

PATRONES DE GASTO EN LOS HOGARES DE LA CIUDAD DE MÉXICO

Carlos M. Jarque

Secretaría de Programación y Presupuesto

En este trabajo se presenta la estimación de un sistema de demanda para la ciudad de México utilizando datos de 1975. Los hogares son clasificados en grupos homogéneos con respecto a los parámetros del sistema, a través del uso de *Análisis de Conglomerados*. Los resultados muestran que la ocupación del jefe del hogar es una de las variables más importantes para determinar el patrón de consumo de los hogares. La estimación es llevada a cabo grupo por grupo, considerando la naturaleza no negativa de los gastos y, por consiguiente, utilizando modelos de *Variable Dependiente Limitada*, esto es, se aplicaron el Modelo Truncado y el Modelo Tobit. Diferencias sustanciales fueron encontradas al estimar los parámetros de demanda utilizando Mínimos Cuadrados Ordinarios (mco) comparados con los de la Estimación por Máxima Verosimilitud aplicada a los modelos de Variable Dependiente Limitada. Se encontró que los resultados de mco, que son frecuentemente utilizados en la práctica pueden, por ejemplo, sobrestimar la propensión marginal al consumo y subestimar considerablemente el gasto de subsistencia, siendo clara la importancia de esto para propósitos de política económica.

1. Introducción

En este trabajo presentamos un estudio de corte transversal de los patrones de gasto en los hogares de la ciudad de México. La información utilizada proviene de la Encuesta de Ingreso-Gasto de los Hogares de 1975 realizada por la Secretaría del Trabajo y Previsión Social y el modelo que se aplica es el Sistema Lineal Extendido de Gasto. Como es bien conocido, los estudios de demanda son de los ejercicios más antiguos en econometría aplicada (algunas referencias históricas se dan en Brown y Deaton, 1972). Sin embargo, aquí nos centramos en el caso de México, para el cual existe poco

$$(4) \quad y_{ji} = x_i' \beta_j + \varepsilon_{ji} \quad ,$$

con $j = 1, \dots, n+1$; $i = 1, \dots, M$; y donde $x_i' = (1, x_i^*)$; x_i^* es el ingreso disponible del i -ésimo hogar; y_{ji} es el gasto del i -ésimo hogar en el bien j ($j = 1, \dots, n$); $y_{(n+1)i} = s_i$; $\beta_j = (\beta_{1j}, \beta_{2j})'$, y ε_{ji} es el disturbio aleatorio correspondiente al i -ésimo hogar para el j -ésimo bien.

Como lo demuestra Howe (1975), el sistema dado en (4) está perfectamente identificado. Así, para valores estimados de las β , podemos obtener estimaciones únicas para $\theta = (\phi_1, \dots, \phi_n, \theta_1, \dots, \theta_n, \mu)$. También notamos que, por definición, existen restricciones que los vectores $\beta_1, \dots, \beta_{(n+1)}$ deben satisfacer. Esto se puede ver de (3) —y de la restricción de que las θ_j suman la unidad— de lo cual se sigue que:

$$(5) \quad \begin{aligned} \beta_{1(n+1)} &= - \sum_{j=1}^n \beta_{1j} & \text{y} \\ \beta_{2(n+1)} &= 1 - \sum_{j=1}^n \beta_{2j} \end{aligned}$$

3.2 Estimación del SLEG

Para la estimación del modelo se utilizan supuestos respecto a la distribución de los disturbios. Antes de discutir esto, haremos algunas consideraciones sobre la Matriz de Varianza-Covarianza (mvc) e introduciremos notación adicional. Definimos $\varepsilon_i = (\varepsilon_{1i}, \dots, \varepsilon_{ni})'$ y denotamos a su mvc como Ω . También definimos $\varepsilon_i^+ = (\varepsilon_i', \varepsilon_{(n+1)i})'$ y denotamos su mvc por Ω^+ . Dado que estamos en un contexto de corte transversal, podemos suponer razonablemente que las covarianzas entre ε_i^+ y ε_k^+ son cero para $i \neq k$. Definimos $1_r'$ como un vector de unos, de dimensión 1 por r . Notamos que debido a que $x_i^* = s_i + y_{1i} + \dots + y_{ni}$ los $n+1$ errores $\varepsilon_{1i}, \dots, \varepsilon_{ni}, \varepsilon_{(n+1)i}$ son interdependientes. Más específicamente, tenemos $1_{(n+1)}' \varepsilon_i^+ = 0$, es decir, $\varepsilon_{(n+1)i} = -1_n' \varepsilon_i$. Por lo tanto, podemos escribir $\varepsilon_i^+ = (\varepsilon_i', -1_n' \varepsilon_i) = H \varepsilon_i$, donde $H = [I_n, -1_n]'$. De esto observamos que:

$$(6) \quad \Omega^+ = H \Omega H' = \begin{bmatrix} \Omega & -\Omega 1_n \\ -1_n' \Omega & 1_n' \Omega 1_n \end{bmatrix}$$

De aquí es claro que el rango de Ω^+ no puede exceder al de Ω y por consiguiente Ω^+ es singular. Lo que sí podemos suponer es que Ω tiene rango completo, de tal forma que el rango de Ω^+ es n .

Ahora procedemos a discutir los *supuestos sobre la distribución* de probabilidad de ε_i^+ .

i) El Modelo Normal

Empezamos con el supuesto distribucional más sencillo, que consiste en suponer normalidad dejando a un lado, por tanto, la no negatividad de la variable dependiente (adelante se hacen comentarios sobre la estimación del modelo bajo especificaciones estocásticas más apropiadas). Dado que la matriz Ω^+ es singular, no podemos expresar $\varepsilon_i^+ \sim N(0, \Omega^+)$ e intentar maximizar la función de verosimilitud sujeta a las restricciones de los parámetros. Una solución apropiada a este problema, es descartar una de las ecuaciones y estimar las n restantes suponiendo normalidad. Este subsistema tendría una mvc no singular (ver Powell, 1974: 48).

Por ejemplo, podemos proceder como en Powell (1973), y eliminar la ecuación relacionada con el ahorro (es decir, la ecuación $n + 1$ en (4)) y obtener estimadores para β_1, \dots, β_n , digamos $\hat{\beta}_1, \dots, \hat{\beta}_n$, y podríamos obtener estimaciones para $\beta(n + 1)$ de las relaciones dadas en (5). Se puede demostrar que bajo el supuesto de normalidad, los Estimadores de Máxima Verosimilitud (EMV) del sistema son iguales a los estimadores de mco ecuación por ecuación. Así, la estimación por mco para cada una de las $n + 1$ ecuaciones en (4) proporciona estimadores que satisfacen las restricciones paramétricas definidas en (5).

Para algunos grupos de hogares, el supuesto de normalidad puede no ser enteramente inapropiado y razonablemente podríamos utilizar mco. Para otros grupos, sin embargo, éste puede no ser el caso, requiriéndose consideraciones adicionales sobre la distribución de los disturbios. Esto se discute a continuación.

ii) El Modelo Truncado

Todas las variables dependientes en el SLEG (a excepción de una) son gastos, por lo que sabemos que son no-negativas, es decir, que y_{ji} está restringida a satisfacer $y_{ji} \geq 0$ para $j = 1, \dots, n$. Hasta ahora hemos supuesto normalidad de los disturbios, lo cual implica que consideramos que y_{ji} está en el intervalo $-\infty < y_{ji} < \infty$. Al suponer esto, estamos permitiendo que exista una probabilidad positiva de tener valores negativos en las observaciones y_{ji} . Típicamente, esta característica del supuesto de normalidad (claramente indeseable), es pasada por alto en trabajos empíricos, pensando que la probabilidad de tener valores negativos es muy pequeña, y que representaría una complicación inútil *restringuir* el rango de las variables dependientes a valores no negativos. Esta es una consideración válida cuando las esperanzas condicionales de y_{1i}, \dots, y_{ni} son "grandes", por ejemplo, en estudios

con información a nivel macroeconómico (podemos pensar en la estimación de una función consumo agregada donde la esperanza del gasto está en miles de millones de pesos y, por tanto, lejos del valor cero).

En este estudio, sin embargo, analizamos el comportamiento de agentes a nivel microeconómico y tenemos que, particularmente para algunos grupos de hogares (por ejemplo aquéllos con bajos ingresos), las esperanzas condicionales de gasto de algunos bienes (por ejemplo educación) son "cercanas" a cero para un gran número de hogares. En estos casos, la probabilidad de observaciones negativas bajo el Modelo Normal no sería "negligible", y por ello sería inapropiado realizar el análisis bajo esa especificación. Una alternativa al Modelo Normal es el uso del *Modelo Truncado* que se presenta a continuación.

