



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

¿QUE MAS HAY DETRAS DE LA EVASION EN EL TRANSPORTE PUBLICO? UN ENFOQUE ECONOMETRICO

PABLO ANDRES GUARDA ROSAS

Tesis para optar al grado de Magíster en Ciencias de la
Ingeniería

Profesores Supervisores:

PATRICIA VIVIANA GALILEA ARANDA

JUAN DE DIOS ORTUZAR SALAS

Santiago de Chile, Agosto, 2015

© 2015, Pablo Guarda Rosas



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

¿QUE MAS HAY DETRAS DE LA EVASION EN EL TRANSPORTE PUBLICO? UN ENFOQUE ECONOMETRICO

PABLO ANDRES GUARDA ROSAS

Tesis presentada a la Comisión integrada por los profesores:

PATRICIA GALILEA ARANDA

JUAN DE DIOS ORTUZAR SALAS

JUAN CARLOS MUÑOZ ABOGABIR

CRISTIAN GUEVARA CUE

JORGE MANZI ASTUDILLO

ALDO CIPRIANO ZAMORANO

Para completar las exigencias del grado de Magíster en Ciencias de la
Ingeniería

Santiago de Chile, Agosto, 2015

Dedicada a mis padres y a mi
hermana por su apoyo incondicional
y por inculcar en mí la vocación de
trabajar por un mundo más justo.

AGRADECIMIENTOS

Quiero agradecer especialmente a mis padres y mi hermana, por su apoyo incondicional en los momentos difíciles y por estar siempre presentes en los hitos importantes de mi vida. Agradezco a Cristóbal Camplá, Daniel Duque, Francisco Jiménez S.J, Dagoberto Lagos S.J, Rosana Latuf y Bernarda Verdugo, por guiarme en el plano espiritual y entregarme herramientas fundamentales para sobrellevar los desafíos de la carrera.

Agradezco a mis amigos y compañeros de universidad, que siempre me motivaron a dar lo mejor e hicieron mucho más grato trabajar en la tesis. Además, quiero agradecer al profesor Juan de Dios Ortúzar y a la profesora Patricia Galilea, por cumplir de forma ejemplar su rol de profesores supervisores. Valoro su cercanía, calidad humana, dedicación y generosidad para compartir su conocimiento experto. Agradezco también el apoyo de los demás miembros de la comisión, su visión externa y su mirada interdisciplinaria enriquecieron mucho esta investigación. Destaco la ayuda de Eduardo Acuña, jefe del Programa Nacional de Fiscalización, su colaboración fue clave para acceder y entender la base de datos utilizada en esta tesis.

Finalmente, agradezco el financiamiento aportado por el Centro de Desarrollo Urbano Sustentable, CEDEUS (Conicyt/Fondap/15110020), el Instituto Milenio de Sistemas Complejos de Ingeniería (ICM: P-05-004-F; CONICYT: FBO16), el Bus Rapid Transit Centre of Excellence (www.brt.cl) y FONDEF D10I1049 – “Una herramienta táctico-estratégica de gestión y planificación de sistemas de transporte público urbano”.

ÍNDICE GENERAL

	Pág.
DEDICATORIA.....	ii
AGRADECIMIENTOS.....	iii
INDICE DE TABLAS.....	vii
INDICE DE FIGURAS.....	viii
RESUMEN.....	ix
ABSTRACT.....	x
1 INTRODUCCION.....	1
1.1 Contexto.....	1
1.2 Objetivos y motivación.....	2
1.3 Estructura.....	3
2 REVISION BIBLIOGRAFICA.....	4
2.1 La psicología de la evasión.....	4
2.1.1 El rol de las emociones.....	4
2.1.2 Efecto de la exposición a la evasión.....	6
2.1.3 Anonimato.....	7
2.2 Políticas actuales para controlar la evasión.....	7
2.2.1 Inspección.....	8
2.2.2 Zonas de prepago.....	9
2.2.3 Barreras físicas.....	11
2.2.4 Estrategia comunicacional.....	12
2.2.5 Rol del conductor.....	13

