

# **ESTIMACIÓN DEL VALOR DEL TIEMPO USANDO MODELOS DE ELECCIÓN DISCRETA CON DATOS DE PREFERENCIAS REVELADAS Y NIVELES DE SERVICIO AGREGADOS**

Marco Batarce

Facultad de Ingeniería y Ciencias, Universidad Diego Portales

marco.batarce@udp.cl

## **RESUMEN**

Este artículo propone un método para estimar el valor del tiempo a partir de modelos de elección discreta utilizando datos agregados sobre las variables de nivel de servicio. El objetivo de calcular la disposición a pagar por el ahorro de tiempo de viaje utilizando los parámetros estimados. Sin embargo, debido a errores de medición, estos parámetros son sesgados cuando se usan datos agregados. Como en los datos de elección de modo se reportan los atributos de la alternativa elegida, es posible utilizar esta información como una muestra de validación y corregir la estimación con información agregada. Para esto, se utilizó tanto una estimación paramétrica como no-paramétrica de la densidad condicional de las variables sin error en las variables con error y se estimó el modelo paramétrico usando máxima verosimilitud en dos etapas. El método se aplica a datos de Santiago.

*Palabras clave:* elección discreta, error de medición, valor del tiempo

## **ABSTRACT**

This paper proposes a method for estimating the value of time from discrete choice models using aggregate data on the level of service variables. The objective to compute the willingness to pay for travel time saving using the estimated parameters. However, because of measurement errors, these parameters are biased when using aggregate data. As the data on mode choice report the attributes of the chosen alternative, it is possible to use this information as a validation sample and correct the estimation with aggregate information. I use both parametric and nonparametric estimation of the conditional density of variables without error on that with error, and estimate the parametric model using a two-step maximum likelihood. The method is applied to data from Santiago.

*Keywords:* discrete choice, measurement error, value of time

## 1. INTRODUCCIÓN

En Chile, el Sistema Nacional de Inversiones Públicas (SNI) basa su metodología para evaluación de proyectos de transporte en el ahorro de recursos económicos valorados a precios sociales. En particular, el tiempo de viaje de las personas es considerado un recurso económico que debe ser ahorrado, ya que puede ser asignado a otras actividades productivas. En este sentido, el tiempo de las personas tiene un costo de oportunidad para la sociedad. Es el SNI quien se encarga de definir cuál es el Valor Social del Tiempo (VST) y publicarlo anualmente para su uso en evaluación de proyectos.

La metodología usada para calcular el VST para viajes urbanos se basa en distinguir dos tipos de propósitos de viaje: por trabajo y el resto de propósitos. El valor del tiempo de viajes por trabajo se estima como el salario por hora promedio nacional. El valor del tiempo de viajes con propósitos que no son trabajo se estima como una fracción fija del valor del tiempo de viaje por trabajo. Por lo tanto, el VST está determinado por el salario promedio de las personas. De esta forma se publica un único VST independiente de la ciudad y de los usuarios que son los beneficiarios del proyecto evaluado.

El enfoque utilizado hasta 2014 suponía que el tiempo de viaje ahorrado podrá ser utilizado por el individuo para trabajar más. Eso sólo es cierto si es posible elegir cuánto tiempo dedicar al trabajo. Pero en el caso chileno la mayor parte de los individuos tienen jornadas laborales fijas y lo que hacen es utilizar el tiempo de viaje ahorrado en actividades de ocio, aunque sea un viaje al trabajo. Por lo tanto, los ahorros de tiempo de viaje no se traducen en aumentos en la producción de los individuos que viajan al trabajo. El supuesto sí sería correcto cuando los viajes se realizan por motivos de trabajo en medio de la jornada laboral.

Lo anterior conduce a plantear que los ahorros de tiempo de viaje se deben valorar de una forma diferente. En primer lugar, se debe reconocer que los individuos están dispuestos a pagar por ahorrar tiempo de viaje cualquiera que sea propósito del viaje.<sup>1</sup> En segundo lugar, se debe considerar que el VST que se utiliza en la evaluación social de proyecto es una decisión de la sociedad en su conjunto y que responde a criterios y restricciones presupuestarias colectivas. Por lo tanto, el VST tiene que representar el valor monetario que la sociedad le asigna al cambio de bienestar de los individuos producto del ahorro de tiempo de viaje. Para esto se plantea calcular el VST con un enfoque basado en la disposición a pagar individual por ahorrar tiempo de viaje, pero consistente con criterios generales de la sociedad.

En esta línea, recientemente el Gobierno chileno modificó el método de cálculo del Valor Social del Tiempo de Viaje, adoptando como valor para los viajes con propósito distinto al de trabajo la disponibilidad a pagar por ahorrar tiempo de viaje. El valor resultante es posteriormente promediado con el valor del tiempo de trabajo, igual a la tasa salarial. Por lo tanto, ahora el valor social del tiempo incluye preferencias individuales, no solo la productividad del trabajo.

