TK) pero esto puede ser aun algo grande.

Para predecir los componentes aleatorios para cada individuo, podemos usar el predictor BLUE

$$\hat{u} = (I_N \otimes CD)\, Z'\, \Phi^{-1}\, (Y - X\hat{\bar{B}}) \qquad (4.5.7)$$

ESTIMACION DE VARIANZAS.

Tanto $\hat{\bar{B}}$ como u, dependen de CD, A y σe^2 que son desconocidas. Para estimarlas podemos escribir la ecuacion (4.5.1) de esta manera

$$Y_{it} = \mathop{SUM}_{k=1}^{K} Bk_i\, Xk_{it} + v_{it}$$

donde

$$Bk_i = \overline{Bk} + uk_i$$

y

$$v_{it} = \mathop{SUM}_{k=1}^{K} 1k_t\, Xk_{it} + e_{it} \qquad (4.5.8)$$

con varianza

$$VR_{it} = E\,[v_{it}^2] = \mathop{SUM}_{k=1}^{K} Xk_{it}^2\, \alpha k + \sigma e^2$$

que en forma vectorial podemos escribir

$$\underline{VR}_i = X\%_i\, \underline{\alpha}$$

donde

$X\%$ es X_i con cada uno de sus elementos al cuadrado

$$\underline{\alpha} = (\alpha_1 + e^2, \alpha_2, \alpha_3, \ldots, \alpha K)'$$

Sea $\hat{e}_i = Y_i - X_i\, b_i$ el vector de residuales del estimador de M.C.O.

y

$\hat{e}\%_i$ el vector de los cuadrados de dichos residuales. Entonces,

$$E\,[\hat{e}\%_i] = M\%_i\, VR_i = F_i\, \alpha$$

donde

$M\%$ contiene los cuadrados de los elementos de $M = I - X_i\, (X_i' X_i)^{-1} X_i'$

y

$$F_i = M\%_i\, X_i$$

Repitiendo el proceso para todos los cortes transversales, tenemos

$$E[\,E\%\,] = F\alpha \qquad (4.5.9)$$

donde

$$E\% = (e\%_1{}', e\%_2{}', \ldots, e\%_N{}')$$

y

$$F = (F_1, F_2, \ldots, F_N)'$$

Aplicando M.C.O. a (4.5.9), obtenemos el estimador

$$\hat{\alpha} = (F'\,F)^{-1}\,F'\,E\%$$

Siguiendo el mismo procedimiento respecto a cada t, obtendremos el estimador insesgado $\hat{\delta}$ para $\delta = (\delta_1 + e^2, \delta_2, \delta_3, \ldots, \delta_K)'$

Un estimador consistente para σe^2 esta dado por

$$\hat{\sigma}_e^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{e\%_{it}\ e^*_{it}}{N\,T} \qquad (4.5.10)$$

donde

e^* son los residuales obtenidos al aplicar M.C.O. separadamente a cada periodo de tiempo

De $\hat{\alpha}$, $\hat{\delta}$ y $\hat{\sigma}_e^2$ se pueden derivar una estimacion para α_1 y δ_1.

En la practica, el modelo Hsiao es poco aceptado en parte por la complicacion de sus calculos y en parte tambien porque sus suspuestos parecen poco realistas. La principal objecion es considerar A y VD como diagonales, pero su eliminacion supone mayores complicaciones en el manejo computacional. Incluso, si existiesen el suficiente numero de observaciones en el tiempo, convendria considerar el efecto del tiempo como un problema de autocorrelacion en los errores.

4.6 ELECCION DE MODELO.

Cuando afrontamos un problema en la practica, nos encontramos con la dificultad de elegir el modelo mas adecuado. En otras palabras, debemos decidir cual es el conjunto de suposiciones mas factibles para nuestro problema. Esto obviamente no es sencillo pues no existe una evaluacion exacta para hacerlo, aunque si puede ayudarnos una guia:

1) ¿Los coeficientes de la pendiente parecen variar sobre los individuos, o es razonable captar la diferencia entre los individuos a traves de la interseccion o de un modelo apropiado del termino de error?

2) Si los coeficientes o la interseccion varian sobre los individuos, ¿las diferencias parecen depender de las variables independientes de los individuos? Si es así, MOVAR o MORAN parecen ser preferibles. Si no, entonces MOCER o MOSCA parecen ser los correctos.

3) Para modelar cambios sobre el tiempo es necesario ver qué metodo es el mejor:

a) Usar la estructura de correlacion constante del MOCER

b) Asumir que los errores siguen un proceso autorregresivo o de promedios moviles (COCO)

c) Usar variables artificiales y usar inferencia condicional para los cambios en la muestra (MOVAR)

4) ¿Cuántas observaciones hay? En modelos donde se asume que los parametros son aleatorios, los tamaños relativos de N y T tendran un importante influjo en la confiabilidad de la estimacion de varianzas para muestras finitas. Si, por ejemplo, N es pequeña, no parece que σ_u^2 para el MOCER, o la estimacion de CD para el MOSCA sean muy confiables. Consecuentemente, tampoco seran confiables las estimaciones hechas a partir de ahi y para su estimacion sera mas conveniente considerarlos como parametros fijos, aun cuando parezca razonable el supuesto de aleatoriedad.

CAPITULO 5

DEMANDA DE AUTOS POR CATEGORIAS.

A manera de ejemplo, trabajaremos con un modelo que estima la demanda de automoviles en Mexico, donde el corte transversal se hara sobre tres categorias de autos (compactos, medianos y de lujo) y las observaciones en el tiempo corresponden a 60 meses, de enero de 1980 a diciembre de 1984.

La estimacion de un modelo encierra bastantes complicaciones, primero en cuanto a determinar las variables que expliquen la variacion real de la variable independiente, luego en recopilar la informacion de cada variable y, finalmente, en procesar toda esta informacion mediante transformaciones para que se ajusten lo mas posible a la realidad.

En este capitulo presentaremos la especificacion del modelo, la definicion de cada una de las variables y, por ultimo, los resultados obtenidos al aplicar algunos modelos de los estudiados en el capitulo 4.

5.1 ESPECIFICACION DEL MODELO.

Todo modelo economico de demanda debe incluir entre sus variables explicatorias el precio del bien en estudio, el ingreso y el precio de los bienes sustitutos. Nosotros decidimos usar como bienes sustitutos el precio del transporte y la tasa de interes bancaria. Ademas de las variables anteriores, incluimos en nuestro modelo una medida del consumo de gasolina.

Muchas otras variables podrian incluirse en nuestro modelo y seguramente el lector pensara algunas de ellas. Pero recordemos que un modelo debe simplificar la realidad, y el efecto de toda aquella informacion no incluida en el mismo sera absorbida por el termino de error.

Asi, pues, podemos escribir para la categoria i:

$$Y_i = f\,(X2_i\,,\ X3_i\,,\ X4_i\,,\ X5_i\,,\ X6_i\,,\ X7_i\,) \qquad (5.1.1)$$

donde

Y_i = demanda de automoviles de la categoria i

$X2_i$ = precio de los automoviles de la categoria i

$X3_i$ = medida del consumo de gasolina de los autos de la categoria i

$X4_i$ = ingreso real (es igual para las tres categorias)

$X5_t$ = precio del transporte (es igual para las tres categorias)

$X6_t$ = tasa de interes activa (es igual para las tres categorias)

$X7_t$ = precio de los automoviles sustitutos

Una especificacion mas concreta del modelo la podremos hacer despues de analizar la definicion de cada una de las variables y las transformaciones que deben hacerse.

5.2 DEFINICION DE LAS VARIABLES.

5.2.1 DEMANDA DE AUTOMOVILES (Y).

¿Hasta donde es posible medir la demanda real de automoviles? La informacion disponible consiste, por un lado, en el volumen de produccion de automoviles por cilindrada (es decir, una medida de la oferta),segun la Direccion General de Estadistica de la Secretaria de Programacion y Presupuesto (S.P.P.), y por otro lado en las ventas mensuales de automoviles por marcas, proporcionada por la Asociacion Mexicana de la Industria Automotriz (A.M.I.A.).

¿Podemos identificar las ventas con la demanda real? El hecho de que en el mercado mexicano la oferta de automoviles es mayor que la demanda (la produccion es mayor que el consumo en el pais), hace admisible identificar las ventas de autos con la demanda.

Aqui conviene anotar que nuestra idea inicial era hacer el corte transversal no por categorias sino por cilindrada, de manera que tendriamos tambien tres tipos de autos: de 4, de 6 y de 8 cilindros. Sin embargo el mercado de autos en Mexico ha venido sufriendo modificaciones en su estructura en los ultimos años. Asi, los modelos Chrysler que antes eran de 6 u 8 cilindros, pasaron a ser de 4 cilindros con la nueva serie de modelos "K" a partir de enero de 1983. De manera similar, el mercado de autos de 8 cilindros desaparece a partir de noviembre de 1984, incrementando la variedad de automoviles de 4 y 6 cilindros. De acuerdo a estos hechos, un modelo que estimara la demanda de automoviles por cilindrada estaria fuera de nuestros objetivos ya que no nos permitiria un estudio confiable en el corte transversal.

Elegimos entonces hacer el corte transversal sobre la categoria de los automoviles (compactos, medianos y de lujo) porque esta clasificacion es semejante al de cilindrada antes del cambio de estructura del mercado, y porque despues del cambio se conservaban las categorias, en cuanto que los coches de lujo seguian siendo de lujo a pesar de tener ahora 4 o 6 cilindros en lugar de 8 que tenian.

Una vez establecidas las categorias, definimos: (5.2.1)

Y_{it} = demanda de automoviles del tipo i en el mes t

donde

Y_{it} esta en unidades vendidas
i = 1, 2, 3 (1 = compactos, 2 = medianos, 3 = de lujo)
t = 1, 2, ... , 60

5.2.2 PRECIO DE LOS AUTOMOVILES (X2).

Definir el precio de los automoviles constituyo uno de los problemas mas serios al especificar nuestro modelo, porque no contamos con las informacion requerida.

La informacion que disponiamos era el volumen de produccion y el costo total de la produccion segun la S.P.P., y tambien con el indice de precios al consumidor para los automoviles en general proporcionado por el Banco de Mexico.

Por otra parte la necesidad de involucrar los bienes sustitutos hacia necesario incluir no solo el precio de los automoviles en cuestion (el precio de los automoviles tipo i explicando la venta de automoviles del mismo tipo), sino tambien el precio de los otros autos (en (5.1.1) la variable X7).

Varias soluciones fueron estudiadas:

1) Costos unitarios.

Podriamos utilizar la variable "costo unitario" (costo unitario = costo total / volumen de produccion) como una aproximacion del precio real, bajo la hipotesis de que la diferencia entre ambos seria una constante.

El modelo involucraria tres variables correspondientes al costo unitario de cada categoria, y se incluirian las tres para cumplir con el requisito de los bienes sustitutos. Sin embargo, al tener la misma tendencia, podria ocasionar multicolinealidad.

2) Indice de precios por categoria.

Teniamos el indice de precios de automoviles en general, pero no para cada categoria. Por otro lado teniamos el costo unitario de cada categoria. Mediante un metodo estadistico conocido por el "Metodo de Ginsburgh"[4], pudimos "desagregar" el indice de precios general en un indice de precios para cada categoria, asociando sus comportamientos al de las variables de costo unitario.

Nota 4) Sobre el Metodo de Ginsburgh puede consultarse "A further note on the derivation of quarterly figures consistent with annual data", Ginsburgh V.A.,1973 (J.R.S.S., 368-378)

Este metodo tenia el grave inconveniente de que, al ser indice de precios, las tres variables de indices generadas para cada categoria, aunque distintas, eran muy semejantes y no reflejaban las diferencias de precios entre las categorias. Por tanto no seria util para comparar precios, y usar las tres para incluir bienes sustitutos crearia un problema grande de multicolinealidad por tener la misma tendencia.

3) Proporcion de precios.

Teniendo el costo unitario de cada categoria, podemos establecer una variable que mida la proporcion del costo de cada categoria. De acuerdo a la hipotesis mencionada de que la diferencia entre el costo y el precio es una constante, entonces la proporcion en costos sera igual a la proporcion en precio.

Asi, pues, decimos:

$$\text{Proporcion de precio tipo } i = \frac{\text{costo unitario tipo } i}{\sum_j \text{costo unitario tipo } j} \times 100.0$$

Esta sola variable representa el peso que tiene el precio de los autos del tipo i con respecto al precio de otros autos, por lo que en esta opcion no necesitamos involucrar las otras variables (en concreto ya no se necesita la variable X7) como bienes sustitutos, evitando asi el problema de multicolinealidad ocasionado por los otros metodos, y especialmente por el segundo.