2.3	Enfoques matemáticos para modelar la evasión en el transporte público	13
3	MODELOS DE REGRESION LINEAL Y DE CONTEO: ESPECIFICACION Y SUPUESTOS	17
3.1	Modelos de regresión lineal múltiple (RLM)	17
3.2	Modelos de regresión de conteo	18
3.2.1	Regresión de Poisson	18
3.2.2	Regresión binomial negativa.....	20
3.3	Supuestos de los modelos	21
3.3.1	Ausencia de multicolinealidad.....	21
3.3.2	Normalidad y homocedasticidad.....	22
3.3.3	Ausencia de sobredispersión.....	22
4	FORMULACION, ESTIMACION Y SELECCION DEL MODELO.....	24
4.1	Generación del banco de datos	24
4.2	Variable dependiente	25
4.3	Variables explicativas	26
4.3.1	Operación de las puertas	26
4.3.2	Variables temporales.....	28
4.3.3	Calidad del servicio.....	29
4.3.4	Operación de la ruta	31
4.3.5	Tipo de paradero	32
4.3.6	Nivel socioeconómico (NSE)	33
4.4	Análisis estadístico descriptivo.....	37
4.5	Estimación de modelos de regresión para explicar la evasión.....	40
4.6	Cumplimiento de los supuestos de los modelos	42

4.7	Comparación del poder predictivo de los modelos.....	43
4.7.1	Gráficos de predicción	44
4.7.2	Indicador de bondad de ajuste.....	45
5	MEJORAS AL MODELO Y ANALISIS DE RESULTADOS	48
5.1	Inclusión del efecto del nivel socioeconómico (NSE) en el modelo	48
5.2	Estimación del modelo para el periodo punta mañana	50
5.3	Análisis de sensibilidad.....	51
6	PROPUESTAS DE POLITICA PUBLICA.....	60
6.1	Tamaño de la flota	60
6.2	Regularidad	61
6.3	Sistema de pago en el bus	61
6.4	Zonas de prepago (ZP).....	62
6.5	Diseño del bus.....	63
6.6	Programación horaria de la inspección	63
6.7	Subsidios	64
7	CONCLUSIONES	66
	BIBLIOGRAFIA	69
	ANEXOS	74
	Anexo A: Descripción de los datos disponibles	75
	Anexo B: Razón de tasas de incidencia y factor de cambio	77
	Anexo C: Especificaciones alternativas de los modelos de regresión lineal múltiple	80
	Anexo D: Modelo de regresión binomial negativa con estratificación por zonas	83
	Anexo E: Variables no significativas en los modelos de regresión	87

INDICE DE TABLAS

	Pág.
Tabla 4-1 Clasificación de los periodos del día en Transantiago	28
Tabla 4-2 Categorías usadas para medir el índice de ocupación (IO) del bus	29
Tabla 4-3 Características de la operación de las unidades de negocio de Transantiago..	31
Tabla 4-4 Nivel de ingreso familiar promedio en comunas de Transantiago 2009	35
Tabla 4-5 Características por quintil de ingreso área urbana Santiago 2009	36
Tabla 4-6 Subconjunto de servicios medidos de cada operador de Transantiago	37
Tabla 4-7 Estadísticas variables no categóricas, Octubre 2012	38
Tabla 4-8 Estadísticas variables categóricas, Octubre 2012	39
Tabla 4-9 Estimación modelos de regresión lineal, Poisson y binomial negativa.....	41
Tabla 4-10 Test estadísticos para validar los supuestos de los modelos de regresión	42
Tabla 4-11 Resultados del test chi-cuadrado para los modelos RLM, RP y NB2	47
Tabla 5-1 Mejoras al modelo de regresión binomial negativa (NB2)	48
Tabla 5-2 Razón de tasas de incidencia y factores de cambio modelo NB2-I.....	53
Tabla 6-1 Características por quintil de ingreso área urbana Santiago 2006.....	64
Tabla A-1 Descripción de variables asociadas a los digitadores y fiscalizadores	75
Tabla A-2 Descripción de variables medidas por los fiscalizadores	75
Tabla C-1 Estimación modelos de regresión lineal múltiple (RLM).....	81
Tabla D-1 Promedio ponderado del nivel de ingreso en cada zona de Transantiago	84
Tabla D-2 Estimación modelos de regresión binomial negativa con efecto ingreso	85
Tabla E-1 Estimación modelos con variable de tipo de servicio, día y frecuencia	87
Tabla E-2 Estimación modelo NB2-I con todas las variables categóricas del NSE	88