Por ahora centrémonos en un solo bien, digamos el bien 1, y limitemos el rango de $y_{1i} = x_i' \beta_1 + \varepsilon_{1i}$ a valores no-negativos (es decir, imponemos la restricción $x_i' \beta_1 + \varepsilon_{1i} \geq 0$). Esto implica que el rango de ε_{1i} se restringe a $\varepsilon_{1i} \geq -x_i' \beta_1$. Adicionalmente, supongamos que ε_{1i} tiene una distribución normal truncada, con punto de truncación $-x_i' \beta_1$, con lo cual y_{1i} sería no-negativa con probabilidad uno. Más formalmente, supongamos que la función de densidad de probabilidad de ε_{1i} está dada por:

$$(7) \quad \text{f.d.p. } (\varepsilon_{1i}) = (2\pi\sigma_1^2 F_{1i}^2)^{-1/2} \exp \{ -\varepsilon_{1i}^2 / (2\sigma_1^2) \}$$

donde $\varepsilon_{1i} \geq -x_i' \beta_1$ y $F_{1i} = F(x_i' \beta_1, \sigma_1^2)$ con

$$F(x_i' \beta_1, \sigma_1^2) = \int_{-\infty}^{x_i' \beta_1} (2\pi\sigma_1^2)^{-1/2} \exp \{ -\lambda^2 / (2\sigma_1^2) \} d\lambda$$

De (7) se puede notar que si $x_i' \beta_1$ fuera grande, entonces $F_{1i} \cong 1$, y la f.d.p. de ε_{1i} sería aproximadamente normal. Asimismo, partiendo de (7) se puede demostrar que, a diferencia del Modelo Normal, la esperanza de ε_{1i} no sería cero. De hecho, la esperanza sería:

$$E[\varepsilon_{1i}] = \sigma_1^2 f_{1i} / F_{1i},$$

de donde se sigue que:

$$E[y_{1i}] = x_i' \beta_1 + \sigma_1^2 f_{1i} / F_{1i}$$

con $f_{1i} = f(x_i' \beta_1, \sigma_1^2)$ y $f(\omega, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-\omega^2 / (2\sigma^2))$.

En (3) definimos $\beta_{1j} = p_j \phi_j - \theta_j \mu p' \phi$ y $\beta_{2j} = \theta_j \mu$. Por otro lado, de la definición del SLEG teníamos que $p_j q_j = (p_j \phi_j - \theta_j \mu p' \phi) + (\theta_j \mu) x^*$. Así, vemos que β_{1j} y β_{2j} intentaban representar, respectivamente, el gasto esperado en

el bien j cuando $x^* = 0$, y la derivada del gasto esperado en el bien j con respecto a x^* . Tomando la expresión para $E[y_{1i}]$ se puede demostrar que (ver McDonald y Moffitt (1980), y Poirier y Melino (1978):

$$E[y_{1i} | x^* = 0] = \beta_{11} + \sigma_1^2 f(\beta_{11}, \sigma_1^2) / F(\beta_{11}, \sigma_1^2)$$

$$y \quad \partial E[y_{1i}] / \partial x^* = \beta_{21} [1 - x_1^* \beta_{11} f_{1i} / F_{1i} - \sigma_1^2 f_{1i}^2 / F_{1i}^2]$$

De aquí se sigue que —antes de calcular θ basado en estimaciones de las β — tenemos que ajustar estas últimas de tal forma que representen las respuestas o estimaciones de los parámetros de demanda requeridos. Más formalmente, después de obtener la estimación (por máxima verosimilitud) de β_1 y σ_1^2 para el Modelo Truncado, digamos $\tilde{\beta}_1 = (\tilde{\beta}_{11}, \tilde{\beta}_{21})$ y $\tilde{\sigma}_1^2$, debemos calcular los valores requeridos $\hat{\beta}_{11}$ y $\hat{\beta}_{21}$ definidos como:

$$(8) \quad \hat{\beta}_{11} = \tilde{\beta}_{11} + \tilde{\sigma}_1^2 f(\tilde{\beta}_{11}, \tilde{\sigma}_1^2) / F(\tilde{\beta}_{11}, \tilde{\sigma}_1^2)$$

$$y \quad \hat{\beta}_{21} = \tilde{\beta}_{21} [1 - \bar{x}' \tilde{\beta}_1 f(\bar{x}' \tilde{\beta}_1, \tilde{\sigma}_1^2) / F(\bar{x}' \tilde{\beta}_1, \tilde{\sigma}_1^2) - \tilde{\sigma}_1^2 f^2(\bar{x}' \tilde{\beta}_1, \tilde{\sigma}_1^2) / F^2(\bar{x}' \tilde{\beta}_1, \tilde{\sigma}_1^2)],$$

$$\text{donde} \quad \bar{x}' = (1, \bar{x}^*) \quad y \quad \bar{x}^* = (x_1^* + \dots + x_M^*) / M$$

iii) El Modelo Tobit

Bajo el supuesto de que los disturbios de, digamos, la ecuación 1, son normales o normales truncados, se tendría una “probabilidad cero” de obtener varias observaciones y_{1i} idénticas e iguales a cero. Esto se debe a que ambas distribuciones son *continuas*. Los supuestos distribucionales anteriores son apropiados cuando ningún hogar o muy pocos reportan gastos nulos. Sin embargo, en nuestros datos tenemos que, particularmente para ciertos grupos, por ejemplo aquéllos con ingresos bajos, un número considerable de hogares reportan gastos iguales a cero en algunos bienes (por ejemplo, durables). Entonces, para tener una especificación “válida” necesitaríamos permitir que las variables dependientes tomaran ese “valor repetido”, con una probabilidad positiva. Para lograr lo anterior, en este estudio utilizamos el Modelo Tobit, que fue propuesto originalmente por Tobin (1958).

Para introducir el modelo, suponemos que un número considerable de hogares reportaron gastos iguales a cero en el bien n . En este caso, el uso del Modelo Tobit implicaría que la f.d.p. de ε_{ni} sería igual a:

$$(9) \quad \text{f.d.p.} (\varepsilon_{ni}) = \left[(2\pi\sigma_n^2)^{-1/2} \exp \left\{ -\varepsilon_{ni}^2 / (2\sigma_n^2) \right\} \right]^{\omega_i} \left[1 - F_{ni} \right]^{1-\omega_i},$$

$$\text{donde } F_{ni} = F(x_1^* \beta_n \sigma_n^2), \omega_i = 1 \text{ si } y_{ni} > 0, \text{ y } \omega_i = 0 \text{ si } y_{ni} = 0$$

Es importante notar que, al igual que en el Modelo Truncado, bajo el Modelo Tobit tendríamos $E[\varepsilon_{ni}] \neq 0$, lo cual tiene implicaciones sobre la esperanza de y_{ni} . Como se demuestra en McDonald y Moffitt (1980) y Poirier y Melino (1978), aquí tendríamos:

$$E[y_{ni}] = x_i' \beta_n F_{ni} + \sigma_n^2 f_{ni} \quad ,$$

$$\text{y} \quad \frac{\partial E[y_{ni}]}{\partial x^*} = \beta_{2n} F_{ni} \quad .$$

Así, siguiendo argumentos similares a los presentados para el Modelo Truncado, una vez calculado el estimador del Modelo Tobit, digamos

$\tilde{\beta}_n = (\tilde{\beta}_{1n}, \tilde{\beta}_{2n})'$ y $\tilde{\sigma}_n^2$, definiríamos:

$$(10) \quad \hat{\beta}_{1n} = \tilde{\beta}_{1n} F(\tilde{\beta}_{1n}, \tilde{\sigma}_n^2) + \tilde{\sigma}_n^2 f(\tilde{\beta}_{1n}, \tilde{\sigma}_n^2)$$

$$\text{y} \quad \hat{\beta}_{2n} = \tilde{\beta}_{2n} F(\tilde{\beta}_{1n}, \tilde{\sigma}_n^2) \quad .$$

Procedimiento utilizado

En este trabajo, para la estimación del SLEG, utilizamos los modelos Truncado y Tobit que hemos presentado. A estos modelos los denominamos Modelos de Variable Dependiente Limitada (VDL). También, para efectos de comparación, se utilizó el Modelo Normal, el que se estima por MCO. En la aplicación de los modelos VDL seguimos el siguiente criterio: si para un bien dado se reportaban gastos iguales a cero, entonces utilizamos el Modelo Tobit; en los otros casos aplicamos el Modelo Truncado. Para ambos modelos utilizamos los EMV, obtenidos (esencialmente) al aplicar los algoritmos descritos en Fair (1977: 1724), separadamente para cada bien.

Teniendo calculados los EMV de los modelos Tobit y Truncado, ajustamos estos a $\hat{\beta}_1, \dots, \hat{\beta}_n$ (como en (8) y (10)); y obtuvimos $\hat{\beta}_{(n+1)}$ a través de las ecuaciones dadas en (5). Con los valores $\hat{\beta}_1, \dots, \hat{\beta}_{(n+1)}$, estimamos los elementos de $\theta = (\phi_1, \dots, \phi_n, \theta_1, \dots, \theta_n, \mu)$ como sigue (véase (3)):

$$(11) \quad \hat{\mu} = 1 - \hat{\beta}_{2(n+1)}$$

$$\hat{\theta}_j = \hat{\beta}_{2j} / (1 - \hat{\beta}_{2(n+1)})$$

$$\text{y} \quad p_j \hat{\phi}_j = (\hat{\beta}_{1j} - \hat{\beta}_{2j} \hat{\beta}_{1(n+1)}) / \hat{\beta}_{2(n+1)} \quad .$$

De aquí, utilizamos $\hat{\mu}$, $\hat{\theta}_j$, y $p_j \hat{\phi}_j$ para la estimación de elasticidades y otras cantidades que proporcionan la base sobre la cual se pueden analizar los patrones de gasto y ahorro de los hogares.

