19
2 ej.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Facultad de Ciencias

"LA MULTICOLINEALIDAD EN EL ANALISIS DE REGRESION: TECNICAS USUALES DE DIAGNOSTICO Y CORRECCION"

T E S I S

Que para obtener el título de:

M A T E M A T I C O

P r e s e n t a :

OCTAVIO LUIS PINEDA

México, D. F., febrero 1988

C O N T E N I D O

Página

I. SUMARIO

El propósito del presente trabajo es triple. En primer lugar, presentar el problema de la multicolinealidad como una 'enfermedad estadística' que acontece muy frecuentemente en el análisis de regresión y en particular en econometría, así como sus más obvias e inmediatas consecuencias en la estimación paramétrica, inferencia estadística, especificación funcional y predicción en modelos econométricos uniecuacionales.

En segundo lugar, y bajo este mismo contexto, describir y discutir las técnicas y métodos más conocidos de detección y diagnóstico para percatarse de su presencia y medir su grado de intensidad.

En tercer y último lugar, presentar y discutir una lista de las técnicas disponibles y más ampliamente utilizadas hoy en día para atacarla.

Conviene destacar, para finalizar, que no es la pretensión de este trabajo adentrarse en la demostración o justificación matemática rigurosa de cada una de las técnicas y métodos descritos en el mismo - que a lo sumo se bosquejan o se menciona la referencia correspondiente para el lector interesado - sino fundamntalmente, buscar su aplicación para entender, reconocer y corregir la multicolinealidad, utilizando para ello casos concretos y ejemplos ilustrativos tomados de la economía y ciencias sociales, básicamente, para así comprender mejor su naturaleza, cómo se manifiesta y afecta a un modelo de regresión o econométrico (uniecuacional) y la manera de resolverla en la práctica.

II. EL PROBLEMA DE LA MULTICOLINEALIDAD EN EL ANALISIS DE REGRESION.

2.1 INTRODUCCION.

La interpretación del modelo de regresión depende implícitamente del supuesto de que las variables explicativas no están fuertemente interrelacionadas entre sí. Es común interpretar a un coeficiente de regresión como la medida del cambio de la variable-respuesta cuando la variable explicativa correspondiente se incrementa en una unidad y todas las demás permanecen constantes. Sin embargo, esta interpretación (como efecto marginal del coeficiente referido sobre la variable dependiente) puede no ser válida cuando existe una fuerte relación lineal entre las variables explicativas del modelo.

Aunque teóricamente es siempre posible incrementar el valor de una variable explicativa y mantener a las otras constantes en un modelo de regresión, en la práctiva, puede no existir información sobre el resultado de dicha manipulación en los datos para estimación. Es más, puede incluso ser imposible hacer variar a una variable explicativa y mantener a las otras constantes en el proceso bajo estudio. Cuando se presenta un nivel peligroso de multicolinealidad, la interpretación simple del coeficiente de regresión como un "efecto marginal", aludido anteriormente, se pierde.

Cuando existe ausencia completa de relación lineal entre las variables explicativas, se dice que éstas son "ortogonales". En la mayoría de las aplicaciones en econometría y las ciencias sociales, las variables explicativas no son ortogonales. Normalmente, la falta de ortogonalidad no es un gran obstáculo para efectuar el análisis. Empero, en algunas circunstancias, las variables explicativas están tan fuertemente interrelacionadas que los resultados de la regresión resulten ambiguos. Típicamente lo que sucede en dichos casos, es que resulta imposible estimar los efectos "aislados" de variables individuales en el modelo de regresión; los valores estimados de los coeficientes son muy sensibles a cambios pequeños en los datos y a la adición o eliminación de variables en la ecuación y los coeficientes de regresión poseen variancias grandes que afectan tanto la inferencia como el pronóstico, basados en el modelo en cuestión.

La condición de no-ortogonalidad severa se conoce en la literatura técnica como problema de "datos colineales", "colinealidad" , "interconexión" o "multicolinealidad". Utilizaremos con más frecuencia esta última manera de nombrar a este problema, por ser la más conocida. La multicolinealidad puede ser extremadamente difícil de detectar. No se trata de un error de especificación que puede descubrirse fácilmente explorando los residuales de la regresión. De hecho, como veremos después, este problema no es un error de modelaje, aunque puede darse, claro está. Sino fundamentalmente, es una condición de deficiencia en los datos.

De cualquier manera, es importante conocer cuando se presenta y estar prevenidos de sus posibles consecuencias. De ahí la recomendación y advertencia al investigador para que sea extremadamente precavido al formular conclusiones sustantivas basadas en un modelo de regresión o econométrico en presencia de multicolinealidad.

A pesar de ser un problema añejo 1/ y de la incuestionable importancia de la multicolinealidad como una " enfermedad estadística" en el análisis de regresión y más concretamente en el proceso de formulación de modelos econométricos - dada su frecuencia y efectos nocivos - no existe en nuestro medio, literatura en nuestro idioma sobre este tema, ni mucho menos accesible para la mayoría de los no-especializados en este campo, para enfrentarla y resolverla.

El fin último de este trabajo, es por consiguiente, coadyuvar modestamente a llenar este enorme vacío poniendo en manos del investigador interesado, sobre todo del no-experto en la materia, y particularmente al del campo de las ciencias sociales, económicas y administrativas, un instrumento de consulta -compendiado y accesible- diseñado de tal forma de proporcionarle

---

1/ Ragnar Frisch fue uno de los primeros en analizar el problema de la multicolinealidad desde 1934, observando que muchas series de tiempo o cronológicas, sujetas como están a un -- conjunto común de fluctuaciones económicas, se movían juntas a través del tiempo , y por consiguiente, resultaban altamente colineales.

Véase, Frisch, R. Statistical Confluence Analysis by Means of Complete Regression Systems. Oslo Norway: University Economics Institute, 1934.

los elementos de juicio suficientes para comprender, manejar y encarar con éxito este problema tan común y frecuente en el curso de toda formulación, análisis o validación de un modelo de regresión o econométrico, derivado o como parte integral de su quehacer de investigación rutinaria dentro del campo del análisis de regresión o econométrico.

2.2 DEFINICION DEL PROBLEMA EN EL CONTEXTO DEL MODELO LINEAL GENERAL.

Antes de comenzar formalmente con nuestra discusión sobre el problema de la multicolinealidad y las técnicas existentes para atacarla, trataremos primeramente de definir el problema.

Sabemos que el modelo general de regresión múltiple, también llamado "modelo general de regresión lineal", o sencillamente, "modelo lineal general", se puede escribir, en forma escalar, como:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{k,i} + U_i \qquad (1)$$

donde: $i = 1,2,3,\ldots,n$

O alternativamente, como un sistema lineal:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k,1} + U_1 \\ Y_2 &= \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_k X_{k,2} + U_2 \\ Y_3 &= \beta_0 + \beta_1 X_{13} + \beta_2 X_{23} + \cdots + \beta_k Y_{k,3} + U_3 \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ Y_n &= \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \cdots + \beta_k X_{k,n} + U_n \end{aligned} \qquad (2)$$

O también, en términos matriciales:

$$Y = X\beta + U \qquad (3)$$

donde las matrices Y, X, y U se definen como:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} & X_{31} \cdots & X_{k,1} \\ 1 & X_{12} & X_{22} & X_{32} \cdots & X_{k,2} \\ 1 & X_{13} & X_{23} & X_{33} \cdots & X_{k,3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{1n} & X_{2n} & X_{3n} \cdots & X_{k,n} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}; \quad U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ \vdots \\ U_n \end{bmatrix} \qquad (4)$$

Sin embargo, la expresión más comúnmente utilizada en el --- análisis de regresión y una de las más empleadas en este trabajo, al menos que se especifique lo contrario, es la "forma de desviación con respecto a la media", misma que se puede obtener , mediante una sencilla transformación a las variables del modelo (1). Esto es:

$$x_{ij} = X_{ij} - \bar{X}_i \ ; \quad y_j = Y_j - \bar{Y} \ ; \qquad \begin{array}{l} i=1,2,\ldots,k \\ j=1,2,\ldots,n \end{array} \tag{5}$$

Si aplicamos la transformación anterior al modelo (1) definido anteriormente, se puede derivar una nueva expresión para el -- mismo en los siguientes términos:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{k,i} + u_i \tag{6}$$

que denominaremos por brevedad como "forma compacta" de (1) y que desarrollada, genera al sistema lineal:

$$\begin{array}{l} y_1 = \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_k x_{k,1} + u_1 \\ y_2 = \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_k x_{k,2} + u_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ y_n = \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_k x_{k,n} + u_n \end{array} \tag{7}$$

y cuya notación matricial resulta idéntica a la expresión(3), esto es:

$$Y = X\beta + u$$

aunque las matrices que ahora la integran difieren considerablemente de las definidas en la expresión (4), es decir: Obser vamos de entrada que la matríz X de datos es ahora de tamaño $n \times k$ en lugar de $n\ x(k+1)$ y el vector -columna β de pará-- metros es ahora de dimensión k en lugar de dimensión (k+1), sin contar que los elementos de estas matrices han sido transfor-- mados previamente a su forma "compacta".

Por otra parte, en el modelo lineal general definido en la ex- presión (3) suponemos implícitamente que se cumplen cada una de las siguientes condiciones o hipótesis generales;

Primera; $E\ (U) = 0$

Segunda; $E\ (U\ U') = \sigma_u^2 I_n$; $\sigma_u^2 =$ constante > 0.

Tercera; X es una matríz (nxk) de números fijos. (8)

Cuarta; Rango (X) = k > n ; donde n= número de observacio- nes en la muestra, k = número de variables explicati- vas en el modelo.

La primera suposición implica que la media de la variable aleato-- ria, término de perturbación o estocástico "U" es cero.

La segunda suposición implica que $E(Ui^2) = \sigma_u^2$ y además -- que $E(Ui\ Uj) = 0$, $i \neq j$, para toda $i, j = 1, 2, \ldots, n$. Esto quiere decir en otras palabras, que la variancia del término estocástico o de perturbación es constante y que la covariancia de dos términos aleatorios distintos es nula.

De las dos suposiciones anteriores se puede inferir, en particular, que no existe relación lineal alguna (esto, no existe correlación) entre las variables explicativas incluidas en el modelo y el término de perturbación. Consideramos importante resaltar este hecho debido a su relevancia en nuestra discusión posterior.

La tercera suposición es equivalente a decir que la matriz X de datos con que estamos trabajando es un "conjunto matemáticamente fijo", lo que significa simplemente que en todas las --- pruebas de hipótesis y procedimientos de estimación, se consideran a las variables explicativas o independientes X1,X2,... ...,Xk como si fueran "constantes". Es decir, en muestras repetidas, se acepta o supone que la única fuente de variación de la variable dependiente "Y" es la proveniente del término estocástico o vector "U", y por consiguiente, todas las propiedades de nuestros estimadores y las pruebas de hipótesis realizadas sobre los valores de éstos, están condicionadas a los valores de la matríz X.

Por último, la cuarta suposición implica simplemente la no- existencia de dependencia lineal entre todas las columnas (o hileras) de la matríz X. En especial, si denotamos como X1,X2,...,Xk a la primera, segunda, tercera,..., k- ésima columna de X, tendremos un conjunto de vectores linealmente independientes. Es decir, dada una combinación lineal de ellos igualada a cero:

$$a_1 X_1 + a_2 X_2 + \cdots + a_k X_k = 0 \tag{9}$$

entonces y solo entonces:

$$a_1 = a_2 = \cdots = a_k = 0 \tag{9-bis}$$

A X1,X2,...,Xk, se les conoce indistintamente como "variables explicativas" o "variables independientes" del moddlo (1). Y a Yi, como la "variable - respuesta" o "dependiente" del referido modelo.

La cuarta suposición es sumamente importante ya que de no cumplirse, no se pueden estimar los parámetros del modelo (1). O en otras palabras, no se puede conocer el efecto individual o separado de las variables explicativas sobre la variable-respuesta.

Entonces, la suposición de independencia de las variables explicativas es asegurar que el determinante $|X'X| \neq 0$, y por lo tanto, el sistema de ecuaciones normales generado a partir del sistema (2) tenga una solución única. Esto resulta evidente — también recordando que nuestro estimador de mínimos cuadrados tiene la forma:

$$\hat{\beta} = (X'X)^{-1} X'Y \qquad (10)$$

y como su cómputo requiere forzosamente de la inversión de la matríz $(X'X)$(') , nos damos cuenta inmediatamente de la -- imposibilidad de lograrlo si el rango de X es menor que k.O sea, si: 1/

$$\text{rango }(X) < K \qquad (10\text{-bis})$$

entonces el número de variables explicativas sería menor que el número de parámetros existentes en el modelo, contrario a la **cuarta** suposición. Y como consecuencia de ello, tendríamos que el determinante:

$$|X'X| = 0 \qquad (11)$$

luego entonces, no existe la inversa de (X'X) y la expresión definida en (10) no tiene o carece de sentido . Y cuando -- esto acontece decimos que estamos ante un caso de "extrema o perfecta multicolinealidad". Esto es, algunas o todas las variables explicativas de nuestro modelo, son perfectamnte coli-

---

(') Utilizando la notación compacta definida anteriormente para los elementos de X, el producto (X'X) **tiene la forma:**

$$(X'X) = \begin{bmatrix} \sum x_1^2 & \sum x_1x_2 & \dots & \sum x_1x_k \\ \sum x_2x_1 & \sum x_2^2 & \dots & \sum x_2x_k \\ \vdots & \vdots & & \vdots \\ \sum x_kx_1 & \sum x_kx_2 & \dots & \sum x_k^2 \end{bmatrix}$$

1/ Para una discusión pormenorizada sobre la condición **respecto del rango** de la matríz X,
Véase, Tintner, G. A Note on Rank, Multicollinearity and Multicollinearity and Multiple Regression. Annals of Mathematical Statistics, Vol.16, 1945, pp.304-308.

neales.

Aunque la multicolinealidad perfecta es teóricamente factible, en la práctica, como se verá después, el determinante $|(X'X)|$ no necesita ser cero para generar un nivel "peligroso" (o inaceptable) de multicolinealidad.(+) Esto sucede cuando la hipótesis:

$$|X'X| \neq 0 \qquad (12)$$

se cumple marginalmente.(++) Cuando esto sucede decimos que las variables explicativas en cuestión son "altamente" pero no "perfectamente" colineales.

Por otro lado, sabemos del análisis de regresión por mínimos cuadrados ordinarios, que la matríz de variancia-covariancia del estimador $\hat{\beta}$ tiene la forma:

$$Var(\hat{\beta}) = \sigma_u^2 (X'X)^{-1} \qquad (13)$$

que pone de manifiesto, otra vez, el papel crucial de la no-singularidad de la matríz (X'X).

(+) Aunque más adelante se da una definición formal respecto a lo que entenderemos por "peligroso" o "inaceptable", por el momento, podemos adelantar que intuitivamente esto significa una fuerte dependencia lineal entre las variables explicativas del modelo, que sin ser perfecta, genera estimadores paramétricos muy poco confiables.

(++) En la práctica, la computadora fija arbitrariamente un valor muy "cercano" a cero para el valor del determinante mencionado (como cota inferior) de tal suerte que cualquier valor mayor que éste, resulta ser diferente de cero para efectos de cómputo, y por ende, puede cumplir la condición (12) y procede a calcular la matríz $(X'X)^{-1}$ y estimar al vector β, aunque de hecho el determiante $|X'X|$ sea cero.

Aunque lo intentaremos mostrar más adelante, es evidente que conforme crece la interdependencia entre las variables explicativas Xi's , la matríz (X'X) tiende a hacerse singular y los elementos de la matríz inversa $(X'X)^{-1}$ tienden a hacerse muy grandes. Y en el límite, la dependencia lineal perfecta entre las variables X1,X2,...,Xk generará un conjunto completamente indeterminado de estimadores paramétricos $\hat{B}o,\hat{B}1,\hat{B}2,..,\hat{B}k$. Más formalmente, esto quiere decir que los elementos diagonales de la matríz $(X'X)^{-1}$ que corresponden a elementos linealmente dependientes de X, tienden a infinito . Y por consiguiente, las variancias respectivas de los estimadores de los coeficientes cuyas variables están afectadas, también crecen desmesuradamente.

III. LAS PRINCIPALES CONSECUENCIAS DE LA MULTICOLINEALIDAD.

Antes de comenzar a discutir las consecuencias inmediatas de la multicolinealidad en diversos aspectos de un modelo de regresión o econométrico, se deben aclarar dos puntos fundamentales en opinión de J.Kmenta[1/]:

1. La multicolinealidad es una cuestión de grado y no de clase. La distinción significativa no es entre la presencia o ausencia de este problema en un modelo, sino entre sus varios grados.

2. Como la multicolinealidad se refiere a una condición sobre las variables explicativas que se suponen no-estocásticas, entonces es una característica de la muestra y no de la población bajo estudio.[2/]

Todo lo anterior implica que la multicolinealidad es una condición de interdependencia entre X1,X2,...,Xk que puede existir independientemente de la dependencia entre la matríz X y la variable Y. Esto es, la multicolinealidad es una propiedad del conjunto de variables explicativas únicamente.[3/] No importando el --

1/ Véase, Kmenta,J. Elements of Econometrics. The Macmillan Company,1971,pp.380-391

2/ Véase, Kmenta,J. Ibid.

3/ Cuando ocurre la multicolinealidad, es como si los miembros del subconjunto de variables explicativas afectadas actuaran en unísono. Como resultado de ello, los datos carecen de suficiente variación independiente para permitirnos determinar el efecto separado de cada variable explicativa. Entonces el problema de la multicolinealidad no es un problema del método de estimación, ni de la población, sino de los datos, un problema básico de la forma de "generación" de las observaciones, un fenómeno muy frecuente en econometría por ejemplo, donde se trabaja con datos de ciencias no-experimentales, como la economía y las ciencias sociales en general.

Véase, Kmenta,J. Ibid.

grado o inclusive la existencia de alguna relación de dependencia o causalidad entre X y Y. 1/

Con estos antecedentes como marco, podemos pasar ahora a describir los principales efectos de la multicolinealidad en un modelo de regresión o econométrico, en la estimación paramétrica, inferencia estadística, especificación funcional y predicción, respectivamente.

1/ A propósito, el hecho de que no exista interdependencia entre las variables explicativas de un modelo no implica que no exista o pueda existir una relación de causalidad o "efecto de interacción" entre éstas y la variable-respuesta. Por ejemplo, una cosa es decir que en el modelo lineal C=f(Y,L) donde C representa al consumo o demanda agregada, Y al ingreso agregado, y L los activos líquidos en una economía, y que estas dos variables explicativas seanaltamente colineales, dada su naturaleza económica, y por lo mismo, no puedan estimarse independientemente, y otra cosa muy distinta, es implicar que no exista un efecto de interacción entre las mismas y la variable dependiente. Esto es, puede darse el caso que la función definida anteriormente sea en realidad de la forma: C = f (Y,L,YL). Entonces el problema de la multicolinealidad es tratar de estimar los coeficientes (β_1 y β_2) asociados a las variables Y,L cuando éstas son "altamente" colineales. Cómo obtener e interpretar un estimador para el coeficiente (β_3) asociado a la variable YL, es un asunto totalmente diferente, que no será abordado en este trabajo. Para los interesados en este tema, le sugerimos la siguiente fuente:

Véase, Sonquist, J.A. y J.N. Morgan. The Detection of Interaction Effects. Survey of Research Center, Monograph No.35, Ann Arbor, Michigan: University of Michigan Press, 1964.

3.1 EFECTO DE LA MULTICOLINEALIDAD EN LA ESTIMACION PARAMETRICA.

Los grandes valores de las variancias de los estimadores provenientes de variables multicolineales, indican muy apropiadamente, el bajo contenido de información de los datos observados y por consiguiente, son un indicativo de la baja calidad de los mismos.

Discutiremos primeramente el efecto de la multicolinealidad en la estimación paramétrica cuando aquélla es perfecta. Para este propósito usaremos el enfoque de L.R.Klein.1/

Si transformamos las variables originales del modelo (1) mediante la transformación:(+)

$$y_t^* = \frac{Y_t - \bar{Y}}{S_Y}, \quad x_{it}^* = \frac{X_{it} - \bar{X}_i}{S_i}; \qquad S_Y = \sqrt{\frac{\sum (Y_t - \bar{Y})^2}{n}} \quad S_i = \sqrt{\frac{\sum (X_{it} - \bar{X}_i)^2}{n}} \tag{14}$$

donde: $t = 1, 2, \ldots, n$; $i = 1, 2, \ldots, K$ y Sy, Si, son las desviaciones estándard de Y y Xi, respectivamente. Obtendremos una versión "estandarizada" del modelo (1), como sigue: (++)

$$y^* = \beta_1 x_1^* + \beta_2 x_2^* + \ldots + \beta_K x_K^* + u_i \tag{15}$$

1/ Véase, Klein, L.R. A Textbook of Econometrics. Prentice-Hall, Inc., 1974, pp. 104-108

(+) Nótese que: $E(x_i^*) = 0$, $E(y_i^*) = 0$, $Var(x_i^*) = 1$, $Var(y_i^*) = 1$ y $cov(x_i^*, y_i^*) = r_{ij}$.

(++) Por simplicidad hemos eliminado los subíndices correspondientes de esta ecuación.

Nótese que esta versión, no debe confundirse con la forma compacta del modelo lineal general definida en (6). Asimismo, obsérvese que las variables estandarizadas definidas en (14) tienen media cero y variancia unitaria. Además ,si denotamos -- "r_{ij}" y "r_{iy}" como los coeficientes de correlación simple entre las variables X_i y X_j y entre X_i y Y_i, respectivamente. Entonces, es claro que $r_{ii} = 1$.

Entonces, según L.R.Klein 1/ , el estimador $\hat{\beta}_i$ de (1) se puede expresar como:

$$\hat{\beta}_i = \frac{\begin{vmatrix} 1 & r_{12} & \dots & r_{1y} & \dots & r_{1k} \\ r_{12} & 1 & \dots & r_{2y} & \dots & r_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{1k} & r_{2k} & \dots & r_{ky} & \dots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} & \dots & r_{1i} & \dots & r_{1k} \\ r_{12} & 1 & \dots & r_{2i} & \dots & r_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{1k} & r_{2k} & \dots & r_{ki} & \dots & 1 \end{vmatrix}} \tag{16}$$

Ahora bien, si dos elementos X_i, X_j del vector $(X_1, X_2, \dots, X_k)$ estuvieran perfectamente correlacionados, esto es, si $r_{ij} = 1$, entonces la i-ésima y j-ésima hilera (o columna) del denominador de la expresión (16) serían idénticas y el determinante se anularía. Esto es evidente, ya que si observamos las hileras referidas con más ddtenimiento:

$$\begin{matrix} r_{i1} & r_{i2} & r_{i3} & \dots & r_{ii} & \dots & r_{ij} & \dots & r_{ik} \\ r_{j1} & r_{j2} & r_{j3} & \dots & r_{ji} & \dots & r_{jj} & \dots & r_{jk} \end{matrix} \tag{17}$$

donde sabemos que $r_{ii} = 1$, $r_{jj} = 1$, pero como $r_{ij} = 1$ por --

---

1/ Véase, Klein, L.R.(1974). Opus cit.

hipótesis, entonces $r_{ij} = 1$, y por consiguiente, tenemos dos hileras idénticas y el determinante anterior se anula.

Observemos ahora la i-ésima y la j-ésima hilera del numerador de la misma expresión para ver que sucede:

$$\begin{matrix} r_{i1} & r_{i2} & r_{i3} \cdots r_{iy} \cdots r_{ij} \cdots\cdots r_{ik} \\ r_{j1} & r_{j2} & r_{j3} \cdots r_{jy} \cdots r_{jj} \cdots\cdots r_{jk} \end{matrix} \qquad (18)$$

En la i-ésima columna de la expresión (18) tenemos a r_{iy}, r_{jy} pero ambos son idénticos, ya que hemos supuesto que $r_{ij} = 1$ y como consecuencia de ello, el numerador de la expresión (16) se hace cero.

Esto es, si tenemos una pareja (X_i, X_j) de variables explicativas que están perfectamente correlacionadas, entonces el estimador $\hat{\beta}_i$ resulta indeterminado. O sea:

$$\hat{\beta}_i = \frac{0}{0} \qquad (19)$$

Se infiere fácilmente que obtendremos la misma indeterminación si existiera una relación lineal exacta entre todas las variables explicativas. Lo que implica simplemente que no podemos determinar el valor de $\hat{\beta}_i$, lo cual era de esperarse, pues si X_i y X_j están perfectamente correlacionadas, no es posible evaluar su efecto separado sobre Y.

Por otra parte, analicemos ahora el efecto de la multicolinealidad en la variancia estimada de $\hat{\beta}_i$. De acuerdo a L.R. Klein 1/ la

---

1/ Véase, Klein, L.R. (1974). Opus cit.

variancia estimada de $\hat{\beta}_i$ se puede expresar como:

$$\widehat{Var}(\hat{\beta}_i) = \hat{\sigma}_u^2 \frac{|(X'X)_{ii}|}{|(X'X)|} \qquad (20)$$

donde el término $\hat{\sigma}_u^2$ corresponde a la variancia estimada del término de perturbación que puede escribirse a su vez como:

$$\hat{\sigma}_u^2 = \hat{\sigma}_e^2 = \frac{\sum e_i^2}{n-k} = \frac{\sum (y_i-\hat{y}_i)^2}{n-k} = \frac{\sum (Y_i-\hat{Y}_i)^2}{n-k} \qquad (20\text{-bis})$$

utilizando para ello, la versión compacta del modelo lineal general.(+) La matríz (X'X)ii de la expresión (20), resulta de eliminar de la matríz (X'X) definida enseguida, la i-ésima hilera e i-ésima columna. Esto es:

$$(X'X) = \begin{bmatrix} \sum x_1^{*2} & \sum x_1^* x_2^* & \dots & \sum x_1^* x_k^* \\ \sum x_2^* x_1^* & \sum x_2^{*2} & \dots & \sum x_2^* x_k^* \\ \cdot & \cdot & \cdots & \cdot \\ \sum x_k^* x_1 & \sum x_k^* x_2^* & \dots & \sum x_k^{*2} \end{bmatrix} \qquad (21)$$

O en otros términos, la matríz (X'X)ii queda definida como:

$$(X'X)_{ii} = \begin{bmatrix} \sum x_1^{*2} & \dots & \sum x_1^* x_{i-1}^* & \sum x_1^* x_{i+1}^* & \dots & \sum x_1^* x_k^* \\ \sum x_2^* x_1^* & \dots & \sum x_2^* x_{i-1}^* & \sum x_2^* x_{i+1}^* & \dots & \sum x_2^* x_k \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \sum x_{i-1}^* x_1^* & \dots & \sum x_{i-1}^* x_{i-1}^* & \sum x_{i-1}^* x_{i+1}^* & \dots & \sum x_{i-1}^* x_k^* \\ \sum x_{i+1}^* x_1^* & \dots & \sum x_{i+1}^* x_{i-1}^* & \sum x_{i+1}^* x_{i+1}^* & \dots & \sum x_{i+1}^* x_k^* \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \sum x_k^* x_1^* & & \sum x_k^* x_{i-1}^* & \sum x_k^* x_{i+1}^* & \dots & \sum x_k^{*2} \end{bmatrix} \qquad (22)$$

Como en la matríz anterior y que también aparece en el numerador de la expresión (20) no existen la i-ésima hilera e i-ésima columna, su determinante no se hace cero o anula en --

(+) Nótese aquí que "e_i" es el término de perturbación estimado, conocido en la literatura técnica como término o elemento "residual" del modelo en cuestión. Esto es, e_i es el estimador correspondiente a u_i que no se conoce. Y se le puede definir formalmente como: $e_i = \hat{u}_i = y_i - \hat{y}_i = Y_i - \hat{Y}_i$

virtud de la perfecta correlación entre Xi y Xj, en tanto su denominador - que es esencialmente el mismo que el definido para el estimador $\hat{\beta}_i$ en la expresión (16) - sí se hace cero. Luego entonces:(+)

$$\widehat{Var}(\hat{\beta}_i) = \frac{\hat{\sigma}_u^2 \left|(X'X)_{ii}\right|}{0} \quad \text{(no está definido)} \tag{23}$$

Afortunadamente, a pesar de todo lo anterior, la multicolinealidad no afecta a la variancia residual del modelo lineal general ($\hat{\sigma}_e^2$) aún cuando los estimadores obtenidos sean indeterminados. Esto es, a pesar de que exista multicolinealidad perfecta entre las variables. Este resultado es particularmente importante y será utilizado posteriormente a lo largo de nuestras discusiones .

Supóngase que rij = 1 y que el modelo de regresión ;

$$y_i = \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_{i-1} x_{i-1} + \beta_{i+1} x_{i+1} + \cdots + \hat{\beta}_k x_k + e_i \tag{24}$$

ha sido estimado. Entonces, de la 1a. y 2a. hipótesis de nuestro modelo general de regresión se infiere que:

$$E(u_t\, x_j) = 0 \tag{25}$$

donde: j =1,2,...,i-1,i+1,...,k ; t=1,2,...,n

(+) Lo cual resulta claro, ya que si estandarizamos todas las variables Xi,Yi del modelo lineal general (1), resulta que rij = $\sum x_i^* y_i^*$, donde $x_i^*; x_j^*$ y y_i^* representan la forma estandarizada de Xi,Xj y Yi. Nótese a propósito, que los elementos de las matrices definidas en (21) y (22) están en -- su forma estandarizada.

o equivalentemente, en términos del elemento residual:

$$E(\hat{u}_t x_j) = E(e_t x_j) = 0 \tag{26}$$

Ahora bien, si recalculamos (24) con Xi incluída, estando presente Xi, dado que rij = 1, entonces debemos tener también que:

$$E(u_t x_i) = 0. \tag{27}$$

Ya que como Xi es perfectamente colineal con Xj por hipótesis, entonces podemos expresarlo como:

$$x_i = \gamma x_j \quad ; \quad \gamma = \text{constante} \tag{27-bis}$$

Luego entonces:

$$E(u x_i) = E(u(\gamma x_j)) = \gamma E(u x_j) = 0 \qquad \text{q.e.d.} \tag{28}$$

Esto quiere que la adición de Xi, o sea, de una variable multicolineal a nuestro modelo de regresión o econométrico, no agrega nada a la correlación global del mismo y en consecuencia,no altera o afecta al estimador de la variancia residual.

<u>Multicolinealidad No-Perfecta</u>

Hasta ahora, hemos considerado el caso téorico de la multicolinealidad perfecta en la estimación paramétrica.Pero en la práctica, manejamos en realidad valores "altos" de rij, digamos en el intervalo:

$$0.5 < |r_{ij}| < 1 \qquad (29)$$

donde ya no tenemos estimadores indeterminados ni variancias enormes debido a que la inversa de la matríz (X'X) sí existe ahora.

Cuando la multicolinealidad es alta pero no perfecta entre las variables explicativas, existe una relación aproximadamente lineal entre las mismas, y aunque podemos obtener en este caso valores numéticos reales para los estimadores $\hat{\beta}_i$, el problma ahora es cómo interpretar estos valores y cómo responder cuando la multicolinealidad es un problema.

Primeramente, debemos percatarnos que frente a la multicolinealidad no-perfecta nos enfretamos a un problema de cómputo. Por ejemplo, el cómputo de $\hat{\beta}_i$, implica la inversión de la matríz de correlación R, definida como:

$$R = \begin{vmatrix} 1 & r_{12} & r_{13} \dots & r_{1k} \\ r_{21} & r_{22} & r_{23} \dots & r_{2k} \\ r_{31} & r_{32} & r_{33} \dots & r_{3k} \\ \cdot & \cdot & \cdot & \cdot \\ r_{k1} & r_{k2} & r_{k3} \dots & r_{kk} \end{vmatrix} \qquad (30)$$

A propósito de esta matríz conviene aclarar que esta versión es la sugerida por L.R.Klein 1/ y difiere ligeramente de la pro

1/ Véase, Klein, L. R. (1974). Opus cit. p.107

proporcionada por otros autores, como por ejemplo, de la uti-- lizada por J.Johnston,[1] que incluye además de los elementos de la matríz definida anteriormente, a una hilera y columna adicionales, conteniendo a los términos de la forma "riy", esto es, a los coeficientes de correlación simple entre Xi y Yi. Aunque para propósitos del presente análisis esta cuestión no tiene mayor relevancia, considramos oportuno hacer esta aclaración para el lector familiarizado con ambos autores, con el objeto de evitar posibles confusiones, en virtud de que ambos utilizan inclusive la misma notación para dicha matríz, aunque nosotros nos apegaremos a la sugerida por L.R.Klein, es decir, a la dada en (30).

Regresando a nuestra discusión, observamos que conforme aumenta la multicolinealidad pero no es perfecta, el determinante de esta matríz tiende a cero. En la práctica, se puede usualmente estimar una inversa para R, si:

$$|R| = \varepsilon = 0 \qquad (31)$$

donde ε (épsilon) es una constante muy pequeña arbritariamente fijada por el computador.

Para problemas de orden pequeño, digamos para K =2,3, hasta 7, normalmente es posible obtener suficientes dígitos de precisión

---

1/ Véase, Johnston,J. Econometric Methods. Mc Graw-Hill Kugakusha, Ltd. Tokyo, Second Edition,1972,p.132

para asegurar que R^{-1} sea calculada de manera correcta, numéricamentehablando. Sin embargo, si k es grande, por ejemplo de orden 8 en adelante, se pierde mucha precisión en las etapas sucesivas del proceso de cómputo, haciendo difícil calcular la inversa de R con la debida precisión. El caso es que aún las computadoras sumamente poderosas no la pueden calcular fácilmente cuando "épsilon" es suficientemente pequeña. 1/

A pesar de que existen muchos criterios para evaluar la exactitud numérica de R^{-1}, la prueba última de evaluación es verificar que se cumpla la conocida condición:

$$R\, R^{-1} = I \qquad (32)$$

donde I es la matríz idéntica.

La fórmula para la desviación estándard de los estimadores de los coeficientes de regresión (véase la expresión (20)) muestra cómo la multicolinealidad cuando es alta pero no completa, en el sentido de correlación unitaria, implica denominadores más y más pequeños en la expresión para la variancia de $\hat{\beta}_i$ (esto es, el determinante de la matríz $(X'X)$ tenderá cero), lo que a su vez genera una tendencia hacia variancias de $\hat{\beta}_i$ $(Var(\hat{\beta}_i))$ cada vez mayores, al menos que el ajuste global de la ecuación(R^2) sea suficientemente grande, que haga que la variancia del término residual ($\hat{\sigma}_e$) sea casi cero. 2/

1/ Véase, Klein, L.R. (1974). Opus cit.

2/ Ibid.

Cuando la intercorrelación entre las variables (rij's) es alta en relación a la correlación (global) múltiple (R^2) de la función estimada, la tendencia hacia cero del denominador de la expresión (20) es más "rápida" que la tendencia del numerador de la misma expresión hacia cero, y por lo tanto, las desviaciones estándard de los estimadores serán grandes. Implicando con ello, valores "t" insignificantes para los coeficientes (por ejemplo, menores que la unidad) 1/

Específicamente, cuando una ecuacion parezca tener un buen ajuste (R^2), digamos entre (R^2 =)0.7 y 0.9 , y las intercorrelaciones entre las variables explicativas sean altas y además algunas o la mayoría de los valores "t" de los coeficientes de regresión sean pequeños o insignificantes, tenemos una señal clara de la existencia de multicolinealidad. 2/

La proposición del penúltimo párrafo en el sentido de que la estimación por mínimos cuadrados ordinarios (+) puede continuar generando resultados confiables cuando el término de perturbación tiene una variancia pequeña, en presencia de multicolinealidad alta pero no perfecta, es compartida también por M.Dutta 3/.

1/ Donde "t" es el estadístico "t" de Student definido como $t = \hat{\beta}_i / \sqrt{\hat{var}(\hat{\beta}_i)}$
Se dice que el estimador $\hat{\beta}_i$ es "significativamente distinto de cero" si "t" calculada (con dicha expresión) resulta mayor que "t" tabulada, para un valor prefijado de significancia.
Véase, Kelejian, H.H. y W.E. Oates. Introduction to Econometrics. Harper and Row. Second Edition. 1981. p.92

2/ Ibid.

(+) Algunas veces nos referiremos a estos estimadores como "OLS" (Ordinary Least Squares).

3/ Véase, Dutta, M. Econometric Methods. South-Western Publishing Co. 1975, pp.147-156

A este respecto M. Dutta [1] propone el siguiente modelo:

$$y = \beta_1 x_1 + \beta_2 x_2 + u \qquad (33)$$

donde u es el término de perturbación, donde β_1 y $\beta_2 \neq 0$.

Supóngase ahora que X1 y X2 son altamente, pero no perfecta-- mente colineales. Entonces la variable X2, por ejemplo, se puede escribir como:

$$x_2 = \delta x_1 + v \qquad (34)$$

donde suponemos que v es un elemento residual muy pequeño pero finito

$$\sum v = 0 \ , \ \sum v x_1 = 0 \qquad (35)$$

Nótese que X1 y X2 en la expresión (34) aparecen en su forma "compacta". Si suponemos además que :

$$\sum x_1^2 = \sum x_2^2 = 1 \qquad (36)$$

entonces tendremos que:

$$\delta = r_{12} \qquad (37)$$

En especial, la matríz (X'X) puede escribirse como:

$$(X'X) = \begin{bmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 x_1 & \sum x_2^2 \end{bmatrix} = \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \qquad (38)$$

1/ Véase, Dutta, M. Opus cit.

Análogamente, la inversa de $(X'X)^{-1}$ se puede escribir como:

$$(X'X)^{-1} = \frac{1}{1-r_{12}^2}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \qquad (39)$$

Haremos ver que el estimador B1 es sesgado (+). Esto es:

$$E(\hat{\beta}1) = B1 \qquad (40)$$

El otro caso puede demostrarse en forma análoga. Como sabemos de la expresión (10), el estimador $\hat{\beta}1$ del modelo (34) se puede expresar como:

$$\hat{\beta}_1 = \frac{\sum x_1 y}{\sum x_1^2}$$

sustituyendo el valor de Y definido en (33) en la última expre-- sión, observamos que:

$$\hat{\beta}_1 = \frac{\sum x_1(\beta_1 x_1 + \beta_2 x_2 + u)}{\sum x_1^2}$$

$$= \frac{\beta_1 \sum x_1^2 + \beta_2 \sum x_1 x_2 + \sum x_1 u}{\sum x_1^2} \qquad (41)$$

Pero como $\sum x_1^2 = 1$, $\sum x_1 x_2 = \delta$ entonces $\hat{\beta}1$ se puede escribir como:

$$\hat{\beta}1 = \hat{\beta}1 + \hat{B}2\ r12 + \sum x_1 u \qquad (42)$$

Si aplicamos el operador E a la expresión anterior, observa-- mos inmediatamente que:

$$\beta 1 = \beta 1 + \beta 2\ r12 \qquad (43)$$

Pero como $B2\ r12 \neq 0$, entonces resulta evidente que la ex-- presión (40) se cumple, esto es:

$$E(\hat{\beta}1) \neq \beta 1$$

(+) Este es el efecto que se lograría al "eliminar" de la ecuación (33) la variable X2, por ejemplo. Cosa parecida ocurriría si eliminaramos la variable X1 de la misma ecuación. Volveremos a retomar este punto en el Apartado 3.3

Por otro lado, de la definición de la matríz (X'X), y de su inversa, definidas respectivamente en las expresiones (38) y (39), es fácil obtener la matríz de variancia-covariancia de $\hat{\beta}$ para el modelo (33):

$$Var(\hat{\beta}) = \sigma_u^2 (X'X)^{-1} = \frac{\sigma_u^2}{1-r_{12}^2}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \quad (44)$$

de donde se desprende que:

$$COV(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\delta\sigma_u^2}{1-\delta^2} = \frac{-r_{12}\sigma_u^2}{1-r_{12}^2} \quad (45)$$

De las expresiones (44) y (45) resulta evidente que conforme aumenta el coeficiente de corrélación simple, la variancia y covariancia de los estimadores tienden a infinito, en valor absoluto.(+)

Excepción hecha del caso en que el término de perturbación tiene una variancia pequeña, en general, podemos concluir, por consiguiente, que nuestros estimadores OLS tienden a ser menos confiables conforme aumenta el nivel o grado de multicolinealidad en el modelo. Aunque más adelante estudiaremos criterios precisos para reconocer estos niveles, un buen punto de partida que usualmente se recomienda, es analizar los elementos **no-diagonales** de la matríz de correlación definida en (30).[1/]

(+) Puesto que hemos visto que la variancia del término de perturbación no se vé afectado por la presencia de multicolinealidad, Y además, como σ_u^2 = constante mayor que cero, si r12 tiende a uno, entonces el término r12/(1-r12^2) de las expresiones (44) y (45) tiende a infinito, y entonces, el valor absoluto de Var($\hat{\beta}_1$) y Cov($\hat{\beta}_1,\hat{\beta}_1$) tienden también a infinito.

1/ Klein, L.R. (1974). Opus cit.

3.2 EFECTO DE LA MULTICOLINEALIDAD EN LA INFERENCIA ESTADISTICA

Utilizaremos un caso de estudio real en este apartado para mostrar la ambigüedad resultante al intentar identificar una variable explicativa importante de entre un conjunto de variables altamente dependientes. El contexto de este estudio se basa en una investigación original realizada en los E.E.U.U. por J.S. Coleman[1/], y también en forma independiente por F. Mosteller [2/], sobre igualdad de oportunidades de educación en escuelas públicas en dicho país.

La motivación de dicha investigación se fundamenta en el Acta de Derechos Civiles de 1964 aprobada por el Congreso Norteamericano, dando como resultado un estudio a nivel nacional enfocado a analizar la falta de igualdad en el acceso a oportunidades de educación para los norteamericanos, por razones de raza, color, religión u origen nacional, en instituciones de educación pública.[3/]

1/ Véase, Coleman, J.S. et al. Equality of Educational Opportunity. Washington, D.C. U.S. Government Printing Office, 1966.

2/ Véase, Mosteller, F. and D.P.Moynihan, Eds. On Equality of Educational Opportunity, Randon House, New York, 1972.

3/ Véase, Chatterjee, S. and B.Price. Regression Analysis by Example. John Wiley and Sons, New York, 1977, pp.144-151

Los datos para esta investigación fueron recabados en 1965 a partir de un muestroo de sección o corte transversal de diversos distritos escolares a través de toda la Unión Americana. Además de reportar estadísticos sumarios de variables tales como el nivel de aprovechamiento, facilidades en las escuelas, etc. **Se utilizó el análisis de regresión para establecer o determinar qué factores podrían considerarse como las más importantes causas del aprovechamiento escolar o académico.**

Los datos de la muestra, de un total de 70 escuelas seleccionadas al azar, incluían variables que medían, aprovechamiento académico, facilidades escolares y preparación académica del profesorado. El objetivo era determinar el impacto de estos insumos escolares en el aprovechamiento general o global del estudiante en las escuelas públicas a nivel nacional.

Supóngase que el índice desarrollado en este estudio para medir aquellos aspectos del medio escolar que se espera afecten el aprovechamiento, es aceptable. Este índice comprende evaluaciones de la planta física, material didáctico, programas especiales de capacitación y motivación docente, etc. La variable "aprovechamiento" se mide mediante un índice construido a partir de calificaciones estandarizadas.(+)

(+) Véase la definición de variable estandarizada dada en la expresión (14) del Apartado 3.1

Existen también otras variables que pueden afectar la relación entre los insumos escolares y el aprovechamiento. Este último puede verse afectado también por el ambiente familiar del alumno y la influencia de su grupo monitor en la escuela. Estas variables deben tomarse muy en cuenta en el análisis antes de evaluar el efecto de los insumos escolares sobre la variable-respuesta (aprovechamiento).

Supongamos que los índices construidos para cada una de las variables del modelo son suficientemente satisfactorios para nuestros propósitos. El modelo en cuestión queda definido como sigue: (+)

$$APROV = \beta o + \beta 1\ FAM + \beta 2\ MON + \beta 3\ ESC + U \qquad (46)$$

donde: (++)

APROV = Indice de aprovechamiento académico del alumno.

FAM = Indice de la influencia del ambiente familiar del alumno en su aprovechamiento académico.

MON = Indice de la influencia del tipo de grupo monitor al que pertenece el estudiante, en su aprovechamiento.

ESC = Indice del impacto de diversos insumos escolares como: planta física, material didáctico y diversas facilidades didácticas, capacitación y motivación docente, etc., en el aprovechamiento del alumno.

(+) Para diferenciarlo de otros modelos que discutieremos a lo largo de este trabajo, algunas veces nos referiremos a este modelo como el "modelo escolar".

(++) Las variables del modelo original han sido modificadas por el autor de esta investigación para facilitar la presente discusión. Sin embargo, para una discusión sucinta sobre el modelo original, véase, Chatterjee, S. y B. Price. Opus cit.

Los datos para una de las variables anteriores aparecen en la Tabla 3.2.1(de la próxima hoja). Todas aparecen en forma estandarizada.

Observemos de la expresión (46), que la contribución de la variable "ESC" puede medirse utilizando el estadístico "t" para el coeficiente β_3 .Recuérdese que este estadístico, en este -- caso, permite verificar si la variable ESC es necesaria o --- no en la expresión (46) toda vez que las variables "FAM" y "MON" ya se encuentran incluídas en la misma. Efectivamente, otra manera de reformular el modelo (46) sería considerar únicamente:

$$APROV = \beta_o + \beta_1 FAM + \beta_2 MON + U \qquad (47)$$

esto es, un modelo orientado a determinar el impacto de la variable FAM y MON sobre APROV, antes de incluir la variable ESC como en (46). Otra manera alternativa de "ajustar" el modelo original, sería considerar la ecuación:

$$\widetilde{APROV} = APROV - \beta_1 FAM - \beta_2 MON = \beta_o + \beta_3 ESC + U \qquad (48)$$

en donde notamos que el miembro izquierdo es un índice "ajustado" para la variable APROV ($\widetilde{APROV}$), donde se ha eliminado FAM y MON. La ecuación resultante es en consecuencia, un modelo de - regresión que mide el impacto de la variable ESC sobre el índice de aprovechamiento "ajustado" ($\widetilde{APROV}$).

T A B L A 3.2.1

DATOS SOBRE LAS VARIABLES EXPLICATIVAS Y EXPLICADA DEL MODELO DEFINIDO EN (46).(INDICES ESTANDARIZADOS).