---

<sup>1</sup> Esto será así siempre y cuando el viaje en sí mismo no sea una actividad que produce bienestar al individuo.

El objetivo de esta investigación es estimar la disponibilidad a pagar por ahorrar tiempo de viaje a partir de la estimación de modelos de elección de modo de viaje. Esta disposición a pagar puede utilizarse para calcular el valor social del tiempo en el análisis costo-beneficio de proyectos de transporte.

La información disponible para la estimación de modelos proviene de encuestas de preferencias reveladas recolectadas en el contexto de la Encuesta de Movilidad de Santiago en el 2012 (EOD). La ventaja de esta encuesta es su representatividad, ya que la muestra disponible para estimación de modelos de elección de modo es más de 38.000 viajeros, y la selección es aleatoria (estratificada). Sin embargo, el uso de encuestas de preferencias reveladas para modelar elección de modo tiene la desventaja de que los atributos (niveles de servicio) de los modos alternativos al elegido no son reportados por el individuo encuestado, y por lo tanto, deben ser estimados para construir la base de datos para la estimación. Una alternativa para determinar los niveles de servicio de los modos alternativos es seleccionar una muestra pequeña y estimar para cada individuo los niveles de servicio. Esto podría tener un problema de representatividad de la muestra. Otra alternativa es usar información de tiempos de viaje y de espera agregada (a nivel de zonas) que se obtiene de modelos de redes de transporte. Esto puede introducir sesgos en las estimaciones porque los regresores tienen errores de medición (Chen et al., 2011). Por lo tanto, para usar datos agregados se debe controlar este error de medición, que es el objetivo de este trabajo.

## 2. MODELOS DE ELECCIÓN DISCRETA BASADOS EN UTILIDAD ALEATORIA

La metodología para la estimación del valor del tiempo se basa la estimación de los parámetros de la función de utilidad de modelos de elección discreta de modo de transporte. Estos modelos, a su vez, se basan en la teoría de la utilidad aleatoria. En el contexto de la elección de modo de transporte, esta teoría puede resumirse en los siguientes supuestos sobre el comportamiento de los individuos que eligen una alternativa de transporte.

- Existe un conjunto (finito) de alternativas de transporte, excluyentes entre sí, para realizar el viaje.
- Las preferencias individuales sobre las alternativas pueden representarse a través de una función de utilidad que depende de las características de las alternativas y de individuo que elige.
- Cada individuo elige el modo de transporte que le genera la mayor utilidad entre todos los disponibles en el conjunto de alternativas.
- Existen variables en la función de utilidad de cada individuo que sólo son observables por él. De esta forma dos individuos con el mismo conjunto de alternativas disponibles y con las mismas características observables pueden elegir distintos modo para realizar un viaje.
- Se asume que las componentes no observables de la utilidad individual son aleatorias o que están distribuidas aleatoriamente en la población.

La componente aleatoria de la utilidad puede tener su origen en distintas fuentes. Por ejemplo, cualquier atributo de las alternativas no observable o medible, o la variación no observada en los

gustos de los individuos. Todos los supuestos anteriores dan origen a la teoría microeconómica de la utilidad aleatoria.

En términos prácticos, la teoría de la utilidad aleatoria implica definir una función de utilidad para cada alternativa de transporte, la cual tiene como variables las características del modo, las del individuo y una componente aleatoria con una distribución sobre la población. Analíticamente, la utilidad aleatoria de la alternativa  $m$  para el individuo  $i$  se escribe como  $V(x_m, z_i, e_{mi})$ , donde  $x_m$  es un vector con características del modo  $m$  (tiempo de viaje, costo, etc.),  $z_i$  es un vector con características del individuo  $i$  (ingreso, edad, etc.), y  $e_{mi}$  es la componente aleatoria. De esta forma la utilidad es una variable aleatoria también.

Dado que se asume que los individuos eligen la alternativa que maximiza su utilidad, entonces el modo  $m$  es elegido si  $V(x_m, z_i, e_{mi}) \geq V(x_k, z_i, e_{ki})$  para cualquier otro modo  $k$  en el conjunto de modos disponibles del individuo. Dado que la utilidad es una variable aleatoria, entonces se puede escribir la probabilidad de que el individuo  $i$  elija la alternativa  $m$  como  $P(m \text{ elige } i) = P(V(x_m, z_i, e_{mi}) \geq V(x_k, z_i, e_{ki}), \text{ para todo } k)$ . Esto es el origen de los distintos modelos de elección discreta.