Optamos, por fin, por el tercer metodo, y definimos: (5.2.2)

$X2_{it}$ = proporcion de precio de los automoviles tipo i en el mes t

donde

$X2_{it}$ esta en porcentaje
i = 1, 2, 3 (como definimos en 5.2.1)
t = 1, 2, ... , 60

5.2.3 CONSUMO DE GASOLINA (X3).

Con el precio de la gasolina en terminos nominales obtenidos de PEMEX y el consumo promedio de gasolina de cada categoria, segun fuentes de las compañias Chrysler y Ford, obtuvimos una medida del consumo nominal en pesos:

$$\text{Consumo nominal categoria } i = \frac{\text{Precio de la gasolina}}{\text{Rendimiento en kms./litro cat. } i}$$

donde el rendimiento de cada categoria en promedio es:

Autos compactos ----> 9 kms/litro
Autos medianos ----> 7 kms/litro
Autos de lujo ----> 5 kms/litro

Falta ponerlo ahora en terminos reales, lo que hacemos dividiendo entre el deflactor:

$$\text{Consumo real categoria i} = \frac{\text{Precio de la gasolina} \times 100.0}{\text{Rendimiento cat. i} \times \text{Deflactor}}$$

Por tanto, definimos nuestra tercera variable de esta manera: (5.2.3)

$X3_{it}$ = consumo real de gasolina para los autos tipo i en el mes t

donde

$X3_{it}$ esta en pesos
$i = 1, 2, 3$ (como definimos en 5.2.1)
$t = 1, 2, \ldots, 60$

5.2.4 INGRESO (X4).

Como una medida del ingreso real, decidimos usar el Producto Interno Bruto Real (PIBR). Sin embargo aqui tambien contabamos con la informacion anual del PIBR proporcionada por el Banco de Mexico y fue necesario "desagregarla" mensualmente por el ya mencionado Metodo de Ginsburgh, asociandola al comportamiento del indice de volumen del sector industrial que el Banco de Mexico si publica mensualmente.

Ademas el PIBR estaba en base 1978, por lo que fue necesaria una sencilla transformacion para ponerla en base 1980 y lograr asi la uniformidad de bases con las otras variables. La trasformacion consistio en dividir cada termino del PIBR entre el promedio del PIBR de 1978, es decir,

$$\text{PIBR base 1980} = \frac{\text{PIBR base 1978}}{43{,}358.8917} \times 100.0$$

Finalmente, definimos: (5.2.4)

$X4_{it}$ = ingreso real en el mes t, igual para toda i

donde

$X4_{it}$ esta en miles de millones de pesos
$i = 1, 2, 3$
$t = 1, 2, \ldots, 60$

5.2.5 PRECIO DEL TRANSPORTE (X5).

El precio del transporte, como bien sustituto, fue obtenido a partir del indice de precios al consumidor, base 1980, de acuerdo al Banco de Mexico. Como podemos observar, las variables hasta ahora definidas (ventas, precio, consumo e ingreso) estan en terminos reales, por lo que nos interesa trabajar con la variacion de los indices del transporte para conocer tambien el impacto real de los precios del transporte, por lo que redefinimos nuestra variable sacando sus diferencias, es decir,

$$\text{Variacion} = \frac{\text{Indice del transporte en el mes } t}{\text{Indice del transporte en el mes } t-1}$$

Por tanto, la quinta variable se define asi: (5.2.5)

$X5_{it}$ = variacion en el indice de precios al consumidor del transporte en Mexico

donde

i = 1, 2, 3 y es igual para toda i
t = 1, 2, ... , 60

5.2.6 TASA DE INTERES ACTIVA (X6).

Considerando el dinero como un bien sustituto, tomamos como el costo del mismo la tasa de interes, y utlizamos la conocida como el "Costo Porcentual Promedio" (C.P.P.) que es la tasa activa que publica mensualmente el Banco de Mexico.

Al encontrar esta variable en terminos nominales, fue necesario ponerla en terminos reales, transformandola esta vez por la siguiente formula:

$$\text{C.P.P. Real } (t) = \frac{\dfrac{\text{C.P.P. Nominal } (t)}{12} - \text{Inflacion } (t)}{1.0 + \text{Inflacion } (t)}$$

Definimos, entonces, nuestra ultima variable de esta manera: (5.2.6)

$X6it$ = C.P.P. Real para el auto tipo i en el mes t

donde

i = 1, 2, 3 y es igual para toda i
t = 1, 2, ... , 60

Finalmente, podemos escribir, de una manera mas explicita, nuestro modelo:

$$Y_{it} = Bo + B1\ X1_{it} + B2\ X2_{it} + B3\ X3_{it} + B4\ X4_{it} + B5\ X5_{it} + B6\ X6_{it} + e_{it}$$

donde

$i = 1, 2, 3$
$t = 1, 2, \ldots, 60$
e_{it} es el termino de error para la categoria i y el mes t

5.3 ESTIMACIONES.

Todas las estimaciones realizadas en este capitulo fueron hechas con el paquete computacional de metodos econometricos conocido como "SHAZAM" (Cfr. Bibliografia), utilizando una computadora VAX/1100 que generosamente presto el centro de "Estudios Economicos BANAMEX". En el apendice B aparecen los programas utilizados con este paquete, asi como los programas en FORTRAN que generaron las tablas de informacion y los programas tambien en FORTRAN que generan las matrices de transformacion definidas en las secciones 4.2.2 y 4.3.2 .

5.3.1 EL PROBLEMA DE LA MULTICOLINEALIDAD.

Al comenzar las estimaciones de nuestro modelo es logico pensar en la presencia de autocorrelacion y heterocedasticidad por las razones ya conocidas. Pero cabe preguntarnos tambien si existe el problema de multicolinealidad, es decir, si existe una alta correlacion entre nuestras variables independientes. Lo primero que haremos sera discutir este asunto.

Para detectar multicolinealidad (y de hecho tambien para corregirla) existe el metodo de "Componentes Principales"[5] que consiste en reducir la matriz de correlacion (la matriz formada por la correlacion entre las K-1 variables independientes) a una matriz de rango menor a traves de un proceso que elimina tantas variables cuantas esten en dependencia de las otras.

NOTA 5) Para una explicacion mas detallada sobre el metodo de "Componentes Principales" puede consultarse como referencia el texto de J. Johnston, "Econometric Methods", pp. 321-331.

Dentro del proceso se involucran las raices caracteristicas de la matriz de correlacion. Como vimos en el capitulo 2, si hay dependencia lineal cuando menos un valor caracteristico sera cero, mientras que si hay independencia perfecta las raices caracteristicas seran todas iguales a uno (recordar que la matriz de correlacion es una matriz positiva definida), de manera que analizando dichos valores podremos saber si existe o no multicolinealidad. La cuestion esta en determinar que tan cercanos a cero deben ser las raices caracteristicas para aceptar la multicolinealidad, pero aqui la experiencia determina el criterio.

Asi, pues, aplicamos el metodo de Componentes Principales a las matrices de correlacion correspondientes a las tres categorias de automoviles, y obtuvimos los siguientes resultados:

1) Autos compactos:

Raices caracteristicas	1.9003	1.5881	0.95738	0.35636	0.19783
Porcentaje acumulado	0.38007	0.69769	0.88916	0.96043	1.00000

2) Autos medianos:

Raices caracteristicas	2.0267	1.6623	0.71151	0.40018	0.19937
Porcentaje acumulado	0.4053	0.73779	0.88009	0.96013	1.00000

3) Autos de lujo:

Raices caracteristicas	1.6620	1.2049	0.99766	0.78589	0.34948
Porcentaje acumulado	0.33241	0.57339	0.77293	0.93010	1.00000

De acuerdo a estos resultados, se admite la existencia de una cierta dependencia en las variables explicatorias de nuestro modelo.

5.3.2 REGRESIONES CONJUNTA Y POR CATEGORIAS.

Pasando ya a las regresiones del modelo, la primera que debemos hacer sera la regresion del modelo conjunto, es decir, aquella que supone los coeficientes constantes y la NO existencia de autocorrelacion y heterocedasticidad. Segun (4.1.3), obtendriamos por M.C.O.

$$b = (X' X)^{-1} X' Y$$

donde X es la matriz que resulta de considerar la informacion de las tres categorias como una sola matriz de informacion.

Los resultados de esta primera regresion son los siguientes:

REGRESION CONJUNTA:

1) R-cuadrada = 0.6626

2) Prueba F = 67.95 con (5,173) grados de libertad

3) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	-259.48	30.397	-8.5362
CONSUMO	-7933.80	1067.300	-7.4333
TRANSPORTE	936.88	2098.200	0.4465
INGRESO	92.66	39.215	2.3628
C.P.P.REAL	-1879.00	897.690	-2.0931
INTERSECCION	9910.80	4668.400	2.1229

4) Suma de los cuadrados de los residuales = 0.13490 E+10

5) Durbin-Watson = 0.6658 lo cual deja ver claramente la existencia de autocorrelacion.

El siguiente paso fue correr la regresion para cada una de las tres categorias por separado, usando el mismo estimador. Los resultados son los siguientes:

REGRESION INDIVIDUAL PARA AUTOS COMPACTOS:

1) R-cuadrada = 0.4762

2) Prueba F = 9.637 con (5,53) grados de libertad

3) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	-173.88	127.730	-1.3614
CONSUMO	-11474.00	2497.300	-4.5944
TRANSPORTE	-772.22	9252.600	-0.0835
INGRESO	135.31	59.603	2.2702
C.P.P.REAL	-5131.10	1529.700	-3.3543
INTERSECCION	8362.40	12503.000	0.6688

4) Suma de los cuadrados de los residuales = 0.25993 E+9

5) Durbin-Watson = 2.0244

REGRESION INDIVIDUAL PARA AUTOS MEDIANOS:

1) R-cuadrada = 0.7111

2) Prueba F = 29.549 con (5,53) grados de libertad

3) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	-16.15	53.839	-0.2841
CONSUMO	-11325.00	1268.500	-8.9280
TRANSPORTE	-2079.90	5881.500	-0.3536
INGRESO	110.43	39.018	2.8302
C.P.P.REAL	-1522.80	994.240	-1.5317
INTERSECCION	3676.70	7790.100	0.4720

4) Suma de los cuadrados de los residuales = 0.11112 E+9

5) Durbin-Watson = 1.3815

REGRESION INDIVIDUAL PARA AUTOS DE LUJO:

1) R-cuadrada = 0.2632

2) Prueba F = 5.572 con (5,53) grados de libertad

3) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	16.69	25.373	0.6577
CONSUMO	-1847.90	375.120	-4.9262
TRANSPORTE	1051.00	3712.100	0.2831
INGRESO	-9395.20	4709.800	-1.9948
C.P.P.REAL	540.03	638.850	0.1094
INTERSECCION	2738.60	3794.500	0.7217

4) Suma de los cuadrados de los residuales = 0.50961 E+8

5) Durbin-Watson = 1.6084

5.3.3 COEFICIENTES CONSTANTES (COCO).

En los resultados anteriores, podemos ver claramente la presencia de autocorrelacion tanto en el modelo conjunto como en las estimaciones individuales para las categorias de autos medianos y de lujo. Por tanto se antoja logico proceder a analizar el modelo conjunto (que sostiene que los coeficientes son constantes) bajo el supuesto de autocorrelacion en los errores, que es justamente el modelo COCO que estudiamos en la seccion 4.1 Utilizando el estimador (4.1.5), obtuvimos:

1) R-cuadrada = 0.8293 , que es una notable mejoria.

2) Los coeficientes de autocorrelacion de las tres categorias son:

$\rho 1 = 0.06$ $\rho 2 = 0.98$ $\rho 3 = 0.66$

3) Sus pruebas t de significancia correspondientes son:

t1 = 0.4577 t2 = 37.82 t3 = 6.74802

4) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	-32.65	40.831	-0.7996
CONSUMO	-5379.00	1505.700	-3.5723
TRANSPORTE	517.23	1246.500	0.4149
INGRESO	142.58	37.853	3.7670
C.P.P.REAL	-1441.20	784.870	-1.8363
INTERSECCION	-4418.80	4375.000	-1.0100

5) Suma de los cuadrados de los residuales = 0.66350 E+9

Estos resultados nos muestran la presencia real de autocorrelacion para los autos medianos y los de lujo, pero no en los compactos. Tambien se reduce mucho la suma de cuadrados de los residuales lo cual favorece la siguiente prueba.

PRUEBA DE HIPOTESIS.

De acuerdo con el estadistico (4.1.7), podemos probar que esta estimacion del modelo COCO es mejor que la del modelo conjunto inicial. Bajo la hipotesis nula de que el modelo conjunto inicial es el indicado, estimamos:

$$F = \frac{(0.13490\ E{+}10\ -\ 0.66350\ E{+}9)\ /\ (3 + 3(4)/2)}{0.66350\ E{+}9\ /\ (180 - 6 - 3 - 3(4)/2)} = 18.94$$

y el valor critico de F con (9,165) grados de libertad es 2.56 .