INDICE DE FIGURAS

	Pág.
Figura 1-1 Evolución tasa mensual de evasión entre Mayo 2007 y Diciembre 2014.....	1
Figura 4-1 Valores observados y predicciones de los modelos RLM, RP y NB2	45
Figura 5-1 Factores de cambio según grupo de operadores (modelo NB2-I).....	57
Figura 5-2 Factores de cambio según grupo de operadores (modelo NB2)	58
Figura 5-3 Factores de cambio según NSE municipalidad (modelo NB2-I).....	59
Figura 5-4 Factores de cambio según NSE municipalidad (modelo NB2-I-AM)	59
Figura A-1 Ficha de toma de datos usada por el Programa Nacional de Fiscalización...	76
Figura C-1 Valores observados y predicciones modelos de regresión lineal múltiple	82

RESUMEN

La evasión es un problema en muchos sistemas de transporte público alrededor del mundo. Las estrategias para reducirla están generalmente enfocadas en aumentar la inspección y las multas. En esta tesis, se analizan otras variables para explicar los altos niveles de evasión en Santiago, con el propósito de orientar las políticas públicas en esta materia. Usando información de corte transversal recolectada por el Programa Nacional de Fiscalización del Ministerio de Transporte y Telecomunicaciones para medir la tasa de evasión mensual en Transantiago, el fenómeno es explicado usando modelos de regresión. Después de estimar varias clases de modelos de regresión y aplicar diferentes test estadísticos, se determinó que el modelo de regresión binomial negativa era la mejor especificación. En base a este modelo, se encontró que las tasas de evasión en los buses aumentan a medida que: más gente entra (o sale) por una puerta, más pasajeros entran por las puertas traseras, los buses tienen niveles más altos de ocupación (y más puertas) y mayores intervalos entre pasadas (*headway*). Además, controlando por estas variables (*ceteris paribus*), los resultados indican que la evasión es mayor durante la tarde y la noche (aunque no es claro que la evasión sea mayor durante el horario punta). En cuanto a las variables socioeconómicas, se encontró que la evasión en paraderos ubicados en áreas de ingreso alto es significativamente menor que en áreas de ingreso bajo. Finalmente, en base a los resultados obtenidos se identificaron cinco métodos para enfrentar la evasión como alternativas a incrementar la inspección y las multas; (i) Aumento de la flota de buses; ii) Mejora de la regularidad del *headway* en la ruta; iii) Implementación de zonas pagas; iv) Cambio del sistema de pago en el bus; v) Modificación del diseño del bus (número de puertas y la cantidad de pasajeros por bus). El enfoque econométrico de este trabajo provee una herramienta poderosa para predecir el cambio en la evasión debido a la implementación de este tipo de estrategias operacionales. Esta metodología también tiene el potencial para ser aplicada en otros sistemas de transporte público en el mundo.