De acuerdo con los supuestos que se hagan sobre la forma funcional de la utilidad y la distribución de probabilidad de la componente aleatoria se obtienen diferentes modelos. Por ejemplo, el modelo logit se obtiene asumiendo que la componente aleatoria es aditiva y separable en la función de utilidad y que distribuye Valor Extremo tipo I. El modelo probit resulta también de asumir que la componente aleatoria es separable, pero que distribuye normal.

Además, habitualmente se asume que la parte observable de la utilidad es una función lineal de las características del modo. De esta forma, si la función de utilidad incluye tiempo y costo, los parámetros asociados a ellos representan la utilidad marginal de tales variables. Por ejemplo, si  $V_i = a_i - bC_i - cT_i + e_i$ , donde  $C_i$  es el costo y  $T_i$  el tiempo de viaje, entonces  $b$  es la utilidad marginal de ingreso y  $c$  es la utilidad marginal de tiempo (en términos simplificados).

La tasa marginal de sustitución entre tiempo y dinero corresponde al valor subjetivo del tiempo y se interpreta como la disposición a pagar por ahorrar tiempo de viaje. Por lo tanto, este valor puede ser calculado como la razón entre los parámetros del tiempo y el costo ( $c/b$ ) de la función de utilidad lineal. En este estudio, el valor subjetivo del tiempo, a su vez, se utilizará para calcular el valor social de tiempo.

Hasta aquí la metodología general es clara. Lo que se requiere es estimar modelos de elección de modo de transporte para viajes urbanos y a partir de la función de utilidad calcular el valor subjetivo del tiempo. Sin embargo, la complejidad de esta tarea radica en especificar una forma funcional de la parte observable de la utilidad (de aquí en adelante solo diremos la utilidad) que represente adecuadamente el comportamiento de los usuarios y una función de distribución de probabilidad que represente correctamente la componente no observable (o aleatoria) de la utilidad (de aquí en adelante, la componente aleatoria).

La función de distribución de probabilidad que se asume habitualmente para la componente aleatoria es la de Valor Extremo tipo I, que resulta en probabilidades de elección tipo logit. Sin embargo, a pesar de tener ventajas en la estimación y en el uso para predecir demanda, los modelos

logit tienen algunas desventajas. Por ejemplo, imponen una estructura de correlación nula entre las componentes aleatorias de las utilidades modales. También las elasticidades presentan un patrón rígido, que puede ser poco realista (ver Train, 2002). Finalmente, si la verdadera distribución de las componentes aleatorias no es Valor Extremo tipo I o no se aproxima, se produce sesgo por mala especificación al estimar los parámetros y, en consecuencia, en el valor subjetivo del tiempo.

Para superar las limitaciones de los modelos logit existen varias alternativas. Una alternativa es el uso de modelos con error compuesto. En este tipo de modelo se asume que la estructura de correlación es arbitraria y se estima junto con los parámetros del modelo. Según Walker et al. (2007), este modelo incluye como casos especiales a modelos con varianzas específicas por alternativa (como logit heterocedástico), modelos con estructuras anidadas (tipo logit jerárquico) y con estructuras de anidadas cruzadas (tipo cross-nested logit).

Otra alternativa más general son los modelos logit mixtos o con parámetros aleatorios. Según McFadden y Train (2000) cualquier modelo de elección discreta basado en la utilidad aleatoria puede ser aproximado tanto como se quiera por un modelo con parámetros aleatorios. De esta forma, es posible reducir sesgos por mala especificación. También estos modelos permiten capturar la heterogeneidad en las preferencias individuales. Un inconveniente de usar un enfoque de parámetros aleatorios con este fin es que la distribución de las preferencias reales no sea bien representada por la distribución supuesta de los parámetros, lo que podría resultar en sesgos por mala especificación, o que exista de la heterogeneidad asociada a componentes sistemáticos (no aleatorios) asociados a alguna característica observable de los individuos.

### **3. INFORMACIÓN DISPONIBLE E IMPLICANCIAS EN LA ESTIMACIÓN**

La información que se utilizará proviene de la encuesta de origen y destino de viajes (EOD) de Santiago del año 2012. Esta encuesta entrega información de los viajes realizados por individuos en la ciudad, además de características personales y del hogar del individuo. La información sobre el viaje incluye solo el valor los atributos (costo, tiempo de viaje, de espera y acceso) del modo elegido por el usuario. Esto significa que no se dispone de los atributos de los modos alternativos. Por lo tanto, se requiere determinar para cada viaje en la muestra los atributos o niveles de servicio de los modos alternativos. Esto es factible de hacer sólo a partir de información agregada de distancia, tiempos de viaje y de espera a nivel de zonas (específicamente en forma de matrices entre zonas). Esto implica introducir sesgos en la estimación debido al error de medición de las variables independientes del modelo.

