

## CORRECCIÓN POR EVASIÓN DE UNA MATRIZ DE VIAJES EN TRANSPORTE PÚBLICO OBTENIDA A PARTIR DE DATOS PASIVOS

Luis-Angel Cantillo, Pontificia Universidad Católica de Chile  
lacantillo@uc.cl

Juan Carlos Muñoz, Pontificia Universidad Católica de Chile  
jcm@ing.puc.cl

Sebastián Raveau, Pontificia Universidad Católica de Chile  
sraveau@ing.puc.cl

Sebastián Tamblay, Pontificia Universidad Católica de Chile  
stm@ing.puc.cl

Paula Iglesias, Ingeniería DICTUC  
piglesias@dictuc.cl

### RESUMEN

Sistemas actuales de transporte público integrado emplean tarjetas electrónicas (*smartcards*) como forma de pago para hacer uso del sistema. En muchos casos, tal como ocurre en el sistema que opera en la ciudad de Santiago, Transantiago, la información recopilada de dichas tarjetas es la empleada para la elaboración de la matriz Origen-Destino con la cual se apoya la labor de planificación del sistema. En consecuencia, por construcción, en estas matrices no están incluidos todos aquellos viajes o etapas de viajes evadidos por los usuarios (no validados). Esto, pudiera generar importantes distorsiones en la estimación de los viajes, en especial en sistemas con altas tasas de evasión. El presente trabajo expone una metodología de corrección de la matriz de viajes en transporte público por evasión, partiendo de una tasa de evasión estimada a nivel de paradero – servicio – sentido; para luego aplicar procedimientos de corrección según tipo de evasión, utilizando información de matrices origen-destino y transacciones de *smartcards*. Finalmente, se incluye un caso de estudio aplicado a Transantiago, cuya tasa de evasión a 2017 se estima superior al 30%.

**Palabras claves:** *Transporte público, matriz origen-destino, evasión.*

### 1. INTRODUCCIÓN

El transporte público se ha convertido en una de las áreas de mayor aplicación de las tecnologías de la información. Al tratarse de sistemas masivos y, cada día más automatizados, que incluyen *Automated Fare Collection* (AFC) y *Automatic Vehicle Location* (AVL), es posible recopilar constantemente gran cantidad de datos pasivos sobre los viajes. Esto sucede en muchos sistemas de transporte en el mundo, que han incorporado nuevas tecnologías a la operación.

Una de dichas nuevas tecnologías consiste en la utilización de AFC, por medio de tarjetas electrónicas empleadas como forma de pago, las que deben ser empleadas para abordar un bus, o acceder a estaciones de buses o metro, y permiten la integración tarifaria. Dichas tarjetas pueden ser identificables y rastreables cada vez que son utilizadas; de esta forma se tiene la

posibilidad de recopilar de forma automática, información sobre el comportamiento de los usuarios del sistema.

Conocer el comportamiento y preferencias de viaje de los usuarios, con el cual elaborar una matriz origen-destino confiable, es un elemento fundamental de una adecuada planificación de un sistema de transporte (Ortuzar & Willumsen, 2011). En este sentido, las tarjetas con tecnología AFC son una gran oportunidad para “rastrear” a los usuarios, y construir con ello la deseada información de viajes a nivel de volumen según pares O-D. De hecho, en la literatura se encuentran propuestas metodológicas, como la desarrollada por Munizaga & Palma (2012) (refinada en Devillaine et al., 2012; y Munizaga et al., 2014), para estimar el destino del viaje, empleando la información AFC y AVL de los buses. De esta forma, es posible estructurar la matriz O-D de viajes validados en un sistema de transporte público que emplee esta tecnología, pero: ¿Qué sucede con todos los viajes no validados?

En diferentes sistemas de transporte masivo a lo largo del mundo, por su diseño, estructura u operación, existe la posibilidad de que algunos pasajeros evadan el pago de la tarifa, al ingresar al sistema sin validar su tarjeta, o que su viaje no quede adecuadamente registrado al no validar en algunas etapas del mismo. Por lo anterior, dichos viajes, o etapas de viaje, no estarán incluidos en la matriz O-D con la que se planifica el sistema, si esta se construye exclusivamente con la información automática. La situación es particularmente preocupante en casos latinoamericanos, donde sistemas como Transantiago (Santiago de Chile) y Transmilenio (Bogotá), presentan cifras preocupantes de evasión. En Transantiago, la tasa de evasión (etapas de viaje no validadas) fue de 31,4% durante el primer trimestre de 2017 (Programa Nacional de Fiscalización, 2017), mientras que en Transmilenio la tasa de evasión bordea el rango 10-15% (Revista Semana, 2016). Las cifras anteriores evidencian que las metodologías de estimación de matriz O-D a partir de datos pasivos pueden estar dejando por fuera una gran cantidad de etapas de viajes; lo cual repercute en una subestimación de los viajes y una planificación distorsionada y alejada de la realidad.

Debido a la estructura integrada del sistema, se pueden presentar dos tipos claros de evasión, con efectos distintos en la construcción de la matriz O-D y en el financiamiento del sistema, y que forman parte de la evasión “global” que se observa en el sistema. Primero, está el caso de los viajes en transporte público en que el usuario no valida en ninguna de las etapas, que llamaremos “evasión completa” y, por otro lado, el de aquellos viajes en que los usuarios validan en al menos una de sus etapas de viaje, pero no lo hacen en el resto, lo que entenderemos como “evasión parcial”.

