METODOS ESTADISTICOS DE EVALUACION DE MODELOS DE REGRESION DISCRETA

Marisa Yadlin A.
Pontificia Universidad Católica de Chile
Casilla 6177, Santiago 22

RESUMEN

Modelos econométricos de regresión con variable endógena discreta sirven en investigaciones del comportamiento de decisión de agentes económicos.

El avance de los modelos que presentamos se debe a su relevancia en estudios de demanda desagregada por transporte.

La literatura se centra en especificación de modelos consistentes con maximización de utilidad y manejables analitica y computacionalmente, en técnicas de estimación y tests de hipótesis en un modelo dado y en criterios estadísticos para comparar entre modelos competidores.

El objetivo puede ser analizar la conducta de viajeros en un punto del tiempo o a través del tiempo, lo que requiere de modelos aplicables a información de corte transversal o longitudinal, respectivamente. En el primer caso, se ha enfatizado la generalización del Modelo Logit Multinomial, de manera que interrelaciones entre los modos sean representadas. En el segundo, la problemática principal ha sido la adaptación de una metodologia de series cronológicas discretas al tratamiento de observaciones sobre paneles disjuntos de individuos.

Examinamos ciertos procedimientos estadìsticos basados en métodos de Máximo Verosimilitud. La confiabilidad de medidas de bondad de ajuste y tests de significancia utilizados para discriminar entre modelos alternativos es discutida.

Conjuntamente se ofrecen algunas ilustraciones de los desarrollos.

"La unidad de toda la ciencia consiste sólamente en su método, no en su materia... No son los hechos ni los números los que forman la ciencia, pero los métodos con que estos son tratados".

Traducido de Pearson (1938).

1. INTRODUCCION

Las inferencias estadísticas acerca de fenómenos en transporte no difieren fundamentalmente de aquellas en otras áreas científicas. No son las cifras por sí mismas, sino lo que hacemos con ellas lo que importa: Debemos asegurarnos que la naturaleza de los modelos probabilísticos formulados y de su análisis estadístico se adecúen a la naturaleza de la situación y de la información recolectada, de modo que las idealizaciones teóricas no desencadenen en conclusiones prácticas engañosas. Por ejemplo, un análisis estadístico de selección de un modelo para predecir en qué modo viaja un individuo puede conducirnos a eliminar la variable explicatoria costo del modo. Si creemos que esta variable debería estar incluida, tendríamos que cuestionarnos si el modelo final que mejor se ajustó no es demasiado sofisticado para nuestros fines y para el tipo de información disponible, si el costo aparentemente no influye porque su variabilidad en los datos es pequeña, etc.

La aplicación de la estadística a demanda por transporte abarca problemas que varían mucho en su carácter y complejidad. En este artículo, prestamos atención a un núcleo relativamente reducido de ideas y métodos, concentrándonos en conceptos generales, mediante argumentos bastante informales, en lugar de en discusiones rigurosas sobre las propiedades de técnicas particulares. Esto último aparece en las referencias citadas. Ofrecemos ilustraciones de modelos de regresión discreta (MRD), realzando el aspecto cualitativo de ellas.

Los MRD (Amemiya, 1981; Maddala, 1982; McFadden, 1981 y 1984; Ortúzar, 1982; Yadlin, 1983 y 1985), son modelos econométricos de regresión con variable endógena discreta, que tienen por objetivo ayudar en la interpretación de datos sujetos a una variabilidad aleatoria apreciable, en el siguiente marco:

Hay $t = 1, \ldots, T$, estratos de agentes económicos, que eligen uno y sólo uno de los bienes en $C_t = \{1, \ldots, J_t\}$. Para simplificar y sin pérdida de generalidad, suponemos que $C_t = C$, $J_t = J$, $t = 1, \ldots, T$. Por ejemplo, C tiene por elementos la posesión o no posesión de automóvil, los destinos de un viaje o los modos de transporte. Los atributos de la alternativa j que enfrenta un agente en el estrato t se resumen en un vector observado Z_{jt}. Las variables latentes o no observadas Y^*_{jt} miden la utilidad o algún índice de deseabilidad de la alternativa j para un agente en t y

$$Y^*_{it} = V_{it} + E_{it}, \tag{1}$$

$$V_{it} = Z_{it}\,\beta = \sum_{k=1}^{K} Z^{k}_{it}\,\beta^{k}, \tag{2}$$

β es un vector de K parámetros de "gusto" desconocidos,

E_{it} son perturbaciones aleatorias.