Hilera	FAM	MON	ESC	APROV
* 1 *	0.60814	0.03509	0.16607	-0.43148
* 2 *	0.75169	0.47924	0.53356	0.79969
* 3 *	-0.92630	-0.61951	-0.78635	-0.92467
* 4 *	-1.25310	-1.21675	-1.04076	-2.19081
* 5 *	0.17399	-0.18517	0.14229	-2.84818
* 6 *	0.20256	0.12764	0.27311	-0.66233
* 7 *	0.24184	-0.09022	0.04967	2.63674
* 8 *	0.59421	0.21750	0.51576	2.35847
* 9 *	-0.61561	-0.48971	-0.63219	-0.91305
* 10 *	0.94391	0.62228	0.93368	0.59445
* 11 *	1.21721	1.00627	1.17381	1.21073
* 12 *	0.41426	0.71103	0.58978	1.87164
* 13 *	0.82782	0.74281	0.72154	-0.10178
* 14 *	-0.75512	-0.64411	-0.56986	-2.87949
* 15 *	-0.37407	-0.13757	-0.21770	3.92590
* 16 *	1.40353	1.14085	1.37147	4.35084
* 17 *	1.64194	1.29229	1.40269	1.57922
* 18 *	-0.21304	-0.07980	-0.21455	3.95699
* 19 *	1.28525	1.22441	1.20478	1.09275
* 20 *	-1.51928	-1.22565	-1.36598	-0.62389
* 21 *	-0.38224	-0.05153	-0.35560	-0.63654
* 22 *	-0.19186	-0.42605	-0.53718	-2.02659
* 23 *	1.27645	0.81427	0.91967	-1.46692
* 24 *	0.52310	0.30720	0.47231	3.15078
* 25 *	-1.59810	-1.01572	-1.48315	-2.18938
* 26 *	0.77914	0.87771	0.76496	1.91715
* 27 *	-1.04745	-0.77536	-0.91397	-2.71428
* 28 *	-1.63217	-1.47709	-1.71347	-6.59852
* 29 *	0.44325	0.60956	0.32833	0.65101
* 30 *	-0.24972	0.07875	-0.17216	-0.13772
* 31 *	-0.33480	-0.39314	-0.37198	-2.43959
* 32 *	-0.20690	-0.13936	0.05626	-3.27802
* 33 *	-1.99375	-1.69587	-1.97838	-2.48058
* 34 *	0.65473	0.79670	0.59865	1.88639
* 35 *	-0.27977	0.10817	-0.26450	5.06459
* 36 *	-0.43990	-0.66022	-0.58490	1.96335
* 37 *	-0.05334	-0.02396	-0.16795	0.26274
* 38 *	-2.06699	-1.31832	-1.72082	-2.94593
* 39 *	-1.02560	-1.15858	-1.19420	-1.38628
* 40 *	0.45847	0.21555	0.31347	-0.20797
* 41 *	0.93979	0.63454	0.69907	-1.07520
* 42 *	-0.93238	-0.95216	-1.02725	-1.66386
* 43 *	-0.35988	-0.30693	-0.46232	0.38117

(Esta tabla se continúa en la próxima hoja)

T A B L A 3.2.1

(C o n t i n u a c i ó n)

DATOS SOBRE LAS VARIABLES EXPLICATIVAS Y EXPLICADA DEL MODELO DEFINIDO EN (46).(INDICES ESTANDARIZADOS).

Hilera	FAM	MON	ESC	APROV
* 44 *	-0.00518	0.35985	0.02485	1.37447
* 45 *	-0.18892	-0.07959	0.01704	-2.82457
* 46 *	0.97211	0.47644	0.57036	3.56363
* 47 *	-2.06993	-1.82915	-2.16738	-2.64141
* 48 *	0.32143	-0.25961	0.21632	0.05387
* 49 *	-1.42382	-0.77626	-1.02473	0.50763
* 50 *	-0.07852	-0.21347	-0.11750	0.64347
* 51 *	-0.14925	-0.03192	-0.36598	2.49414
* 52 *	0.32665	0.79149	0.71369	0.61955
* 53 *	-1.49102	-1.02073	-1.38103	0.61745
* 54 *	-0.94757	-1.28991	-1.24791	-1.00743
* 55 *	0.24550	0.83794	0.59596	-0.37469
* 56 *	-0.41630	-0.60312	-0.34951	-2.52824
* 57 *	1.38143	1.54542	1.59429	0.02372
* 58 *	1.03806	0.91637	0.97622	2.51077
* 59 *	-0.88039	-0.47532	-0.77693	-4.22716
* 60 *	1.08555	0.65700	0.29401	1.96847
* 61 *	-1.95142	-1.94199	-1.89645	1.25669
* 62 *	2.83384	2.47398	2.79222	-0.16843
* 63 *	1.86753	1.55229	1.89057	-0.54158
* 64 *	-1.11172	-0.69732	-0.80197	-2.23973
* 65 *	1.41958	1.11491	1.24558	3.62654
* 66 *	0.53940	0.16152	0.33477	0.97032
* 67 *	0.22491	0.74800	0.66182	-3.16093
* 68 *	1.48244	1.47079	1.54285	-1.90801
* 69 *	2.05425	1.80369	1.96066	0.64598
* 70 *	1.24058	0.64484	0.87372	-1.75915

Fuente: Los datos que aparecen aquí no son los auténticos, sino que fueron especialmente generados por Chatterjee,S. y B.Price,opus cit.para efectos de este ejemplo.

Si utilizamos la información de la Tabla 3.2.1 se puede estimar el modelo (46) obteniéndose los resultados sumarios que aparecen en la Tabla 3.2.2 (que aparece en la próxima hoja), así como una gráfica (Figura 3.2.1, que se muestra enseguida de la Tabla 3.2.2) que muestra los residuales estandarizados (e_i^*) contra los valores estimados de APROV.

Primeramente, observemos de la Figura 3.2.1, dado el comportamiento de los residuales, que no existen huellas de mala especificación funcional en el modelo estimado. La observación localizada en la parte inferior izquierda de la gráfica tiene un valor residual de aproximadamente 2.5 desviaciones estándard de la media cero y debería de observarse con más cuidado. Sin embargo, si se elimina de la muestra, los resultados de la regresión prácticamente no manifiestan cambio alguno. Por consiguiente, la observación se retiene en el análisis.(+)

Por otra parte, los resultados de la Tabla 3.2.2 muestran que alrededor del 20% de la variación total en el índice de aprovechamiento (esto es, en la variable-respuesta) queda explica ---

---

(+) Se omiten los detalles de este procedimiento para simplificar la presente exposición. A este respecto, conviene señalar que se ha seguido la metodología propuesta por Chatterjee, S. y B.Price. Opus cit. pp.19-50 quienes sugieren una serie de criterios precisos para reconocer la presencia de problemas de especificación en un modelo de regresión o econométrico, a partir del análisis de los elementos residuales del modelo estimado, así como la manera de "ajustar" el modelo en cuestión, a fin de eliminar dicho problema.(Aunque conviene aclarar que el transfondo de dichos procedimientos, es hacer que se cumplan las primeras dos condiciones del modelo lineal general).

T A B L A 3.2.2

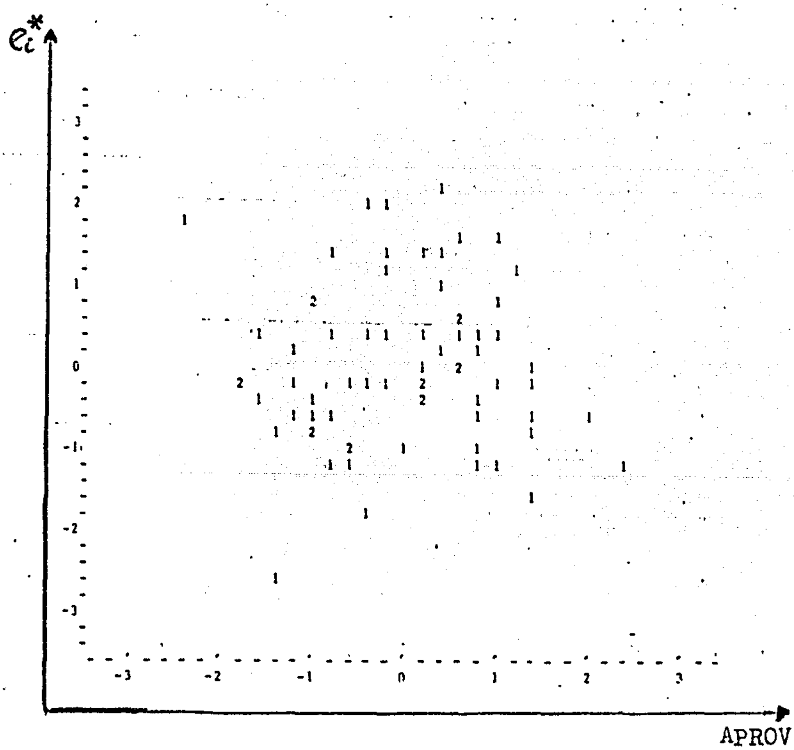
RESULTADOS SUMARIOS DE LA REGRESION DEL MODELO (46)

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDARD	VALORES " t "
. FAM	1.101	1.411	0.780
. MON	2.322	1.481	1.568
. ESC.	- 2.281	2.220	-1.027
CONSTANTE	- 0.070	0.251	0.279
N= 70	R^2= 0.206,	S= 2.070	

Fuente: Obtenidos a partir de los datos de la Tabla 3.2.1

Figura 3.2.1

GRAFICA DE LOS RESIDUALES ESTANDARIZADOS (e_i^*) CONTRA LOS VALORES ESTIMADOS DE LA VARIABLE DEPENDIENTE (APROV), PARA EL MODELO (46).



Fuente: Chatterjee,S. y B.Price.Opus cit.p.148

da conjuntamente, por las tres variables explicativas incluídas en la expresión (46). El valor de F calculada es de 5.72 con — (3.66) grados de libertad, que resulta ser significativa al 1%.

Por tanto, aún cuando la variación total explicada del modelo se estima en sólo un 20%, se acepta que las variables explicativas correspondientes son válidas. Aunque los valores individuales de las "t" sean todos muy pequeños. Es decir, los estadísticos sumarios indican que aunque las tres variables del modelo (46) tomadas conjuntamente, son muy importantes, a juzgar por los valores "t" individuales de las mismas, se desprende que alguna de dichas variables puede eliminarse del modelo, con tal de mantener a las — otras dos restantes.

Estos resultados muestran una situación típica en donde existe un nivel peligroso de multicolinealidad. Las variables explicativas están tan altamente correlacionadas entre sí que cualquiera de ellas puede servir como una aproximación de las otras dos, sin afectar el poder explicativo global del modelo. Los valores pequeños de "t" confirman que cualquiera de las variables explicativas pueden eliminarse de la ecuación (46).

Por consiguiente, el análisis de regresión no nos permite obtener ninguna información para evaluar la importancia de los insumos escolares en el aprovechamiento académico o escolar. La causa de todo esto, es claramente la multicolinealidad. Las correlaciones simples entre las variables explicativas FAM y MON; FAM y ESC; y MON y ESC de .960, .986 y .982 respectivamente, así lo confirman.

La multicolinealidad en este modelo era de esperarse. Es la naturaleza de estas variables que hace que cada una de ellas esté determinada por, y contribuya a determinar a las otras variables. No es ilógico pensar, por consiguiente, que no se tratan de tres, sino de una sola variable.

Desafortunadamente la conclusión anterior no nos ayuda a contestar la pregunta inicial sobre el efecto de los insumos escolares (variable ESC) sobre el aprovechamiento académico (variable- respuesta APROV). Existen dos posibilidades a este respecto:

Primero, que la multicolinealidad se presente por deficiencia de los datos, en cuyo caso puede subsanarse recabando información adicional.

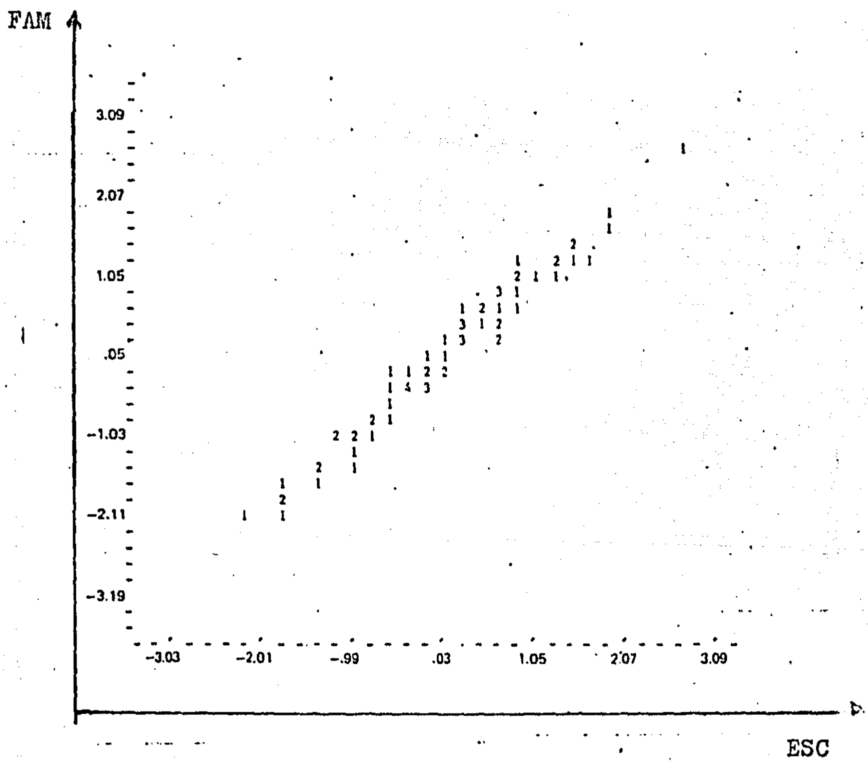
Segundo, que la multicolinealidad se presente debido a que las interrelaciones entre las variables en cuestión sean una característica inherente al fenómeno bajo estudio. Ambas situaciones se discuten brevemnnte a continuación.

En el primer caso, la muestra debió ser escogida de tal suerte de asegurar que la correlación entre las variables explicativas no fuesen tan grande.Por ejemplo, considérense las variables --FAM y ESC. Si graficamos los valores de ámbas variables en un --eje cartesiano (con FAM en el eje vertical y ESC en el eje hori--zontal) obtendremos una gráfica, como la que aparece en la Figu--ra 3.2.2 (que aparece en la próxima hoja), donde notamos que --todas las observaciones se encuentran diseminadas alrededor de una línea recta que cruza los valores medios de las variables: FAM y ESC, esto es, el punto (0,0) del eje cartesiano.También se observa que no existen datos sobre la variable ESC en esta muestra con valores en la parte superior izquierda e inferior derecha del --plano de la figura. Y por lo mismo, no existe información en la muestra sobre la variable APROV cuando el valor FAM es alto y el de ESC es bajo, o cuando FAM es bajo y ESC es alto.

Pero es sólo con datos recabados bajo estas dos condiciones que los efectos de FAM y ESC sobre APROV podrían, en un momento da--do, determinarse.Por ejemplo, supóngase que existiese alguna ---observación en la parte superior izquierda del cuadrante de la Figura 3.2.2. Entonces sería factible, cuando menos, comparar el valor promedio de la variable APROV con valores altos y ba---jos de las variable ESC cuando la variable FAM se mantiene --constante.

E I G U R A 3.2.2

GRAFICA DE LA VARIABLE "FAM" CONTRA "ESC" PARA EL MODELO (46).



Fuente: Chatterjee,S. y B.Price.Opus cit.p.149

Al agregar la tercera variable explicativa, MON, al modelo, implica la existencia de 8 combinaciones distintas posibles de datos que debían ser incluidos en la muestra. Si denotamos con el símbolo "+" a un valor arriba del promedio, y con "-" a un valor abajo del promedio, tendremos las ocho posibilidades que se muestran en la Tabla 3.2.3 (que aparece en la siguiente hoja)

Al analizar la Tabla 3.2.3, nos percatamos que las grandes correlaciones encontradas en el análisis, sugieren que sólo las combinaciones 1 y 8 están representadas en los datos. Si la muestra bajo discusión resultáse de esta manera por azar o casualidad, la receta para eliminar la multicolinealidad sería entonces recabar información adicional sobre las otras combinaciones de datos no consideradas en la muestra original. Por ejemplo, los datos basados en las combinaciones 1 y 2 únicamente podrían usarse para evaluar el efecto sobre APROV, manteniendo a FAM y MON como "constantes", ambas variables tomadas con valores arriba del promedio. Si estas fueran las únicas combinaciones representadas en los datos, el análisis consistiría en una regresión simple de APROV contra ESC. Y los resultados arrojarían una respuesta parcial, es decir, una evaluación del aprovechamiento escolar cuando las variables FAM y MON se consideran arriba del promedio.

Sin embargo, la receta de recolección de información adicional para resolver la multicolinealidad no debe considerarse como —

una "panacea". Con frecuencia , es prácticamente imposible recabar más datos debido fundamentalmente a problemas presupuestarios, de tiempo y/o de personal. Es siempre mejor estar "prevenidos" de antemano de posibles deficiencias en los datos. En consecuencia, siempre que se pueda, la información deberá recabarse de acuerdo a un diseño. Desafortunadamente, el diseño "a priori" no es siempre factible. En determinadas investigaciones económicas o estudios de observación social como el presente, los valores de las variables explicativas usualmente no se conocen hasta que el dato se elige de la muestra, después de un proceso de medición sumamente lrgo y costoso. Con dicho procedimiento, resulta muy difícil la obtención de una muestra "balanceada" como la que se recomienda.

La segunda razón de la existencia o aparición de la multicolinealidad en el modelo puede radicar en que la relación entre las variables explicativas sea una característica inherente o intrínseca al proceso o fenómeno muestreado. De tal forma que si las variables FAM, MCN y ESC existen en la población sólo como combinaciones de datos de la forma 1 y 8 de la Tabla 3.2.3, por ejemplo, entonces sería imposible pretender estimar los efectos individuales de dichas variables sobre la variable-respuesta APROV.
El único recurso viable para continuar con el análisis de estos efectos, en este caso, sería buscar las "causas subyacentes" que pudiesen explicar las interrelaciones existentes entre las variables explicativas. Mediante este proceso, se podrían descubrir

inclusive, otras variables explicativas que pudiesen afectar más determinantemente a la variable-respuesta en consideración, y reformular incluso, el modelo en cuestión.

3.3. EL EFECTO DE LA MULTICOLINEALIDAD EN LA ESPECIFICACION FUNCIONAL.

Según D.E. Farrar y R.R. Glauber 1/ "...La correcta especificación normalmente es más importante para la construcción correcta de modelos que la elección de una técnica correcta de estimación".

Sostiene que la especificación comienza ordinariamente en la mente del constructor del modelo. A través de una combinación de teoría, información anterior e intuición, el investigador escoge las variables explicativas que supuestamente le permitirán explicar el comportamiento de la variable dependiente o -- variable- respuesta. Y por lo general, el trabajo no termina con la primera especificación tentativa. Antes de que una e-- cuación se considere como "aceptable" debe ser sometida a prue bas a través de un conjunto de datos empíricos. Si ésta es deficiente de alguna manera, se modifica y se vuelve a probar otra vez. 2/

1/ Véase, Farrar, D.E. y Glauber, R.R. Multicollinearity in Re-- gression Analysis: The Problem Revisited. Review of Econo-- mics and Statistics, Vol.49, 1967, pp.92-107

2/ Ibid.

También, afirman que independientemente de la naturaleza de la dependencia real entre Y y cada variable de un conjunto relativamente grande de variables independientes contenidas en la matríz X, el aumento de la interdependencia dentro de X al aumentar su tamaño, hace que disminuya rápidamente la estabilidad y por ende, la significación de la contribución de cada variable independiente en la variación total del modelo. 1/

Por otra parte, de acuerdo a T.C.Liu 2/ ."...Las limitaciones de información y no las teóricas son las principales responsables de la tendencia persistente a subespecificar un modelo de regresión o econométrico".

Farrar y Glauber concluyen que "...independientemente de la situación particular, la esencia del problema de multicolinealidad es la cantidad de información requerida para una construcción modelística satisfactoria contra la contenida en una muestra de datos disponible. Y si al modelo se le pretende presentar en toda su complejidad, la solución del problema de multicolinealidad requerirá aumentar la muestra disponible para introducir información adicional. 3/

Vamos a hacer ver a continuación, cómo la multicolinealidad afecta a un modelo subespecificado y cuál es su consecuencia en la estimación paramétrica. A este respecto, consideramos que es —

1/ Véase, Farrar y Glauber. Opus cit.

2/ Véase, Liu, T.C. Underidentification, Structural Estimation and Forecasting. Econometrica, 28, October, 1960, p.856

3/ Véase, Farrar y Glauber. Opus cit.

oportuno señalar que es muy frecuente en la especificación modelística, tener que reducir el número de variables explicativas necesarias a un número mucho menor por insuficiencia de información. Esto imposibilita entre otras cosas, la posible eliminación de una variable multicolineal en el modelo, expresándola como combinación lineal de otras que no lo son, y asî eliminar la colinealidad en el modelo afectado.

Para comenzar, conviene aclarar que la exclusión arbitraria de una variable altamente colineal de un modelo de regresión, como se acostumbra muy a menudo, como una forma "rápida" de deshacerse de la multicolinealidad, contribuye a la incorrecta especificación de la estructura global del mismo, creando al mismo tiempo, sesgo en los estimadores paramétricos, como observaremos enseguida.

Usaremos el enfoque de M.Dutta 1/ para demostrar que dado el modelo:

$$y = \beta_1 x_1 + \beta_2 x_2 + u \qquad (49)$$

donde X1 y X2 están altamente correlacionadas entre sí. Si alguna de ellas se elimina del mismo, el estimador del parámetro que quede en él, resultará sesgado. **Como se muestra enseguida (Nótese de paso que el modelo anterior se encuentra en su forma compacta).**

Supóngase pues que para "corregir" el problema de la multico-

1/ Véase, Dutta,M. Econometric Methods. Opus cit.

linealidad, eliminamos la variable X2. Nuestro modelo se transformaría entonces en :

$$y = \beta_1 x_1 + v \tag{50}$$

donde v es la nueva variable aleatoria. Nuestra estimación de mínimos cuadrados aplicado al modelo anterior, nos dice que:

$$\hat{\beta}_1 = \frac{\sum y x_1}{\sum x_1^2} \tag{51}$$

pero como la variable Y de la expresión (51) en realidad tiene la forma definida en (49), entonces el estimador $\hat{\beta}_1$ se puede escribir también como:

$$\hat{\beta}_1 = \frac{\sum x_1 (\beta_1 x_1 + \beta_2 x_2 + u)}{\sum x_1^2} \tag{52}$$

Aplicando el operador E a la expresión anterior:

$$\begin{aligned} E(\hat{\beta}_1) &= E\left[\frac{\sum x_1 (\beta_1 x_1 + \beta_2 x_2 + u)}{\sum x_1^2}\right] \\ &= \frac{1}{\sum x_1^2} E\left[\sum x_1 (\beta_1 x_1 + \beta_2 x_2 + u)\right] \\ &= \beta_1 + \beta_2 \cdot \frac{\sum x_1 x_2}{\sum x_1^2} = \beta_1 + \beta_2 \cdot \delta \end{aligned} \tag{53}$$

donde: $\delta \equiv r_{12} \neq 0$

Es decir:

$$E(\hat{\beta}1) = \beta 1 + \beta 2 \tag{54}$$

Y por lo tanto :

$$E(\hat{\beta}1) \neq \beta 1 \text{ , q.e.d.} \tag{55}$$

(+) Lo cual concuerda perfectamente con el resultado obtenido en el Apartado 3.2 (página 27).

Según M. Dutta [1] si la multicolinealidad entre las variables independientes es severa en un modelo especificado correctamente, la relación puede resultar en canjear cierto nivel de sesgo en el estimador por una variancia menor del mismo, en virtud de — que la variancia del estimador que queda en el modelo en presencia de la variable multicolineal es mayor que cuando se calcula en ausencia de ella. O sea, si denotamos a la variancia del — estimador $\hat{\beta}i$ como;

(Antes) Var $(\hat{B}i^{**})$; (Con multicolinealidad

(Después) Var $(\hat{B}i^{*})$; (Sin multicolinealidad

entonces lo anterior implica que;

$$\text{Var}\,(\hat{\beta}i^{**}) > \text{Var}\,(\beta i^{*}) \tag{56}$$

Lo cual es fácil de demostrar para el caso que nos ocupa, ya que la variancia de $\hat{\beta}1$ en el modelo (49), que denotaremos como Var($\hat{\beta}1$), (esta última calculada en presencia de la variable multicolineal X2) se puede demostrar fácilmente que es igual a;

$$Var(\hat{\beta}_1^{*}) = \frac{\sigma_u^2 \, \Sigma x_2^2}{\Sigma x_1^2 \, \Sigma x_2^2 - (\Sigma x_1 x_2)^2} \tag{56-bis}$$

y como por otro lado, sabemos que la variancia de $\hat{\beta}1$ para el modelo (50), que podemos denotar como Var($\hat{\beta}1^{**}$), (calculada aho

1/ Véase, Dutta, M. Econometric Methods. Opus cit.

ra después de eliminar del modelo a la variable multicolineal X2), se puede escribir como:

$$\mathrm{Var}\,(\hat{\beta}_1^{**}) = \frac{\sigma_u^2}{\sum x_1^2} \qquad (57)$$

Ahora bien, a partir de (56-bis) y (57) resulta evidente que:

$$\frac{1}{\mathrm{Var}\,(\hat{\beta}_1^{**})} + \frac{\left(\sum x_1 x_2\right)^2}{\sigma_u^2 \sum x_2^2} = \frac{1}{\mathrm{Var}\,(\hat{\beta}_1^{*})} \qquad (58)$$

de donde se infiere que:

$$\frac{1}{\mathrm{Var}\,(\hat{\beta}_1^{**})} < \frac{1}{\mathrm{Var}\,(\hat{\beta}_1^{*})} \qquad (59)$$

puesto que el segundo término del miembro izquierdo de la igualdad (58) es siempre positivo. Y por lo tanto, se sigue que la desigualdad (56) es cierta. O sea:

$$\mathrm{Var}\,(\hat{\beta}_1^{**}) > \mathrm{Var}\,(\hat{\beta}_1^{*})$$

En otras palabras, la variancia del estimador que queda en el modelo después de eliminar la variable colineal, es menor que la variancia que tenía el mismo estimador antes, o sea, en presencia de dicha variable colineal.

Para concluir consideramos conveniente señalar ------ el hecho de que en muchas ocasiones en economía, por razones de índole teórica, puede suceder que se postule o especifique un modelo como el que sigue:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 \Delta X_t + U_t \qquad (60)$$

en donde definimos cada variable del modelo anterior como:

Yt = Consumo agregado en el año corriente (t)

Xt = Ingreso disponible en el año corriente (t)

Xt-1 = Ingreso disponible en el año anterior (t-1)

ΔXt = Xt - Xt-1 = Diferencia entre el ingreso del año corriente y el ingreso del año anterior (i.e.esta variable representa el "efecto de las expectativas"de la sociedad derivadas de cambios recientes en su nivel de ingresos)

Notamos de entrada, que en el modelo mencionado, existe multicolinealidad perfecta, ya que las variables Xt y Xt son linealmente dependientes. Lo cual hace imposible, teóricamente, estimar por separado los valores de los parámetros de dicho modelo, como hemos destacado anteriormente.

No obstante, en la prádtica, y con el auxilio de la computadora, es posible usualmente,estimar el modelo (60) pero con la desventaja de obtener estimadores muy poco confiables y variables no-significativas.

Para resolver este tipo de problemas hay necesidad de reformular o reespecificar completamente el modelo original en términos de las variables originales a fin de romper el "candado" de multicolinealidad como el que afecta al modelo (60), de la manera como discutiremos en forma pormenorizada en el Capítulo V al abordar el análisis de las técnicas y métodos disponibles hoy en día para la corrección de la multicolinealidad.

Sin embargo, y a propósito del modelo (60), conviene observar asimismo, que incluye como una de sus variables explicativas, a una variable con retraso (o "retrasada"). Esto es muy común en economía. Esto es, un modelo econométrico puede incluir aparte de determinadas variables colineales, una serie o subconjunto de variables retrasadas. A propósito, un modelo econométrico con retraso distribuido de orden "K" sobre la variable explicativa "Xt" se define como:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{k+1} X_{t-k} + u_t \qquad \text{(60-bis)}$$

cuyo tratamiento requiere del uso de técnicas especiales. Existen para el efecto dos ampliamente conocidas y utilizadas en econometría, cada una con ventajas y limitaciones particulares. Estas son la de L.M.Koyck 1/ y la de S.Almon 2/.

1/ Véase, Koyck, L.M. Distriguted Lags and Investment Analysis. North Holland: Amsterdam, 1954.

2/ Véase, Almon, Shirley. The Distributed Lag Between Capital Appropiations and Expenditures. Econometrica, 33, January 1965, pp.178-196

La técnica de Koyck en particular tiene la ventaja de minimizar la pérdida de información ocasionada por la magnitud del retraso (pues su aplicación requiere nadamás de un periodo de retraso de la variables - explicativa objeto) aunque tiene la limitante o desventaja de generar autocorrelación en el modelo reformulado o corregido.

Por su parte, la técnica de Almon a diferencia de la de Koyck, requiere de más de un periodo de retraso de la variable -explicativa objeto para la estimación del modelo corregido (esto es, requiere usualmente de mayor información que la de Koyck) pero tiene la ventaja de no violar ninguno de los supuestos del modelo lineal general, y por lo mismo, permite la estimación de estimadores confiables y estables para la inferencia estadística y predicción, por ejemplo.

En consecuencia, - y adelantándonos un poco al proceso de corrección de la multicolinealidad - ante el tipo de situación anterior, se recomienda primeramente aplicar a cada una de las variables retrasadas incluidas en el modelo en cuestión, de ser posible, la técnica de Almon (por las razones expuestas) y posteriormente, alguno de los métodos para corregir la multicolinealidad que serán discutidos en el Capítulo V, y de esta manera, reformular o reespecificar en definitiva, el modelo en cuestión.

3.4 EFECTO DE LA MULTICOLINEALIDAD EN LA PREDICCION

J.R. Meyer y E.Kuh [1] sostienen que "...Aquellos dedicados a realizar pronósticos no deben preocuparse por la presencia de la multicolinealidad."

Esto es comprensible, ya que aunque el intervalo de predicción tiende a ser mayor cuando existe multicolinealidad, el pronóstico en sí prácticamente es el mismo.

Lo anterior se desprende por el hecho de que el valor proyectado (o pronostico) $\hat{y}_o$ correspondiente o asociado a un valor Xo del modelo lineal general se puede escribir como:

$$\hat{y}_o = X_o' \hat{\beta} \tag{61}$$

donde Xo es el vector correspondiente de observación definido como:

$$X_o' = (X10,\ X20, \ldots, Xk0) \tag{62}$$

Nótese que el valor proyectado definido en (61) ha sido obtenido a partir del modelo general en su forma compacta y cuya expresión matricial es la conocida ecuación:

$$Y = X\beta + u \tag{3}$$

Por otra parte, sabemos del análisis de regresión por mínimos cuadrados, que la varianza de $\hat{y}_o$ se puede expresar como:

$$Var(\hat{y}_o) = [X_o'(X'X)^{-1}X_o + 1]\,\sigma_u^2 \tag{63}$$

---

[1] Véase, Meyer, J.R. y Kuh, E. <u>How Extraneous are Extraneous Estimates</u> ?. Review of Economics and Statistics, Vol.39, November, 1957, pp.380-393.

y la banda del error o "banda de confianza" para $\hat{y}_0$ se puede definir como:

$$\pm\ t(n-k-1,\ \alpha/2)\sqrt{Var(\hat{y}_0)} \qquad (64)$$

donde α es el nivel de significación previamente fijado.

Entonces, conforme la multicolinealidad aumenta (pero no es perfecta como sucede en la vida real, o en la mayoría de los casos) la matríz $(X'X)$ tiende a hacerse singular y los elementos de $(X'X)^{-1}$ tenderán a hacerse infinitos como ya hemos señalado anteriormente y entonces la banda del error tenderá a ser mayor en presencia de multicolinealidad que en ausencia de ella, toda vez que el único factor de la variancia de $\hat{y}_0$ que crece con la multicolinealidad es precisamente la matríz $(X'X)^{-1}$, ya que el término $\hat{\sigma}_u^2$ (esto es, la variancia del término de perturbación) permanece constante, como ya hemos destacado anteriormente. De la misma manera, notamos que el valor de pronóstico $\hat{y}_0$ no se vé afectado, ya que como se observa de la expresión (61) sólo depende de $\hat{\beta}$ y del vector Xo, o sea, tiene la forma:

$$\hat{y}_0 = \hat{\beta}_1 X_1^0 + \hat{\beta}_2 X_2^0 + \cdots + \hat{\beta}_k X_k^0 \qquad (65)$$

cuyo valor no es de esperarse se vea afectado aún existiendo alto nivel de multicolinealidad en el modelo, con tal que la relación estructural entre las variables explicativas X1,X2,..., Xk, se mantenga inalterable.(+)

(+) Podemos agregar de paso, que todo lo dicho sobre la banda del error respecto al pronóstico ($\hat{y}_0$) también se cumple para el valor medio esperado definido como $\hat{\mu}_0 = \hat{y}_0$, y cuya variancia se puede escribir $Var(\hat{\mu}_0) = \hat{\sigma}_u^2 X_0 (X'X)^{-1} X_0$.

Si aplicamos estos resultados al caso de un modelo de una variable explicativa (o dos variables en total), podemos visualizar esta situación mejor mediante una gráfica como la que se muestra en la hoja que sigue (Figura 3.4.1).

Resulta entonces claro que es probable reducir sustancialmente el intervalo de predicción definido como:

$$\hat{y}_0 \pm t_{(n-k-1,\ \alpha/2)} \sqrt{Var(\hat{y}_0)} \qquad (66)$$

mediante la eliminación de un cierto número de variables colineales del modelo afectado, pues de esta manera, reducimos la magnitud de la variancia de $\hat{y}_0$. Sin embargo, como ya hemos discutido en el apartado anterior, esta "receta" en particular para eliminar la multicolinealidad, genera a su vez, sesgo y subespecificación en nuestro modelo.

Pero como B.R. Chiswick y S.T. Chiswick [1] afirman , el problema de la multicolinealidad es indeseable si el objetivo es la obtención de estimadores confiables de las variables explicativas que lo conforman.(Lo cual no hace más que corroborar lo pernicioso del efecto de la multicolinealidad en la inferencia estadística discutido en el Apartado 3.2).

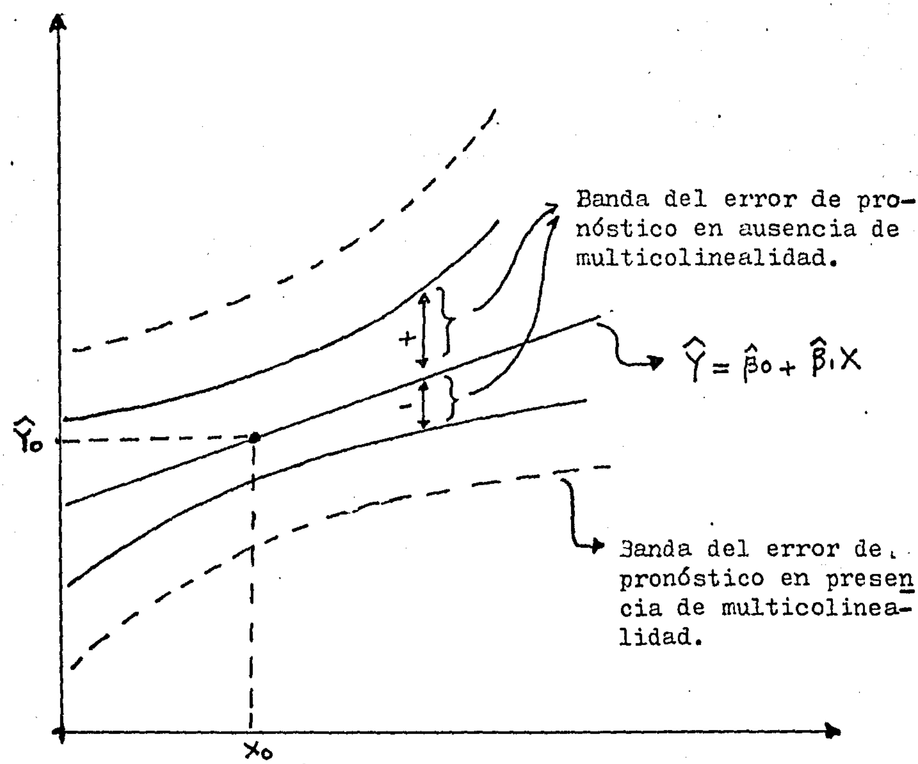
Por consiguiente, la multicolinealidad no representa realmente un problema, si el objetivo del análisis es la predicción. Y si

---

1/ Chiswick,B.R. y Chiswick,S.J. Statistics and Econometrics:A Problem Solving Text. University Park Press,1975,pp.180-191

F I G U R A 3.4.1

EFECTO DE LA MULTICOLINEALIDAD SOBRE LA BANDA DEL ERROR DE PRONOSTICO.



Fuente: Elaboración del autor de esta investigación.

se puede suponer además que la misma estructura altamente multicolineal que existe en la población de la que extrajo la muestra para efectuar la regresión (observaciones 1,2,3,....,n) se mantiene o existe también en la población de la que se extrae o selecciona la muestra para la predicción (observación n+1,n+2,.. ...) 1/.

Esto implica, en otras palabras, que dado un modelo econométrico de la forma;

$$Y = X\beta + u \qquad (3)$$

estimado a partir de una muestra de "n" observaciones y en donde existen al menos dos variables colineales Xi y Xj, entonces nuestra predicción tendrá sentido si la relación funcional existente entre ambas permanece inalterable para la observación n+1, n+2, etc. O sea, si en particular la estructura multicolineal existente en la muestra, sigue prevaleciendo en la población.

Meyer y Kuh 2/ confirman lo anterior argumentando que en tanto que la eliminación de variables explicativas colineales tiende a reducir el intervalo de predicción, como acabamos de observar, el pronóstico real, por hipótesis, cambia muy poco, como también se ha destacado. Y además, agregan que aunque el pronosticador pragmático es indiferente al grado de multicolinealidad, y el más sofisticado no lo es, ambos obtienen pronósticos similares y los errores reales de pronóstico son aproximadamente iguales.

1/ Véase, Chiswick y Chiswick. Opus cit.

2/ Véase, Meyer y Kuh. Opus cit.

Meyer y Kuh [1] terminan sugiriendo el uso combinado de datos de corte transversal y de series de tiempo y la utilización de "estomadores "ajenos"(o "exógenos") [2] en el proceso de estimación para eliminar el problema de la multicolinealidad. [3]

Para clarificar los conceptos antes expuestos, consideramos oportuno y conveniente incluir en este apartado, un ejemplo -----

---

1/ Véase, Meyer y Kuh.Opus cit.

2/ Se refiere a ciertos tipos de estimadores paramétricos calculados a partir de datos de corte transversal que luego son -- "reutilizados" en forma "exógena" para la estimación de otros parámetros de un determinado modelo-objeto, usando para este último propósito, datos de series temporales, con el objeto de eliminar o "romper" el candado de la multicolinealidad en ciertos modelos econométricos.Aunque hay que reconocer -- que algunos autores señalan diversos peligros e incovenientes sobre su uso generalizado en econometría para atacar -- dicho problema.Sobre el particular, además de la referencia anotada en el pie de página anterior,

Véase, Kuh,E. The Validity of Cross-Sectionally Estimated Behavior Equations in Time-Series Applications. Econometrica, Vol.27, No.2, April 1959,pp.197-214

3/ Para una discusión pormenorizada sobre la manera de aplicar la técnica de combinación de datos de corte transversal y de series temporales ("Pooling Technique") para efectuar -- la estimación de un modelo en presencia de multicolineali-- dad se sugieren al respecto, las siguientes referencias; --

Véase, Tobin,J.A. A Statistical Demand Function for Food in the U.S.A. Royal Statistical Society, Vol.113-A,1959,pp.113-141

Véase, Wold, H.O.A. y Lars Juréen.Demand Analysis; A Study in Econometrics.New York; John Wiley and Sons, 1953,pp.192-195

Véase, Stone, J.R.N. The Measurement of Consumers' Expenditure and Behavior in the United Kingdom,1920-1938.Cambridge, England; Cambridge University Press,1954.

Véase, Balestra,P. y M.Nerlove. Pooling Cross-Section and -- Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas.Econometrica, Vol.34,No.3,July 1966, pp.585-612 concreto del efecto de la multicolinealidad en la predicción. Este nos permitirá observar, entre otras cosas, que el valor del pronóstico ($\hat{y}_0$) no cambia prácticamente a pesar de la existencia de fuerte colinealidad entre las variables explicativas. Además, este caso ilustrativo resulta de particular interés en este trabajo, ya que nos servirá de referencia más adelante al tratar los temas de detección y corrección de la multicolinealidad.

Hemos seleccionado un caso de estudio basado en información macroeconómica referente a importaciones de la economía francesa. Esta información fue originalmente analizada por E.Malinvaud [1/]. La presente discusión se basa en dicho análisis. Las variables involucradas son: Las importaciones totales (IMPORT), la producción doméstica (DOPROD), la formación de capital (STOCK), así como el consumo doméstico(CONSUM), todas ellas expresadas en miles de millones de francos franceses, a precios corrientes, durante el período 1949-1966. Los datos se reproducen en la Tabla 3.4.1(que aparece en la próxima hoja). El modelo en cuestión queda formulado en los siguientes términos:

$$IMPORT = \beta_0 + \beta_1 \cdot DOPROD + \beta_2 \cdot STOCK + \beta_3 \cdot CONSUM + u \quad (67)$$

Los resultados de la regresión del modelo anterior aparecen en la Tabla 3.4.2 (enseguida de la Tabla 3.4.1).La gráfica de residuales contra el tiempo, que aparece en la Figura 3.4.2 (después de la Tabla 3.4.2) muestra un patrón claro que sugie--

1/ Véase, Malinvaud,E. Statistical Methods of Econometrics.North Holland Publishing Co.,1970,pp.80-85,pp.204-205

T A B L A 3.4.1

DATOS SOBRE LA ECONOMIA FRANCESA (1949-1966).VARIABLES EXPLICATIVAS Y EXPLICADA PARA EL MODELO DEFINIDO EN (67). (CIFRAS EN MILES DE MILLONES DE FRANCOS FRANCESES)

AÑOS	IMPORT	DOPROD	STOCK	CONSUM
49	15.9	149.3	4.2	108.1
50	16.4	161.2	4.1	114.8
51	19.0	171.5	3.1	123.2
52	19.1	175.5	3.1	126.9
53	18.8	180.8	1.1	132.1
54	20.4	190.7	2.2	137.7
55	22.7	202.1	2.1	146.0
56	26.5	212.4	5.6	154.1
57	28.1	226.1	5.0	162.3
58	27.6	231.9	5.1	164.3
59	26.3	239.0	0.7	167.6
60	31.1	258.0	5.6	176.8
61	33.3	269.8	3.9	186.6
62	37.0	288.4	3.1	199.7
63	43.3	304.5	4.6	213.9
64	49.0	323.4	7.0	223.8
65	50.3	336.8	1.2	232.0
66	56.6	353.9	4.5	242.9

Fuente: Malinvaud,E. Statistical Methods of Econometrics.Opus cit.

T A B L A 3.4.2

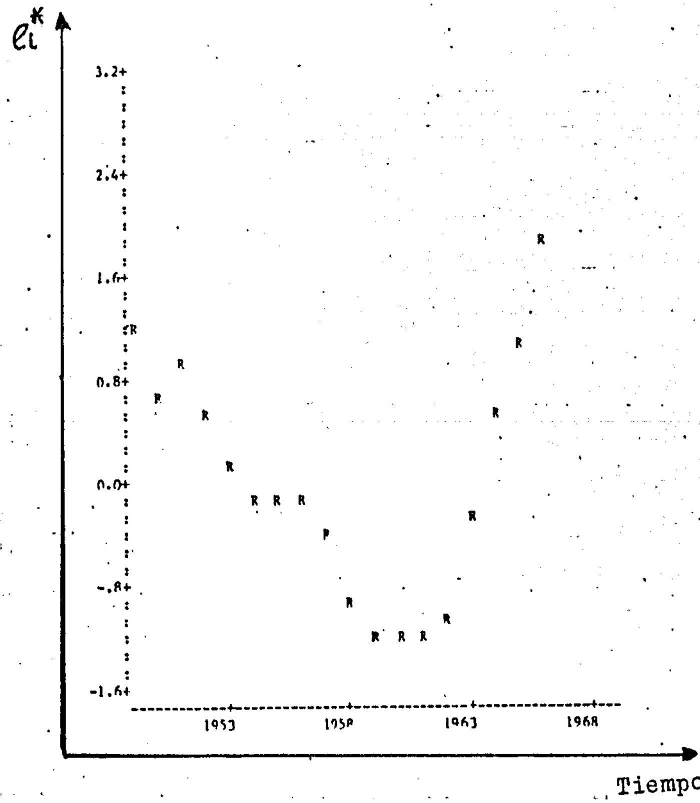
RESULTADOS SUMARIOS DE LA REGRESION DEL MODELO (67) ESTIMADO PARA EL PERIODO 1949-1966.

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDARD	VALORES " t "
DOPROD	0.032	0.187	0.171
STOCK	0.414	0.322	1.286
CONSUM	0.243	0.285	0.853
CONSTANTE	-19.730	4.125	-4.783
n=18	R^2=.973	S=2.258	F (3,14)= 168.45

Fuente: Estimados a partir de los datos de la Tabla 3.4.1

F I G U R A 3.4.2

GRAFICA DE RESIDUALES ESTANDARIZADOS CONTRA EL TIEMPO PARA EL MODELO (67), PERIODO 1949-1966.



Fuente: Chatterjee, S, y B.Price. Opus cit.p.153

TABLA 3.4.3

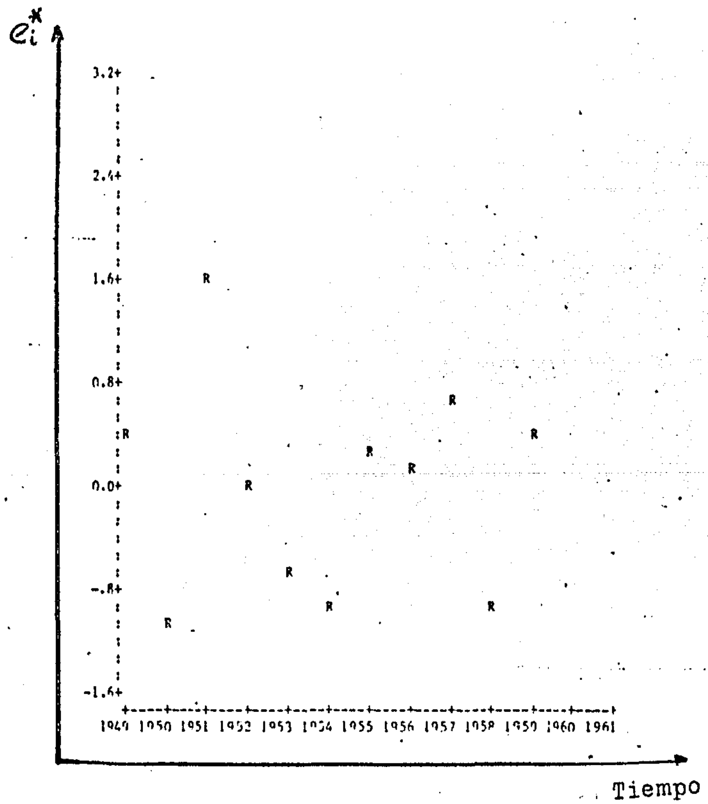
RESULTADOS DE LA REGRESION DEL MODELO (67) ESTIMADO PARA EL PERIODO 1949 - 1959.