En esta investigación, se propone busca hacerse cargo de estos efectos, mediante una metodología de corrección de una matriz de viajes elaborada a partir de datos pasivos, que estime los viajes (o etapas de viajes) no registrados, asociados a la evasión<sup>1</sup>. La metodología que se propone consta de tres fases. En primer lugar, un modelo econométrico de estimación

---

<sup>1</sup> Entenderemos por “Evasión” al fenómeno de no validar en general, más que específicamente al hecho de evitar pagar la tarifa, debido a que, para todos los efectos de este análisis y, en general, de los indicadores que se elaboran al respecto, no resulta necesario ni factible distinguir entre “no validaciones” que habrían tenido costo cero (por tratarse de etapas posteriores o anteriores a una validación paga, dentro de un viaje) y no validaciones con las que, efectivamente, se está evitando el pago del viaje.

de la tasa de evasión a nivel de paradero – servicio – sentido. Con esta información, se corrigen los dos tipos de evasión que se identifican relevantes, la evasión parcial y la evasión completa, que son respectivamente las fases 2 y 3 de la metodología.

Con esta fase, se puede conocer la tasa de evasión en el punto de origen de la etapa de viaje, a ser distribuidos entre los posibles destinos, y de forma agregada se conoce el total de etapas evadidas. Como segunda fase, la corrección de la evasión parcial, y finalmente la corrección de la evasión completa.

A continuación, en el capítulo 2 se revisan antecedentes a la metodología propuesta, que se encontraron en la literatura. En el capítulo 3 se describe cada una de las etapas de la metodología que se propone, para la corrección de la matriz. El capítulo 4 presenta el caso de estudio aplicado a Santiago de Chile, y su sistema de transporte público, Transantiago. Finalmente, en el capítulo 5 se exponen las principales conclusiones de la investigación, junto con las limitantes de la metodología y futuras líneas de investigación.

## 2. ANTECEDENTES METODOLOGICOS

Dentro de la literatura revisada se observa consenso respecto a que existen diferentes formas de evasión, que generan distintos tipos de distorsión en la matriz OD (Gallegos et al., 2016). Por lo anterior, se requiere empezar por hacer una revisión de los perfiles de evasores y tipos de evasión relevantes de considerar, para luego analizar las metodologías con las cuales se ha abordado cada uno de estos tipos.

### 2.1. Tipos de evasión

El Ministerio de Transportes y Telecomunicaciones de Chile (MTT) desarrolló en 2010 un informe en que identificó diversos tipos de evasores, y caracterizó cada perfil (Factor Estratégico, 2010). Este mismo estudio, define una tipología de evasores, basados en los mecanismos y motivaciones de evasión. Como conclusión se proponen dos grandes categorías de evasor. El primer tipo es el evasor “circunstancial”, cuya característica es que sólo evaden bajo ciertas situaciones que les impide validar, por ejemplo falta de saldo en la *smartcard* o bus muy lleno que les impide acercarse a validar. El segundo tipo son los evasores “crónicos o duros”, aquellos que sistemáticamente evaden el pago de la tarifa.

Otra investigación que propone una categorización de los evasores es Delbosc & Currie (2016), quienes a partir de un trabajo desarrollado en Melbourne (Australia), clasifican a los evasores en “deliberados” e “involuntarios”. Los primeros se caracterizan por ser reincidentes y presentar una menor aversión, mientras que los segundos son poco reincidentes y suelen evadir bajo ciertas circunstancias de negligencia personal o del sistema, especialmente cuando la forma de pago les es complicada.

En el estudio australiano, podemos identificar una clara relación con la categorización obtenida en Chile, donde los evasores “deliberados” equivalen a los “crónicos”, y los “involuntarios” a

los “circunstanciales”. Con ello tenemos, dos tipos distintos de evasores con comportamiento y motivaciones diferentes.

Por otro lado, debido a la estructura de los sistemas de transporte público integrados, es posible tener viajes con diferente número de etapas, entre las cuales se hace transbordo, ya sea con tarifa integrada o sin ella. Tal es el caso actual de muchas ciudades en el mundo. Al cruzar las categorías de evasores con los tipos de viaje, se puede llegar a una clasificación de “tipos de evasión”, dependiendo el tipo de evasor y del tipo de viaje que realiza, tal como se muestra en la Tabla 1.

**Tabla 1** Tipos de evasión según categoría de evasor y viaje.

Tipo de viajes	Comportamiento del usuario	
	Evasor duro	Evasor circunstancial
<b>1 etapa</b>	Evasión Completa	Evasión Completa
<b>2 o más etapas</b>	Evasión Completa	Evasión Completa ó Parcial

Fuente: Elaboración propia.

Un evasor duro o deliberado, se entenderá como aquel que evade todas las etapas de su viaje, independiente de cuantas sean o de las circunstancias. Por su parte, el evasor circunstancial ó involuntario, sólo recurrirá a evadir cuando las condiciones de pago sean adversas, pudiendo dejar sin validar una o más etapas de su viaje, o validarlo íntegramente. De esta forma, podemos clasificar como “Evasión Completa” aquella situación en que ninguna de las etapas del viaje es validada por el usuario, lo que ocurrirá para todos los viajes asociados a evasores duros o deliberados, y también se observará en algunos de los viajes asociados a evasores circunstanciales o involuntarios. Por su parte, se pudiera observar “Evasión parcial” cuando evasores circunstanciales o involuntarios no validen en algunas etapas de su viaje, mientras que sí lo hacen en las restantes. El evasor circunstancial pudiera, también, generar viajes completamente validados, ya sean de una o más etapas.