Como resulta altamente significativa, rechazamos la hipotesis nula y aceptamos la alternativa que nos indica que el modelo COCO resulta mas adecuado.

Es importante hacer notar algunas diferencias sobre los dos modelos aqui comparados. En primer lugar la R^2 en el modelo COCO es mayor lo que nos dice que explica en mayor proporcion la varianza de la variable dependiente. En segundo lugar notamos el cambio en los coeficientes, pero sobre todo en la significancia de los mismos (de acuerdo con la prueba t); asi la variable "porcentaje del precio" deja de ser significativa y la "tasa de interes" es ahora apenas significativa; el termino de interseccion cambia de signo y deja de ser significativo. Sin embargo el resto de las variables conserva su significancia y sus signos son los esperados, aunque de ello hablaremos mas tarde, en las conclusiones.

5.3.4 MODELOS DE INTERSECCION VARIABLE.

Nos debemos preguntar ahora si los modelos de interseccion variable estudiados en las secciones 4.2 y 4.3 pueden ser los adecuados, es decir, si pueden aun mejorar la estimacion del modelo COCO.

Observando los resultados de las regresiones individuales, vemos que efectivamente los terminos de interseccion varian, pero tambien lo hacen los coeficientes de las variables independientes. Mas aun, si nos fijamos bien, el coeficiente del ingreso (y por tanto su prueba t) es positivo para los autos compactos y medianos, pero casualmente es negativo para los autos de lujo.

Por otra parte debemos reconocer, independientemente de lo anterior, que las pruebas t para los terminos de interseccion en las tres categorias es sumamente baja (<0.73) por lo que en ningun caso resulta significativo, a diferencia del primer modelo conjunto ya estudiado.

Todo ello nos lleva a pensar que la diferencia entre las categorias no se puede explicar sencillamente por la variacion en sus intersecciones, sino mas bien en el cambio de sus pendientes. Por tanto no resultaran convenientes las estimaciones de este tipo de modelos (MOCER y MOVAR respecto a las categorias y/o respecto al tiempo).

Estudiaremos, entonces, la estimacion de los modelos que explican la diferencia entre los individuos (en este caso entre las categorias) en la variacion de sus pendientes.

5.3.5 MODELOS DE PENDIENTE VARIABLE.

De acuerdo a la seccion 4.4, un primer modelo de pendiente variable es MORAN (SUR en los textos de habla inglesa) que establece una estructura de correlacion de los errores entre los distintos individuos y su estimacion por el metodo de Minimos Cuadrados Generalizados, donde la matriz de covarianza (S) es calculada a partir de los vectores de error resultantes al aplicar M.C.O. a cada individuo por categoria.

El programa de computacion "SHAZAM" posee un comando que ejecuta precisamente este modelo, y la matriz de covarianza la calcula iterativamente con el mismo modelo. La instruccion clave que nos permite esta estimacion es "SYSTEM i" donde i es el numero de individuos; en nuestro caso i=3 categorias.

Al ejecutar este programa, obtuvimos los siguientes resultados:

1) R^2 del sistema = 0.8354

2) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC.PR.COMP.	-165.59	120.340	-1.3761
CONSUMO COMP.	-11519.00	2363.000	-4.8745
TRANSPORTE	-606.76	8765.800	-0.0692
INGRESO	135.72	56.342	2.4088
C.P.P.REAL	-5112.80	1449.500	-3.5274
INTERSECCION	7960.30	11827.000	0.6730
PORC.PR.MEDIAN.	-25.73	53.006	-0.4854
CONSUMO MEDIAN.	-11394.00	1176.800	-9.6822
TRANSPORTE	-1971.00	5568.400	-0.3539
INGRESO	111.53	35.664	3.1273
C.P.P.REAL	-1559.20	941.360	-1.6563
INTERSECCION	3783.20	7278.700	0.5197
PORC.PR.DE LUJO	17.20	25.374	0.6777
CONSUMO DE LUJO	-1845.80	374.940	-4.9229
TRANSPORTE	1012.20	3711.200	0.2727
INGRESO	-88.57	43.338	-2.0438
C.P.P.REAL	560.63	639.230	0.8770
INTERSECCION	2755.20	3793.900	0.7262

3) Numero de iteraciones necesarias = 5 en el calculo de S.

4) La matriz de covarianza entre los errores es la siguiente:

$$
S = \begin{matrix} 4,406\ 000 & & \\ 331\ 480 & 1,188\ 440 & \\ 180\ 140 & 430\ 000 & 863\ 600 \end{matrix}
$$

Det S = 0.62509 E+19

5) Suma de cuadrados de los residuales = 0.402092 E+9

Si comparamos estos resultados con los obtenidos en las regresiones individuales para cada categoria vemos que casi no hay mucha diferencia y podemos pensar, a priori, que el considerar la existencia de correlacion entre los errores no aporta significativamente mejores resultados. Sin embargo esto deberemos probarlo.

Por otra parte tambien deberemos probar si el presente modelo es mejor que el modelo COCO, es decir, si aceptamos la hipotesis de que los coeficientes de la pendiente varian con los individuos en lugar de ser constantes.

Ambas cosas haremos a continuacion.

PRUEBAS DE HIPOTESIS.

Primero probaremos la hipotesis nula de que los coeficientes son constantes (es decir, que el modelo COCO es el adecuado) contra la hipotesis alternativa de que son variables (y por tanto MORAN sera mejor). Para comparar ambos modelos, usaremos el estadistico 4.4.8:

$$
F = \frac{(0.6635\text{ E+9} - 0.402092\text{ E+9}) \;/\; (18 - 6 - 3)}{(0.402092\text{ E+9}) \;/\; (180 - 18 - 6)} = 11.26
$$

que igualmente resulta muy significativa para una F con (9,156) grados de libertad, lo que nos lleva a optar por la hipotesis alternativa de que los coeficientes varian en las categorias.

Para probar si la aportacion del modelo MORAN es significativa con respecto a las estimaciones individuales iniciales, usaremos el estadistico equivalente, donde ahora los residuales restringidos seran la suma de los residuales individuales y sus grados de libertad seran (NT-NK) por tener NT observaciones y NK parametros a estimar. Por tanto los grados de libertad del numerador seran la resta de estos grados de libertad y los del MORAN, resultando (N(N+1)/2) grados de libertad. Asi pues,

$$F = \frac{(0.42201\ E{+}9 - 0.402092\ E{+}9)\ /\ (3(4)/2)}{(0.402092\ E{+}9)\ /\ (180 - 18 - 6)} = 1.28$$

El valor critico para F(6,156) con 5 % de nivel de significancia es, aproximadamente, 2.18 por lo que nuestro estadistico es muy bajo y se acepta la hipotesis de que el modelo de estimaciones individuales es el adecuado.

Las pruebas anteriores nos indican que se trata, efectivamente, de un modelo donde los coeficientes de la pendiente varian con las categorias, pero que la correlacion existente entre los errores de diversas categorias no aporta eficiencia significativamente.

Como dijimos en la seccion 4.4.1, la estimacion por MORAN generalmente es mejor que otros estimadores, excepto cuando no existe correlacion alguna entre los individuos, o cuando la matriz de informacion (X_i) es la misma para todos los individuos; en ambos casos la estimacion por MORAN sera exactamente igual a la estimacion por M.C.O. de cada individuo por separado.

En nuestro caso podemos ver en la matriz de covarianza (S) que la covarianza entre los errores si es distinta de cero. Sin embargo las matrices de informacion de cada categoria son casi iguales, ya que tienen en comun tres variables (precio del transporte, ingreso y tasa de interes), y solo difieren en las dos primeras variables (porcentaje de los precios y consumo de gasolina); ademas estas dos variables tienen una misma tendencia entre las diferentes categorias.

Todo esto nos hace pensar que nuestro modelo es un ejemplo del segundo caso que Zellner menciona, es decir, de regresiones efectivamente no relacionadas.

Podemos concluir que la mejor estimacion para nuestro modelo es la estimacion individual de cada categoria. Pero cabe hacer un ultimo analisis para saber cual de las estimaciones individuales posibles es la mejor.

5.3.6 ESTIMACION INDIVIDUAL.

De acuerdo con los resultados obtenidos en la seccion 5.3.2 de regresiones por categorias, y como lo expresamos ya al analizar el modelo COCO, la existencia de autocorrelacion en los errores para los autos medianos y de lujo es manifiesta, pero no en los compactos (la prueba de Durbin-Watson = 2.0244 lo comprueba).

Vimos conveniente por ello analizar el modelo con la estructura de autocorrelacion en los errores pero sujeto a la restriccion de que los coeficientes son constantes. Sabiendo ahora que esta restriccion no es la conveniente, es necesario ver si logramos una mejor estimacion individual bajo el supuesto de autocorrelacion pero ahora de manera independiente entre las categorias.

Con el fin de presentar los resultados juntos y de forma clara, repetimos los resultados de los autos compactos obtenidos por M.C.O. y despues presentamos las estimaciones bajo autocorrelacion para los autos medianos y de lujo.

REGRESION POR M.C.O. PARA AUTOS COMPACTOS

1) R-cuadrada = 0.4762

2) Prueba F = 9.637 con (5,53) grados de libertad

3) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	-173.88	127.730	-1.3614
CONSUMO	-11474.00	2497.300	-4.5944
TRANSPORTE	-772.22	9252.600	-0.0835
INGRESO	135.31	59.603	2.2702
C.P.P.REAL	-5131.10	1529.700	-3.3543
INTERSECCION	8362.40	12503.000	0.6688

4) Suma de los cuadrados de los residuales = 0.25993 E+9

5) Durbin-Watson = 2.0244

AUTOS MEDIANOS BAJO AUTOCORRELACION

1) R-cuadrada = 0.7977

2) Coeficiente de autocorrelacion = 0.98

3) Prueba t de significancia = 37.83

4) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	56.711	49.999	1.1342
CONSUMO	-15648.00	3624.300	-4.3175
TRANSPORTE	-322.63	2003.400	-0.1610
INGRESO	36545.00	4819.100	7.5835
C.P.P.REAL	-2374.20	896.950	-2.6470
INTERSECCION	150410.00	23042.000	-6.5275

5) Suma de los cuadrados de los residuales = 0.77792 E+8

AUTOS DE LUJO BAJO AUTOCORRELACION

1) R-cuadrada = 0.5367

2) Coeficiente de autocorrelacion = 0.66

3) Prueba t de significancia = 6.748

4) Analisis de varianza:

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDAR	PRUEBA t
PORC. PRECIO	24.75	23.397	1.0577
CONSUMO	-1409.80	656.900	-2.1462
TRANSPORTE	-2528.90	2567.300	-0.9850
INGRESO	15.56	32.350	0.4810
C.P.P.REAL	114.74	613.320	0.1871
INTERSECCION	5752.20	2618.600	2.1967

5) Suma de los cuadrados de los residuales = 0.31761 E+8

La suma de cuadrados de los residuales de todo el modelo es la suma de los residuales individuales, es decir, 0.369483 E+9, y con ellos haremos la ultima prueba bajo la hipotesis nula de que las regresiones por M.C.O. individuales son las indicadas, contra la alternativa de nuestras ultimas estimaciones. El estadistico es el mismo de antes:

$$F = \frac{(0.42201 - 0.369483) / 3}{(0.369483) / 169} = 8.01$$

que es de nuevo altamente significativo para una F (3,169). Por tanto nuestras ultimas estimaciones son las mas convenientes.

5.4 INTERPRETACION DE LOS RESULTADOS

Mediante las pruebas de hipotesis estudiadas, hemos concluido que el modelo de demanda de automoviles, definido como lo hemos hecho al presente, es de coeficientes variables y su mejor estimacion es la individual, considerando cada categoria como independiente y por tanto aceptando para los autos medianos y de lujo la existencia de autocorrelacion en sus errores.

Pasemos ahora a interpretar los resultados de los coeficientes de cada una de las variables para completar el estudio.

1) PORCENTAJE DE PRECIOS.

El signo negativo en los autos compactos revela que mientras mas suban de precio esos autos (cuando menos respecto al precio de los otros automoviles), menor sera la demanda; el signo positivo en las otras categorias nos puede hacer pensar lo contrario o que no importa mucho a los consumidores de autos medianos y de lujo si el precio de estos autos es mayor o menor en proporcion a otros.

De cualquier manera, en las tres categorias observamos que el porcentaje de precios no es significativa (la prueba t mas alta de las tres es -1.3614). Por tanto no importara tanto la interpretacion que se de a los signos.