Las variables de elección observadas

$$Y_{it} = \begin{cases} 1 \text{ si un individuo en t elige i} \\ 0 \text{ si no} \end{cases} \tag{3}$$

se relacionan con los Y^*_{it} por el principio de maximización de utilidad

$$Y_{it} = 1 \text{ si y sólo si } Y^*_{it} > Y^*_{jt},\ \forall\ j \in C,\ j \neq i. \tag{4}$$

Si F es la Función de la Distribución Acumulativa (FDAC) de los $E_t = (E_{jt};\ j \in C), t = 1,...T$, F depende de un vector α de L parámetros desconocidos, que representan dependencias entre los E_{it}.

Poniendo $\theta = (\beta, \alpha)$, la familia $\mathcal{F}$ de modelos probabilísticos

$$P_{it} = P_{it}(\theta) = P_r\,[\text{un sujeto en t elija i} \in C] = P_r\,[Y_{it}=1] \tag{5}$$

está totalmente determinada, excepto por θ, por F.

Las preguntas que surgen primeramente son:

(a) ¿Son los datos consistentes con la familia $\mathcal{F}$ u otra familia $\mathcal{F}_1$ se ajusta mejor?.

(b) Asumiendo provisionalmente que los datos son generados por uno de los modelos en $\mathcal{F}$, ¿qué podemos concluir acerca de la especificación de V_{in}, del parámetro θ, de los valores de observaciones futuras surgidas del mismo modelo?.

Recurrimos a procedimientos estadísticos que sean cercanamente óptimos para evaluar dos o más familias de modelos o los modelos dentro de una familia dada. Sin embargo, los criterios de optimalidad deben usarse críticamente, en conexión con conocimientos teóricos y empíricos previos. Un análisis óptimo en la comparación de $\mathcal{F}$ con $\mathcal{F}_1$ puede ser ineficiente para juzgar $\mathcal{F}$ versus $\mathcal{F}_m$, aunque $\mathcal{F}_1$ y $\mathcal{F}_m$ sean indistinguibles para algunos efectos prácticos. Por ejemplo, $\mathcal{F}$, $\mathcal{F}_1$ y $\mathcal{F}_m$ son los Modelos Probit Multinomial Correlacionado (MPMC), Probit Multinomial Independiente (MPMI; Hausman et al, 1978) y Logit

Multinomial (MLM; McFadden, 1981), respectivamente, y la técnica de comparación es el Test de Razón de Verosimilitud (Yadlin, 1986 a), que describimos más abajo. O una técnica buena puede arrojar estimaciones de θ fuera de los limites compatibles con el sistema real. Por ejemplo, al calibrar un MRD para partición modal, el coeficiente β^{*} asociado a la variable costo de viaje Z^{*}_{it} puede tener signo incorrecto (Fernández et al, 1983). Así, una familia de modelos debe considerarse como una base esencialmente tentativa para el estudio de un sistema y un procedimiento que es aceptable bajo supuestos amplios es preferible a uno que es óptimo sólo bajo supuestos muy restrictivos.

Examinamos la pregunta (a) en MRD y algunas de sus soluciones en términos de métodos Máximo Verosimiles. Esquematizamos los elementos de estos a continuación.

2. LA FUNCION DE VEROSIMILITUD

Disponemos de observaciones $X = (X_1, \ldots\ldots, X_N)$ con densidad $p(x, \theta)$, con dominio $\mathcal{X}$, para cada $\theta \in \Theta$ fijo. $\mathcal{X} \subset R^{N}$ es el espacio muestral de valores posibles de x y $\Theta \subset R^{K+L}$ es el espacio paramétrico de valores de θ (Lehmann 1959 y 1983; Silvey, 1980). Por ejemplo, extraemos una muestra aleatoria estratificada según los $Z_t = (Z_{jt};\ j \in C)$, $t = 1,\ldots,T$, o sea, una muestra basada en las variables independientes. El tamaño muestral del estrato t

es R_t, fijo a-priori y $N = \sum_{t=1}^{T} R_t$.

Observamos $Y = (Y_{jtm};\ j \in C,\ m = 1,\ldots, R_t,\ t = 1,\ldots\ldots,T)$,
$Y_{jtm} = 1$ si el m-ésimo sujeto en el estrato t decide j, 0 sino.
Aqui, $X = Y$, $\mathcal{X} = (y = (y_{jtm};\ j \in C,\ m = 1,\ldots, R_t,\ t=1,\ldots,T);\ y_{jtm} = 1 \text{ o } 0,\ \sum_{j\in C} y_{jtm} = 1)$.

Los Y_{jtm} son variables de Bernoulli independientes con densidad

$$p(y, \theta) = \prod_{t=1}^{T} \prod_{m=1}^{R_t} \prod_{j\in C} [P_{jt}(\theta)]^{y_{jtm}} = \prod_{t=1}^{T} \prod_{j\in C} [P_{jt}(\theta)]^{s_{jt}}, \quad (6)$$

$S_{jt} = \sum_{m=1}^{R_t} Y_{jtm}$ es el número de decisiones i en el estrato t, $\sum_{j\in C} S_{jt} = R_t$. Para $R_t = 1$, $S_{jt} = Y_{jt}$.