Cabe señalar que en el procedimiento de estimación VDL seguido, la forma mediante la cual se asegura que las restricciones dadas en (5) sean satisfechas, es a través de los coeficientes de la función de ahorro, habiendo —desde luego— otras opciones. Aquí decidimos seguir este camino dado que, al aplicar los modelos VDL a los n bienes (que tienen gastos no-negativos) y obtener de las restricciones los coeficientes de la ecuación de ahorro (cuyo rango no está restringido) podemos hacer una comparación directa —para cada bien— entre los resultados de la estimación MCO y EMV en el modelo VDL utilizado.¹

Es conveniente notar que si las covarianzas entre los n ε_{ji} (es decir, los elementos fuera de la diagonal en Ω — ver (6)) fueran iguales a cero, la función de verosimilitud del sistema, L , estaría definida por el producto de tres componentes. Un primer componente, L_1 , estaría constituido por el producto de $i = 1$ a N de funciones de la forma dada en (7). Esto sería para los bienes a los que se aplica el Modelo Truncado. Un segundo componente, L_2 , estaría definido por el producto de $i = 1$ a N de funciones de la forma dada en (9), para los bienes a los que se les aplica el Modelo Tobit. Finalmente, la verosimilitud de la ecuación de ahorro, L_3 .

Como se podrá apreciar, el método de estimación que hemos seguido lo podemos denominar de quasi máxima verosimilitud, dado que maximiza L_1L_2 y estima los parámetros de la ecuación de ahorro a través de las restricciones en (5). Cabe notar que la maximización de $L = L_1L_2L_3$ sujeta a (5) no fue realizada dado que requería de algoritmos de cómputo no disponibles. Tampoco se intentó la aplicación de Modelos Multivariados VDL Restringidos debido a que los resultados disponibles sobre éstos son casi nulos, a pesar de que constituyen una de las áreas con activa investigación en los últimos años.

Al comparar las propiedades de los estimadores del Modelo Normal, es decir MCO, con los de EMV en los modelos VDL aplicados, podemos observar lo siguiente:

a) Si el “problema de truncación” (o de “repetición de valores”) es *severo* en el sentido de que bajo el supuesto de Normalidad la probabilidad de obtener observaciones negativas es grande (o bien se tienen muchas observaciones iguales a cero, lo cual no puede ocurrir con una densidad continua), entonces el estimador de MCO sería inconsistente (para detalles véase Greene (1981: 510-512)). En este caso el EMV, para cada bien con el Modelo Truncado (o con el Modelo Tobit en el caso de valores repetidos) sería consis-

¹ Nótese que, por definición, el disturbio de la ecuación de ahorro se puede escribir como $\varepsilon_{(n+1)j} = -1'_n \varepsilon_j$. Al aplicar el Modelo Truncado para algunos bienes, tendríamos que el rango de sus correspondientes ε_{ji} estaría restringido. Sin embargo, para otros bienes se aplica el Modelo Tobit, para el cual se tiene $-\infty \leq \varepsilon_{ji} \leq \infty$, aunque y_{ji} sea mayor o igual a cero. Por tanto, al ser $-\varepsilon_{(n+1)j}$ la suma de todos los valores ε_{ji} , se tendría que $\varepsilon_{(n+1)j}$'s están en el intervalo irrestricto $(-\infty, \infty)$.

tente y asintóticamente eficiente siempre que no se utilice un enfoque de sistema. Asimismo, cuando se aplica un enfoque de sistema y las covarianzas en Ω son iguales o cercanas a cero, entonces esta estimación sería atractiva dado que sería de quasi máxima verosimilitud.

b) Por otra parte, si el "problema de truncación" (o de "repetición de valores") *no es severo*, la densidad Normal sería una "buena" aproximación a la verosimilitud de cada bien y el estimador de MCO sería casi igual al de EMV. En este caso no sería necesario considerar las covarianzas.

Así, vemos que, en ambos casos (cuándo la truncación es un problema severo y cuándo no lo es) el EMV para cada bien con modelos VDL tiene ciertas ventajas, o por lo menos no se puede decir que sea peor que el estimador de MCO. Por ello, resulta atractivo obtener los EMV en modelos VDL cuando se tienen los programas de cómputo necesarios. En la sección 5 presentamos ambos estimadores, el EMV sobre el Modelo Truncado o Tobit (según corresponda) y el de MCO, para analizar la magnitud empírica de las posibles diferencias.

3.3 Clasificación de hogares

En la discusión anterior supusimos que los N hogares fueron clasificados en L grupos, tal que los parámetros del SLEG fueran aproximadamente iguales para los hogares pertenecientes a un mismo grupo y el problema analizado fue cómo estimar los parámetros del modelo en un grupo dado. Ahora discutiremos cómo fueron formados dichos grupos.

La evidencia empírica muestra que el comportamiento del consumo de los hogares (analizado en términos de los parámetros en θ) depende de variables como la edad y ocupación del jefe del hogar, así como el tamaño de la familia y su nivel de ingreso (véase por ejemplo Luch, Powell y Williams (1977), Pollak y Wales (1980) y Williams (1977)). Estas variables se utilizan típicamente como criterio de estratificación para agrupar a los hogares. Sin embargo, esto se hace de una manera informal y muchas veces arbitraria, tanto en la determinación del número de grupos como en la definición de los puntos de corte en las variables de estratificación. En nuestro caso se intentó dar un tratamiento más formal al problema de clasificación de hogares. Los detalles se describen por completo en Jarque (1980, 1987) y aquí se presenta únicamente la motivación del procedimiento.

Empezamos con la ecuación (4) y para simplificar, eliminamos temporalmente el sufijo j , que usábamos para referirnos al bien j . Ahora permitimos al vector de parámetros β variar de un hogar a otro. Así, introducimos el sufijo i en β llegando a:

$$(12) \quad y_i = x_i' \beta_i + \varepsilon_i$$

En la ecuación (12) tenemos $2N$ parámetros, pero solo N observaciones están disponibles para la estimación. Por ello, para tener un modelo estima-

ble, debemos hacer supuestos adicionales con relación al proceso de generación de las β_i . Afortunadamente, como se mencionó, en nuestro caso tenemos evidencia de que ciertas variables socioeconómicas (edad y ocupación del jefe del hogar, entre otras), afectan a β_i , y es natural que las utilicemos. Suponemos que tenemos p de ellas y denotamos la i -ésima observación en esas variables por el vector z_i de dimensión p por 1.

Aunque la relación funcional precisa entre β_i y z_i no es conocida, se puede demostrar que —para propósito de agrupamiento— una aproximación lineal es razonablemente buena. Así, podemos escribir:

$$(13) \quad \beta_i = \gamma_0 + \gamma z_i + u_i$$

donde γ_0 (vector) y γ (matriz) contienen coeficientes desconocidos y u_i es el vector de términos residuales. La sustitución de (13) en (12) nos lleva a una ecuación de la cual podemos obtener estimaciones de γ_0 y γ (por ejemplo por Mínimos Cuadrados Generalizados) y, de aquí, predicciones del vector β_i (de dimensión 2 por 1), digamos $\hat{\beta}_i^+ = \hat{\gamma}_0 + \hat{\gamma}z_i$.

Esto puede repetirse para cada uno de los n bienes y las B_i^+ resultantes podrían conformarse en un vector B_i^+ de dimensión $2n$. Tendríamos entonces N vectores B_i^+ (uno para cada hogar), y el problema sería agrupar a los hogares tal que sus vectores B_i^+ fueran cercanos. Los hogares que quedaran en el mismo grupo podrían ser considerados homogéneos en términos de los parámetros del modelo. Hay que notar que el agrupamiento sería realizado en un espacio de $K = 2n$ dimensiones. Esto es importante porque deseamos que los hogares dentro de un grupo sean homogéneos en sus respuestas de demanda para *todos* los bienes y no solamente para un subconjunto de ellos.

Por supuesto, hay muchos criterios y procedimientos para agrupar N entidades con respecto a K variables. Sin embargo, en Jarque (1980), el análisis se hace en relación a este problema de estimación econométrica y se muestra que un criterio conveniente es maximizar el "poder explicativo relativo total", R^2 , el cual se define como uno menos la traza de la matriz de la suma de cuadrados dentro de los grupos de las B_i^+ estandarizadas por la desviación estándar muestral. Afortunadamente, existen algoritmos disponibles para la solución de este problema computacional (por ejemplo, véase Sparks (1973), Hartigan (1975) y Ward (1963)). Asimismo, estos algoritmos proveen de información sobre incrementos en R^2 debido a incrementos en el número de grupos L y ésta puede ser utilizada para determinar la L óptima, definida como aquella más allá de la cual sólo se obtienen muy pequeños incrementos en R^2 (nótese que cuando L es igual al número de hogares N , $R^2 = 1$ ya que, en este caso, la diagonal de la matriz de suma de cuadrados dentro de los grupos es el vector cero). Los resultados de la aplicación de este procedimiento se presentan en la subsección 5.2.

4. Los datos

Los datos utilizados para la estimación del SLEG provienen de la Encuesta de Ingreso-Gasto de los Hogares de 1975, llevada a cabo por el Centro Nacional de Información y Estadística del Trabajo, de la Secretaría del Trabajo y Previsión Social.² Para nuestro estudio utilizamos la información de 494 hogares entrevistados en la ciudad de México.

Los datos originales se dividieron en gastos en $n = 7$ bienes, denominados: alimentos, vestido, vivienda, durables, educación, servicios médicos y otros. El regresor utilizado x^* es el ingreso familiar disponible. Cabe señalar que todas las variables de gasto y de ingreso estaban dadas en pesos por mes, y que antes de la estimación fueron expresadas en términos per capita. Respecto a las *variables socioeconómicas*, contienen tamaño de la familia, y edad y ocupación del jefe del hogar. Esta última variable fue dividida en cuatro categorías: obrero, comerciante, técnico y desempleado (categoría que incluye a los económicamente inactivos). Los resultados empíricos se dan en la siguiente sección.