Al especificar el modelo, pensabamos que el precio sí sería significativo (al comprar un coche es determinante su precio) y el hecho de que haya resultado no significativo puede deberse a una mala definicion de la variable, o al metodo con el cual aproximamos el precio o, en ultimo caso a que en el mercado de verdad no son determinantes los precios. Esta ultima razon puede ser perfectamente valida en un mercado como el nuestro afectado por la crisis economica, como mencionaremos mas adelante en las conclusiones.

2) CONSUMO DE GASOLINA.

Por el contrario, encontramos que los coeficientes de consumo de gasolina en las tres categorias son negativas y altamente significativas. Esto quiere decir, como lo esperabamos, que los compradores de autos se fijan en el consumo de gasolina del coche, sobre todo porque el precio de la misma ha subido considerablemente en los ultimos años (vease en la grafica del precio de la gasolina, en el apendice A, el aumento fuerte de su precio en estos años).

3) PRECIO DEL TRANSPORTE.

Al especificar el modelo, pensabamos que un comprador de autos en Mexico definitivamente no toma en cuenta el precio del transporte ya que en nuestro pais este esta altamente subsidiado (a la fecha el costo del pasaje en el Metro sigue siendo de un peso) por el gobierno, y quien considera la posibilidad de comprar un coche esta en una situacion economica muy arriba de quien solo puede recurrir al transporte publico.

Los resultados nos confirmaron lo anterior, ya que los coeficientes en los tres casos es negativo pero nada significativos (la t más grande es de -0.9850).

4) INGRESO.

En todo modelo de demanda se espera que el ingreso (el dinero del que se dispone) sea positivo y significativo, es decir, que mientras mas dinero haya mas autos se pueden comprar. Esto se confirmaba en las categorias de compactos y medianos a lo largo de las diversas estimaciones, pero en el caso de los autos de lujo muchas veces el signo era negativo.

Sin embargo en la estimacion definitiva, aquella que considera la estructura de autocorrelacion en los errores para los autos medianos y de lujo, el signo del ingreso para los autos de lujo fue positivo, aunque no significativo, y positivo y significativo para los compactos y medianos.

Esto tiene una logica. A los consumidores de autos de lujo no les afecta si el ingreso nacional aumenta o disminuye (recordar que nuestro ingreso se refiere al ingreso nacional, no al ingreso individual) ya que ellos podran seguir comprando este tipo de autos cualquiera que sea la situacion economica nacional. De hecho todos podemos constatar que este tipo de personas siguen comprando autos de lujo a pesar de que el ingreso nacional disminuya. Esto no sucede con los consumidores de autos compactos y medianos, cuyo ingreso individual es un reflejo claro del ingreso nacional, y las crisis les afecta directamente.

5) TASA DE INTERES.

Como en la variable anterior, aqui tambien podemos observar una diferencia entre las estimaciones para autos compactos y medianos y los de lujo. Para los primeros sus coeficientes son negativos y significativos lo que es coherente con la realidad, ya que un aumento en el C.P.P. (en el costo del dinero) les afecta, y no podran comprar facilmente un automovil (mientras mas les cueste un prestamo bancario, menos podran comprar un auto). En cambio, para los autos de lujo, el signo es positivo pero no significativo (t = 0.1871) y se puede justificar de la misma manera como lo hicimos para el ingreso.

CAPITULO 6

CONCLUSION.

Como vimos en los resultados, el modelo que aplicamos es un caso en el que la informacion combinada de series de tiempo y corte transversal no aporta mucha eficiencia a la estimacion de cada individuo por separado. Tambien dijimos que ello se debia a la similitud de las matrices de informacion de cada categoria de automoviles.

Alguno podria decir que se trata de un mal ejemplo ya que no se obtuvo mayor eficiencia al combinar la informacion, pero no somos de esta opinion, ya que fue precisamente la aplicacion de los diversos modelos de estimacion lo que nos condujo a estas conclusiones; luego los modelos estudiados funcionan y el nuestro en concreto ha servido de ejemplo para demostrarlo.

Ciertamente serian deseables mejores resultados: no tener multicolinealidad, unas R^2 mas altas, coeficientes mas significativos en variables que esperabamos lo fueran (como en el caso del precio), una mejor eficiencia al combinar la informacion de las diversas categorias, etc.

Esto no lo discutimos en lo absoluto, y pensamos que si los resultados no fueron mejores se debe a varias razones, entre las cuales destacamos las siguientes:

1) En primer lugar la estructura del mercado de automoviles en Mexico afronta cambios continuos, como vimos cuando definimos las variables del modelo, comprometiendo la estabilidad de cualquier modelo que se quiera hacer.

2) El mercado de automoviles en Mexico es sumamente irregular, con altas y bajas extremosas, unas veces como consecuencia de las irregularidades en el comportamiento economico del pais, pero otras sin razon alguna.

3) La situacion politica y economica apremiante que vive el pais induce a la gente al consumo de bienes, sin importar mucho otros factores que en circunstancias normales afectan de manera decisiva toda relacion de demanda, y asi vemos que en algunos circulos de la sociedad se siguen comprando coches de gran valor a pesar de la crisis economica. Esto hace tambien dificil poder explicar y modelar la demanda de autos en funcion de unas variables.

4) La informacion fue obtenida con dificultad, teniendo que recurrrir en ocasiones a metodos estadisticos que nos "aproximaban" a la realidad, pero no con una certeza absoluta (recordar el Metodo de Ginsburgh).

5) Igualmente, y de forma principal, la carencia de datos reales en los precios de los automoviles quiza sea una de las causas mas importantes de su no significancia en el modelo.

Indudablemente, podemos entonces decir, nuestro modelo es susceptible de ser mejorado, y no solo lo aceptamos sino que se propone como un trabajo de investigacion posterior donde se pueden alcanzar diversos objetivos:

1) Eliminar la presencia de multicolinealidad.

2) Mejorar las R^2 de las estimaciones.

3) Obtener una informacion mas precisa y conveniente.

4) Lograr una informacion que sí aumente la eficiencia al combinar el corte transversal y las series de tiempo, quiza cambiando la estructura del corte transversal.

5) Establecer comparaciones con otros modelos.

6) Probar la resistencia de los estimadores.

etcetera.

Finalmente, en lo personal pienso que el mayor beneficio de esta tesis es la satisfaccion de haber abordado, con la profundidad que me fue posible, un tema que me gusta y que considero apenas comienzo a conocer.

Como dije en la introduccion, mi primer objetivo, afianzarme en un campo de mi carrera, se ha logrado con creces, y al terminar satisfecho este trabajo puedo decir con Sócrates que mientras más he conocido, "lo único que he sabido es que no sé nada".

APENDICE A

TABLAS Y GRAFICAS DE LA INFORMACION

En este apendice presentamos las tablas y las graficas de los datos procesados en nuestro modelo. Para un mayor conocimiento de estos datos debe consultarse la seccion 5.2 donde definimos cada una de las variables del modelo.

La primera tabla, titulada "TABLA DE INFORMACION" presenta los datos definitivos de las variables del modelo que fueron procesados en las estimaciones.

La segunda tabla, titulada "TABLA PREVIA DE INFORMACION", muestra las variables auxiliares que sirvieron para generar las variables definitivas; esas variables son: costo unitario de los automoviles, ingreso (PIB Real) base 1978, costo porcentual promedio nominal, precio de la gasolina, deflactor y la inflacion.

En el apendice B se pueden encontrar los programas en FORTRAN que generaron ambas tablas mediante la lectura y transformacion de unos archivos fuente.

En la bibliografia se pueden consultar las fuentes precisas de donde obtuvimos todos los datos anteriores.

En cuanto a las graficas, ayudaran a visualizar el comportamiento de las variables del modelo en las diversas categorias. Estan construidas con las Tablas de informacion, usando el programa de computacion "S.A.I." del Centro de Estudios Economicos de Banamex.

Las graficas que presentamos son las siguientes:

1) Ventas de automoviles, por categorias.

2) Proporcion de precios, por categorias.

3) Consumo de gasolina, por categorias.

4) Indice de precios del transporte.

5) Ingreso (PIB)

6) Costo Porcentual Promedio

7) Ingreso contra ventas de automoviles, por categorias.

8) Costo unitario contra venta de autos, por categorias.

9) Consumo de gasolina contra venta de autos, por categorias.

T A B L A   D E   I N F O R M A C I O N.

T	VENTAS COMPAC	VENTAS MEDIAN	VENTAS LUJO	% COSTO COMPACTO	% COSTO MEDIANOS	% COSTO DE LUJO	CONSUMO COMPACTOS	CONSUMO MEDIANOS	CONSUMO DE LUJO	INDICE TRANSP.	INGRESO BASE 80	C.P.P. REAL
1	13623.	7763.	3894.	24.31560	32.12114	43.36366	0.36550	0.46993	0.65791	85.00	95.72	-0.58577
2	11441.	7109.	4023.	23.86180	32.15858	43.97963	0.34505	0.44749	0.62649	87.10	96.17	-0.28310
3	13913.	6374.	4075.	24.60581	31.60996	43.90422	0.33945	0.43644	0.61101	96.70	99.76	-0.20204
4	10569.	5557.	3657.	24.55723	31.55087	43.89190	0.33196	0.42680	0.59752	100.20	96.94	-0.05076
5	12555.	6668.	3451.	28.33152	29.90496	41.76352	0.32610	0.41928	0.58699	101.00	101.59	0.00159
6	14495.	8255.	4328.	24.89264	31.73166	43.37569	0.32067	0.41229	0.57720	100.80	101.32	-0.09805
7	16115.	9052.	2978.	26.29159	32.26527	41.44315	0.31438	0.40420	0.56589	102.40	100.75	-0.25674
8	12923.	8913.	854.	26.60414	30.80532	42.59053	0.30618	0.39366	0.55112	103.40	101.64	-0.08350
9	10710.	6012.	1630.	25.15871	32.26462	42.57687	0.30006	0.38580	0.54012	104.30	98.33	0.64260
10	11030.	9643.	2451.	24.99260	34.12928	40.87312	0.29728	0.38222	0.53511	106.40	104.98	0.21555
11	14725.	7981.	3823.	22.50150	34.56640	42.93209	0.29329	0.37709	0.52793	106.90	100.62	0.14780
12	10363.	9570.	4904.	23.38195	32.81691	43.79913	0.28889	0.37143	0.52000	107.80	102.19	-0.07966
13	16171.	8957.	4001.	24.78347	30.41867	44.79786	0.28244	0.36314	0.50840	107.90	99.95	-0.20689
14	16019.	8635.	3995.	24.85904	31.66308	43.47788	0.27439	0.35279	0.49390	112.60	99.21	-0.01853
15	14521.	10668.	4729.	24.22366	32.54270	43.23364	0.26842	0.34511	0.48315	124.40	106.95	0.06930
16	11960.	9476.	4671.	24.61736	31.99041	43.59223	0.26299	0.33813	0.47339	127.40	102.90	0.01743
17	14846.	9698.	4655.	23.53747	32.49047	43.97205	0.25736	0.33089	0.46325	128.70	107.22	0.25457
18	14854.	10838.	5408.	23.38932	32.88602	43.72465	0.25329	0.32566	0.45593	129.00	107.82	0.28112
19	17685.	11612.	4583.	24.00652	33.15733	42.83615	0.24935	0.32060	0.44884	128.90	110.35	0.11469
20	14226.	9767.	3291.	24.96277	32.22574	42.81149	0.24441	0.31424	0.43994	130.70	107.54	0.02069
21	14459.	7027.	2535.	23.59699	34.51263	41.89738	0.23871	0.30691	0.42968	131.80	107.95	0.09510
22	13314.	[illegible]	3273.	23.21953	32.40920	44.37127	0.23350	0.30021	0.42029	132.00	110.70	-0.00344
23	15310.	8321.	5267.	24.20147	33.93006	42.60847	0.22753	0.29254	0.40956	132.30	106.86	0.06887
24	16465.	5880.	4597.	22.71736	33.84187	41.64077	0.30707	0.39481	0.55273	138.00	106.80	-0.12807
25	18444.	7059.	3690.	24.28545	34.45883	41.25573	0.46138	0.59320	0.83048	145.40	106.44	-0.42612
26	15833.	4913.	3537.	23.38843	32.43727	43.67429	0.43767	0.56272	0.78781	156.40	107.84	-0.27038
27	18230.	12283.	3388.	23.90019	31.24183	44.85798	0.41985	0.53980	0.75572	160.30	117.87	-0.21968
28	15938.	9330.	1805.	25.63458	29.86483	44.50058	0.40420	0.51969	0.72757	162.50	109.73	-0.39707
29	15439.	9459.	1756.	26.47083	29.70819	43.82098	0.38345	0.49301	0.69022	168.00	115.07	-0.38185
30	16316.	7956.	1522.	24.85445	33.70946	41.43608	0.36344	0.46728	0.65420	184.30	114.54	-0.23491
31	9368.	7298.	1914.	25.98243	32.89793	41.11964	0.34740	0.44665	0.62531	185.90	114.67	-0.21370
32	18021.	6508.	571.	23.68788	32.08625	44.22587	0.55219	0.70996	0.99395	221.10	114.49	-0.37825
33	15280.	5571.	2143.	22.48005	29.83179	47.68816	0.49953	0.64225	0.89915	219.20	109.64	-0.16191
34	14246.	3657.	1813.	24.43245	27.63588	47.93166	0.47613	0.61216	0.85703	220.40	111.93	-0.16207
35	17018.	4097.	3722.	24.11688	30.76763	45.11548	0.45446	0.58431	0.81803	220.80	110.99	-0.14062
36	7795.	2540.	2303.	22.84064	33.11499	44.04437	0.86915	1.11748	1.56447	261.10	111.57	-0.55838
37	13361.	3149.	2652.	24.24506	30.38165	44.87329	0.79037	1.01619	1.42267	302.70	109.32	-0.53711
38	10953.	3524.	1925.	27.50537	31.66792	40.82670	0.71712	0.92202	1.29082	308.80	117.22	-0.07176
39	13315.	6316.	2546.	25.37365	30.47919	44.14735	0.68332	0.87856	1.22998	312.80	117.61	0.03370
40	13246.	3451.	1916.	24.75067	31.75316	43.49617	0.78471	1.00892	1.41249	333.40	117.51	-0.17777
41	9143.	2962.	2177.	25.85696	31.26311	42.88194	0.74020	0.95168	1.33235	348.00	119.88	0.14757
42	10399.	3679.	2116.	23.81669	31.80983	44.37349	0.71109	0.91426	1.27996	358.00	118.76	0.28303
43	7719.	3026.	1862.	23.46239	31.21446	45.32315	0.68666	0.88259	1.23563	359.80	119.03	0.02295
44	9428.	6059.	2657.	22.79827	33.70658	43.49514	0.65526	0.84249	1.17967	366.70	119.79	0.23190
45	6343.	2105.	2103.	21.17965	30.85659	47.96376	0.63157	0.81202	1.13683	379.50	115.86	0.48607
46	9726.	3862.	2856.	23.23064	28.85468	47.91468	0.76668	0.98573	1.38007	382.00	118.13	0.35790
47	10341.	3106.	2958.	23.83100	30.51018	[illegible]	0.74160	0.95477	1.33667	416.70	117.83	-0.15767
48	[illegible]	3315.	2450.	[illegible]	30.01100	[illegible]	0.70189	0.90237	[illegible]	428.00	115.43	-0.09936
49	[illegible]	3275.	2208.	21.28658	24.98318	54.05026	0.67332	0.86570	1.21198	442.10	116.42	-0.22680
50	10040.	6798.	1724.	26.23598	33.33745	40.32657	0.63328	0.81423	1.13990	451.20	120.14	-0.10588
51	9345.	5072.	2125.	24.50857	26.24718	49.24425	0.60164	0.77353	1.08294	471.70	123.18	0.03133
52	8022.	4030.	1928.	23.81590	26.38601	49.80008	0.76940	0.98922	1.38491	500.40	116.09	-0.00988
53	13099.	4905.	2923.	26.08891	26.94288	46.96821	0.73760	0.94834	1.32768	509.00	122.62	0.20216
54	12871.	5227.	2338.	22.45619	21.10727	56.43654	0.71399	0.91799	1.28519	513.30	122.05	0.12871
55	9828.	5032.	2436.	21.38861	19.44083	59.17056	0.68915	0.88605	1.26048	526.60	122.67	0.22076
56	12602.	5203.	2060.	28.55815	24.67025	46.77160	0.66726	0.85791	1.20107	536.90	123.26	0.37074
57	13877.	2533.	1309.	30.20333	23.58670	46.20996	0.64893	0.83433	1.16807	543.20	117.88	0.31244
58	12776.	4512.	4543.	26.83679	22.36350	50.79973	0.63018	0.81023	1.13432	552.00	123.13	0.13073
59	12116.	5290.	4493.	26.72847	29.78348	43.48806	0.60896	0.78295	1.09612	557.00	123.16	0.13441
60	11383.	7581.	3835.	26.64359	31.99648	41.35993	0.58877	0.75698	1.05978	565.60	120.65	-0.05335

