
CRÍTICA AL USO DEL MÉTODO ACM EN MODELOS DE GENERACIÓN DE VIAJES

Cristian A. Guevara
Universidad de Los Andes
San Carlos de Apoquindo 2200, Las Condes, Santiago, Chile.
Teléfono: (56 2) 412 9477, Fax: (56 2) 214 9551
E-mail: aguevara@uandes.cl

Alan Thomas
SECTRA, Gobierno de Chile
Teatinos 950 Piso 16, Santiago, Chile.
Teléfono: (56 2) 671 0935, Fax: (56 2) 696 6477
E-mail: athomas@sectra.cl

RESUMEN

En este trabajo se muestra que la versión del método “Análisis de Clasificación Múltiple” (ACM) utilizada habitualmente en ingeniería de transporte para la estimación de modelos de generación de viajes, lleva implícita un supuesto que usualmente es vulnerado y que cuya trasgresión implica generalmente un sesgo en la composición socioeconómica de los viajes modelados además de la sobreestimación del número estimado de viajes futuros. Para ilustrar los efectos señalados, se analizan y comparan diversos métodos ACM descritos en la literatura mediante simulaciones de Monte Carlo y datos reales de la Encuesta Origen Destino 2001 de Santiago de Chile. Se termina concluyendo que la versión mencionada del método ACM debe ser descartada y que, en cambio, la estimación de las tasas de viaje como el promedio simple por categoría resulta ser más robusto frente a diferentes estructuras del modelo subyacente. Finalmente, se argumenta sobre la necesidad de utilizar enfoques más sofisticados que el ACM para dar cuenta de los fenómenos asociados a la generación de viajes.

Palabras Clave: Análisis de Clasificación Múltiple, Tasas ACM, Modelos de Generación.

ABSTRACT

We analyse various Multiple Classification Analysis (MCA) methods to model trip production (generation). We first show that the MCA version most widely used in transport engineering implies a set of rarely feasible assumptions, the transgression of which may drive a significant overestimation of the future number of trips and a systematic bias in its socioeconomic composition. To illustrate this effect, we use Monte Carlo simulation and real data from Santiago de Chile to compare the various MCA approaches, concluding that the aforementioned form should be discarded. Our analysis also shows that the MCA method which is more robust to the structure of the underlying model is the simple calculation of trip rates as averages for each category. Finally, we hint at the need to use more sophisticated formulations than MCA to model trip production.

Key Words: Multiple Classification Analysis, Trip Generation, Cross Classification.

1. INTRODUCCIÓN

En 1983 Stopher y McDonald hicieron una aplicación pionera en la modelación de la generación de viajes en sistemas de transporte, de un método de uso habitual en las Ciencias Sociales. La aplicación de dicho método, conocido como “Análisis de Clasificación Múltiple” (ACM), fue posteriormente replicada en diversos estudios, descrita en libros y explicada en manuales de modelación de sistemas de transporte (TMIP, 2004; Sectra, 1998; Ortúzar y Willumsen, 1994).

Sin embargo, se ha detectado que el método ACM aplicado de la forma descrita en Stopher y McDonald (1983), sobreestima considerablemente los viajes en el sistema para los cortes temporales futuros, lo cual hace necesario revisar sus fundamentos. Con dicho objetivo, en este trabajo se estudian diversos métodos ACM que han sido descritos en la literatura, para luego compararlos utilizando experimentos de Monte Carlo y una base de datos reales. Finalmente, se detallan las principales conclusiones y recomendaciones respecto del tema tratado.

2. MÉTODOS ACM DESCRITOS EN LA LITERATURA

2.1. Primer Método ACM Descrito por Stopher y McDonald (ACM_SM1)

Comenzaremos revisando el primero de dos métodos ACM descritos en Stopher y McDonald (1983) el cual será consecuentemente denominado como ACM_SM1. Este método es ampliamente usado en ingeniería de transporte y es también detallado en Ortúzar y Willumsem (1994), SECTRA (1998) y TMIP (2004).

Esta técnica puede ser aplicada a múltiples variables pero, para simplificar su análisis y comparación con otros métodos, consideraremos un caso especial en el cual se modela la generación de viajes basados en el hogar de ida (BHI) clasificando a los hogares en función de su estrato de ingreso (i) y estrato de motorización (m) (número de vehículos que posee) en diversas categorías (im). En este contexto, el número de viajes por hogar con este método se calcularía como la suma de: a) el promedio global de viajes por hogar (\hat{t}); la diferencia entre el promedio de viajes por hogar para hogares de ingreso i y el promedio global de viajes por hogar ($\hat{t}_i - \hat{t}$); y c) la diferencia entre el promedio de viajes por hogar para hogares de motorización m y el promedio global de viajes por hogar ($\hat{t}_m - \hat{t}$).

$$\hat{t}_{im} = \hat{t} + (\hat{t}_i - \hat{t}) + (\hat{t}_m - \hat{t}) \quad (1)$$