5. Resultados empíricos

5.1 Prueba de variación sistemática en los parámetros

Primero consideramos los 494 hogares en muestra, y definimos y_{ji} , β_j , x_i^j y ε_{ji} como en (4). Ahora tenemos $n = 7$ y $N = 494$. También definimos $z_i^j = (z_{i1}, z_{i2}, z_{i3}, z_{i4})$, donde z_{i1} es la ocupación del jefe del hogar (0 si es desempleado, 1 si es técnico, 2 si es comerciante y 3 si es obrero); z_{i2} es el cuadrado del ingreso,³ z_{i3} es el tamaño de la familia; y z_{i4} es la edad del jefe, cada una de ellas referida al i -ésimo hogar o familia.

Tomamos una ecuación determinada, por ejemplo la ecuación j , es de-

² En esta encuesta se utilizó un diseño de muestra estratificado y multietápico. Una descripción completa de las características de la encuesta se da en CBNJET (1977). Es importante notar que, en análisis econométrico de regresión, basado en datos de corte transversal provenientes de una encuesta por muestreo, no resulta indispensable tomar en cuenta el diseño muestral en la obtención de estimadores del modelo, aunque sí en el cálculo de sus varianzas (por ejemplo, véase el capítulo 2 de la tesis doctoral del autor, titulada *Contributions to The Econometrics of Cross-Sections*, presentada en 1981 en la Facultad de Economía de la Universidad Nacional de Australia). Lo anterior se debe a que si un conjunto de hogares en muestra tienen *la misma* β , no es necesario ponderar las observaciones por el número de hogares que cada observación representa. Esto no debe confundirse con la "ponderación" que uno podría hacer a las β estimadas en cada uno de los L grupos para obtener una β promedio para el total de la población, en cuyo caso se ponderaría por la importancia de cada grupo en la población estudiada.

³ Utilizamos el ingreso al cuadrado $(x^*)^2$ en lugar del ingreso (x^*) para evitar multicolinealidad en (14). Recordar que $x_i^j = (1, x_i^j)$.

cir $y_{ji} = x_i' \beta_j + \varepsilon_{ji}$, y postulamos que β_j para el i -ésimo hogar se relacionaba con z_i . Primero formulamos una relación funcional lineal entre β_j y z_i ; digamos $\beta_j = \Gamma_j(1, z_i')$. Así, obtenemos el "modelo aumentado de regresión":

$$(14) \quad y_{ji} = (x_i', (z_i' \otimes x_i')) \gamma_j + u_{ji+}$$

donde $u_{ji+} = x_i' \varepsilon_{ji}$ y $\gamma_j = \text{Vec} \{ \Gamma_j \}$

Estimamos la ecuación (14) mediante mco y probamos la hipótesis de variación sistemática en los parámetros (es decir, la significancia de los coeficientes de los regresores $(z_i' \otimes x_i')$) utilizando el estadístico F correspondiente (véase White, 1980: 820). Los resultados mostraron que, salvo en el caso de la regresión para servicios medicos, hay evidencia de la existencia de variación en los parámetros (todos los valores de F —excepto los de servicios médicos— exceden el punto 20.1 correspondiente al punto crítico a 1% en una x_8^2). Lo anterior sugiere que la estimación de (4) empleando todas las $N = 494$ observaciones, discriminando el efecto de las variables socioeconómicas z_i , ocasionaría errores de especificación en el modelo.

Es posible detenerse en esta etapa y basar el análisis en los resultados del modelo que se proporciona en (14). Sin embargo, aquí no nos limitamos a la forma lineal en la variación paramétrica. Más bien, usamos un procedimiento *bietápico*, en el cual la *primera etapa* consiste en la clasificación de los hogares en grupos homogéneos, como se describió en la subsección 3.3, y la *segunda etapa* se refiere a la estimación del SLEG, como se describió en la subsección 3.2.

5.2 Primera etapa: clasificación de hogares

El punto de partida en la clasificación de los hogares es la determinación del número de grupos en los que deben clasificarse. Para ello, evaluamos el criterio R^2 en la clasificación C^* que maximiza su valor para un amplio rango de valores L (a saber, $L = 2, 3, \dots, 20$) (para calcular C^* utilizamos el algoritmo de Sparks (1973), que se basa en el procedimiento de las K -medias de Beale (1969)).

Los resultados se proporcionan en el cuadro 1. Aunque la elección de L es un tanto arbitraria, decidimos tomar $L = 14$ debido a que arroja un valor de R^2 superior a 0.9 y, además, porque más allá sólo se obtienen modestos incrementos. (Por ejemplo, el incremento en R^2 de $L = 13$ a $L = 14$ sería de 0.015; en forma similar, el incremento de $L = 14$ a $L = 15$ sería sólo de 0.001).

Se tomó la clasificación C^* correspondiente a $L = 14$ y se calcularon algunas características de los grupos que la integran. Los resultados se presentan en el cuadro 2. En la columna uno se da la categoría ocupacional que

CUADRO 1

Poder explicativo relativo total

<i>L</i>	<i>R</i> ²	ΔR^2	<i>L</i>	<i>R</i> ²	ΔR^2
1	0.000	—	11	0.871	0.010
2	0.453	0.453	12	0.884	0.013
3	0.618	0.165	13	0.895	0.011
4	0.699	0.081	14	0.910	0.015*
5	0.740	0.041	15	0.911	0.001
6	0.780	0.040	16	0.912	0.001
7	0.827	0.047	17	0.913	0.001
8	0.843	0.016	18	0.916	0.003
9	0.857	0.014	19	0.919	0.003
10	0.861	0.004	20	0.919	0.000

ocurre más frecuentemente en cada grupo y su frecuencia correspondiente. Por ejemplo, el grupo 2 está integrado por 23 hogares de los cuales 22 tienen como jefe de familia a un cesepleado (D). Utilizamos *O*, *C* y *T* para denotar, respectivamente, al obrero, comerciante y técnico). Otras de las variables que se incluyen en el cuadro son ingreso familiar promedio, tamaño medio de la familia y edad promedio del jefe de familia. *Una interesante característica del agrupamiento es la marcada separación de los hogares por categorías ocupacionales, más que por niveles de ingreso* (por ejemplo, los grupos 1 y 3 tienen valores similares para el ingreso familiar, el tamaño y la edad del jefe de familia y difieren debido a que el grupo 1 está formado por desempleados en tanto que el grupo 3 lo integran hogares de obreros). Esto parece indicar que la ocupación ejerce una de las principales influencias en la determinación del comportamiento de gasto.

Por supuesto podríamos estimar el SLE para cada uno de estos grupos de hogares de "comportamiento homogéneo" los cuales son "óptimos" desde un punto de vista estadístico. Sin embargo, en algunos casos se tienen pocas observaciones para la estimación econométrica, por ejemplo, en los grupos 13 y 14 tenemos, respectivamente 5 y 4 hogares (véase la última columna del cuadro 2). Asimismo, en otros grupos tenemos dos o más categorías ocupacionales; por ejemplo, el grupo 11 tiene 21 técnicos, 2 comerciantes, 2 obreros y 1 desempleado. Con base en los resultados del análisis de conglomerados, decidimos redefinir los grupos para tener (i) un número suficiente de observaciones para la estimación y (ii) grupos de estudio fácilmente identificables.

Esto precisó la fusión de algunos grupos y la reasignación de algunos hogares dentro de los grupos. Como se señaló anteriormente, la división más importante de los hogares es por categoría ocupacional, así que este fue el punto de partida más obvio para la definición final de los grupos.

CUADRO 2

Características de grupos de análisis de conglomerados

<i>Grupo</i>	<i>Ocupación</i>	<i>Ingreso</i>	<i>Tamaño de la familia</i>	<i>Edad</i>	<i>Número en el grupo</i>
1	D(23/23)	572.	8.0	55.7	23
2	D(22/23)	801.	3.3	65.9	23
3	O(41/42)	540.	9.8	49.0	42
4	O(70/70)	464.	7.1	38.3	70
5	O(68/69)	835.	3.6	31.8	69
6	O(47/47)	859.	4.4	54.6	47
7	C(13/28)	668.	10.4	45.9	28
8	C(41/41)	911.	5.4	42.2	41
9	T(62/65)	1154.	5.8	38.9	65
10	T(37/41)	1701.	3.1	31.9	41
11	T(21/26)	3534.	3.6	39.4	26
12	T(6/10)	5538.	3.1	49.3	10
13	T(5/5)	7196.	3.0	44.8	5
14	T(4/4)	9614.	4.0	41.0	4

Hogares de desempleados

En lo referente a los hogares de desempleados, el desglose lógico fue por tamaño de la familia, con un valor de división igual a 6 (los "grupos mayormente desempleados", es decir los grupos 1 y 2, eran tales que el grupo 1 tenía hogares de tamaño básicamente mayores o iguales a 6 y el grupo 2 menores que 6. Asimismo, no había diferenciales de ingreso o edad claras entre estos grupos). Definimos los hogares grandes (G) como aquellos que tienen tamaños de familia mayores o iguales a 6 y los pequeños (P) como aquellos con familias menores a 6. Por lo tanto, dividimos los hogares desempleados en 2 grupos: G y P.

Hogares de obreros

En lo referente a los hogares de obreros nos encontramos con que los grupos 3 y 4 eran hogares de tamaño grande (G); en el grupo 3 la mayoría de los jefes de familia tenían más de 45 años de edad y en el grupo 4 la mayoría era menor de 45 años. Definimos los hogares en los que el jefe de familia es mayor de 45 años como viejos (V) y aquellos en los que el jefe de familia es menor de 45 años como jóvenes (J). (Resulta interesante observar que Llich, Powell y Williams, 1977: 122) emplearon este punto de corte

en su clasificación por edades. También observamos que los grupos 5 y 6 eran hogares de tamaño pequeño (P) estando integrado el grupo 5 por hogares jóvenes (J) y el grupo 6 por hogares viejos (V). (En general, los grupos 3 y 4 contenían hogares con ingresos menores que los grupos 5 y 6, reflejando con ello que los hogares pequeños tienen ingresos más altos.) Debido a lo anterior, decidimos dividir los hogares de obreros en 4 grupos: GJ, GV, PJ y PV.