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDARD	VALORES "t"	MATRIZ DE CORRELACION DOPROD	STOCK	CONSUM	IMPORT
. DOPROD	-0.051	0.070	-0.728	1	0.026	0.997	0.965
. STOCK	0.587	0.095	6.179		1	0.036	0.251
. CONSUM	0.287	0.102	2.813			1	0.972
. CONSTANTE	-10.130	1.212	-8.358				-
n = 11,	R^2 =.992	S=.489					

Fuente: Estimaciones a partir de los datos de la Tabla 3.4.1

F I G U R A 3.4.3

GRAFICA DE RESIDUALES ESTANDARIZADOS CONTRA EL TIEMPO PARA EL MODELO (67), PERIODO 1949-1959.



Fuente: Chatterjee,S. y B.Price.Opus cit.p.154

la Tabla 3.4.3).

De la Tabla 3.4.3 es evidente que la variación explicada del modelo es alta, 0.99. Sin embargo, el coeficiente de DOPROD es negativo y no es estadísticamente significativo, contrario a las expectativas iniciales. Esto es, era de esperarse, económicamente hablando, que si mantuviésemos constantes las variables STOCK y CONSUM, un incremento en la variable DOPROD implicaría necesariamente un incremento en la variable IMPORT, probablemente en el rubro de materias primas e insumos o equipo de manufactura, como resultado de una expansión en la producción doméstica.

Por consiguiente, la multicolinealidad se torna de una posibilidad, en un problema real para este modelo. Ya que se observa de la Tabla 3.4.3 que el coeficiente de correlación simple entre CONSUM y DOPROD es de 0.997 . Y profundizando más en el análisis se puede demostrar que la variable CONSUM es aproximadamente igual a 2/3 de DOPROD a través de todo el periodo de 11 años (1949-1959). Ya que la relación estimada entre estas dos variables resulta ser:(+)

$$\widehat{CONSUM} = 6.258 + 0.686\ DOPROD \qquad (68)$$

(+) Resultado de efectuar una regresión simple de la variable CONSUM sobre la variable DOPROD, utilizando para ello los datos de la Tabla 3.4.1, para el periodo 1949-1959.

Aún ante la presencia de multicolinealidad severa,como en este caso, el modelo de regresión estimado puede proporcionar buenos pronósticos.La ecuación de pronóstico puede escribirse por lo tanto como:

$$\widehat{\text{IMPORT}} = -10.13 - 0.051\text{DOPROD} + 0.587\text{STOCK} + 0.287\text{CONSUM} \qquad (69)$$

Nótese que el ajuste del modelo anterior a los datos históricos es bastante bueno a juzgar por el valor de R^2 (véase la Tabla 3.4.3) y el comportamiento aleatorio de sus residuales(véase la Figura 3.4.3). Sin embargo, para pronosticar debemos estar seguros que el carácter y la fuerza de la relación global entre las variables explicativas y la explicada permanecerá inalterada en períodos futuros.Esta cuestión de confianza es un problema en cualquier modelo de pronóstico exista o nó multicolinealidad.Para propósitos de este ejemplo, supondremos que la relación global entre las variables que conforman el modelo (69) sigue siendo válida en periodos futuros.(+)

Implícita en la suposición anterior se encuentra la hipótesis de la relación existente entre las variables DOPROD y CONSUM. O sea, el modelo de pronóstico (69) será bueno,es decir,produ-

---

(+) Para propósitos de esta exposición se soslayan las dificultades suscitadas debido a nuestro hallazgo inicial en el sentido de que la formación del Mercomún había alterado las relaciones estructurales del modelo bajo discusión, desde 1960.De donde resulta importante y oportuno advertir al lector que cualquier cambio en la estructura de un modelo puede afectar sustancialmente al pronóstico,aún cuando el ajuste histórico haya resultado excelente.

cirá buenos pronósticos en tanto los valores futuros de las variables DOPROD, STOCK y CONSUM sean tales que:

$$\text{CONSUM} \doteq 2/3\ \text{DOPROD} \tag{70}$$

Por ejemplo, pronostiquemos el cambio en la variable IMPORT para el año siguiente correspondiente a un incremento en DOPROD de 10 unidades, mientras se mantienen las variables STOCK y CONSUM a sus niveles corrientes. El modelo resultante tiene la forma:

$$\widehat{\text{IMPORT}}_{1960} = \widehat{\text{IMPORT}}_{1959} - 0.51 \tag{71}$$

Nótese que la ecuación anterior se obtiene de restar las siguientes ecuaciones:

$$\widehat{\text{IMPORT}}_{1960} = -10.13 - 0.051(239.0+10.0) + 0.587\text{STOCK} + 0.287\text{CONSUM} \tag{72}$$

$$\widehat{\text{IMPORT}}_{1959} = -10.13 - 0.051(239.0) + 0.587\text{STOCK} + 0.287\text{CONSUM} \tag{73}$$

que resultan de sustituir, respectivamente, en la ecuación (69) el valor de DOPROD para 1969 y 1959 - esto es, 249.0 (=239.0+10.0) y 239.0 - manteniendo a las variables STOCK y CONSUM como "constantes".

Obsérvese que si utilizamos el modelo (71) para predecir la variable IMPORT para 1960, obtendríamos un valor de pronóstico que resultaría "subestimado" respecto del verdadero valor histórico reportado para dicha variable en 1960 (véase Tabla 3.4.1), que utilizando el modelo definido en (69).

Aunque el valor subestimado en sí no difiere gran cosa del valor histórico real de dicha variable, como se verá más adelante, es muy importante resaltar el hecho de que desde un punto de vista macroeconómico, el modelo (71) resulta inconsistente puesto que implica que dado un incremento (de 10 unidades en este caso) en el nivel de producción doméstica, o sea, en la variable DOPROD, se produce o corresponde, un cambio negativo o contracción en el nivel de importaciones, esto es:

$$\widehat{IMPORT}_{1960} = \widehat{IMPORT}_{1960} - \widehat{IMPORT}_{1959} = -0.51 \qquad (74)$$

lo cual es un contrasentido, o cuando menos, no es de esperarse, de acuerdo a la teoría económica.

No obstante, si tomamos en cuenta la relación entre las variables DOPROD y CONSUM definida en (70), entonces la variable CONSUM se incrementará en 15 unidades y el modelo resultante adoptará la forma:

$$\widehat{IMPORT}_{1960} = \widehat{IMPORT}_{1959} -0.51 + 4.305 = \widehat{IMPORT}_{1959} + 3.795 \qquad (75)$$

Nótese aquí que la expresión anterior representa un modelo econométrico más congruente que el anterior ya que el cambio en el nivel de importaciones ahora es positivo y por ende, es de esperarse que produzca un mejor pronóstico que el obtenido utilizando el modelo (71).

Esta situación se puede visualizar fácilmente observando la Tabla 3.4.1 que muestra los valores históricos reales de las variables IMPORT, DOPROD, STOCK y CONSUM para los años 1959 y 1960 que se reproducen por comodidad del lector en el siguiente cuadro:

$$\boxed{\begin{array}{llll} IMPORT_{1959} = 26.3 & ; \quad DOPROD_{1959} = 239.0 & ; \quad STOCK_{1959} = 0.7 \\ CONSUM_{1959} = 167.6 & & \\ IMPORT_{1960} = 31.1 & ; \quad DOPROD_{1960} = 258.0 & ; \quad STOCK_{1960} = 5.6 \\ CONSUM_{1960} = 176.8 & & \end{array}} \qquad (76)$$

Ahora bien, sustituyendo el valor histórico de la variable $IMPORT_{1959}$ en la ecuación (75), notamos que el pronóstico para la variable -- $IMPORT_{1960}$ usando dicho modelo es:

$$IMPORT_{1960} = IMPORT_{1959} + 3.795 = 26.3 + 3.795 = 30.095 \qquad (77)$$

que implica una diferencia de un poco más del 3% (3.23) del valor histórico real (= 31.1).

Sin embargo, si hubiésemos utilizado la ecuación (74) el pronóstico para $IMPORT_{1960}$ sería:

$$IMPORT_{1960} = IMPORT_{1959} - 0.51 = 26.3 - 0.51 = 25.79 \qquad (78)$$

en cuyo caso, notamos que el valor de pronóstico difiere en un --- poco más del 17% (17.1) del valor histórico real de la variable -- explicada. O sea, con este modelo, se tiende a subestimar más el valor histórico real, o sencillamente, el valor real de la variable explicada IMPORT, como ya habíamos hecho notar previamente.

Por otra parte, y siguiendo con esta misma línea de ideas, si nos olvidamos por un momento de la relación existente entre las variables CONSUM y DOPROD definida en (70) y simplmmente sustituimos, en el modelo de pronóstico definido en (69), los valores históricos para cada una de las variables explicativas que lo conforman y que se proporcionan para facilidad de esta exposición, en el cuadro de la expresión (76), obtendremos un valor pronosticado para la variable $IMPORT_{1960}$ tal que:

$$\widehat{IMPORT}_{1960} = -10.13 - 0.051(258.0) + 0.587(5.6) + 0.287(176.8)$$
$$= -10.13 - 13.16 + 3.3 + 50.74 = 30.75 \qquad (79)$$

cuyo valor difiere muy ligeramente del obtenido con el modelo "reformulado" definido en (75) ya que sólo difiere en un poco más del 1% (1.12) del valor histórico real (31.1).

En consecuencia, estos ejemplos muestran claramente que los datos no-ortogonales o colineales, no tienen ninguna importancia en el proceso de predicción, obteniéndose por consiguiente, esencialmente, el mismo valor de pronóstico ($\hat{y}_0$) con o sin multicolinealidad en el modelo, siempre que se haya eliminado previamente del mismo, cualquier problema de especificación. El único inconveniente, como ya hemos señalado al comienzo de este apartado, es que la banda del error de pronóstico ante la presencia de multicolinealidad, tiende a ser mayor que sin ella, y por lo tanto, el proceso de inferencia y pruebas de hipótesis sobre el valor de pronóstico ($\hat{y}_0$), se torna menos confiable.

IV. DETECCION DE LA MULTICOLINEALIDAD: TECNICAS DE DIAGNOSTICO.

Como cualquier otro problema estadístico, nuestra primera tarea para atacar la multicolinealidad es detectar su presencia y su grado de severidad. 1/

Con el propósito anterior discutiremos aquí las técnicas más comunes existentes en la literatura para detectar su presencia así como su grado de severidad.

Empezaremos inicialmente con el análisis de las reglas empíricas más conocidas, para luego pasar a discutir las técnicas más sofisticadas.

4.1 REGLAS EMPIRICAS PARA DETECTAR LA MULTICOLINEALIDAD

Regla i) $\boxed{\text{Si } |rij| \geq .8\text{, entonces la multicolinealidad existente es peligrosa.}}$ 2/ (80)

Donde rij es el conocido coeficiente de correlación simple entre Xi y Xj.

1/ Cuando hablamos del "grado de severidad" de la multicolinealidad en un modelo, nos referimos usualmente a la magnitud de su peligrosidad en el mismo. Esto quedará de manifiesto después de cubrir el material de este capítulo en donde analizaremos criterios precisos para determinar cuando ésta es peligrosa.

2/ Citada por Farrar y Glauber. Opus cit.

Esta es la más popular de todas las "reglas del pulgar" en el análisis de regresión y quizás la definición operacional más simple de multicolinealidad "inaceptable". Como se observa, consiste llanamente en limitar los coeficientes de correlación simple, para cada pareja de variables explicativas dentro del modelo, mediante una cota superior de $r = 0.8$ ó 0.9.

Regla ii) <u>La interrelación o multicolinealidad no es necesariamente peligrosa al menos que sea alta en relación al coeficiente de correlación múltiple global.</u> 1/

Esto quiere decir, en otras palabras, que:

Si <u>rij > Ry , entonces la multicolinealidad existente es peligrosa</u>	(8i)

donde Ry es el coeficiente de correlación múltiple entre Y y las variables X1, X2,...,Xk (Nótese también -- que este coeficiente es igual a la raíz cuadrada del conocido "coeficiente de determinación" R^2).

Pero de acuerdo con Farrar y Glauber 2/, tanto la regla i) como la ii) son altamente vulnerables cuando se trabaja con un espacio de más de dos dimensiones. Debido a que la multicolinealidad total es completamente consistente con bajos niveles de corre--

---

1/ Citada por Klein, L.R. <u>An Introduction to Econometrics</u>. Englewood Cliffs, N.J., Prentice- Hall, 1962. pp. 64, 101.

2/ Véase, Farrar y Glauber. Opus cit.

lación simple por parejas, entre los miembros del conjunto de las X's.

Esto es, puede acontecer que la variable Xi, en nuestro modelo general (en forma compacta) que se puede escribir como:

$$y = \beta_1 x_1 + \dots + \beta_i x_i + \dots + \beta_j x_j + \dots + \beta_k x_k + u_i \qquad (82)$$

sea tal que:

$$x_i = \alpha_1 x_1 + \dots + \alpha_{i-1} x_{i-1} + \alpha_{i+1} x_{i+1} + \dots + \alpha_k x_k \qquad (83)$$

es decir, que Xi sea una combinación lineal perfecta de las -- restantes variables explicativas con excepción de Xj, y darse el caso que:

$$r_{ij} \angle .8, \quad \forall j = 1, 2, \dots, i-1, i+1, \dots, k \qquad (84)$$

y también que:

$$r_{ij} \angle .8 , \ \forall i, j = 1, 2, \dots, k \qquad (85)$$

O sea, cumplirse que Xi sea completamente colineal con respecto a todas las demás variables explicativas en el modelo, pero no peligrosamente multicolineal por parejas (en especial, con respecto a Xj).

Regla iii) <u>Regla Empírica de Farrar y Glauber</u> 1/

<u>Una variable Xi es peligrosamente multicolineal sola-- mente si su coeficiente de correlación multiple con -- respecto a las otras variables independientes (Ri), es</u>

1/ Véase, Farrar y Glauber. Opus cit

mayor que el coeficiente de correlación múltiple de Y(Ry) con respecto a todas las variables explicativas del modelo. 1/

Es decir:

Si	Ri $\geqslant$ Ry, entonces la multicolinealidad existente es peligrosa	(86)

O alternativamente:

Si	Ri/Ry $\geqslant$ 1 , entonces la multicolinealidad existente es peligrosa (Ry $\neq$ 0)	(87)

Regla iv) Regla Empírica de Haitovsky 1/

Como la regla empírica de Farrar y Glauber y la sugerida por -- Klein dependen en parte del coeficiente de correlación múltiple entre Y y el conjunto de variables X1,X2,...,Xk, es posible que cierto nivel de colinealidad con respecto a Xi en este conjunto no sea considerado peligroso cuando Ry en el denominador de la expresión (87) cambie debido a un cambio en la variable dependiente. O sea, cuando Ry1 $\neq$ Ry2 , ambas pruebas no -- son concluyentes.

1/ Haitovsky, Y. Multicollinearity in Regression Analysis:Comment. Review of Economics and Statistics, Vol.LI,November 1969, pp.486-489

Entonces la evidencia de peligrosidad de la multicolinealidad propuesta por Farrar y Glauber es diferente para las dos regresiones. Para demostrar esto, Dutta 1/ recurre a un ejemplo muy sencillo.

Supóngase que Ri =0.8 en dos casos separados, Caso I y Caso II. Y además que Ry1=0.4 en el caso I ; Al efectuar una regresión de Y1 sobre el conjunto X1, X2,...,Xk , y Ry2=0.9 en el Caso II; Al efectuar una regresión de Y2 sobre X1,X2,...,Xk.

Observamos entonces que aunque el conjunto de variables explicativas o independientes X1,X2,...,Xk no es menos multicolineal en relación a Y1, que con respecto a Y2, su efecto puede resultar más serio en uno u otro caso, de acuerdo al criterio utilizado para "medir" su nivel de peligrosidad. Así tenemos que;

$$0.8/0.4 > 1 \qquad (\text{Caso I}) \qquad (88)$$

$$0.8/0.9 \not> 1 \qquad (\text{Caso II}) \qquad (89)$$

y por consiguiente, el caso I resulta con nivel peligroso de multicolinealidad, mientras que en el caso II no existe como problema.

Lo anterior implica que las reglas ii) y iii) antes referidas no juzgan en términos de la existencia de multicolinealidad en el conjunto de variables X1,X2,...,Xk , sino en términos de su

---

1/ Véase, Dutta, M. <u>Econometric Methods</u>. Opus cit.

severidad en relación a un parámetro dado, esto es, en relación a la variable dependiente específica "Y" con que se trabaja.

Y por consiguiente, Y.Haitovsky 1/ sugiere la siguiente regla: Reemplazar el numerador de la expresión (87) por el coeficiente de correlación parcial (Rij. 1 2 3 ...k) entre las variables explicativas Xi y Xj cuando las demás variables independientes permanecen constantes. O en otras palabras:

Si $\frac{R_{ij.123\ldots k}}{R_y} \geq 1$, entonces la multicolinealidad existente es peligrosa ($R_y \neq 0$)	(90)

Sin embargo, en nuestra opinión, este método también adolece del mismo defecto que señala M.Dutta. Puesto que la severidad de la multicolinealidad no puede variar al cambiar la variable dependiente. Pues la multicolinealidad es un fenómeno que depende de las variables explicativas exclusivamente, y no presupone o implica una relación causal con respecto a la variable - respuesta, como quedará de manifiesto en nuestras discusiones subsiguientes.

1/ Véase, Haitovsky, Y. Opus cit.

4.2 METODO DE FARRAR Y GLAUBER [1]

Si denotamos como r^{ii} al i-ésimo elemento diagonal de la inversa de la matríz $(X'X)$, es decir, al correspondiente o asociado a la variable Xi, se puede hacer ver que es posible escribir: [2]

$$r^{ii} = \frac{\left|(X'X)_{ii}\right|}{\left|X'X\right|} \quad ; \; i=1,2,\ldots,k \qquad (91)$$

donde el numerador y denominador de la expresión (91) se definen a su vez como se indica en las expresiones (22) y (21), respectivamente.

Como ilustración de su manejo, la aplicaremos al caso k=4 , i.e. para un modelo de 4 variables explicativas. Para i= 3, el elemento diagonal r^{33} asociado a la variable X3 se puede escribir como:

$$r^{33} = \frac{\begin{vmatrix} r_{11} & r_{12} & r_{14} \\ r_{12} & r_{22} & r_{24} \\ r_{41} & r_{42} & r_{44} \end{vmatrix}}{\begin{vmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{vmatrix}} \qquad (92)$$

1/ El método aquí descrito y atribuido a Farrar y Glauber (opus cit.) se inspira en realidad en una investigación anterior realizada por Glauber y otro colega, con un enfoque pragmático, orientada a "localizar la multicolinealidad" mediante un paquete de regresión.
Véase, Glauber,R.R. y E.A. Beaton. Statistical Laboratory Ultimate Regression Package Programming. Cambridge, Massachussets: Harvard Statistical Laboratory,1962.

2/ Véase, Klein,L.R. (1974). Opus cit.p.75

Entonces si la variable explicativa X_i se considera ortogonal con respecto a las demás variables explicativas, es fácil inferir que:

$$\left|(X'X)_{ii}\right| = \left|X'X\right| \tag{93}$$

ya que tanto el numerador como denominador de la expresión (91) están formados por determinantes cuyos elementos no-diagonales son cero y sus elementos diagonales son iguales a la unidad. Esto es, $r_{ij}=0$, para todo $i \neq j$ y $r_{ii}=1$, dado que $X_i.X_j=0, X_i'.X_i=1$ por el supuesto de ortogonalidad entre las variables $X_1, X_2, \ldots, X_k$. Y por consiguiente, la expresión (91) se reduce simplemente a:

$$r^{ii} = 1 \tag{94}$$

Si aplicamos estos resultados a nuestro ejemplo, la expresión (92) se transforma en:

$$r^{33} = \frac{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}} = 1 \tag{95}$$

pues se supone que X_3 es completamente ortogonal respecto de X_1, X_2 y X_4.

Por otra parte, si X_i fuese perfectamente dependiente con respecto a los demás miembros de la matríz X, el denominador se haría cero mientras que el numerador, como no contiene a X_i por definición, no se verá afectado. Esto es, $\left|X'X\right| = 0$, pero en cambio el deter---

minante $\left|(X'X)ii\right| \neq 0$. Entonces r^{ii} quedaría indefinida:

$$r^{ii} = \frac{\left|(X'X)_{ii}\right|}{\left\| X'X \right\|} = \frac{\left|(X'X)_{ii}\right|}{0} \qquad (96)$$

Ahora bien, recordando nuestra definición de variancia estimada de $\hat{\beta}_i$ dada en (20), resulta evidente que utilizando el concepto dado en (91) podemos escribir una nueva versión para la misma en los siguientes términos:

$$\widehat{Var}(\hat{\beta}_i) = \hat{\sigma}_u^2 \, r^{ii} \qquad (97)$$

entonces cuando el elemento r^{ii} crece desmesuradamente en la expresión anterior, la variancia estimada de $\hat{\beta}_i$ también crecerá desmesuradamente.

Farrar y Glauber 1/ sostienen que aún cuando el espectro:

$$1 \leq r^{ii} < \infty \qquad (98)$$

está poco explorado, los elementos diagonales, por su tamaño, nos proporcionan una visión heurística o "cruda" de la severidad relativa de la multicolinealidad así como de la ubicación de ésta dentro del conjunto de variables explicativas del modelo.

1/ Véase, Farrar y Glauber. Opus cit.

Por consiguiente, dichos autores afirman que definiendo la multicolinealidad en términos de su "alejamiento de la ortogonalidad" ofrece dos ventajas:(+)

1o. Distingue claramente entre la naturaleza esencial del problema de la multicolinealidad (no independencia o interdependencia entre las variables explicativas) y los síntomas o efectos que produce gracias a ésta relación de dependencia.

2o. La ortogonalidad se presta con facilidad a la formulación de hipótesis estadísticas y como tal, lleva directamente al desarrollo de estadísticos de prueba ajustados para el número número de variables y observaciones en el conjunto X, contra los que se puede calibrar la severidad en términos de su alejamiento de la ortogonalidad. Desarrollándose con suficiente detalle, dichos estadísticos pueden ofrecer una visión profunda de la localización y el patrón de la multicolinealidad, así como de la severidad de la interdependencia que socaba la calidad experimental de determinada información

Lo antérior se puede entender en cierta forma observando lo que dichos autores han dado por llamar una "relación heurística"

(+)La utilidad del concepto de ortogonalidad aplicado a la detección de la multicolinealidad quedará de manifiesto al analizar el Método de las Componentes Principales que será discutido extensivamente en el próximo apartado.

entre el concepto de ortogonalidad y el determinante de la matríz (X'X) que queda expresado en el espectro:

$$0 \leq |(X'X)| \leq 1 \tag{99}$$

A este respecto, Farrar y Glauber 1/ afirman que transformando a $|(X'X)|$ en un estadístico aproximadamente Ji-cuadrado, se puede obtener una escala adecuada contra la cual se puede calibrar el alejamiento de la ortogonalidad de las variables explicativas contenidas en el conjunto X, y por ende, el gradiente entre la singularidad y ortogonalidad (de la matríz de correlación del modelo de regresión original). Específicamente, el estadístico propuesto tiene la forma: 2/

$$\boxed{\chi^2_{|X'X|}(\nu) = -\left[n-1-\frac{1}{6}(2k+5)\right]\log|X'X|} \tag{100}$$

donde n y k se definen como antes y el miembro izquierdo de la expresión anterior es una variable distribuida aproximadamente como una Ji-cuadrada con $\nu = 1/2\ k(k-1)$ grados de libertad. 3/

Farrar y Glauber sostienen además que, si se acepta la hipótesis normal multivariada, los niveles de probabilidad ofreden una medida cardinal del grado de interdependencia en el conjunto X. Para lo cual proponen una medida cardinal de la severidad de la multicolin

---

1/ Véase, Farrar y Glauber. Opus cit.

2/ Para la derivación y aplicación de este estadístico, Véase, Barlett, B.S. Test of Significance in Factor Analysis. British Journal of Psicology, Statistical Section, 3, 1950, p. 83

3/ Ibid.

lidad eb una variable Xi, que definen como: 1/

$$\boxed{\omega_i = \left(r^{ii} - 1\right)\left(\frac{n-k}{k-1}\right)} \qquad (101)$$

donde ω_i es un estadístico que se distribuye como una variable F con (n-k,k-1) grados de libertad, donde n y k se definen como anteriormente. Entonces, es de esperarse que si F "calculada", definida como el valor absoluto de la expresión (101), resulta mayor que F "tabulada", esto es, F(tablas) = F(n-k,k-1), entonces la variable en cuestión es multicolineal, para un nivel de significación prefijado.

Farrar y Glauber 2/ eventualmente recomiendan que un criterio más confiable para probar la presencia de multicolinealidad en un modelo de regresión o econométrico puede obtenerse de la siguiente manera:

a) Determínense los elementos r^{ii} y los correspondientes coeficientes Ri^2 para cada una de las variables explicativas del modelo.

b) Verifíquese la significancia de cada uno de estos coeficientes, a través de un estadístico de Fisher como el definido en la expresión (104).(Más adelante).

1/ Véase, Farrar y Glauber. Opus cit.

2/ Ibid.

c) Obsérvense los estadísticos F definidos en el paso anterior -- que sean marginalmente significativos o completamente no-significativos para, de entre ellos, detectar a las variables sospechosas de multicolinealidad.

Para facilitar el proceso anterior, los autores referidos recomiendan que a partir del valor del elemento r^{ii}, se obtenga el -- valor del coeficiente Ri^2 mediante la siguiente expresión: 1/

$$\boxed{r^{ii} = 1 - 1/Ri^2} \tag{102}$$

Ahora bien, sabemos por otra parte, que el estadístico F definido como: 2/

$$F = \frac{R^2/K}{(1-R^2)/(n-K-1)} \tag{103}$$

sirve como una prueba estadística para medir la influencia o impacto estadístico de las variables explicativas X1,X2,...,Xk sobre la variable-respuesta "Y". O en otras palabras, mide la significancia global del modelo lineal general definido en (1).

En consecuencia, si aplicamos el estadístico definido en (103) para probar la significancia de cada uno de los coeficientes Ri podemos obtener una nueva versión del mismo, como sigue:

$$\boxed{F_i = \frac{R_i^2/(K-1)}{(1-R_i^2)/(n-K)}} \tag{104}$$

Nótese que con el estadístico anterior se pretende determinar el -- impacto o influencia de las variables explicativas X1,X2,...,Xi-1, Xi+1,...,Xk sobre la variable explicativa Xi considerada como "variable-respuesta".

1/ Véase, Farrar y Glauber. Opus cit.

2/ Véase, Johnston, J. Econometric Methods. McGraw-Hill, Kogakusha Ltd, Tokyo, 1972, pp. 159-168

Farrar y Glauber [1] concluyen que observando el valor para cada "Fi" de la expresión anterior, se podrá detectar el grado de multicolinealidad inherente a cada variable explicativa Xi, a pesar de que la mayoría, sino es que todos, resulten significativos y nos indidarán el área sobre la que debemos dirigir nuestros esfuerzos para buscar mejor y más provechosa información, y eliminar así, la multicolinealidad existente en el modelo.

1/ Véase, Farrar y Glauber. Opus cit.

4.3 METODO DE LAS COMPONENTES PRINCIPALES. 1/

Ya hemos visto a lo largo de nuestra exposición hasta este momento, que la multicolinealidad se asocia usualmente con coeficientes estimados inestables. Esta situación comó producto de la presencia de fuerte dependencia lineal entre las variables explicativas. No es en esencia un problema de especificación. Por lo mismo, la investigación empírica de los problemas derivados de un conjunto de datos colineales debe comenzar solamente después de que el modelo en cuestión haya sido satisfactoriamente

---

1/ El método de las componentes principales usado en este apartado se encuentra en la mayoría de los libros de análisis estadístico multivariado. Se observará que el análisis de las componentes principales se orienta a caracterizar la naturaleza de la información e identificación de las dependencias (cuando existen) entre las variables explicativas de un modelo de regresión o econométrico. A continuación se enlista una bibliografía teórica y aplicada sobre esta técnica, para el lector interesado:

Véase, Press, S.J. Applied Multivariate Analysis. Holt Rinehart and Winston, N.Y. 1972

Véase, Kendall, M.G. A Course in Multivariate Analysis. Charles Griffin, London, 1957.

Véase, Girshick, M.A. Principal Components. Journal of the American Statistical Association, Vol.31, 1936, pp.519-528

Véase, Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. Journal of Educational Psycology, Vol.24, 1933, pp.417-441

Véase, Massy, W.F. Principal Components Regression in Exploratory Statistical Research. Journal of the American Statistical Association, Vol.60, No.309, March 1965, pp.234-256

Véase, Adams, F.G. Consumer Attitudes, Buying Plans and Purchase of Durable Goods: A Principal Component Analysis. Review of Economics and Statistics, Vol.46, 1964, pp.347-355 especificado.

Sin embargo, pueden aparecer signos o indicaciones de multicolinealidad suscitados durante el proceso de adición, eliminación o transformación de variables u observaciones en busca del modelo adecuado. Estas indicaciones que implican inestabilidad en los coeficientes estimados se reflejan en general como:

1. Grandes variaciones en los coeficientes estimados cuando se agrega o elimina una variable explicativa.

2. Grandes variaciones en los coeficientes estimados cuando una observación (o dato) se altera o elimina de la muestra bajo estudio. Ahora bien, una vez que la gráfica de residuos estandarizados indique que el modelo se encuentra satisfactoriamente especificado, la multicolinealidad puede darse si se observa que :

3. Los signos algebraícos de los coeficientes estimados no corresponden o concuerdan con las expectativas teóricas o lógicas.

4. Los coeficientes de las variables que se espera sean importantes, tienen grandes desviaciones estándard.

Para visualizar y entender mejor estos conceptos, analizaremos brevemente el caso del modelo de la "economía francesa" discutido en el Apartado 3.4.

Recordaremos que en dicho modelo, definido en (67) como:

$$IMPORT = \beta o \ + \ \beta 1 \ DOPROD + \beta 2 \ STOCK + \beta 3 \ CONSUM + U$$

era de esperarse que el coeficiente estimado de la variable DOPROD fuese positivo, de acuerdo a la teoría económica.Sin embargo, en la práctica, el coeficiente de dicha variable --- previa eliminación del problema de especificación, resultó ser negativo y no significativo, como se desprende de la Tabla 3.4.3. También notamos que la presencia o ausencia de ciertas varia-- bles tiene un efecto sustancial en los otros coeficientes.(+)

No obstante, puede darse el caso de que los signos algebraícos de los coeficientes de un modelo concuerden con las expectativas lógicas, pero suceder que las desviaciones estándard de los mis— mos sean tan grandes que ocasione que éstos resulten no-significativos. Como ejemplo de esta situación, podemos citar el caso del modelo (46) discutido en el Apartado 3.3. Se recordará allí que los signos algebraícos del modelo en cuestion definido como:

$$APROV = \beta o + \beta 1 \ FAM + \beta 2 \ MON + \beta 3 \ ESC + U$$

era de esperarse que fuesen todos positivos, por la naturaleza misma de las variables FAM, MON y ESC. Se observa entonces, a -- diferencia del caso anterior, que los signos algebraicos para — este modelo, sí concuerdan con las expectativas lógicas, a juz-- gar por los resultados de la Tabla 3.2.2 .No obstante, sus des-- viaciones estándard son tan grandes que hace que todos resulten

(+) Es decir, los coeficientes de este modelo resultan inesta-- bles.Se omiten los detalles de este proceso para obviar esta exposición.

no-significativos. Aunque era de esperarse que todos los coeficientes fuesen importantes, esto es, significativos.

Por otra parte, y como ya hemos apuntado anteriormente, la presencia de multicolinealidad también se manifiesta por el tamaño de los coeficientes de correlación simple entre las variables explicativas. Un coeficiente de correlación grande es indicativo de una fuerte dependencia lineal entre dos variables. Los coeficientes de correlación, por parejas, para las variables del modelo (46) resultaron muy grandes a juzgar por los valores obtenidos: .960 (FAM y MON), .986 (FAM y ESC) y .986 (MON y ESC) manifestando con ello , una fuerte dependencia lineal entre las mismas. De la misma manera, para el caso del modelo (67) de la "economía francesa" el coeficiente de correlación entre DOPROD y CONSUM resultó ser de .997 como se desprende de la Tabla 3.4.3, evidenciando también una fuerte dependencia lineal entre ambas variables, contrastando con los valores de los coeficientes de correlación entre la interacción de dichas variables y la variable STOCK. Esto es, de .036 (entre CONSUM y STOCK) y de .026 (DOPROD y STOCK).

Sin embargo, en la práctica, la fuente de multicolinealidad en un modelo puede ser más sutíl que una simple relación entre dos variables. Una relación lineal puede involucrar a muchas de las variables explicativas. No es posible detectar dicha relación con un simple coeficiente de correlación simple, o en general, mediante reglas empíricas como las estudiadas anteriormente. Como ejemplo de esta situación analizaremos el siguiente caso:

Supóngase que deseamos investigar el efecto o impacto de los gastos de publicidad (PBt,PBt-1: en el año corriente y año anterior, respectivamente), gastos de promoción (PRt,PRt-1: en el año corriente y año anterior, respectivamente) y gastos de ventas (GVt), sobre las ventas totales (VTt) de una cierta empresa,utilizando para ello la información de la Tabla 4.3.1 (que aparece en la próxima hoja) y que presenta los datos históricos para dichas variables para un total de 23 años durante los cuales dicha empresa ha operado en condiciones muy estables. El modelo de "publicidad" propuesto queda formulado en los siguientes términos: 1/

$$VT_t = B_o + B_1\ PB_t + B_2\ PR_t + B_3\ GV_t + B_4\ PB_{t-1} + B_5\ PR_{t-1} + U \qquad (105)$$

Los resultados de la regresión del modelo (105) se expresan en la Tabla 4.3.2 (enseguida de la Tabla 4.3.1). Se observa tanto de la gráfica de residuales contra los valores estimados de la variable explicada (VT), esto es, de la Figura 4.3.1 (enseguida de la Tabla 4.3.2), como de la gráfica correspondiente de los residuales estandarizados contra el tiempo, o sea, la Figura 4.3.2 (inmediatamente después de la Figura 4.3.1), que no existen problemas de mala especificación en dicho modelo.

Por otra parte, de la Tabla 4.3.2 observamos que los coeficientes de correlación entre las variables explicativas son pequeñas.

1/ Tomado de Chatterjee,S. y B.Price.Opus cit.p.157

TABLA 4.3.1

DATOS ANUALES SOBRE GASTOS DE PUBLICIDAD (PB),GASTOS DE PROMOCION (PR),GASTOS DE VENTAS (GV) Y VENTAS TOTALES (VT), EN MILLONES DE DOLARES CORRIENTES,PARA EL MODELO (105).

Hilera	PB	PR	GV	PB-1	PR-1	VT
* 1 *	1.98786	1.0	0.30	2.01722	0.0	20.11371
* 2 *	1.94418	0.0	0.30	1.98786	1.0	15.10439
* 3 *	2.19954	0.8	0.35	1.94418	0.0	18.68375
* 4 *	2.00107	0.0	0.35	2.19954	0.8	16.05173
* 5 *	1.69292	1.3	0.30	2.00107	0.0	21.30101
* 6 *	1.74334	0.3	0.32	1.69292	1.3	17.85004
* 7 *	2.06907	1.0	0.31	1.74334	0.3	18.87558
* 8 *	1.01709	1.0	0.41	2.06907	1.0	21.26599
* 9 *	2.01906	0.9	0.45	1.01709	1.0	20.48473
* 10 *	1.06139	1.0	0.45	2.01906	0.9	20.54032
* 11 *	1.45999	1.5	0.50	1.06139	1.0	26.18441
* 12 *	1.87511	0.0	0.60	1.45999	1.5	21.71606
* 13 *	2.27109	0.8	0.65	1.87511	0.0	28.69595
* 14 *	1.11191	1.0	0.65	2.27109	0.8	25.83720
* 15 *	1.77407	1.2	0.65	1.11191	1.0	29.31987
* 16 *	0.95878	1.0	0.65	1.77407	1.2	24.19041
* 17 *	1.98930	1.0	0.62	0.95878	1.0	26.58966
* 18 *	1.97111	0.0	0.60	1.98930	1.0	22.24466
* 19 *	2.26603	0.7	0.60	1.97111	0.0	24.79944
* 20 *	1.98346	0.1	0.61	2.26603	0.7	21.19105
* 21 *	2.10054	1.0	0.60	1.98346	0.1	26.03441
* 22 *	1.06815	1.0	0.58	2.10054	1.0	27.39304

Fuente: Chatterjee,S. y B.Price.Opus cit. p.158

TABLA 4.3.2

RESULTADOS SUMARIOS DE LA REGRESION DEL MODELO (105)

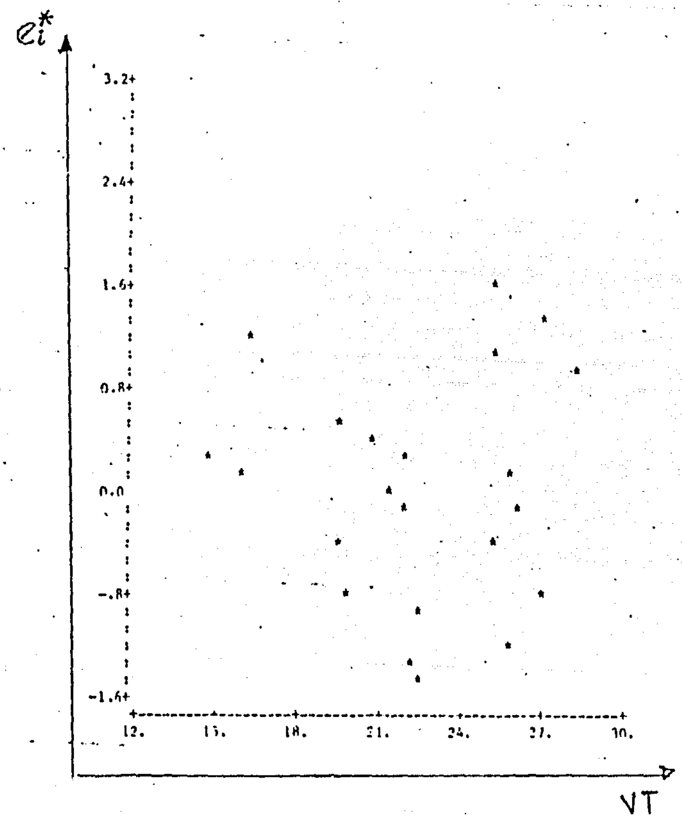
VARIABLE	COEFICIENTE ESTIMADO	DESV ESTAND.	VALORES "t"	MATRIZ DE CORRELACION PB	PR	GV	PB-1	PR-1	VT
PB	5.3606	4.0277	1.331	1	-.35695	-.12852	-.13974	-.49599	-.17041
.PR	8.3722	3.5864	2.334		1	.06259	-.31646	-.29636	.54081
.GV	22.5210	2.1424	10.512			1	-.16643	.20811	.81087
.PB-1	3.8545	3.5778	1.077				1	-.35776	-.30516
.PR-1	4.1247	3.8952	1.059					1	-.05204
CONST.	14.1900	18.7150	-.758						-

n=22 , R^2=.9169, S = 1.3200

Fuente: Estimaciones a partir de los datos de la Tabla 4.3.1

F I G U R A 4.3.1

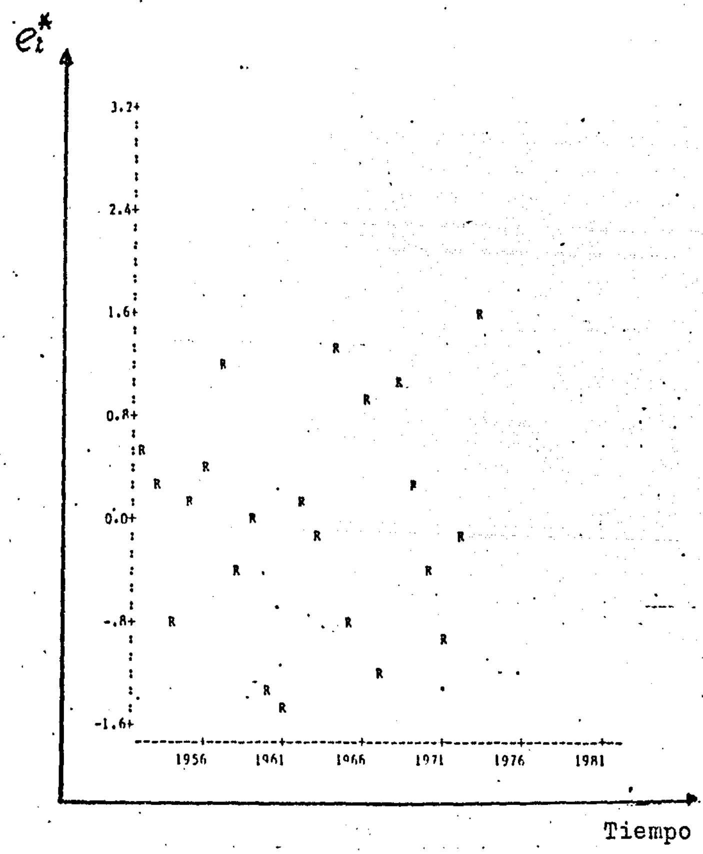
GRAFICA DE RESIDUALES ESTANDARIZADOS CONTRA VALORES ESTIMADOS DE LA VARIABLE-RESPUESTA (VT), PARA EL MODELO (105).



Fuente: Chatterjee, S. y B.Price. Opus cit.p.159

F I G U R A 4.3.2

GRAFICA DE RESIDUALES ESTANDARIZADOS CONTRA EL TIEMPO, PARA EL MODELO (105).



Fuente: Chatterjee, S. y B. Price. Opus cit. p.160

Sin embargo, si realizamos un poco de experimentación para verificar la estabilidad de los coeficientes estimados, eliminando por ejemplo, la variable PB_t(gasto de publicidad) del modelo,se producirían muchos cambios: El coeficiente de la variable PR_tse reduciría de 8.37 a 3.83 (se desplomaría en un 54.2%), y los coeficientes rezagados de PBt-1, PRt-1 cambiarían de signo y ya no serían significativos. Pero el coeficiente de determinación de"ventas totales" (VT) es estable y R^2 no cambia mucho.(+)

La evidencia empírica sugiere entonces la existencia de algún tipo de relación involucraddo las variables PBt y PBt-1 y PRt-1, puesto que la regresión de PBt respecto de PBt-1 y PRt-1, cuya ecuación estimada es:

$$\hat{PBt} = 4.63 - 0.87\ PRt - 0.86\ PBt\text{-}1 - 0.95\ PRt\text{-}1 \qquad (106)$$

posee un coeficiente $R^2 = .973$, manifestando así una fuerte dependencia lineal entre dichas variables.

Investigando con más detalle las operaciones de la firma bajo estudio, se detectó que la misma ejerció una estricta política de control presupuestal a lo largo de los 23 años de vida estable de la misma.En particular, se determinó una regla empírica aproximada impuesta en materia presupuestaria por la firma,--

(+) Estos cómputos se omiten para obviar esta exposición.Es instructivo y deseable que al lector interesado los realice por su cuenta a fin de familiarizarse con los conceptos y procedimientos hasta aquí disdutidos.

consistente en hacer que "la suma de las variables PBt, PBt-1, PRt y PRt-1, fuese aproximadamente igual a 5 unidades," por cada período de 2 años. Es decir:

$$PBt + PRt + PBt\text{-}1 + PRt\text{-}1 \doteq 5 \qquad (107)$$

lo cual evidentemente explica la multicolinealidad en el modelo (105). Por lo tanto, se desprende que una búsqueda completa de la multicolinealidad en un modelo de regresión o econométrico -- en general, incluye la verificación de los valores de R^2 para -- cada variable explicativa (Xi) contra, o con respecto a, las demás. Lo cual concuerda con nuestras observaciones y criterios anteriormente discutidos para detectar la presencia de multicolinealidad.

Sin embargo, como se observa del modelo (105), llegar a descubrir - sin conocer los antecedentes específicos del caso - que el origen o fuente de la multicolinealidad en dicho modelo, se debía -- precisamente a que sus variables guardaban en realidad una dependencia o "patrón" lineal como el que se muestra en la expresión (107) no es tarea fácil. La técnica de las componentes principales nos ayudará a resolver este tipo de problemas. Antes de aplicarla a la resolución de un problema en particular, discutiremos brevemente en qué consiste.

Dado nuestro modelo general definido (matricialmente) en (3) como:

$$Y = XB + U$$

con las mismas hipótesis sobre U establecidas anteriormente, excepto que ahora, tanto sus elementos como los de las matrices Y y X han sido estandarizados previamente. De tal suerte que en -

particular ,los productos $X'X$ y $X'Y$ son ahora matrices de coeficientes de correlación.(+) Entonces existe una matríz C (no-singular de orden k) tal que;

$$C'(X'X)C = \Lambda \quad \text{y} \quad C'C = CC' = I \qquad (108)$$

donde Λ es una matríz diagonal cuyos elementos son las raíces características "ordenadas" de la matríz (X'X). Si denotamos a estas raíces como $\lambda_1, \lambda_2, \ldots, \lambda_k$, entonces podemos escribir o expresar a la matríz Λ como ;

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \lambda_k \end{bmatrix} ; \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k \qquad (109)$$

Nótese que las columnas de la matríz C son los vectores característicos normalizados correspondientes a las raíces características $\lambda_1, \lambda_2, \ldots\ldots, \lambda_k$ y la matríz C puede usarse para estimar un nuevo conjunto de variables explicativas, esto es;

$$(W(1), W(2), \ldots, W(k)) = W = XC = (X(1), X(2), \ldots, X(k)) \qquad (110$$

que son funciones lineales de las variables explicativas originales.Las nuevas variables W's se conocen como "componentes principales".Entonces la ecuación (3) podemos reescribirla en términos de las componentes principales como;

$$Y = W\alpha + U \qquad (111)$$

donde W es una matríz de tamaño (nxk) y α es un vector-columna

---

(+)Nótese en particular, que con este supuesto coinciden la matríz (X'X) y la matríz de correlación R definida en (30), o sea, (X'X)= R.

de dimensión k, tales que:

$$W = XC \quad y \quad \alpha = C'B \tag{112}$$

El concepto de componentes principales y raíces característi-- cas puede utilizarse para detectar y analizar la colinealidad (o multicolinealidad) existente entre las variables explicati- vas de un modelo de regresión o econométrico. La reformulación del modelo lineal general en términos del modelo de regresión definido en (111) debe concebirse, en consecuencia, como una "re-- parametrización" de la ecuación (3), o sea, un replanteamiento del modelo lineal general en términos de variables explicati-- vas ortogonales. Es decir, en términos de variables W_i $(i=1,2,..,k)$ tales que:

$$W_i' \cdot W_j = 0 \ (i \neq j), \quad W_i' \cdot W_i = \lambda_i \tag{113}$$

donde la raíz característica λ_i corresponde a la variancia mues- tral de las componentes principal W_i. Luego entonces, si $\lambda_i = 0$, entonces todas las observaciones en la i-ésima son también cero. Y como la i-ésima componente principal es una función lineal de las X's, digamos de la forma:

$$W_i = C_{1i}X_1 + C_{2i}X_2 + \dots + C_{ki}X_k \tag{114}$$

se sigue que, cuando $\lambda_i = 0$, existe una relación lineal perfecta entre las variables explicativas del modelo original como la -- definida en la expresión anterior.