La expresión (1) puede re-escribirse, después de algo de álgebra, como se muestra en la expresión (1') donde v^h corresponde a los viajes observados que genera el hogar h y $I_{im}^h = 1$ si el hogar h pertenece a la categoría im y cero en caso contrario. M , I y H corresponden al número de estratos de motorización, ingreso y el número de hogares, respectivamente. H_{im} corresponde al número de hogares pertenecientes a la categoría im .

$$\hat{t}_{im} = \underbrace{\sum_{h=1}^H \sum_{k=1}^M 1_{ik}^h v^h / H_i}_{\hat{t}_i} + \underbrace{\sum_{h=1}^H \sum_{k=1}^I 1_{km}^h v^h / H_m}_{\hat{t}_m} - \underbrace{\sum_{h=1}^H v^h / H}_{\hat{t}} \quad (1')$$

Stopher y McDonald (1983) señalan que (1) corresponde a una aplicación del método de ANOVA. Como puede verificarse en cualquier libro de econometría (ver por ejemplo, Greene, 2003), ANOVA corresponde al estimador de Mínimos Cuadrados Ordinarios (MCO) de un modelo que es lineal en variables dicotómicas que representan la pertenencia a determinados estratos socioeconómicos. Ciertos especialistas dedicados a la aplicación de la estadística a las Ciencias Sociales derivaron expresiones simplificadas de este modelo que son aplicables a casos especiales. ACM_SM1 es uno de dichos casos especiales. Corresponde al estimador MCO de un modelo lineal en las variables de estratificación y donde el número de observaciones por categoría (ingreso-motorización) es exactamente **el mismo** (ver por ejemplo Glass y Stanley, 1986). Desgraciadamente Stopher y McDonald olvidaron declarar dicho importante supuesto en su artículo. Este pequeño descuido sería el origen de 25 de años de modelos de generación mal especificados y quizás una explicación parcial para sobre-estimación de la demanda de diversos proyectos de transporte en todo el mundo.

El problema es que el número de observaciones por categoría difícilmente podrían ser iguales si se utiliza muestreo aleatorio de hogares. En nuestro ejemplo, debido a que la tasa de motorización y el ingreso están positivamente correlacionados, existirán más observaciones pertenecientes a las categorías extremas (ingreso bajo-motorización baja e ingreso alto-motorización alta) en desmedro del resto de las categorías.

El efecto de la trasgresión de este supuesto es que los viajes predichos para cortes temporales futuros serán sobreestimados y que existirá un sesgo sistemático en su composición socioeconómica. Para mostrar porqué ocurre esto, re-escribamos la expresión (1) imponiendo la condición que hace que ésta sea equivalente al estimador MCO, es decir, que número de observaciones por categoría es el mismo. Definamos además \bar{t}_{im} como el promedio de viajes por hogar para la categoría im . Tengamos en mente además que \bar{t}_{im} es un estimador consistente del verdadero t_{im} , por cualquier ley de grandes números.

$$\hat{t}_{im} = \frac{1}{M} \sum_{k=1}^M \bar{t}_{ik} + \frac{1}{I} \sum_{k=1}^I \bar{t}_{km} - \hat{t} \quad (1'')$$

Re-escribamos ahora la expresión (1) considerando que el número de observaciones por categoría no es el mismo.

$$\hat{t}_{im} = \sum_{k=1}^M \frac{H_{ik}}{H_i} \bar{t}_{ik} + \sum_{k=1}^I \frac{H_{km}}{H_m} \bar{t}_{km} - \hat{t} \quad (1''')$$

Luego, si el número de observaciones por categoría no es el mismo, el error en que se incurre al utilizar la expresión (1) para estimar las tasas de viaje, corresponde a la diferencia entre las expresiones (1''') y (1''). Para ver como se comporta este sesgo, basta notar que para ingresos bajos H_{im} debiera ser decreciente con la motorización y para ingresos altos H_{im} debiera ser creciente con la motorización. Por otra parte \bar{t}_{im} debiera ser siempre creciente con la motorización y el ingreso. Luego, por ejemplo, para el estrato de ingreso más bajo, el primer término en (1'') va a ser subestimado por el término respectivo de (1''') porque los \bar{t}_{im} más pequeños (de los

hogares con motorización baja) van a estar multiplicados por un factor más grande ($H_{im}/H_i > 1/M$) y, por el contrario, los \bar{t}_{im} más grandes (de los hogares con motorización más alta) van a estar multiplicados por un factor más pequeño. Algo equivalente ocurre con otros estratos y factores. En otras palabras, debido a que las categorías extremas están sobre-representadas, \hat{t}_i y \hat{t}_m son menores de lo que “debieran” ser para estratos de ingreso y motorización bajos; y mayores que lo que “debieran” ser para estratos de ingreso y motorización altos, generando así un sesgo importante en las tasas estimadas con el método ACM_SM1 (1).

Para el corte temporal en el cual el modelo es calibrado, este sesgo no afectará el número total de viajes estimados, pero su composición socioeconómica va a ser espuria. Ya que las tasas de los hogares más pudientes son sobre-estimadas y las de los hogares más pobres son sub-estimadas, la composición socioeconómica de los viajes modelados se verá afectada en concordancia. Para los cortes temporales futuros, este sesgo significará además la sobreestimación del impacto en el número de viajes que tiene el corrimiento de hogares desde las categorías socioeconómicas más bajas a las más altas, a medida que la economía crece. Este efecto es discutido en mayor profundidad en la sección 3, mediante simulaciones de Monte Carlo, y en la sección 4, utilizando datos reales de la ciudad de Santiago de Chile.