Hogares de comerciantes

En lo que respecta a los comerciantes, tenemos que el grupo 7 estaba formado por hogares grandes (G) y el grupo 8 por hogares principalmente pequeños (P). (Los diferenciales de ingreso y edad fueron menos significativos entre estos grupos, salvo por el hecho de que los hogares de tamaño pequeño tienden a recibir ingresos ligeramente más altos.) Por lo tanto, dividimos los hogares de comerciantes en 2 grupos: CG y CP.

Hogares de técnicos

Por último, en lo tocante a hogares de técnicos, observamos que los grupos 9 y 10 eran básicamente hogares de "bajo ingreso" (con respecto a otros hogares de técnicos), con ingresos menores a 3 000 pesos mensuales *per capita*. Los hogares con ingresos por debajo de este nivel se denotan mediante el símbolo 11. Entre los grupos 9 y 10 no se evidencian diferencias claras en la edad; pero el grupo 9 estuvo formado básicamente por hogares de tamaño grande (G) y el grupo 10 por hogares de tamaño pequeño (P).

Adicionalmente, el grupo 11 estaba formado por hogares de "ingresos medios", con ingresos entre 3 000 y 5 000 pesos. Los hogares con ingresos dentro de este intervalo se denotan mediante el símbolo 12. Por último, teníamos los hogares de "altos ingresos" en los grupos 12, 13 y 14 cuyos ingresos eran superiores a los 5 000 pesos. Estos se denotan con el símbolo 13. Así, integramos 4 grupos (obsérvese que anteriormente teníamos 6) de hogares de técnicos: 11G, 11P, 12 e 13.

En resumen, llegamos a 12 grupos de estudio que contenían hogares con un comportamiento de consumo relativamente homogéneo, determinados éstos por ocupación y tamaño de la familia, y edad (en el caso de obreros) e ingreso (en el caso de técnicos).

El siguiente paso consistió en estimar el SLEG para cada uno de los 12 grupos.

5.3 Segunda etapa: estimación del SLEG

Para cada uno de los 12 grupos, estimamos el SLEG mediante el Modelo Normal, esto es MCO, y EMV en modelos VDL. Respecto a los modelos VDL, para

los rubros alimentos, vestido, vivienda y otros, empleamos el Modelo Truncado; para los bienes durables, educación y servicios médicos se tuvieron reportes repetidos de gastos nulos, así que utilizamos el Modelo Tobit. Los resultados de las ecuaciones estimadas reflejaron los mismos comportamientos para los 12 grupos, así que sólo necesitamos discutir uno de ellos en detalle. Nos centraremos en el cuadro 3, que presenta los resultados para los hogares de obreros de tamaño pequeño y edad viejo.

Estimación por mco

La primera mitad del cuadro 3 contiene los resultados de estimaciones por mco en el Modelo Normal de regresión ordinaria (véase la subsección 3.2 (i)). Para cada bien, los números en paréntesis debajo de las estimaciones

CUADRO 3

Regresiones estimadas: obrero, pequeño, viejo

	<i>Alimentos</i>	<i>Vestido</i>	<i>Vivienda</i>	<i>Durables</i>	<i>Educación</i>	<i>Servicios médicos</i>	<i>Otros</i>
Resultados utilizando el modelo de regresión ordinaria (MCO)							
$\hat{\beta}_1$ (MCO)	178.0 (2.65)	-25.7 (-.75)	148.8 (4.65)	-0.88 (-.19)	-10.0 (-.78)	-9.8 (-1.35)	-65.3 (-2.53)
$\hat{\beta}_2$ (MCO)	0.300 (3.63)	0.125 (2.71)	0.053 (2.57)	0.013 (2.74)	0.041 (3.62)	0.030 (3.46)	0.249 (7.37)
SD_u	184.4	81.6	137.8	19.7	46.3	25.0	92.6
LM_N	0.62	0.27	7.87	110.29*	183.86*	23.34*	2.38
LM_H	7.03	27.87*	0.10	1.38	0.10	4.21	7.93
LM_{NH}	7.65	28.14*	7.98	111.67*	183.97*	27.55*	10.32
$F(x^2 \hat{\beta}, \sigma^2)$	0.008	0.153	0.077	0.291	0.284	0.254	0.051
Resultados utilizando el modelo VDL							
$\tilde{\beta}_1$ (VDL)	154.2 (2.56)	-204.3 (-1.65)	-5.1 (-0.03)	-12.9 (-1.44)	-31.6 (-1.55)	-14.7 (-1.55)	-173.8 (-2.73)
$\tilde{\beta}_2$ (VDL)	0.312 (6.18)	0.193 (3.96)	0.093 (1.37)	0.017 (2.34)	0.045 (2.76)	0.030 (3.79)	0.299 (8.19)
SD_u	184.5	92.2	140.3	19.8	46.4	25.8	97.5
$LM_{N(\cdot)}$	0.02	3.22	3.15	130.06*	229.69*	1776.03*	2.20
$LM_{H(\cdot)}$	5.35	0.02	0.16	4.10*	15.63*	55.10*	0.25
$LM_{NH(\cdot)}$	6.29	4.93	4.88	137.94*	246.04*	1777.27*	2.27
$\hat{\beta}_1$	223.7	43.0	160.4	5.6	12.2	6.0	46.5
$\hat{\beta}_2$	0.280	0.057	0.042	0.009	0.025	0.019	0.170

$\hat{\beta}_1$ (MCO) y $\hat{\beta}_2$ (MCO), se refieren al estadístico t que se obtiene utilizando la MVC de MCO que es compatible con heteroscedasticidad; y SD_u contiene la raíz cuadrada de la suma de residuales estimados al cuadrado entre el tamaño de muestra.

El renglón correspondiente a LM_N , contiene los valores del estadístico para probar normalidad (N) de los disturbios, sugerido en Jarque y Bera (s/f). Con la hipótesis mantenida de homoscedasticidad (H), LM_N se distribuye asintóticamente como χ^2_2 bajo N . En el cuadro, LM_H es el estadístico de homoscedasticidad (véase Breusch y Pagan, 1979), calculado bajo el supuesto de que la varianza de los disturbios es una función del ingreso. Dado N , LM_H se distribuye asintóticamente como χ^2_1 bajo H . En seguida, $LM_{NH} = LM_N + LM_H$ es la prueba estadística conjunta LM para normalidad y homoscedasticidad NH de los disturbios sugerida en Jarque y Bera (1980). Bajo NH , LM_{NH} se distribuye asintóticamente como χ^2_2 .

Al 1% del nivel de significancia, tenemos $\chi^2_1(.99) = 6.63$, $\chi^2_2(.99) = 9.21$ y $\chi^2_2(.99) = 11.3$. Utilizamos estos puntos y denotamos a los valores estadísticos significativos de LM_N , LM_H y LM_{NH} con una estrella (*). Así, observando el renglón para LM_{NH} , encontramos que disturbios NH se rechazan para vestido, durables, educación y servicios médicos. Es interesante notar que éstos son los bienes para los cuales las probabilidades bajo NH de obtener una observación negativa son mayores (véase el renglón correspondiente a $F(\bar{x}'\hat{\beta}, \hat{\sigma}^2)$ y obsérvese que las probabilidades son 0.153, 0.291, 0.284 y 0.254, respectivamente). Aparte de vestido, la contribución más importante a LM_{NH} proviene de LM_N . Por ejemplo, para durables tenemos $LM_{NH} = 111.67$, con $LM_N = 110.29$ y $LM_H = 1.38$. Para durables, educación y servicios médicos, aproximadamente la mitad de los hogares reportaron gastos nulos, así que no es sorprendente que (bajo el supuesto de que H se mantiene) la normalidad sea rechazada. De lo anterior es claro que para estos bienes se debe utilizar un modelo tipo Tobit.

Estimación por VDL

Ahora observamos la segunda mitad del cuadro 3. Ésta presenta los resultados de EMV de los parámetros utilizando modelos VDL. Para cada bien j , los valores $\hat{\beta}_1$ (VDL) y $\hat{\beta}_2$ (VDL) denotan los EMV de β_{1j} y β_{2j} . Como se mencionó, para alimentos, vestido, vivienda y otros, estos se basan en la función de verosimilitud de un Modelo Truncado; y para durables, educación y servicios médicos, en la función de verosimilitud del Modelo Tobit. En ambos casos, utilizamos esencialmente los algoritmos de Fair (1977: 1724). Debajo de los valores de $\hat{\beta}_1$ (VDL) y $\hat{\beta}_2$ (VDL), se da en paréntesis la prueba estadística t para muestras grandes; SD_u denota la raíz cuadrada del segundo momento muestral acerca de la media de los residuales estimados; $LM_{N(\cdot)}$, $LM_{H(\cdot)}$, $LM_{NH(\cdot)}$ denotan las pruebas estadísticas para disturbios N y/o H en mode-

los VDL, es decir, estos son $LM_{N(\text{Trun})}$, $LM_{H(\text{Trun})}$ y $LM_{NH(\text{Trun})}$, o $LM_{N(\text{Tobit})}$ y $LM_{H(\text{Tobit})}$, $LM_{NH(\text{Tobit})}$ sugeridos en Jarque y Bera (1982) (véase también Bera, Jarque y Lee, 1984) de acuerdo con el modelo que se utilice. Nuevamente, para la prueba de homoscedasticidad, especificamos que la varianza era una función del ingreso. Valores con una estrella (*) denotan significancia al 1%.