Cada tipo de evasión, es decir, evasión completa o evasión parcial, implica un manejo diferente en la metodología de corrección planteada en este estudio. Antes de entrar en esta materia, es necesario revisar lo que entenderemos como evasión “global”, que identificaremos como la reunión de todos los tipos de evasión recién identificados.

## 2.2. Estimación de evasión global

Como punto de partida se requiere contar con una adecuada cuantificación del fenómeno de la evasión a lo largo de la ciudad. Por supuesto, medir la evasión en forma detallada en una ciudad de gran tamaño es sumamente costoso, y, entonces, disponer de indicadores puntuales, para distintos momentos del tiempo y con regular actualización, no suele resultar factible. Lo que se requiere, entonces, es contar con adecuadas estimaciones del fenómeno; para ello se busca explicar la evasión observada, inicialmente a partir de variables socioeconómicas y operaciones, y en lo posible no contempladas en modelos anteriormente planteados en otros estudios.

Al revisar la literatura correspondiente, se encuentra que los niveles de evasión en el transporte público han sido explicados por diversos factores, como el nivel de servicio del sistema, las tasas de inspección y multas, el periodo del día y el sistema de pago, entre otros (Torres-Montoya, 2013). Además, Lee (2011) constató diferencias significativas en los niveles de evasión a lo largo del día, y se ha detectado también una fuerte correlación entre los niveles de inspección y la evasión (Clarke et al., 2010). Recientemente, Guarda et al. (2015) estimaron modelos de regresión de conteo para Santiago de Chile, utilizando una base de datos del Programa Nacional de Fiscalización de Transportes, y explican el número de evasores que ingresan por una puerta del bus en un determinado paradero. Los resultados indican que la evasión aumenta cuando; i) más gente entra o sale por la puerta; ii) la entrada de pasajeros es por las puertas traseras; iii) los buses están más ocupados y el intervalo entre pasadas es mayor, iv) los buses tienen más puertas; y v) el nivel de ingresos de la comuna en donde está ubicado el paradero disminuye.

Referente al tipo de modelo, se ha probado en la literatura el uso de regresión logística para estimar la probabilidad de evadir (Barabino *et al.*, 2015), donde se buscó explicar una variable *dummy* de propensión a evadir, a partir de variables sociodemográficas categóricas. Otra alternativa es el empleo de modelos de conteo (Guarda et al., 2015), que abarcan modelos de regresión para trabajar con procesos de este tipo, como el caso de conteo de personas. Para implementar esta alternativa la variable dependiente pasaría a ser “número de evasores”, en lugar de considerar la “tasa de evasión”. Con esta alternativa se encontró directa relación entre el número de pasajeros que aborda el bus y el número de personas que evade, con lo cual esta variable se lleva la mayor parte de la capacidad explicativa del modelo reduciendo el efecto de otras variables relevantes que afectan la disposición a evadir; y si por el contrario se excluye de los modelos la variable “personas que abordan”, se pierde significativamente capacidad predictiva. Por lo anterior, se prefirió intentar con otras formas funcionales, que evitaran este problema. Esta misma investigación, sugiere probar modelos de regresión logística para la estimación de la tasa de evasión, ya que permite limitar la variable dependiente al rango de valores entre 0 y 1, como efectivamente sucede con la evasión.

Por otro lado, investigaciones como Buneder (2016) y Troncoso & De Grange (2017), han abordado el tema utilizando series de tiempo, pero para ello se requiere información con diferentes cortes temporales; lo cual, limita las variables a incluir, por no contar con la respectiva información. Al respecto, cabe señalar que el objetivo de estos estudios es diferente, ya que mientras estas investigaciones buscan identificar como cambian las tasas globales de evasión en el tiempo, y que factores la afectan, la presente busca predecir cómo se reparte la tasa global de evasión entre los diferentes paraderos-servicio-sentido-periodo, en aras de elaborar un método de corrección de la matriz de viajes que considere esa estructura.

### 2.3. Corrección por tipo de evasión

La principal referencia dentro de la literatura al respecto es la investigación de Gallegos *et al.* (2016), quienes desarrollaron una metodología de corrección de matriz de viajes por evasión, partiendo precisamente del cruce de bases de datos de encuestas O-D y datos pasivos de *smartcards*. Dicha investigación se enfocó en el caso de estudio de Transantiago, para el cual propusieron dos tipos de corrección.

- Una corrección para usuarios que realizan combinación bus – metro, y evaden la primera etapa de su viaje, es decir, la etapa en bus, pero pagan al ingresar al metro, por cual están realizando una no validación parcial.

El procedimiento que proponen, compara la matriz O-D de validaciones de Munizaga & Palma (2012), que únicamente incluye las etapas de viajes que se pagan (validan); con la encuesta O-D de metro 2013 (Ipsos, 2013), en donde los encuestados declaran todas sus etapas de viaje, incluso sin haberlas pagado.

Bajo el supuesto, de que muchos usuarios evaden en la etapa en bus, previa al metro, por falta de saldo en la tarjeta, proponen el cruce de ambas bases de datos, para llevar a cabo la calibración de un modelo biproporcional, que defina los factores de corrección, de los viajes entre cada zona y las estaciones de metro.

Se destaca el aporte de esta aproximación, por darle el mejor uso posible a la información disponible, en casos en que esta es limitada. Por lo cual es punto de referencia para análisis similares, en sistemas de transporte con escasa información adicional a la recopilada en las *smartcards*.