Notemos que toda la información que aporta el vector observado y acerca de θ, está contenida en los S_{jt}. Como no queda ningún conocimiento adicional sobre los $P_{jt}(\theta)$ en Y, después de obtener el estadistico suficiente $S = (S_{jt};\ j\in C,\ t=1,\ldots,T)$, podemos limitarnos a observar S.

Aquí $X = S$, $\mathcal{X} = (s = (s_{jt};\ j \in C,\ t = 1, \ldots, T) : s_{jt} = 0, 1, \ldots, R_t, \sum_{j \in C} s_{jt} = R_t)$.

Si ponemos $S_t = (S_{jt};\ j \in C)$, $S = (S_t;\ t = 1, \ldots, T)$, S_t es un vector Multinomial $(R_t, (P_{jt}(\theta);\ j \in C))$, los S_t son independientes con densidad

$$p(s, \theta) = \prod_{t=1}^{T} \prod_{j \in C} R_t\,!\,[P_{jt}(\theta)]^{s_{jt}} / s_{jt}! \qquad (7)$$

La función de verosimilitud (fv) $L(\theta, x)$ es la densidad de X, considerada no como función de x sino de θ con dominio Θ, para $x \in \mathcal{X}$ fijo. Si $X = x_o$, $L(\theta, x_o)$ o, equivalentemente, $\log L(\theta, x_o)$ nos sugiere cuán probable sería la observación x_o si θ fuera el verdadero valor del parámetro para cada $\theta \in \Theta$.

La Razón de Verosimilitud (RV) $R(\theta_1, \theta_2, x_o) = L(\theta_1, x_o) / L(\theta_2, X_o)$ o $R = L_1 / L_2$, nos indica cuánto más plausible es que θ_1 versus θ_2 constituya el valor del parámetro subyacente a x_o. R nos permite decidir entre dos familias de hipótesis o modelos, la de hipótesis nula H y la de hipótesis admisibles K, generalmente formuladas como $\theta \in \Theta_o \subset \Theta$ versus $\theta \in \Theta_1 \subset \Theta$. Por ejemplo, el test del MLM versus el Modelo de Valor Extremo Generalizado (MVEG) se traduce en H: $\theta \in R^K \times (0)$ o $\beta \in R^K$, $\alpha = 0$ versus K: $\theta \in R^K \times [0, 1)^L$ o $\beta \in R^K$, $\alpha \in [0, 1)^L$ (Hausman y McFadden, 1984; Ortúzar, 1983; Sobel, 1980; Yadlin 1987). Ahora, no tiene sentido cuestionarse cuán plausible es que θ_1 o θ_2 hayan generado observaciones distintas x_1 y x_2, contrastando $L(\theta_1, x_1)$ con $L(\theta_2, x_2)$, aún cuando x_1 y x_2 correspondan a muestras independientes. A veces, H y K constituyen familias separadas de modelos. Por ejemplo, en demanda por transporte, H y K postulan que las elecciones provienen de un MLM y de un MPMC, respectivamente. En este caso, la inspección de R para un x_o fijo, es válida siempre que H y K sean hipótesis simples, o sea, especifiquen totalmente los modelos, de manera que no sea necesario usar x_o en la estimación de parámetros desconocidos. Si H o K son compuestas, R no sirve de indicador de cuánto mejor explica H o K la muestra dada x_o. Nuevamente, calcular R con dos muestras independientes no suprime esta dificultad. Así, Gunn y Bates (1982) explican que la validación de un modelo ajustado con una muestra, $\hat{P}_{it}$, consiste en asegurarse que las selecciones observadas en una segunda muestra sean consistentes con las predicciones obtenidas mediante los $\hat{P}_{it}$. Los autores afirman que la comparación de dos modelos alternativos se resuelve con dos conjuntos de datos independientes para estimación y predicción, ya que la RV puede usarse pese a diferencias en la estructura de los modelos o en el número de parámetros. Esta aserción es falsa, a menos que los modelos calibrados se interpreten como dos modelos competitores exactos en su estructura y en el valor de los parámetros estimados. O sea, los modelos estimados corresponden a H versus K simples. Aún así, la RV constituye un índice de cuánto mejor reproduce la segunda muestra uno de los dos modelos y no un test para detectar cuál de ellos es el verdadero. Por ejemplo, Ortúzar et al (1987) recurren a este índice para concluir que el Modelo Logit Jerárquico es 195 veces más probable que el MLM en una muestra de validación. Aunque este enfoque es correcto, es difícil darle un significado práctico, debido principalmente a que la RV no está acotada: ¿Es 195 veces más probable mucho mejor que 150 veces más probable?.