La presencia de este error de medición viola los supuestos básicos de los métodos de estimación disponibles, que es la independencia entre los regresores y la componente aleatoria de la función de utilidad. Para ver esto, consideremos que la utilidad del modo  $m$  depende del tiempo,  $T_m$ , y el costo de viaje,  $C_m$ , y que el verdadero tiempo de viaje,  $\tau_m$ , está medido con un error  $\varepsilon_m$ , tal que  $T_m = \tau_m + \varepsilon_m$ . Además, consideremos que dicho error es aleatorio (si fuera sistemático, se podría intentar eliminar) con media cero, varianza  $\sigma_m^2$  e independiente de  $\tau_m$ . Entonces la utilidad del modo  $m$  será

$$\begin{aligned}
V_m &= a_m + bC_m + c\tau_m + e_m \\
&= a_m + bC_m + c(T_m - \varepsilon_m) + e_m \\
&= a_m + bC_m + cT_m - c\varepsilon_m + e_m \\
&= a_m + bC_m + cT_m + u_m
\end{aligned}$$

Luego, al estimar el modelo usando el tiempo con error,  $T_m$ , la componente aleatoria no es independiente de este regresor. Basta tomar la esperanza del producto del tiempo de viaje y el error para ver que no se cumple la independencia.

$$\begin{aligned}
E(T_m u_m) &= E[(\tau_m + \varepsilon_m)(e_m - c\varepsilon_m)] \\
&= -cE(\varepsilon_m^2) \\
&= -c\sigma_m^2
\end{aligned}$$

Por lo tanto, el uso de niveles de servicio agregados (zonales) puede inducir sesgos porque los estimadores no serán consistentes.

Para solucionar este problema se pueden usar distintas técnicas como variables instrumentales y función de control. En este caso ambos enfoques requerirían encontrar una variable que esté correlacionada con el tiempo de viaje y no lo esté con el error de medición. En caso de no existir dicha variable, no es posible utilizar estos métodos. Una alternativa es suponer una distribución para el error de medición, introducirlo explícitamente en la función de utilidad y estimar un modelo con error compuesto donde el origen de la componente que no distribuye Valor Extremo tipo I es el término  $c\varepsilon_m$ . Mientras la distribución del error supuesta se aproxime a la verdadera distribución no existirían problemas de mala especificación y el modelo podría ser consistentemente estimado. Este enfoque además permitiría estimar heterocedasticidad y correlación entre alternativas, aunque no se podría identificar separadamente ninguna estas características de los errores de los modos de las de los errores de medición.

#### 4. METODOLOGÍA

El problema del sesgo por error de medición en variables también se puede reducir utilizando la información de niveles de servicio reportada por los encuestados en la EOD junto con la información a nivel de zonas. La metodología requiere tener una estimación de la distribución del valor real de la variable condicional en el valor de la misma con error de medición. El supuesto básico es el siguiente:

SUPUESTO 1:

$$\Pr(y_m = 1 | x_m^*, x_m) = \Pr(y_m = 1 | x_m^*),$$

donde  $y_m$  es igual a 1 si el individuo elige la alternativa  $m$  cero si no,  $x^*$  son las variables sin error, pero no observadas, y  $x$  son las variables medidas con error y observadas.

El supuesto significa que condicional en las verdaderas variables no observadas  $x^*$ , el error de medición contenido en  $x$  no proporciona información adicional con respecto a la distribución condicional de la variable que representa la elección del usuario. En términos de la esperanza de la variable dependiente, el supuesto anterior implica que

$$E(y_m | x^*, x) = E(y_m | x^*) = P_m(x^*, \beta)$$

La información sobre los parámetros del modelo,  $\beta$ , puede ser deducida de la distribución conjunta de las variables observadas  $x$  e  $y_m$ , a partir de la siguiente relación (Chen et al., 2011)

$$E(y_m | x) = \int E(y_m | x^*, x) f(x^* | x) dx^* = \int P_m(x^*, \beta) f(x^* | x) dx^*$$

Notar que  $E(y_m | x) = \Pr(y_m = 1 | x)$ , ya que  $y_m$  es una variable binaria. Por lo tanto, conociendo la densidad de la variable sin error condicional en la variable observada con error,  $f(x^* | x)$ , se puede estimar los parámetros del modelo sin sesgo. Para estimar la densidad  $f(x^* | x)$  se requiere de una muestra de validación o una muestra de datos auxiliares (por ejemplo, Chen et al., 2005).