- La segunda corrección implementada en Gallegos et al (2016), es el caso de los usuarios que realizan viajes únicamente en bus, es decir, sin trasbordar a metro. Para estos casos, y debido a la falta de otra fuente de información, asumieron todos se comportan como evasores duros, por lo cual en viajes únicamente en bus, sólo existe evasión completa. De esta forma, aplican una corrección de etapas de viaje, de modo que iterativamente logren consistencia entre las tasas de evasión medidas por paradero, y las etapas corregidas con la metodología.

La metodología, desarrollada en este estudio, es la principal referencia y punto base para las propuestas de corrección que proponemos en la actual investigación.

### 3. PROPUESTA METODOLÓGICA

La metodología propuesta en la investigación se divide en tres grandes fases, iniciando con la estimación de la evasión a nivel de paradero – servicio – sentido – periodo, para luego conformar la matriz O-D de evasores totales, y finalmente realizar la corrección por evasores parciales.

#### 3.1. Estimación de tasa de evasión

De forma resumida, la tasa de evasión es el número de personas que no validan, entre el total de personas que ingresan, ya sea a nivel de bus, servicio, paradero ó zona; y la metodología aplicada para esta etapa de la investigación fue la siguiente; conformar una base de datos, establecer la muestra de calibración y con ella calibrar modelos de estimación de evasión y validar los modelos. Como se indicó en la revisión de la literatura, se consideró adecuado, la utilización de un modelo de regresión logística para la estimación de la tasa global de evasión.

La regresión logística corresponde a una técnica estadística que ha sido ampliamente utilizada en la práctica (Hosmer & Lemeshow, 2000). Este tipo de modelos estadísticos busca encontrar la probabilidad continua de ocurrencia de un evento en base a la utilización de regresores. El modelo calcula una función logística (1) que estima la probabilidad de ocurrencia de un fenómeno ( $P_i$ ), en relación con la dependencia de que dicho fenómeno no ocurra  $1-P(i)$ , para ello se utilizan como datos de entrada los vectores de variables dependientes  $x_i$ , con los cuales se ha de calibrar los coeficientes  $\beta$  (2).

$$P_i = \frac{e^{\beta X_i}}{1 + e^{\beta X_i}} = \text{Tasa de Evasión} \quad (1)$$

El proceso de calibración de los parámetros ( $\beta$ ) se realiza por medio del método de máxima verosimilitud (Eliason, 1993), el cual maximiza la probabilidad estimada de obtener los resultados observados. En resumen, se maximiza la función (2), donde  $NVi$  son el número de individuos que evadió en la observación  $i$ , y  $Vi$  el número de individuos que validó en dicha observación.

$$\max_{\{\beta\}} L = \prod_i (P_i(\beta))^{NVi} (1 - P_i(\beta))^{Vi}$$

### 3.2. Corrección por evasión parcial

Para el caso de evasión parcial, la corrección que aplica la metodología, toma como punto de partida la propuesta de Gallegos *et al* (2016), aplicando una serie de modificaciones:

- La corrección se aplicará a los usuarios que combinan bus y metro, también en la etapa final, es decir, en el viaje metro – bus. Esto se justifica, debido a la observación del sistema, donde se identificó a usuarios que aún luego de pagar en metro, no validan en bus, principalmente por dos razones; el bus está muy lleno y no se les permite acercarse al validador, o desconocen que se requiera validar nuevamente, cuando ya validaron previamente en el sistema.
- Se trabajará con la tasa de evasión global a nivel de paradero – servicio – sentido – periodo, estimada con los modelos propuestos, y no con la tasa medida por DICTUC (2012), que fue la fuente de calibración de los modelos.
- En la etapa de convertir la matriz de validaciones de la base de datos pasivos de las *smartcards*, a una matriz zonal que pueda compararse con la información de la Encuesta Origen- Destino, se incluyó el desarrollo llevado a cabo por Tamblay *et al* (2016), en el cual se estima la zona de origen de los usuarios que llegan a un paradero determinado, en base a variables socioeconómicas de las zonas aledañas al paradero; en asumir, que los usuarios de una zona son únicamente lo de los paraderos que encuentran dentro de ella.

### 3.3. Corrección por evasión total

Una vez que se aplicaron los algoritmos enfocados a corregir la evasión parcial, corresponde proponer una metodología para inferir orígenes y destinos para el resto de viajes evadidos (i.e.,

evasión completa). El método propuesto busca la elaboración de una matriz O-D de evasores, de forma que sea consistente con la evasión por arco, definiéndose arco como todo par de paraderos en el sistema. Para ello considera que se tenga en cuenta:

- i) Tasas de evasión completa a nivel de paradero-servicio-sentido-período. Esta puede calcularse como la diferencia entre la tasa de evasión global y parcial, obtenida según los algoritmos ya presentadas.
- ii) Información de viajes pagos que abordan en cada paradero-servicio-sentido-período.
- iii) Distribución de paraderos de bajada de las etapas de usuarios que validan, a nivel de paradero-servicio-sentido-período. Esta puede obtenerse, por ejemplo, desde información como la procesada por Munizaga y Palma (2012) (refinada en Devillaine et al., 2012; y Munizaga et al., 2014). Además, se supondrá que la distribución de paraderos de bajada de las etapas de evasores totales coincide con la de los usuarios que pagan pero que no usan metro en ningún momento de su viaje y abordan en ese mismo paradero-servicio-sentido-período.
- iv) Modelo de elección de rutas, que permite asignar las matrices alternativas de evasores sobre una red modificada de transporte público (que no incluya las estaciones de Metro, donde la evasión es despreciable). En este punto es posible aprovechar desarrollos disponibles mercado; un caso particular para la aplicación a Santiago de Chile, es la herramienta STEPTRANS (Raveau et al, 2017).