2.2. Segundo Método ACM Descrito por Stopher y McDonald (ACM_SM2)

El segundo método ACM descrito en Stopher y McDonald (1983) será denominado como ACM_SM2. Este método es presentado por los autores como una corrección del método ACM_SM1 para casos en que existe, lo que los autores denominan como “interacción” entre las variables. Este método corresponde en la práctica a una aproximación numérica que busca adaptar el modelo a casos en que el número de observaciones por categoría no es la misma.

Este método aparece también descrito en Ortúzar y Willumsem (1994), pero no en las metodologías SECTRA (1998) ni en TMIP (2004). Difiere del ACM_SM1 en que, en vez utilizar el promedio simple de los viajes por hogar para cada estrato de ingreso y estrato de motorización, se utiliza el promedio de viajes ponderados por el peso de cada estrato en el número total de hogares. Formalmente, la expresión (2) resume la forma de aplicar el método ACM_SM2.

$$\hat{t}'_{im} = \underbrace{\sum_{h=1}^H \sum_{k=1}^M \left(\frac{H_k}{H} \right) 1_{ik}^h v^h / H_i}_{\hat{t}'_i} + \underbrace{\sum_{h=1}^H \sum_{k=1}^I \left(\frac{H_k}{H} \right) 1_{km}^h v^h / H_m}_{\hat{t}'_m} - \underbrace{\sum_{h=1}^H v^h / H}_{\hat{t}} \quad (2)$$

Puede notarse que los pesos considerados en el método ACM_SM2 van a tender a suavizar los factores mostrados en (1'''), mejorando así los coeficientes estimados, pero difícilmente transformándolos en los estimadores MCO. El efecto neto debiera ser una mejora parcial en los estimadores. Esto será verificado más adelante en los ejemplos numéricos de las secciones 3 y 4.

2.3. Método ACM de Mínimos Cuadrados Lineal (ACM_MCL)

Consideremos ahora un modelo de MCO en el cual el número de viajes por hogar es lineal en el estrato de ingreso y el estrato de motorización más un término de error. Para determinar el estimado MCO de este modelo, es necesario resolver primero el problema de colinealidad perfecta entre las variables. En la expresión (3) se muestra uno de los posibles modelos equivalentes, donde los estratos $i=1$ y $m=1$ fueron elegidos para la normalización.

$$v^h = \beta_0 + \sum_{i \neq 1} \beta_i 1_{i.}^h + \sum_{m \neq 1} \beta_m 1_{.m}^h + \varepsilon^h \quad \forall h \quad (3')$$

Ya que el método ACM_MCL no es más que una aplicación de MCO, es posible tomar de éste todas las herramientas estadísticas y paquetes computacionales disponibles en el mercado. En particular, si se asume alguna distribución del error (ε), por ejemplo Normal $(0, \sigma^2)$, sería posible utilizar toda la gama de *test* estadísticos existentes (F , R^2 , etc.) para, por ejemplo, identificar las variables de estratificación o el tamaño de los estratos o categorías.

Finalmente, cabe recalcar que el método ACM_SM1, ACM_SM2 y ACM_MCL son numéricamente iguales si el número de observaciones por categoría es el mismo. Si este no es el caso, ACM_SM2 puede verse como una aproximación numérica de ACM_MCL, que corresponde al método ACM correctamente aplicado en ese caso. Por lo tanto, si el método correcto es plenamente aplicable, resulta indudablemente innecesario utilizar otro método que sólo se le aproxima numéricamente.

2.4. Método ACM de Promedio Simple por Categoría (ACM_PSC)

El último método analizado corresponde al método ACM_PSC, consistente en calcular directamente el promedio de viajes por hogar para cada categoría ingreso-motorización. Este método, también denominado simplemente como “Análisis por Categorías” (Ortúzar y Willumsem, 1994), es equivalente a estimar un modelo MCO con efectos fijos donde se considera una variable dicotómica para cada categoría (ver por ejemplo Goodman, 1973). Sin embargo, para lograr la identificación del modelo MCO es necesario fijar unas de las categorías en cero, tal como se muestra en (4) para el ejemplo clásico analizado en este trabajo.

$$v^h = \varphi_0 + \sum_{i \neq 1; m \neq 1} \varphi_{im} 1_{im}^h + \varepsilon'^h \quad \forall h \quad (4)$$

Este método posee la ventaja de considerar efectos no-lineales en la tasa de viajes por hogar respecto de su pertenencia a diferentes estratos de ingreso y motorización. Sin embargo, tiene la desventaja de utilizar menos información para estimar cada uno de los coeficientes, obteniendo por lo tanto estimaciones menos eficientes o sesgadas por tamaño muestral.