El Test de la Razón de Verosimilitud (TRV) apunta al problema anterior. Consiste en construir la RV de dos modelos sin parámetros desconocidos, en una muestra dada, y transformarla a algún estadístico con distribución conocida. Por lo tanto, es dable llevar a cabo un test para decidir entre los modelos con nivel de significancia y potencia, o sea, con probabilidades de rechazar incorrecta o correctamente H versus K computables.

Cuando hay parámetros desconocidos debemos recurrir al Estimador Máximo Verosímil (EMV) $\hat{\theta}$ de θ (Kendall et al, 1973). $\hat{\theta}$ es aquel $\theta \in Ⓗ$ que maximiza L (θ, x_o), x_o fijo. De todos los puntos en Ⓗ, $\hat{\theta}$ es aquel con mayor credibilidad de haber producido x_o. Bajo condiciones de regularidad $\hat{\theta}$ es asintóticamente Normal. En aplicaciones hay que constatar si las condiciones se cumplen y si N es suficientemente grande.

El Test de la Razón de Verosimilitud Generalizado (TRVG) es análogo al TRV. En ambos se conoce la distribución de la que proviene la muestra, pero el primero se aplica cuando H o K no especifican θ y se estima a través de EMV. Las familias de modelos en H deben estar anidadas en las de K. Es decir, obtenemos $\theta \in Ⓗ_1$ imponiendo r restricciones independientes sobre $\theta \in Ⓗ_*$. La RVG es L $(\hat{\theta}, x_o)$ /L $(\tilde{\theta}, x_o)$, donde $\hat{\theta}$ y $\tilde{\theta}$ son los EMV de θ no restringido (bajo K) y restringido (bajo H), respectivamente. Para nuestra observación x_o, determinamos sus mejores chances bajo K y bajo H. De acuerdo al principio de la RV, mientras mayor la RVG, mayor la evidencia en contra de H contenida en x_o. Por ejemplo, en el TRVG del MLM versus el MVEG, $\hat{\theta}$ es el EMV de $(\beta, \alpha) \in R^{K+L}$, $\tilde{\theta}$ es el EMV de $\beta \in R^K$ y rechazamos la hipótesis H que el modelo es un MLM, o sea, $\alpha = 0$ (no hay interrelación entre las alternativas), a favor de la hipótesis K, que el modelo es un MVEG o sea, $\alpha \neq 0$, si la RVG es grande, dada la muestra. Si la RVG no se traduce en un estadístico con distribución conocida y el tamaño muestral N es grande, el teorema de Wilks (1938) establece que, sujeto a regularidad, 2 log LRG tiene aproximadamente una distribución de Chi-Cuadrado con r grados de libertad, bajo H (para todo $\theta \in Ⓗ_1$). La demostración asume la teoría asintótica (N grande) para los EMV y que K se encuentra cercana a H. Más precisamente, si H es de la forma $\theta = \theta_o$ y t está acotado, K dice que $\theta = \theta_o + t/\sqrt{N}$, que es una secuencia de alternativas, la cual se aproxima a θ_o a una tasa de $1/\sqrt{N}$. En el ejemplo anterior, α bajo K está en una vecindad de 0. Si N es grande, $\hat{\theta}$ y $\tilde{\theta}$ están cerca de θ_o, por lo tanto, cercanos entre ellos y se puede linearizar el problema mediante una expansión en serie de Taylor alrededor de $\hat{\theta}$. Notemos que no se trata de un estimador y un test para N fijo, sino de una secuencia de procedimientos Máximo Verosímiles, los que se van haciendo óptimos a medida que N crece. Además, el enfoque hacia K relativa a H es local, lo que descarta totalmente el tratamiento de hipótesis sobre familias separadas. Los tests de Wald, Lagrangiano y de Chi-Cuadrado son asintóticamente equivalentes al TRVG, aunque pueden no concordar con este último para N pequeño (Wald, 1943; Silvey, 1959). Los primeros son más fáciles de computar, ya que involucran sólo el cálculo de $\hat{\theta}$ o $\tilde{\theta}$, lo que no significa que H no esté contenida en K. Por ende, estos tests no constituyen tests de significancia para evaluar una familia de modelos versus otra disjunta, siendo su potencia despreciable.

Cox (1961) diseñó un método para realizar tests de familias separadas de hipótesis de modelos, combinando éstos para deducir un modelo híbrido. Como los modelos originales juegan roles simétricos, la potencia es muy baja. Además, el test conduce a cuatro resultados, dos de los cuales no son conclusivos. Atkinson (1970) y Quandt (1974) han seguido esta línea. Sus métodos facilitan el aspecto computacional, pero exhiben las otras limitaciones. Aunque la comparación de los valores de medidas de bondad de ajuste sugiere cuál de dos o más modelos, anidados o no, es más apropiado, no provee tests discriminatorios tampoco. Como el nivel de significancia y la potencia no son determinables, carecemos de un índice de la confiabilidad del rango que hemos asignado a los modelos. Además, aceptaremos el modelo con el mayor número de parámetros, a menos que corrijamos por los grados de libertad. Un ejemplo es el Porcentaje Correctamente Predicho (PCP; Tardiff, 1976), paralelo al coeficiente de Determinación R^2 en modelos lineales.