Los datos de la encuesta EOD proporcionan la muestra de validación, lo que permite estimar la densidad en un primer paso y, dada la estimación la densidad  $\hat{f}$ , estimar los parámetros del modelo en un segundo paso. Se propone dos alternativas. Una es estimar la densidad condicional paramétricamente mediante regresión lineal y asumir que la distribución de los errores de la regresión siguen una distribución normal. La otra alternativa es estimar no paramétricamente la densidad. En ambos casos se estima los parámetros del modelo mediante máxima verosimilitud. Para una muestra de  $n$  individuos, el estimador máximo verosímil es definido como

$$\hat{\beta} = \arg \max_{\beta} \sum_{h=1}^n \sum_{m \in M_h} y_{hm} \ln \hat{P}(y_{hm} = 1 | x_h) = \sum_{h=1}^n \sum_{m \in M_h} y_{hm} \ln \int P_m(x_h^*, \beta) \hat{f}(x_h^* | x_h) dx^* ,$$

donde  $y_{hm}$  toma el valor 1 si el individuo  $h$  elige la alternativa  $m$  y 0 en otro caso,  $M_h$  es el conjunto de alternativas disponibles para el individuo  $h$ , la densidad  $\hat{f}$  es una estimación de  $f$ , y  $P_m(x_h^*, \beta)$  es el modelo paramétrico de la probabilidad de elección, por ejemplo un modelo logit simple o uno con parámetros aleatorios. Notar que si  $\hat{f}$  es una estimación no-paramétrica de  $f$ , el estimador de  $\beta$  es semi-paramétrico (por ejemplo, Hu y Ridder, 2010).

Para la estimación se requiere hacer supuestos adicionales. El principal problema es que para un vector  $x$  de variables independientes (atributos de los modos) se observa sólo una componente, la del modo elegido. Por ejemplo, si los atributos son tiempos de viaje, hay dos alternativas y la elegida es la alternativa 1, entonces del vector  $x=(x_1, x_2)$  se tiene una medida sin error sólo de  $x_1$ . Es decir, la muestra de validación sólo incluye un subconjunto de los regresores medidos con error. Esto dificulta la estimación de la función de distribución condicional puesto que solo hay información de la distribución marginal de  $x_1$  condicional en que es elegida la alternativa 1.

Para formalizar los supuestos para estimación consideremos que hay dos alternativas indexadas por  $i \in \{1, 2\}$  y que el único atributo de las alternativas es el tiempo de viaje. Definimos la variable dependiente  $y = (y_1, y_2)$  como un vector de ceros con un uno en la posición de la alternativa elegida. La variable independiente es el tiempo de viaje que denotamos por  $x^* = (x_1^*, x_2^*)$  cuando es medida sin error y por  $x = (x_1, x_2)$  cuando es medida con error.

Como la muestra de validación se observa sólo condicional en la alternativa elegida, para estimar la densidad de la variable sin error condicional en la variable observada con error,  $f(x_j^* | x_j)$ , se necesita un supuesto adicional.

SUPUESTO 2:  $f(x_j^* | x_j, y_j = 1) = f(x_j^* | x_j)$ .

Note que este supuesto no implica que la distribución del valor del atributo de la alternativa  $j$  sea independiente de la elección de esa alternativa. De hecho, no puede ser así por construcción del modelo. Lo que el supuesto dice es que la distribución de  $x_j^*$  condicional en  $x_j$  a nivel de población es la misma que a nivel de la muestra que elige la alternativa  $j$ .

Además, como muestra de validación sólo entrega información de las distribución marginal de las variables, y evidentemente el tiempo de viaje en una alternativa está correlacionado con el de la otra, no es posible identificar la distribución conjunta de  $x^* | x$  sin supuestos adicionales.

SUPUESTO 3:  $x_i^* | x_i$  es independiente de  $x_j^* | x_j$  con  $i \neq j$ .

Este supuesto permite estimar la distribución de errores condicional en la variable medida con error y recuperar la distribución condicional de las variables sin error. De esta forma, la densidad condicional conjunta es

$$f(x^* | x) = f_{x_1^* | x_1}(x_1^* | x_1) \cdot f_{x_2^* | x_2}(x_2^* | x_2).$$

Por lo tanto, con la muestra de validación se debe estimar la densidad marginal condicional de las variables medidas con error.

Si se utiliza estimación no paramétrica no son necesarios más supuestos. En particular, si se utiliza kernels para estimar la densidad condicional  $f$ , con una muestra de validación  $\{(X_{vk}^*, X_{vk}): k = 1, \dots, n_v\}$ , entonces  $E(y_m | x) = \Pr(y_m = 1 | x)$  se estima como

$$\hat{P}(y_m = 1 | x) = \frac{1}{n_v a^d} \sum_{k=1}^{n_v} \frac{P_m(X_{vk}^*, \beta) K\left(\frac{(X_{vk}-x)}{a}\right)}{\frac{1}{n_v a^d} \sum_{k=1}^{n_v} K\left(\frac{(X_{vk}-x)}{a}\right)},$$

donde  $K$  es un kernel de  $\mathbb{R}^d$  en  $\mathbb{R}$  (en este caso  $d=2$ ), y  $a$  es un ancho de banda tal que  $n_v a^d \rightarrow \infty$  y  $a \rightarrow 0$  cuando  $n_v \rightarrow \infty$ .