El modelo ACM_PSC puede escribirse como el ACM_MCL más los parámetros δ de interacción (no-lineales) por categoría que se indican en la expresión (4'). De esta manera, el método ACM_MCL puede considerarse como un modelo restringido del método de ACM_PSC con lo cual si se asume una distribución del error, por ejemplo, Normal $(0, \sigma^2)$, sería posible utilizar un *test F* para verificar la conveniencia de utilizar uno u otro método.

$$v^h = \beta_0 + \sum_{i \neq 1} \beta_{i.} 1_{i.}^h + \sum_{m \neq 0} \beta_{.m} 1_{.m}^h + \sum_{i \neq 1; m \neq 0} \delta_{im} 1_{im}^h + \varepsilon^{''''h} \quad \forall h \quad (4')$$

En caso de ser aceptada la hipótesis nula, esto es, que los modelos ACM_MCL y ACM_PSC son estadísticamente equivalentes, el modelo a utilizar debiera ser el ACM_MCL, puesto que es más eficiente a causa de que requiere estimar un número menor de coeficientes con la misma información. En caso de ser rechazada la hipótesis nula, el modelo a utilizar debiera ser el ACM_PSC debido a que, independiente de la linealidad o no linealidad del modelo subyacente, entrega coeficientes consistentes. Esto último no se cumpliría con el método ACM_MCL, debido a que, si el modelo subyacente es no lineal, el término omitido estaría correlacionado con las variables consideradas en el modelo, lo que ocasionaría un problema de endogeneidad que haría inconsistentes las estimaciones (Guevara y Ben-Akiva, 2006). Esto significa que el método ACM_PSC resulta ser más robusto a diferentes especificaciones del modelo subyacente.

3. COMPARACIÓN DE MÉTODOS ACM UTILIZANDO SIMULACIONES DE MONTE CARLO

En esta sección se analiza la base teórica y las implicancias prácticas de los métodos ACM descritos en la sección 2, mediante dos experimentos de Monte Carlo. El modelo subyacente o “real” utilizado en el primer experimento considera que las tasas de generación son lineales en el ingreso y la motorización. En el segundo caso se considera una relación no lineal. En ambos experimentos se utilizó una muestra de 1000 hogares distribuidos por categoría de ingreso y motorización, donde se consideró además que estas variables estaban positivamente relacionadas de forma tal que el número de hogares en las categorías extremas fuese mayor, tal como puede apreciarse en la Tabla 1.

3.1. Modelo Subyacente Lineal

En la Tabla 2 se presentan las “tasas reales” de generación de viajes por hogar consideradas en este experimento, las cuales son lineales en el estrato de ingreso (1, 2 y 3) y de motorización (0, 1 y 2) del hogar. En efecto, puede verse que la tasa aumenta en 0.2 viajes por cada estrato de motorización adicional (independiente del estrato de ingreso) y 0.6 viajes por cada estrato de ingreso extra (independiente del estrato de motorización).

El número de viajes generados “observados” por cada hogar se construyó como la suma de la “tasa real” (Tabla 2) y un término de error *iid* Uniforme (-0.5, 0.5). Cabe destacar que, ya que los cuatro métodos ACM analizados buscan minimizar el error al cuadrado, la consistencia, errores de especificación, o eficiencia relativa de uno u otro no dependen de la estructura de error considerada. Luego, considerar estructuras más sofisticadas solo ayudaría a obscurecer sus resultados principales.

A continuación se calculó las tasas de viaje ACM_SM1 aplicando el método descrito anteriormente a los viajes “observados”. En la Tabla 3 Puede verse que el método ACM_SM1 estima tasas más altas que las reales para las categorías de más alto ingreso y tasa de motorización. Por el contrario, este método estima tasas menores que las reales, para los estratos de ingreso y motorización menores.

El método ACM_SM1 permite reproducir los viajes totales observados, pero no los viajes observados para cada estrato de ingreso o motorización. Más aún, en la estimación de viajes para los cortes temporales futuros ni siquiera se logra reproducir correctamente el número de viajes totales. Como se discutió anteriormente, todo escenario de distribución de hogares que considere crecimiento económico, se traducirá en un “corrimiento” de hogares desde los estratos del ingreso y motorización más bajos hacia los más altos, lo que causaría una sobreestimación del impacto en el total de viajes.

Para el experimento desarrollado se observa que la sobreestimación del impacto en los viajes por hogar, para los hogares que incrementan su estrato de motorización o ingreso, va desde un 17% a un 183%. En efecto, considérese por ejemplo el caso en el cual el único cambio corresponde al “corrimiento” de 100 hogares desde la categoría de ingreso dos y motorización cero, a la de ingreso dos y motorización uno. Si se consideran las tasas reales, el impacto en el número total de viajes sería de $(1,8-1,6)*100 = 20$ viajes. En cambio, si se consideran las tasas ACM_SM1, el impacto en el número total de viajes sería de $(1,92-1,43)*100 = 48$ viajes, es decir, un 183% de sobreestimación del cambio en el número de viajes.

Finalmente, cabe destacar que debido a que la sobreestimación de viajes se produciría de forma más pronunciada para las categorías de ingreso-motorización superiores, se debiera observar además una sobreestimación de la partición modal de transporte privado que es predicha por el modelo de equilibrio transporte que es alimentado por modelo de generación. Este impacto es discutido con mayor detalle en la aplicación con datos reales descrita más adelante en la sección 4.