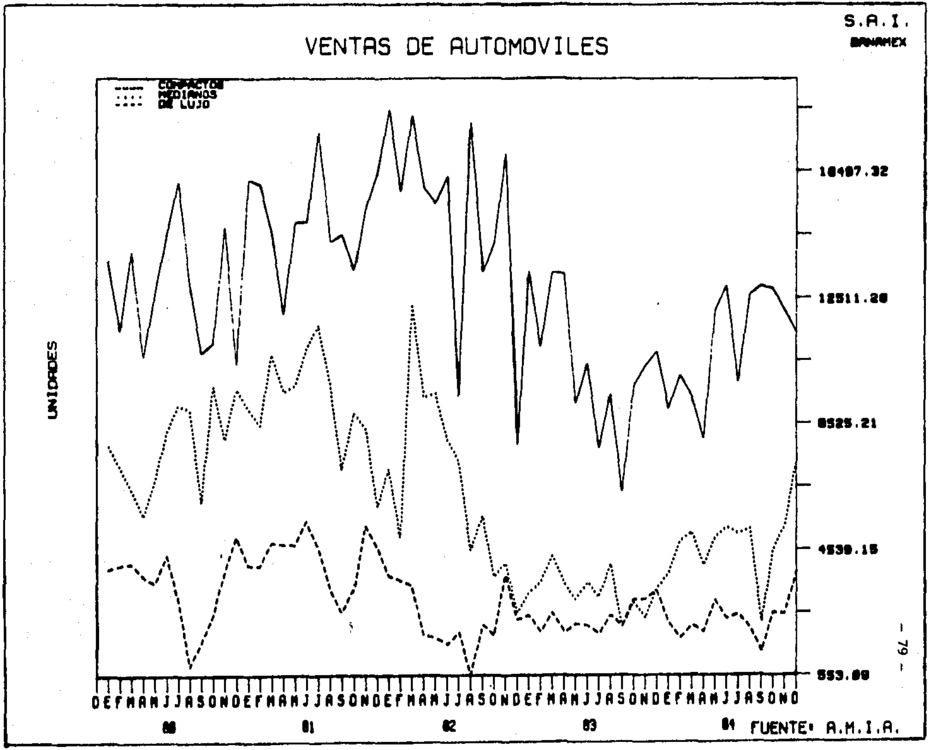
NOTAS: VENTAS EN UNIDADES.
CONSUMO EN PESOS/KM.
INGRESO EN MILES DE MILLONES DE PESOS.

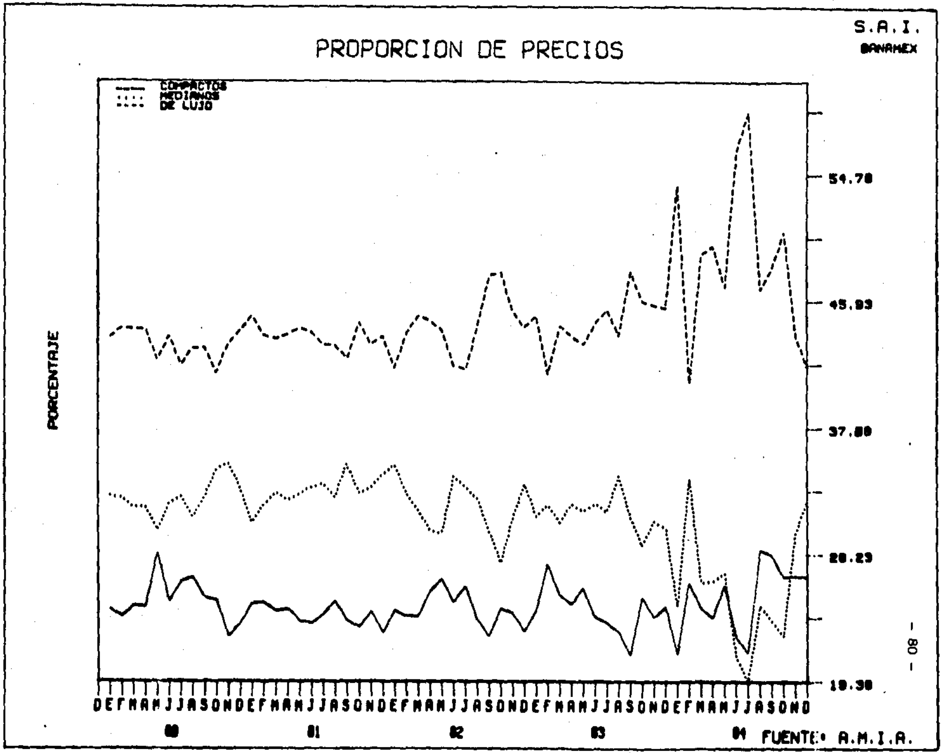
FUENTES: BANCO DE MEXICO
WANAMEX
I.M.I.A.
PEMEX

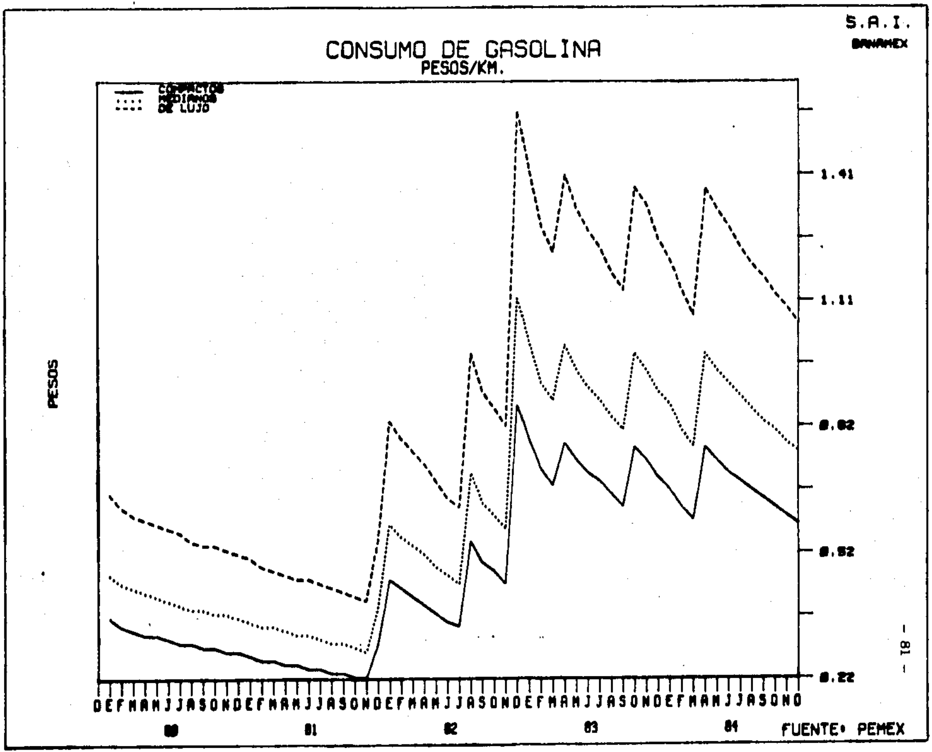
T A B L A   P R E V I A   D E   I N F O R M A C I O N