Bajo ciertas condiciones de regularidad del modelo paramétrico y del kernel se puede probar que el estimador es consistente y asintóticamente normal. La matriz de varianza del estimador no está dada por el resultado clásico de máxima verosimilitud, sino que debe ser corregida para tomar en cuenta el efecto del primer paso ya sea paramétrico o no. Esta corrección incluye la función de influencia de  $f$  sobre el estimador (Newey, 1994a, 1994b; Newey y McFadden, 1994). En esta etapa de esta investigación los errores estándar de los parámetros se presentan sin la corrección.

Si se utiliza estimación paramétrica de la densidad condicional  $f$ , entonces se necesita un supuesto para establecer la forma paramétrica. Para la aplicación que se presenta más adelante se asume lo siguiente:

SUPUESTO 4:  $\ln x_j^* | x_j \sim N(\theta_1 + \theta_2 \ln x_j, \sigma_j)$ .

El Supuesto 4 implica que la densidad condicional se obtiene estimando  $\theta_1$  y  $\theta_2$  mediante una regresión lineal de  $\ln x_j^*$  en  $\ln x_j$  y  $\sigma_j$  con la desviación estándar de los residuos de la regresión asumiendo que los residuos distribuyen normalmente.

## 5. SIMULACION DE MONTE CARLO

Para verificar la bondad del método propuesto se desarrolló una serie de simulaciones de Monte Carlo. Se aplicó el método paramétrico a una muestra simulada de 2000 individuos eligiendo de un conjunto de tres alternativas. Las funciones de utilidad dependen de dos variables,  $x_1$  y  $x_2$ , que se observan con error. Estas variables fueron generadas como una muestra de variables aleatorias uniformemente distribuidas tal que  $x_1 \sim U[0,10]$  y  $x_2 \sim U[0,5]$ . Las variables sin error,  $x_1^*$  y  $x_2^*$ , son generadas sumando un término de error distribuido normal estándar. Para las simulaciones se distingue, además, dos posibles errores sistemáticos que podrían aparecer en datos de elección de modo, ya que las variables de servicio observadas provienen de otros modelos, como los de asignación a la red. Los modelos usados en las simulaciones son los siguientes:

Modelo de error de medición 1:  $x_1^* = x_1 + \varepsilon_1$ ;  $x_2^* = x_2 + \varepsilon_2$

Modelo de error de medición 2:  $x_1^* = 1.2x_1 + \varepsilon_1$ ;  $x_2^* = 0.8x_2 + \varepsilon_2$

Modelo de error de medición 3:  $x_1^* = x_1^{0.7} + \varepsilon_1$ ;  $x_2^* = x_2 + \varepsilon_2$

Las elecciones (variable dependiente) se generaron como variables aleatorias multinomiales con probabilidades de éxito dadas por un modelo logit y utilidades calculadas con las variables independientes sin error. Las funciones de utilidad son  $U_m = \alpha_m + \beta_1 x_{1m}^* + \beta_2 x_{2m}^*$ , donde  $\alpha_m$  corresponde a la constante modal de la alternativa  $m$ . La Tabla 1 muestra el valor de los parámetros

usados y los resultados de 500 simulaciones de Monte Carlo. El método propuesto reduce el sesgo de los parámetros asociados a las variables medidas con error en las tres diferentes formas del error. Sin embargo, la varianza de los parámetros aumenta, lo que es esperable dado que se utiliza estimación en dos etapas. Las constantes específicas de las alternativas presentan sesgo del mismo orden en los dos modelos. Es interesante notar que el sesgo se reduce más significativamente cuando la relación entre las variables con error y las sin error incluye alguna distorsión sistemática, como en los modelos de error de medición 2 y 3. En el caso de las simulaciones la regresión lineal permite obtener una buena estimación de dicha relación, sin embargo en aplicaciones reales esto puede ser difícil de lograr. Es en este caso donde la estimación no-paramétrica de la densidad condicional es más valiosa.

Tabla 1. Resultados de la simulación de Monte Carlo: Media de los parámetros estimados, sesgo, error cuadrático medio (MSE) y varianza, para el modelo que corrige los errores de medición, el modelo sin corrección y el modelo con los datos reales.