ABSTRACT

Fare evasion is a problem in many public transport systems around the world. Policies to reduce it are generally aimed at improving control and increasing fines. The aim of this thesis is to analyse other variables that may explain the high evasion levels in Santiago in order to guide public policy in this matter. Using cross sectional data collected by the National Enforcement Programme of the Ministry of Transport and Telecommunication on the monthly fare evasion rate on buses in Santiago, the phenomenon is explained using regression models. After testing several classes of regression models and apply different statistical tests, we found the negative binomial regression model to be the best specification. With this model, we found that bus fare evasion rates increase if: more people board (or alight) at a given bus door, more passengers board by a rear door, buses have higher occupancy levels (and more doors) and longer headways between them. Additionally, by controlling these variables (*ceteris paribus*), results indicate that evasion is greater during the afternoon and evening (but it is not clear that fare evasion is higher during peak hours). In the context of Transantiago, the high fare evasion rates measured in low-income areas could be legitimately linked to an inability to afford public transport, and could explain, partially at least, the average differences on fare evasion observed between municipalities in the regression models. Based on the results, five main methods to address evasion were identified as alternatives to increase inspection and enforcement; (i) Increasing the bus fleet, (ii) Improving the bus headway regularity; iii) Implementing off-board payment stations; iv) Changing the payment system on board; v) Modifying the bus design (number of doors or capacity). The econometric approach used in this work provides a powerful tool to predict the reduction of fare evasion due to the implementation of these types of operational strategies. This methodology could be also applied to other bus public transport systems.

1 INTRODUCCION

1.1 Contexto

En Febrero de 2007 Chile inauguró un sistema de transporte público integrado en su capital, llamado Transantiago, el cual ha sido ampliamente criticado desde su inicio debido a los problemas de su implementación (Muñoz et al., 2008). El diseño del sistema de buses fue modificado radicalmente, pasando desde un sistema informal donde los pasajeros pagaban en efectivo a los conductores, a un sistema donde el pago se realiza mediante tarjetas electrónicas (tarjeta BIP) y que no está operado por el conductor. El sistema antiguo de “micros amarillos”, en el cual muchos conductores eran dueños de sus propios buses y sus sueldos dependían del número de pasajes pagados, cambió a un sistema de buses operados por grandes compañías reguladas por el Estado y en que los conductores perciben un ingreso fijo y tienen condiciones laborales mucho más dignas. La evasión no era medida en el sistema previo, pero en Transantiago ha llegado a constituir un problema serio (Figura 1-1).