A continuación se aplicó el método AMC_SM2 a la misma base de datos, resultando las tasas estimadas que son reportadas en la Tabla 4. Comparando los resultados de la Tabla 4 con los de la Tabla 3 y las tasas reales reportadas en la Tabla 2, puede verse que, para el caso lineal estudiado en esta sección, la aplicación del método ACM_SM2 reduce en cierta medida el problema de sobreestimación de viajes que ocurre con el corrimiento de hogares. Sin embargo, por tratarse de un método aproximado, éste no permite reproducir el número de viajes totales, ni por estrato ni por categoría. El efecto y relevancia de este último problema se mostrará más adelante en la sección 4, en aplicación con datos reales para la ciudad de Santiago.

Por su parte, en la Tabla 5 se muestran las tasas que resultan de la aplicación del método ACM_MCL al problema lineal. Puede verse que en este caso las tasas obtenidas son muy similares a las reales, en contraste con las que se obtenían con el método ACM_SM1. Adicionalmente, no se observa el sesgo positivo que implica una sobrestimación de tasas para categorías superiores, ni el sesgo negativo que implica una subestimación de tasas para categorías menores.

En la Tabla 6 se muestran las tasas que resultan de la aplicación del método ACM_PSC al problema lineal descrito en el la Tabla 1. Puede verse que, al igual que para el método ACM_MCL, en este caso las tasas obtenidas son muy similares a las reales, en comparación con las que se obtenían con el método ACM_SM1. Del mismo modo, no se observa el sesgo de sobrestimación de tasas para estratos de ingreso o motorización superior y subestimación para categorías menores.

El paso siguiente de la investigación correspondió a repetir el experimento de Monte Carlo descrito en las Tablas 1 y 2, generando diferentes vectores de error. El objetivo es comparar la precisión alcanzada por los métodos analizados independientemente de la componente de error existente. Se observó que, a partir de la décima simulación, el ordenamiento en cuanto a precisión de los métodos era el mismo. Sin embargo, para garantizar la robustez del análisis realizado, se consideraron 20 simulaciones de Monte Carlo para realizar la comparación. Para cada una de las 20 simulaciones se calculó el promedio de la diferencia porcentual absoluta entre las tasas “reales” y las modeladas por categoría, además del error, ponderado por el número de observaciones de cada categoría. Los resultados de estas 20 simulaciones fueron luego promediados y son presentados en la Tabla 7.

La primera conclusión que se deriva la Tabla 7 es que la estimación de tasas mediante el método ACM_SM1 es la más deficiente, con errores que superan el 25%. Por otra parte, el método ACM_SM2 permite reducir en cierta medida el problema, pero no alcanza a ser superior a la aplicación del método ACM_MCL ni al de ACM_PSC.

Se observa además que el método ACM_MCL es claramente superior a todas las alternativas. Es mejor que el ACM_SM1 pues el supuesto de igual número de observaciones por categoría esta muy lejos de cumplirse (Tabla 1). Es superior al ACM_SM2 pues dicho método corresponde sólo a una aproximación numérica que intenta dar cuenta de la disparidad en el número de observaciones por categoría. Finalmente el método ACM_MCL es superior al de ACM_PSC pues implica la estimación de menos coeficientes (cinco en vez de nueve) con una misma cantidad de observaciones, es decir, hace una estimación más eficiente (con menor varianza) de los coeficientes, al considerar mayor número de observaciones por parámetro estimado.

3.2. Modelo Subyacente No Lineal

En esta sección se investiga un caso en el cual la tasa “real” de generación de viajes no es lineal en el ingreso y la motorización sino que el efecto de un cambio en el estrato de ingreso dependen del estrato de motorización del hogar y viceversa. Al igual que en el caso lineal, las tasas “observadas” fueron construidas como la suma de la tasa “real” correspondiente a cada categoría (Tabla 8) y un término de error $U(-0,5 ; 0,5)$.

En la Tabla 9 se presenta el análisis de precisión promedio de los diversos métodos en el caso no lineal para 20 simulaciones de Monte Carlo. En este caso se puede observar que, tal como ocurría con el modelo lineal, el ACM_MCL es mejor que el ACM_SM2 y éste que el ACM_SM1. Sin embargo, en este caso los errores asociados a dichos métodos son varias veces superiores al que se obtiene con el método ACM_PSC. Lo que ocurre en este caso es que, como se señaló anteriormente, el método de ACM_PSC no supone que las tasas sean lineales, como si lo hacen los otros métodos.

Puede notarse además que los errores del método ACM_PSC son los mismos, tanto para el modelo lineal como para el no lineal. Lo que ocurre es que su nivel de error está asociado al número de observaciones que hay en cada categoría y al error de simulación, aspectos que no fueron cambiados al pasar del modelo lineal al no lineal.

4. COMPARACIÓN DE MÉTODOS ACM UTILIZANDO DATOS REALES OBTENIDOS DE LA EOD 2001 DE SANTIAGO DE CHILE

El ejercicio final correspondió a contrastar los métodos ACM investigados usando datos reales obtenidos en la Encuesta Origen Destino (EOD) 2001 realizada en la ciudad de Santiago. Los detalles respecto de la EOD 2001 pueden revisarse en Sectra (2003).