Evidencia empírica que confirma tentativamente los comentarios de esta Sección y revela la conveniencia de restringir las comparaciones a familias anidadas, se encuentra en Yadlin (1983 y 1986b) y Yadlin et al (1986). Dedicamos las Secciones siguientes a algunas ilustraciones de la literatura en transporte.

3. MODELOS PARA INFORMACION DE CORTE TRANSVERSAL

Con este tipo de datos, es común asumir que los E_t son independientes. Por ejemplo, E_t independientes Normal J-Variada con esperanza 0 y matriz de Varianzas-Covarianzas V (N_J (0, V)), t = 1, ..., T, implica el MPM y E_{it} independientes idénticamente distribuidos (iid) Valor Extremo (VE), i ∈ C, t = 1,..., T, implica el MLM, que presenta tractabilidad analítica y computacional

$$P_{it}(\beta) = \exp(V_{it}) \Big/ \sum_{j \in C} \exp(V_{jt}), \qquad (8)$$

V_{it} está dado en (2) y $\alpha = 0$, por independencia de los E_{it}.

Notemos que los Z^k_{it}, $k = 1,\dots, K$, deben estar asociados a las alternativas $i \in C$, para que no se cancelen en (8). Esto impide incluir en todos los V_{it}, $i \in C$, $t = 1,\dots, T$, la misma constante β^k ($Z^k_{it} = 1$, $i \in C$, $t = 1,\dots, T$) o alguna variable Z^k_{it} que mida características propias del individuo t sólamente. De esta forma, el término constante β^k está referido siempre a una alternativa base $i \in C$ ($Z^k_{jt} = 1$ si $j = i$; 0 sino) o se contemplan varias constantes β^k, no necesariamente iguales, específicas a algunas alternativas $i \in C$, ($Z^k_{jt} = 1$, si j es alguna de estas alternativas; 0 sino). Asimismo, un atributo individual I_t aparece sólo en los V_{it} de algunos $i \in C$, ($Z^k_{jt} = I_t$ si j es alguna de estas alternativas; 0 sino) o combinado con algún atributo propio de las alternativas. Por ejemplo, si C = (1 = Metro, 2 = Bus, 3 = Liebre), el modo 1 se distingue claramente de los otros dos. La variable comfort del modo se puede incorporar como β^1 en V_{1t}, $t = 1,\dots,T$, y la variable ingreso del individuo t como Z^m_{it} o como Z^m_{it} = Ingreso del individuo t / Costo del modo i, $i \in C$. Recalcamos estas limitaciones, debido a que su motivo es puramente aritmético y no conceptual, por lo que su empleo para efectuar tests de especificación del MLM (Horowitz, 1981) es equivocado.

Un test de especificación del MLM si puede originarse en el axioma de Independencia de Alternativas Irrelevantes (IIA) que este modelo cumple: Las chances relativas de elegir i sobre j son independientes de otras opciones disponibles en C. Como consecuencia, el MLM no permite introducir interrelaciones entre las alternativas. Hausman y McFadden (1984) implementan un procedimiento de Hausman (1978) para desarrollar un test de este tipo. Los autores toman un subconjunto propio $D \subset C$, para ver si el comportamiento de selección condicionado a D, obedece la propiedad de IIA y derivan un estadístico de test T fácil de calcular: $T = (\hat{\beta}_D - \hat{\beta}_C)^T (\hat{V}_D - \hat{V}_C)^{-1} (\hat{\beta}_C - \hat{\beta}_D)$, donde $\hat{\beta}_C$ y $\hat{\beta}_D$ son los EMV de los parámetros asociados a las alternativas en C y en D, computados con toda la muestra y con la submuestra correspondiente a D, respectivamente, y $\hat{V}_D$ y $\hat{V}_C$ son los EMV de sus matrices de Varianzas - Covarianzas. Si el tamaño muestral es grande, $\hat{\beta}_D - \hat{\beta}_C$ se acerca a 0 y la distribución de T se aproxima a una Chi-Cuadrado, bajo H. Por lo tanto, se rechaza H para valores grandes de T. Las desventajas de este test son que es muy sensible al subconjunto D que se tome, puede rechazar H por especificaciones falsas distintas de la IIA y su potencia es baja, a menos que la violación de la IIA sea sustancial.