Con los puntos (i), (ii) y (iii) pueden estimarse niveles de flujo de etapas de viajes completos evadidos a nivel de arcos de la red. Luego, incorporando el punto (iv), se buscarían las matrices O-D más consistentes con los flujos sobre los arcos.

En otras palabras, se buscaría la matriz O-D que, al asignarse con los modelos de elección de rutas, predice de mejor manera, los flujos sobre los arcos, de las etapas evadidas. Cabe destacar que, para que esta metodología prediga matrices O-D con una estructura de viajes consistente con los patrones generales de la ciudad (y no genere, por ejemplo, sólo viajes cortos en que cada uno coincida con una etapa de un evasor completo), se incluirían adicionalmente medidas de atraktividad de las distintas zonas con las que se buscaría ajustar. Estas podrían obtenerse desde la Encuesta Origen Destino 2012 para Santiago (SECTRA, 2015), que contiene información tanto de viajes pagos como evadidos.

#### 4. CASO DE ESTUDIO

La metodología propuesta en esta investigación, se aplicó al caso de estudio de Transantiago, el sistema de transporte de la ciudad de Santiago (Chile). A continuación, se detallan los resultados obtenidos de la aplicación.

##### 4.1. Generación de la base de datos



Se trabajó con la base de datos elaborada para el Directorio de Transporte Público Metropolitano (DICTUC, 2012), en la cual se hizo medición de evasión al abordar los buses de Transantiago, la cual destaca por ser la más completa por tener mediciones de personas que validan y no validan a nivel de bus – paradero – servicio – sentido – periodo, para los siete operadores de Transantiago; por lo anterior, se decidió emplear esta base de datos para la calibración de los modelos de estimación de evasión.

La base de datos incluye mediciones en diferentes periodos del día durante los 5 días hábiles de la semana, con mediciones a bordo del bus, registrando los ascensos y descensos, identificando si validaban o no. La base de datos la conforman 573.249 registros, distribuidos en 143.312 registro en punta mañana (hora de fin de viaje entre 8:00-9:00), 166.242 registros en punta tarde (hora inicio de viaje entre 18:00-19:00) y los 263.695 restantes en periodo fuera de punta (hora media de viaje entre 10:00-12:00). Por otro lado, la distribución entre operadores indica que el operador con menos registro es unidad de negocio (UN) 6 (9%), seguido por UN7 (11%), UN4 (13%), UN5 (15%), UN3 (15%), UN2 (18%) y UN1 con 19% de la muestra.

Considerando que los modelos a estimar se tratarán a nivel de paradero – servicio – sentido – periodo, fue necesario agrupar los registros, ya que estaban al detalle de bus ara trabajar al nivel de agregación que, posteriormente, se utilizaría en los modelos. Por lo anterior, la base de datos se redujo a 17.300 registros para el periodo punta mañana. Luego de realizar una depuración de los datos se llegó a una base de información de 16.467 registros para el mismo periodo. De estos registros, el 90% se utilizó para la calibración del modelo, y el 10% restante se mantuvo como muestra de validación, dicha muestra se seleccionó de forma aleatoria, verificando que existiera suficiente representatividad en cuando a operadores y comunas.

Análogamente, en el caso del periodo fuera de punta, se preparó una base de datos de calibración que contó con 21.094 registros, habiendo apartado una muestra de validación que constó de un total de 2.106 registros. Finalmente, en el periodo de punta tarde se utilizaron 17.350 registros en la calibración y se dejaron 3.317 registros para la muestra de validación. Es importante indicar que las observaciones de paraderos se les asignó un “peso” que corrige la representatividad en relación a la cantidad de pasajeros del mismo.

## **4.2. Formulación de los modelos**

Se definió estimar 3 modelos, uno para cada uno de los periodos del día a analizar. Esta decisión se debe a la hipótesis previa, que el comportamiento asociado a viajes obligatorios, es decir, con motivo trabajo o estudio (realizados en periodos punta); puede ser distinto al observado en viajes con otros propósitos que suelen realizarse en periodos fuera de punta.

de que ciertas variables evaluadas en los modelos, podían llegar a tener efecto diferente (signo diferente) en la evasión, dependiendo el periodo del día; como efectivamente sucedió con algunas de las variables incluidas.

### **4.2.1. Variable dependiente:**

Para los modelos de regresión logística estimados, se definió la variable dependiente como la tasa de evasión del paradero, calculada según la siguiente ecuación:

$$TasaEvasión_{ijkl} = \frac{N^{\circ}evasores_{ijkl}}{N^{\circ}evasores_{ijkl} + N^{\circ}noevasores_{ijkl}}$$

$$\begin{aligned} \forall i = 1, \dots, I & \quad I = \text{Paraderos} \\ \forall j = 1, \dots, J & \quad J = \text{Servicios} \\ \forall k = 1, \dots, K & \quad K = \text{Sentidos} \\ \forall l = 1, \dots, L & \quad L = \text{Periodos} \end{aligned}$$

#### 4.2.2. Variables explicativas:

Las variables explicativas empleadas en la formulación del modelo pueden clasificarse en cuatro categorías: socioeconómicas, de accesibilidad, operacionales y un identificador de la empresa operadora del servicio. Al respecto, cabe señalar que se buscó utilizar variables respecto a las cuales se disponga de información a nivel poblacional zonal, pues el objetivo final de contar con modelos de estimación de evasión es aportar información necesaria para el procedimiento de corrección de la matriz de viajes, que ha de ser diseñado para incluir los viajes evadidos. A continuación, se realiza una descripción de las variables evaluadas para ser utilizadas en los modelos (no todas permanecen en los modelos definitivos presentados en este informe para los tres períodos analizados).