A diferencia de lo que ocurría en las simulaciones de Monte Carlo, en este caso no se conocen las tasas “reales” sino que éstas son precisamente el objeto de estimación a partir de los datos disponibles. Con este fin, a partir de la base datos sin expandir, se calculó las tasas de viajes BHI generados por hogar entre 7:15 y 9:15 AM de un día laboral en temporada normal, utilizando los métodos ACM_SM1, ACM_SM2, ACM_MCL y ACM_PSC, las cuales son reportados en la Tabla 10.

En la Tabla 10 puede observarse que, salvo por el estrato de ingreso 5, las tasas obtenidas con el método ACM_SM1 tienen un comportamiento “apropiado” en el sentido de siempre aumentar con el ingreso y la motorización. Sin embargo, tal como ocurría en las simulaciones de Monte Carlo, esta propiedad, aparentemente positiva, implica una distorsión importante en los coeficientes estimados que producen la sobreestimación de los viajes generados en los cortes temporales futuros. En efecto, puede verse que, comparado con el ACM_PSC, el método ACM_SM1 tiende a reducir las tasas para las categorías de ingreso y motorización más bajas y amplificarlas para los más altos.

Respecto del método ACM_SM2, puede verse que éste corrige las tasas calculadas mediante el método ACM_SM1, en el sentido correcto, esto es, “aplanando” la matriz de tasas y haciéndolo así más semejante a la obtenida con los métodos ACM_MCL y ACM_PSC. Sin embargo, esta mejora esconde aún un problema en la predicción de viajes futuros, la cual se verá con claridad más adelante en este trabajo, cuando se analicen el impacto de considerar escenarios de hogares para cortes temporales futuros.

Las matrices de tasas estimadas con ACM_MCL y ACM_PSC son muy similares¹. Para decidir sobre la utilización de una u otra se calculó un *test* F para verificar si los modelos estimados con dichos métodos eran estadísticamente iguales. Los coeficientes necesarios para realizar este *test* corresponden a:

- El número de observaciones ($n=9.038$).
- El número de coeficientes del modelo no restringido, que en este caso corresponde al modelo de ACM_PSC ($k=15$).
- El número de restricciones que se imponen para llegar al modelo restringido ACM_MCL ($m=7$).
- El R^2 de los modelos ACM_MCL ($R^2=0,11$) y ACM_PSC ($R^2=0,12$).

El valor del *test* F en este caso es 2,74, lo cual supera el valor crítico $F[J; n-K]=F[8; 9,023]=1.94$ con 95% de confianza. Esto implica que el modelo de ACM_PSC es

¹ El detalle de los modelos estimados no es presentado por razones de espacio. Puede ser enviado por correo electrónico a quienes lo soliciten.

estadísticamente diferente al modelo lineal o ACM_MCL y, de acuerdo a lo discutido anteriormente, el método de ACM_PSC debiera ser el seleccionado para estimar las tasas de generación buscadas. Cabe destacar que, en una aplicación real, incluso en caso de haberse aceptado la hipótesis nula, quedaría por investigar, antes de aceptar el modelo ACM_MCL, la conveniencia de utilizar modelos que consideren no-linealidades sólo para algunas categorías.

En forma adicional, puede observarse que las tasas estimadas con los métodos ACM_MCL y ACM_PSC son muy similares, salvo para las categorías de ingreso bajo-motorización alta e ingreso alto-motorización baja. Esto concuerda con las categorías para las cuales se tiene un número menor de observaciones, lo que indicaría que el efecto se explica por sesgo por tamaño de la muestra. En una aplicación real en la que se quisiera utilizar este método, el problema debiera ser subsanando mediante la fusión de categorías con pocas observaciones.

El paso final en el esta investigación correspondió a estimar el numero de viajes BHI por categoría predichos por los diversos métodos ACM analizados, para los cortes temporales 2005 y 2010, utilizando sendos escenarios de uso de suelo desarrollados para la ciudad de Santiago. En las Tablas 11 y 12 se presenta el número de viajes BHI por estrato de motorización para el método ACM_PSC, que es usado como referencia para calcular la diferencia porcentual en la predicción de los otros métodos. Este mismo análisis puede realizarse por estrato de ingreso, en vez de por motorización, caso en el cual las conclusiones que se obtienen son equivalentes, y luego, no son presentadas en este trabajo por razones de espacio.

Puede verse que, tanto para el corte temporal 2005 como para el 2010, el número total de viajes predichos por el método ACM_PSC es superado por el total predicho por el método ACM_SM1 y que éste a su vez es superado por el total predicho por el ACM_SM2. En cambio, el método ACM_MCL predice un número total comparativamente muy similar al predicho por el ACM_PSC.

Por otra parte, puede verse que el método ACM_SM1 predice un número menor de viajes BHI generados desde hogares con motorización cero (cautivos de transporte público), que el método ACM_PSC. En cambio, el ACM_SM1 predice un número mayor de viajes BHI, generados desde hogares con motorización 1 y 2+ (que tienen la opción de transporte privado) que el método ACM_PSC.