<u>De lo anterior se infiere que cuando λ_i es pequeña (aproximada-- mente cero) existe una relación lineal aproximadamente exacta entre las variables explicativas. Es decir, una raíz característi-</u>

ca pequeña es un indicador de la existencia de multicolinealidad en un modelo. Además de que la expresión (114) identifica o refleja el "patrón" o forma exacta de la relación lineal que ocasiona u orgina la multicolinealidad.

Conviene de paso hacer referencia en esta pequeña discusión teórica sobre la técnica que nos concierne, a algunos aspectos interesantes sobre la "precisión de las funciones lineales construidas a partir del estimador $\hat{\beta}$". 1/

Si denotamos a $\hat{\alpha}$ y $\hat{\beta}$ como los estimadores de mínimos cuadrados de α y β, respectivamente, entonces se puede demostrar que:

$$\hat{\alpha} = C'\hat{\beta} \qquad \text{y viceversa} \qquad \hat{\beta} = C\hat{\alpha} \tag{115}$$

donde:

$$\hat{\alpha} = (W'W)^{-1} W'Y \tag{116}$$

se sigue entonces que la matríz de variancia-covariancia de $\hat{\alpha}$ se puede expresar como:

$$\text{Var}(\hat{\alpha}) = \Lambda^{-1} \sigma_u^2 \tag{117}$$

y la correspondiente matríz de variancia-covariancia de $\hat{\beta}$ es:

$$\text{Var}(\hat{\beta}) = C\Lambda^{-1}C' \sigma_u^2 \tag{118}$$

1/ Para una discusión teórica detallada , así como la justificación matemática de los conceptos aquí vertidos sobre este particular,
Véase, Silvey, S.D. Multicollinearity and Imprecise Estimation. J. Royal Statistical Society, Vol. 31, 1969, pp. 539-552

En síntesis, la variancia de la función $\hat{\varphi}$ resulta ser una simple combinación lineal de los recíprocos delas raíces características. Se sigue entonces que la función $\hat{\varphi}$ tiene una"buena precisión", si ninguna de las raíces características es cercana a cero, o bien, si ninguna excede a su correspondiente ri^2, cuando ri es muy pequeña (o sea, cuando λ_i es tal que: $0 < \lambda_i \leq r_i^2$; con ri pequeña). Además, siempre es posible seleccionar un vector"T", y por ende,construir una función de $\hat{\beta}$ de tal forma de eliminar el efecto de una o dos raíces características pequeñas, y lograr que la función definida en (120) tenga una variancia pequeña. 1/

A propósito de la presente discusión, conviene mencionar aquí una regla empírica citada por Chatterjee y Price 2/ para detectar la multicolinealidad en un modelo y que utiliza precisamente el concepto de las componentes principales:

"Si alguna de las raíces características es menor que 0.01 ó la suma de los recíprocos de todas las raíces del modelo en cuestión es mayor que cinco veces el número de variables explicativas en dicho modelo, entonces dichas variables se consideran'colineales' (i.e. existe multicolinealidad en el modelo referido).De lo con-

---

1/ Aunque estos conceptos se desarrollan con un poco más de detalle al tratar el Método de la Dirección Optima de Silvey en este mismo capítulo, en el presente apartado - adelantándonos un poco a dicha discusión - buscaremos su aplicación a casos concretos para diagnosticar la multicolinealidad así como para descubrir el "patrón" de multicolinealidad inherente en los modelos estudiados.
Véase, Silvey,S.D.Opus cit.pp.539-552

2/ Véase, Chatterjee,S. y B.Price.Opus cit.p.200

trario, las variables se consideran como'no-colineales'(i.e.no existe multicolinealidad)" 1/

Antes de proceder a abordar un caso concreto de aplicación de -- la técnica antes descrita para detectar la multicolinealidad en un modelo de regresión y como ésta permite descubrir el "patrón" de multicolinealidad existente en el mismo, conviene hacer algunas consideraciones adicionales importantes a este respecto.

Los indicadores para detectar la multicolinealidad que hemos analizado hasta ahora pueden derivarse u obtenerse mediante procedimientos computacionales normales de análisis de regresión. Empero, la técnica de las componentes principales implica procedimientos de cómputo que usualmente no se incluyen en los paquetes de regresión estándares. Como se puede notar de la suscinta exposición teórica sobre la misma, expuesta líneas arriba, esta técnica se basa en el hecho de que el modelo lineal general se puede reexpresar en términos de un conjunto de variables explicativas ortogonales (véase la expresión (111)).

Ya hemos señalado que estas nuevas variables ortogonales o "componentes principales" (i.e. las Wi's definidas en (110)), se -- pueden obtener como una combinación lineal del conjunto de --- variables explicativas originales.

1/ Véase, Chatterjee, S. y B. Price. Opus cit. p. 200

Sea T un vector arbitrario de constantes de orden (kxl). La función lineal definida como:

$$\varphi = T'\beta \tag{119}$$

tiene como estimador de mínimos cuadrados a:

$$\hat{\varphi} = T'\hat{\beta} \tag{120}$$

y una matríz de variancia-covariancia de la forma:

$$Var(\hat{\varphi}) = T'C\Lambda^{-1}C'T\sigma_u^2 \tag{121}$$

Sea Ci la i-ésima columna de la matríz C. Entonces el vector T puede escribirse como:

$$T = \sum_{i=1}^{k} riCi = r1C1 + r2C2 + \ldots + rkCk \tag{122}$$

donde r1,r2,...,rk son constantes apropiadamente seleccionadas. Entonces la ecuación (121) se transforma en:

$$Var(\hat{\varphi}) = S'\Lambda^{-1} S\,\sigma_u^2 \tag{123}$$

donde la matríz S es tal que: S = C'T.

O equivalentemente:

$$Var(\hat{\varphi}) = \left(\sum \frac{r_i^2}{\lambda_i}\right)\sigma_u^2 \tag{124}$$

En otras palabras:

$$Var(\hat{\varphi}) = \left(\frac{r_1^2}{\lambda_1} + \frac{r_2^2}{\lambda_2} + \ldots + \frac{r_k^2}{\lambda_k}\right)\sigma_u^2 \tag{125}$$

Sin embargo, es oportuno observar, que el modelo definido en (111) a partir de las componentes principales, carece de una interpretación simple, ya que como se nota, dichas variables son en realidad una "mezcla" o combinación de las variables explicativas originales. Aunque, como quedará de manifiesto más adelante, la técnica de las componentes principales proporciona un enfoque unificado para obtener información sobre el patrón de multicolinealidad en un modelo y nos servirá de base como una técnica de estimación alternativa para corregirla - aspecto este último que se discutirá en detalle en el próximo capítulo -.

Para desarrollar el método de las componentes principales, primero notemos que la base de cualquier análisis de la multicolinealidad subyace en la estructura de las correlaciones de las variables explicativas. Como las correlaciones no se ven afectadas por la escala, basta y conviene trabajar con variables estandarizadas. Ya hemos visto que una variable estandarizada se obtiene restando a cada observación su valor promedio y dividiendo esta diferencia entre la desviación estándard de las observaciones.(+)

Esto es, si denotamos como X_i^* y y_i^* a las variables estandarizadas correspondientes a las variables Xi y Yi, respectivamente, entonces resulta evidente que se cumplen las siguientes condiciones

---

(+) Véase la definición de variable estandarizada dada en la expresión (14).

para las variables estandarizadas x_i^* y y_i^*:

$$E(x_i^*)=0,\ E(y_i^*)=0,\ var(x_i^*)=1,\ Var(y_i^*)=1$$

$$y\ cov(x_i^*, y_i^*)=r_{ij} \qquad (126)$$

Por otra parte, conviene recordar que un conjunto de variables es linealmente independiente si no existe relación lineal entre las mismas. Es decir, si X1,X2,...,Xk son linealmente independientes, entonces:

$$\theta_1 X1 + \theta_2 X2 + \theta_3 X3+\ldots+\theta_k Xk = 0 \qquad (127)$$

si y solo si: $\theta_1 = \theta_2 =\ldots = \theta_K = 0$.

A propósito, es fácil hacer ver que si X1,X2,...,Xk son ortogonales, entonces también son linealmente independientes. Esto es, si

$$Xi \,.\, Xj = 0 \ ,\ (i \neq j\ ),\quad Xi\,.Xj' = 1 \qquad (128)$$

para todo i,j = 1,2,...,k, entonces (son ortogonales y) también cumplen la condición (127).(+)

Entonces la matríz de variancia-covariancia del vector $\hat{\beta}$ del modelo lineal general (3) definido a partir de un conjunto X1,X2,... ...,Xk de variables ortogonales, previamente estandarizadas, es de la forma:

$$var(\hat{\beta}) = \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & \ldots & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} \qquad (129)$$

(+) Lo contrario no es necesariamente cierto. Considérense los vectores (3,0,2), (7,0,9) y (4,1,2) que son linealmente independientes pero no ortogonales. En cambio, los vectores (1,0,0), (0,1,0) y (0,0,1) son ortogonales y linealmente iddependientes.

La matríz de variancia-covariancia, calculada a partir de los -- datos previamente estandarizados provenientes de la Tabla 3.4.1 resulta ser:

$$\begin{array}{c} \\ DOPROD^* \\ STOCK^* \\ CONSUM^* \end{array} \begin{array}{c} \begin{array}{ccc} DOPROD^* & STOCK^* & CONSUM^* \end{array} \\ \left[\begin{array}{ccc} 1 & .026 & .997 \\ .026 & 1 & .036 \\ .997 & .036 & 1 \end{array}\right] \end{array} \qquad (129\text{-bis})$$

Por otra parte, sabemos de nuestra discusión téorica de la técnica de las componentes principales (véanse las expresiones (111) y (112)) que la variable- respuesta IMPORT se puede escribir como una combinación lineal de sus componentes principales, que denotaremos como Z1,Z2 y Z3 y que se puede hacer ver que pueden definirse mediante la transformación :(+)

$$\begin{aligned} Z_1 &= .7063\ DOPROD^* + .0435\ STOCK^* + .7065\ CONSUM^* \\ Z_2 &= -.0357\ DOPROD^* + .9990\ STOCK^* - .0258\ CONSUM^* \\ Z_3 &= -.7070\ DOPROD^* - .0070\ STOCK^* + .7072\ CONSUM^* \end{aligned} \qquad (130)$$

La matríz de variancia-covariancia para estas nuevas variables -- tiene la forma general:

$$\begin{array}{c} \\ Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{array} \begin{array}{c} \begin{array}{cccc} Z_1 & Z_2 & \cdots & Z_k \end{array} \\ \left[\begin{array}{cccc} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{array}\right] \end{array} = \Lambda \qquad (130\text{-bis})$$

(+) **Las constantes que definen la transformación (130) corresponden, respectivamente, a las componentes de los vectores característicos normalizados correspondientes a las variables independientes.**

Para el ejemplo que estamos manejando, la matriz Λ definida en (131) aplicado a los datos del modelo (67) resulta ser:

$$\begin{array}{c|ccc} & Z_1 & Z_2 & Z_3 \\ \hline Z_1 & 1.999 & 0 & 0 \\ Z_2 & 0 & .998 & 0 \\ Z_3 & 0 & 0 & .003 \end{array} \qquad (131)$$

esto es, $\lambda_1 = 1.999$; $\lambda_2 = .998$; $\lambda_3 = .003$. Donde se percibe claramente la existencia de multicolinealidad dado el valor pequeño de λ_3 respecto de λ_1 y λ_2.

Para las variables del modelo "escolar" definido en (46), la matríz Λ tiene la forma;

$$\Lambda = \begin{bmatrix} 2.952 & 0 & 0 \\ 0 & .040 & 0 \\ 0 & 0 & .008 \end{bmatrix} \qquad (132)$$

en donde también detectamos inmediatamente, la presencia de multicolinealidad manifiesta por el valor de $\lambda_3 = .008$, en relación con los valores de $\lambda_1 = 2.952$ y $\lambda_2 = .040$, lo cual era de esperarse.

Análogamente, para el modelo de "publicidad" definido en (105) la matríz correspondiente sería:

$$\Lambda = \begin{bmatrix} 1.701 & 0 & 0 & 0 & 0 \\ 0 & 1.288 & 0 & 0 & 0 \\ 0 & 0 & 1.145 & 0 & 0 \\ 0 & 0 & 0 & .859 & 0 \\ 0 & 0 & 0 & 0 & .007 \end{bmatrix} \qquad (133)$$

cuyo coeficiente $\lambda_5 = .007$ muestra claramente la presencia de multicolinealidad en dicho modelo.

Una información adicional que es posible obtener a partir de este análisis es el siguiente: Como las λ_i's son las variancias de las componentes principales, si λ_i es aproximadamente igual a cero, la correspondiente componente principal (Z_i) es aproximadamente igual a una constante.

De lo anterior se desprende que la ecuación que define la componente principal afectada proporciona una idea del tipo de relación entre las variables explicativas que está causando la multicolinealidad que afecta el modelo en cuestión. Como ya hemos destacado anteriormente.

Con estos elementos como marco, podemos proceder a discutir a continuación la aplicación de la técnica de las componente principales a dos modelos particulares - previamente analizados - con el propósito de identificar el "patrón de colinealidad" existente entre las variables explicativas de los mismos.

Comenzaremos abordando el modelo de la "economía francesa" definido en (67) mediante la ecuación:

$$IMPORT = \beta o + \beta 1 .DOPROD + \beta 2.STOCK + \beta 3.CONSUM + U$$

Como sabemos de la matríz Λ definida en (135) que $\lambda 3 = .003 \doteq 0$, entonces, la componente principal asociada a esta raíz (Z3) es casi constante. Esta constante es el valor medio de la variable Z3 que es cero, ya que todas las componentes principales tienen media cero, pues son combinaciones lineales de variables estandarizadas cuyas medias son cero. En consecuencia, de la expresión (130) tendremos que:

$$Z3 = - .7070\ DOPROD^* - .0070 STOCK^* + .7072 CONSUM^* = 0 \qquad (134)$$

que en forma simplificada equivale a:

$$CONSUM^* \doteq DOPROD^* \qquad (135)$$

que nos representa la relación aproximada sobre las versiones estandarizadas de las variables CONSUM y DOPROD. Lo cual es consistente con nuestro hallazgo inicial basado únicamente en el coeficiente de correlación simple entre las variables CONSUM y DOPROD, que era de r = .997.

Como $\lambda 3$, es la única raíz característica más pequeña, el método de las componentes principales nos indica que la estructura de la dependencia entre las variables explicativas del modelo (67), a juzgar por la manera en que se refleja en los datos, no es más compleja que una simple relación lineal entre las variables CONSUM y DOPROD que era de r = .997.

De igual manera, para el modelo de "publicidad" definido en (105) como:

$$VT_t = Bo + B1\ PBt + B2\ PRt + B3\ GVt + B4\ PBt\text{-}1 + B5\ PRt\text{-}1 + U \qquad (105)$$

tenemos que la raíz más pequeña de la matríz definida en (133) es $\lambda 5 = .007$. Entonces la componente principal asociada a esta raíz es Z5, definida como sigue:(+)

$$Z5 = .514\ PBt^{*} + .489\ PRt^{*} - .010\ GVt^{*} + .428\ PBt^{*}\text{-}1 + .559\ PRt^{*}\text{-}1 \qquad (136)$$

Si hacemos Z5 =0 y despejamos a la variable PBt^{*} de la expresión anterior, tendremos que:

$$PBt^{*} = -.951\ PRt^{*} - .833\ PBt^{*}\text{-}1 - 1.087\ PRt^{*}\text{-}1 \qquad (137)$$

que confirma nuestro previo hallazgo sobre la relación entre — las variables PBt, PRt, PBt-1, PRt-1. Además, como las raíces restantes: $\lambda 4 = .859$, $\lambda 3 = 1.145$, $\lambda 2 = 1.288$, $\lambda 1 = 1.701$ poseen grandes valores, podemos estar seguros que la única fuente de colinealidad en el modelo (105) es la representada por la relación dada en (137).

Para finalizar este apartado, conviene destacar que, a través — del mismo, hemos discutido diferentes técnicas para detectar la presencia de colinealidad en un modelo a través de la magnitud

---

(+) Se omite la descripción del sistema de componentes principales correspondiente a este modelo en obvio de simplificación de — esta exposición.

de diversos indicadores, llámese coeficiente de correlación simple, de determinación,etc., o inclusive raíz característica.Aunque usualmente hablamos en términos de "grande" o "pequeño", no hay manera de establecer valores "frontera" que precisen dichos conceptos. El concepto de magnitud debe servirnos únicamente para proporcionarnos una indicación de que todo se encuentra "en orden" o que "algo anda mal".

Por lo tanto, el único criterio razonable para juzgar el tamaño o magnitud de un indicador para detectar la multicolinealidad,es decidir si la ambigüedad resultante como consecuencia de la presencia de multicolinealidad es o nó de importancia fundamental para el modelo en cuestión.

De cualquier manera, e independientemente del nivel de peligrosidad de la multicolinealidad en un determinado modelo, consideramos que el método de las componentes principales, dada su naturaleza y la manera como opera, es una técnica muy poderosa que nos permite no sólo detectar la presencia de multicolinealidad en el modelo afectado, sino también y sobre todo, conocer el patrón de colinealidad existente entre las variables explicativas que lo conforman, lo cual la convierte en una de las más efectivas para analizar y atacar el problema de multicolinealidad, aunque claro está, implique un poco más de computo e investigación que las técnicas tradicionales.Pero consideramos que el esfuerzo adicional es recompensado con creces por los beneficios obtenidos.Cosa parecida podemos decir del Método de Regresión de Cordillera que es el tema del próximo apartado, y que combinado con la técnica objeto de este apartado, ofrece un enfoque muy poderoso y efectivo no sólo para detectar sino también para corregir la multicolinealidad en un modelo de regresión o econométrico, como quedará de manifiesto en el transcurso de esta investigación. 1/

1/ Este comentario desde luego es válido para todas las técnicas y métodos que serán abordados a través del presente trabajo, que consideramos para efectos prácticos, como las más conocidas y ampliamente utilizadas actualmente en econometría y en el análisis de regresión en general. Sin embargo, para poder generalizar o extrapolar esta afirmación al universo de técnicas hasta hoy investigadas sobre la materia, tendríamos en justicia que analizar y discutir en detalle, algunos otros enfoques que aunque indudablemente novedosos e interesantes, su tratamiento e inclusión se omiten aquí en obvio de tiempo y por exceder los alcances de esta obra. De cualquier manera, a continuación se enlistan algunas referencias para el lector interesado en conocer y profundizar sobre algunos de dichos enfoques:

Véase, Besley, Kuh and Welsh. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley and Sons. 1980.

Véase, Willan and Watts. Meaningful Multicollinearity Measures, Technometrics, Vol.20, No.4, November 1978.

Véase, Swindel, B.F. Instability of Regression Coefficients Illustrated. The American Statistician, 28(2), 1974, pp.63-65

4.4 METODO DE LA REGRESION DE CORDILLERA 1/

Antes de pasar a discutir la forma de detección de la multicolinealidad mediante este método, conviene describir brevemente su metodología.

Considérese el modelo original en su forma matricial (3):

$$Y = X\beta + U$$

donde como sabemos, Y es el vector columna de observaciones — correspondiente a la variable-respuesta de dimensión n, X es la matríz dd observaciones correspondiente a las k variables explicativas, de tamaño nxk, β es el vector columna de coeficientes de regresión (o parámetros) de dimensión k y U es el vector columna de residuales (o términos de perturbación) de dimensión n y que satisface todas las propiedades ya conocidas. (Véase (8)). Se supone además que X y Y han sido escaladas de tal forma que — los productos X'X y X'Y sean matrices de coeficientes de correlación simple.

1/ El nombre en inglés de este método es "Ridge Regression Analysis" y fue utilizado formalmente por vez primera por A.E. Hoerl en su investigación: Application of Ridge Analysis to Regression Problem. Chem. Engr. Progress, 58, 1962, pp. 54-59, por su semejanza con el análisis efectuado con superficies de respuesta de segundo-orden en muchas variables en un trabajo anterior del referido autor titulado: Optimum Solutions of Many Variables. Chem. Eng. Progress, 55, 1959, pp. 69-78, en donde aplicó esta técnica. Sin embargo, la presente discusión se basa fundamentalmente en el trabajo conjunto de dicho autor con R.W. Kennard, titulado: Ridge Regression, Biased Estimation for Nonorthogonal Problems. Technometrics, 12, 1970, pp. 69-82.

Sabemos de (10) que el estimador de mínimos cuadrados del modelo general es de la forma:

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

Entonces, según Hoerl y Kennard 1/ se puede demostrar que:

$$MSE(\hat{\beta}) = E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right]^2 = \sigma_u^2 \sum_{j=1}^{K} \lambda_j^{-1} \qquad (138)$$

donde $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_K$ son las raíces características de la matríz (X'X). El miembro izquierdo de (138) se le conoce como "error cuadrático medio total" de $\hat{\beta}$ (respecto de β)(+). Y representa una medida compuesta de la desviación(o distancia)al cuadrado de los estimadores de sus valores paramétricos verdaderos. En el contexto del análisis de regresión por mínimos cuadrados ordinarios, este término equivale a la varianza del estimador $\hat{\beta}$ (i.e. Var ($\hat{\beta}$)).

Por otra parte, como ya hemos discutido anteriormente, la multicolinealidad es sinónimo de raíces características pequeñas. Se infiere de la ecuación (138), que cuando una o más de las λ's es pequeña, el error cuadrático medio total de $\hat{\beta}$ es grande, implicando con ello, imprecisión en el proceso de estimación por mínimos cuadrados. <u>El método de regresión de cordillera es un intento de construir un estimador alternativo de $\hat{\beta}$ con un error cuadrático medio total más pequeño que la varianza de $\hat{\beta}$.</u> 2/

1/ Véase, Hoerl y Kennard. Opus cit.

(+) MSE es la abreviación en inglés de 'Mean Square Error'. Utilizamos esta notación para facilitar la consulta de fuentes originales sugeridas al lector interesado. A propósito, el error cuadrático medio (MSE) de un estimador $\hat{\alpha}$ del parámetro α se puede definir también como: $MSE(\hat{\alpha}) = Var(\hat{\alpha}) + [E(\hat{\alpha}) - \alpha]^2$. Nótese que cuando $\hat{\alpha}$ es insesgado, es decir, $E(\hat{\alpha}) = \alpha$, entonces $MSE(\hat{\alpha}) = Var(\hat{\alpha})$. Asimismo, si $\hat{\alpha}$ es un estimador "eficiente" de α, es de esperarse que para "n" grande, se cumpla que $MSE(\hat{\alpha}) \leq MSE(\tilde{\alpha})$ donde $\tilde{\alpha}$ es cualquier otro estimador "consistente" de α.

Véase, Kelejian, H.H. y W.E. Oates, <u>Introduction to Econometrics.</u> Second Edition, Harper and Row, 1981, Opus cit. p.208

2/ Véase, Hoerl y Kennard. Opus cit.

Hoerl y Kennard 1/ proponen la construcción de una clase de estimadores "indizados" respecto a un parámetro "p" ($p > 0$).

El "estimador de cordillera" de $\hat{\beta}$, para un valor dado de p, se define como:

$$\hat{\beta}(p) = (X'X + pI)^{-1} X'Y = (X'X + pI)^{-1} X'X \hat{\beta} \qquad (139)$$

el valor esperado de $\hat{\beta}(p)$ es:

$$E\left[\hat{\beta}(p)\right] = (X'X + pI)^{-1} X'X \beta \qquad (140)$$

la matríz de variancia-covariancia de $\hat{\beta}(p)$ puede expresarse -- como:

$$Var\left[\hat{\beta}(p)\right] = (X'X + pI)^{-1} X'X (X'X + pI)^{-1} \sigma_u^2 \qquad (141)$$

la variación no-explicada (o suma de los cuadrados de los errores) del modelo (3) asociado al estimador $\hat{\beta}(p)$, se define como:

$$(Y - X\hat{\beta}(p))'(Y - X\hat{\beta}(p)) = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + \qquad (142)$$
$$+ (\hat{\beta}(p) - \hat{\beta})' X'X (\hat{\beta}(p) - \hat{\beta})$$

y el error cuadrático medio total correspondiente, según estos -- autores, se puede escribir como:

$$E\left[(\hat{\beta}(p) - \beta)'(\hat{\beta}(p) - \beta)\right] = \sigma_u^2 \text{ traza}\left[(X'X + pI)^{-1} X'X (X'X + pI)^{-1}\right] + \qquad (143)$$
$$+ p^2 \beta'(X'X + pI)^{-2} \beta$$
$$= \sigma_u^2 \sum_{i=1}^{k} \lambda_i (\lambda_i + p)^{-2} + p^2 \beta'(X'X + pI)^{-2} \beta$$

Nótese que el primer término del miembro derecho de la expresión anterior es la suma de las variancias de las componentes de $\hat{\beta}(p)$ (variancia total) y el segundo, representa el cuadrado del sesgo correspondiente.

1/ Véase, Hoerl y Kennard. Opus cit.

Entonces para $p > 0$, $\hat{\beta}(p)$ es sesgado y este sesgo aumenta conforme crece el valor de "p". Por otra parte, la varianza total de $\hat{\beta}(p)$ es una función decreciente de "p". La idea de la regresión de cordillera es encontrar un valor "p" tal que la reducción en la variancia total no sea rebasada por el incremento del sesgo. Esto es, Hoerl y Kennard 1/ demuestran que existe un valor de $p > 0$, tal que:

$$\boxed{MSE\left[\hat{\beta}(p)\right] = E\left[(\hat{\beta}(p) - \beta)'(\hat{\beta}(p) - \beta)\right] < E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right] = Var(\hat{\beta})} \qquad (144)$$

sugieren que un valor "p" puede seleccionarse observando la "traza de cordillera" y algunos estadísticos sumarios complementarios de $\hat{\beta}(p)$.

La "traza de cordillera" como se observa en la Figura 4.4.1 (hojas más adelante) es una gráfica simultánea del vector $\beta(p)$ contra el parámetro "p", con "p" localizado en el intervalo $[0,1]$. Si la multicolinealdidad es peligrosa, los "estimadores de cordillera" ($\hat{\beta}i(p)$) variarán dramáticamente conforme "p" aumente lentamente su valor desde cero. El vector $\hat{\beta}(p)$ eventualmente se estabilizará.

El comportamiento de $\hat{\beta}(p)$ como una función de "p" se puede observar fácilmente de la traza de cordillera. El valor de "p" seleccionado es el mínimo valor que hace que $\hat{\beta}(p)$ sea estable. Además, en este valor mínimo de "p", la variancia no -explicada (suma de los cuadrados de los errores) estará cerca de su valor mínimo y la

---

1/ Véase, Hoerl y Kennard. Opus cit.

matríz de variancia-covariancia de $\hat{\beta}(p)$ tendrá la apariencia de un sistema ortogonal. 1/

Los estimadores de cordillera han sido generalizados de varias maneras. Algunas veces se les conoce genéricamente como "estimadores de compactamiento" (o "achicamiento") porque la regresión de cordillera tiende a "compactar" a los estimadores de los --- coeficientes de regresión hacia cero.

Para visualizar una posible generalización de estos conceptos, considérese el modelo de regresión expresado en términos de sus componentes principales, definido en (111), esto es:

$$Y = W\alpha + u$$

donde W es la matriz (nxk) y α es el vector-columna (de dimensión k) definidos en (112), tales que:

$$W = XC \quad , \quad \alpha = C'\beta$$

y C es la matríz no-singular de orden k definida en (108) tal que:

$$C'X'XC = \Lambda \quad , \quad C'C = CC' = 1$$

y donde la matríz Λ se define como en (109), esto es:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \quad ; \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k .$$

1/ Véase, Hoerl y Kennard. Opus cit.

Entonces según Hoerl y Kennard 1/ el error cuadrático medio total de $\hat{\beta}(p)$ se puede escribir como:

$$E\left[(\hat{\beta}(p)-\beta)'(\hat{\beta}(p)-\beta)\right] = \sigma_u^2 \sum \frac{\lambda_i}{(\lambda_i+p)^2} + \sum \frac{p^2 \alpha_i^2}{(\lambda_i+p)^2} \qquad (145)$$

donde α es un vector-columna tal que $\alpha = \{\alpha 1, \alpha 2, \alpha 3, \ldots, \alpha k\}$

En lugar de tomar un sólo valor para "p", podríamos considerar varios valores diferentes para "p", digamos p1, p2, p3,...,pk. Consideremos parámetros de "compactación" separados (i.e. factores de achicamiento) para cada uno de los coeficientes de regresión. Entonces el escalar "p" se convierte en un vector " $\vec{p}$. ". El error cuadrático medio total definido en (145) se transformaría ahora en : 2/

$$E\left[(\hat{\beta}(\vec{p})-\beta)'(\hat{\beta}(\vec{p})-\beta)\right] = \sigma_u^2 \sum \frac{\lambda_i}{(\lambda_i+p_i)^2} + \sum \frac{p_i \alpha_i^2}{(\lambda_i+p_i)^2} \qquad (146)$$

La expresión anterior de conformidad con Hoerl y Kennard 3/ se minimiza haciendo:

$$p_i = \frac{\sigma_u^2}{\alpha_i^2} \qquad (147)$$

y estos autores sugieren a este respecto, un procedimiento itera--

---

1/ Véase, Hoerl y Kennard. Opus cit.

2/ Ibid.

tivo para obtener los estimadores de compactación:

En el primer paso, se procede a estimar " pi " usando los estimadores de mínimos cuadrados ordinarios de σ^2 y αi.

En el segundo, se calcula un nuevo valor ddl vector $\hat{\alpha}(\vec{p})$ definido como:

$$\hat{\alpha}(\vec{p}) \equiv (W'W + \vec{P})^{-1} W'Y \qquad (148)$$

usando para ello los valores p1, p2,...,pk obtenidos del paso anterior. Este proceso se repite hasta obtener cambios insignificantes en las componentes del vector $\hat{\alpha}(\vec{p})$. Entonces, utilizando la versión "vectorizada" de la expresión (112), se puede estimar $\hat{\beta}(\vec{p})$ (esto es, el vector de estimadores de compactación) como sigue:

$$\hat{\beta}(\vec{p}) = C\,\hat{\alpha}(\vec{p}) \qquad (149)$$

Los dos tipos de estimadores de cordillera definidos anteriormente (uno a partir de un solo valor de "p" y el otro a partir de varios valores de "p") así como otras técnicas **de estimación alternativas a la de mínimos cuadrados y relacionadas con este tema son discutidos en detalle por Dempster et al.** 1/
En dicha investigación se comparan y evalúan todas ellas mediante el Método de Monte Carlo. En general, se observa que la selección del mejor método de estimación para un problema particular

---

1/ Dempster, A.P. et al. A Simulation Study of Alternatives to Ordinary Least Squares. Research Report S-35, Departmnnt of Statistics, Harvard University, 1975.

depende del modelo y datos específicos que se trate. Dempster et al 1/ terminan por sugerir un tipo de análisis que podría utilizarse para identificar el mejor método de estimación para un conjunto dado de datos.

Para propósitos de la presente discusión adoptaremos la versión más simple del método de la regresión de cordillera, esto es, utilizando un solo parámetro de cordillera "p", seleccionado después de un exámen gráfico de la traza de cordillera. El atractivo de este enfoque, radica en que su representación gráfica permite observar los efectos de la multicolinealidad en los coeficientes estimados. 2/

En consecuencia, se desprende de la discusión teórica anterior que el análisis de regresión de cordillera proporciona un método alternativo de estimación (al de mínimos cuadrados ordinarios u OLS) que puede utilizarse ventajosamente cuando las variables explkcativas son "altamente no- ortogonales". Se ha visto que existen varias maneras de definir y calcular los estimadores de cordillera. Hemos decidido presentar el método asociado con la "traza de cordillera", por considerarla como un enfoque gráfico y concebirla más bien como una técnica exploratoria.

1/ Dempster, A.P. et al. A Simulation Study of Alternatives to Ordinary Least Squares. Opuscit.

2/ Para el lector interesado en ahondar más en el análisis de cordillera le recomendamos consultar la fuente que aparece enseguida y que compendia una lista muy completa de referencias sobre la materia:
Véase, Hocking, R. R. The Analysis and Selection of Variables in Linear Regression, Biometrics, 32, 1976, pp. 1-49

El análisis de regresión de cordillera mediante la "traza de cordillera" representa un enfoque unificado a los problemas de detección y estimación cuando se sospecha la existencia de multicolinealidad. Aunque los estimadores obtenidos bajo este análisis son sesgados, tienden a tener un error cuadrático medio (i.e. $MSE[\hat{\beta}(p)]$) menor que el de los estimadores OLS (i.e. $Var(\hat{\beta})$) como ya se destacó oportunamente (véase la expresión (144)).

Los estimadores de cordillera son estables en el sentido de que no se ven afectados por ligeras variaciones en los datos de estimación, debido a la propiedad (144) sobre sus errores cuadráticos medios, los valores de los estimadores de cordillera se espera estén más cerca de los valores reales de los parámetros βi's que los estimadores OLS. Asimismo, los pronósticos de la variable-respuesta correspondientes a valores de las variables explicativas, no incluidas en la muestra de datos utilizada para la estimación del modelo en cuestión , tienden a ser más exactos.

Como el caso del método de las componentes principales, los criterios para decidir cuándo utilizar el método de regresión de cordillera en lugar del método OLS, depende de los valores reales de los parámetros dn el modelo bajo estudio.

Los estimadores de cordillera de los parámtros βi's pueden obtenerse resolviendo un sistema de ecuaciones muy parecidas al de las "ecuaciones normales". En otras palabras, supóngase el modelo lineal general, en su forma estandarizada previamente definida en la la expresión(15) como:

$$y^{*} = \beta_1 x_1^{*} + \beta_2 x_2^{*} + \cdots + \beta_k x_k^{*} + u$$

Entonces el sistema de "ecuaciones normales" para determinar los coeficientes de cordillera de los parámetros del modelo (6) se define como :

$$\begin{aligned}
(1+p)\tilde{\beta}_1 + r_{12}\tilde{\beta}_2 + \cdots\cdots + r_{1k}\tilde{\beta}_k &= r_{1y} \\
r_{12}\tilde{\beta}_1 + (1+p)\tilde{\beta}_2 + \cdots\cdots + r_{2k}\tilde{\beta}_k &= r_{2y} \\
\cdots\cdots\cdots\cdots \\
r_{1k}\tilde{\beta}_1 + r_{2k}\tilde{\beta}_2 + \cdots\cdots + (1+p)\tilde{\beta}_k &= r_{ky}
\end{aligned} \qquad (150)$$

donde "rij" corresponde al coeficiente de correlación simple entre Xi y Xj y "riy" al coeficiente de correlación simple entre Xi y Y.

La solución del sistema (150) es el conjunto de "estimadores de cordillera" $\tilde{b}1$, $\tilde{b}2$, ..., $\tilde{b}k$. El parámetro esencial que distingue el análisis de regresión de cordillera del de mínimos cuadrados, es precisamente el parámetro "p". (Nótese que cuando p= 0 , los $\tilde{b}$'s se reducen a los estimadores OLS). El parámetro "p" se conoce como el "parámetro de sesgo", conforme "p" incrementa su valor a partir de cero, el sesgo de los estimadores aumenta. Si "p" sigue aumentando sin acotamiento, los estimadores referidos tienden a cero.

Se ha destacado anteriormente que existe un valor positivo de "p" para el cual todos los estimadores de cordillera se estabilizan con respecto a pequeños cambios en los datos de estimación.

En la práctica, se escoge un valor de "p", calculando los estimadores $\tilde{b}1$, $\tilde{b}2$,..., $\tilde{b}k$ para un rango de valores "p", dentro del intervalo $[0,1]$ y graficando los resultados correspondientes contra el parámetro "p". A continuación se sugiere una metodología para seleccionar el valor "óptimo" de "p", a través de un ejemplo concreto. Pero antes, conviene hacer algunas consideraciones importantes.

Existen dos métodos para detectar la multicolinealidad que están indiscutiblemente asociados con la regresión de cordillera. El primero, se relaciona con el efecto de la multicolinealidad en el error entre los estimadores OLS y los valores reales de los parámetros β's. El segundo, se refiere a la inestabilidad básica que los estimadores OLS manifiestan como resultado de variaciones pequeñas en los datos de estimación.

El primer método de detección utiliza el concepto de "Factor de Inflación de la Variancia" (FIV)(+) que se define líneas abajo. La precisión de un estimador OLS se mide por su variancia que es proporcional a σ_u^2, esto es, a la variancia del término de perturbación o estocástico del modelo de regresión. La constante de proporcionalidad correspondiente resulta ser precisamente dicho factor. En consecuencia, existe un FIV para cada estimador OLS. El FIV para $\tilde{b}i$ se define como:

$$\boxed{FIVi = \frac{1}{(1 - R_i^2)} \quad , \; 0 \leq Ri^2 \leq 1} \qquad (151)$$

donde Ri^2 es el cuadrado del coeficiente de correlación múltiple de la variable Xi respecto de las demás variables explicativas.

(+) Las siglas en inglés de este factor son 'VIF' (Variance Inflation Factor).

Conforme Ri^2 tiende a 1, indicando la presencia de multicolinealidad en el modelo, el FIV de la variable Xi afectada tiende a infinito. Como una regla empírica se acepta en general que:

$$\boxed{\text{Si} \quad FIVi \gtrapprox 0 \Rightarrow \text{La multicolinealidad es peligrosa}} \qquad (152)$$

Si la condición anterior se cumple para alguna "i", entonces la presencia de multicolinealidad en el modelo puede estar ocasionando problemas en la estimación paramétrica. El concepto de "FIV" puede utilizarse también para obtener una expresión para el cuadrado de la distancia esperada de los estimadores OLS de sus valores verdaderos. Esta distancia es otra medida de precisión de los estimadores OLS. Mientras más pequeña sea, más precisos serán dichos estimadores. Si denotamos a esta distancia al cuadrado como L^2, entonces podemos escribir:

$$L^2 = \sigma_u^2 \sum_{i=1}^{K} FIVi \qquad (153)$$

O alternativamente como:

$$L^2 = \sigma_u^2(FIV1 + FIV2 + \ldots\ldots + FIVk) \qquad (154)$$

Obsérvese que si todas las variables explicativas del modelo lineal general fuesen ortogonales, entonces todos los FIV's serían iguales a la unidad y la expresión anterior se transformaría en:

$$L^2 = \sigma_u^2(k) = k \qquad (155)$$

En el otro extremo, si todas las variables fuesen totalmente dependientes, entonces la suma de los FIV's sería un número M, tan grande como se quisiera, de tal suerte que L^2 sería tal que:

$$L^2 = \sigma_u^2(M) = M\,\sigma_u^2 \doteq \infty \qquad (156)$$

Y por consiguiente, L^2 oscila en el intervalo:

$$k\sigma_u^2 \leq L^2 < \infty \qquad (157)$$

Con base en estos resultados, se puede entonces definir la siguiente razón:

$$\boxed{R_L = \frac{\sigma_u^2 \sum FIV_i}{k\sigma_u^2} = \frac{\sum_{i=1}^{k} FIV_i}{k}} \qquad (158)$$

que mide el error cuadrado de los estimadores OLS en relación al tamaño del error, en caso de que todas las variables del modelo original fuesen ortogonales. Por consiguiente, R_L puede utilizarse también como un índice de detección de la multicolinealidad. Nótese de la definición de R_L que esta razón se encuentra en el intervalo :

$$1 \leq R_L \leq \infty \qquad (159)$$

Por ejemplo, si usamos los datos del modelo de la "economía francesa" definido en (67) , se puede hacer ver que los FIV's para los coeficientes de cordillera de las variables DOPROD, STOCK y CONSUM son 185.99, 1.02 y 186.10, respectivamente. R_L resulta ser de 124.37. Es decir, el error al cuadrado de los estimadores OLS es 124 veces mayor que si dichas variables explicativas fuesen ortogonales. La magnitud de la razón R_L señala claramente la presencia de multicolinealidad.

El uso de R_L y los FIV's es posiblemente mucho más de lo que se requeriría para detectar la multicolinealidad en el modelo antes mencionado. Recuérdese que el coeficiente de correlación simple entre las variables DOPROD y CONSUM era de .997. Sin embargo, este ejemplo ha sido bueno para propósitos ilustrativos. En el próximo capítulo veremos una aplicación más compleja del mismo.

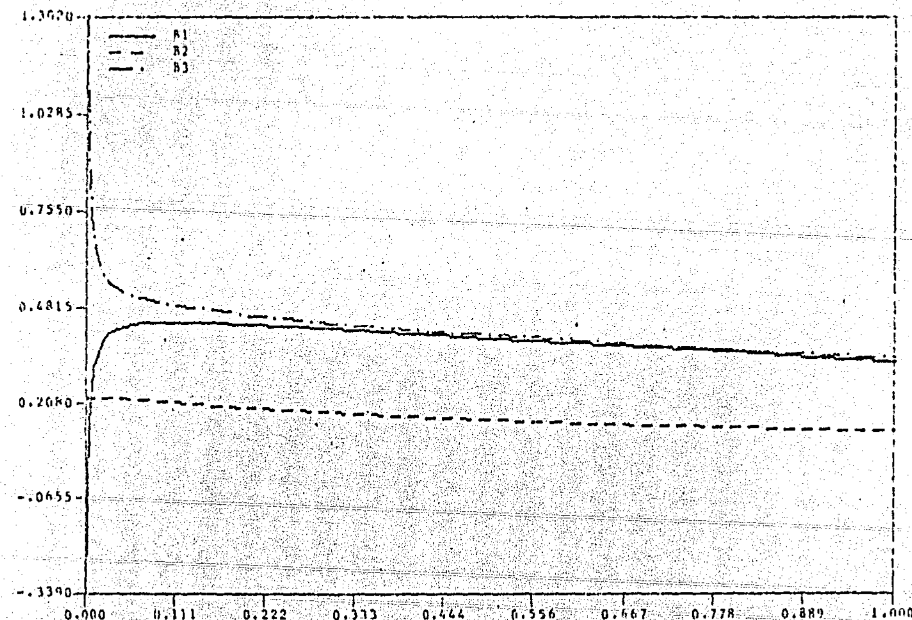
El segundo método para detectar la multicolinealidad basada en el análisis de regresión de cordillera tiene que ver con la inestabilidad generada en los estimadores CLS como producto de cambios o variaciones pequeñas en los datos de estimación. Esta inestabilidad puede observarse en la "traza de cordillera".

Ya hemos anotado que la traza de cordillera es una gráfica simultánea de los k coeficientes de regresión contra el parámetro "p". Hemos señalado que "p" es el "parámetro de sesgo".

Conviene destacar que para valores pequeños de p = .001, .002, etc., los estimadores de cordillera correspondientes pueden verse como producto de un conjunto de datos ligeramente alterado. En la próxima hoja, se presenta la Figura 4.4.1 que muestra la traza de cordillera para el modelo (67). (Esta gráfica fue construida a partir de los datos de la Tabla 4.4.1 que le sucede y que contiene los estimadores de cordillera para 29 valores de "p" que van del 0 al 1). Típicamente, los valores de "p" que se escogen o seleccionan se encuentran concentrados cerca de la parta más baja de dicho rango. Si los coeficientes estimados muestran grandes fluctuaciones para pequeños valores de "p", entonces existe inestabilidad, y por consiguiente, también muy probablemente, multicolinealidad peligrosa en el modelo.

F I G U R A 4.4.1

TRAZA DE CORDILLERA PARA EL MODELO (67)



Fuente; Chatterjee,S. y B. Price.Opus Cit. p.184

T A B L A 4.4.1

ESTIMADORES DE CORDILLERA PARA EL MODELO (67)

Hilera	P	B1	B2	B3
001	0.000	-.339	0.213	1.302
002	0.001	-.117	0.215	1.080
003	0.002	0.010	0.216	0.952
004	0.003	0.092	0.217	0.879
005	0.004	0.150	0.217	0.811
006	0.005	0.193	0.217	0.768
007	0.006	0.225	0.217	0.735
008	0.007	0.251	0.217	0.709
009	0.008	0.272	0.217	0.687
010	0.009	0.296	0.217	0.669
011	0.010	0.344	0.217	0.654
012	0.020	0.379	0.217	0.575
013	0.030	0.406	0.214	0.543
014	0.040	0.420	0.213	0.525
015	0.050	0.427	0.211	0.513
016	0.060	0.432	0.209	0.504
017	0.070	0.434	0.207	0.497
018	0.080	0.436	0.206	0.491
019	0.090	0.436	0.204	0.486
020	0.100	0.436	0.202	0.481
021	0.200	0.428	0.186	0.450
022	0.300	0.411	0.173	0.427
023	0.400	0.396	0.161	0.408
024	0.500	0.381	0.151	0.391
025	0.600	0.367	0.142	0.376
026	0.700	0.354	0.135	0.361
027	0.800	0.342	0.128	0.342
028	0.900	0.330	0.121	0.336
029	1.000	0.319	0.115	0.325

Fuente: Estimaciones a partir de los datos de la Tabla 3.4.1

Lo que es evidente de la traza de cordillera, a juzgar por los valores de la Tabla 4.4.1, es que los estimadores de los parámetros $\tilde{\beta}1$ y $\tilde{\beta}3$ son muy inestables para valores pequeños de "p". Se observa, por ejemplo, que el estimador de $\tilde{\beta}1$ cambia rápidamente de un inaceptable valor de -.3394 a un valor estable de alrededor de .42. Asimismo, el estimador de $\tilde{\beta}3$ se compacta de 1.3028 hacia un valor estable de aproximadamente .52 . En contraste, el coeficiente de la variable STOCK, $\tilde{\beta}2$, no se vé afectado por la multicolinealidad y permanece estable a lo largo de todo el rango en alrededor de .21.

El próximo paso en el análisis de regresión es seleccionar un valor de "p" y los correspondientes estimadores de los coeficiantes de -- regresión.

A continuación se proporciona una serie de lineamientos generales para seleccionar el valor de "p", sugeridos por Hoerl y Kennard.[1/]

1. Para un determinado valor de "p", el sistema se estabilizará y se comportará como un sistema ortogonal.

2. Los coeficientes de regresión no manifestarán valores absolutos irrazonables dn relación a factores para los que representan -- tasas de cambio.

3. Los coeficientes con signos 'impropios' en p =0 , tenderán a re-- cuperar su signo 'correcto'.

1/ Véase, Hoerl y Kennard.Opus cit.

4. La variación no-explicada (o suma de los cuadrados de los residuos) no observará un valor irrazonablemente grande. No será grande respecto al mínimo valor de la variación explicada, ni tampoco (grande) en relación a una variancia considerada "razonable" para el proceso que generó la información para estimación.

Como ya hemos destacado de la Tabla 4.4.1 para los datos del modelo de la "economía francesa" (133) que estamos discutiendo, la traza de cordillera, parece estabilizarse para un valor de "p" ligeramente mayor que .01.