Entre las variables socioeconómicas evaluadas está el **costo del metro cuadrado** (para diferentes usos de suelo), **porcentaje de metros cuadrados de uso de suelo** (para diferentes usos de suelo), **Índice de prioridad social**, que es un indicador de calidad de vida a nivel de comuna que incluye dimensiones de ingresos, educación y salud; y el **Ingreso promedio per cápita** a nivel de zonas.

Dentro de la categoría de variables de accesibilidad se evaluaron, el **acceso a puntos de recarga y acceso a metro** por tratarse de un modo de transporte que a priori mantiene mayor control y por ende menores índices de evasión.

Entre las variables operacionales se incluyeron la **tasa de ocupación** de los buses, **número de puertas promedio** del servicio – sentido – periodo, el **porcentaje de viajes de transbordo** del paradero, **frecuencia** del servicio – sentido – periodo, **Índice de Cumplimiento de Frecuencia (ICF)** que es el porcentaje de cumplimiento de las frecuencias ofrecidas vs la reportada en el programa de operación, **Índice de Cumplimiento de la Regularidad (ICR)** y finalmente se evaluó **frecuencia\*ICF**.

Se probó con variables asociadas a tipologías de servicio, por un lado, se clasificaron los servicios en troncales, alimentadores y tronco-alimentadores; por otro lado, se categorizaron según longitud del recorrido ó según la dirección de los recorridos.

Por último, se incluyeron en los modelos evaluados variables categóricas por operador, para los casos de sistemas con múltiples empresas operadoras.

#### 4.2.3. Estimación y validación de los modelos

El conjunto de variables explicativas presentes en el mejor de los modelos estimado, para cada uno de los períodos de análisis, se presenta en la Tabla 2. Dichas variables fueron

seleccionadas para su inclusión, siguiendo los criterios de consistencia de signos y significancia estadística de los parámetros estimados (Ortúzar y Willumsen, 2011). En la tabla se presenta el valor de coeficiente y test t para cada una de las variables, así como también el valor de la log-verosimilitud del modelo de cada periodo.

Tabla 2 Estimación de los modelos de regresión logística.

Variable	Punta Mañana		Fuera de Punta		Punta Tarde	
	Coef.	Test-T	Coef.	Test-T	Coef.	Test-T
<b>Socioeconómicas</b>						
Índice de Prioridad Social (IPS)	0,0101	6,17	0,0116	6,15	0,0082	3,65
Ingreso Promedio Zonal Persona (M\$CLP)	-0,0010	-5,12	-0,0015	-7,08	-0,0012	-4,70
<b>De ubicación y accesibilidad</b>						
Punto de recarga Bip a 200 metros	-0,0782	-5,10	-0,0718	-2,39	-0,0538	-1,44
Estación de metro a 200 metros	-0,2768	-4,93	-0,3929	-6,18	-0,5763	-7,30
Frecuencia del paradero (veh/h)	-0,0030	-4,67	-0,0031	-3,48	-0,0032	-2,34
<b>Operacionales</b>						
Tasa de ocupación (%)	1,3834	17,31	0,8309	5,79	1,0783	8,96
Frecuencia Corregida (Veh/h)	-0,0077	-2,48	0,0204	4,10	-0,0063	-1,87
N° de puertas promedio	0,2444	6,84	0,0282	0,82	0,2277	5,07
Porcentaje de viajes de transbordo (%)	0,3511	4,96	0,2490	3,23	-0,2673	-2,88
<b>Operadores</b>						
UN1	0,4564	10,36	0,4875	8,51	-0,0219	-0,31
UN3	0,4131	8,57	0,5567	9,61	-0,1932	-2,88
UN4	0,2879	5,60	0,5548	9,72	-0,2874	-4,02
UN5	0,2749	6,10	0,3789	7,48	-0,2894	-4,20
UN6	0,3979	5,61	0,5571	7,37	-0,1575	-1,92
UN7	0,6585	12,27	0,5949	9,01	-0,1612	-1,75
Constante	-2,8576	-15,76	-2,1408	-11,70	-1,6187	-7,25
<b>Log-verosimilitud</b>	<b>-8891,94</b>		<b>-72116,87</b>		<b>-131025,87</b>	

Fuente: Elaboración propia.

En la tabla se comprueba la capacidad explicativa de las variables del modelo, utilizando como medida el test t-student (para todos los casos su valor es superior a 1,645, valor referencial para un 90% de confianza). Es importante aclarar que, aunque se presenta variables que no son significativas para un periodo en particular, se decidió mantenerlas, ya que son muy significativas en los otros periodos, y el mantener las mismas variables en los 3 periodos permite una mejor comparación entre ellos, y permite revisar cómo cambia el efecto que produce cada una de las variables según el periodo del día que se analice.

Referente a los signos, se encuentra que:

- Las comunas con mejores condiciones económicas, de educación y salud, presentan un IPS menor, por lo cual se considera correcto el signo positivo del coeficiente de esta variable, ya que se espera que personas con mejores condiciones económicas y mejor calidad de vida tiendan a evadir menos.