Primero consideramos las estimaciones para la ecuación de gasto en *alimentos*. Para esta ecuación vemos que $F(\bar{x}'\hat{\beta}, \hat{\sigma}^2)$ es pequeña (igual a 0.008), así que no se espera tener serios problemas al olvidar, en la especificación de la verosimilitud, el hecho de que el gasto es no-negativo. Éste es el caso, y los resultados de MCO son muy similares a los resultados EMV en modelos VDL. Para ver esto, podemos comparar los valores $\hat{\beta}$, SD_u y LM entre ambas técnicas de estimación. En los 12 tipos de hogares, tenemos que, para alimentos la probabilidad $F(\bar{x}'\hat{\beta}, \hat{\sigma}^2) \leq .025$, y encontramos que los resultados por MCO y por VDL son muy similares; así, podemos concluir que —para este bien— la truncación no es un problema severo.

Observando la ecuación de gasto en vestido, tenemos que $F(\bar{x}'\hat{\beta}, \hat{\sigma}^2)$ es 0.153. Para este bien, tuvimos un valor significativo de LM_H ; pero utilizando el Modelo Truncado, aceptaríamos NH (truncada) eliminando este problema. Similarmente, la ecuación estimada para vivienda, no da evidencia de no-normalidad (N) ni de heteroscedasticidad (H) en los disturbios. Lo mismo ocurre con la ecuación para otros.

Ahora analizamos los tres bienes para los cuales se utilizó el modelo Tobit; estos son, durables, educación y servicios médicos. Aquí, encontramos que en los tres casos $LM_{NH(\text{Tobit})}$ rechazaría NH (véase el renglón $LM_{NH(\cdot)}$ en el cuadro 3). Vale la pena notar dos puntos con relación a este resultado:

i) El primer punto se refiere a las propiedades de los estadísticos cuando son calculados con *muestras finitas*. En un modelo de regresión ordinario se pueden obtener los puntos de significancia en muestras finitas para LM_{NH} por simulación (ver Bera y Jarque, 1982). Esto quiere decir que, para cualquier tamaño de muestra y conjunto de regresores, podemos realizar pruebas con niveles de significancia prácticamente iguales a los deseados. En contraste, para las pruebas $LM_{NH(\text{Trun})}$ y $LM_{NH(\text{Tobit})}$ sólo se conoce la distribución asintótica; y basar estas pruebas en puntos de significancia asintóticos, nos puede llevar a niveles considerablemente diferentes de los supuestos. En el presente estudio, debido a la división de la muestra para alcanzar grupos de hogares homogéneos en su comportamiento de demanda, el tamaño de muestra varía entre 13 y 70, y muy bien podría ser que éstos no fueran suficientemente grandes para hacer un uso válido de la distribución asintótica. Sin embargo, los desarrollos actuales de análisis de este tipo de modelos VDL nos limita al uso de resultados para muestras grandes. Esperamos que futuras investigaciones en este tema lleven a la obtención de la distribución en muestras finitas para estos estadísticos de prueba.

ii) En segundo lugar, debemos apuntar que, en algunos casos, obtuvimos valores “extremadamente grandes” en los estadísticos; por ejemplo,

obtuvimos $LM_{NH(Tobit)} = 1777.27$ para servicios médicos (ver cuadro 3).^{4,5}

Uno podría argumentar que esos valores, aun después del ajuste de los puntos de significancia asintóticos, nos llevarían a rechazar la hipótesis NH . En estos casos se puede volver a estimar la ecuación utilizando EMV con una verosimilitud que incorpore \bar{N} (por ejemplo, a través del uso de una familia de transformaciones) y/o \bar{H} . Ésta es un área que requiere de mayor investigación y cuyos descubrimientos proporcionarán un avance muy significativo en metodología econométrica.

Hasta este punto los resultados indican que en estudios microeconómicos de demanda con datos a nivel de hogar, el problema de truncación puede ser severo y, por consiguiente, el uso de MCO proporciona estimadores de β inconsistentes. Asimismo, el supuesto de normalidad es rechazado con frecuencia al aplicar el modelo clásico de regresión, indicando que el uso de modelos VDL puede ser más apropiado. Al aplicar estos modelos obtuvimos que, especialmente para el Modelo Tobit, se rechaza el supuesto de normalidad truncada lo cual resalta la importancia de investigar el análisis de estos modelos bajo supuestos alternativos.

La presencia de \bar{N} y \bar{H} hace inconsistentes los estimadores $\tilde{\beta}_1$ (VDL) y $\tilde{\beta}_2$ (VDL). En nuestro estudio, con frecuencia se encontró una fuerte evidencia de violación de NH (es decir un valor muy grande de $LM_{NH(\cdot)}$) en las ecuaciones de durables, educación y servicios médicos. Por ello, lo apropiado del subsecuente análisis depende de la magnitud de la inconsistencia. En sentido estricto las estimaciones para estos bienes *deben ser tomadas con precaución*.

Dada la estructura del SLEG, los coeficientes β calculados para un bien, también tomarían parte en el cálculo de respuestas de demanda para otros bienes (por ejemplo, ver la ecuación [11]). Esto parecería indicar que el uso de los resultados basados en modelos VDL quedaría aquí. Afortunadamente, los coeficientes asociados a las ecuaciones de durables, educación y servicios médicos son pequeños en magnitud relativa y las fluctuaciones en éstos tienen un efecto pequeño en la demanda estimada y respuestas de ahorro para otros bienes (por ejemplo, véase el cuadro 3 y nótese que $\tilde{\beta}_2$ (VDL) para durables, educación y servicios médicos, es respectivamente 0.017, 0.045 y 0.030; mientras que para alimentos, vestido, vivienda y otros, tenemos 0.312, 0.193, 0.093 y 0.299). Así, decidimos proceder con nuestro análisis dejando de lado la posible presencia de inconsistencia en los estimado-

⁴ Estos valores grandes fueron totalmente inesperados. Para estar seguros de que nuestros programas computacionales no eran erróneos, generamos datos bajo NH con $N = 100$ y calculamos la prueba estadística $LM_{NH(Tobit)}$. Los valores obtenidos fueron de magnitud razonable (relativa a la distribución asintótica correspondiente χ^2). En esta simulación encontramos que al utilizar el punto crítico asintótico, el porcentaje de rechazos de NH estuvo muy cerca del nivel de significancia teórico.

⁵ Los valores extremadamente grandes de pruebas estadísticas ocurrieron en casos en los que muchos hogares de "altos ingresos" reportaron "cero gastos", lo cual es un evento al que la hipótesis nula (NH) asigna una muy baja probabilidad.

res de máxima verosimilitud de modelos vDL, habiendo notado que, en todo caso, los resultados se deben tomar con cautela.

Como se mencionó en la sección 3, los coeficientes estimados en los modelos vDL necesitan ser ajustados de tal modo que representen las respuestas apropiadas. Por consiguiente calculamos, para cada bien, los valores $\hat{\beta}_{1j}$ y $\hat{\beta}_{2j}$ utilizando la ecuación (8) para alimentos, vestido, vivienda, y otros; y la ecuación (10) para durables, educación y servicios médicos. Los valores obtenidos se dan en los dos últimos renglones del cuadro 3 (para tipo de hogar obrero PV). Es evidente que la falta de ajuste de los coeficientes puede conducir a resultados engañosos; por ejemplo, todos los valores $\hat{\beta}_1$ (vDL) menos uno son negativos; no obstante, cuando se ajustan, todos pasan a ser positivos (ver el renglón correspondiente a $\hat{\beta}_1$), propiedad que, dicho sea de paso, no tuvieron los estimadores mco (ver primer renglón del cuadro 3). Los valores ajustados fueron entonces utilizados para el cálculo de los elementos en θ los cuales representan respuestas de ahorro y demanda (ver las ecuaciones (11)) y para el cálculo de elasticidades. Los resultados se discuten en la subsección 5.4.

5.4 Respuestas estimadas de demanda

Participación marginal presupuestal y propensión marginal al consumo.

Los valores estimados de las participaciones marginales presupuestales (θ_j) se dan en el cuadro 4 para cada uno de los 12 grupos socioeconómicos. Nuevamente, la primera mitad contiene resultados utilizando mco, y la segunda mitad incluye aquellos utilizando la estimación de máxima verosimilitud en modelos vDL (en adelante se referirá como estimación vDL-MV). De las 84 estimaciones tenemos que 5 estimaciones de mco y 5 de vDL-MV son negativas, y que éstas ocurrieron en los casos donde los gastos reportados fueron muy pequeños.

En la última columna del cuadro 4, observamos que todas las propensiones marginales al consumo estimadas ($\hat{\mu}$) fueron menores que uno, y que, para cada grupo socioeconómico, la $\hat{\mu}$ estimada utilizando mco fue mayor a la de la estimación vDL-MV. Por ejemplo, en el grupo comerciante P, el estimador de mco es 0.840, el cual es 30% mayor que el estimador de vDL-MV, que es igual a 0.646. *Esto parece sugerir que en estudios de presupuestos familiares, la omisión en el análisis econométrico del hecho de que los gastos son no-negativos, puede llevar a una considerable sobreestimación de la propensión marginal al consumo.* Se podría argumentar que los modelos vDL utilizados son muy particulares (suponen que Ω es diagonal) y que la bondad de mco no necesariamente debe juzgarse en comparación con éstos. Si bien lo anterior es cierto, el hecho de haber obtenido diferencias importantes resalta la necesidad de estudiar detenidamente las técnicas econométricas utilizadas en estudios de demanda.