Por otra parte, si estimamos los valores de los FIV's para $\tilde{\beta}_1$, $\tilde{\beta}_2$ y $\tilde{\beta}_3$ notamos que éstos se proximan a la unidad, como se observa de la Tabla 4.4.2 (en la próxima hoja), es decir, presentan un valor típico de un sistema de variables ortogonales, para un valor de p = .03 y .04.

Entonces seleccionamos p =.04 como la solución. Es deseable escoger el valor mínimo de "p" para el cual se obtiene la estabilidad, puesto que el tamaño de "p" está directamente ligado a la magnitud del sesgo en el estimador correspondiente.

Para este valor, k = .04 , el improcedente signo negativo del estimador de $\tilde{\beta}_1$ desaparece y dicho coeficiente se estabiliza en alrededor de .42. La variación no-explicada se incrementa solamente de .0081 (para p =0) hasta .0117, en tanto R^2 pasa de .992 a .988. Por tanto, la solución de p = .04 parece satisfactoria desde todos

T A B L A 4.4.2

FACTORES DE INFLACION DE LA VARIANCIA (FIV's) PARA LOS COEFICIENTES DE CORDILLERA DEL MODELO (67).

Hilera	P	B1	B2	B3
* 1 *	0.000	185.941	1.019	186.059
* 2 *	0.001	98.972	1.008	99.030
* 3 *	0.002	61.369	1.003	61.405
* 4 *	0.003	41.779	0.999	41.804
* 5 *	0.004	30.295	0.996	30.313
* 6 *	0.005	22.955	0.993	23.002
* 7 *	0.006	18.058	0.991	18.068
* 8 *	0.007	14.572	0.988	14.581
* 9 *	0.008	12.015	0.986	12.025
* 10 *	0.009	10.090	0.984	10.096
* 11 *	0.010	8.601	0.982	8.605
* 12 *	0.020	2.559	0.962	2.560
* 13 *	0.030	1.592	0.943	1.503
* 14 *	0.040	0.979	0.925	0.979
* 15 *	0.050	0.723	0.908	0.723
* 16 *	0.060	0.579	0.891	0.578
* 17 *	0.070	0.489	0.874	0.488
* 18 *	0.080	0.429	0.858	0.428
* 19 *	0.090	0.386	0.842	0.386
* 20 *	0.100	0.355	0.827	0.355
* 21 *	0.200	0.240	0.695	0.240
* 22 *	0.300	0.204	0.592	0.204
* 23 *	0.400	0.192	0.510	0.182
* 24 *	0.500	0.166	0.444	0.165
* 25 *	0.600	0.152	0.391	0.152
* 26 *	0.700	0.140	0.346	0.140
* 27 *	0.800	0.130	0.309	0.130
* 28 *	0.900	0.121	0.277	0.121
* 29 *	1.000	0.113	0.250	0.113

Fuente: Estimaciones a partir de los datos de la Tabla 3.4.1

T A B L A 4.4.3

ESTIMADORES DE CORDILLERA PARA EL MODELO DE LA ECONOMIA FRANCESA (67)(+)

VARIABLE	ESTIMADORES OLS (K = 0)		ESTIMADORES DE CORDILLERA (K = .04)	
	$\tilde{b}$	b	$\tilde{b}$	b
DOPROD	-.3394	-.0514	.4196	.0635
STOCK	.2130	.5869	.2127	.5859
CONSUM	1.3028	.2868	.5249	.1156
CONSTANTE	0	-10.1300	0	-8.5537
R^2:	.992		.988	

(+) Esta tabla presenta los valores de los coeficientes estimados $\tilde{b}i$ para el modelo (133) en su versión estandarizada, así como los respectivos valores de los estimadores bi correspondientes al modelo original. (Recuérdese que bi se puede derivar a partir del valor de bi como: $bi = \tilde{b}i(Sy/Si)$, donde Sy y Si son las desviaciones estándard de la variable-respuesta y la i-ésima variable explicativa (Xi), respectivamente, del modelo original.

Fuente: Estimaciones a partir de los datos de la Tabla 3.4.1

los angulos. Los resultados se resumen en la Tabla 4.4.3 (que aparece en la próxima hoja)

El modelo resultante para p = .04 se puede expresar en consecuencia como ;

$$\widehat{\text{IMPORT}} = -8.5537 + .0635\ \text{DOPROD} + .5859\ \text{STOCK} + .1156\ \text{CONSUM} \qquad (160)$$

La ecuación anterior ofrece una representación más plausible de la relación entre las variables explicativas y la variable-respuesta. Además, como se observará posteriormente, no difiere sustancialmente de la que se obtendría utilizando las dos primeras componentes principales. Aunque, claro está , los dos métodos parezcan diferentes.

Para cerrar este capítulo conviene destacar que los dos últimos métodos de estimación discutidos - el de regresión de cordillera y el de componentes principales - permiten obtener información adicional sobre los datos bajo estudio.

Hemos visto que las raíces características de la matríz de correlación de las variables explicativas del modelo original desempeñan un papel muy importante para detectar la multicolinealidad y analizar sus efectos. Y también que los estimadores obtenidos mediante estos métodos, aunque sesgados, pueden ser más precisos que los correspondientes estimadores OLS en términos de sus errores cuadráticos medios.

Sin embargo, es imposible evaluar con precisión la ganancia para un problema específico al aplicar cualquiera de dichos métodos, pues una comparación de los mismos respecto de la regresión de mínimos cuadrados (OLS) requiere del conocimiento de los valores verdaderos de los parámetros. No obstante, cuando se sospecha la existencia de multicolinealidad severa en un modelo, vale la pena estimar al menos un conjunto adicional de estimadores por cualquiera de estos métodos - aparte de los estimadores minimo-cuadráticos - ya que dichos estimadores podrían sugerir una interpretación distinta de la información que no se había contemplado anteriormente.

V. CORRECCION DE LA MULTICOLINEALIDAD: ROMPIMIENTO DEL CANDADO DE LA MULTICOLINEALIDAD.

Existen diversas técnicas para eliminar la multicolinealidad de un modelo de regresión multiple (o econométrico) donde al menos dos de sus variables explicativas están afectadas. Sin embargo, como veremos, dichas técnicas representan en general un costo para la estimación paramétrica.

Procederemos entonces a describir enseguida algunas de las técnicas más ampliamente conocidas y utilizadas actualmente en la literatura para dicho propósito.

5.1 METODO DE ELIMINACION DE VARIABLES.

Este método lo denominaremos de esta manera, en virtud de no existir en la literatura, un nombre particular o específico para él y también por el hecho de que consiste precisamente en eliminar del modelo afectado, aquella o aquellas variables que poseen un nivel peligroso de multicolinealidad.

Pero como ya se ha destacado anteriormente, este técnica en especial para atacar la multicolinealidad - que entre paréntesis es la más socorrida en el análisis de regresión por su facilidad de aplicación - tiene el inconveniente de producir cierta pérdida del poder explicativo del modelo bajo estudio, así como sesgo en los estimadores de los parámetros de las variables no multicolineales que permanecen en dicho modelo, después de eliminar del mismo, las variables colineales.

Por consiguiente, el método de "eliminación de variables" puede considerarse justifidable siempre y cuando tratemos de obtener estimadores más confiables de parámetros particulares o de probar la significancia estadística de variables específicas, como ya se mencionó anteriormente (véase la expresión (56)), pero no es recomendable cuando se utiliza para fines de predicción por ejemplo, por las razones expuestas en el párrafo anterior y -- también las mencionadas en el Apartado 3.4.(+)

5.2 PROCEDIMIENTO DE CORRECCION DE LA MULTICOLINEALIDAD MEDIANTE ESTIMADORES EXOGENOS.

De acuerdo con Farrar y Glauber 1/ al tratar con el problema de la multicolinealidad, cuando no disponemos de información adicional, pero tenemos en cambio, estimadores de algunos de los parámetros (calculados de fuentes ajenas o exógenas, digamos) podemos entonces proceder a "reestimar" nuestro modelo, usando estos estimadores para incorporarlos en nuestro proceso de estimación, con

---

(+) Recuérdese que cuando se elimina arbitrariamente una variable colineal de un modelo, puede acontecer que la subespecificación resultante altere seriamente la estructura del modelo original haciendo por tanto muy poco confiables los valores pronosticados obtenidos a partir del mismo, pues una de las condiciones básicas de una buena predicción, es precisamente que la permanencia de la relación estructural (multicolineal o nó) existente en el pasado entre las variables explicativas, siga siendo válida también en el futuro.

1/ Véase, Farrar y Glauber. Opus cit.

el propósito de estimar los parámetros restantes y entonces eliminar la multicolinealidad del modelo resultante.

Esto es precisamente lo que discutiremos en este apartado. Utilizaremos para ello, el enfoque de J.Johnston.1/

Supóngase que en nuestro modelo lineal general expresado en forma matricial:

$$Y = X\beta + U$$

tenemos únicamente estimadores de (k-r) de los k elementos del vector β . O Sea, se desconocen "r" estimadores de los k parámetros.

Reenumeremos enseguida las variables Xi's de tal suerte que los coeficientes estimados tomen el lugar de las últimas (k-r) variables de la matríz X, haciendo que los coeficientes por estimar ocupen los primeros "r" lugares en la **misma** . Ahora estamos en posibilidad de abordar la técnica en cuestión.

Primeramente, removeremos de la variable Y, la influencia lineal de las variables Xr1, Xr2, ..., Xrk (i.e. las variables explicativas correspondientes a los parámetros: $\beta_{r+1}, \beta_{r+2}, \ldots, \beta_{k-1}, \beta_k$). Si llamamos $b_{r+1}, b_{r+2}, \ldots, b_k$ a los respectivos estimadores de los parámetros $\beta_{r+1}, \beta_{r+2}, \ldots, \beta_k$ entonces podemos es--

---

1/ Véase, Johnston,J.Opus cit.

Se supone en esta discusión que $n > k > r$ donde"n"es el número de observaciones,"k"el número total de parámetros en el modelo lineal general y "r" el número de parámetros desconocidos o estimados exógenamente.

cribir;

$$y^* = y - b_{r+1} x_{r+1} - \ldots - b_k x_k \tag{161}$$

donde y^* es nuestra variable dependiente "corregida". Podemos, por tanto, reescribir la expresión anterior en términos matriciales como sigue;

$$Y^* = Y - X_{k-r} b_{k-r} \tag{162}$$

donde la matríz X_{k-r} se define como;

$$X_{k-r} = \left[X_{r+1} \; X_{r+2} \ldots X_k \right] \tag{163}$$

cuyo tamaño es de n x (k-r). Y el vector-columna b_{k-r} se define como;

$$b_{k-r} = \left\{ b_{r+1}, b_{r+2}, \ldots, b_k \right\} \tag{164}$$

Resulta útil para nuestra discusión subsiguiente, definir las -- siguientes matrices y vectores;

$$\begin{aligned}
X_r &= \left[X1, X2, \ldots, Xr \right] &&: \text{ de tamaño nx r} \\
X &= \left[X_r \; X_{k-r} \right] &&: \text{ de tamaño nxk} \\
\beta &= \left\{ B_r \; , \; B_{k-r} \right\} &&: \text{ vector-columna de tamaño k} \\
b &= \left\{ b_r \; , \; b_{k-r} \right\} &&: \text{ vector-columna de tamaño k}
\end{aligned} \tag{165}$$

Supóngase ahora que b_{k-r} es un estimador insesgado del parámetro B_{k-r} . Es decir:

$$E\,(b_{k-r}) = B_{k-r} \tag{166}$$

y que su matríz de variancia-covariancia se conoce y es de la forma:

$$Var(b_{k-r}) = E\,(b_{k-r} - B_{k-r})(b_{k-r} - B_{k-r})' \tag{167}$$

y que el estimador b_{k-r} es independiente del conjunto de datos que se usaron para estimar b_r .

Nuestro problema consiste entonces en estimar:

$$b_r = (X_r' X_r)^{-1} X_r' Y^* \tag{168}$$

sustituyendo (162) en (168) tenemos que:

$$b_r = (X_r' X_r)^{-1} X_r' (Y - X_{k-r} b_{k-r}) \tag{169}$$

Si desarrollamos (169) obtendremos:

$$\begin{aligned} b_r &= (X_r' X_r)^{-1} \left[X_r' Y - X_r' X_{k-r} b_{k-r} \right] \\ &= (X_r' X_r)^{-1} X_r' Y - (X_r' X_r)^{-1} X_r' X_{k-r} b_{k-r} \\ &= X_r^{-1} (X_r')^{-1} X_r' Y - X_r^{-1} (X_r')^{-1} X_r' X_{k-r} b_{k-r} \\ &= X_r^{-1} \cdot I \cdot Y - X_r^{-1} \cdot I \cdot X_{k-r} b_{k-r} \end{aligned} \tag{170}$$

Esto es:

$$b_r = X_r^{-1} Y - X_r^{-1} X_{k-r} b_{k-r} \tag{171}$$

pero de nuestra definición de Y dada en (3) sabemos que podemos escribir:

$$Y = X_r B_r + X_{k-r} B_{k-r} + U \tag{172}$$

ya que Y puede expresarse como:

$$Y = X \beta + U = \begin{bmatrix} X_r & X_{k-r} \end{bmatrix} \begin{bmatrix} B_r \\ B_{k-r} \end{bmatrix} \tag{173}$$

pues X y β se definen como en (165).

Si sustituimos (173) en (171) obtenemos:

$$\begin{aligned} b_r &= X_r^{-1}(X_r B_r + X_{k-r} B_{k-r} + U) - X_r^{-1} X_{k-r} b_{k-r} \\ &= X_r^{-1} X_r B_r + X_r^{-1} X_{k-r} B_{k-r} + X_r^{-1} U - X_r^{-1} X_{k-r} b_{k-r} \\ &= B_r - X_r^{-1} X_{k-r} (B_{k-r} - b_{k-r}) + X_r^{-1} U \\ &= B_r + X_r^{-1} X_{k-r} (b_{k-r} - B_{k-r}) + X_r^{-1} U \end{aligned} \tag{174}$$

de donde b_r se puede expresar ahora como:

$$b_r = B_r + X_r^{-1} X_{k-r} (b_{k-r} - B_{k-r}) + X_r^{-1} U \tag{175}$$

Pero como sabemos que $E(U) = 0$, $E(b_{k-r}) = B_{k-r}$ por hipótesis, entonces aplicando el operador E a la expresión anterior tendremos que:

$$\begin{aligned} E(b_r) &= E(B_r) + X_r^{-1} X_{k-r} \left[E(b_{k-r} - B_{k-r}) \right] + X_r^{-1} E(U) \\ &= B_r + X_r^{-1} X_{k-r} \left[E(b_{k-r} - B_{k-r}) \right] + X_r^{-1} E(U) \\ &= B_r + X_r^{-1} X_{k-r} (B_{k-r} - B_{k-r}) + X_r^{-1} E(U) \\ &= B_r + 0 + 0 \end{aligned} \tag{176}$$

Por consiguiente:

$$E\ (\ b_r\ ) = B_r \tag{177}$$

q.e.d.

y nuestro estimador b_r es un estimador insesgado de B_r.

Procederemos enseguida a determinar la matríz de variancia-covariancia para $\hat{\beta}$. Específicamente, demostraremos que:

$$\mathrm{Var}(b) = \mathrm{Var}\ (b_r) + A\ \mathrm{Var}(b_{k-r})\ A' \tag{178}$$

donde A es una matríz definida como:

$$A \equiv (X_r'\ X_r)^{-1}\ X_r'\ X_{k-r} \tag{179}$$

Por otra parte, de la suposición de independencia de los dos conjuntos de datos para estimar b_r y b_{k-r} respectivamente, tenemos que:

$$\mathrm{Var}(b_r\ ) = E\left\{(b_r - B_r)(\ b_r - B_r)'\right\} \tag{180}$$

De la expresión (174) tenemos que $(b_r - B_r)$ se puede expresar como:

$$\begin{aligned} b_r - B_r &= X_r^{-1}\ U + X_r^{-1}\ X_{k-r}\ (b_{k-r} - B_{k-r}) \\ &= (X_r'\ X_r)^{-1}\ X_r'\ U + (X_r'\ X_r)^{-1}\ X_r'\ X_{k-r}(b_{k-r} - B_{k-r}) \end{aligned} \tag{181}$$

de donde trasponiendo la expresión anterior tenemos que:

$$(b_r - B_r)' = (X_r'\ U)'(X_r'\ X_r)^{-1} + (b_{k-r} - B_{k-r})'\ X_{k-r}'\ X_r\ (X_r'\ X_r)^{-1} \tag{182}$$

multiplicando las expresiones (181) y (182):

$$
\begin{aligned}
(b_r - B_r)(b_r - B_r)' &= (X_r' X_r)^{-1} X_r U U' X_r (X_r' X_r)^{-1} + \\
&+ (X_r' X_r)^{-1} X_r X_{k-r} (b_{k-r} - B_{k-r}) U' X_r (X_r' X_r) + \\
&+ (X_r' X_r)^{-1} X_r' U (b_{k-r} - B_{k-r})' X_{k-r}' X_r (X_r' X_r)^{-1} + \\
&+ (X_r' X_r)^{-1} X_r' X_{k-r} (b_{k-r} - B_{k-r})(b_{k-r} - B_{k-r})' X_{k-r}' X_r (X_r' X_r)^{-1}
\end{aligned}
\qquad (183)
$$

Aplicando el operador E la expresión anterior notamos que:

$$
\begin{aligned}
E\left\{(b_r - B_r)(b_r - B_r)'\right\} &= (X_r' X_r)^{-1} X_r' E(U'U) X_r (X_r' X_r)^{-1} + \\
&+ (X_r' X_r)^{-1} X_r' X_{k-r} E\left\{(b_{k-r} - B_{k-r}) U'\right\} X_r (X_r' X_r)^{-1} + \\
&+ (X_r' X_r)^{-1} X_r' E\left\{U(b_{k-r} - B_{k-r})'\right\} X_{k-r}' X_r (X_r' X_r)^{-1} + \\
&+ (X_r' X_r)^{-1} X_r' X_{k-r} E\left\{(b_{k-r} - B_{k-r})(b_{k-r} - B_{k-r})'\right\} X_{k-r}' X_r (X_r' X_r)^{-1}
\end{aligned}
\qquad (184)
$$

$$
= \sigma_u^2 (X_r' X_r)^{-1} X_r' X_r (X_r' X_r)^{-1} + (X_r' X_r)^{-1} X_r' X_{k-r} \mathrm{Var}(b_{k-r}) X_{k-r}' X_r (X_r' X_r)^{-1}
$$

Si hacemos $A = (X_r' X_r)^{-1} X_r' X_{k-r}$ en la expresión anterior, entonces podemos concluir que:

$$
\mathrm{Var}(b) = \mathrm{Var}(b_r) + A\, \mathrm{Var}(b_{k-r})\, A' \qquad (185)
$$

Entonces se ha hecho ver que la nueva matríz de variancia-covariancia de b es la suma de la matríz de variancia-covariancia del estimador b_r más un término de "ajuste o corrección". Es evidente que la presencia de este término es necesaria debido a la incertidumbre asociada a los coeficientes utilizados en el "ajuste" de la variable dependiente Y.

Pero si observamos la expresión (184) nos percatamos que no podemos proceder a realizar la estimación de (178) si desconocemos el valor de σ_u^2. Podemos obtener un estimador de su valor utilizando la conocida expresión: (+)

$$\hat{\sigma}_u^2 = \frac{e'e}{n-k} \qquad (186)$$

pero donde "e" queda ahora definida como:

$$e = Y^* - X_r\, b_r = Y - X_r\, b_r - X_{k-r}\, b_{k-r} \qquad (187)$$

Nótese de paso que en la expresión (186) dividimos entre (n-k) y no entre (n-r) pues el modelo "ajustado" (162) depende en realidad de k parámetros estimados.

Es decir, el poder conocer el valor de $\hat{\sigma}_u^2$ definido en (186) nos permitirá entonces obtener la variancia de b_r definida en (185), "corregida" por el efecto de la inclusión de los (k-r) estimadores ajenos de los parámetros B_{r+1}, B_{r+2}, ..., B_k y utilizados para romper el "candado de multicolinealidad" en el modelo original, y que se espera naturalmente, sea menor que la variancia asociada al modelo "sin ajuste", para de esta manera, mejorar la precisión de los nuevos estimadores.

Cabe notar, no obstante, que esta técnica sólo puede aplicarse cuando se dispone o puede conseguir información adicional para estimar "exógenamente" a cierto número de parámetros de nuestro modelo objeto, como ya se ha destacado.

(+) Nótese que esta definición es equivalente a la proporcionada en la expresión (20-bis) en el Capítulo I.

Sin embargo, disponer de información adicional no es posible en muchas ocasiones y tenemos que conformarnos solamente con la muestra disponible. Pero inclusive cuando esto fuese posible, ninguna de las técnicas hasta ahora discutidas, incluyendo a la de Farrar y Glauber, nos da una indicación precisa de cómo romper el "candado de multicolinealidad" en el modelo. Esto es precisamente el propósito de nuestro próximo método a discutir, esto es, el sugerido por S.D. Silvey.1/

1/ Silvey, S.D. Multicollinearity and Imprecise Estimation. Journal of the Royal Statistical Society, Series B, Vol. 31, 1969, pp. 539-552.

Sin restarle los méritos y la importancia téorica y práctica del Método de la Dirección Optima de Silvey -objeto del próximo apartado- conviene resaltar (adelantándonos un poco a su discusión) que para poder cumplir con dicho cometido, dicho método requiere de información adicional para poder aplicarse, a diferencia de otros métodos como el de las Componentes Principales - cuyos pormenores serán asimismo analizados en este mismo capítulo -que también sirven al mismo propósito, pero sin necesidad de requerir de información fuera de la muestra original.

5.3 METODO DE LA DIRECCION OPTIMA DE SILVEY [1]

Bajo las suposiciones del modelo general regresión se puede demostrar la validez de la siguiente proposición.

Proposición: La matríz (X'X) es simétrica de orden k y es positiva -definida. Y por consiguiente, tiene k raíces características diferentes de cero. [2]

Sea $V = [V1, V2, \ldots, Vk]$ la matríz asociada a las raíces características de (X'X) donde Vi denota el vector-columna asociado a la raíz λi, o sea:

$$(X'X)\ Vi = \lambda i\ Vi \tag{188}$$

y también tenemos que la matríz V cumple con la condición:

$$V'V = VV' = I \tag{189}$$

pues es una matríz ortogonal.

Se puede demostrar que cualquier combinación lineal de los coe--

---

1/ Véase, Silvey, S.D. Opus cit.

2/ Para una demostración de esto:

Véase, Hadley, G. Linear Algebra. Addisson-Wesley, 1966, pp. 245-249 ficientes estimados ($\hat{B}i's$) del modelo lineal general:

$$Y = X\beta + U \tag{3}$$

Se puede escribir como una función lineal tal que: 1/

$$C'\hat{\beta} = C1\hat{\beta}1 + C2\hat{\beta}2 + \ldots + Ck\hat{\beta}k \tag{190}$$

donde C es un vector-columna de k elementos conocidos. Como C contiene k elementos se puede expresar como una combinación lineal de k vectores característicos ortogonales (Vi's). O sea:

$$C = V\alpha = V1\alpha_1 + V2\alpha_2 + \ldots\ldots + Vk\alpha_k \tag{191}$$

dondo α es un vector-columna tal que :

$$\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_k\} \tag{192}$$

Usando (191) se puede demostrar que la variancia de $C'\hat{B}$ se puede escribir como: 1/

$$\begin{aligned} \mathrm{Var}(C'\hat{B}) &= \sigma_u^2\, \alpha' V' (X'X)^{-1} V\alpha = \\ &= \sigma_u^2\, \alpha' \Lambda^{-1} \alpha \end{aligned} \tag{193}$$

donde Λ es la conocida matríz de raíces características definida previamente en (109), esto es:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_k \end{bmatrix} ; \quad \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k$$

1/ Véase, Silvey, S.D. Opus cit.

Nótese que la expresión (193) se puede escribir también como sigue;

$$Var(C'\hat{\beta}) = \sigma_u^2 (\alpha_1\ \alpha_2 \ldots \alpha_k) \begin{bmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 1/\lambda_k \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} \qquad (194)$$

$$= \sigma_u^2 \left(\frac{\alpha_1}{\lambda_1}\ \frac{\alpha_2}{\lambda_2} \cdots \frac{\alpha_k}{\lambda_k} \right) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix}$$

$$= \sigma_u^2 \left(\frac{\alpha_1^2}{\lambda_1} + \frac{\alpha_2^2}{\lambda_2} + \cdots + \frac{\alpha_k^2}{\lambda_k} \right)$$

O alternativamente como;

$$Var(C'\hat{\beta}) = \sigma_u^2 \sum \frac{\alpha_i^2}{\lambda_i} \qquad (195)$$

Entonces la variancia de $C'\hat{\beta}$ depende inversamente de los valores característicos de la matríz $(X'X)$. Esto es, mientras más pequeña sea λ_i mayor será su contribución a la variancia y por ende contribuirá mayormente a socabar la precisión de la función $C'B$ (como ya se destacó en el Apartado 4.3.)

Supóngase que estamos interesados en un estimador en particular, digamos en $\hat{\beta}_i$. Entonces el vector C definido en (191) tendrá un sólo "uno" en la i-ésima posición y "ceros" en los otros lugares y, por consiguiente, dicha expresión se puede expresar también como;

$$\alpha = V'C = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{ik} \end{bmatrix} \qquad (196)$$

donde $v_{i1}, v_{i2}, \ldots, v_{ik}$ son los elementos de la i-ésima hilera -- de la matríz V. Esto se puede entender mejor observando que la matríz V se puede expresar como:

$$V = \begin{bmatrix} v_{11} & v_{12} & \ldots & v_{i1} & \ldots & v_{1k} \\ v_{21} & v_{22} & \ldots & v_{i2} & \ldots & v_{2k} \\ v_{31} & v_{32} & \ldots & v_{i3} & \ldots & v_{3k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ v_{k1} & v_{k2} & \ldots & v_{ik} & \ldots & v_{kk} \end{bmatrix} \tag{197}$$

y por consiguiente, el producto V'C resulta ser:

$$V'C = \begin{bmatrix} v_{11} & v_{21} & \ldots & v_{k1} \\ v_{12} & v_{22} & \ldots & v_{k2} \\ \cdot & \cdot & \cdot & \cdot \\ v_{i1} & v_{i2} & \ldots & v_{ik} \\ \cdot & \cdot & \cdot & \cdot \\ v_{1k} & v_{2k} & \ldots & v_{kk} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ \vdots \\ v_{ik} \end{bmatrix} \tag{198}$$

y de (196) es evidente que:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{ik} \end{bmatrix} \tag{199}$$

de donde es claro que:

$$\alpha_i = v_{ij} \Rightarrow \alpha_i^2 = v_{ij}^2 , \quad \forall j = 1, 2, \ldots, k \tag{200}$$

Y por consiguiente de (193) es fácil inferir que la variancia de $\hat{\beta}i$ resulta ser:

$$\mathrm{Var}(\hat{\beta}i) = \sigma_u^2 \left(\frac{v_{i1}^2}{\lambda_1} + \frac{v_{i2}^2}{\lambda_2} + \cdots + \frac{v_{ik}^2}{\lambda_k} \right) \qquad (201)$$

S.D. Silvey 1/ afirma que si pudiésemos agregar más información a la muestra original contenida en la matríz X de observaciones de tal suerte de obtener una nueva matríz aumentada X^* tal que el producto $(X^{*'} X^*)$ tuviese raíces características (μi) que fuesen mayores que las correspondientes raíces (λi) de $(X'X)$, entonces los errores estándard de los estimadores de los coeficientes de regresión se reducirían. 2/

Es decir:

$$\text{Si} \qquad \mu i \geq \lambda i \qquad (202)$$

entonces los errores estándard de los estimadores de los coeficientes de regresión serán más pequeños. Esto es, la precisión de la función definida en (190) tenderá a aumentar, pues la variancia de cada $\hat{\beta}i$ definida en (201) tenderá a reducirse, para toda i=1,2,...,k, y por tanto, la variancia de $C'\hat{\beta}$ será más pequeña.

De lo anterior, resulta claro que la raíz λi puede incrementarse agregando información adicional en la dirección del vector característico Vi , dejando a los demás vectores y raíces caracte-

1/ Véase, Silvey, S.D. Opus cit.

2/ Ibid.

inalterados. Esto se demuestra enseguida:

Supóngase que a la matríz X se le aumenta otra hilera de la -- forma dV_i' , donde "d" es un escalar distinto de cero, para formar la matríz X^* tal que:

$$X^* = \begin{bmatrix} X \\ dV_i' \end{bmatrix} \tag{203}$$

entonces podremos escribir el producto $(X^{*\prime}X^*)$ como:

$$(X^{*\prime}\, X^*) = X'X + d^2\, V_i\, V_i' \tag{204}$$

de donde:

$$\begin{aligned} (X^{*\prime} X^*)\, V_i &= (X'X + d^2\, V_i . V_i')\, V_i \\ &= (X'XV_i + d^2\, V_i' V_i')\, V_i \\ &= X'X\, V_i + d^2 V_i\, V_i'\, V_i \\ &= (\lambda_i + d^2) V_i \end{aligned} \tag{205}$$

ya que $V_i'\, V_i = 1$ y $(X'X)V_i = \lambda_i V_i$ por hipótesis .Entonces V_i es un vector característico de la matríz $(X^{*\prime}\, X^*)$ correspondiente a la raíz característica $(\lambda_i + d^2)$.

Por otra parte, como:

$$\begin{aligned} (X^{*\prime}\, X^*) V_j &= (X'X + d^2 V_i\, V_i') V_j \\ &= X'X\, V_j + d^2 V_i\, V_i' V_j \\ &= \lambda_j\, V_j \end{aligned} \tag{206}$$

puesto que $V_i V_j = 0$ y $(X'X) V_j = \lambda_j V_j$ por hipótesis.

Por consiguiente, Vj ($j \neq i$) es un vector característico de $(X^{*\prime}X^{*})$ correspondiente a la raíz característica λ_j.

O en otras palabras, se ha demostrado que los vectores característicos de $(X^{*\prime}X^{*})$ son los mismos de $(X'X)$ y que todas sus raíces características son las mismas excepto que λ_i se incrementa en $(\lambda_i + d^2)$.

El resultado principal obtenido por Silvey 1/ en su investigación es en el sentido de que la dirección óptima de la nueva observación -- que mejorará la precisión de la función $C'\hat{\beta}$ es precisamente la del vector definido como:

$$(I + X'X)^{-1}C \qquad (207)$$

A reserva de mostrar con datos reales las bondades de este método frente a otros como el de componentes principales y el de regresión de cordillera (+), podemos observar de entrada que una de sus fuertes limitaciones radica en el hecho de que presupone la disposición de información adicional -además de la muestra original -para poder operar, y por consiguiente, eliminar la multicolinealidad de un modelo econométrico o de regresión.Puesto que como ya hemos destacado anteriormente,uno de los más grandes escollos en el análi-

---

1/ Véase, Silvey,S.D.Opus cit. para los detalles de la demostración correspondiente.

(+)Ejercicio que aunque evidentemente resultaría muy interesante e ilustrativo desde un punto de vista teórico y práctico, consideramos rebasa los alcances de la presente investigación.Motivo por el cual se deja como una inquietud personal para el lector interesado y curioso en profundizar en los detalles sobre este particular.

sis de regresión es precisamente ampliar la muestra bajo estu-- dio, por el costo en términos económicos, tiempo y esfuerzo que ello implica en general, haciendo prácticamente imposible o in-- viable su obtención.

Por consiguiente, y con la reserva señalada en el párrafo precedente , este método podrá considerarse cuando existá la certeza, y también la posibilidad económica,de allegarse de información adicional para eliminar la multicolinealidad en el modelo bajo estudio.

5.4 METODO DE LA RESTRICCION PARAMETRICA.

Este es un método ampliamente conocido aunque poco aplicado. El propósito de llamarlo de esta manera y de incluirlo aquí resultará evidente y además sugestivo para el lector no familiarizado después de entender la manera cómo trabaja, y sobre todo, comprender su operación y alcances como un instrumento "puente" o "antesala" para poder aplicar la técnica de las componentes principales, objeto de nuestro próximo apartado y con el auxilio del cual estudiaremos cómo se puede corregir el problema de la multicolinealidad.

Sea nuestro modelo de regresión general expresado en forma compacta y previamente definido en (6) como:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

nuestro objetivo fundamental de estimación consiste en determinar una solución única al sistema de ecuaciones simultáneas:1/

$$\begin{aligned}
1.\ & \sum x_1 y = \beta_1 \sum x_1^2 + \beta_2 \sum x_1 x_2 + \dots + \beta_k \sum x_1 x_k \\
2.\ & \sum x_2 y = \beta_1 \sum x_2 x_1 + \beta_2 \sum x_2^2 + \dots + \beta_k \sum x_2 x_k \\
& \dots\dots\dots\dots\dots\dots \\
k.\ & \sum x_k y = \beta_1 \sum x_k x_1 + \beta_2 \sum x_k x_2 + \dots + \beta_k \sum x_k^2
\end{aligned} \qquad (208)$$

1/ Utilizamos aquí el enfoque de K.H.Kelejian y E.O.Oates, únicamente expresando nuestras variables en forma compacta. Véase, Kelejian, K.H. y Oates, E.O. Introduction to Econometrics: Principles and Applications. Second Edition, Harper and Row Publishers, 1981, pp.132-136

Siempre y cuando se cumplan las condiciones siguientes:

$$\begin{aligned} &1.\ \sum (x_1 \hat{u}_t) = 0 \\ &2.\ \sum (x_2 \hat{u}_t) = 0 \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ &k.\ \sum (x_k \hat{u}_t) = 0 \end{aligned} \qquad (209)$$

Obsérvese que el sistema (208) consiste de k ecuaciones y k incógnitas. Si alguna de las variables explicativas del modelo (6) resulta una combinación lineal perfecta de las restantes, entonces este sistema será inconsistente, y la estimación de los pará metros $\beta 1, \beta 2, \ldots, \beta k$ resulta imposible, como ya hemos seña-- lado a lo largo de nuestras discusiones anteriores.1/

En otras palabras, el sistema anterior tendrá una solución única siempre y cuando el determinante de la matríz:

$$(X'X) = \begin{bmatrix} \sum x_1^2 & \sum x_1 x_2 & \ldots & \sum x_1 x_k \\ \sum x_2 x_1 & \sum x_2^2 & \ldots & \sum x_2 x_k \\ \cdot & \cdot & \cdot & \cdot \\ \sum x_k x_1 & \sum x_k x_2 & \ldots & \sum x_k^2 \end{bmatrix} \qquad (210)$$

sea diferente de cero. Y esto no podrá ocurrir si alguna de las

---

1/ Para una discusión sobre el problema de singularidad respecto al sistema de ecuaciones en econometría y su relación con el problema de multicolinealidad:
Véase, Klein, L.R. y Mitsuga Nakamura. Singularity in the Equation System of Econometrics: Some Aspects of the Problem of Multicollinearity. International Economic Review, Vol.3, No.3, September 1962, pp. 274-299.

variables explicativas es perfectamente dependiente con respecto, a cuando menos algunas de las otras variables explicativas, puesto que ello ocasionaría que éste se anulara.

Con todo lo anteriormente revisado como marco, pasaremos ahora a discutir el método que nos atañe.

Supóngase que la variable Xi es perfectamente dependiente de Xj. Entonces sabemos que no puede existir una solución única al sistema (208) puesto que tendríamos (k-1) ecuaciones con k incógnitas por estimar.

Sin embargo, si pudiéramos de alguna manera "establecer" una restricción en los parámetros respectivos de estas variables multicolineales, como por ejemplo, la restricción:

$$\beta i = w\beta j \quad ; \; w = \text{constante} \qquad (211)$$

entonces tendríamos k ecuaciones independientes (puesto que a las (k-1) ecuaciones normales le estamos agregando ahora la ecuación correspondiente a la restricción (211) y k parámetros por estimar. Luego entonces, es posible encontrar una solución única al sistema (208) siempre que $|(X'X)| \neq 0$.

No obstante, para poder aplicar correctamente este método, es necesario conocer, a través de un análisis a priori o de evidencia empírica independiente, que las restricciones impuestas a los coeficientes sean consistentes con la realidad.

Una de las limitaciones más fuertes con que se enfrentan los investigadores que tratan de aplicar este método para erradicar la multicolinealidad se debe a que por lo regular no se conoce a priori la forma de las restricciones que subyacen entre los coeficientes de las variables multicolineales. Dichas restricciones no pueden "inventarse" o generarse por "inspiración divina". Hacerlo de esta manera por tratar de eliminar el problema de multicolinealidad, nos llevaría únicamente a empeorar el panorama inicial al producir estimadores carentes de sentido para la adecuada toma de decisiones. 1/

Sin embargo, si este método lo aplicamos en combinación con la técnica de las componentes principales -anteriormente discutido- tenemos las siguientes ventajas:

1o. Podemos confirmar la existencia del nivel peligroso de multicolinealidad.

2o. Conocer las variables explicativas que generan la multicolinealidad.

3o. Identificar el tipo de relación lineal aproximado (o "patrón" de multicolinealidad) que guardan los coeficientes de las va-

---

1/ Para una discusión en torno a ciertos criterios para juzgar la adecuación de restricciones paramétricas en un modelo econométrico:
Véase, Toro-Vizcarrondo, C. y T.D. Wallace. A Test of Mean Square Error Criteria for Restrictions in Linear Regression. Journal of the American Statistical Association. Vol. 63, 1968, pp. 558-572.

riables multicolineales.

Y con este marco, podemos aplicar el método de la restricción paramétrica con eficacia y eficiencia.Por tal motivo, el pró-- ximo apartado se enfocará a la discusión detallada del método de las componentes principales del que nos serviremos para i- lustrar la manera de utilizar el método objeto de este aparta- do para corregir la multicolinealidad de un modelo de regresión o econométrico,con las ventajas antes mencionadas.

5.5 METODO DE LAS COMPONENTES PRINCIPALES.

Este método lo aplicaremos a uno de los casos discutidos con antelación. Concretamente, al modelo de la "economía francesa" definido en (67) como;

$$IMPORT = \beta o + \beta 1 \cdot DOPROD + \beta 2 \cdot STOCK + \beta 3 \cdot CONSUM + U$$

Recordaremos que desde un principio, podíamos esperar que los efectos marginales de DOPROD y CONSUM fuesen iguales. Es decir, en base a la teoría económica y antes de analizar los datos, podíamos suponer que;

$$\beta 1 = \beta 3 \qquad (212)$$

o equivalentemente;

$$\beta 1 - \beta 3 = 0 \qquad (213)$$

Entonces utilizando la información de la Tabla 3.4.1, podemos efectuar una regresión de la variable IMPORT contra STOCK y una nueva variable (NEWVAR), que creamos, definida como;

$$NEWVAR = DOPROD + CONSUM \qquad (214)$$

Nótese que la variable NEWVAR tiene significancia sólo como una manipulación técnica para encontrar un estimador del valor común de los coeficientes $\beta 1$ y $\beta 3$.(+)

(+) Se supone que todos los coeficientes del modelo (67) han sido previamente estimados, véase Tabla 3.4.2

Luego entonces, podemos definir un nuevo modelo en los siguientes términos:

$$\text{IMPORT} = \gamma_0 + \gamma_1 \cdot \text{STOCK} + \gamma_2 \text{NEWVAR} + U \qquad (215)$$

Los resultados del modelo anterior se presentan en la Tabla 5.5.1 (eb la próxima hoja). El coeficiente de correlación simple entre las variables STOCK y NEWVAR es de 0.299 y la matríz Λ para di-- cho modelo resulta ser:

$$\Lambda = \begin{bmatrix} 1.030 & 0 \\ 0 & .970 \end{bmatrix} \qquad (216)$$

la cual manifiesta claramente la ausencia de multicolinealidad en el modelo referido.

Por otra parte, las gráficas de los residuos estandarizados con-- tra el tiempo y contra los valores estimados de la variable-res-- puesta, como se desprende del análisis de las Figuras 5.5.1 y 5.5.2, respectivamente (y que aparecen enseguida de la Tabla 5.5.1), tampoco observan señales de problemas de especificación.

En consecuencia, el modelo estimado después de corregir el pro-- blema de multicolinealidad tiene la forma:

$$\widehat{\text{IMPORT}} = -9.007 + .086\ \text{DOPROD} + 0.612\ \text{STOCK} + 0.086\ \text{CONSUM} \qquad (217)$$

Nótese a este respecto qud también sería perfectamenteposible -- probar la restricción $\beta 1 = \beta 3$ como una hipótesis estadística. Aún cuando ésta haya sido impuesta en base a la teoría (económica) --

T A B L A 5.5.1

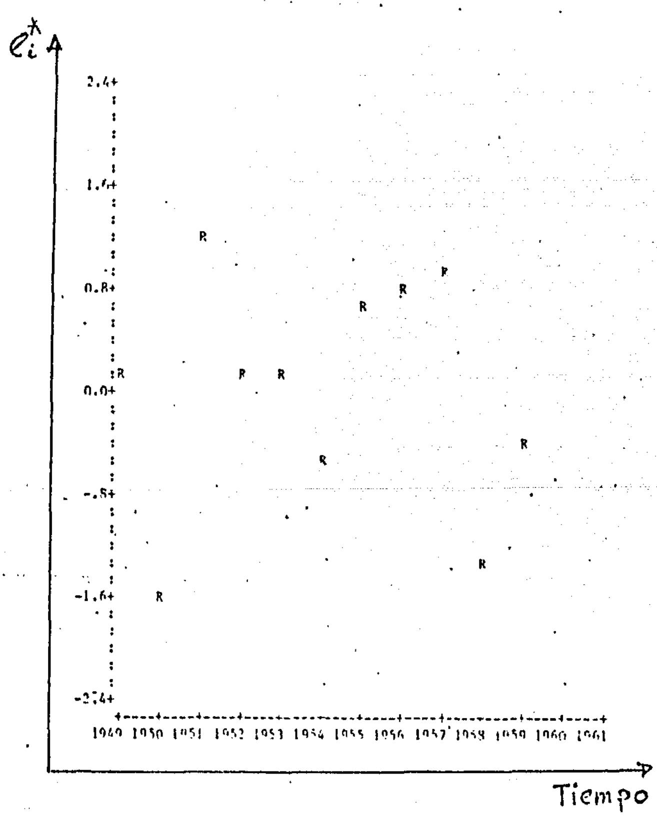
RESULTADOS SUMARIOS DE LA REGRESION DEL MODELO (215)

VARIABLE	COEFICIENTE ESTIMADO	DESVIACION ESTANDARD	VALORES "t"
.STOCK	.612	.109	5.60
. NEWVAR	.086	.003	24.23
.CONSTANTE	-9.007	——	——
n =18	R^2= .987	,	S=.569

Fuente; Estimaciones a partir de los datos de la Tabla 3.4.1

FIGURA 5.5.1

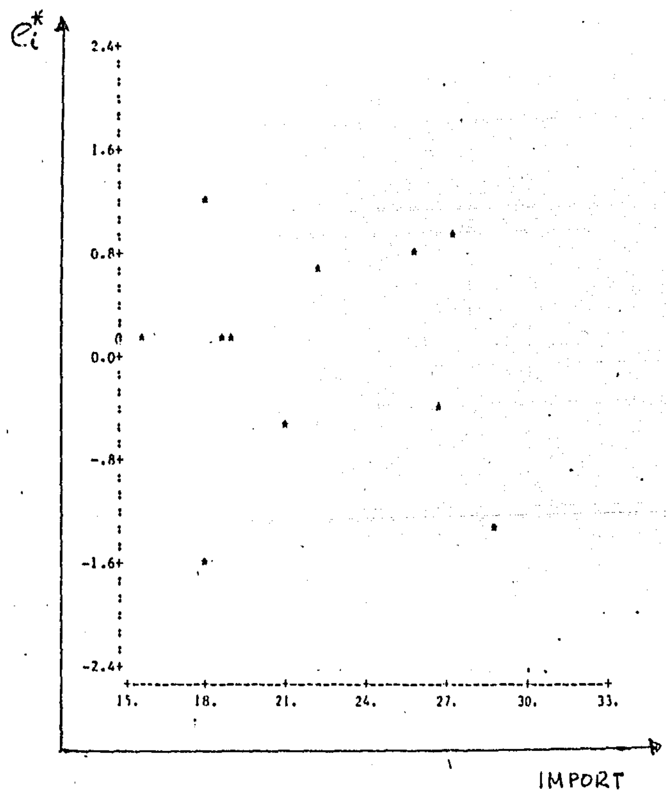
GRAFICA DE RESIDUALES ESTANDARIZADOS CONTRA EL TIEMPO, PARA EL MODELO (218).



Fuente: Chatterjee, S y B.Price. Opus cit.p.165

F I G U R A 5.5.2

GRAFICA DE RESIDUALES ESTANDARIZADOS CONTRA VALORES ESTIMADOS DE LA VARIABLE-RESPUESTA (IMPORT), PARA EL MODELO (218).



Fuente: Chatterjee, S. y B. Price. Opus cit. p. 166

existente, resulta interesante evaluar el efecto de dicha res-- tricción sobre el poder explicativo del modelo original (con sin restricciones). Así, observamos que los valores de R^2 para el -- modelo "completo" o modelo original definido en (67) y el mode--- lo "reducido" definido en (215) estimado con la restricción ba-- jo estudio son, respectivamente, .992 y .987 (como se desprende de las Tablas 3.4.3 y 5.5.1 respectivamente).

El estadístico F calculado para probar la hipótesis nula $H_0: \beta_1 = \beta_3$ resulta ser de 3.36 con (1,7) grados de libertad. Como la F tabu-- lada correspondiente al $\alpha = 0.05$ de significancia nos arroja un valor de 5.59, entonces se acepta la hipótesis nula, o sea, se adepta la hipótesis de que la restricción $\beta_1 = \beta_3$ es signifi-- cativa en la estructura del modelo original.

Cabe aclarar que el resultado anterior es consecuencia directa de aplicar la razón F como una prueba de hipótesis estadística para determinar la significancia de alguna restricción que --- involucre a un subconjunto de parámetros (de al menos dos ele-- mentos) de un modelo regresión múltiple y que se puede demostrar que tiene la forma: 1/

$$F = \frac{[VAN(MR) - VAN(MC)]/(k+1-p)}{VAN(MC)/(n-k-1)} \qquad (218)$$

1/ Al lector interesado en los detalles de este resultado así -- como otros tópicos relacionados con el mismo, lo remitimos a consultar las siguientes fuentes:

Véase, Plackett, R.L. Regression Analysis. Oxford University Press, London, 1960.

Véase, Rao, C.R. Linear Statistical Inference and Applications. Wiley, New York, 1973.

Véase, Seber, G.F. Linear Regression Analysis. Wiley, New York, 1977.

donde:

VAN(MC) = Variación No-explicada del modelo "completo" = $\sum(Y_i - \hat{Y}_i)^2$
= suma de las desviaciones al cuadrado de la variable-respuesta (Y_i) respecto de su correspondiente valor estimado($\hat{Y}_i$).