Procedimientos más clásicos se han aplicado para docimar el MLM versus modelos alternativos, consistentes con maximización de utilidad, pero que se desvian de la IIA, permitiendo representar relaciones entre los bienes y procesos de decisión en mallas, mediante: 1. Introducir dependencias entre los E_{it}, $i \in C$, relajando $\alpha = 0$ en F. Por ejemplo, en el MPMC, V es no diagonal y las coordenadas de α son las covarianzas en V; 2. En $V_{\ast\ast}$, permitir que β dependa de los $i \in C$ y que aparezcan términos interactivos entre los $Z_{i\ast}$, $i \in C$. Por ejemplo, el Modelo Logit Universal (MLU; McFadden et al, 1976). 1. y 2. no son excluyentes, aunque conducen a dos clases de familias de modelos excluyentes: Familias de modelos que constituyen una generalización paramétrica del MLM y lo contienen como caso particular ($\alpha = 0$) y familias de modelos separados del MLM. Por ejemplo, el MVEG, el MLU y el Modelo de Valor Extremo Global (MVEGl; Yadlin 1985 y 1986a) pertenecen a la primera clase, mientras que el MPM, a la segunda.

Los métodos Máximo Verosímiles son los más frecuentes en la literatura. Típicamente, se formula la fv de acuerdo a (6), que corresponde a una muestra estratificada según las variables independientes. Si la muestra es aleatoria no estratificada con $N = \sum_{t=1}^{T} R_{\ast}$, N fijo a - priori, $R_{\ast}$ libres, la fv es proporcional a (6), siempre que la FDAC de los $Z_{i\ast}$ no dependa de θ, obteniéndose el mismo EMV $\hat{\theta}$. Hay que tener presente que esto no ocurre cuando la muestra es estratificada según las respuestas observadas, o sea, cuando un individuo t participa en la muestra porque ya ha elegido cierto $i \in C$ y luego es entrevistado sobre sus valores de $Z_{i\ast}$, $i \in C$. Las encuestas realizadas por la Dirección General del Metro de Santiago proveen un ejemplo. Cosslett (1981), Manski y Lerman (1977) y Manski y McFadden (1981) indican que la fv debe ponderarse por las probabilidades a-priori de selección, las que pueden estimarse por las particiones de mercado de los $i \in C$.

Suponiendo que la fv es compatible con el muestreo y que N es grande, efectuar un TRVG es apropiado para evaluar el MLM frente a la primera clase de familias, únicamente. Citamos algunos análisis adecuados a continuación. Hausman y McFadden (1984) discuten el TRVG y los tests de Wald y Lagrangiano para contrastar el MLM con el MVEG. McFadden et al (1976) diseñan TRVG de especificación del MLM versus generalizaciones de éste, como el MLU, el modelo saturado (cuyos parámetros son todos los $P_{i\ast}$) y un modelo que considera β aleatorio (Bayesiano). Yadlin (1986a y b) y Yadlin et al (1986) muestran que el nivel de significancia empírico coincide con el nominal y que la potencia es adecuada, cuando H y K estipulan que el modelo verdadero es el MLM y el MVEG o el MVEGl, respectivamente, pero que lo contrario sucede si se docima el MVEG versus el MVEGl que son disjuntos. Además, se aprecia que las medidas de bondad de ajuste tienen bajo poder discriminatorio, aún cuando se trate de corregir el sesgo producido por la asignación de máxima probabilidad, prediciendo las selecciones mediante simulación Multinomial. Por otra parte, los tests del MLM versus el MPM ilustran la implementación incorrecta del TRVG. Hausman y Wise (1978), Horowitz (1981 a y

b) usan el TRVG y los tests de Wald y Lagrangiano para llevar a cabo un test de H:MLM versus K:MPMC, reemplazando H por H:MPMI. Los autores argumentan que el MPMI, con correlaciones nulas y varianzas iguales a la varianza de una variable VE, es virtualmente equivalente al MLM. Este razonamiento es confuso y de los tests sólo se puede deducir la calidad del MPMI relativa al MPMC.

4. MODELOS PARA INFORMACION DE CORTE LONGITUDINAL

Como los estratos constituyen un panel, o sea, observaciones Y_{it} para la misma muestra aleatoria, en períodos de tiempo $t = 1,...,T$, $T \geq 2$, y el foco es la relación intertemporal entre las elecciones de un sujeto, se asume que los E_t son dependientes.

Con el objetivo de explicar decisiones presentes a partir de decisiones pasadas, los modelos deben ser suficientemente flexibles en cuanto a la inclusión de variables explicatorias que varían a través del tiempo y de estructuras de correlación serial entre los $Y*_{it}$. Además, existen otros dos problemas que deben ser investigados. El primero se refiere a una regularidad empíricamente comprobada: Los individuos que han experimentado un suceso en el pasado tienen más chances de repetirlo en el futuro que aquellos que no lo han experimentado nunca. Esta situación de contagio o hábito puede ser causada por una influencia genuina del comportamiento pasado sobre el futuro o por una heterogeneidad de la propensidad de los individuos a comportarse de cierta manera. Se requiere distinguir entre estas dos causas, denominadas dependencias de estado real y espuria, respectivamente. El segundo problema se refiere a cómo inicializar el proceso: Hay que determinar si las condiciones iniciales o historia del proceso previas al muestreo son exógenas o no.