CUADRO 4

**Participaciones marginales presupuestales (θ_j) y
propensión marginal al consumo (μ)**

Tipo de hogar	Alimentos	Vestido	Vivienda	Durables	Educación	Servicios médicos	Otros	μ
Resultados utilizando estimación MCO								
<i>Desempleado</i>								
G	0.480	0.124	0.110	0.010	0.008	-0.004	0.271	0.485
P	0.404	0.098	0.201	0.001	0.030	0.087	0.178	0.664
<i>Obrero</i>								
G,J	0.409	0.198	0.205	0.020	0.043	-0.010	0.133	0.634
G,V	0.230	0.175	0.175	0.032	-0.003	0.012	0.376	0.616
P,J	0.371	0.149	0.213	0.000	0.020	0.012	0.232	0.612
P,V	0.368	0.154	0.066	0.016	0.051	0.037	0.305	0.815
<i>Comerciante</i>								
G	0.473	0.127	0.133	0.001	0.004	-0.001	0.262	0.715
P	0.301	0.222	0.154	0.033	-0.007	0.001	0.278	0.840
<i>Técnico</i>								
11,G	0.180	0.077	0.331	0.023	0.051	0.027	0.308	0.835
11,P	0.187	0.192	0.236	0.048	0.023	0.061	0.253	0.735
12	0.165	0.121	0.180	0.076	0.073	0.091	0.294	0.601
13	0.125	0.101	0.190	0.091	0.081	0.087	0.325	0.528
Resultados utilizando estimación VDL-MV								
<i>Desempleado</i>								
G	0.552	0.096	0.127	0.007	0.036	0.002	0.178	0.416
P	0.474	0.080	0.178	0.001	-0.002	0.083	0.180	0.561
<i>Obrero</i>								
G,J	0.502	0.160	0.170	0.021	0.033	-0.009	0.120	0.499
G,V	0.277	0.154	0.185	0.030	-0.003	0.013	0.342	0.525
P,J	0.431	0.148	0.192	0.002	0.013	0.014	0.195	0.494
P,V	0.471	0.093	0.068	0.014	0.041	0.031	0.278	0.611
<i>Comerciante</i>								
G	0.511	0.113	0.126	0.003	0.005	-0.002	0.242	0.593
P	0.393	0.160	0.170	0.037	-0.001	0.003	0.235	0.646
<i>Técnico</i>								
11,G	0.209	0.084	0.293	0.022	0.039	0.022	0.327	0.721
11,P	0.191	0.203	0.239	0.031	0.029	0.050	0.257	0.687
12	0.176	0.159	0.185	0.080	0.066	0.083	0.251	0.503
13	0.137	0.120	0.131	0.094	0.080	0.089	0.349	0.412

Analizando las participaciones marginales presupuestales, observamos que una sobre o subestimación puede ocurrir. Por ejemplo, las estimaciones por MCO de la participación marginal presupuestal para alimentos son más pequeñas que las de VDL-MV; mientras que las de otros son mayores con MCO (excluyendo a técnicos). En resumen, estos resultados parecen sugerir que se debe proceder con cautela al utilizar resultados cuantitativos de estu-

dios a nivel hogar donde no se contempla la naturaleza no-negativa de los gastos, y se estima el modelo con MCO.

Gastos de subsistencia

Ahora consideramos los gastos de subsistencia estimados, es decir $p_j\phi_j$ (ver (11)). Éstos se dan en el cuadro 5 para cada uno de los 12 grupos, en donde observamos que todos son positivos. Fue interesante encontrar que, para

CUADRO 5

Gastos de subsistencia estimados ($p_j\phi_j$)

Tipo de hogar	Alimentos	Vestido	Vivienda	Durables	Educación	Servicios		Total
						médicos	Otros	
Resultados utilizando estimación MCO								
<i>Desempleado</i>								
G	337.5	76.4	124.2	18.8	8.4	14.7	165.3	745.3
P	385.7	113.4	285.0	10.7	14.3	27.8	215.7	1052.6
<i>Obrero</i>								
G,J	338.6	97.9	120.2	14.3	25.5	8.5	95.1	700.3
G,V	337.6	98.2	139.3	14.2	17.6	10.8	179.5	797.4
P,J	425.0	116.1	193.0	17.9	12.0	16.2	179.9	960.4
P,V	528.4	120.8	211.6	14.8	38.8	25.5	225.5	1165.8
<i>Comerciante</i>								
G	326.7	104.2	170.0	13.9	25.1	9.3	167.6	816.6
P	503.4	160.7	274.8	31.1	17.6	25.7	251.8	1265.4
<i>Técnico</i>								
11,G	457.2	139.4	318.0	40.8	51.2	23.8	321.8	1352.4
11,P	583.1	221.9	350.2	53.8	27.3	45.6	411.0	1692.9
12	635.2	504.3	517.8	84.3	88.7	67.9	596.0	2494.2
13	670.0	603.0	1001.9	180.4	100.2	83.1	803.4	3442.0
Resultados utilizando estimación VDL-MV								
<i>Desempleado</i>								
G	328.3	81.4	127.5	21.4	15.1	17.8	174.6	766.1
P	415.6	143.2	308.7	13.2	17.2	43.1	243.0	1184.0
<i>Obrero</i>								
G,J	348.3	91.7	113.5	15.2	39.4	14.7	91.5	714.7
G,V	353.7	109.1	155.8	19.4	21.7	12.8	216.3	889.1
P,J	463.0	131.0	208.4	19.2	15.2	20.5	197.3	1054.8
P,V	593.5	116.6	214.5	17.1	44.6	31.1	264.8	1282.6
<i>Comerciante</i>								
G	344.3	106.7	187.4	16.8	35.3	12.9	179.0	882.4
P	615.3	196.3	319.2	47.9	21.4	28.2	314.3	1542.7
<i>Técnico</i>								
11,G	559.6	190.3	513.5	57.0	88.0	39.8	557.3	2005.9
11,P	603.1	251.4	401.4	61.8	40.9	61.0	431.9	1851.5
12	703.7	631.0	531.5	80.1	132.0	103.2	872.4	3053.9
13	780.5	702.8	1132.3	201.3	97.6	81.5	891.7	3887.7

cada bien, los gastos de subsistencia estimados con MCO fueron menores a los de la estimación VDL-MV. A menudo la subestimación por MCO fue considerable. Por ejemplo, para el grupo socioeconómico de comerciantes P el gasto total de subsistencia fue 1265.4 utilizando MCO, el cual es 21% menor que la estimación de VDL-MV, igual a 1542.7 (los valores están en pesos de 1975 *per capita* al mes). *Esto sugiere que si se aplica MCO en casos donde se debe utilizar un modelo VDL, podemos subestimar severamente los gastos de subsistencia de los hogares.* El resultado es importante porque si las estimaciones por VDL-MV son correctas, significaría que los hogares tienen menor libertad al asignar su ingreso (ya que una mayor proporción está comprometida) con relación a lo que las estimaciones por MCO sugerirían.

Con las estimaciones presentadas se pueden calcular elasticidades de demanda, de gasto total, precio y ahorro (ver Lluch, Powell y Williams, 1977: 16-20). Desde luego, como era de esperarse, los resultados numéricos difieren dependiendo del uso de θ_j , μ y $p_j\phi_j$ obtenidos por MCO o por VDL-MV. Sin embargo, los resultados cualitativos, es decir, el estudio de tendencias o patrones nos lleva —en general— a las mismas conclusiones. Un resumen de estas conclusiones se presenta en la segunda parte de la siguiente sección.

6. Conclusiones

En este trabajo, estimamos el SLEG para estudiar patrones de gasto en la ciudad de México. Los datos se obtuvieron de la Encuesta de Ingreso-Gasto de los Hogares de 1975.

6.1 Conclusiones Metodológicas

Los puntos principales que surgieron de la metodología econométrica empleada fueron:

1) Los *parámetros* del modelo están relacionados con variables socioeconómicas y el no tomarlo en cuenta llevaría a errores de especificación en el modelo de demanda. En particular, la estimación de un solo sistema con datos de todos los hogares en muestra es inapropiado.

2) Al formar grupos de consumidores homogéneos mediante *Análisis de Conglomerados*, algunos hogares tuvieron que ser reasignados para obtener dominios de estudio fácilmente identificables y un número suficiente de observaciones en cada grupo.

3) Los resultados del *Análisis de Conglomerados* muestran que la categoría ocupacional es uno de los factores de mayor influencia en el comportamiento de demanda.

4) Cuando la probabilidad de obtener una observación negativa bajo normalidad $F(\bar{x}'\beta, \sigma^2)$, es menor a 0.025, los resultados de MCO y EMV en el *Modelo Truncado* son básicamente los mismos. En estos casos, olvidar el

hecho de que los gastos son no-negativos y por lo tanto utilizar MCO, no necesariamente lleva a severas inconsistencias. En otros casos donde $F(\bar{x}'\hat{\beta}, \hat{\sigma}^2)$ es superior a 0.025, encontramos que se pueden tener diferencias significativas entre ambas estimaciones.

5) La prueba estadística LM_N , aplicada a situaciones donde se reportan gastos nulos, casi siempre lleva a rechazar la normalidad de los disturbios, lo que indica que esta prueba puede ser utilizada para identificar la necesidad del uso del *Modelo Tobit*.