	Modelo Error 1				Modelo Error 2				Modelo Error 3			
	Parámetro	Sesgo	MSE	Var	Parámetro	Sesgo	MSE	Var	Parámetro	Sesgo	MSE	Var
<b><math>\beta_1 = -0.2</math></b>												
Corregido	-0,2022	-0,0022	0,0002	0,0002	-0,2008	-0,0008	0,0001	0,0001	-0,2006	-0,0006	0,0007	0,0007
No corregido	-0,1844	0,0156	0,0004	0,0001	-0,2195	-0,0195	0,0005	0,0001	-0,0848	0,1152	0,0134	0,0001
Sin errores	-0,2009	-0,0009	0,0001	0,0001	-0,1995	0,0005	0,0001	0,0001	-0,2010	-0,0010	0,0004	0,0004
<b><math>\beta_2 = -0.5</math></b>												
Corregido	-0,5210	-0,0210	0,0014	0,0010	-0,5204	-0,0204	0,0020	0,0015	-0,5259	-0,0259	0,0018	0,0011
No corregido	-0,4591	0,0409	0,0022	0,0005	-0,3671	0,1329	0,0182	0,0005	-0,4610	0,0390	0,0021	0,0006
Sin errores	-0,5005	-0,0005	0,0005	0,0005	-0,4996	0,0004	0,0006	0,0006	-0,5008	-0,0008	0,0005	0,0005
<b><math>\alpha_1 = -0.5</math></b>												
Corregido	-0,5331	-0,0331	0,0082	0,0071	-0,5293	-0,0293	0,0070	0,0061	-0,5357	-0,0357	0,0073	0,0060
No corregido	-0,4616	0,0384	0,0062	0,0048	-0,4589	0,0411	0,0061	0,0044	-0,4635	0,0365	0,0056	0,0043
Sin errores	-0,5014	-0,0014	0,0053	0,0053	-0,4973	0,0027	0,0048	0,0048	-0,5007	-0,0007	0,0045	0,0045
<b><math>\alpha_2 = -0.5</math></b>												
Corregido	0,5418	0,0418	0,0063	0,0046	0,5407	0,0407	0,0065	0,0049	0,5502	0,0502	0,0076	0,0051
No corregido	0,4589	-0,0411	0,0049	0,0032	0,4559	-0,0441	0,0052	0,0032	0,4591	-0,0409	0,0050	0,0033
Sin errores	0,5005	0,0005	0,0036	0,0036	0,4981	-0,0019	0,0035	0,0035	0,5033	0,0033	0,0036	0,0036

## 6. APPLICACIÓN A DATOS DE SANTIAGO

En esta sección se presenta los resultados de la estimación de los modelos de elección de modo usando datos de Santiago. Se asume que los modos disponibles son cinco: auto, transporte público (bus y metro), taxi-colectivo, taxi y auto como pasajero. La función de utilidad es lineal en los atributos de los modos: costo, tiempo de viaje, tiempo de espera y tiempo de acceso. Se asume que el costo y el tiempo de acceso no tienen error de medición. Mientras que el tiempo de viaje y el de espera tienen error de medición ya que son medidos a nivel de zona. Para la función de utilidad se

incluye el costo de viaje dividido por el ingreso familiar per cápita de forma que el valor del tiempo del modelo corresponde a una fracción del ingreso.

En los modelos estimados se intenta distinguir por el propósito de los individuos ya que puede ser relevante si existe un efecto directo sobre la forma de financiar el viaje o elegir el modo de viaje. Esto es, si la decisión de viajar está tomada por otra persona que no es miembro del hogar del viajero (por ejemplo, el empleador). La información contenida en la encuesta no permite identificar estos casos. Sin embargo, se puede distinguir entre viajes por trabajo y el resto de los propósitos de viaje. Para esto se define una variable dummy que vale uno cuando el propósito del viaje es “por trabajo” y cero en otro caso. Así, se diferencia en la función de utilidad modal el costo de viaje de un viaje de trabajo o de ocio. La Tabla 2 presenta los resultados del modelo logit multinomial sin corrección de error de medición y con corrección (las constantes específicas por alternativa no se presentan).

Tabla 2. Estimación de modelo logit multinomial

Variable	Sin corrección		Con corrección paramétrica		Con corrección no paramétrica	
	Estimación	Test t	Estimación	Test t	Estimación	Test t
Costo	-0,0418	-13,88	-0,0356	-7,55	-0,0351	-11,61
Costo * Trabajo	-0,0009	-0,03	0,0835	4,83	0,0431	2,74
Tiempo de viaje	-0,0277	-15,34	-0,0077	-9,60	-0,0040	-3,89
Tiempo de espera	-0,0799	-22,24	0,0000	0,00	-0,0139	-4,23
Tiempo de acceso	0,0131	3,53	0,0258	9,87	0,0276	7,72
No. Observaciones	38.855		38.855		38.988	
Log-Likelihood	-23.823		-24.374		-24.507	
Valor del tiempo (\$/h)	7.918		3.427		1.856	

Los resultados de los modelos con corrección del error de medición muestran un valor del tiempo bajo respecto a los modelos sin corrección y más cercano a 1.480 (CLP/hora), el valor utilizado actualmente por el SNI.