Heckman (1978 y 1981) desarrolla un modelo muy amplio, extendiendo el MPM, y que puede calibrarse por EMV.

Modelos más simples para medir cambios en Tablas de Contingencia, que también son analizables por Máximo Verosimilitud, pueden ser adaptados con el fin de investigar conductas de decisión en dos puntos del tiempo (Amemiya, 1978; Bishop el al, 1975; Fienberg et al, 1976).

Koppelman et al (1982) han propuesto un índice de transferencia, basado en las fv correspondientes a dos períodos, para evaluar la estabilidad temporal de un modelo. Ortúzar et al (1985a) utilizan este índice con datos de 1981 y 1983 sobre elección modal en Santiago.

Frecuentemente, se dispone de paneles disjuntos de sujetos, o sea, de muestras diferentes observadas en distintos puntos del tiempo. Cuando los paneles son muy similares en aquellos aspectos que inciden en la conducta como función de conductas pasadas, la misma metodologia para datos de panel puede ser adaptada. Ortúzar et al (1985b) formulan un MPM para modelar elecciones modales en dos muestras, recolectadas en dos periodos, cuyos miembros poseen atributos semejantes y el mismo conjunto de alternativas y de restricciones. Si las muestras no son comparables, es apropiado suponerlas independientes y sólo es factible estudiar si han ocurrido modificaciones temporales en el parámetro β o en las probabilidades $P_{i\bullet}$. Un test muy simple se puede construir como consecuencia de la distribución asintótica de los EMV de β y de $P_{i\bullet}$. Una ilustración típica es la de modelos para contextos previo y posterior a la introducción de un nuevo modo. Train (1978) investiga si los parámetros y las particiones modales son iguales antes y después de la apertura del tren subterráneo en San Francisco, U.S.A. Escobar et al (1984) implementan un test en términos de los EMV pre y post Metro, cuyas diferencias son aproximadamente Normales, para analizar el impacto del Metro en Santiago.

REFERENCIAS

AMEMIYA, T. (1978). A note on the estimation of a time dependent Markov Chain model, Department of Economics, Stanford University.

AMEMIYA, T. (1981). Qualitative response models: A survey, Journal of Economic Literature 19, 1843-1536.

ATKINSON, A.C. (1970). A method for discriminating between models, Journal of the Royal Statistical Society B 32, 323-353.

BISHOP, T., FIENBERG, S. y HOLLAND, P. (1975). Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge.

COSSLET, S. (1981). Efficient estimators of discrete choice models, en C. Manski y D. McFadden (ed.) Structural Analysis of Discrete Choice Models, MIT Press, Cambridge.

COX, D.R. (1961). Test of separate families of hipotheses, Proceedings of the Fourth Berkeley Symposium 1, 105-123.

ESCOBAR, L. y YADLIN, M. (1984). Uso de simulación en el análisis de robustez de Modelos de Demanda Discreta, Actas del 2° Congreso Latino Iberoamericano de Investigación Operativa e Ingeniería de Sistemas, Volumen Avances en Modelamiento, 13-28, Buenos Aires.

FERNANDEZ, J.E., COEYMANS, J.E. & ORTUZAR, J. de D. (1983). Evaluating extensions to the Santiago underground system, 11th PTCR Summer Anual Meeting, University of Sussex.

FIENBERG, S.E. y LARNITZ, K.(1970). Log linear representation for paired and multiple comparison models, Biometrika 76, 245-254.

GUNN, H.F. y BATES, J.J. (1982). Statistical aspects of travel demand models, Transportation Research A 16, 371-383.

HAUSMAN, J. (1978). Specification tests in econometrics, Econometrica 46, 1251-1272.

HAUSMAN, J. y WISE, D. (1978). A conditional Probit Model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences, Econometrica 46, 403-426.

HAUSMAN, J. y McFADDEN, D. (1984). Specification tests for the Multinominal Logit Model, Econometrica 52, 1919-1240.

HECKMAN, J. (1978). Simple statistical models for discrete panel data developed and applied to test the hipothesis of the true state dependence against the hipothesis of spurious state dependence, Annals de Insee 30-31, 227-269.

HECKMAN, J. (1981). Statistical models for the analysis of discrete panel data in C.F. Manski y D. McFadden (ed.) Structural Analysis of Discrete Data, MIT Press, Cambridge.

HOROWITZ, J. (1981a). Identification and diagnosis errors in the Multinomial Logit Model, Transportation Research B 15, 345-360.