VAN(MR) = Variación No-explicada del modelo reducido = $\sum(Y_i - \hat{Y}_i^*)$
= suma de las desviaciones al cuadrado de la variable-respuesta del modelo original o completo (Y_i) respecto del valor estimado de la variable -respuesta del modelo "reducido" ($\hat{Y}_i^*$).

donde F es una distribución de Fisher con $[(k+1 - p), (n - k - 1)]$ grados de libertad. (Siendo n = número de observaciones, k = número de parametros en el modelo completo y p = número de parámetros en el modelo reducido).

Debemos enfatizar en esta discusión que la restricción $\beta 1 = \beta 3$ es por supuesto sólo un ejemplo de las muchas restricciones posibles que pueden utilizarse al especificar un modelo de regresión o econométrico.Existe una infinidad de posibilidades a este respecto. 1/

1/ La clase general de posibilidades se encuentra determinada por el espectro de hipótesis más comúnmente empleadas, que son:

1. Todos los parámetros son cero. Es decir, no existe relación lineal entre la variable dependiente (o respuesta) y las variables independientes (o explicativas).

2. Un subconjunto de parámetros (de al menos dos elementos) es nulo. Esto es, todos estos parámetros son cero.

3. Un subconjunto de parámetros son iguales entre sí.

Véase, Plackett, R.L. (1960); Rao, C.R. (1973) y Seber, G.F. (1977), Opus cit.

Las restricciones se justifican normalmente en base a la teoría subyacente que se trate.(Para el caso que acabamos de analizar naturalmente es en base a la teoría económica).Se puede afirmar en general que estas teorías son frecuentemente las responsa-- bles del problema de la multicolinealidad.Además , cualquier restricción en particular puede considerarse como una hipó-- tesis estadística sujeta a comprobación por los métodos conven-- cionales de prueba de hipótesis.

Recapitulando lo expuesto hasta aquí en relación al método de las componentes principales podemos observar que este puede -- interpretarse de dos maneras: .

La primera se refiere al análisis del grado de no-ortogonali-- dad de las variables explicativas multicolineales (enfoque usa-- do previamente en el Apartado 4.3 para detectar la multicolinea- lidad) y la segunda, se relaciona con el análisis de las relacio-- nes existentes entre los coeficientes de las variables explica-- tivas colineales.Esta es la estamos tratando de introducir en la presente discusión y que trataremos de ampliar un poco más en lo que sigue.

Para una mayor comprensión de esta segunda interpretación o fa-- ceta de esta técnica, retomaremos nuevamente el modelo de la "eco- nomía francesa" definido en (67) para mostrar como opera y se apli- ca para corregir el problema de la multicolinealidad en un mode-- lo econométrico específico. Es decir:

$$IMPORT = \beta o + \beta 1 \cdot DOPROD + \beta 2 \cdot STOCK + \beta 3 \cdot CONSUM + U$$

Entonces sabemos de nuestra discusión del Apartado 4.3. (véase — la expresión (130) que las componentes principales de las variables explicativas estandarizadas del modelo anterior son:

$$Z1 = .7063\ DOPROD^{*} + .0435\ STOCK^{*} + .7065\ CONSUM^{*}$$
$$Z2 = -.0357\ DOPROD^{*} + .9990\ STOCK^{*} - .0258\ CONSUM^{*}$$
$$Z3 = -.7070\ DOPROD^{*} - .0070\ STOCK^{*} + .7072\ CONSUM^{*}$$

y donde las variancias de Z1, Z2 y Z3 son respectivamente $\lambda 1 = 1.999$ $\lambda 2 = .998$ y $\lambda 3 = .003$. Donde las λi's son las raíces características de la matríz de correlación de las variables explicativas del modelo (67). Entonces la forma estandarizada de este último resulta ser:

$$IMPORT^{*} = \tilde{\beta}1 \cdot DOPROD^{*} + \tilde{\beta}2 \cdot STOCK^{*} + \beta 3 \cdot CONSUM^{*} + U \tag{219}$$

O alternativamente, en términos de las componentes principales:

$$IMPORT^{*} = \alpha_1 Z1 + \alpha_2 Z2 + \alpha_3 Z3 + U \tag{220}$$

Obsérvese que la ecuación anterior es equivalente a la (219). No obstante, las Zi's son ortogonales. Esta equivalencia es una — consecuencia inmediata de la relación directa entre las α's y $\tilde{\beta}$'s definida en (112). En particular se puede hacdr ver que:

$$\begin{aligned} \alpha_1 &= .7063\,\tilde{\beta}_1 + .0435\,\tilde{\beta}_2 + .7065\,\tilde{\beta}_3 \\ \alpha_2 &= -.0357\,\tilde{\beta}_1 + .9990\,\tilde{\beta}_2 - .0258\,\tilde{\beta}_3 \\ \alpha_3 &= -.7070\,\tilde{\beta}_1 - .0070\,\tilde{\beta}_2 + .7072\,\tilde{\beta}_3 \end{aligned} \tag{221}$$

e inversamente:

$$\tilde{\beta}_1 = .7063\alpha_1 - .0357\alpha_2 - .7070\alpha_3$$

$$\tilde{\beta}_2 = .0435\alpha_1 + .9990\alpha_2 - .0070\alpha_3 \qquad (222)$$

$$\tilde{\beta}_3 = .7065\alpha_1 - .0258\alpha_2 + .7072\alpha_3$$

Las mismas relaciones son válidas para los estimadores minimocuadráticos de los αi's y $\tilde{\beta}i$'s, esto es, para las ai's y $\tilde{b}i$'s. Entonces estos estimadores pueden obtenerse efectuando la regresión de la variable IMPORT* contra las variables explicativas indicadas en las ecuaciones (219) y (220), respectivamente.

Obsérvese que el modelo de regresión (220) definido en términos de las componentes principales no admite una interpretación sencilla. Las variables explicativas (Zi's) en este caso representan combinaciones lineales de las variables explicativas originales (Xi's). Por consiguiente, las α's a diferencia de las $\tilde{\beta}$'s, no permiten una interpretación simple como efectos marginales sobre la variable dependiente respecto de cambios unitarios en las variables explicativas originales.

Por lo tanto, debemos utilizar la regresión en términos de las componentes principales (como la de la expresión (220)) sólo como un medio para analizar el problema de la multicolinealidad. Los resultados finales de la estimación deberán reexpresarse invariablemente en términos de las $\tilde{\beta}i$'s para su reinterpretación.

Con este marco, estamos ahora en posibilidad de usar la técnica de las componentes principales para reducir la multicolinealidad en los datos de estimación.Esta reducción se logra usando un subconjunto del total de componentes principales del modelo original con el propósito de explicar la variación en la variable-respuesta. Cuando se utilizan todas las (tres) componentes principales, la solución por mínimos cuadrados se reproduce exactamente aplicando el sistema de ecuaciones (222).

Como la variable Z3 tiene una variancia igual a .003 , la función lineal definida por Z3 es aproximadamente igual a cero, y es la fuente de multicolinealidad en los datos.Entonces si excluimos a Z3 del modelo(220) y consideramos las regresiones resultantes de la variable $IMPORT^*$ contra Z1 sola y Z1 y Z2,respectivamente.Entonces tendremos que;

$$IMPORT^* = \alpha_1 Z1 + U \qquad (223)$$

y también;

$$IMPORT^* = \alpha_1 Z1 + \alpha_2 Z2 + U \qquad (224)$$

Ambos modelos generan estimadores para los tres coeficientes originales $\tilde{\beta}1$, $\tilde{\beta}2$ y $\tilde{\beta}3$. Estos estimadores son sesgados puesto que parte de la información (Z3 en (224) y Z2 y Z3 en (223)) ha sido eliminada en los dos casos.

Los estimadores para α_1 ó α_1 y α_2 se obtienen efectuando una regrésión de la variable $IMPORT^*$ contra Z1 y después contra Z1 y Z3. No obstante, existe un procedimiento computacional que explota el concepto de ortogonalidad inherente a Z1, Z2 y Z3.(+) Por ejemplo, el mismo estimador para α_1 se podría obtener mediante los modelos (220), (223) y (224). Análogamente, el estimador para α_2 podría obtenerse de (220) y (224).

Se sigue también de esto que si obtenemos los estimadores minimocuadráticos de los $\tilde{\beta}i$'s , los estimadores de las α's se pueden obtener mediante la ecuación (221). Entonces los estimadores de los parámetros $\tilde{\beta}i$'s de las componentes principales correspondientes a los moddlos (223) y (224) pueden calcularse sencillamente utilizando las ecuaciones (222) y haciendo cero las α's apropiadas.

A continuación veremos como se lleva a cabo este proceso para el presente caso de estudio.

Los estimadores para el modelo (220) se proporcionan en la Tabla 5.5.3 (que aparece en la próxima hoja). Entonces de la eduación (221) podemos obtener los estimadores para las α's , que resultan ser, respectivamente, de: $\alpha 1$= .6900, $\alpha 2$=.1913 , $\alpha 3$=1.1597 . Si sustituimos $\alpha 1$= .6900 y $\alpha 2 = \alpha 3$ =0 en las ecuaciones (222) obtenemos los estimadores de las $\tilde{\beta}i$'s correspondientes solamente a la primera componente principal (Z1). Véase la Tabla 5.5.4 (enseguida de la Tabla 5.5.3).

(+) En cualquier modelo de regresión donde el conjunto total de sus variables explicativas son ortogonales, los estimadores de los coeficientes de regresión no se alteran cuando subconjuntos propios (no-vacíos) de dichas variables se agregan o eliminan del modelo en cuestión.

T A B L A 5.5.3

.RESULTADOS SUMARIOS DE LA REGRESION DEL MODELO (67)(+)
(Incluyendo los coeficientes Beta)

VARIABLE	COEFICIENTES BETA ($\tilde{\beta}_i$)	COEFICIENTES OLS (β_i)	DESV. STAND. (β_i)	VALORES "t"
.DOPROD	-.3394	-.0514	.0703	-.731
.STOCK	.2130	.5869	.0046	6.203
.CONSUM	1.3028	.2868	.1022	2.806
.CONSTANT	0	-10.1300	1.2122	-8.355

(+) La mayoría de los paquetes de regresión, por ejemplo el SPSS (Statistical Package for Social Sciences), etc. producen junto con los coeficientes minimocuadráticos (OLS), los coeficientes estandarizados (Beta) asociados o correspondientes a los parámetros del modelo original; los estimadores mencionados satisfacen la relación $b_i = \tilde{b}_i\ (S_y/S_i)$, donde S_y y S_i se definen como anteriormente (Véase, Tabla 4.4.3).

Fuente: Estimaciones a partir de los datos de la Tabla 3.4.1

T A B L A 5.5.4

RESULTADOS DE LAS REGRESIONES DE LOS MODELOS (67),(220),(223) y (224)
(Coeficientes Beta y Coeficientes Minimocuadráticos)

VARIABLE	1a. COMP. PRINCIPAL $\tilde{b}$	1a. COMP. PRINCIPAL b	1a. y 2a. COMP. PRINCIPAL $\tilde{b}$	1a. y 2a. COMP. PRINCIPAL b	TOTALIDAD DE LAS COMP. PRINCIPALES/MINIMOS CUADRADOS $\tilde{b}$	TOTALIDAD DE LAS COMP. PRINCIPALES/MINIMOS CUADRADOS b
. DOPROD	.4873	.0738	.4804	.0727	-.3394	-.0514
. STOCK	.0300	.0820	.2211	.6091	.2130	.5869
. CONSUM	.4875	.1073	.4825	.1062	1.3028	.2868
. CONSTANTE	0	-7.7350	0	-9.1057	0	-10.1300
	R^2=. 952;		R^2=.988 ;		R^2=.992	

Fuente: Estimaciones a partir de los datos de la Tabla 3.4.1

Equivalatemente, la definición de Z1 de la expresión (130) puede usarse para obtener una nueva formulación del modelo (219) como sigue:

$$\begin{aligned} IMPORT^{*} &= .6900\ (.7063\ DOPROD^{*} + .0435\ STOCK^{*} + .7065\ CONSUM^{*}) \\ &= .4873\ DOPROD^{*} + .0300\ STOCK^{*} + .4875\ CONSUM^{*} \end{aligned} \qquad (225)$$

Análogamente, los estimadores de las $\tilde{\beta}i$'s ($\tilde{\beta}1$ y $\tilde{\beta}2$) se obtienen efectuando la regresión de la variable IMPORT con respecto a las dos primeras componentes principales definidas en (224), o bien, sustituyendo los valores $\alpha 1 = .6900$ y $\alpha 2 = .1913$ y $\alpha 3 = 0$ en las dos primeras ecuaciones de la expresión (222).

De la Tabla 5.5.4 resulta evidente que usando diferentes subconjuntos propios (no-vacíos) de componentes principales obtenemos resultados sustancialmente diferentes.

Ya hemos mencionado que los estimadores OLs son insatisfactorios. Por ejemplo, el signo negativo del coeficiente de la variable $DOPROD^{*}$ no era de esperarse, y por lo mismo, no puede interpretarse lógicamente o de acuerdo a la teoría económica. Además, existe un nivel peligroso de multicolinealidad que se refleja a través de la tercera componente principal Z3. Esta variable tiene una variancia de casi cero ($\lambda 3 = .003$), y es por lo tanto, aproximadamente igual a cero.

De las otras dos componentes principales (Z1 y Z2), es muy claro que la primera se asocia con el efecto combinado de DORROD y CONSUM. La segunda componente principal se asocia únicamente con STOCK. Esta conclusión se desprende observando los valores de la Tabla 5.5.4. Ya que los coeficientes de DOPROD y CONSUM se determinan completamente efectuando la regresión de $IMPORT^*$ contra Z1 solamente. La adición de Z2 hace que el coeficiente de STOCK aumente de .0826 a .6091. Asimismo, R^2 crece de .952 a .988. Si escogemos el modelo basado en las primeras dos componentes principales, la ecuación resultante expresada en términos de las variables originales es:

$$\widehat{IMPORT} = -9.1057 + .0727\ DOPROD + .6091\ STOCK + .1062\ CONSUM \qquad (226)$$

El modelo anterior ofrece una representación más plausible y, sobre todo, más lógica que la obtenida mediante mínimos cuadrados ordinarios. Además, nuestro análisis nos permite la obtención de una cuantificación explícita (en términos de variables estandarizadas) de la dependencia lineal existente entre las variables explicativas. Específicamente, haciendo Z3 = 0 y sustituyendo este valor en la expresión (130) tendremos:

$$-.7070\ DOPROD^* - .0070\ STOCK^* + .7072\ CONSUM^* \doteq 0 \qquad (227)$$

Los valores estandarizados de DORROD y CONSUM son esencialmente los mismos. Esta información puede ser útil cuanti y cualitativamente si la ecuación (226) se usa para la predicción o para el análisis de políticas.

Entonces siguiendo con nuestra discusión sobre la segunda interpretación de la técnica de las componentes principales, reiteramos que dicha técnica se asocia a la noción de imposición de restricciones entre las $\tilde{\beta}i$'s que ha sido introducida en discusiones anteriores.

Así, tenemos que los estimadores de la ecuación (224) pueden obtenerse haciendo $\alpha_3 = 0$ en las ecuaciones del sistema (222).

$$-.7070\ \tilde{\beta}1\ -.0070\ \tilde{\beta}2\ +.7072\ \tilde{\beta}3 = 0 \qquad (238)$$

o equivalentemente:

$$\tilde{\beta}1 \doteq \tilde{\beta}3 \qquad (229)$$

La ecuación (228) en términos de los parámetros originales, resulta ser:

$$-6.67\ \beta 1\ +\ 4.54\beta 3 = 0 \qquad (230)$$

o también:

$$\beta 1 = .6872\ \beta 3 \qquad (231)$$

Recuérdese que ya sabíamos que la restricción definida en (212):

$$\beta 1 = \beta 3$$

era una restricción válida desde un punto de vista téorico, o sea, supusimos que esta restricción era el resultado de un análisis cualitativo basado en el conocimiento del proceso bajo estudio. Y se impuso sin examinar la información. Ahora bien, utilizando datos concretos, hemos "descubierto" que la regresión de las --

componentes principales Z1 y Z2 genera un resultado que es equivalente a imponer la restricción (231) al modelo original. Esto es, el resultado sugiere que el efecto marginal de la producción doméstica (DOPROD) sobre las importaciones (IMPORT) es alrededor de 69% del efecto marginal del consumo doméstico (CONSUM) sobre las importaciones (IMPORT).

En otras palabras, el método de las componentes principales permite disponer de estimadores alternativos de los coeficientes de regresión así como de información útil sobre el proceso subyacente que ha generado los datos para la estimación. La estructura de la dependencia lineal entre las variables explicativas se hace explícita. Las componentes principales con variancias pequeñas (raíces características) muestran la relación lineal entre las variables originales que representan la fuente de la multicolinealidad. También la eliminación de la multicolinealidad mediante la remoción de una o más componentes principales de un modelo de regresión es equivalente a imponer restricciones sobre los parámetros o coeficientes de regresión. Adicionalmente, a través de la técnica de las componentes principales es posible identificar qué restricciones son consistentes con el modelo propuesto y la información contenida en los datos, y de esta manera, el proceso de eliminación no resulta arbitrario o producto de una mera "corazonada", como sucede por lo regular, con las consecuencias ya anotadas en el Apartado 5.1.

A propósito, y ya que estamos abordando el tema de las restriccio--

nes sobre los parámetros de un modelo de regresión, conviene --- destacar un aspecto importante sobre las"combinaciones lineales" de éstos. Es decir, sobre las funciones lineales que se pueden construir a partir de los estimadores de los parámetros Bi's.

Supongamos que el modelo general definido en (6) como:

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

ha sido cuidadosamente especificado de tal suerte que los coe-- ficientes de dicha ecuación son de interés primordial para el aná lisis de políticas y la toma de decisiones. Hemos visto que la presencia de multicolinealidad en un modelo puede imposibilitar, o al menos dificultar, la estimación de parámetros individuales. Sin embargo, se puede demostrar que siempre es posible estimar con exactitud algunas funciones lineales de las βi's. 1/

No obstante, la pregunta obvia que resulta es: ¿ Cuáles de estas funciones lineales pueden estimarse?, y de éstas, ¿ Cuáles son de interés para nuestro análisis?. Aunque más adelante en este mismo apartado utilizaremos los datos del modelo (67) de la"economía francesa" para estimar ciertas combinaciones paramétricas de valor especial para nuestro estudio, por el momento, demostra-- remos de una manera indirecta, que siempre existen funciones li-- neales de las βi's que pueden ser estimadas con precisión. Es de--

---

1/ Véase, Silvey, S.D. Multicollinearity and Imprecise Estimation. Opus cit.

cir, considérese el modelo:

$$IMPORT = \beta 0 + \beta 1 \cdot DOPROD + \beta 2 \cdot STOCK + \beta 3 \cdot CONSUM + U$$

Ya hemos observado que existe una relación histórica entre las variables CONSUM y DOPROD definida en la expresión (70) como:

$$CONSUM = (2/3)DOPROD$$

Si sustituimos la expresión (70) en la ecuación (67) obtenemos el modelo:

$$IMPORT = \beta 0 + (\beta 1 + (2/3)\beta 3) \cdot DOPROD + \beta 2 \cdot STOCK + U \qquad (232)$$

Es decir, al eliminar la variable CONSUM de la ecuación (67) podemos obtener estimadores precisos de los parámetros $\beta 1+(2/3)\beta 3$ y $\beta 2$. Donde la multicolinealidad ya no existe más. Los resultados de la regresión del modelo (232) se muestran en la Tabla 5.5.5 (que aparece en la próxima hoja) en donde observamos que el valor correspondiente de R^2 (.983) prácticamente no cambia (respecto por ejemplo, al valor reportado en la Tabla 5.5.4). Por otra parte, sabemos de la Tabla 3.4.3 que la correlación entre las variables DOPROD y STOCK es de .026. Si construyeramos las gráficas correspondientes de residuos observaríamos que éstas guardan un comportamiento satisfactorio.(+)

En resumen, para este caso hemos usado información adicional a los datos para mostrar que el parámetro de DOPROD en el modelo (67)

(+) Se omiten estas gráficas para simplificar la presente exposición.

TABLA 5.5.5

RESULTADOS DE LA REGRESION DEL MODELO (232)

FUNCION LINEAL	COEFICIENTE ESTIMADO	DESV. ESTAND.	VALORES "t"
$\beta 1 + (2/3)\beta 3$	.145	.007	20.70
$\beta 2$	.622	.128	4.86
R^2=.983		S= .667	

Fuente: Estimaciones a partir de la Tabla 3.4.1

es una combinación lineal de la forma:

$$\beta 1 + (2/3)\beta 3 \qquad (233)$$

De igual manera, hemos hecho ver que esta función lineal puede -- ser estimada con exactitud aún cuando exista multicolinealidad en los datos. Que el valor de la expresión anterior tenga o nó una utilidad o "sentido económico" (como en este caso) es desde luego otra cuestión. Al menos, es importante saber que el estimador --- del coeficiente de DOPROD en el modelo (232) no está midiendo únicamente el efecto marginal de la variable DOPROD sobre la variable IMPORT sino también incluye parte del efecto de la variable CONSUM.

Siguiendo con este orden de ideas, utilizaremos el concepto de -- las componentes principales discutido en el Apartado 4.3 para identificar las combinaciones lineales de las βi's que pueden ser exáctamente estimadas. Lo aplicaremos al modelo de "publicidad" definido en (105), esto es:

$$VT_t = \beta o + \beta 1 PBt + \beta 2 PRt + \beta 3 GVt + \beta 4 PBt-1 + \beta 5 PRt-1 + Ut$$

Aunque los conceptos aquí vertidos son menos intuitivos que los discutidos anteriormente en relación al método de las componentes principales, trataremos de simplificar esta exposición para mayor comprensión de los no especializados en la materia. Un desarrollo formal para los especializados se ofrece en las referencias bibliográficas recomendadas. Aunque parte de ella ya se describió en el Apartado 4.3.

Comenzaremos recordando que la transformación definida en (110) convierte a las variables explicativas estandarizadas en un nuevo conjunto de variables ortogonales. Para simplificar esta exposición denominaremos como : X1, X2,...,X5 y Y respectivamente, a las variables explicativas y explicada del modelo (105).

Entonces, si aplicamos la transformación referida a las variables arriba mencionadas, podemos convertir a X1,X2,...,X5 en un conjunto de variables ortogonales Z1,Z2,...,Z5 mediante el sistema:

$$\begin{aligned}
Z1 &= -.532\ X1 + .232\ X2 + .389\ X3 - .395\ X4 + .595\ X5 \\
Z2 &= -.024\ X1 + .825\ X2 - .022\ X3 - .260\ X4 - .501\ X5 \\
Z3 &= -.668\ X1 + .158\ X2 - .217\ X3 + .692\ X4 - .057\ X5 \qquad (234) \\
Z4 &= .074\ X1 - .037\ X2 + .895\ X3 + .338\ X4 - .279\ X5 \\
Z5 &= .514\ X1 + .489\ X2 - .010\ X3 + .428\ X4 + .559\ X5
\end{aligned}$$

Nótese aquí que los coeficientes de la ecuación que define a Z1 son las componentes del vector característico correspondiente a la raíz característica más grande de la matríz de correlación (R) asociada a las variables explicativas originales. Asimismo, los coeficientes que definen a las variables Z2 hasta Z5 son, a su vez, las componentes de los vectores característicos correspondientes a las raíces características restantes, en orden decreciente de magnitud.

Por consiguiente, el modelo (105) puede expresarse en términos --

de sus variables estandarizadas X1,X2,...,X5 y Y, como:

$$Y = \tilde{\beta}1\ X1 + \tilde{\beta}2\ X2 + \tilde{\beta}3\ X3 + \tilde{\beta}4\ X4 + \tilde{\beta}5\ X5 + U' \qquad (235)$$

Los coeficientes del modelo estandarizado anterior se conocen -- usualmente en la literatura como "Coeficientes Beta". Representan los efedtos marginales de las variables explicativas estandariza-- das sobre la variable- respuesta, en términos de unidades estan-- darizadas de desviación. Por ejemplo, el parámetro $\beta 1$ mide el ---- cambio en unidades estandarizadas de las ventas totales (VT_t) co-- rrespondientes a un incremento de una unidad de desviación están-- dard de la variable "PB_t" (gastos de publicidad), etc.

Sabemos de la expresión (111) que la ecuación anterior escribirse alternativamente en términos de las componentes principales defi-- nidas en (149) como:

$$Y = \alpha_1 Z1 + \alpha_2 Z2 + \alpha_3 Z3 + \alpha_4 Z4 + \alpha_5 Z5 + U' \qquad (236)$$

La equivalencia de la ecuación (235) y la (236) es consecuencia directa de la relación existente entre las variables Xi's y Zi's definidas mediante la transformación (234) y la relación existen-- te entre las α's y β's y sus valores estimados que denominare-- mos en lo sucesivo como "a's" y "$\tilde{b}$'s" se define asimismo mediante la siguiente transformación:

$$\begin{aligned}
\tilde{b}_1 &= -.532\,a_1 - .024\,a_2 - .668\,a_3 + .074\,a_4 + .514\,a_5 \\
\tilde{b}_2 &= .232\,a_1 + .825\,a_2 + .158\,a_3 - .037\,a_4 + .489\,a_5 \\
\tilde{b}_3 &= .389\,a_1 - .022\,a_2 - .217\,a_3 + .895\,a_4 - .010\,a_5 \\
\tilde{b}_4 &= -.395\,a_1 - .260\,a_2 + .692\,a_3 + .338\,a_4 + .428\,a_5 \\
\tilde{b}_5 &= .595\,a_1 - .501\,a_2 - .057\,a_3 - .279\,a_4 + .559\,a_5
\end{aligned} \qquad (237)$$

Obsérvese que la transformación anterior involucra los mismos — "pesos" usados para la definición de la transformación (234). Esto es consecuencia de la propiedad teórica establecida en (115).

La ventaja de trabajar con el modelo transformado (236) es que las variables Zi's son ortogonales. Por ejemplo, la precisión de los estimadores de los parámetros α's se puede evaluar fácilmente por la variancia de las a's. Sabemos por la expresión (119) que la variancia estimada de una "a_i" en particular se puede escribir como:

$$\widehat{Var}(\hat{\alpha}_i) = \widehat{Var}(a_i) = \frac{\hat{\sigma}_u^2}{\lambda_i} \qquad (238)$$

es decir, resulta inversamente proporcional a la i-ésima raíz característica. Si recordamos que la matríz Λ para el modelo de "publicidad" definida previamente en la expresión (133) como:

$$\Lambda = \begin{bmatrix} 1.701 & 0 & 0 & 0 & 0 \\ 0 & 1.288 & 0 & 0 & 0 \\ 0 & 0 & 1.145 & 0 & 0 \\ 0 & 0 & 0 & .859 & 0 \\ 0 & 0 & 0 & 0 & .007 \end{bmatrix}$$

notaremos enseguida que todas las variancias de las "a_i's" — pueden estimarse con precisión mediante (238) con excepción de la asociada a la raíz $\lambda 5$, que resulta desmesuradamente grande en relación a las otras variancias, denotando con ello, la presencia de multicolinealidad entre las variables explicativas (estandarizadas) que conforman la componente principal Z5.

Nuestro interés en las α's es solamente porque son un medio para analizar las $\tilde{\beta}$'s. Sabemos por (125) que la variancia estimada de una $\tilde{b}i$ cualquiera puede expresarse como:

$$\hat{Var}(\hat{\tilde{\beta}}i) = \hat{Var}(\tilde{b}i) =$$

$$= \left[\frac{r_1^2(i)}{\lambda_1} + \frac{r_2^2(i)}{\lambda_2} + \cdots + \frac{r_5^2(i)}{\lambda_5} \right] \hat{\sigma}_u^{*2} \qquad (239)$$

donde $r_1^2(i), \ldots, r_5^2(i)$ corresponden respectivamente, a los coeficientes o constantes que definen la combinación lineal que da — origen a $\tilde{b}i$ de conformidad con la transformación (237) y $\hat{\sigma}_u^{*2}$ es el valor "corregido" de $\hat{\sigma}_u^2$ que definiremos más adelante. El valor de la variancia de $\tilde{b}1$ tiene entonces la forma:

$$\hat{Var}(\tilde{b}1) = \left[\frac{(-.532)^2}{(1.701)} + \frac{(-.024)^2}{(1.288)} + \frac{(-.668)^2}{(1.145)} + \frac{(.074)^2}{(.859)} + \frac{(.514)^2}{(.007)} \right] \hat{\sigma}_u^{*2} \qquad (240)$$

O en otras palabras:

$$\hat{Var}(\tilde{b}1) = (0.428 + 0.120 + 0.714 + 0.317 + 102.420) \hat{\sigma}_u^{*2} \qquad (241)$$

esto es:

$$Var(\tilde{b}1) = (1.579 + 102.420) \hat{\sigma}_u^{*2} \qquad (242)$$

O alternativamente:

$$\hat{Var}(\tilde{b}1) = (1.579) \hat{\sigma}_u^{*2} + (102.420) \hat{\sigma}_u^{*2} \qquad (243)$$

Obsérvese de (240) que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_5$ y que sólo λ_5 es desproporcionadamente pequeña ($\lambda_5 = .007$). Luego entonces, como pode--

mos notar de las expresiones (241)-(243), es sólo el último término de la variancia estimada de $\tilde{b}_1$ (que involucra a λ_5) que puede destruir la precisión de $\tilde{b}_1$ por su enorme magnitud.

Por otra parte, como las expresiones para las variancias de las otras $\tilde{b}_i$'s pueden también obtenerse a partir de (239) un requisito para obtener la variancia más pequeña es equivalente a pedir que el coeficiente que modifica al factor $1/\lambda_5$ sea el más pequeño.

Si investigamos detenidamente las ecuaciones que definen la transformación (237) que asocia (ó "mapea") a los estimadores $\{a_i\}$ hacia $\{\tilde{b}_i\}$ entonces podemos concluir que $\tilde{b}_3$ es el estimador más preciso, puesto que el coeficiente o constante asociada con el factor $1/\lambda_5$ en la expresión para la variancia de $\tilde{b}_3$ es $(.01)^2$ ó .001, o sea, es el más pequeño.

Se puede extender un poco más este tipo de análisis para identificar funciones lineales significativas de las $\tilde{\beta}_i$'s que pueden estimarse con más precisión que las $\tilde{\beta}_i$'s individuales. Por ejemplo, podríamos estar más interesados en estimar la combinación $(\tilde{\beta}_1 - \tilde{\beta}_2)$ que los parámetros $\tilde{\beta}_1$ y $\tilde{\beta}_2$ separadamente.

Para el modelo (105), la combinación lineal $(\tilde{\beta}_1 - \tilde{\beta}_2)$ mediría por ejemplo, el incremento en las ventas totales (VT_t) correspondientes a un incremento unitario en el presupuesto de publicidad para el año corriente (PBt) acompañado al mismo tiempo, de una reducción o baja unitaria en los gastos de promoción para el año corriente (PRt). La variancia de $(\tilde{b}_1 - \tilde{b}_2)$ para este caso —

puede calcularse simplemente restando los valores de los coeficientes correspondientes definidos en la transformación (237) y utilizando los coeficientes restantes de las a's como anteriormente. Las raíces $\lambda_1, \lambda_2, \ldots, \lambda_5$ siguen siendo las mismas. Específicamente se tiene que:

$$\tilde{b}1 - \tilde{b}2 = -0.764\,a_1 - 0.849\,a_2 - 0.826\,a_3 + 0.111\,a_4 + 0.025\,a_5 \qquad (244)$$

y entonces la variancia de ($\tilde{b}1 - \tilde{b}2$) se puede escribir como:

$$\text{Var}\,(\tilde{b}1 - \tilde{b}2) =$$

$$= \left[\frac{(-.764)^2}{\lambda_1} + \frac{(-.849)^2}{\lambda_2} + \frac{(-.826)^2}{\lambda_3} + \frac{(.111)^2}{\lambda_4} + \frac{(.025)^2}{\lambda_5}\right] \hat{\sigma}_u^2 \qquad (245)$$

<u>Otra vez, observamos que el coeficiente que modifica al factor ($1/\lambda 5$) hace posible estimar con precisión al parámetro ($\tilde{\beta}1 - \tilde{\beta}2$). Generalizando este procedimiento concluimos que cualquier combinación lineal de las $\tilde{\beta}$'s que contenga un pequeño coeficiente asociado al factor "problema" ($1/\lambda 5$ en este caso) en la expresión correspondiente a la variancia, puede ser estimado con precisión.</u>

Por último, conviene insistir que aunque la técnica de las componentes principales es bastante poderosa, el proceso de cómputo involucrado para este tipo de análisis es más complejo y no es posible realizarlo con la "paquetería estándar" (SOFTWARE) disponible para el análisis de regresión, como ya hemos destacado con anterioridad. Puesto que los datos deben ser procesados mediante una subrutina de componentes principales que trabaje con la matríz de correlación de las variables explicativas del modelo original (sin estandarizar) para estimar las raíces características y los coeficientes de la transformación correspondiente como la que se define en (237).

No obstante, la mayoría de los paquetes producen estimadores de los coeficientes "Beta" como parte normal de su proceso de cómputo. En la próxima hoja se presenta la Tabla 5.5.6 en donde se muestran los estimadores paramétricos para el modelo (105) que hemos estado discutiendo hasta ahora. Usando esta información, notamos que el estimador de ($\tilde{\beta}1 - \tilde{\beta}2$) es .583-.973 =-.390 .Y que los valores "t" para probar la significancia estadística (igual a cero) de los parametros βi y $\tilde{\beta} i$ son idénticos.

Nótese de paso que el coeficiente beta, $\tilde{\beta} i$, es un versión escalada del parámetro βi, y que al construir los valores "t" definidos indistintamente como:

$$b_i/\sqrt{\widehat{Var}(b_i)} = \tilde{b}_i/\sqrt{\widehat{Var}(\tilde{b}_i)} \qquad (246)$$

este valor "t" no se altera. Sin embargo, el valor de $\hat{\sigma}_u^{*2}$ que aparece en las expresiones (240)-(243) y (245) refleja este factor de escala. El valor de este término es el valor de la varianza del término de perturbación (del modelo de regresión original (σ_u^2)) dividido entre la suma de las desviaciones al cuadrado respecto de la media de la variable-respuesta. O sea, entre la "variación total" ($\Sigma(Y_i-\bar{Y})^2$). Es decir:

$$\hat{\sigma}_u^{*2} = \frac{\hat{\sigma}_u^2}{\sum (Y_i-\bar{Y})^2} \qquad (247)$$

T A B L A 5.5.6

RESULTADOS DE LA REGRESION DEL MODELO (237)
(Versión estandarizada del modelo (105))

VARIABLE	COEFICIENTE BETA ($\tilde{b}_i$)	DESV. STAND, ($\tilde{b}_i$)	VALORES "t" (para b_i y $\tilde{b}_i$)
PB^*_t	.5830	.4380	1.331
PR^*_t	.9734	.4170	2.334
GV^*_t	.7859	.07476	10.512
PB^*_{t-1}	.3953	.3670	1.077
PR^*_{t-1}	.5035	.4754	1.059
n=22 ,	R^2=.9169,	S=.0721 ($\hat{\sigma}^{*2}_u$ = .0052)	

Fuente: Estimaciones a partir de los datos de la Tabla 4.3.1

Para el problema que estamos manejando, tendremos que:

$$\hat{\sigma}_u^{*2} = \frac{1.7424}{335.451} = .0052 \tag{248}$$

donde los valores para el numerador y denominador de la expresión anterior se infieren de la información contenida en la Tabla 4.3.2. que corresponde a la regresión original. Alternativamente, el valor de $\hat{\sigma}_u^{*2}$ así como el de los coeficientes "beta" puedan obtenerse mediante la regresión del modelo (235) que representa la versión estandarizada del modelo original (105). Los resultados son los que se presentan en la Tabla 5.5.6.

Por cualquiera de los procedimientos antes señalados, se puede conocer el valor de $\hat{\sigma}^{*2}$, y por tanto, es posible calcular el valor de la variancia de $\tilde{b}1$ y $(\tilde{b}1 - \tilde{b}2)$, por ejemplo. De la expresión (245) resulta evidente que:

$$\hat{Var}(\tilde{b}1 - \tilde{b}2) = \left[\frac{(-.764)^2}{(1.701)} + \frac{(-.849)^2}{(1.288)} + \frac{(-.826)^2}{(1.145)} + \frac{(.111)^2}{(.859)} + \frac{(.025)^2}{(.007)}\right](.0052) \tag{249}$$

$$= (1.5129 + 0.0893)\ (.0052)$$

$$= (0.0079 + 0.0005) = 0.0084$$

por consiguiente:

$$\hat{Var}(\tilde{b}1 - \tilde{b}2) = 0.0084 \tag{250}$$

Análogamente, podemos calcular la variancia de $\tilde{b}1$, utilizando la expresión (243), resulta evidente que:

$$\hat{Var}(\tilde{b}1) = (1.579)(.0052) + (102.420)(.0052)$$

$$= 0.0079 + 0.536 = 0.5405 \tag{251}$$

Y por consiguiente, es posible determinar intervalos de con--- fianza para dichos estimadores. Por ejemplo, un intervalo del 95% de confianza para el estimador $(\tilde{b}1 - \tilde{b}2) = (.5830 - .9734) = -.390$ será ;

$$-.390 \pm (2.12)\sqrt{.0083} \qquad (252)$$

O en otros términos:

$$(-.583 \ , \ -.197) \qquad (253)$$

Similarmente, para $\tilde{b}1 = .5830$ un intervalo del 95% de con--- fianza será:

$$.5830 \pm (2.12)\sqrt{0.5405} \qquad (254)$$

es decir:

$$(-.976 \ , \ 2.142) \qquad (255)$$

Naturalmente que existen otras funciones lineales de las $\tilde{\beta}i$'s que podrían estimarse con precisión. Por ejemplo, cualquier función que produzca un coeficiente pequeño asociado al factor $(1/\lambda 5)$. en su variancia, es una posibilidad. De la expresión (237) es -- fácil notar que todas las diferencias que involucran a $\tilde{b}_1, \tilde{b}_2, .., \tilde{b}_5$ podrían considerarse. No obstante, algunas diferencias tienen sentido, en cambio otras nó. Por ejemplo, el parámetro $(\tilde{\beta}1 - \tilde{\beta}2)$ nos indica que el efecto de desplazar una unidad de gasto de promoción hacia publicidad en el año corriente, representa una pérdida entre 0.197 y 0.583 unidades estandarizadas en las ventas — totales. Pero, sin embargo, $(\tilde{\beta}1 - \tilde{\beta}4)$ no lo es, porque representa el impacto en las ventas totales, de desplazar una unidad de -- gasto en publicidad del año corriente hacia el gasto en publi-- cidad del año anterior.Lo cual es un absurdo, etc.etc.

En general, cuando las constantes o "pesos" de las ecuaciones de la transformación (237) se conocen, así como los valores de las raíces características, siempre es posible "escudriñar" dichas constantes e identificar aquellas combinaciones lineales de los coeficientes o parámetros originales (βi's) que pueden estimarse con precisión.Desde luego que de todas esas funciones, solo algunas resultarán de interés y tendrán sentido para nuestra investigación como ya hemos observado.

En resumen, y para concluir con este apartado, podemos afirmar que cuando se presenta la multicolinealidad y no se puede obtener información adicional (esto es, de tipo extra-muestral) por la razón que sea, siempre es posible estimar algunos coeficientes de regresión y además, alguna función de éstos,con precisión. Para determinar cuáles coeficientes y qué combinaciones lineales de los mismos pueden estimarse, se recomienda precisamente, la técnica de las componentes principales,cuyos alcances hemos -- intentado introducir en este apartado. Aunque debemos reconocer que la técnica aludida no resuelve por sí sola el problema de multicolinealidad cuando ésta se presente en un modelo, sí re-- presenta un poderoso instrumento para identificar las dependencias estructurales entre las variables explicativas y qué fun-- ciones paramétricas son estimables, abriendo con ello la posibilidad indiscutible de poder aplicar correcta y efectivamente, el método de la restricción paramétrica o de eliminación de variables inclusive, para corregir o eliminar , esto es, romper el "candado" de la multicolinealidad en un modelo de regresión o econométrico (uniecuacional) cualquiera.

5.6 METODO DE SELECCION DE VARIABLES Y LA REGRESION DE CORDILLERA 1/

En nuestras discusiones anteriores sobre el modelo lineal general supusimos que las variables que se incluían en la ecuación habían sido seleccionadas previamente. Nuestro análisis se enfocaba en el examen de la ecuación para averiguar si la especificación funcional era o nó correcta, y si los supuestos sobre el término de perturbación o aleatorio eran o nó válidos. El análisis presuponía que el conjunto de variables por incluirse en la ecuación ya estaba decidido.

Sin embargo, en muchas aplicaciones econométricas o del análisis de regresión en general, el conjunto de variables por incluirse en el modelo no está predeterminado, y frecuentemente, la primera parte del análisis consiste, precisamente, en seleccionar dichas variables.

Existen algunas ocasiones donde por razones teóricas o de otra

---

1/ Aunque el método de selección de variables no representa en sí mismo una técnica diseñada ex-profeso para atacar la multicolinealidad en un modelo de regresión, el propósito de tratarlo aquí es orientarlo a la eliminación o minimización de dicho problema, y bajo el contexto del método referido, buscar y resaltar su ligazón con el análisis de regresión de cordillera estudiado en el Apartado 4.4 .Para mayor abundamiento sobre el método de selección de variables se recomienda la fuente que se cita enseguida:

Véase, Hocking, R.R. Misspecification in Regression. The American Statistician, 28 ,1974, pp.39-40.

índole es relativamente fácil determinar las variables por incluirse en la ecuación. Sin embargo, en otras, en donde no existe una teoría específica de respaldo, el problema de selección de variables se vuelve muy importante.

El problema de selección de variables y el de especificación funcional de la ecuación están muy ligados entre sí. Las interrogantes que deben contestarse al formular un modelo de regresión o econométrico son: ¿ Qué variables deberán de incluirse en la ecuación?. ¿ Y de qué manera deberán de incluirse? . Esto es, ¿ deberemos incluirlas en su forma original Xi ó como una variable transformada como : Xi^2 , 1/Xi , logXi, etc..o.una combinación adecuada de dichas formas ?, etc.

Aunque sería deseable poder resolver ambos problemas a la vez, se recomienda atacarlos en forma secuencial: Es decir, primeramente determinar qué variables se incluirán en la ecuación y posteriormente, investigar la forma exacta en que dichas variables deberán aparecer en la ecuación.

El enfoque anterior es una simplificación evidentemente, pero permite hacer más tratable el problema de selección de variables, es decir, una vez incluidas en la ecuación (o modelo) las variables correctas, se procede a aplicar los métodos convencionales del análisis de regresión para determinar la forma exacta en que aparecerán las variables explicativas en la ecuación final.

Considérese el modelo lineal general definido en (1) como:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U_i$$

donde β_i son parámetros y U_i es el conocido término de perturbación . En vez de tratar con todo el conjunto de k variables (sobre todo cuando k es grande), podríamos optar por eliminar un cierto número de ellas y construir una "nueva ecuación" con un subconjunto de las k variables. Uno de los propósitos de este apartado se aboca a determinar qué variables deberán permanecer o retenerse en la nueva ecuación.

Denotaremos a las variables retenidas como $X_1, X_2, \ldots, X_p$ y a las que se eliminan como $X_{p+1}, X_{p+2}, \ldots, X_k$.

Examinemos brevemente el efecto de la eliminación de variables bajo las dos condiciones generales siguientes:

1. El modelo que relaciona a la variable-respuesta Y con todas las X_i's tiene coeficientes β_i's ($\beta_0, \beta_1, \ldots, \beta_k$) distintos de cero.(Esto es, el modelo (1) con $\beta_i \neq 0$).

2. El modelo (1) con $\beta_0, \beta_1, \beta_2, \ldots, \beta_p \neq 0$ pero $\beta_{p+1}, \beta_{p+2}, \ldots, \beta_k = 0$.

Supóngase que en lugar de estimar el modelo (1), estimamos el sub-modelo (o modelo "reducido") siguiente:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + U_i \qquad (256)$$

Describiremos el efecto de estimar el modelo (1) con todas y -- parte de las variables explicativas contenidas en el mismo, bajo las dos condiciones anotadas anteriormente. Esto es, analizaremos los efectos de incluir ciertas variables en una ecuación cuando deberían haberse eliminado (porque sus parámetros poblacionales son en realidad cero) así como el efecto de dejar "afuera" determinadas variables explicativas, que deberían estar -- "adentro" (porque sus parámetros poblacionales son en realidad distintos de cero)

Examinaremos el efecto de la eliminación de variables sobre los estimadores de los parámetros y los valores pronosticados de Y. Veremos que la solución al problema de selección de variables resulta evidente cuando se conocen los efectos de la eliminación de variables esenciales en una ecuación o modelo de regresión.(+)

Denotemos como $b_0^*, b_1^*, b_2^*, \ldots, b_k^*$ a los estimadores de las βi's cuando el modelo (1) es estimado con todas sus variables explicativas; X1, X2,...,Xk. Y como ; $b_0, b_1, b_2, \ldots, b_p$ a los estimadores correspondientes a los parámetros del modelo (256).

(+) Aunque la mayor parte del material sobre eliminación de variables y su impacto sobre la precisión de los estimadores de un modelo de regresión ya ha sido tratado con anterioridad y formalmente en el Apartado 5.1, consideramos oportuno y conveniente revisar someramente algunos aspectos sobresalientes sobre este tópico por su importancia y estrecha vinculación con el tema objeto de este apartado.

Sean $\hat{Y}^*$ y $\hat{Y}$ los valores pronosticados correspondientes obtenidos considerando el modelo (1) y el modelo (256), respectivamente, correspondiente a un mismo conjunto de datos (X1, X2, .., Xk).

Entonces se puede demostrar que los siguientes resultados son válidos: 1/

1o. Los estimadores $b_0, b_1, b_2, \ldots, b_p$ son estimadores sesgados de $\beta_0, \beta_1, \ldots, \beta_p$ al menos que las restantes β_i's en el modelo ($\beta_{p+1}, \beta_{p+2}, \ldots, \beta_k$) sean cero o que el conjunto de variables $X_1, X_2, \ldots, X_p$ sea ortogonal respecto al conjunto de variables $X_{p+1}, X_{p+2}, \ldots, X_k$.

2o. Los estimadores $b_0^*, b_1^*, b_2^*, \ldots, b_p^*$ son menos precisos que los estimadores $b_0, b_1, b_2, \ldots, b_p$. O sea:

$$\mathrm{Var}(b_i^*) \geq \mathrm{Var}(b_i); \; i=0,1,2,\ldots,p \qquad (257)$$

Es decir, la variancia de los estimadores del modelo reducido (256) no son mayores que las variancias correspondientes de los estimadores del modelo completo (1). Cosa que ya hemos demostrado previamente. <u>En otras palabras, la eliminación de variables disminuye las variancias de los estimadores de los parámetros del modelo reducido.</u> (+)

Como b_i es un estimador sesgado de β_i y b_i^* no lo es, una mejor

---

1/ Véase, Hocking, R.R. (1974). Opus cit.