Se observa que los valores del tiempo obtenidos con los modelos sin corrección del error en la medición del tiempo de viaje son significativamente más altos a los utilizados ahora por el Sistema Nacional de Inversiones. La diferencia se debe a que el valor social del SNI está determinado por el salario de los individuos más una componente que representa el valor del tiempo de viaje para los viajes de ocio, la que es una fracción del salario.

El valor del tiempo como disposición a pagar por ahorrar tiempo de viaje está compuesto por el valor del tiempo de ocio más la desutilidad de viajar. Desde este punto de vista, estos resultados indicarían que los usuarios perciben una gran pérdida de bienestar por viajar y que están dispuesto a pagar una alta suma por reducir su tiempo viajando. Sin embargo, hay que considerar que estos valores están calculados con el ingreso familiar per cápita de la muestra de estimación de \$264.405, que implica un salario por hora del orden de \$5.000 para un hogar de cuatro personas y un único trabajador. Esto indica que los usuarios estarían dispuestos a pagar más de dos veces su salario por

hora para ahorrarse una hora de viaje. Esto contradice la teoría si se asume que la desutilidad del tiempo dedicado a trabajar es mayor que la del tiempo dedicado a viajar. Además, si la disposición por ahorrar tiempo de viaje fuera tan alta como indican los resultados de la estimación, se observaría una mayor diferencia de tarifa entre modos rápidos (taxi colectivo) y lentos (bus).

Además se estimaron modelos con componente de error normal (Tabla 3) sin correlación. Estos modelos permiten estimar varianzas específicas por modo, de manera que las componentes aleatorias de la función de utilidad no son idénticamente distribuidas, aunque sí independientes. Los resultados indican que el modelo no corregido tiende a sobreestimar el valor subjetivo del tiempo, aunque ambos modelos entregan un valor muy alto en comparación con los modelos logit simple. Esto puede deberse a un problema de mala especificación del modelo o algunos problemas de convergencia detectados en la estimación, que no se discutirán en este trabajo, ya que el objetivo principal es mostrar la aplicabilidad del método propuesto.

Tabla 3. Estimación de modelo logit con componente de error normal

Variable	Sin corrección		Con corrección paramétrica	
	Estimación	Test t	Estimación	Test t
Costo	-0.0427	-7.55	-0.0377	-7.55
Costo * Trabajo	0.0785	4.83	0.1140	4.83
Tiempo de viaje	-0.0288	-9.60	-0.0231	-9.60
Tiempo de espera	-0.0871	0.00	0.0000	0.00
Tiempo de acceso	0.0138	9.87	0.0400	9.87
No. Observaciones	38.855		38.855	
Log-Likelihood	-23.783		-24.321	
Valor del tiempo (\$/h)	10.700		9.721	

## 7. CONCLUSIONES

En este trabajo se propone un método para estimar modelo de elección discreta usando variables agregadas, usualmente obtenidas de modelos de asignación a redes de transporte, que elimina la inconsistencia del estimador de máxima verosimilitud estándar cuando se usan variables con error de medición. El método se basa en usar los niveles de servicio reportados por los encuestados como muestra de validación.

Los resultados de las simulaciones de Monte Carlo muestran que el método reduce el sesgo en la estimación de los parámetros. La aplicación a datos de Santiago muestra que utilizar variables agregadas entrega resultados muy sesgados con valores poco consistentes con la teoría. En cambio, el uso del método de corrección, tanto paramétrica como no paramétrica, resulta en parámetros de la función de utilidad más consistentes con el rango de ingreso familiar en la muestra de estimación.

Finalmente, el método también es aplicable a modelo con parámetros aleatorios y entrega resultados en tiempos de estimación razonables (2 a 3 horas) si se aplica el método de corrección paramétrico.

## **AGRADECIMIENTOS**

Esta investigación fue financiada por CONICYT a través del Proyecto FONDECYT N ° 11160857.

## **REFERENCIAS**

- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242-262.
- Chen, X., Hong, H., y Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, 49(4), 901-937.
- Chen, X., Hong, H., & Tamer, E. (2005). Measurement error models with auxiliary data. *The Review of Economic Studies*, 72(2), 343-366.
- Hu, Y., & Ridder, G. (2012). Estimation of nonlinear models with mismeasured regressors using marginal information. *Journal of Applied Econometrics*, 27(3), 347-385.
- Jara-Díaz, S. y M. Farah (1987) Transport demand and users' benefits with fixed income: the good/leisure trade-off revisited. *Transportation Research Part B*, Vol. 21, No. 2.
- Newey, W. K. (1994a). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, 1349-1382.
- Newey, W. K. (1994b). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10(02), 1-21.
- Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4, 2111-2245.
- Train, K.E. (2002) *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Train, K.E. y D. McFadden (1978) The goods/leisure trade-off and disaggregate work trip mode choice models. *Transportation Research Vol. 12*, pp. 349-353.