T	COSTO UN. COMPACTOS	COSTO UN. MEDIANOS	COSTO UN. DE LUJO	INGRESO BASE 78	C.P.P. NOMINAL	GASO-LINA.	DEFLAC-TOR	INFLA-CION
1	116.6721	155.0559	208.0700	43618.70	17.90	2.80	131.60	5.0152
2	117.9970	159.0247	217.4300	43619.80	18.39	2.80	138.20	2.5326
3	122.0389	156.2382	217.7900	45248.20	19.20	2.80	141.70	2.2583
4	121.0352	155.5068	216.3300	43970.80	19.83	2.80	144.90	1.7944
5	148.7958	157.0596	219.3400	46080.30	20.39	2.60	147.50	1.6949
6	131.0533	167.0577	228.3600	45959.00	20.47	2.80	150.00	2.0000
7	138.3683	169.3314	218.1400	45697.30	20.82	2.80	153.00	2.6797
8	137.3790	159.3067	220.2500	46100.70	21.40	2.80	157.10	2.0369
9	131.3691	168.4724	222.3200	44599.90	21.51	2.80	160.30	0.9357
10	138.2949	135.8386	226.1800	47618.30	22.42	2.80	161.80	1.3397
11	137.2352	210.8182	261.8400	45640.60	22.77	2.80	164.00	1.5244
12	141.8174	199.0258	295.9300	46353.10	24.25	2.80	166.50	2.2823
13	150.2116	193.5525	271.7100	45335.90	25.56	2.80	170.30	2.9360
14	155.6110	198.2025	272.1600	45002.50	25.98	2.80	175.30	2.2248
15	153.9359	206.6015	276.7400	48511.30	26.59	2.80	179.20	2.0647
16	158.2314	207.3069	282.4900	46673.00	26.91	2.80	182.90	2.1870
17	153.4280	211.7873	286.6300	48633.00	27.22	2.80	186.90	1.6051
18	155.6626	218.8658	291.0003	48904.00	27.66	2.80	189.90	1.5798
19	163.2691	225.5040	291.3300	50055.50	28.42	2.80	192.90	2.0218
20	174.7566	225.6024	299.7100	48777.80	29.50	2.80	196.80	2.3862
21	168.6307	246.6578	299.4000	48554.70	30.45	2.80	201.50	2.2335
22	171.6980	241.0737	329.2400	50213.70	31.22	2.80	206.00	2.6214
23	183.8358	250.2067	325.7200	48463.60	31.77	2.80	211.40	2.4125
24	175.1988	230.9923	335.0200	48443.20	31.81	3.87	216.50	3.1871
25	183.6961	260.6480	312.0603	48279.90	32.34	6.00	223.40	5.4163
26	203.4228	276.2206	371.9100	48915.30	33.43	6.00	235.50	4.2463
27	212.1544	277.3233	398.1900	51463.30	33.67	6.00	245.50	3.8697
28	250.6223	291.9303	435.0700	49773.80	34.39	6.00	255.00	5.4118
29	267.5295	300.2690	442.8800	52196.00	36.26	6.00	268.80	5.5060
30	266.0589	360.8489	443.5600	51955.00	39.59	6.00	283.60	4.6192
31	286.4316	360.1363	650.1400	52011.90	43.23	6.00	296.70	4.8536
32	272.8297	369.5596	509.3800	51931.40	46.42	10.00	311.10	10.5432
33	288.3818	382.6923	611.7600	49732.00	47.48	10.00	343.90	4.9142
34	309.4296	350.0000	607.0400	50770.00	45.99	10.00	360.80	4.7672
35	313.1318	425.0000	623.1900	50342.60	45.51	10.00	378.00	4.5767
36	356.3818	487.6986	648.6600	50604.70	46.12	20.00	395.30	9.9671
37	393.0027	437.8619	708.8700	49815.00	50.29	30.00	434.70	10.2139
38	436.1535	502.1590	647.3900	50901.50	54.24	30.00	479.10	4.3468
39	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	502.40	4.4948
40	[illegible]	584.4119	800.2100	[illegible]	57.31	34.00	525.40	4.0155
41	[illegible]	[illegible]	801.3100	[illegible]	58.14	24.00	557.00	4.0934
42	504.5600	673.9038	960.0699	53866.90	58.63	24.00	579.80	3.5876
43	507.0360	674.5628	979.4600	53991.00	58.73	24.00	600.60	4.7619
44	511.8811	736.8012	976.5800	54336.40	58.23	24.00	629.20	3.7508
45	494.2208	720.0295	1119.2200	52551.40	57.78	24.00	652.80	2.9718
46	607.7870	695.0874	1106.0500	53578.50	57.14	30.00	672.20	3.2431
47	608.2169	780.3653	1163.6200	53499.30	56.82	30.00	694.00	5.8069
48	678.2034	828.0597	1252.7400	52356.30	56.44	30.00	734.30	4.2353
49	722.5984	837.2213	1834.8000	52806.30	55.95	30.00	765.40	6.3235
50	818.0544	1045.7190	1254.2900	54494.40	55.16	30.00	813.80	5.2593
51	857.5816	918.4174	1723.1100	55874.70	53.11	30.00	856.60	4.2610
52	829.6003	919.0574	1734.7300	52657.60	51.10	40.00	893.10	4.3108
53	890.5340	919.6837	1603.2400	55621.30	50.12	40.00	931.60	3.3061
54	960.6008	921.6970	2464.4302	55359.40	50.38	40.00	962.40	3.6056
55	1016.3970	923.6359	2811.2600	55551.80	50.69	40.00	997.10	3.2795
56	1070.7750	925.0000	1753.6799	55458.00	50.93	40.00	1029.80	2.8258
57	1162.2170	971.9723	1946.6600	53466.90	50.60	40.00	1058.90	2.9748
58	1147.4370	971.9723	2138.3099	55849.40	49.34	40.00	1090.40	3.4850
59	1229.2330	1309.7321	2000.0000	55655.40	48.30	40.00	1128.40	3.4296
60	1288.3771	1547.2209	2000.0000	54726.20	47.54	40.00	1167.10	4.2413

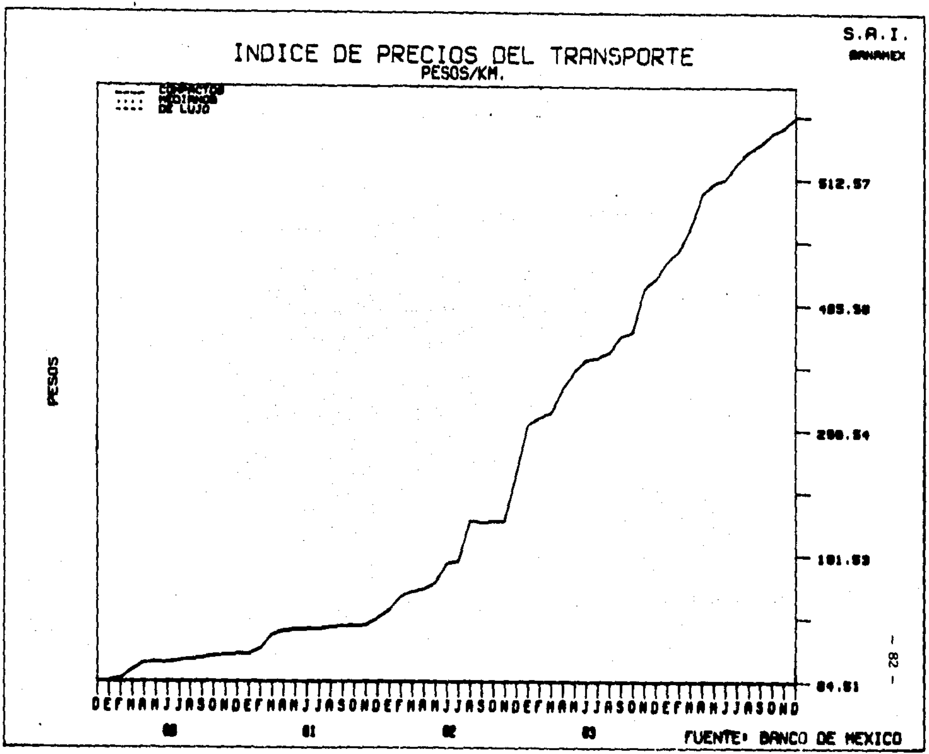
NOTA: COSTO UNITARIO EN MILES DE PESOS
INGRESO EN MILES DE MILLONES DE PESOS
GASOLINA EN PESOS.

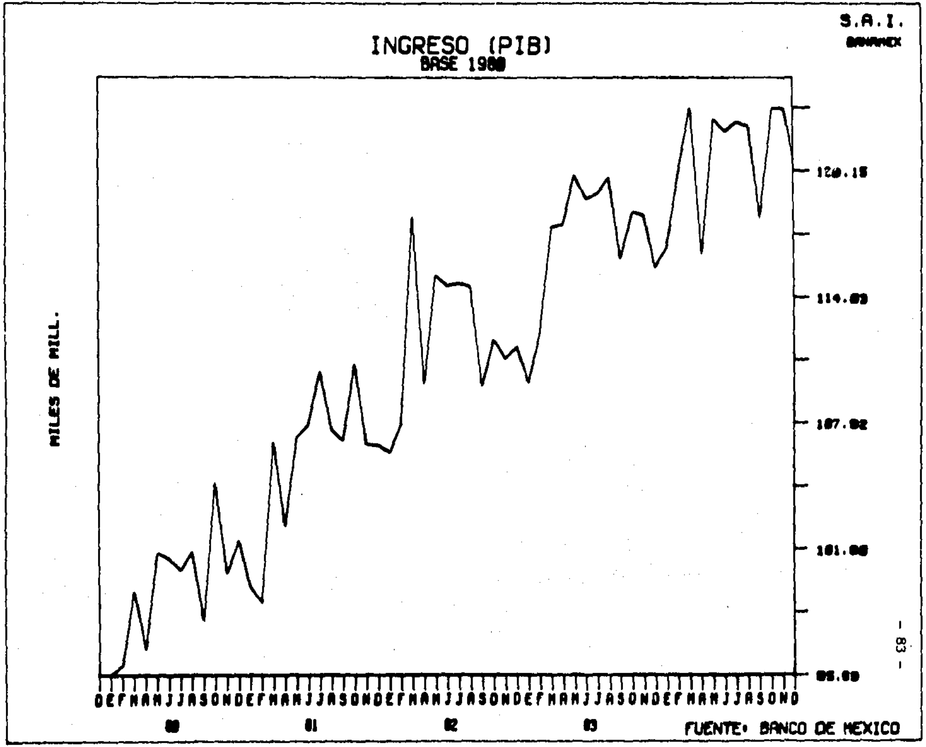

S.A.I.
BANAMEX
VENTAS DE AUTOMOVILES
COMPACTOS
MEDIANOS
DE LUJO
UNIDADES
16497.32
12511.26
8525.21
4539.15
553.09
D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D
80
81
82
83
84
FUENTE: A.M.I.A.


S.A.I.
BANAMEX
PROPORCION DE PRECIOS
COMPACTOS
MEDIANOS
DE LUJO
PORCENTAJE
54.70
45.93
28.23
19.30
D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D
80
81
82
83
84
FUENTE: A.M.I.A.

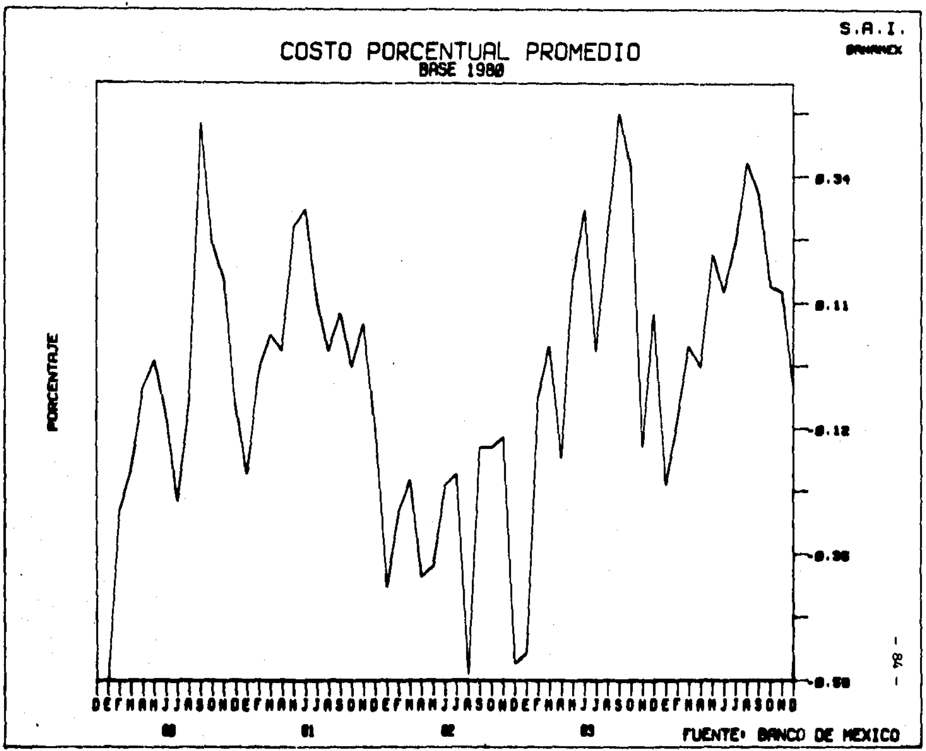


S.A.I.
BANAMEX
CONSUMO DE GASOLINA
PESOS/KM.
COMPACTOS
MEDIANOS
DE LUJO
PESOS
1.41
1.11
0.82
0.52
0.22
D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D
80
81
82
83
84
FUENTE: PEMEX