(+) Véase la discusión del Apartado 3.3 y específicamente, la expresión (56).

comparación de la precisión de los estimadores, se puede obtener comparando el error cuadrático medio (MSE) de b_i con las varianzas de b_i^*. Entonces tendremos una forma alternativa de expresar la condición (257) como sigue:

$$MSE(b_i) \leq Var(b_i^*) \qquad (258)$$

la cual es válida siempre que las variables eliminadas posean estimadores con valores más pequeños en magnitud (o valor absoluto) que los correspondientes valores de sus desviaciones estándard.(Esto es, siempre que las variables eliminadas sean no-significativas). Nótese de paso que el estimador de σ_u^2 correspondiente al modelo reducido está sesgado hacia arriba.[1/]

Bajo este orden de ideas, revisaremos brevemente el efecto de la eliminación de variables en la predicción. Como consecuencia de nuestra discusión anterior, se desprende que el estimador $\hat{Y}$ es sesgado al menos que las variables eliminadas posean parámetros

---

1/ Véase, Hocking, R.R. (1974). Opus cit.

nulos o que las variables retenidas sean ortogonales respecto al conjunto de variables removidas.

La variancia de $\hat{Y}$, o sea del valor pronosticado del modelo (258) por consiguiente, es menor o igual que la variancia del valor pronosticado de $\hat{Y}^*$ del modelo (1). Es decir:

$$Var(\hat{Y}) \leq Var(\hat{Y}^*) \qquad (259)$$

La expresión equivalente a (257) para las variancia de $\hat{Y}$ y $\hat{Y}^*$ será:

$$MSE(\hat{Y}) \leq Var(\hat{Y}^*) \qquad (260)$$

Entonces la justificación para la selección de variables puede resumirse en los siguientes términos: Aún cuando las variables eliminadas tengan parámetros no-nulos, los parámetros de las variables retenidas en el modelo reducido pueden estimarse con menor variancia que siendo estimados dentro del modelo completo. Y esto también se cumple para la variancia del valor pronosticado.

El precio pagado por la eliminación de variables es la introducción de sesgo en los estimadores del modelo reducido. No obstante, existen condiciones, como hemos discutido anteriormente (en el análisis de cordillera), donde el error cuadrático medio (MSE) de los estimadores resulta menor que la variancia de los estimadores insesgados. Es decir, la ganancia en precisión no resulta afectada por el cuadrado del sesgo. Por otra parte, si alguna de las variables retenidas en la ecuación son irrelevantes o no-esenciales, esto es, tienen parámetros nulos menores que los valo res de sus respectivas desviaciones estándard, entonces la retención de dichas variables en la ecuación referida genera una pérdida de precisión tanto en la estimación como en la predicción.

Antes de pasar a discutir el método de selección de variables propiamente dicho, conviene hacer dos consideraciones previas;

Primeramente, normalmente no tiene mucho sentido hablar del "mejor conjunto" de variables explicativas por incluir en un modelo de regresión.Porque simplemente no existe un "mejor conjunto" de variables. Una ecuación de regresión puede servir para varios propósitos.El conjunto de variables que resulta el "mejor" para un propósito puede no serlo para otro.Luego entonces, el propósito para el cual se construye una ecuación debe tenerse siempre presente durante el proceso de selección de variables. Veremos más adelante que dicho propósito determina los criterios para seleccionar y evaluar las contribuciones de las diferentes variables.

Segunda, como no hay un "mejor conjunto" de variables, existen varios subconjuntos de éstas que pueden ser adecuados y ser utilizados para formar una ecuación.Un buen procedimiento de selección de variables debe enfocarse a determinar varios subconjuntos y no concretarse a generar un único "mejor" conjunto. Los diversos subconjuntos de variables adecuados nos proporcionan información sobre la estructura de los datos y nos ayudan a comprender el proceso subyacente en los mismos. De hecho,el proceso de selección de variables debe contemplarse como un análisis intensivo de la estructura correlacional de las variables independientes y cómo afectan, individual y conjuntamente, a la variable dependiente bajo estudio. Esta es la razón por la cual hemos considerado importante citar este método en conexión con la corrección de multicolinealidad en un modelo de regresión. De cualquier manera, los dos puntos anteriores afectan la metología que presentaremos a continuación en relación con la selección de variables.

Para comenzar, recordaremos que una ecuación de regresión puede orientarse a varios usos alternativos. Estos pueden resumirse en tres grandes categorías:

i) Para la descripción de un fenómeno y formulación de un modelo

ii) Para propósitos de estimación y predicción

iii) Como instrumento de control

En el primer caso (i), la ecuación se utiliza para describir un proceso particular o como un modelo para expresar un sistema de interacción complejo. El propósito de la ecuación puede ser meramente descriptivo para esclarecer la naturaleza de esta interacción compleja. Sin embargo, para este fin existen dos requerimientos básicos en conflicto que deben cumplirse: a) Explicar la mayor parte de la variación total de la variable-respuesta, que implica por lo regular la inclusión de un gran número de variables y, b) apegarse al principio de "simplificación" que recomienda que, en obvio de entendimiento, debemos buscar describir el proceso con el mínimo número de variables posible. En las ocasiones en donde la descripción sea nuestro objetivo central, trataremos de seleccionar el mínimo número posible de variables independientes que permitan explicar la parte más sustencial de la variación de la variable dependiente.

En el segundo caso (ii), una ecuación se construye algunas veces para propósitos de predicción. O sea, a partir de la ecuación de regresión, se busca predecir el valor de una observación futura o estimar la respuesta-promedio correspondiente a una observación dada. Cuando una ecuación se utiliza para este propósito, las variables se eligen buscando minimizar el error cuadrático medio (MSE) de predicción.

Por último, en el caso (iii) una ecuación puede emplearse como una herramienta de control. El propósito de construir una ecuación de este tipo puede ser determinar la magnitud por la que debemos alterar el valor de una cierta variable independiente para obtener un valor específico correspondiente en la variable dependiente (o variable-respuesta objeto). Aquí, el modelo de regresión se concibe como una función-respuesta, donde Y es la variable-respuesta. Para propósitos de control, es deseable que los estimadores de los coeficientes del modelo en cuestión se midan con precisión, es decir, que sus errores estándard sean pequeños.

Sin embargo, y como ya lo hemos señalado, las categorías antes mencionadas para los usos de un modelo de regresión o econométrico son sumamente amplias. Frecuentemente éstas se entrelazan y una ecuación puede utilizarse para uno o todos los usos referidos. Lo que conviene enfatizar es que el propósito central para el cual se construye la ecuación, determina el criterio que se buscará optimizar en su formulación. Se desprende de esto, y como ya lo habíamos destacado anteriormente, qud un subconjunto de variables que puede ser el mejor para un propósito, puede no serlo para otro. El concepto del "mejor" subconjunto de variables por incluirse en una ecuación, requiere de una calificación específica. A continuación se describen brevemente algunos criterios para comprender mejor este concepto.

Para juzgar la adecuación de las diversas ecuaciones estimadas necesitamos de criterios estadísticos precisos. Muchos de éstos se describen en la literatura estadística. Discutiremos brevemente dos que consideramos de los más útiles.

Una medida que se utiliza a menudo para juzgar la adecuación de una ecuación estimada, es el cuadrado promedio de los residuales (RMS). En una ecuación de k variables explicativas, esto --

1/ Una lista exhaustiva de estos criterios se puede encontrar en la fuente citada a continuación:
Véase, Hocking, R.R. The Analysis and Selection of Variables in Linear Regression. Biometrics, 32, 1976, pp.1-49.

estadístico se define como: 1/

$$(RMS)_k = \frac{VAN}{n-k-1} \qquad (261)$$

donde VAN se define como:

$$VAN = \sum_{i=1}^{n} (Yi - \hat{Y}i)^2 \qquad (262)$$

Entonces dada dos ecuaciones, aquella **cuyo RMS sea el menor, se** escoge en especial si el objeto es la extrapolación. Nótese de paso que el estadístico RMS no es otra cosa que el estimador para la variancia del término de perturbación definido en (20-bis).

Es evidente que RMS se puede relacionar con el coeficiente de determinación R^2 y el coeficiente de determinación "ajustado" R_a^2 que como ya sabemos, son estadísticos que nos sirven para juzgar la adecuación del ajuste global de un modelo.(+) La relación entre RMS y R^2 es tal que:

$$R^2 = 1 - (n-k-1)\frac{RMS}{VAT} \qquad (263)$$

donde VAT (variación total) se define como:(++)

$$VAT = \sum_{i=1}^{n} (Yi - \bar{Y})^2 \qquad (264)$$

1/ RMS: son las siglas en inglés de 'Residual Mean Square'. Véase, Hocking, R.R. Opus cit.

(+) R_a^2 se define como $R_a^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1-R^2)$, donde n y k se definen como antes.

(++) En un modelo de regresión cualquiera se cumple que: VAT = VAE + VAE (Variación total= variación explicada + variación no-explicada), donde VAN y VAT se definen como en (262) y (264), respecitvamente, y VAE como VAE $= \sum(\hat{Y}i - \bar{Y})^2$.

Similarmente, se puede hacer ver que:

$$R_a^2 = 1 - (n-1)\frac{RMS}{VAT} \qquad (265)$$

Por otra parte, hemos señalado anteriormente que los valores pronosticados de una ecuación de regresión basados en un subconjunto de variables son generalmente sesgados. Para juzgar el desempeño de una ecuación consideramos el error cuadrático medio (MSE) del valor pronosticado en lugar de la variancia.

Entonces a partir del concepto de MSE podemos definir un nuevo estadístico (que denotaremos como "J_k") para medir el error cuadrático medio total estandarizado de predicción para los datos observados, o sea:

$$J_k = \frac{1}{\sigma_u^2}\sum_{i=1}^{n} MSE(\hat{y}_i) \qquad (266)$$

o alternativamente:

$$J_k = \frac{1}{\sigma_u^2}\left[MSE(\hat{y}_1) + MSE(\hat{y}_2) + \ldots + MSE(\hat{y}_n)\right) \qquad (267)$$

donde MSE ($\hat{y}_i$) es el error cuadrático medio del i-ésimo valor pronosticado $\hat{y}_i$ de un modelo de k variables explicativas y σ_u^2 es la variancia del término de perturbación.

Entonces el término MSE($\hat{y}_i$) tiene dos componentes, la variancia de predicción que resulta de la estimación (i.e. Var ($\hat{y}_i$)) y una componente de sesgo resultante de la eliminación de las variables.(+)

(+) Este término es de la forma general: $[E(\hat{y}_i^*) - y_i]^2$; donde $\hat{y}_i^*$ es el estimador de y_i del modelo del que se han eliminado un cierto número de variables.

En particular, se puede demostrar que J_k se puede expresar como: 1/

$$J_k = V_k + \frac{B_k}{\sigma_u^2} \tag{268}$$

donde el primer y segundo término del miembro derecho de la expre-sión anterior corresponden, respectivamente, a las componentes de varianza y sesgo. Específicamente, se puede hacer ver que: 2/

$$J_k = \frac{E(VAN)}{\sigma_u^2} + [2(k+1) - n] \tag{269}$$

Para estimar J_k podemos usar el estadístico C_k definido como:

$$C_k = \frac{VAN}{\hat{\sigma}_u^2} + [2(k+1) - n] \tag{270}$$

donde $\hat{\sigma}_u^2$ es un estimador de σ_u^2, y usualmente, se obtiene del modelo original (con k variables). Se puede demostrar, según Mallows 3/ que el valor esperado de C_k es $(k+1)$, es decir:

$$E(C_k) = k + 1 \tag{271}$$

cuando no existe ningún sesgo en la ecuación estimada del modelo

---

1/ Para la demostración de esta propiedad:
Véase, Mallows, C.L. Some Comments on C_p. Technometrics, 15, 1973, pp. 661-75

2/ Ibid.

3/ Véase, Mallows, C.L. Some Comments on C_p. Ibid.

original con k variables. Por consiguiente, la desviación de C_k respecto de $(k + 1)$ puede utilizarse como una medida de sesgo.

Luego entonces, el estadístico C_k mide el desempeño de las variables en términos del error cuadrático medio estandarizado de predicción (J_k), y toma en cuenta, tanto el sesgo como la variancia correspondiente.

Entonces aquellos subconjuntos de variables que producen valores de C_k cercanos a $(k+1)$ son los subconjuntos deseables. La selección de "buenos" subconjuntos se realiza gráficamente. Para este propósito se construye una gráfica de C_k contra $(k+1)$ para los distintos subconjuntos de variables. En especial, se traza la línea $C_k = k + 1$ en dicha gráfica. Entonces aquellos subconjuntos de variables correspondientes a puntos cercanos a la línea $C_k = k + 1$, serán los subconjuntos "buenos" o "deseables" para integrar la ecuación.1/ El uso de las gráficas de C_k se ilustra y discute con mayor detalle en el ejemplo que se proporciona más adelante en este mismo apartado. Un estudio más profundo sobre el estadístico C_k (ó "C_p") puede encontrarse en Daniel y Wood 2/.

1/ Véase, Chatterjee, S, y B.Price. Opus cit. p.199

2/ Véase, Daniel, C. y F.S. Wood, Fitting Equations to Data. Wiley. New York, 1971.

Para poder discutir los procedimientos de selección de variables debemos distinguir entre dos grandes y diferentes situaciones:

1. Que las variables independientes no sean colineales, o sea, que no exista evidencia de multicolinealidad peligrosa.

2. Que las variables independientes sea colineales, esto es,que los datos sean altamente colineales.

Dependiendo de la estructura de correlación de las variables independientes, proponemos diferentes enfoques para el procedimiento de selección de variables.Si los datos analizados no son colineales procederemos de una manera, y por el contrario, si lo son, procederemos de otra.

Como primera providencia en el procedimiento de selección de variables se recomienda calcular las raíces características de la matríz de correlación, a partir de las variables independientes. Como ha sido explicado en el Apartado 4.3, la presencia de raíces características pequeñas indica la presencia de multicolinealidad . Además de observar las raíces características individuales también debemos fijarnos en la suma de los recíprocos de las las raíces características. Hemos mencionado que si alguna de las raíces características es menor que 0.01 ó la suma de los recíprocos de las mismas es mayor que digamos, cinco veces el número de variables explicativas en el modelo,entonces las consideramos como si fueran 'colineales'.Si estas condiciones no se cumplen, entonces las variables se toman como 'no-colineales.'

A continuación estudiaremos dos grandes categorías de procedimientos sobre selección de variables: A) Procedimiento de evaluación de todas las ecuaciones posibles y B) Procedimientos de selección de variables, por pasos.(+)

A. Procedimiento de Evaluación de todas las Ecuaciones Posibles.

El primer procedimiento que abordamos es muy directo y se aplica muy bien tanto a datos colineales como a los no-colineales. Este implica estimar todos los subconjuntos posibles de ecuaciones (derivados del total de variables explicativas del modelo original) utilizando para ello un mismo conjunto de datos. Si el modelo es de k variables, entonces el número de ecuaciones por estimar será de 2^k. (++)

O sea, en este espectro de ecuaciones se incluyen a aquellas que contienen a las k variables y aquella que no contiene a ninguna variable explicativa. (Para esta última se tiene simplemente que : $Y_i = \bar{Y}$).

(+) Para el lector interesado, su nombre en inglés: Selection of Variables, Stepwise Procedures.

(++) De la teoría de conjuntos, sabemos que si un conjunto A tiene n elementos, entonces el número total de subconjuntos de A es 2^n, incluyendo al vacío y a él mismo.

Este método proporciona claramente al investigador la máxima cantidad de información posible respecto a la naturaleza de las relaciones entre Y y el conjunto de las X'S. Sin embargo, el número de ecuaciones o informaciones suplementaria o adicional que debe analizarse, resultan prohibitivas. Puesto que para un modelo de sólo 6 variables, tendríamos que estimar y revisar $2^6 = 64$, y para un modelo de 7 u 8 variables explicativas, tendríamos queconsiderar $2^7=128$ y $2^8= 256$ ecuaciones, respectivamente, lo cual no resulta factible ni práctico, desde cualquier ángulo que se le mire.

Por consiguiente, cuando se utiliza este método, las ecuaciones más prometedores se "aíslan" mediante los estadísticos C_k ó RMS. Las ecuaciones aisladas son posteriormente analizadas a través de un examen de sus observaciones "desfasadas"(+), o verificando la posible existencia de autocorrelación, o bien, estudiando la necesidad de transformar dichas variables antes de decidir finalmente qué subconjunto de variables que se analizan puedan sugerir interpretaciones de los datos que pueden haber sido omitidas en un procedimiento más estricto de selección de variables.

A pesar de que cuando el número de variables es grande y la evaluación de todas las ecuaciones posibles puede no ser prácticamente factible, existen en la actualidad ciertos métodos abrevia

---

(+) Estas observaciones son aquellas cuyos residuales estandarizados se encuentran o localizan "fuera de foco" dentro de la gráfica de residuales e_i^* contra $\hat{Y}i$.

dos ("atajos") que no requieren de calcular el conjunto total de ecuaciones en el proceso de búsqueda de los subconjuntos deseables y facilitan en gran medida la tarea de selección.[1/] No — obstante, cuando el número de variables crece, aún estos procedimientos requieren todavía una considerable cantidad de cómputo.

Sin embargo, existen otros procedimientos o métodos de selección de variables que no implican la evaluación de todas ecuaciones posibles y que analizaremos enseguida. Aunque conviene resaltar de entrada que estos procedimientos no proporcionan al investigador o analista con tanta información que estimando todas las ecuaciones posibles, le permiten reducir el volumen de cómputo requerido y puede representar en un momento dado la única solución práctica. Estos procedimientos se conocen genéricamente como "Procedimientos de Selección de Variables por Pasos" y resultan muy eficientes para trabajar con datos no-colineales. En consecuencia, no se recomienda su uso para datos colineales. A pesar de ello, consideramos conveniente y oportuno su discusión aquí por su utilidad en el proceso de "depuración" de variables de un modelo afectado con multicolinealidad, sobre todo cuando K es grande.

1/ Véase, La Motte, L.R. y R.R.Hocking. Computational Efficiency in the Selection of Regression Variables. Technometrics, 12, 1970, pp. 83-93. Asimismo,

Véase, Furnival, G.M. y R.W. Wilson Jr. Regression By Leaps — and Bounds. Technometrics, 16, 1974, pp. 449-512

B. Procedimientos de Selección de Variables por Pasos.

Estos procedimientos se aplican cuando existe un gran número de variables explicativas y se trabaja con un subconjunto propio del total de variables explicativas. Tienen la característica primordial de que introducen o eliminan una variable a la vez de la ecuación e implican examinar sólo un subconjunto de variables del universo total de subconjuntos de los k variables explicativas originales. Esto es, para un modelo de k variables, implicará considerar a lo más (k+1) ecuaciones, en contraste con las 2^k ecuaciones requeridas para examinar todas las ecuaciones posibles. Estos procedimientos se subclasifican en dos grandes categorías: B.1) El procedimiento de selección de variables (por pasos) hacia adelante (FS) (+) y B.2) El procedimiento de selección de variables (por pasos) por eliminación hacia atrás (BE)(++).

B.1) Procedimiento de Selección de Variables (Por Pasos) Hacia Adelante (FS).

Este procedimiento comienza con una ecuación que no contiene ninguna variable explicativa, sólo el término constante. La primera variable en incluirse en la ecuación es aquella con el más alto coeficiente de correlación simple con respecto a la variable dependiente "Y". Si el coeficiente de esta variable resulta significativamente distinto de cero, entonces se retiene en la ecuación. (De lo contrario, se toma la segunda variable con el más alto coeficiente de correlación simple, etc. y así sucesivamente,

(+) En inglés: FS= 'Forward Selection Procedure'

(++) En inglés: BE= 'Backward Elimination Procedure' hasta encontrar una variable cuyo coeficiente sea significativamente distinto de cero.

La variable que ingresa a la ecuación como "segunda" variable es aquella que posee el máximo coeficiente de correlación simple con respecto a "Y(1)" (donde "Y(1)", es el resultado de ajustar a Y por el efecto de la primera variable ,o sea, la variable con el más grande coeficiente de correlación con respecto a los elementos residuales del paso anterior). Entonces, se prueba la significancia del coeficiente de regresión de la segunda variable. Si éste resulta significativo, se busca una tercera variable de la misma manera. Si no lo es, se sigue el mismo procedimiento sugerido en el paso anterior.

El procedimiento termina cuando la última variable que ingresa a la ecuación posee un coeficiente de regresión no-significativo, o bien, cuando todas las variables se han incluido en la ecuación. La significancia del coeficiente de regresión de la última variable introducida en la ecuación se verifica mediante el estadístico "t" calculado a partir de la última ecuación. A este respecto, conviene señalar que la mayoría de los algoritmos de selección de variables (por pasos) hacia adelante usan un valor "piso" de "t" muy bajo para juzgar la significancia estadística de los coeficientes de las nuevas variables que ingresan a la ecuación, y por lo tanto, los procedimientos de selección hacia adelante consideran todas las variables explicativas y nos proporcionan un total de k ecuaciones posibles.

B.2 Procedimientos de Selección de Variables (por pasos) por Eliminación hacia Atrás (BE).

Este procedimiento principia con la ecuación completa (de k variables), y sucesivamente, las va eliminando una a la vez. Las variables se eliminan en base a su contribución a la reducción del estadístico VAN (i.e. la suma de los cuadrados de los errores definido en (262)). Esto es equivalente a eliminar la variable con el mínimo valor "t" (i.e. el cociente del valor del estimador del coeficiente de regresión de la variable en cuestión entre el valor de la desviación estándard de dicho estimador) de entre todas las variables contenidas en la ecuación original.

Si todos los valores "t" resultan significativos en el proceso anterior, entonces se retienen todas las variables de la ecuación. Suponiendo que exista una o más variables con valores "t" no-significativos, entonces el procedimiento sugiere eliminar aquella variable con el mínimo valor "t" no-significativo. Entonces se estima una nueva ecuación con las restantes (k-1) variables y se vuelven a examinar los valores "t" para los nuevos coeficientes dd regresión. El procedimiento finaliza cuando todos los valores "t" son significativos o bien cuando todas las variables, excepto una, han sido eliminadas.

En la mayoría de los algoritmos hacia atrás, la cota inferior de "t" se fija tan alta que el procedimiento analiza todas las va--

riables, es decir, empieza con una ecuación de k variables para luego terminar con una ecuación de una sola variable explicativa. O sea, el procedimiento de eliminación hacia atrás implica estimar a lo más, k ecuaciones posibles.

B.3 <u>Método de Selección de Variables por Pasos.</u>(+)

Este método es esencialmente un procedimiento de selección "hacia adelante" (FS) pero con la condición adicional que, en cada etapa, se considera la posibilidad de eliminar una variable como en la eliminación "hacia atrás" (BE).

En este método, una variable que ingresa en las primeras etapas puede llegar a ser eliminada en etapas posteriores. Los criterios para la inclusión o exclusión de las variables son los mismos que los usados en los procedimientos "hacia adelante" y "hacia atrás". Frecuentemente, se utilizan diferentes niveles de significancia para determinar la inclusión o exclusión de variables de la ecuación.

Sin embargo, y para terminar con la discusión sobre este particular, conviene resaltar que en general, los "procedimientos -- por pasos" descritos anteriormente deben ser utilizados con precaución. No deben, por tanto, usarse mecánicamente para determinar el "mejor" subconjunto de variables. El orden en que las va--

(+) En inglés: 'Stepwise Method" .

riables ingresan o salen de la ecuación no debe interpretarse como producto de la importancia relativa de las variables. Si estas recomendaciones se toman en cuenta, los "procedimientos por pasos" representan herramientas poderosas para la selección de variables no-colineales.

Los tres procedimientos descritos anteriormente arrojan prácticamente la misma selección de variables para datos no-colineales. Y naturalmente, implican mucho menos cómputo que el análisis de todas las ecuaciones posibles.

Se han diseñado diversas reglas empíricas de "alto" o "parada" para los procedimientos por pasos. Una regla de "alto" que en la práctica ha resultado muy efectiva se describe enseguida: 1/

Para el Procedimiento Hacia Adelante:

Detenerse si el mínimo valor de "t" es tal que $|t| < 1$

Para el Procedimiento Hacia Atrás:

Detenerse si el mínimo valor de "t" es tal que $|t| > 1$

En el ejemplo que se cita más abajo, se ilustrará el efecto de estas dos reglas de "alto" para la selección de variables. Sin embargo, se recomienda en general el procedimiento "hacia atrás" sobre el procedimiento "hacia adelante". Una razón obvia de esto

1/ Véase, Chatterjee, S. y B. Price. Opus cit.

es que en el procedimiento "hacia atrás," la ecuación con todas las variables, se estima y se puede examinar aún cuando no se utilice como ecuación final. Aunque no se recomienda el uso de los "procedimientos por pasos" en casos de datos colineales se ha podido comprobar empíricamente que el procedimiento "hacia atrás" resulta mejor para manejar el problema de la multicolinealidad que el procedimiento "hacia adelante".[1/]

En las aplicaciones de los procedimientos por pasos, se generan varias (o muchas) ecuaciones, cada una con un número distinto de variables. Las diversas ecuaciones generadas pueden evaluarse utilizando estadísticos como C_k ó RMS. De la misma manera, los residuales de las diferentes ecuaciones deben examinarse. Aquellos con gráficas de residuales "no satisfactorias" deberán rechazarse. Solo mediante un análisis total y comprensivo se podrá efectuar una selección adecuada de variables y obtener una ecuación de regresión o modelo útil. Este enfoque de selección de variables se ilustra a través del siguiente ejemplo:

Se requiere formular un modelo para estudiar las cualidades que permitan la caracterización de buenos supervisores por el personal que es supervisado por éstos. El modelo se construye con el propósito de entender el proceso de supervisión y la importancia relativa de las diferentes variables explicativas. En términos de los usos de una ecuación de regresión o modelo económé--

---

1/ Véase, Mantel, N. Why Stepdown Procedures in Variable Selection. Technometrics, 12, 1970, pp. 591-612 trico, este caso implicaría que deseamos conocer los estimadores de los diferentes coeficientes de regresión, en contraste con un modelo utilizado sólo con propósitos de predicción. El modelo en cuestión, tiene la forma: 1/

$$Y = Bo + B1\ X1 + B2\ X2 + B3\ X3 + B4\ X4 + B5\ X5 + B6\ X6 + U \quad (272)$$

donde Y y Xi (i=1,2,...,6) se definen como:

Y : Calificación global del trabajo realizado por el Supervisor.

X1 : Maneja adecuadamente las quejas de los empleados.

X2 : No concede privilegios especiales a subalternos.

X3 : Brinda oportunidades al personal a su mando para aprender labores nuevas.

X4 : Promueve ascensos o promociones entre su personal conforme a desempeño individual.

X5 : Es demasiado crítico respecto a desempeños pobres.

X6 : Grado de estímulo o apoyo a subalternos para el logro de mejores puestos.

Los datos para el modelo anterior se presentan en la Tabla 5.6.1 (que se reproduce en la próxima hoja). De igual manera, la matríz de correlación para dicho modelo se presenta en la Tabla 5.6.2 (que aparece inmediatamente después de la Tabla 5.6.1). Por otro la

---

1/ Véase, Chatterjee, S. y B. Price. Opus cit.

TABLA 5.6.1

DATOS SOBRE LAS VARIABLES EXPLICATIVAS Y EXPLICADA DEL MODELO (272).

Hilera	Y	X1	X2	X3	X4	X5	X6
1	43	51	30	39	61	92	45
2	63	64	51	54	63	73	47
3	71	70	68	69	76	86	48
4	61	63	45	47	54	84	35
5	81	78	56	66	71	83	47
6	43	55	49	44	54	49	34
7	58	67	42	56	66	68	35
8	71	75	50	55	70	66	41
9	72	82	72	67	71	83	31
10	67	61	45	47	62	80	41
11	64	53	53	58	58	67	34
12	67	60	47	39	59	74	41
13	69	62	57	42	55	63	25
14	68	83	83	45	59	77	35
15	77	77	54	72	79	77	46
16	81	90	50	72	60	54	36
17	74	85	64	69	79	79	63
18	65	60	65	75	55	80	60
19	65	70	46	57	75	85	46
20	50	58	68	54	64	78	52
21	50	40	33	34	43	64	33
22	64	61	52	62	66	80	41
23	53	66	52	50	63	80	37
24	40	37	42	58	50	57	49
25	63	54	42	48	66	75	33
26	66	77	66	63	88	76	72
27	78	75	58	74	80	78	49
28	48	57	44	45	51	83	38
29	85	85	71	71	77	74	55
30	82	82	39	59	64	78	39

Fuente: Chatterjee, S. y B. Price. Opus cit. p. 61.

TABLA 5.6.2

MATRIZ DE CORREALCION PARA EL MODELO (274)

	X1	X2	X3	X4	X5	X6
X1	1.000					
X2	0.558	1.000				
X3	0.597	0.493	1.000			
X4	0.609	0.445	0.640	1.000		
X5	0.188	0.147	0.116	0.377	1.000	
X6	0.225	0.343	0.532	0.574	0.283	1.000

Fuente: Estimaciones a partir de los datos de la Tabla 5.6.1

la matríz Λ para este modelo resulta ser:

$$= \begin{bmatrix} 3.236 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.033 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.750 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.535 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.253 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.193 \end{bmatrix} \quad (273)$$

Esto es, las raíces características correspondientes para el modelo bajo estudio son:

$$\lambda_1 = 3.236, \quad \lambda_2 = 1.033, \quad \lambda_3 = 0.750, \quad \lambda_4 = 0.535$$
$$\lambda_5 = 0.253, \quad \lambda_6 = 0.193 \quad (274)$$

Como observamos, ninguna de las raíces es menor que .01, ni la suma de sus recíprocos:

$$\sum_{i=1}^{6} \frac{1}{\lambda_i} = \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + \frac{1}{\lambda_4} + \frac{1}{\lambda_5} + \frac{1}{\lambda_6} \right) = 13.624 \quad (275)$$

es una cifra tal que no exceda a 30 (es decir, cinco veces el número de variables explicativas en el modelo (272)) concluímos por tanto que los datos de este modelo no son necesariamente colineales y por lo mismo, podemos aplicar los procedimientos por pasos anteriormente discutidos.

Los resultados sumarios de aplicar el procedimiento (por pasos) hacia adelante, se presentan en la Tabla 5.6.3 (en la próxima hoja). En donde se detallan para cada una de las ecuaciones sucesivas,

TABLA 5.6.3

SELECCION DE VARIABLES MEDIANTE EL PROCEDIMIENTO DE SELECCION HACIA ADELANTE, PARA EL MODELO (272)

VARIABLE EXP. EN LA ECUACION	(K+1)	RMS	Ck	Rango
X1	2	6.993	1.41	1
X1 X3	3	6.817	1.11	1
X1 X3 X6	4	6.734	1.60	1
X1 X3 X6 X2	5	6.820	3.28	1
X1 X3 X6 X2 X4	6	6.920	5.07	1
X1 X3 X6 X2 X4 X5	7	7.068	7.00	1

Fuente: Estimaciones a partir de los datos de la Tabla 5.6.1

Los valores de los estadísticos RMS y C_k .La última columna — muestra el rango del subconjunto obtenido mediante el procedimiento (por pasos) "hacia adelante", respecto al mejor subconjunto del mismo tamaño (en base al estadístico RMS).

A continuación, se proporcionan dos criterios empíricos de "alto" aplicados para este problema: 1/

1.Detenerse si el valor mínimo de "t" es tal que: $\lvert t_c \rvert < t_{.05}(n-k-1)$

Es decir, parar si el valor mínimo de todas las t's calculadas resulta menor que su valor tabulado para un nivel de $\alpha = 5\%$ de significancia, con $[n-(k-1)]$ grados de libertad.

2.Detenerse si el valor mínimo de "t" es tal que: $\lvert t_c \rvert < 1$

O sea, parar si el valor mínimo de todas las t's calculadas resulta menor que la unidad.

Nótese que el primer criterio es más estricto que el segundo y termina con la selección de las variables X1 y X3. El segundo es menos estricto y finaliza con la inclusión de las variables X1, X3 y X6.

Por otra parte, la aplicación del procedimiento de selección por eliminación "hacia atrás" nos arroja los resultados presentados en

---

1/ Obsérvese que estos mismos criterios se aplican para el procedimiento "hacia adelante".

Véase, Chatterjee,S. y B.Price.Opus cit.

la Tabla 5.6.4 (que se muestra en la hoja siguiente) y que resulta idéntica, en estructura, a la Tabla 5.6.3.

Para este procedimiento se utilizaron los siguientes criterios de "alto": 1/

1. Detenerse si el valor mínimo de "t" es tal que: $\lvert t_c \rvert > t_{.05}(n-k-1)$
2. Detenerse si el valor mínimo de "t" es tal que: $\lvert t_c \rvert > 1$

Con el primer criterio se seleccionan las variables X1 y X3. Y con el segundo, las variables X1, X3 y X6. Los procedimientos "hacia adelante" y "hacia atrás" arrojan el mismo subconjunto de ecuaciones para este problema, aunque este no es siempre el caso.(+)

En consecuencia, para describir la variable-respuesta objeto del modelo (272) se recomienda la siguiente ecuación:

$$\hat{Y} = 13.58 + 0.62X1 + 0.31X3 - 0.19X6 \qquad (276)$$

Si graficaramos los residuales de esta ecuación, observaríamos que su comportamiento es satisfactorio.(++)

1/ Véase, Chatterjee, S. y B. Price. Opus cit.

(+) Si se aplicara el método de selección de variables por pasos (B.3) discutido anteriormente, se observaría que se obtendría el mismo resultado. Se omiten los resultados en obvio de simplificacion de esta exposición.

(++) Se omite la gráfica correspondiente por simplicidad de la exposición.

TABLA 5.6.4

SELECCION DE VARIABLES MEDIANTE EL PROCEDIMIENTO DE SELECCION POR ELIMINACION HACIA ATRAS, PARA EL MODELO (272)

VARIABLES EXP. EN LA ECUACION	(K+1)	RMS	Ck	RANGO
X1 X2 X3 X4 X5 X6	7	7.068	7.00	-
X1 X2 X3 X4 X6	6	6.928	5.07	1
X1 X2 X3 X6	5	6.820	3.28	1
X1 X3 X6	4	6.734	1.60	1
X1 X3	3	6.817	1.11	1
X1	2	6.993	1.41	1

Fuente: Estimaciones a partir de los datos de la Tabla 5.6.1

Como el presente problema tiene solo seis variables explicativas, el número total de ecuaciones que pueden estimarse conteniendo al menos una variable explicativa es 63 (i.e. $2^6 - 1$). Los valores de C_k para las 63 ecuaciones se muestran en la Tabla 5.6.5 (que aparece en la próxima hoja).Los valores del estadístico C_k se grafican contra los valores de $(k + 1)$ en la Figura 5.6.1 (que se presenta enseguida de la Tabla 5.6.5).Los mejores sub-conjuntos de variables seleccionados en base a los valores del estadístico referido se muestran en la Tabla 5.6.6 (que aparece enseguida de la Figura 5.6.1).

Se observa que los subconjuntos de variables seleccionadas mediante C_k son distintos de aquellos obtenidos por los "procedimientos por pasos" así como aquellos seleccionados mediante el estadístico RMS. Esta discrepancia trae a colación un aspecto muy importante respecto al estadístico C_k que el lector debe tomar muy en cuenta. Cuando se aplica dicho estadístico,se necesita un estimador de σ_u^2 .Normalmente, el estimador de σ_u^2 se obtiene de la suma de los cuadrados de los residuales del modelo original o "completo". Si éste tiene un gran número de variables explicativas sin poder explicativo (i.e.sus parámetros poblacionales son cero), el estimador de σ_u^2 derivado del modelo original será grande.La pérdida de grados de libertad no es recuperada por la reducción en la variancián no-explicada.

Si el estimador $\hat{\sigma}_u^2$ es grande, entonces C_k es pequeña. Para que el estadístico C_k trabaje bien se requiere un buen estimador de σ_u^2.Cuando esto no es posible, C_k sólo es de utilidad limitada . Para el presente ejercicio , el ---

T A B L A 5.6.5

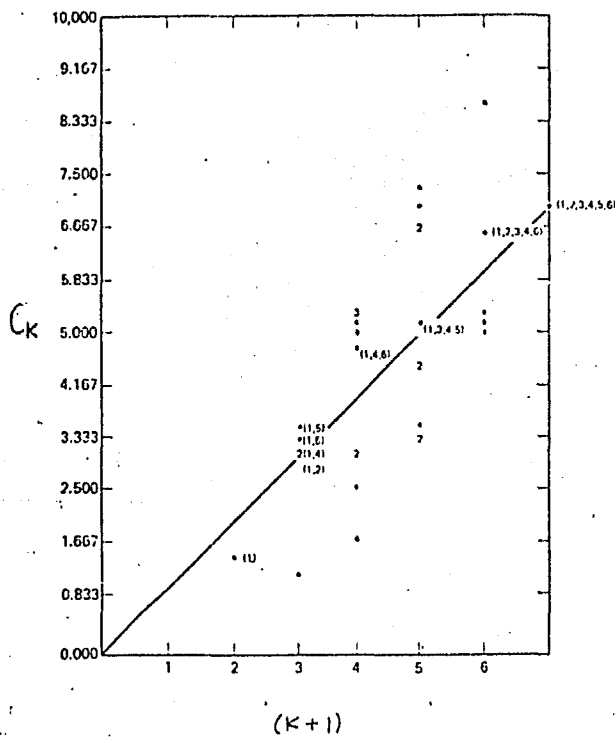
VALORES DEL ESTADISTICO Ck, PARA TODAS LAS ECUACIONES POSIBLES, PARA EL MODELO (272).

Variables	Ck	Variables	Ck	Variables	Ck	Variables	Ck
1	1.41	1 5	3.41	1 6	3.33	1 5 6	5.32
2	43.40	2 5	45.62	2 6	46.39	2 5 6	47.91
1 2	3.26	1 2 5	5.26	1 2 6	5.22	1 2 5 6	7.22
3	26.56	3 5	27.94	3 6	24.82	3 5 6	23.02
1 3	1.11	1 3 5	3.11	1 3 6	1.60	1 3 5 6	3.46
2 3	26.96	2 3 5	28.53	2 3 6	24.62	2 3 5 6	25.11
1 2 3	2.51	1 2 3 5	4.51	1 2 3 6	3.28	1 2 3 5 6	5.14
4	30.06	4 5	31.62	4 6	27.73	4 5 6	29.50
1 4	3.19	1 4 5	5.16	1 4 6	4.70	1 4 5 6	6.69
2 4	29.20	2 4 5	10.92	2 4 6	25.91	2 4 5 6	27.71
1 2 4	4.99	1 2 4 5	6.97	1 2 4 6	6.63	1 2 4 5 6	8.61
3 4	23.25	3 4 5	25.23	3 4 6	16.50	3 4 5 6	18.42
1 3 4	3.09	1 3 4 5	5.09	1 3 4 6	3.35	1 3 4 5 6	5.29
2 3 4	24.56	2 3 4 5	26.53	2 3 4 6	17.57	2 3 4 5 6	19.51
1 2 3 4	4.49	1 2 3 4 5	6.48	1 2 3 4 6	5.07	1 2 3 4 5 6	7.00
5	57.91	6	57.95	5 6	58.76		

Fuente; Estimaciones a partir de los datos de la Tabla 5.6.1

F I G U R A 5.6.1

GRAFICA DE LOS VALORES DEL ESTADISTICO Ck CONTRA EL PARAMETRO (k + 1),PARA EL MODELO (272).



Fuente: Chatterjee,S. y B. Price. Opus cit.p.207

TABLA 5.6.6

SELECCION DE VARIABLES MEDIANTE EL ESTADISTICO Ck
PARA EL MODELO (272)

VARIABLES EXP. EN LA ECUACION	(K+1)	RMS	Ck	RANGO
X1	2	6.993	1.41	1
X1 X4	3	7.093	3.19	2
X1 X4 X6	4	7.163	4.70	5
X1 X3 X4 X5	5	7.080	5.04	6
X1 X2 X3 X4 X5	6	7.139	6.48	4
X1 X2 X3 X4 X5 X6	7	7.068	7.00	-

Fuente: Estimaciones a partir de los datos de la Tabla 5.6.1

estadístico RMS para el modelo completo con seis variables es mayor que RMS para el modelo con tres variables X1, X2 y X6. Por consiguiente, los valores de C_k se distorsionan y no son muy útiles para la selección de variables en el presente caso.

El tipo de situación descrita anteriormente queda de manifiesto observando el comportamiento de los valores del estadístico RMS en las Tablas 5.6.3 y 5.6.4 (+). Este comportamiento sugiere que las últimas variables no están contribuyendo significativamente a la reducción de la variancia no-explicada (VAN). La aplicación útil del estadístico C_k requiere de un monitoreo paralelo del estadístico RMS para evitar posibles distorsiones.(++)

Procederemos a discutir ahora la selección de variables cuando existan datos colineales, o en otras palabras, cuando se presenta el problema de multicolinealidad entre las variables invo--

---

(+) En las Tablas 5.6.3 y 5.6.4 notamos que el estadístico RMS tiende a decrecer primeramente, para luego aumentar o crecer en etapas posteriores. De cualquier manera, para ambos casos, el subconjunto de variables que determinan el mínimo valor de RMS es el integrado por las variables X1, X2 y X3.

(++) Nótese de la Tabla 5.6.6 que el estadístico RMS primero aumenta para después disminuir (y que el valor del estadístico C_k tiende a crecer sistemáticamente en todos los casos: Tablas 5.6.3, 5.6.4 y 5.6.6). Obsérvese que el subconjunto de variables seleccionadas en base al estadístico C_k es $\{X1, X4, X6\}$ que difiere del subconjunto $\{X1, X2, X3\}$ seleccionado en base a los procedimientos "hacia adelante" (Tabla 5.6.3) y "hacia atrás" (Tabla 5.6.4) discutidos anteriormente.

lucradas.

Existen dos enfoques para manejar esta situación. El primero, que no recomendamos mucho, es el método de eliminación de variables.(+) El segundo, que sí se recomienda, es el que utiliza la regresión de cordillera ("Ridge Regression") como herramienta principal.

Ya hemos señalado con anterioridad (Apartado 4.4) que uno de los objetivos de la regresión de cordillera es producir o generar una ecuación de regresión o modelo con coeficientes estables. Estos son estables en el sentido de que no se ven afectados por ligeras variaciones en los datos, y por ende, son confiables y permiten la adecuada inferencia y predicción.

Por otra parte, según Chatterjee y Price 1/, los objetivos de un buen método o procedimiento de selección de variables son:

i) Seleccionar un conjunto de variables que faciliten o permitan un claro entendimiento del fenómeno o proceso bajo estudio.

ii) Formular una ecuación o modelo que permita obtener buenas predicciones o estimaciones de la variable-respuesta correspondiente a valores de las variables explicativas no necesariamente incluidas en el estudio (o mejor dicho, en la muestra bajo estudio).

(+) Salvo que se aplique bajo las circunstancias mencionadas en el Apartado 5.1 (Método de Eliminación de Variables) y previo análisis del costo de adoptar este procedimiento, en lugar del método de la restricción paramétrica, por ejemplo, que es menos peligroso y más poderoso, aunque un poco más tardado.

1/ Véase, Chatterjee, S. y B. Price. Opus cit.

Se observa pues, que los objetivos de un buen método de selección de variables y los del análisis de regresión de cordillera resultan muy similares, y en consecuencia, este último puede emplearse o aplicarse para lograr el primero, es decir, lograr una buena selección de variables.

La selección de variables se realiza examinando la "traza de cordillera" cuyas características y parámetros ya han sido introducidos en el Apartado 4.4 . La traza de cordillera se utiliza para eliminar las variables de la ecuación. Las reglas específicas para dicha eliminación son, según Hoerl y Kennard 1/:

1. Eliminar aquellas variables cuyos coeficientes son estables pero pequeños. Como el análisis de regresión de cordillera se aplica a datos estandarizados, las magnitudes de los diversos coeficientes son comparables directamente.

2. Eliminar aquellas variables con coeficientes inestables que "no sostienen" sus poderes de predicción, es decir, las que posean coeficientes inestables que tiendan a cero.

3. Eliminar una o más variables con coeficientes inestables. Las

1/ Véase, Hoerl, A.E. y R.W. Kennard. Ridge Regression: Applications to Nonorthogonal Problems. Opus cit.

<u>"r" variables que permanezcan del conjunto original de k variables, integrarán la ecuación final objeto.</u>

El subconjunto de variables (r) que quedan después del proceso de eliminación descrito anteriormente, deberá ser examinado para ver si constituye un conjunto ortogonal. Un método gráfico -- para realizar esto se obtiene graficando el término definido como: (+)

$$\sum_{i=1}^{r} \hat{\beta}_i^2(p) = \hat{\beta}1^2(p) + \hat{\beta}2^2(p) + \ldots + \hat{\beta}r^2(p) \qquad (277)$$

contra el parámetro "p". Nótese que en términos geométricos, la expresión anterior representa la distancia al cuadrado de los estimadores de cordillera respecto al origen. Se puede demostrar que para un sistema ortogonal esta distancia adopta la forma: 1/

$$\frac{\sum_{i=1}^{r} \hat{\beta}_i^2(o)}{(1+p)^2} = \frac{1}{(1+p)^2}\left(\hat{\beta}_1^2(o) + \hat{\beta}_2^2(o) + \cdots + \hat{\beta}_r^2(o)\right) \qquad (278)$$

donde el término $\hat{\beta}i(0)$ corresponde al estimador de mínimos cuadrados de la i-ésima variable y "p" es el parámetro de cordille-

---

(+) El término $\hat{\beta}i(p)$ es el estimador del "coeficiente de cordillera" de la i-ésima variable correspondiente al parámetro "p".

1/ Véase, Hoerl y Kennard. Opus cit.

ra. Entonces si las variables explicativas retenidas constitu-- yen o forman un conjunto aproximadamente ortogonal, las gráfi-- cas de las expresiones (277) y (278) serán prácticamente idénti cas.

Si las gráficas son idénticas se procede al paso siguiente del análisis, pero si son muy diferentes, es necesario eliminar va-- riables adicionales. Las variables eliminadas son aquellas con coe ficientes inestables. Este procedimiento debe conducir a un con-- junto aproximadamente ortogonal. La ecuación final se estima a-- plicando la regresión de cordillera a la ecuación completa en vez de aplicar OLS al modelo reducido. Esto es equivalente a descar-- tar las variables afectadas haciendo que sus valores sean igua-- les a sus promedios para todas las predicciones. A continuación se ilustra este procedimiento mediante un caso concreto de estu- dio.

Este estudio se debe a McDonald y Schwing[1/] y se refiere a la -- formulación de un modelo para explicar la mortalidad total de -- un cierto grupo humano de una región determinada, en términos de variables climáticas, socioeconómicas y de contaminación ambien- tal. En la Tabla 5.6.7 (de la próxima hoja) se enlistan y descri- ben un total de 15 variables que caen dentro de las tres grandes categorías antes referidas. La variable dependiente (Y) corres-- ponde a la "mortalidad ajustada", en términos de años, por todas -- las causas a que se refieren las variables explicativas en la ta

---

1/ Véase, McDonald, G.C. y R.C. Schwing. Instability of Regression Estimates Relating Air Pollution to Mortality. Technometrics, 15, 1973, pp. 463-81

T A B L A 5.6.7

DESCRIPCION DE LAS VARIABLES EXPLICATIVAS (MOSTRANDO SUS MEDIAS Y DESVIACIONES ESTANDARD) QUE CONFORMAN EL MODELO (279).