HOROWITZ, J. (1981b). Testing the Multinominal Logit Model againts the Multinomial Probit Model without estimating the Probit parameters, Transportation Science 15, 153-163.

KENDALL, M.G. y STUART, A. (1973). The Advanced Theory of Statistics, Vol 2, 3ra. ed., Hafner, Nueva York.

KOPPELMAN, F.S. y WILMOT, C.G. (1982). Transferability analysis of disaggregate choice models, Transportation Research Record 895, 18-24.

LEHMANN, E. (1959). Testing Statistical Hypotheses, Wiley, Nueva York.

LEHMANN, E. (1983). Theory of Pont Estimation, Wiley, Nueva York.

McFADDEN, D., TYE, W. y TRAIN, K. (1976). An application of diagnostic tests for the Independence from Irrelevant Alternatives property of the Multinomial Logit Model, Transportation Research Record 673, 39-46.

McFADDEN, D. (1981). Econometric models of probabilistic choice, in C. Manski y D. McFadden (ed.) Structural Analysis of Discrete Data, MIT Press, Cambridge.

McFADDEN, D. (1984). Econometric analysis of qualitative choice, en Z. Griliches y M.D. Intriligator (ed.) Handbook of Econometrics, Vol 2, North Holland, Amsterdam.

MADDALA, G.S. (1982). Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press, Nueva York.

MANSKI, C. y LERMAN, S. (1977). The estimation of choice probabilities from choice-based samples, Econometrica 45, 1977-1988.

MANSKI, C. y McFADDEN, D. (1981). Alternative estimates and sample designs for discrete choice, en C. Manski y D. McFadden (ed.) Structural analysis of Discrete Data, MIT Press, Cambridge.

ORTUZAR, J. de D. (1982). Fundamentals of discrete multimodal choice modelling, Transport Review 2, 47-78.

ORTUZAR, J. de D. (1983). Nested logit models for mixed mode travel demand in urban corridors, Transportation Research 17 A, 283-299.

ORTUZAR, J. de D. y ACHONDO, F.J. (1985a). A pseudo-panel discrete mode choice model: Preliminary findings, Departamento de Ingeniería de Transporte. Pontificia Universidad Católica de Chile.

ORTUZAR, J. de D. y ACHONDO, F.J. (1985b). La estabilidad temporal de modelos desagregados de partición modal. Actas del 2° Congreso de Ingeniería de Transporte, Sociedad Chilena de Ingeniería de Transporte, Santiago, 200-213.

ORTUZAR, J. de D. y SERRA, M. (1987). Un Modelo Logit Jerárquico para el corredor Las Condes-Centro de Santiago, Revista Ingeniería de Sistemas V, 11-24.

PEARSON, K. (1938). The Grammar of Science, Everyman, Londres.

QUANDT, R.E. (1974). A comparison of methods for testing non-nested hypotheses, Review of Economic and Statistics 56, 92-99.

SILVEY, D. (1959). The Lagrangian multiplier test, Annals of Mathematical Statistics 30, 389-407.

SILVEY,S. (1980). Statistical Inference, Penguin, Middlesex.

SOBEL, K.L. (1980). Travel demand forecasting by using the Nested Multinomial Logit Model, Transportation Research Record 775, 48-55.

TARDIFF, T.J. (1976). A note on goodness of fit statistics for Probit and Logit Models, Transportation 5, 377-388.

TRAIN, K. (1978). A validation test of disaggregate mode choice models, Transportation Research B 12, 107-174.

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large, Transactions of the American Mathematical Society 54, 426-482.

WILKS, S.S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses, Annals of Mathematical Statistics 9, 60-62.

YADLIN, M. (1985). Development of a model for probabilistic discrete decisions, Tesis de Doctorado, Departamento de Estadística, Universidad de California, Berkeley.

YADLIN, M. (1986a). Some Procedures for the statistical and empirical analyses of probabilistic discrete decision models, Informe Técnico PUC IFM - 86 /B, Facultad de Matemáticas, Pontificia Universidad Católica. Aceptado con modificaciones en Transportation Research B.

YADLIN, M. (1986b). Some analytical and empirical comparisons of substitutibility specifications in discrete demand models, Sixth Latinoamerican Meeting of the International Econometric Society, Córdova, Argentina.

YADLIN, M. y ARELLANO, R. (1986). Análisis empírico de errores de especificación del mecanismo probabilístico de modelos de decisión discreta. Aceptado en las Actas del III Congreso Latino-Iberoamericano de Investigación Operativa e Ingeniería de Sistemas, Santiago, Chile.

YADLIN, M. (1987). On some extensions of qualitative response models, International Joint Statistical Meeting, San Francisco, E.E.U.U. Resumen publicado en Institute of Mathematical Statistics Bulletin, Julio 1987.