Lo anterior implica que, si los vectores obtenidos con el método ACM_SM1 son utilizados en un modelo de equilibrio, el **número y porcentaje relativo de viajes predichos en transporte privado será sobreestimado y el de transporte público subestimado**. Este efecto sería el resultado de la subestimación de usuarios cautivos de transporte público (entre 19% y 18%); la sobreestimación de usuarios con disponibilidad de automóvil (entre 19% y 30%); y el efecto negativo en la participación de transporte público que tiene el aumento irreal en los niveles de congestión, el cual es producido por el aumento de viajes en transporte privado. Cabe destacar que la sobrestimación de viajes, de vehículos y del consumo de recursos en el sistema, implicaría una **sobre-valoración de beneficios privados y sociales de proyectos de infraestructura de transporte, que podría conducir a evaluaciones sociales o privadas erróneas**.

Con respecto al método ACM_SM2, en las Tablas 11 y 12 puede verse que el problema de sobrestimación de viajes es más semejante entre diversos estratos de motorización (se “corrige

para arriba”), reduciendo en cierta medida el efecto descrito para el caso del método ACM_SM1. De esta manera el porcentaje relativo de viajes en transporte privado no se vería tan afectado. En cambio, el valor absoluto si será sobreestimado debido a la sobrestimación de viajes con disponibilidad de automóvil y el efecto de la sobreestimación de la congestión.

Por último, puede verse que el método ACM_MCL, que fue descartado mediante el *test* F descrito anteriormente, entrega resultados muy similares a los del ACM_PSC, en comparación con el nivel de error producido por los otros métodos. Lo que sucede es que la inconsistencia por atributos omitidos en que caería en método ACM_MCL dista mucho de tener la gravedad de los serios problemas de especificación que presentan los métodos ACM_SM1 Y ACM_SM2.

5. CONCLUSIONES Y RECOMENDACIONES

La primera conclusión que se puede obtener del presente trabajo es que el método ACM_SM1, la técnica más utilizado para estimar viajes basados en el hogar de ida en modelos clásicos de cuatro etapas, debe ser descartado porque se basa en supuestos que raramente se verifican y que cuya violación se traduce en sesgos importantes en la modelación de sistemas de transporte.

Por su parte, el método ACM_SM2 debe ser igualmente descartado, pues corresponde a una corrección numérica del método ACM_SM1 que pretende hacerlo más similar al ACM_MCL, modelo este último que es fácilmente estimable.

Los métodos ACM_MCL y ACM_PSC son muy superiores, en cuanto a precisión y sustento teórico, al resto de los métodos analizados. La selección de uno u otro dependerá del caso investigado, lo cual puede ser probado estadísticamente. Sin perjuicio de lo anterior puede afirmarse que el modelo ACM_PSC es más robusto frente a la especificación del modelo subyacente ya que, en el peor de los casos, sólo se estaría perdiendo eficiencia (y no consistencia) al considerarlo por sobre el ACM_MCL.

Finalmente, cabe destacar el bajo nivel de ajuste ($R^2 = 0,1166$) alcanzado por los modelos de generación que dependen sólo de las categorías de ingreso y motorización. Si se agrega como variable de estratificación el tamaño de hogar (cubriendo así la gran mayoría de los modelos ACM estimados en el mundo) se obtienen mejoras estadísticamente significativas, pero el modelo resultante posee aún baja capacidad explicativa ($R^2 = 0,1332$). Esto refleja la necesidad de investigar modelos más sofisticados que no consideren sólo variables dicotómicas por categoría. De forma adicional, en esta investigación se exploraron preliminarmente especificaciones no-lineales considerando variables continuas y variables dicotómicas entre las cuales se consideró la accesibilidad y la composición de hogar. Mediante este ejercicio se lograron mejoras sustanciales ($R^2 = 0,3012$), pero indudablemente queda aún mucho camino por explorar a este respecto.

AGRADECIMIENTOS

Quisiéramos agradecer los valiosos comentarios realizados por Juan de Dios Ortúzar, Joaquín De Cea, Reinaldo Guerra, Luis Rizzi y Marisol Castro, los cuales permitieron mejorar sustancialmente versiones preliminares de este documento. Posibles errores u omisiones que persistan son de nuestra exclusiva responsabilidad.

REFERENCIAS

Glass, G.V y G.C. Stanley (1986) **Métodos Estadísticos Aplicados a las Ciencias Sociales**. Prentice Hall, Mexico.

Greene, W. (2003) **Econometric Analysis**, Prentice Hall, 5th Edition.

Guevara C.A. y M. Ben-Akiva (2006) Endogeneity in Residential Location Choice Models. **Transportation Research Record: Journal of the TRB**, No 1977, Washington, DC. 2006. pp. 60-66

Ortúzar, J. de D. y L. Willumsen (1994) **Modelling Transport**. John Willey & Sons, London.

SECTRA (2003) **Actualización de Encuestas Origen y Destino de Viajes, V Etapa**. Informe final del estudio homónimo preparado por DICTUC para SECTRA, MIDEPLAN, Chile.