Número de Variable	D e s c r i p c i ó n :	Media	Desv.Est.
1	Precipitación Anual Promedio (en pulgadas).	37.37	9.98
2	Temperatura Promedio en Enero (en grados Fahrenheit).	33.98	10.17
3	Temperatura Promedio en Julio (en grados Fahrenheit).	74.58	4.76
4	Porcentaje de Población con -- más de 65 Años de Edad.	8.80	1.46
5	Número de Familias Existentes en la población Total objeto.	3.26	0.14
6	Promedio de Educación (en años cursados y terminados).	10.97	0.85
7	Porcentaje de Unidades Habita-- cionales en Buen Estado.	80.92	5.15
8	Población por Milla Cuadrada.	3,876.05	1,454.10
9	Porcentaje de Población que -- no es Blanca.	11.87	8.92
10	Porcentaje de Empleo en Actividades de tipo Admvo. o Burocrático.	46.08	4.61
11	Porcentaje de Familias con Ingreso Inferior a los 3,000 dólares.	14.37	4.16
12	Contaminación Relativa por Hidrocarburos.	37.85	91.98
13	Contaminación Relativa por Dióxido de Nitrógeno.	22.65	46.33
14	Contaminación Relativa por Dióxido de Sulfuro.	53.77	63.39
15	Porcentaje de Humedad Relativa.	57.67	5.37

Fuente: McDonald,G.C. y R.C.Schwing. Instability of Regression Estimates Relating Air Pollution to Mortality. Opus cit.

bla mencionada. El modelo en cuestión queda formulado como:

$$Y = \beta o + \beta 1 X1 + \beta 2 X2 + \ldots + \beta 15 X15 + U \qquad (279)$$

Conviene destacar, antes de proseguir, que no se pretende aquí entrar en la discusión de los aspectos epidemiológicos de este estudio ni mucho menos, sino solamente utilizar los datos de este modelo como un ejemplo ilustrativo para la selección de variables. Una discusión pormenorizada sobre el mismo, para el lector interesado, se encuentra en la referencia antes proporcionada.

La matríz de correlación para este modelo se presenta en la Tabla 5.6.8 (que aparece en la próxima hoja). Como era de esperarse, por la naturaleza de las variables involucradas, algunas son altamente colineales entre sí. La evidencia de multicolinealidad se observa claramente si examinamos las raíces características de la matríz de correlación para este modelo. Estas resultan ser:(+)

$\lambda 1 = 4.5272$	$\lambda 6 = .9605$	$\lambda 11 = .1605$
$\lambda 2 = 2.7547$	$\lambda 7 = .6124$	$\lambda 12 = .1275$
$\lambda 3 = 2.0545$	$\lambda 8 = .4729$	$\lambda 13 = .1142$
$\lambda 4 = 1.3487$	$\lambda 9 = .3708$	$\lambda 14 = .0460$
$\lambda 5 = 1.2227$	$\lambda 10 = .2163$	$\lambda 15 = .0049$

Observamos inmediatamente la existencia de dos raíces características muy pequeñas $\lambda 14 = .0460$ y $\lambda 15 = .0049$. También nota--

(+) Calculadas a partir de la matríz de correlación que aparece en la Tabla 5.6.8

TABLA 5.6.8

MATRIZ DE CORRELACION PARA LAS VARIABLES EXPLICATIVAS DEL MODELO (279)

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	.0922	.5033	.1011	.2634	-.4904	-.4903	-.0035	.4132	-.2973	.5066	-.5318	-.4373	-.1069	-.0773	.5095
2		.3463	-.3981	-.2097	.1163	.0139	-.1001	.4538	.2380	.5651	.3508	.3210	-.1078	.0673	-.0300
3			-.4340	.2623	-.2385	-.4155	-.0610	.5753	-.0714	.6193	-.3565	-.3177	-.0993	-.4526	.2770
4				-.5091	-.1339	.0549	.1620	-.4378	-.1177	-.1098	-.0205	-.0021	.0172	.1124	-.1746
5					-.3951	-.4095	-.1843	.4194	-.4757	.2509	-.3582	-.3584	-.0041	-.1357	.3573
6						.5515	-.2439	-.2008	.7037	-.4033	.2368	.2244	-.2341	.1765	-.5110
7							.1806	-.4091	.3376	-.6806	.3859	.3476	.1180	.1224	-.4248
8								-.0057	-.0318	-.1629	.1203	.1855	.4371	-.1250	.2655
9									-.0044	.7049	-.0257	.0184	.1593	-.2150	.5637
10										-.1352	.2037	.1600	-.0135	.0607	-.2848
11											-.1298	-.1025	-.0965	-.1572	.4104
12												.9839	.2823	-.0202	-.1772
13													.4094	-.0459	-.0774
14														-.1026	.4257
15															-.0885

Fuente: McDonald, G.C. y R.C.Schwing. Instability of Regression Estimates Relating Air Pollution to Mortality. Opus cit.

mos que la suma de los recíprocos de las raíces de la expresión anterior es tal que:

$$\sum_{i=1}^{15} \frac{1}{\lambda_i} = 263.06 \tag{281}$$

O sea, casi 3.5 veces mayor que cinco veces el número de variables explicativas en el modelo (es decir, 5x 15 =75). En consecuencia, se desprende que existe multicolinealidad en el mismo. Si examinamos el vector característico correspondiente a la raíz característica más pequeña, es decir, λ_{15}=.0049, nos percatamos que existe una fuerte relación o dependencia lineal entre las variables 12,13 y 14. Específicamente, esta dependencia adopta la forma:

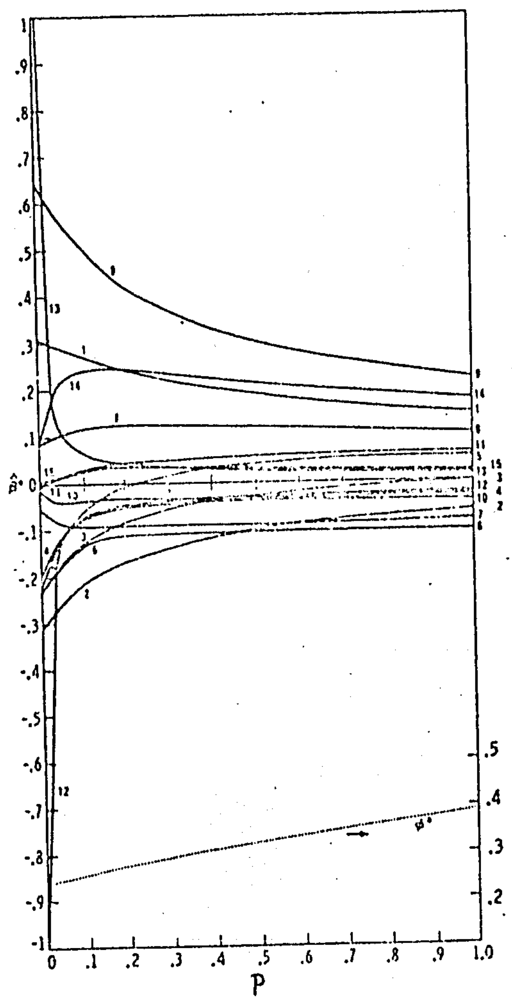
$$-0.689\, X_{12}^{*} + 0.712\, X_{13}^{*} - 0.108\, X_{14}^{*} \doteq 0 \tag{282}$$

donde X_i^{*} (i=1,2,3,...,15) es la variable estandarizada correspondiente a la variable Xi (i=1,2,...,15) y los coeficientes del resto de variables explicativas son aproximadamente cero, razón por la cual se omiten de la expresión anterior.

En base a la "traza de cordillera" construida para este problema y que se muestra en la Figura 5.6.2 (de la hoja siguiente), procederemos a la eliminación de variables.

F I G U R A 5.6.2

TRAZA DE CORDILLERA PARA EL MODELO (279).



Fuente: McDonald, G.C. y R.C. Schwing. Instability of Regression Estimates Relating Air Pollution to Mortality. Opus cit.

Apegándonos al primer criterio para el efecto(+), eliminamos primeramente las variables 4,7,10,11 y 15. Pues aunque estas variables tienen coeficientes muy estables, a juzgar por lo "aplanado" de sus "trazas de cordillera", éstos resultan muy pequeños.

El segundo criterio recomienda eliminar aquellas variables con coeficientes inestables que tienden a cero. Reexaminando la traza de cordillera, nos percatamos que las variables 12 y 13 caen dentro de esta categoría.

El tercer criterio sugiere eliminar aquellas variables con coeficientes inestables, y revisando otra vez la traza de cordillera, se observa que las variables 3 y 5 caen en esta categoría. Las variables que permanecen después de esta "depuración" mediante la traza de cordillera son: 1,2,6,8,9 y 14.

Para verificar si las variables retenidas efectivamente pueden considerarse como un "conjunto ortogonal ", comparamos la longitud (o distancia) observada al cuadrado del vector de coeficientes de cordillera respecto de la longitud esperada al cuadrado del vector de coeficientes de cordillera que se obtendría si el conjunto de variables explicativas fuese ortogonal. En otras palabras, debemos comparar las gráficas de las curvas (277) y (278) contra el parámetro "p". Ambas curvas se muestran en la Figura 5.6.3

(+) Recuérdese que el primer criterio establece que debemos eliminar primeramente aquellas variables cuyos coeficientes son estables pero pequeños.

(que aparece en la hoja siguiente).Donde notamos que la curva — punteada corresponde a la curva teórica (ortogonal) definida en la expresión (278) y la contínua a la curva real o verdadera (o sea la calculada a partir de los coeficientes de cordillera es— timados) definida en (277).

La Figura 5.6.2 muestra que la traza de cordillera es estable — para el valor $p = 0.2$. La ecuación estimada en términos de las variables estandarizadas resulta ser;

$$Y^{*} = .243X1^{*} - .168X2^{*} - .114X5^{*} + .123X8^{*} + .423X9^{*} + .243X14^{*} \qquad (283)$$

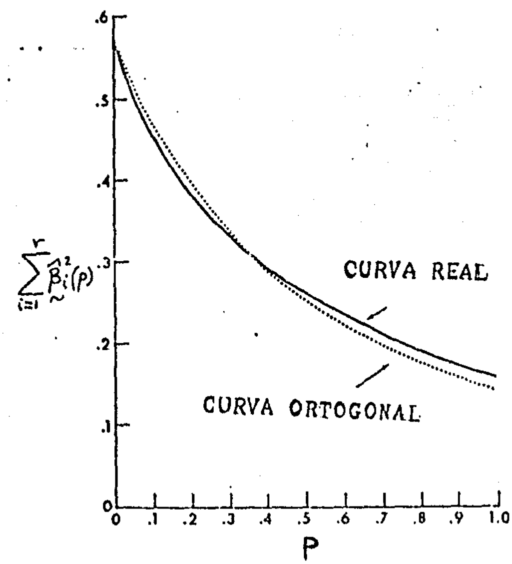
Por lo tanto, hemos visto que de un conjunto de 15 variables se han seleccionado un subconjunto de sólo 6 variables para formar una ecuación de coeficientes estables.Los valores de los estadísticos que fueron utilizados para juzgar la ecuación anterior son $C_k = 5.52$ y $R^2 = 0.571$.(+)

Para finalizar, resulta claro de nuestra discusión en este apar— tado que el proceso de selección de variables,y en especial, su vinculación para este propósito con el análisis de regresión de cordillera, debe realizarse con cuidado y precaución,pues resulta una mezcla de "arte y ciencia".Hemos intentado, por lo mismo, delinear en esta exposición algunos enfoques para este propósi— to en lugar de intentar exponer un procedimiento formal.Sobre —

(+) La ecuación estimada (283) observa un comportamiento bastan— te adecuado bajo los distintos estadísticos y criterios re— feridos a lo largo de nuestras discusiones para juzgar el — desempeño de una ecuación.Se omiten los cálculos.

F I G U R A 5.6.3

LONGITUD AL CUADRADO DEL VECTOR DE COEFICIENTES DE CORDILLERA PARA LAS VARIABLES RETENIDAS DESPUES DEL PROCESO DE ELIMINACION DE VARIABLES MEDIANTE EL ANALISIS DE CORDILLERA APLICADO AL MODELO (279).



Fuente; McDonald,G.C. y R.C. Schwing. Instability of Regression Estimates Relating Air Pollution to Mortaliy. Opus cit.

todo, al abordar el caso de la selección de variables multicolineales.

En conclusión, debemos hacer hincapié en el hecho de que la selección de variables no debe realizarse en forma mecánica y como un fin en sí mismo, sino como una exploración en la estructura de los datos bajo estudio, y por consiguiente, el "explorador" o investigador en cuestión debe guiarse en una mezcla de teoría, intuición y sentido común para poder tener éxito en su empresa.

VI. CONCLUSIONES

El peligro de la multicolinealidad en el análisis de regresión es particularmente agudo en aplicaciones económicas y sociales como hemos destacado, debido a la naturaleza misma de los problemas, así como cuando se manejan muestras pequeñas, principalmente aquellas de series cronológicas o de corte transversal que se generan "en vivo" o en la realidad, y no por experimentos controlados y muestreo aleatorio, como sucede en las ciencias exactas, físicas y biológicas, en donde por lo mismo, su ocurrencia es menos frecuente.

Como resultado de lo anterior, el investigador de las ciencias económicas y sociales necesita conocer cómo responder adecuadamnte al'infortunio' de la multicolinealidad y evitar, o al menos reducir al mínimo, sus adversos efectos en la estimación paramétrica e inferencia estadística, principalmente.

Desafortunadamente, como ya se ha observado a través de todo este trabajo, no existe un tratamiento uniforme, ni mucho menos directo para eliminarla o erradicarla de un modelo de regresión o econométrico afectado, como sería el caso de cualquier otra enfermedad estadística resultante de la violación de algunos de los supuestos generales del modelo lineal general, como por ejemplo, la autocorrelación o heteroscedasticidad, por citar las más importantes.

A pesar de lo anterior y previo análisis de las diversas técnicas y métodos discutidos aquí para diagnosticar y corregir el problema de la multicolinealidad, el enfoque que en nuestra particular opinión resulta más conveniente y efectivo para cumplir con ambos propósitos, es el que explota el concepto de ortogonalidad implícita o inherente a las variables --- explicativas del modelo lineal general.

Concretamente, nos referimos a la técnica de las componentes principales, que por sus características particulares resulta un instrumento sumamente poderoso tanto para el diagnóstico y análisis profundo de la multicolinealidad como para su erradicación,a través del auxilio de otras técnicas como la de la restricción paramétrica o la de la regresión de cordillera, que consideramos muy convenientes y adecuadas para la eliminación de la colinealidad y que aunque por sí solas pueden coadyuvar a eliminarla, su mejor efectividad se obtiene cuando se puede conocer y analizar la estructura colineal existente entre las variables explicativas afectadas, objetivo este último, que se puede lograr precisamente, aplicando la técnica de las componentes principales, como hemos tratado de resaltar a través de prácticamente todo este trabajo. De ahí se explica el énfasis y relevancia que se le ha asignado a través del mismo.

Sin embargo, no obstante todo lo anterior, podemos sentir un cierto alivio después de percatarnos que la colinealidad se vuelve menos seria y por lo mismo menos dañina, al aumentar el tamaño de la muestra, ya que al suceder esto, la variancia del término de perturbación tiende a disminuir forzosamente y nuestros estimadores se vuelven más confiables, como hemos señalado.

Aunque no debemos olvidar en el señalamiento anterior, que las muestras grandes son más costosas, tanto para recabar como para analizarse. A este propósito, podemos agregar que una ventaja de la técnica de las componentes principales, además de las señaladas en el párrafo anterior,con respecto a otras técnicas - como por ejemplo, la de la Dirección Optima de Silveyradica en que no requiere de información adicional para su aplicación y por lo mismo, resulta más práctica y por supuesto menos costosa en términos económicos y de tiempo, aunque hay que reconocer que su aplicación implica un poco más de esfuerzo que las técnicas tradicionales.

Conviene enfatizar, por otra parte, que con los diversos métodos, procedimientos y técnicas discutidos a lo largo de este trabajo hemos intentado presentar un compendio, lo más exhaustivo posible de las técnicas existentes y más ampliamente utilizadas por el investigador en la materia actualmente.

Empero, y a pesar de todas nuestras recomendaciones, el método o técnica en particular que ha de aplicarse en cada caso, dependerá en última instancia de la naturaleza propia del -- problema, de la severidad de la multicolinealidad en el modelo en cuestión, y finalmente, pero no menos importante, del propio juicio del investigador.

Para terminar, conviene hacer hincapié en el hecho de que esta aportación ha sido un intento modesto de analizar el fenómeno de la multicolinealidad en el análisis de regresión en una "cápsula" y tratar de presentar de una manera coherente, con un lenguaje accesible,pragmático y por demás"desmistificado", las técnicas más conocidas en la literatura especializada a fin de diagnosticarla y atacarla, con la esperanza de que pueda ser utilizada como un instrumento de consulta por el investigador interesado, en especial al de las ciencias económicas y sociales, a quien fundamentalmente está dirigido a quien no dudamos le sea de alguna utilidad, a pesar de sus limitaciones y posibles omisiones y errores involuntarios, por los cuales asumo responsabilidad plena y suplico disculpas y comprensión de antemano al lector.

VII. BIBLIOGRAFIA

1. Adams, F.G. Consumer Attitudes, Buying Plans and Purchase of Durable Goods: A Principal Component Analysis. Review of Economics and Statistics, Vol. 46, 1964, pp. 347-355.

2. Almon, S. The Distributed Lag Between Capital Appropiations and Expenditures. Econometrica, 33, January 1965, pp. 178-196.

3. Anderson, R.L. and T.A. Bancroft. Statistical Theory in Research. McGraw-Hill Book Co. Inc., 1952, p. 171.

4. Balestra, P. and M. Nerlove. Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. Econometrica, Vol. 34, No. 3, July 1966, pp. 585-612.

5. Barlett, B.S. Test of Significance in Factor Analysis. British Journal of Psychology, Statistical Section, 3, 1950, p. 83

6. Besley, Kuh and Welsh. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley and Sons. 1980.

7. Brennan, M.J. Preface to Econometrics: An introduction to Quantitative Methods in Economics. South-Western Publishing Company, 1965, pp. 348-350.

8. Coleman, J.S. et al. Equality of Educational Opportunity. Washington, D.C. United States Printing Office, 1966.

9. Chatterjee, S. and B. Price. Regression Analysis by Example. John Wiley and Sons, New York, 1977, pp. 144-151

10. Chiswick, B.R. and S.J. Chiswick. Statistics and Econometrics: A Problem-Solving Text. University Park Press, 1975, pp. 180-191.

11. Daniel, C. and F.S. Wood. Fitting Equations to Data. Wiley, New York, 1971.

12. Dempster, A.P. et al. A Simulation Study of Alternatives to Ordinay Least Squares. Research Report S-35, Dept. of Statistics, Harvard University, 1975.

13. Dhrymes,P.J. Econometrics: Statistical Foundations and Applications. Harper and Row Publishers,1970,pp.15-24.

14. Dutta, M. Econometric Methods. South-Western Publishing Co.,1975,pp.147-156.

15. Farrar,D.E. and R.R.Glauber.Multicollinearity in Regression Analysis: The Problem Revisited.Review of Economics and Statistics.Vol.49,1967,pp.92-107.

16. Frisch, R. Statistical Confluence Analysis by Means of Complete Regression Systems.Oslo, Norway: University Economics Institute,1934.

17. Furnival, G.M. and R.W.Wilson Jr.Regression by Leaps and Bounds. Technometrics,16,1974,pp.499-512.

18. Girshick,M.A. Principal Components. Journal of the American Statistical Association, Vol31,1936,pp.519-528.

19. Glauber,R.R. and E.A. Beaton. Statistical Laboratory Ultimate Regression Package Programming. Cambridge, Massachussets: Harvard Statistical Laboratory,1962.

20. Haavelmo,T. Remarks on Frisch's Confluence Analysis and Its use in Econometrics, in Statistical Inference in Dynamic Economic Models, Edited by T.C.Koopmans.New York: John Wiley and Sons,1950,Chapter 5.

21. Hadley, G.Linear Algebra. Addison- Wesley.1961.pp.237-240.

22. Haitovsky, Y. Multicollinearity in Regression Analysis: Comment.Review of Economics and Statistics.Vol.LI, November 1969,pp.486-489.

23. Hocking,R.R. The Analysis and Selection of Variables in Linear Regression.Biometrics,32,1976,pp.1-49.

24. Hocking,R.R. Misspecification in Regression. The American Statistician, 28,1974,pp.39-40.

25 . Hoerl, A.E. Optimum Solutions of Many Variables. Chem. Eng. Q. Progress, 55,1959, pp.69-78.

26 . Hoerl, A.E. Application of Ridge Analysis to Regression Problem. Chem. Engr. Progress, 58,1962, pp.54-59.

27 . Hoerl, A.E. and R.W. Kennard. Ridge Regression, Biased Estimation for Nonorthogonal Problems. Technometrics, 12,1970, pp.69-82.

28 . Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. Journal of Educational Psychology, Vol.24,1933, pp.417-441.

29 . Johnston, J. Econometric Methods. McGraw-Hill. Kogakusha, Ltd. Tokyo. 1972. pp.159-168

30 . Kane, E.J. Economic Statistics and Econometrics: An Introduction to Quantitative Economics. Harper and Row Publishers, 1968, pp.277-280.

31 . Kelejian, K.H. and E.O. Oates. Introduction to Econometrics: Principles and Applications. Second Edition. Harper and Row Publishers, New York, 1981, pp.132-136.

32 . Kendall, M.G. A Course in Multivariate Analysis. Charles Griffin, London, 1957, pp.70-75.

33 . Klein, L.R. and Mitsuga Nakamura. Singularity in the Equation System of Econometrics: Some Aspects of the Problem of Multicollinearity. International Economic Review, Vol.3, No.3, September 1962, pp.274-299.

34 . Klein, L.R. An Introduction to Econometrics. Englewood Cliffs, N.J. Prentice-Hall, Inc. 1962, pp.64,101.

35 . Klein, L.R. A Textbook of Econometrics. Prentice-Hall, Inc. 1974. pp.104-108.

36 . Kmenta, J. Elements of Econometrics. The Macmillan Company. 1971, pp.380-391.

37. Koyck, L.M. Distributed Lags and Investment Analysis. North-Holland, Amsterdam, 1954.

38. Kuh, E. and Meyer, J. How Extraneous are Extraneous Estimates?. Review of Economics and Statistics, Vol. 39, 1957, pp. 380-393.

39. Kuh, E. The Validity of Cross-Sectionally Estimated Behavior Equations in Time-Series Applications. Econometrica, Vol. 27, No. 2, April 1959, pp. 197-214.

40. La Motte, L.R. and R.R. Hocking. Computational Efficiency in the Selection of Regression Variables. Technometrics, 12, 1970, pp. 83-93.

41. Liu, T.C. Underidentification. Structural Estimation and Forecasting. Econometrica, 28, October 1960, p. 856.

42. Malinvaud, E. Statistical Methods of Econometrics. North-Holland Publishing Company, 1970, pp. 80-85, pp. 204-205.

43. Mantel, N. Why Stepdown Procedures in Variable Selection? Technometrics, 12, 1970, pp. 591-612.

44. Mallows, C.L. Some Comments on Cp. Technometrics, 15, 1973, pp. 661-75.

45. Massy, W.F. Principal Components Regression in Exploratory Statistical Resea . Journal of the American Statistical Association, Vol. 60, No. 309, March 1965, pp. 234-256.

46. McDonald, G.C. and R.C. Schwing. Instability of Regression Estimates Relating Air Pollution to Mortality. Technometrics, 15, 1973, pp. 463-81.

47. Mosteller, F. and D.P. Moynihan. Eds. On Equality of Educational Opportunity. Random House, New York, 1972.

48. Plackett, R.L. Regression Analysis. Oxford University Press, London, 1960.

49. Press,S.J. Applied Multivariate Analysis.Holt,Rinehart and Winston, New York,1972.

50. Rao,C.R. Linear Statistical Inference and Applications.Wiley, New York,1973.

51. Seber,G.F. Linear Regression Analysis.Wiley,New York,1977.

52. Silvey,S.D. Multicollinearity and Imprecise Estimation. Journal of the Royal Statistical Society, SeriesB, Vol.31,1969,pp.539-552.

53. Sonquist,J.A. and J.N.Morgan. The Detection of Interaction Effects.Survey of Research Center, Monograph No.35, Ann Arbor, Michigan: University of Michigan Press, 1964.

54. Stone,J.R.N. The Measurement of Consumers' Expenditure and Behavior in the United Kingdom,1920-1938.Cambridge, England: Cambridge University Press,1954.

55. Stone,J.R.N. The Analysis of Market Demand.Journal of the Royal Statistical Society,Vol.108,1945,Parts III and IV,pp.286-382.

56. Swindel,B.F.Instability of Regression Coefficients Illustrated.The American Statistician, 28(2), 1974,pp.63-65.

57. Tintner,G. A Note on Rank, Multicollinearity and Multiple Regression. Annals of Mathematical Statistics,Vol.16, 1945,pp.304-308.

58. Tobin, J.A. A Statistical Demand Function for Food in the U.S.A. Journal of the Royal Statistical Society, Vol.113-A,1950,pp.113-141.

59. Toro-Vizcarrondo,C. and T.D.Wallace. A Test of Mean Square Error Criteria for Restrictions in Linear Regression. Journal of the American Statistical Association, Vol.63,1968,pp.558-572.

60. Walter,A.A. An Introduction to Econometrics. Macmillan and Co.Ltd.,1970,p.124.

61. Willan and Watts.Meaningful Multicollinearity Measures. Technometrics, Vol.20, No.4, November 1978.

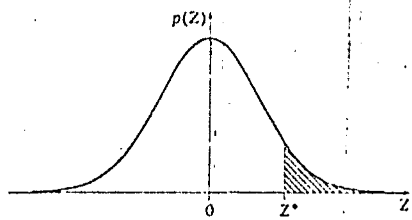
62. Wold,H.O.A. and Lars Juréen.Demand Analysis: A Study in Econometrics.New York: John Wiley and Sons,1953, pp.192-195.

VIII. A P E N D I C E (+)

(+) El material de este apéndice fue tomado de Mirer, Thad W. <u>Economic and Econometrics</u>. Macmillan Publishing Co. New York, pp. 356-365

Use of Table A-1: The Standard Normal Distribution

The main body of this table gives selected values of Z^* and the corresponding $Pr(Z \geq Z^*)$, where Z is a random variable having a standard normal distribution (see Section 10.1). In the following figure, Z^* is a specific value of Z, and $Pr(Z \geq Z^*)$ is given by the shaded area.



This table is used to answer two types of probability problems. First,

Find α such that $Pr(Z \geq Z^*) = \alpha$

In this case the Z^* value is given, and the table is used to determine the corresponding probability, which is denoted by α. Second,

Find Z^c such that $Pr(Z \geq Z^c) = \alpha$

In this case the probability amount (α) is given, and the table is used to determine the critical value of Z (denoted generally by Z^c, but corresponding to Z^* in the table) that bounds exactly α probability in the right-hand tail. In both cases, interpolation can be used if greater precision is desired.

The lower section of the table gives several probability statements that are of special interest for statistical inference.

For example, given $Z^* = 1.50$, we find that $Pr(Z \geq 1.50) = .067$. Also, given that $Pr(Z \geq Z^c) = .25$, we find that Z^c lies between 0.65 and 0.70; interpolating, we may say that $Z^c \approx 0.675$.

Table A-1 Standard Normal Distribution

Z^*	$Pr(Z \geq Z^*)$	Z^*	$Pr(Z \geq Z^*)$	Z^*	$Pr(Z \geq Z^*)$
0.00	.500	1.00	.159	2.00	.023
0.05	.480	1.05	.147	2.05	.020
0.10	.460	1.10	.136	2.10	.018
0.15	.440	1.15	.125	2.15	.016
0.20	.421	1.20	.115	2.20	.014
0.25	.401	1.25	.106	2.25	.012
0.30	.382	1.30	.097	2.30	.011
0.35	.363	1.35	.089	2.35	.009
0.40	.345	1.40	.081	2.40	.008
0.45	.326	1.45	.074	2.45	.007
0.50	.309	1.50	.067	2.50	.006
0.55	.291	1.55	.061	2.55	.005
0.60	.274	1.60	.055	2.60	.005
0.65	.258	1.65	.049	2.65	.004
0.70	.242	1.70	.045	2.70	.003
0.75	.227	1.75	.040	2.75	.003
0.80	.212	1.80	.036	2.80	.003
0.85	.198	1.85	.032	2.85	.002
0.90	.184	1.90	.029	2.90	.002
0.95	.171	1.95	.026	2.95	.002
				3.00	.001

$Pr(Z \geq 1.282) = .10$
$Pr(Z \geq 1.645) = .05$
$Pr(Z \geq 1.960) = .025$
$Pr(Z \geq 2.326) = .01$
$Pr(Z \geq 2.576) = .005$

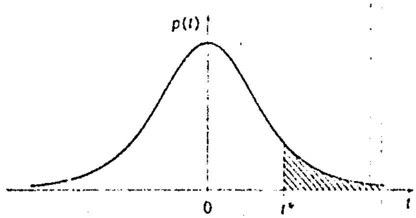
Note: Table entries give Z^* and corresponding $Pr(Z \geq Z^*)$.
Source: Computed using Fortran subroutines from the IMSL Library.

Use of Table A-2: The t Distribution

This table gives selected right-hand tail probabilities for a random variable having a t distribution with df degrees of freedom (see Section 10.2). Each entry in the table is the answer to a problem of the form:

Find α such that $Pr(t \geq t^*) = \alpha$

In hypothesis testing, this α is known as the P-value. In the following figure, t^* is a specific value of the random variable t and the probability α is given by the shaded area.



In the table, various t^* values are given in the column headings, and the corresponding α is found in the row labeled by df.

For example, if $df = 15$, we find that $Pr(t \geq 1.50) = .077$ and $Pr(t \geq 2.50) = .012$.

Table A-2 P-Values for the *t* Distribution

	$t^* = 0.50$	1.00	1.25	1.50	1.75	2.00	2.25	2.50	3.00
$df = 1$	.352	.250	.215	.187	.165	.148	.133	.121	.102
2	.333	.211	.169	.136	.111	.092	.077	.065	.048
3	.326	.196	.150	.115	.089	.070	.055	.044	.029
4	.322	.187	.140	.104	.078	.058	.044	.033	.020
5	.319	.182	.133	.097	.070	.051	.037	.027	.015
6	.317	.178	.129	.092	.065	.046	.033	.023	.012
7	.316	.175	.126	.089	.062	.043	.030	.020	.010
8	.315	.173	.123	.086	.059	.040	.027	.018	.009
9	.315	.172	.121	.084	.057	.038	.026	.017	.007
10	.314	.170	.120	.082	.055	.037	.024	.016	.007
11	.313	.169	.119	.081	.054	.035	.023	.015	.006
12	.313	.169	.118	.080	.053	.034	.022	.014	.006
13	.313	.168	.117	.079	.052	.033	.021	.013	.005
14	.312	.167	.116	.078	.051	.033	.021	.013	.005
15	.312	.167	.115	.077	.050	.032	.020	.012	.004
16	.312	.166	.115	.077	.050	.031	.019	.012	.004
17	.312	.166	.114	.076	.049	.031	.019	.011	.004
18	.312	.165	.114	.075	.049	.030	.019	.011	.004
19	.311	.165	.113	.075	.048	.030	.018	.011	.004
20	.311	.165	.113	.075	.048	.030	.018	.011	.004
21	.311	.164	.113	.074	.047	.029	.018	.010	.003
22	.311	.164	.112	.074	.047	.029	.017	.010	.003
23	.311	.164	.112	.074	.047	.029	.017	.010	.003
24	.311	.164	.112	.073	.046	.028	.017	.010	.003
25	.311	.163	.111	.073	.046	.028	.017	.010	.003
26	.311	.163	.111	.073	.046	.028	.017	.010	.003
27	.311	.163	.111	.073	.046	.028	.016	.009	.003
28	.310	.163	.111	.072	.046	.028	.016	.009	.003
29	.310	.163	.111	.072	.045	.027	.016	.009	.003
30	.310	.163	.110	.072	.045	.027	.016	.009	.003
40	.310	.162	.109	.071	.044	.026	.015	.008	.002
50	.310	.161	.109	.070	.043	.025	.014	.008	.002
60	.309	.161	.108	.069	.043	.025	.014	.008	.002
70	.309	.160	.103	.069	.042	.025	.014	.007	.002
80	.309	.160	.107	.069	.042	.024	.014	.007	.002
90	.309	.160	.107	.069	.042	.024	.013	.007	.002
100	.309	.160	.107	.068	.042	.024	.013	.007	.002
125	.309	.160	.107	.068	.041	.024	.013	.007	.002
150	.309	.159	.107	.068	.041	.024	.013	.007	.002
200	.309	.159	.106	.068	.041	.023	.013	.007	.002
∞	.309	.159	.106	.067	.040	.023	.012	.006	.001

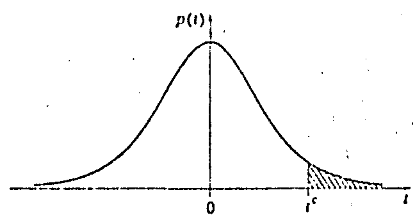
Note: Table entry gives $Pr(t \geq t^*)$ for t^* in column heading.
Source: Computed using Fortran subroutines from the IMSL Library.

Use of Table A-3: The t Distribution

This table gives critical values for a random variable having a t distribution with df degrees of freedom (see Section 10.2). The table is arranged to answer problems of two forms. First:

Find t^c such that $Pr(t \geq t^c) = \alpha$

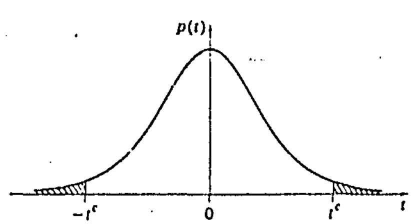
Various one-tailed α values are given in the column headings, and the corresponding critical values, t^c, are found in the row for a particular number of degrees of freedom. In the following figure, the shaded area corresponds to the given α, and t^c is the t value that bounds α probability in the right-hand tail of the distribution.



The second problem is of the form:

Find t^c such that $Pr(|t| \geq t^c) = \alpha$

The probability amount α is split between the two tails of the distribution, each containing $\alpha/2$. Various two-tailed α values are given in the column headings, and the corresponding positive critical value, t^c, is given in the appropriate row (the other critical value is $-t^c$). In the following figure the two shaded areas together correspond to the given α, and t^c is the t value that bounds $\alpha/2$ probability in the right-hand tail.



For example, if $df = 15$ and $\alpha = .05$, $t^c = 1.753$ for a one-tailed test, and $t^c = 2.131$ for a two-tailed test. Note that t^c for a one-tailed test with $\alpha = .05$ is the same as t^c for a two-tailed test with $\alpha = .10$.

Table A-3 Critical Values for the t Distribution

One-tailed α = .10	.05	.025	.01	.005
Two-tailed α = .20	**.10**	**.05**	**.02**	**.01**
df = 1 3.078	6.314	12.706	31.821	63.657
2 1.886	2.920	4.303	6.965	9.925
3 1.638	2.353	3.182	4.541	5.841
4 1.533	2.132	2.776	3.747	4.604
5 1.476	2.015	2.571	3.365	4.032
6 1.440	1.943	2.447	3.143	3.707
7 1.415	1.895	2.365	2.998	3.499
8 1.397	1.860	2.306	2.896	3.355
9 1.383	1.833	2.262	2.821	3.250
10 1.372	1.812	2.228	2.764	3.169
11 1.363	1.796	2.201	2.718	3.106
12 1.356	1.782	2.179	2.681	3.055
13 1.350	1.771	2.160	2.650	3.012
14 1.345	1.761	2.145	2.624	2.977
15 1.341	1.753	2.131	2.602	2.947
16 1.337	1.746	2.120	2.583	2.921
17 1.333	1.740	2.110	2.567	2.898
18 1.330	1.734	2.101	2.552	2.878
19 1.328	1.729	2.093	2.539	2.861
20 1.325	1.725	2.086	2.528	2.845
21 1.323	1.721	2.080	2.518	2.831
22 1.321	1.717	2.074	2.508	2.819
23 1.319	1.714	2.069	2.500	2.807
24 1.318	1.711	2.064	2.492	2.797
25 1.316	1.708	2.060	2.485	2.787
26 1.315	1.706	2.056	2.479	2.779
27 1.314	1.703	2.052	2.473	2.771
28 1.313	1.701	2.048	2.467	2.763
29 1.311	1.699	2.045	2.462	2.756
30 1.310	1.697	2.042	2.457	2.750
40 1.303	1.684	2.021	2.423	2.704
50 1.299	1.676	2.009	2.403	2.678
60 1.296	1.671	2.000	2.390	2.660
70 1.294	1.667	1.994	2.381	2.648
80 1.292	1.664	1.990	2.374	2.639
90 1.291	1.662	1.987	2.368	2.632
100 1.290	1.660	1.984	2.364	2.626
125 1.288	1.657	1.979	2.357	2.616
150 1.287	1.655	1.976	2.351	2.609
200 1.286	1.653	1.972	2.345	2.601
∞ 1.282	1.645	1.960	2.326	2.576

Note: Table entry gives t^c corresponding to $Pr(t \geq t^c) = \alpha$ for one-tailed tests and $Pr(|t| \geq t^c) = \alpha$ for two-tailed tests.

Source: Computed using Fortran subroutines from the IMSL Library.

Use of Table A-4: The Chi-Square Distribution

This table gives critical values for a random variable having a chi-square distribution with *df* degrees of freedom (see the appendix to Chapter 10). Each entry in the table is the answer to a problem of the form:

Find $(\chi^2)^c$ such that $Pr(\chi^2 \geq (\chi^2)^c) = \alpha$

Various values of α are given in the column headings, and the corresponding $(\chi^2)^c$ is found in the row labeled by *df*. Critical values for areas in the left side of the distribution can be found by subtraction.

For example, if $df = 35$, $(\chi^2)^c = 53.22$ for $\alpha = .025$ and $(\chi^2)^c = 20.56$ for $\alpha = .975$. Thus, $(\chi^2)^c = 53.22$ bounds .025 probability in the right-hand tail and $(\chi^2)^c = 20.56$ bounds .025 probability in the left side.

Table A-4 Critical Values for the Chi-Square Distribution

	α = .990	.975	.950	.900	.100	.050	.025	.010
df = 1	—	—	—	0.02	2.71	3.84	5.02	6.64
2	0.02	0.05	0.10	0.21	4.60	5.99	7.38	9.22
3	0.11	0.22	0.35	0.58	6.25	7.82	9.36	11.32
4	0.30	0.48	0.71	1.06	7.78	9.49	11.15	13.28
5	0.55	0.83	1.15	1.61	9.24	11.07	12.84	15.09
6	0.87	1.24	1.63	2.20	10.65	12.60	14.46	16.81
7	1.24	1.69	2.17	2.83	12.02	14.07	16.02	18.47
8	1.64	2.18	2.73	3.49	13.36	15.51	17.55	20.08
9	2.09	2.70	3.32	4.17	14.69	16.93	19.03	21.65
10	2.55	3.24	3.94	4.86	15.99	18.31	20.50	23.19
11	3.05	3.81	4.57	5.58	17.28	19.68	21.93	24.75
12	3.57	4.40	5.22	6.30	18.55	21.03	23.35	26.25
13	4.10	5.01	5.89	7.04	19.81	22.37	24.75	27.72
14	4.65	5.62	6.57	7.79	21.07	23.69	26.13	29.17
15	5.23	6.26	7.26	8.55	22.31	25.00	27.50	30.61
16	5.81	6.90	7.96	9.31	23.55	26.30	28.86	32.03
17	6.40	7.56	8.67	10.08	24.77	27.59	30.20	33.44
18	7.00	8.23	9.39	10.86	25.99	28.83	31.54	34.83
19	7.63	8.90	10.11	11.65	27.21	30.15	32.87	36.22
20	8.25	9.59	10.85	12.44	28.42	31.42	34.18	37.59
21	8.89	10.28	11.59	13.24	29.62	32.68	35.49	38.96
22	9.53	10.98	12.34	14.04	30.82	33.93	36.79	40.31
23	10.19	11.69	13.09	14.85	32.01	35.18	38.09	41.66
24	10.85	12.40	13.84	15.66	33.20	36.42	39.38	43.00
25	11.51	13.11	14.61	16.47	34.38	37.66	40.66	44.34
26	12.19	13.84	15.38	17.29	35.57	38.89	41.94	45.66
27	12.87	14.57	16.15	18.11	36.74	40.12	43.21	46.99
28	13.55	15.30	16.92	18.94	37.92	41.34	44.47	48.30
29	14.24	16.04	17.70	19.77	39.09	42.56	45.74	49.61
30	14.94	16.78	18.49	20.60	40.26	43.78	46.99	50.91
35	18.49	20.56	22.46	24.79	46.06	49.81	53.22	57.36
40	22.14	24.42	26.51	29.05	51.80	55.75	59.34	63.71
45	25.88	28.36	30.61	33.36	57.50	61.65	65.41	69.98
50	29.68	32.35	34.76	37.69	63.16	67.50	71.42	76.17
55	33.55	36.39	38.96	42.06	68.79	73.31	77.38	82.31
60	37.46	40.47	43.19	46.46	74.39	79.03	83.30	88.40
65	41.42	44.60	47.45	50.89	79.97	84.82	89.18	94.44
70	45.42	48.75	51.74	55.33	85.52	90.53	95.03	100.44
75	49.46	52.94	56.05	59.80	91.06	96.21	100.84	106.41
80	53.52	57.15	60.39	64.28	96.57	101.88	106.63	112.34
85	57.62	61.38	64.75	68.78	102.07	107.52	112.40	118.25
90	61.74	65.64	69.13	73.29	107.56	113.14	118.14	124.13
95	65.88	69.92	73.52	77.82	113.03	118.75	123.86	129.99
100	70.05	74.22	77.93	82.36	118.49	124.34	129.56	135.82

Note: Table entry gives $(\chi^2)^c$ corresponding to $Pr(\chi^2 \geq (\chi^2)^c) = \alpha$.

Source: Computed using Fortran subroutines from the IMSL Library. Some values differ slightly from those in other published tables.

Use of Tables A-5 and A-6: The F Distribution

These tables give critical values for a random variable having an F distribution with n degrees of freedom in the numerator and d degrees of freedom in the denominator (see the appendix to Chapter 10). Each entry in the tables is the answer to a problem of the form:

Find F^c such that $Pr(F_{n,d} \geq F^c) = \alpha$

Table A-5 contains the F^c values for the .05 level of significance, and Table A-6 contains the F^c values for $\alpha = .01$. The entry for a particular distribution is in the column headed by n and the row labeled by d.

For example, if $n = 4$ and $d = 50$, $F^c = 2.56$ for $\alpha = .05$ and $F^c = 3.72$ for $\alpha = .01$.

Table A-5 Critical Values for the F Distribution ($\alpha = .05$)

	$n = 1$	2	3	4	5	6	8	10	15
$d = 1$	161.4	199.5	215.7	224.6	230.2	234.0	238.9	241.9	245.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.40	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.79	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.64	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.35	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.14	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.98	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.85	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.75	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.67	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.60	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.54	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.45	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.38	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.20
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.32	2.18
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.30	2.15
23	4.28	3.42	3.03	2.80	2.64	2.53	2.37	2.27	2.13
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.25	2.11
25	4.24	3.39	2.99	2.76	2.60	2.49	2.34	2.24	2.09
26	4.23	3.37	2.98	2.74	2.59	2.47	2.32	2.22	2.07
27	4.21	3.35	2.96	2.73	2.57	2.46	2.31	2.20	2.06
28	4.20	3.34	2.95	2.71	2.56	2.45	2.29	2.19	2.04
29	4.18	3.33	2.93	2.70	2.55	2.43	2.28	2.18	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16	2.01
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.08	1.92
50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	2.03	1.87
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.99	1.84
70	3.98	3.13	2.74	2.50	2.35	2.23	2.07	1.97	1.81
80	3.96	3.11	2.72	2.49	2.33	2.21	2.06	1.95	1.79
90	3.95	3.10	2.71	2.47	2.32	2.20	2.04	1.94	1.78
100	3.94	3.09	2.70	2.46	2.31	2.19	2.03	1.93	1.77
125	3.92	3.07	2.68	2.44	2.29	2.17	2.01	1.91	1.75
150	3.90	3.06	2.66	2.43	2.27	2.16	2.00	1.89	1.73
200	3.89	3.04	2.65	2.42	2.26	2.14	1.98	1.88	1.72
∞	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.83	1.67

Note: Table entry gives F^c corresponding to $Pr(F_{n,d} \geq F^c) = .05$.
Source: Computed using Fortran subroutines from the IMSL Library.

Table A-6 Critical Values for the F Distribution ($\alpha = .01$)

	$n = 1$	2	3	4	5	6	8	10	15
$d = 1$	4052.	4999.	5403.	5625.	5764.	5859.	5981.	6056.	6157.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.40	99.43
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.23	26.87
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.55	14.20
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	10.05	9.72
6	13.75	10.92	9.78	9.15	8.75	8.47	8.10	7.87	7.56
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.62	6.31
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.81	5.52
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.26	4.96
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.85	4.56
11	9.65	7.21	6.22	5.67	5.32	5.07	4.74	4.54	4.25
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.30	4.01
13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	4.10	3.82
14	8.86	6.51	5.56	5.04	4.69	4.46	4.14	3.94	3.66
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.80	3.52
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.69	3.41
17	8.40	6.11	5.19	4.67	4.34	4.10	3.79	3.59	3.31
18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.51	3.23
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.43	3.15
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.37	3.09
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.31	3.03
22	7.95	5.72	4.82	4.31	3.99	3.76	3.45	3.26	2.98
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.21	2.93
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.17	2.89
25	7.77	5.57	4.68	4.18	3.85	3.63	3.32	3.13	2.85
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	3.09	2.81
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	3.06	2.78
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	3.03	2.75
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	3.00	2.73
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.98	2.70
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.80	2.52
50	7.17	5.06	4.20	3.72	3.41	3.19	2.89	2.70	2.42
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.63	2.35
70	7.01	4.92	4.07	3.60	3.29	3.07	2.78	2.59	2.31
80	6.96	4.88	4.04	3.56	3.26	3.04	2.74	2.55	2.27
90	6.93	4.85	4.01	3.53	3.23	3.01	2.72	2.52	2.24
100	6.90	4.82	3.98	3.51	3.21	2.99	2.69	2.50	2.22
125	6.84	4.78	3.94	3.47	3.17	2.95	2.66	2.47	2.19
150	6.81	4.75	3.91	3.45	3.14	2.92	2.63	2.44	2.16
200	6.76	4.71	3.88	3.41	3.11	2.89	2.60	2.41	2.13
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.32	2.04

Note: Table entry gives F^c corresponding to $Pr(F_{n,d} \geq F^c) = .01$.
Source: Computed using Fortran subroutines from the IMSL Library.