6) En nuestro estudio, dividir la muestra resultó en submuestras de tamaño pequeño dificultando la determinación de la significancia de los valores de las pruebas estadísticas $LM_{NH(Tobit)}$ y $LM_{NH(Trun)}$. Sin embargo, en algunos casos los valores de éstas fueron extremadamente grandes, sugiriendo la presencia de \bar{N} y/o \bar{H} en los modelos. Esto da señales de la necesidad de *proceder con cuidado* en aplicaciones de modelos VDL (la presencia de \bar{N} y/o \bar{H} hace inconsistentes a los estimadores de MV en su formulación clásica).

7) Al utilizar modelos VDL, la falta de ajuste de los estimadores de MV, tal que representen las respuestas apropiadas, puede llevar a resultados erróneos.

8) Se encontraron diferencias sustanciales en los valores numéricos de las respuestas de demanda estimadas por MCO y por VDL-MV. En particular, las estimaciones de la propensión marginal al consumo basadas en MCO son significativamente más grandes, y los gastos de subsistencia son significativamente más pequeños que al utilizar la estimación VDL-MV. Sin embargo, a pesar de estas importantes diferencias cuantitativas en ambas técnicas de estimación, las conclusiones cualitativas en los patrones sistemáticos de demanda y respuestas del ahorro basadas en la estimación de MCO o en VDL-MV, son básicamente las mismas.

6.2 Conclusiones empíricas

Del análisis de los datos de la Encuesta de Ingreso-Gasto de 1975 y de las estimaciones econométricas realizadas se obtuvieron muchos resultados de los cuales sólo algunos han sido mencionados en el texto. A continuación se presenta un resumen de las conclusiones empíricas más importantes (dada la restricción de espacio) relacionadas con la demanda de los hogares en la ciudad de México.

1) Dentro de una categoría ocupacional, existe una tendencia de la *participación marginal presupuestal* por alimentos a decrecer a mayor ingreso; es decir, dado un incremento en el gasto total, los hogares con mayores ingresos asignarían una menor proporción de este incremento para gastarlo en alimentos.

2) Se encuentran efectos significativos de la categoría ocupacional en las *participaciones marginales presupuestales* por alimentos, durables, edu-

cación y servicios médicos. En particular, los técnicos tienen menores participaciones marginales en alimentos y mayores en durables, educación y servicios médicos. Adicionalmente, los obreros y comerciantes tuvieron, en general, significativamente mayor propensión marginal al consumo.

3) Los *gastos de subsistencia* per cápita difieren dependiendo de la categoría ocupacional; los técnicos tienen gastos de subsistencia significativamente mayores. También encontramos una tendencia general en los gastos totales de subsistencia de los obreros a ser menores que los de los comerciantes. Como en Lluich, Powell y Williams (1977: 150), los resultados sugieren que los gastos de subsistencia representan estándares mínimos aceptables para hogares que se identifican con un grupo socioeconómico dado.

4) Hay una tendencia de la *elasticidad de la demanda* por alimentos con respecto al gasto total a ser la más pequeña, como era de esperarse, dado que es un bien básico. Además, esta elasticidad es menor para familias pequeñas y para técnicos.

5) *Las elasticidades del propio precio* aumentan en valor absoluto a mayor ingreso y a mayor tamaño de la familia. Lo anterior parece indicar que familias con ingreso alto y/o las de mayor tamaño tienen más libertad para ajustar su demanda ante incrementos en los precios.

6) *La elasticidad del precio cruzado de alimentos* para servicios médicos y otros, disminuye en valor absoluto al aumentar el ingreso. Es decir, los hogares con ingresos altos sustituyen en menor medida alimentos por servicios médicos y otros ante aumentos en el precio del bien básico alimentos.

Para concluir, notamos que algunas de nuestras estimaciones fueron de signo incorrecto o de una magnitud aparentemente inadecuada (por ejemplo, propensiones marginales al consumo relativamente bajas). Para explicar esto podríamos por supuesto utilizar cifras de otras encuestas validando la calidad de los datos. Pero, adicionalmente, se debe contemplar la técnica de estimación utilizada VDL-MV. Como se mencionó, ésta arroja estimaciones inconsistentes en casos donde el supuesto *NH* no se mantiene. Otro elemento a considerar es el hecho de haber utilizado técnicas univariadas de VDL, ante la escasez de métodos disponibles para afrontar el problema multivariado. Por otro lado, podríamos cuestionar la validez del SLEG como una formulación teórica adecuada de la descripción del comportamiento de demanda y ahorro de los hogares.

En resumen, *para los datos disponibles*, nuestro estudio no da un soporte contundente por el SLEG y sugiere la necesidad de considerar otros sistemas de demanda. Nuestro trabajo también sugiere que, en estudios microeconómicos, resulta en ocasiones conveniente utilizar técnicas de estimación que van más allá de mínimos cuadrados, y que un área importante para futura investigación es el manejo de modelos de Variables Dependientes Limitadas, con disturbios no normales y/o heteroscedásticos y con formulación multivariada. Técnicas apropiadas para esto podrían dar

resultados numéricos suficientemente exactos que permitieran servir de insumos para responder de manera más precisa y eficiente a preguntas de política económica.

Bibliografía

- Beale, E.M.L. (1969), "Euclidean cluster analysis", trabajo presentado en la 37ava sesión del International Statistical Institute.
- Bera, A.K., C.M. Jarque y L.F. Lee (1984), "Testing for the normality assumption in limited dependent variable models", en *International Economic Review* 25 (1984), 563-578.
- Bera, A.K. y C.M. Jarque (1982), "Model Specification Tests: A simultaneous Approach", en *Journal of Econometrics* 20 (1982), 59-82.
- Breusch, T.S. y A.R. Pagan (1979), "A simple test for heteroscedasticity and random coefficient variation", en *Econometrica* 47 (1979), 1287-1294.
- Brown, A. y A. Deaton (1972), "Surveys in applied economics: Models of consumer behaviour", en *The Economic Journal* 82 (1972), 1145-1236.
- CENIET (1977), *Encuesta de ingresos y gastos familiares 1975*, Serie Estudios núm. 1, Centro Nacional de Información y Estadísticas del Trabajo, México.
- Deaton, A.S. (1974), "A reconsideration of the empirical implications of additive preferences", en *Economic Journal* 84 (1974), 338-348.
- _____ (1975), *Models and Projections of demand in post-war Britain*, London, Chapman and Hall.
- Deaton, A. y J. Muellbauer (1980), "An almost ideal demand system", en *American Economic Review*, 70 (1980), 312-326.
- Fair, R.C. (1977), "A note on the computation of the Tobit estimator", en *Econometrica*, 45 (1977), 1723-1727.
- Greene, W.H. (1981), "On the asymptotic bias of the ordinary least squares estimator of the Tobit Model", en *Econometrica*, 49 (1981), 505-513.
- Hartigan, J.A. (1975), *Clustering Algorithms*, New York, Wiley and Sons.
- Howe, H. (1975), "Development of the extended linear expenditure system from simple saving assumptions", en *European Economic Review*, 6 (1975), 305-310.
- Jarque, C.M. (1980), "Estimation of varying parameter models using cluster analysis", en *Working Papers in Economics and Econometrics*, 38 (1980) The Australian National University.
- _____ (1981), "A test for heteroscedasticity in a limited dependent variable model", en *Australian Journal of Statistics*, 23 (1981), 159-163.
- _____ (1987), "Sample Splitting and Applied Econometric Modeling", en *Journal of Business and Economic Statistics*, 5, 2, 43-58.
- Jarque, C.M. y A.K. Bera (s/f), "A test for normality of observations and regression residuals", de próxima publicación en *International Statistical Review*.
- _____ (1980), "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", en *Economics Letters* 6 (1980), 255-259.
- _____ (1982), "Efficient specification tests for limited dependent variable models" en *Economics Letters*, 9 (1982), 153-160.
- Lluch, C. (1973), "The extended linear expenditure system", en *European Economic Review*, 4 (1973), 21-32.
- Lluch, C., A.A. Powell y R.A. Williams (1977), *Patterns in Household Demand and Saving*, New York, Oxford University Press.

- McDonald, J.F. y R.A. Moffitt (1980), "The uses of Tobit analysis", en *Review of Economics and Statistics*, LXH (1980), 318-321.
- Poirier, D.J. y A. Melino (1978), "A note on the interpretation of regression coefficients within a class of truncated distributions", en *Econometrica*, 46 (1978), 1207-1209.
- Pollak, R.A. y T.J. Wales (1980), "Demographic variables in demand analysis", trabajo presentado en Fourth World Congress of the Econometric Society, Aix-en-Provence.
- Powell, A.A. (1973), "Estimation of Lluch's extended linear expenditure system from cross-sectional data", en *Australian Journal of Statistics*, 15 (1973), 111-117.
- _____ (1974), *Empirical Analytics of Demand Systems*, London, Heath and Comany.
- Sparks, D.M. (1973), "Euclidean cluster analysis", en *Journal of the Royal Statistical Society*, 22 (1973), 126-130.
- Tobin, J. (1958), "Estimation of relationships for limited dependent variables", en *Econometrica*, 26 (1958), 24-36.
- Ward, J.H. (1963), "Hierarchical grouping to optimize an objective function", en *Journal of the American Statistical Association*, 58 (1963), 236-244.
- White, H. (1980), "A heteroscedasticity consistent covariance matrix estimator and a direct test for heteroscedasticity", en *Econometrica*, 48 (1980), 817-838.
- Williams, R.A. (1977), "Engel Curves and demand systems: Demographic effects on consumption patterns in Australia", *Preliminary Working Paper* núm. SP-07, Impact Project, Melbourne.