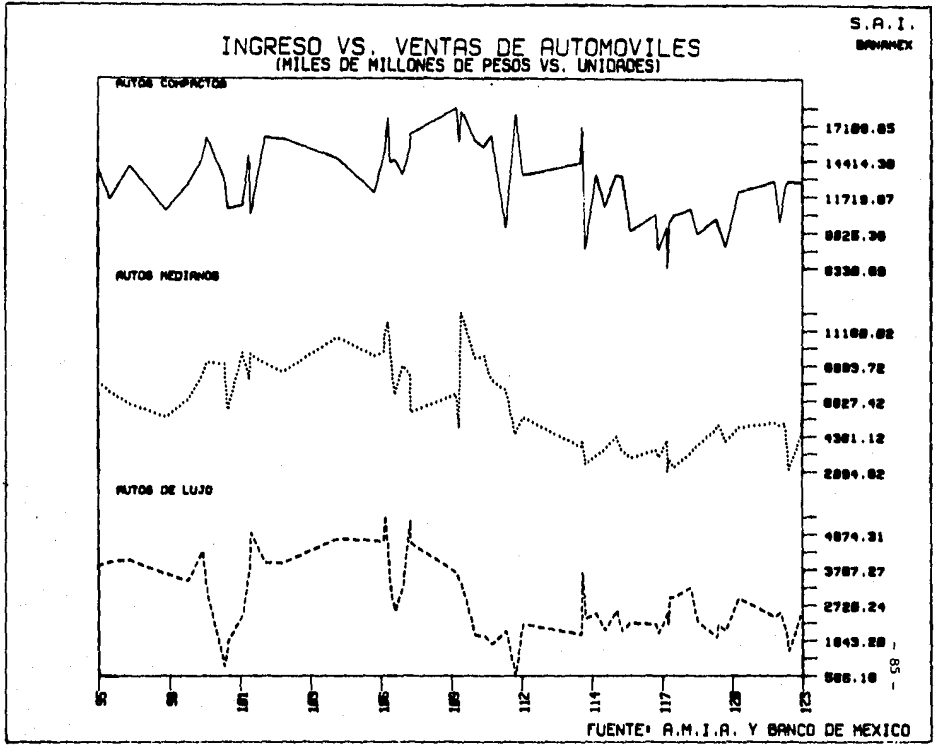

S.A.I.
BANAMEX
INDICE DE PRECIOS DEL TRANSPORTE
PESOS/KM.
DE LUJO
PESOS
512.57
D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D
FUENTE: BANCO DE MEXICO

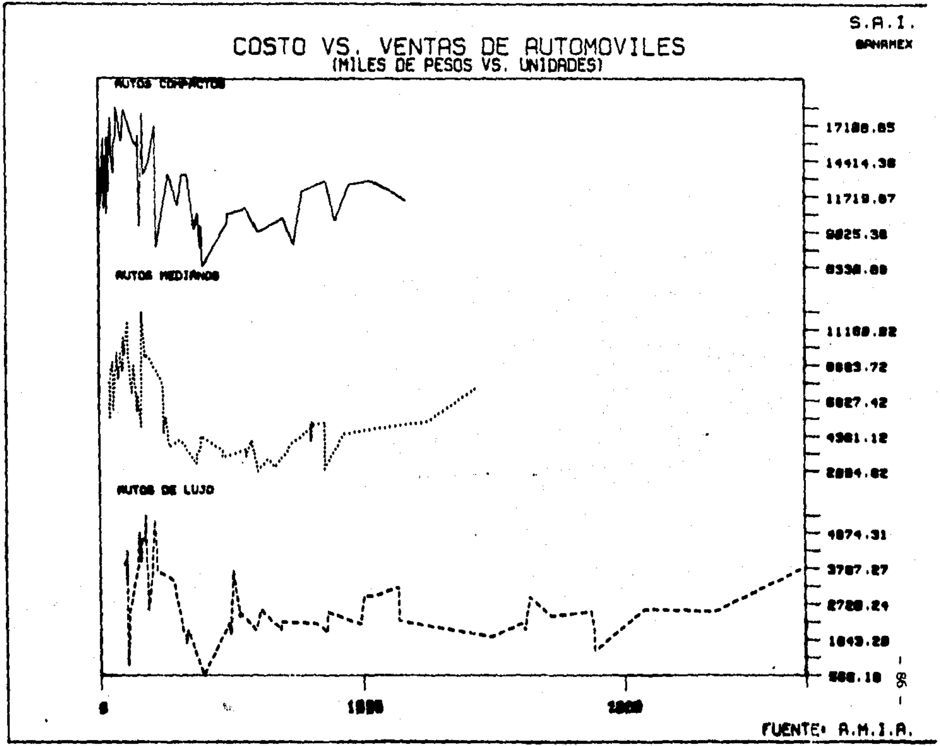


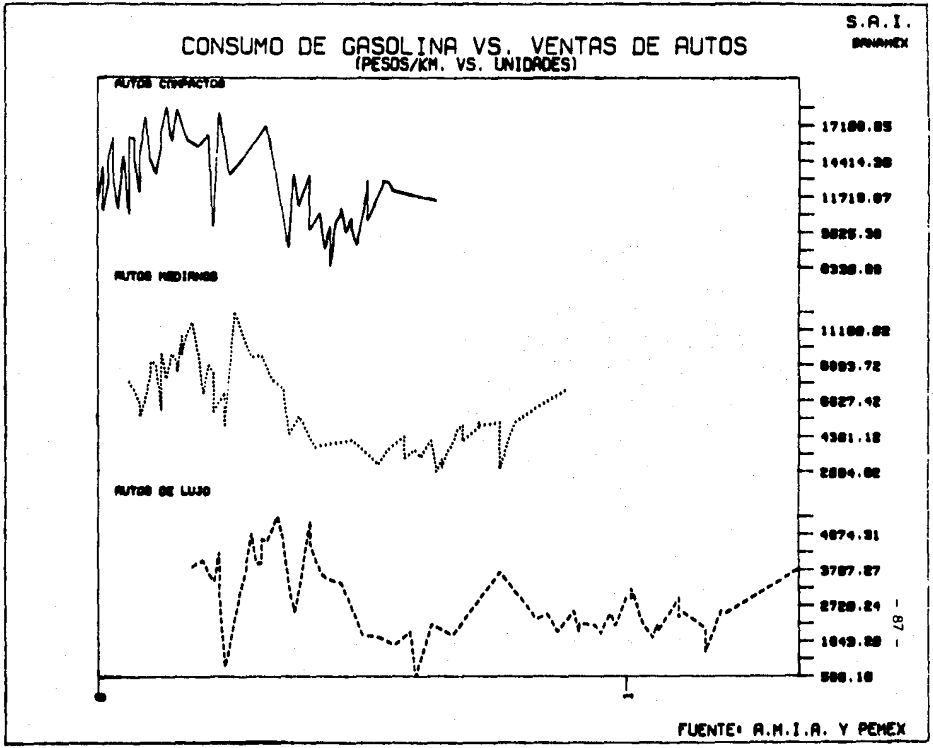
S.A.I.
BANAMEX
INGRESO (PIB)
BASE 1980
MILES DE MILL.
120.15
114.03
107.92
101.80
95.69
D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D
80
81
82
83
FUENTE: BANCO DE MEXICO



COSTO PORCENTUAL PROMEDIO
BASE 1980
S.A.I.
PORCENTAJE
FUENTE: BANCO DE MEXICO


S.A.I.
BANAMEX
INGRESO VS. VENTAS DE AUTOMOVILES
(MILES DE MILLONES DE PESOS VS. UNIDADES)
AUTOS COMPACTOS
AUTOS MEDIANOS
AUTOS DE LUJO
14414.30
11719.07
2728.24
112
114
117
FUENTE: A.M.I.A. Y BANCO DE MEXICO


S.A.I.
BANAMEX
COSTO VS. VENTAS DE AUTOMOVILES
(MILES DE PESOS VS. UNIDADES)
AUTOS COMPACTOS
AUTOS MEDIANOS
AUTOS DE LUJO
14414.38
11719.87
4074.31
0
FUENTE: A.M.I.A.


S.A.I.
BANAMEX
CONSUMO DE GASOLINA VS. VENTAS DE AUTOS
(PESOS/KM. VS. UNIDADES)
AUTOS COMPACTOS
AUTOS MEDIANOS
AUTOS DE LUJO
FUENTE: A.M.I.A. Y PEMEX

APENDICE B

PROGRAMAS DE COMPUTACION

En este apendice presentamos los programas de computacion que utilizamos en la elaboracion de esta tesis, tanto para la generacion de las tablas de informacion como para las estimaciones del modelo. Tambien presentamos un programa que genera las matrices de transformacion (corrigen por las medias) usadas para los modelos MOVAR del capitulo 4.

Todos los programas, incluyendo los de estimacion del Paquete SHAZAM, estan en FORTRAN. En los programas de SHAZAM simplificamos el listado poniendo "[Tabla de informacion]" en lugar de listar los datos completos, entendiendo que en ese lugar deben inserirse los datos de la tabla mencionada.

Los programas que presentamos son los siguientes:

1) Generacion de las tablas de informacion

2) Estimacion conjunta

3) Estimacion individual y MORAN

4) Estimacion con autocorrelacion

5) Matrices de transformacion

6) Ejemplo de matrices de transformacion (N=2, T=5).

PROGRAMA 1

TABLAS DE INFORMACION

OBJETIVO: Lee datos de archivos fuente y genera tablas de informacion.

EJECUCION: FORTRAN.

PROGRAMA:

```
      REAL*4 INF(60)
      DIMENSION VENTC(60),VENTM(60),VENTL(60),TRAN(60),CPP(60),CPPR(60)
      DIMENSION PORCC(60),PORCM(60),PORCL(60),SUMA(60),PIB(60),PIBR(60)
      DIMENSION CONSC(60),CONSM(60),CONSL(60),GASO(60),DEF(60),DEFL(60)
      DIMENSION COSTC(60),COSTM(60),COSTL(60)
      OPEN (UNIT=1,FILE='COMPACTOS.',STATUS='OLD')
      OPEN (UNIT=2,FILE='MEDIANOS.',STATUS='OLD')
      OPEN (UNIT=3,FILE='LUJO.',STATUS='OLD')
      OPEN (UNIT=4,FILE='COSTUNIT.COM',STATUS='OLD')
      OPEN (UNIT=5,FILE='COSTUNIT.MED',STATUS='OLD')
      OPEN (UNIT=6,FILE='COSTUNIT.LUJ',STATUS='OLD')
      OPEN (UNIT=7,FILE='DEFLACION',STATUS='OLD')
      OPEN (UNIT=8,FILE='PIBR.MEN',STATUS='OLD')
      OPEN (UNIT=9,FILE='NOVA.',STATUS='OLD')
      OPEN (UNIT=10,FILE='CPP.',STATUS='OLD')
      OPEN (UNIT=11,FILE='TRANSP.',STATUS='OLD')
      OPEN (UNIT=12,FILE='DEFLA.DAT',STATUS='OLD')
      OPEN (UNIT=13,FILE='TABLA.INFORMACION',STATUS='NEW')
      OPEN (UNIT=14,FILE='TABLA.ANEXA',STATUS='NEW')
      DO I=1,80
      READ(1,*,END=25)VENTAC(I)
      READ(2,*)VENTAM(I)
      READ(3,*)VENTAL(I)
      READ(4,*)COSTOC(I) *1000.0
      READ(5,*)COSTOM(I) *1000.0
      READ(6,*)COSTOL(I) *1000.0
      READ(7,*)DEFL(I)
      READ(8,*)PIB(I)
      READ(9,*)GASO(I)
      READ(10,*)CPP(I)
      READ(11,*)TRAN(I)
      READ(12,*)INF(I)
      ENDDO
 25   DO I = 1,60
      SUMA(I) = COSTC(I) + COSTOM(I) +COSTOL(I)
```

```
      PORCC(I) = COSTC(I)*100 / SUMA(I)
      PORCM(I) = COSTM(I)*100 / SUMA(I)
      PORCL(I) = COSTL(I)*100 / SUMA(I)
      DEFLAC(I) = DEFL(I)*100.0 / 154.60833
      PIBR(I) = PIB(I)*100.0 / 45358.8917
      CPPR(I) = (((CPP(I)/12.0)-INF(I)) / (1.0 + INF(I))
      CONSC(I) = GASO(I)*100.0 / (9.0*DEFLAC(I))
      CONSM(I) = GASO(I)*100.0 / (9.0*DEFLAC(I))
      CONSL(I) = GASO(I)*100.0 / (9.0*DEFLAC(I))
      WRITE(13,18)I,VENTC(I),VENTM(I),VENTAL(I),PORCC(I),PORCM(I),
     1PORCL(I),CONSC(I),CONSM(I),CONSL(I),TRAN(I),PIBR(I),CPPR(I)
      WRITE(14,20) PIB(I),CPP(I),GASO(I),DELF(I),INF(I)
      ENDDO
      PRINT *, 'YA TERMINE'
 18   FORMAT(1X,F3.0,3F8.1,6F10.5/9X,2F9,2,F10.5)
 20   FORMAT(1X,4F9.2,F11.4)
      STOP
      END
```

PROGRAMA 2

ESTIMACION CONJUNTA

OBJETIVO: Estimacion conjunta del modelo.