- Se cumple el supuesto inicial de que en zonas con mayor ingreso económico la tasa de evasión es inferior. Lo anterior es consecuente con el signo negativo de dicha variable.
- El tener un punto de recarga cerca permite reducir la evasión por falta de saldos en la tarjeta BIP, lo que concuerda con el signo negativo que se observa para este coeficiente. Adicionalmente, se encuentra que tiene mayor peso y significancia en los periodos punta mañana que en punta tarde, lo cual es comprensible ya que la situación de falta de saldo es más limitante en el viaje de la mañana que en el de la tarde, cuando pudiera haberse pasado por puntos de recarga a lo largo del día.
- Los paraderos con estación de metro cercana tienen mayor probabilidad de que sus pasajeros provengan del metro (transbordo), en donde, al menos esa parte de los usuarios se habría visto obligado a pagar, por lo cual se entiende que se presente menor evasión en dichos paraderos.
- Paraderos con menor “frecuencia total” tienden a encontrarse menos conectados y por ende el servicio ofrecido es inferior, y parece que esto está relacionado con una mayor evasión.
- Una mayor tasa de ocupación en los buses dificulta a las personas acercarse al lector de la tarjeta BIP con lo cual en ocasiones evaden incluso sin involuntariamente. Adicionalmente, se puede pensar que existe menor presión social al evadir, cuando dicha acción se hace en grupo. Por esta razón se considera adecuado el signo positivo obtenido para esta variable y su gran significancia.
- Similar al caso de “frecuencia de paradero”, en “frecuencia corregida”, se esperaba a priori que servicios con mayor frecuencia y mejor Índice de Cumplimiento de Frecuencia (ICF), ofrezcan un mejor servicio, que aumente la conformidad del usuario con el sistema y lo desincentive a evadir la tarifa; sin embargo, llama la atención que para el periodo fuera de punta el signo cambie respecto a los otros dos periodos. Puede deberse en parte a que durante este periodo el sistema de metro no se encuentra a capacidad, con lo cual usuarios de buses que siempre pagan la tarifa se cambian de modo, mientras que usuarios que buscan evadir emplean los servicios de buses que le compiten al metro, que suelen ser aquellos con mayor frecuencia. Esto es sólo una posible razón para el comportamiento observado, que no se puede confirmar sólo a partir de estos resultados.
- Entre mayor número de puertas tenga un bus, más cantidad de puntos de acceso factible tiene un evasor, en especial, en caso de que las puertas traseras se abran para dejar algún pasajero en el paradero. Esto se refleja en el signo positivo obtenido para esta variable, y concuerda con lo encontrado en la literatura.
- Los coeficientes de las variables de operadores son difíciles de analizar, ya que no es posible relacionarlo a características socioeconómicas de las zonas donde operan,

pues por la estructuración del sistema, cada operador recorre muchas comunas en simultaneado con otros operadores, adicionalmente, el efecto de las condiciones socioeconómicas de la comuna está recogido en la variable de **Índice de Prioridad Social (IPS)**. Por otro lado, tampoco es posible asumir relación con la calidad del servicio ofrecido por el operador, ya que este efecto es recogido en parte por la variable **Frecuencia Corregida**, que incluye el indicador de cumplimiento de frecuencia (ICF) como medidor de la calidad de servicio ofrecido. Finalmente, el análisis que se le puede dar a estas variables de operadores es reflejar un poco las políticas implementadas por unos u otros para controlar la evasión, como el caso de instalación de torniquetes, operativos de fiscalización, y otros; de los cuales lamentablemente no se cuenta con información completa, pero es uno de los temas a continuar investigando.

## 5. CONCLUSIONES

En conclusión, se presenta una metodología que permite la corrección de matrices de viajes, utilizando información de *smartcards* y encuestas origen destino. La metodología propone la estimación de un modelo econométrico para estimar tasa de evasión a nivel de paradero-servicio-sentido, a partir de variables de fácil obtención y actualización de la información, pero que se limita a efectos del entorno y la operación del servicio, sin incluir razones intrínsecas al usuario, que sería interesante abordar en futuros desarrollos. En el caso de aplicación, se encontró significancia estadística a variables socioeconómicas, operacionales y de accesibilidad, que pueden ser probadas en otras bases de datos.

Posterior, se presentaron las etapas de corrección de la evasión parcial y evasión completa, de forma independiente. Para estas etapas se mostró la metodología de forma conceptual, en las que actualmente se está trabajando en la aplicación al caso de Santiago de Chile, a partir de las cuales lograr mayores conclusiones.

Finalmente, es importante destacar los efectos en política pública de la presente investigación. El acercar la matriz origen-destino un poco más a la realidad, permite mejoras en la planificación del sistema de transporte público en la ciudad de Santiago de Chile. De igual forma, el estudio permite dimensionar el fenómeno de la evasión y sus efectos.

### 5.1. Limitaciones de la metodología.

Entre las limitaciones de la metodología expuesta, se encuentra principalmente, la no inclusión de los evasores parciales “sólo bus”, es decir, que no realizan combinación con metro u otro sistema que permita asegurar la no evasión en esa etapa de viaje. Al revisar la distribución de los viajes validados, es decir, registrados en la base de datos de las *smartcards*, se puede identificar el porcentaje de viajes “sólo bus” con más de una etapa de viaje, dicho dato para el caso de Santiago es de 14.4% (Ipsos, 2013). Este dato es importante, ya que depende que tan grande sea, el si se justifica desarrollar una metodología que los aborde, o simplemente despreciar la posibilidad de este tipo de evasión en ese tipo de viaje.

Otra limitación consiste en que el modelo de estimación de la evasión, aplica para un corte temporal determinado, pero se ha encontrado que en muchos sistemas del mundo la evasión fluctúa en el tiempo (Buneder, 2017); con lo cual es necesario herramientas de actualización del modelo, que permita aplicarlo a otros periodos de tiempo.