SECTRA (1998) **Metodología de Análisis de Sistemas de Transporte Urbano**. Obtenido en línea desde http://www.sectra.cl/contenido/metodologia/transporte_urbano/Analisis_sistema_transporte_urbano.htm. Elaborado por Fernández y De Cea Ingenieros Limitada para SECTRA.

Stopher, P.R. y K. McDonald (1983) Trip generation by cross-classification: an alternative methodology. **Transportation Research Record** 944, 84-91.

TMIP (2004) **Report on Findings of the Second Peer Review Panel for Southern California Association of Governments (SACG)**. US DOT Travel Model Improvement Program

Tabla 1: Hogares por Categoría, Simulaciones Monte Carlo

# Hogares	Motorizacion			Total general
Ingreso	0	1	2	
1	352	132	3	487
2	109	213	99	421
3	2	24	66	92
Total general	463	369	168	1000

Tabla 2: Tasas “Reales”, Modelo Lineal, Simulaciones Monte Carlo

Tasas REALES	Motorizacion		
Ingreso	0	1	2
1	1,00	1,20	1,40
2	1,60	1,80	2,00
3	2,20	2,40	2,60

Tabla 3: Tasas ACM_SM1, Modelo Lineal, Simulaciones Monte Carlo

ACM_SM1	Motorizacion		
Ingreso	0	1	2
1	0,697	1,19	1,75
2	1,43	1,92	2,49
3	2,13	2,62	3,19

Tabla 4: Tasas ACM_SM2, Modelo Lineal, Simulaciones Monte Carlo

ACM_SM2	Motorizacion		
Ingreso	0	1	2
1	0,982	1,19	1,34
2	1,59	1,80	1,94
3	2,18	2,39	2,54

Tabla 5: Tasas ACM_MCL Modelo Lineal, Simulaciones Monte Carlo

ACM_MCL	Motorizacion		
Ingreso	0	1	2
1	0,994	1,21	1,39
2	1,59	1,80	1,98
3	2,15	2,37	2,55

Tabla 6: Tasas ACM_PSC, Modelo Lineal, Simulaciones Monte Carlo

ACM_PSC	Motorizacion		
Ingreso	0	1	2
1	1,00	1,20	1,31
2	1,57	1,81	1,98
3	2,23	2,35	2,55

Tabla 7: Error Diferentes Métodos, Modelo Lineal, Simulaciones Monte Carlo

	ACM_SM1	ACM_SM2	ACM_MCL	ACM_PSC
Error Promedio	27%	5,3%	2,4%	4,8%
Error Promedio Ponderado	25%	3,1%	1,7%	2,0%

Tabla 8: Tasas Reales Modelo No Lineal, Simulaciones Monte Carlo

Tasas REALES	Motorizacion		
Ingreso	0	1	2
1	0,800	1,20	2,10
2	1,60	1,80	2,00
3	2,70	2,40	2,80

Tabla 9: Error Diferentes Métodos, Modelo No-Lineal, Simulaciones Monte Carlo

	ACM_SM1	ACM_SM2	ACM_MCL	ACM_PSC
Error Promedio	39%	23,5%	18,3%	4,8%
Error Promedio Ponderado	32%	19,4%	4,9%	2,0%

Tabla 10: Tasas de Generación de Viajes BHI, Diversos Métodos, EOD 2001

ACM_SM1	Motorización			ACM_SM2	Motorización		
Ingreso	0	1	2	Ingreso	0	1	2
1	0,0850	0,218	0,412	1	0,230	0,152	0,261
2	0,323	0,457	0,651	2	0,450	0,371	0,481
3	0,599	0,733	0,926	3	0,715	0,636	0,746
4	0,794	0,927	1,12	4	0,949	0,871	0,980
5	0,757	0,890	1,08	5	0,986	0,908	1,02

ACM_MCL	Motorización			ACM_PSC	Tasa Simple		
Ingreso	0	1	2	Ingreso	0	1	2
1	0,159	0,0741	0,119	1	0,146	0,202	0,400
2	0,407	0,322	0,367	2	0,393	0,369	0,600
3	0,699	0,614	0,659	3	0,708	0,606	0,590
4	0,912	0,827	0,871	4	0,955	0,809	0,844
5	0,879	0,794	0,839	5	1,05	0,731	0,863

**Tabla 11: Número de Viajes BHI y Porcentaje Respecto de ACM_PSC
Estimados por Diversos Métodos, Escenario Santiago 2005**

Motorización	ACM_SM1	ACM_SM2	ACM_MCL	ACM_PSC
0	-19%	8%	-0,53%	4.19.022
1	19%	9%	0,92%	329.156
2+	30%	18%	-0,39%	172.105
Total	3,8%	10%	0,015%	920.283

**Tabla 12: Número de Viajes BHI y Porcentaje Respecto de ACM_PSC
Estimados por Diversos Métodos, Escenario Santiago 2010**

Motorización	ACM_SM1	ACM_SM2	ACM_MCL	ACM_PSC
0	-18%	4,6%	-1,8%	459.660
1	18%	9,9%	2,2%	448.738
2+	30%	18%	-0,44%	272.236
Total	6,8%	9,7%	0,055%	1.180.635