EJECUCION: SHAZAM

PROGRAMA:

```
SIZE 40
SOLOMON 6 180  DEMANDA DE AUTOS POR CATEGORIAS.
IO LS DS
SMPL 2 180
NAME VENTAS PORCENTAJE CONSUMO TRANSPOR INGRESO CPPR
GENR TRANSPOR = TRANSPOR / LAG(TRANSPOR)
DATA
...
   [ TABLA DE INFORMACION CONJUNTA ]      (vease nota)
...
PC 2 3 4 5 6 /CR NC=4
OLS 1 2 3 4 5 6 / MAX
AUTO 1 2 3 4 5 6 / DN GS
MERCURY
```

NOTA: La "tabla de informacion conjunta" se refiere a la matriz formada al unir las matrices de informacion de cada categoria, de forma que su dimension es de (180,5).

PROGRAMA 3

ESTIMACION INDIVIDUAL Y MORAN

OBJETIVO: Componentes principales, estimacion individual y por el MORAN.

EJECUCION: SHAZAM

PROGRAMA:

```
SIZE 40
SOLOMON 13 60  DEMANDA DE AUTOS POR CATEGORIAS.
IO LS DS
SMPL 2 60
NAME TIEMPO VENTAC VENTAM VENTAL PORCC PORCM PORCL
  CONSC CONSM CONSL TRANSPOR INGRESO CPPR
GENR TRANSPOR = TRANSPOR / LAG(TRANSPOR)
DATA (1X,F3.0,3F8.1,6F10.5/9X,2F9,2,F10.5)
...

   [ TABLA DE INFORMACION ]
...

PC PORCC CONSC TRANSPOR INGRESO CPPR / CR NC=4
PC PORCM CONSM TRANSPOR INGRESO CPPR / CR NC=4
PC PORCL CONSL TRANSPOR INGRESO CPPR / CR NC=4
OLS VENTAC PORCC CONSC TRANSPOR INGRESO CPPR / MAX
OLS VENTAM PORCM CONSM TRANSPOR INGRESO CPPR / MAX
OLS VENTAL PORCL CONSL TRANSPOR INGRESO CPPR / MAX
SYSTEM 3
OLS VENTAC PORCC CONSC TRANSPOR INGRESO CPPR
OLS VENTAM PORCM CONSM TRANSPOR INGRESO CPPR
OLS VENTAL PORCL CONSL TRANSPOR INGRESO CPPR
OPTIONS DN
MERCURY
```

PROGRAMA 4

ESTIMACION CON AUTOCORRELACION

OBJETIVO: Estimacion del modelo con estructura de autocorrelacion.

EJECUCION: SHAZAM

PROGRAMA:

```
SIZE 40
SOLOMON 13 60  DEMANDA DE AUTOS POR CATEGORIAS.
IO LS DS
SMPL 2 60
NAME TIEMPO VENTAC VENTAM VENTAL PORCC PORCM PORCL
  CONSC CONSM CONSL TRANSPOR INGRESO CPPR
GENR TRANSPOR = TRANSPOR / LAG(TRANSPOR)
DATA (1X,F3.0,3F8.1,6F10.5/9X,2F9,2,F10.5)
...

   [ TABLA DE INFORMACION ]
...

AUTO VENTAC PORCC CONSC TRANSPOR INGRESO CPPR / MAX
AUTO VENTAM PORCM CONSM TRANSPOR INGRESO CPPR / MAX
AUTO VENTAL PORCL CONSL TRANSPOR INGRESO CPPR / MAX
MERCURY
```

PROGRAMA 5

MATRICES DE TRANSFORMACION

OBJETIVO: Genera las matrices de transformacion para MOVAR.
El programa es un ejemplo para cuando N=2 y T=5

EJECUCION: FORTRAN

PROGRAMA:

```
      DIMENSION A(10,10),B(10,10),Q(10,10),Q1(10,10),Q2(10,10),Q3(10,10),
      INTEGER T
     * Para otros casos, cambiar las siguientes dimensiones:
      N = 2
      T = 5
      NT = N*T
      UN = 1.0/N
      UT = 1.0/T
      UNT = 1.0/NT
      DO I=1, NT
            DO J=1, NT
                  Q1(I,J) = -UNT
                  Q2(I,J) = -UNT
                  Q3(I,J) = UNT
                  IF(I.EQ.J)THEN
                    A(I,J) = 1.0
                    B(I,J) = 1.0
                  ENDIF
            ENDDO
      ENDDO
      DO L=1, N
            DO I=(L-1)*T+1, L*T
                  K = I-(L-1)*T
                  DO J=(L-1)*T+1, L*T
                    A(I,J) = A(I,J) - UT
                    Q1(I,J) = UT - UNT
                  ENDDO
            ENDDO
      ENDDO
      DO I=1,NT
            DO J=1,NT
                Q(I,J) = B(I,J) - Q1(I,J)
            ENDDO
      ENDDO
      PRINT *, 'LA MATRIZ A ES:'
```

```
      PRINT 10,A
      PRINT *, 'LA MATRIZ B ES:'
      PRINT 10,B
      PRINT *, 'LA MATRIZ Q ES:'
      PRINT 10,Q
      PRINT *, 'LA MATRIZ Q1 ES:'
      PRINT 10,Q1
      PRINT *, 'LA MATRIZ Q2 ES:'
      PRINT 10,Q2
      PRINT *, 'LA MATRIZ Q3 ES:'
      PRINT 10,Q3
10    FORMAT (NT(1X,F5.2)//)
      STOP
      END
```

EJEMPLO DEL PROGRAMA 5

MATRICES DE TRANSFORMACION (N=2, T=5)

OBJETIVO: Presentar los resultados obtenidos al ejecutar el programa 5.

RESULTADOS:

LA MATRIZ A ES:

0.8	-0.2	-0.2	-0.2	-0.2	0.0	0.0	0.0	0.0	0.0
-0.2	0.8	-0.2	-0.2	-0.2	0.0	0.0	0.0	0.0	0.0
-0.2	-0.2	0.8	-0.2	-0.2	0.0	0.0	0.0	0.0	0.0
-0.2	-0.2	-0.2	0.8	-0.2	0.0	0.0	0.0	0.0	0.0
-0.2	-0.2	-0.2	-0.2	0.8	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.8	-0.2	-0.2	-0.2	-0.2
0.0	0.0	0.0	0.0	0.0	-0.2	0.8	-0.2	-0.2	-0.2
0.0	0.0	0.0	0.0	0.0	-0.2	-0.2	0.8	-0.2	-0.2
0.0	0.0	0.0	0.0	0.0	-0.2	-0.2	-0.2	0.8	-0.2
0.0	0.0	0.0	0.0	0.0	-0.2	-0.2	-0.2	-0.2	0.8

LA MATRIZ B ES:

0.5	0.0	0.0	0.0	0.0	-0.5	0.0	0.0	0.0	0.0
0.0	0.5	0.0	0.0	0.0	0.0	-0.5	0.0	0.0	0.0
0.0	0.0	0.5	0.0	0.0	0.0	0.0	-0.5	0.0	0.0
0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	-0.5	0.0
0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	-0.5
-0.5	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0
0.0	-0.5	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0
0.0	0.0	-0.5	0.0	0.0	0.0	0.0	0.5	0.0	0.0
0.0	0.0	0.0	-0.5	0.0	0.0	0.0	0.0	0.5	0.0
0.0	0.0	0.0	0.0	-0.5	0.0	0.0	0.0	0.0	0.5

LA MATRIZ Q ES:

0.7	-0.1	-0.1	-0.1	-0.1	-0.4	0.1	0.1	0.1	0.1
-0.1	0.7	-0.1	-0.1	-0.1	0.1	-0.4	0.1	0.1	0.1
-0.1	-0.1	0.7	-0.1	-0.1	0.1	0.1	-0.4	0.1	0.1
-0.1	-0.1	-0.1	0.7	-0.1	0.1	0.1	0.1	-0.4	0.1
-0.1	-0.1	-0.1	-0.1	0.7	0.1	0.1	0.1	0.1	-0.4
-0.4	0.1	0.1	0.1	0.1	0.7	-0.1	-0.1	-0.1	-0.1
0.1	-0.4	0.1	0.1	0.1	-0.1	0.7	-0.1	-0.1	-0.1
0.1	0.1	-0.4	0.1	0.1	-0.1	-0.1	0.7	-0.1	-0.1
0.1	0.1	0.1	-0.4	0.1	-0.1	-0.1	-0.1	0.7	-0.1
0.1	0.1	0.1	0.1	-0.4	-0.1	-0.1	-0.1	-0.1	0.7

LA MATRIZ Q1 ES:

0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	-0.1	-0.1	-0.1
0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	-0.1	-0.1	-0.1
0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	-0.1	-0.1	-0.1
0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	-0.1	-0.1	-0.1
0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	0.1	0.1	0.1	0.1	0.1
-0.1	-0.1	-0.1	-0.1	-0.1	0.1	0.1	0.1	0.1	0.1
-0.1	-0.1	-0.1	-0.1	-0.1	0.1	0.1	0.1	0.1	0.1
-0.1	-0.1	-0.1	-0.1	-0.1	0.1	0.1	0.1	0.1	0.1
-0.1	-0.1	-0.1	-0.1	-0.1	0.1	0.1	0.1	0.1	0.1

LA MATRIZ Q2 ES:

0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1
-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1
-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1
-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1
-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4
0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1
-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1
-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1
-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4	-0.1
-0.1	-0.1	-0.1	-0.1	0.4	-0.1	-0.1	-0.1	-0.1	0.4

LA MATRIZ Q3 ES:

-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1

BIBLIOGRAFIA

DIELMAN, Terry E.
"Pooled Cross-Sectional and Time Series Data: a survey of current statistical methodology"
The American Statistician, Vol 37, No. 2 ,May, 1983,

GRAYBILL
"Introduction to Matrices with Applications in Statistics"
Wadsworth Publishing Co., Inc. ,1969,

GUENTHER, WILLIAM C.
"Introduccion a la Inferencia Estadistica"
Mc Graw Hill, 1968, Madrid

HADLEY,G.
"Linear Algebra"
Addisson Wesley ,1961,

INTRILIGATOR, Michael D.
"Econometric Models, Techniques & Applications"
Prentice Hall ,1978, Englewood Cliffs, New Jersey

JOHNSTON, J.
"Econometric Methods", 2nd. Edition
Mc Graw Hill ,1972, New York

JUDGE George G., GRIFFITHS William, HILL Carter & LEE Tsoung-Chao
"The theory and practice of Econometrics"
John Wiley & Sons ,1980

KMENTA, Jan
"Elements of Econometrics"
Macmillan, 1971, New York

MUNDLAK, Yair
"On the pooling of time series and cross section data"
Econometrica, Vol. 46, No.1 ,January, 1978

PINDYCK, Robert S and RUBINFELD Daniel L.
"Econometric Models and Economic Forecasts"
Mc Graw Hill ,1981, Singapore

SNEDECOR, George and COCHRAN, William
"Statistical Methods"
Iowa State

SOUZA, Rita
"Algebra Lineal"
Universidad Anahuac ,1980, Mexico D.F.

THEIL, Henry
"Principles of Econometrics"
John Wiley & Sons, Inc. ,1971, New York

WHITE, KENNETH J.
"SHAZAM, an econometric computer program", Version 4.1
University of British Columbia, Mayo, 1982 Vancouver, Canada

OTRAS FUENTES:

ASOCIACION MEXICANA DE LA INDUSTRIA AUTOMOTRIZ
"Venta mensual-anual de automoviles por categorias y marcas 1975-1984"
Estadisticas A.M.I.A., 1985

BANCO DE MEXICO
"Indice de precios al consumidor"
Ediciones de enero de 1980 a diciembre de 1984

BANCO NACIONAL DE MEXICO
"Producto Interno Bruto desagregado mensualmente"
Centro de Estudios Economicos, 1985

BANCO NACIONAL DE MEXICO
"Costo Porcentual Promedio"
Centro de Estudios Economicos, 1985

PETROLEOS MEXICANOS
Precios al consumidor
Enero de 1980 a diciembre de 1984

SECRETARIA DE PROGRAMACION Y PRESUPUESTO
"Encuesta Industrial Mensual: Volumen de la produccion de los principales productos manufacturados"
Direccion General de Estadistica, Instituto Nacional de Estadistica, Geografia e Informatica, 1981-1985

SECRETARIA DE PROGRAMACION Y PRESUPUESTO
"Encuesta Industrial Mensual: Valor de la produccion de los principales articulos manufacturados"
Direccion General de Estadistica, Instituto Nacional de Estadistica, Geografia e Informatica, 1981-1985