## 5.2. Futuras líneas de investigación.

Referente a los modelos de estimación de tasas de evasión, se propone la inclusión de variables asociadas a las características personales del evasor, información de fiscalización y presencia de mecanismos de control de evasión, como el caso de los torniquetes. Dichas variables se consideran interesantes de incluir, y no se hizo en la presente investigación, por falta de información al respecto.

Desarrollar metodologías que permitan estimar, el dato grueso de la distribución entre evasión completa y evasión parcial, a nivel de sistema. De esta forma es posible incluir en la metodología de corrección a los evasores parciales “sólo bus”, que están por fuera de lo abordado en esta investigación.

De igual forma se propone abordar desde un enfoque cualitativo, que permita realizar perfiles de usuarios más detallados, inclusive encontrando otro tipo de evasor, adicional a los dos abordados en esta metodología.

Finalmente, sería de gran interés, contar con bases de datos, que adicional al conteo de evasores, den información de la persona, ya sea sexo, edad, etc. Que permita incluir variables de la persona, en los modelos de estimación de evasión.

## 6. REFERENCIAS

- Buneder, C. (2017). Temporal and spatial analysis of fare evasion in Transantiago. Tesis para optar al grado de Magíster en Ciencias de la Ingeniería, mención Transporte, Pontificia Universidad Católica de Chile.
- Barabino, B., Salis, S., & Useli, B. (2015). What are the determinants in making people free riders in proof-of-payment transit systems? Evidence from Italy. *Transportation Research Part A: Policy and Practice*, 80, 184-196.
- Clarke, R. V., Contre, S., & Petrossian, G. (2010). Deterrence and fare evasion: Results of a natural experiment. *Security Journal*, 23(1), 5-17.
- Delbosc, A., Currie, G. (2016) Cluster analysis of fare evasion behaviours in Melbourne, Australia. *Transport Policy* 50, 29-36.
- Devillaine, F., Munizaga, M., And Trepanier, M. (2012). Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2276 (1), 48-55.
- DICTUC (2012) Mediciones de perfil de carga con evasión en la ciudad de Santiago. Estudio realizado por DICTUC para la Coordinación Transantiago.
- Eliason, S. R. (1993). Maximum likelihood estimation: Logic and practice (Vol. 96). Sage Publications.

- Factor Estratégico (2010) Informe Final Estudio Cualitativo de Opinión: el Fenómeno de la Evasión, Ministerio de Transportes y Telecomunicaciones, Chile.
- Gallegos, N. (2016) Construcción de factores de corrección por evasión para la matriz origen-destino de viajes en transporte público en Santiago obtenida a partir de transacciones Bip!, Tesis para optar al grado de Magíster en Ciencias de la Ingeniería, mención Transporte, Universidad de Chile.
- Guarda, P., Galilea, P., Paget-Seekins, L., & de Dios Ortúzar, J. (2016). What is behind fare evasion in urban bus systems? An econometric approach. *Transportation Research Part A: Policy and Practice*, 84, 55-71.
- Hosmer, D. W., & Lemeshow, S. (2000). Interpretation of the fitted logistic regression model. *Applied Logistic Regression*, Second Edition, 47-90.
- Ipsos (2013) Encuesta Origen Destino de viajes 2013. Realizado por Ipsos para Metro S.A.
- Lee, J. (2011). Uncovering San Francisco, California, Muni's proof-of-payment patterns to help reduce fare evasion. *Transportation Research Record: Journal of the Transportation Research Board*, (2216), 75-84.
- Munizaga, M., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
- Munizaga, M., Devillaine, F., Navarrete, C. And Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, 44, 70-79.
- Ortúzar, J. de D. & Willumsen, L. G. (2011). *Modelling Transport* (Fourth Edi.). Chichester: John Wiley & Sons, Ltd.
- Programa Nacional de Fiscalización. (13 de 01 de 2017). Fiscalización Transporte. Obtenido de [www.fiscalizacion.cl/wp-content/uploads/2016/10/Indice-Evasion-Jul-Sep16.pdf](http://www.fiscalizacion.cl/wp-content/uploads/2016/10/Indice-Evasion-Jul-Sep16.pdf)
- Raveau, S., Muñoz, J.C., Prato, C., Soto, A., Tamblay, S., & Iglesias, P. (2017). A Behavioural Planning Tool for Modelling Public Transport Systems. *Transit Data* 2017.
- Revista Semana. (13 de junio de 2017). Transmilenio pierde \$555 millones diarios por ‘colados’. Obtenido de <http://www.semana.com/nacion/articulo/no-pagar-en-transmilenio-le-cuesta-1700-millones-bogota/385385-3>
- SECTRA (2015). Informe Final del estudio “Encuesta Origen Destino de Viajes 2012”. Secretaría Interministerial de Planificación de Transporte, Santiago.
- Tamblay, S., Galilea, P., Iglesias, P., Raveau, S., & Muñoz, J. (2016). A zonal inference model based on observed smart-card transactions for Santiago de Chile. *Transportation Research Part A: Policy and Practice*, 44-54.
- Torres-Montoya, M. (2014). Tackling fare evasion in Transantiago: an integrated approach. In *Transportation Research Board 93rd Annual Meeting* (No. 14-4641).
- Troncoso, R., & de Grange, L. (2017). Fare evasion in public transport: A time series approach. *Transportation Research Part A: Policy and Practice*, 100, 311-318.