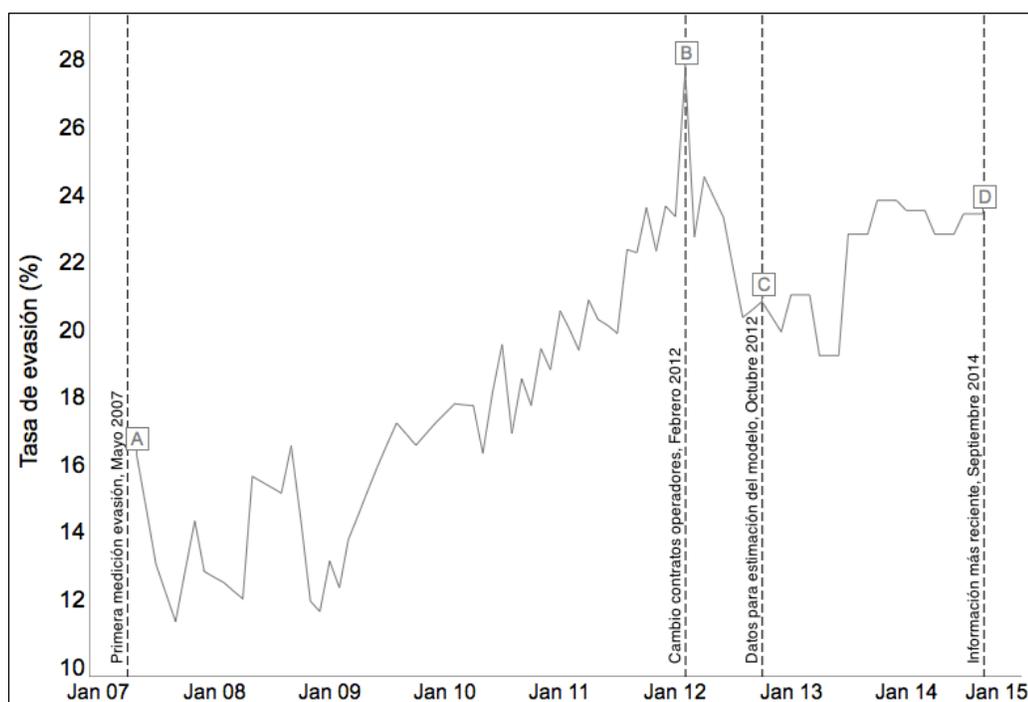


Figura 1-1 Evolución tasa mensual de evasión entre Mayo 2007 y Diciembre 2014

Fuente: Gráfico construido con R Studio. Datos obtenidos de DTPM (2014).

1.2 Objetivos y motivación

Las políticas para disminuir la evasión del pago de pasajes en el mundo han estado muy enfocadas en mejorar los sistemas de fiscalización e incrementar las multas a los infractores (Killias et al., 2009). De la misma forma, muchos estudios en la literatura se han focalizado en el diseño de estrategias de reducción de la evasión basadas en la fiscalización (Barabino et al., 2013; Lee, 2011; Thorlacius et al., 2010). Sin embargo, entender a los evasores como actores racionales que ponderan únicamente los costos del pasaje y la probabilidad de ser atrapados, no considera el contexto social donde la evasión tiene lugar (Boyd et al., 1989; Kooreman, 1993). En realidad, la evasión parece ser una combinación de muchos otros factores tales como el nivel de ingreso de los pasajeros, la percepción del servicio de transporte público, el tipo de sistema de pago del pasaje, entre otros (Buccioli et al., 2013; Torres-Montoya, 2014). Por esta razón, enfocar la política pública en solo aumentar el nivel de fiscalización podría sobreestimar el beneficio de este tipo de estrategias y generar costos innecesarios de implementación.

En este contexto, el objetivo principal de esta investigación es analizar y cuantificar el impacto conjunto en la evasión de algunos factores no vinculados con la fiscalización, con el propósito de identificar nuevas estrategias para enfrentar este problema. Para llevar esto a cabo, se estimaron modelos de regresión relativamente sofisticados, que permitieron identificar variables clave que explican la evasión, tales como el nivel de ingreso del área donde los paraderos están ubicados, el periodo del día, el nivel de servicio del transporte público (*headway* y nivel de ocupación) y la operación de las puertas en los paraderos (número de pasajeros entrando y saliendo por las puertas). Pese a que diversas agencias de transporte público en el mundo han constatado y reconocen que las variables operacionales inciden en los niveles de evasión (por ejemplo, el nivel de ocupación de los buses o la acumulación de pasajeros en los paraderos), hasta ahora no se han formulado modelos estadísticos para medir su impacto conjunto. A lo largo de la tesis se presentará evidencia estadística obtenida rigurosamente, que contribuye a ampliar el entendimiento actual sobre la evasión y guiar las políticas públicas dirigidas a reducirla.

1.3 Estructura

Esta tesis se organiza en siete capítulos. El segundo capítulo presenta una revisión de teorías desarrolladas en el campo de la psicología social para entender los procesos que intervienen en la concepción de la conducta ilícita; también se describen algunas estrategias implementadas en otros países para enfrentar la evasión y se revisan los estudios existentes en la literatura para modelarla. El tercer capítulo presenta el marco teórico de los tres modelos de regresión utilizados en esta tesis; regresión lineal múltiple, regresión de Poisson y regresión binomial negativa. El cuarto capítulo describe los datos disponibles y las variables incluidas en la formulación de los modelos, presenta los resultados de la estimación de los modelos y los criterios de selección de la mejor especificación. En el quinto capítulo se validan los resultados obtenidos con la especificación seleccionada y se realiza un análisis de sensibilidad para analizar el impacto de las variables explicativas sobre los niveles de evasión. Luego, en base a los resultados obtenidos con los modelos de regresión, el capítulo 6 propone políticas públicas para reducir la evasión. Finalmente, el capítulo 7 presenta las conclusiones de la tesis y algunas líneas futuras de investigación.

2 REVISION BIBLIOGRAFICA

2.1 La psicología de la evasión

A diario se conocen grandes fraudes corporativos que son ampliamente cuestionados por la sociedad. No obstante, mucho más frecuentes son actos individuales a pequeña escala que transgreden las leyes y que, independiente de su nivel de gravedad, no dejan de ser cuestionables desde un punto de vista ético y legal (Karstedt y Farrall, 2006). Entonces, si los involucrados tienen claro que estos actos violan los principios morales de la sociedad y suelen tener consecuencias en el entorno ¿Qué hace que gente común y corriente realice estas acciones? ¿Cómo puede la gente ser seducida para actuar inmoralmente? ¿Dónde está la línea entre actuar o no conforme a principios que debieran regir la sociedad? Desafortunadamente, existen fuerzas sociales y situacionales que condicionan las decisiones de cada individuo y lo impulsan a cruzar esta línea, independiente de cuál sea su nivel socioeconómico, rango etario o clase social (Zimbardo, 2008).

Algunos macro-componentes como las concepciones morales en la sociedad son buenos predictores de la frecuencia de los actos ilícitos. Sin embargo, también hay factores a menor escala que tienen gran incidencia en la conducta individual (Gino et al., 2009). En particular, la influencia de dinámicas grupales, las emociones o el contexto situacional permiten comprender por qué las decisiones frente a dilemas éticos no son inmutables. En este contexto, para estudiar la conducta de los evasores de Transantiago, es fundamental revisar algunos estudios desarrollados en el campo de la ética del comportamiento (Bazerman y Gino, 2012). Estas teorías identifican algunos factores claves que explican la concepción de cualquier tipo de acto ilícito y, por consiguiente, también son útiles para estudiar el caso de la evasión en Transantiago.

2.1.1 El rol de las emociones

La literatura plantea que en el desarrollo del acto ilícito intervienen distintos procesos psicológicos que se producen tanto a nivel interno como externo a los individuos. Por un lado, hay procesos externos asociados a la influencia social según los cuales bajo ciertas

condiciones, la gente tiende a reemplazar sus estándares individuales por los del grupo (Myers, 2012). Por otra parte, existen factores internos, como las emociones experimentadas por el resultado de tomar una decisión no ética (Ruedy et al., 2013).

Las emociones son un componente crítico en cualquier decisión, y particularmente importantes en las decisiones que conciernen a la ética de los individuos (Ruedy et al., 2013), ya que en estos casos se ven comprometidas las creencias y valores de cada persona. Mucha literatura en el campo de la psicología social plantea que los actos no éticos solo generan emociones negativas en quien los realiza. Sin embargo, estos estudios han estado enfocados en situaciones donde hay una víctima prominente (Ruedy et al., 2013). En estos casos existe un daño evidente sobre otras personas o el entorno, por lo cual se activan mecanismos cognitivos relacionados con el juicio moral del individuo, que normalmente disminuyen la inclinación a cometer un acto contrario a sus principios éticos. En contraste, ante dilemas éticos donde el individuo no percibe un daño evidente en su entorno, no es claro que sienta emociones negativas por esta acción. Incluso, en algunos casos las personas podrían disfrutar al participar en comportamientos catalogados como “inmorales”; este fenómeno se conoce como *cheater's high* (Ruedy et al., 2013).

En este sentido, la relación entre la conducta ilícita y la percepción de las instituciones públicas también debería ser considerada (Karstedt y Farrall, 2006). A comienzos del 2013, la consultora Collect Gfk (2010) reveló que el 61% de los pasajeros en Transantiago evalúa negativamente su desempeño (EMOL, 2013). Por lo tanto, una proporción importante de la evasión podría ser una forma de protesta frente a la insatisfacción con el sistema y, en tal caso, los pasajeros no experimentarían una emoción negativa por cometer el acto ilícito. Por otra parte, los pasajeros podrían no estar conscientes de las pérdidas financieras que la evasión genera en el sistema, por lo cual la evasión podría ser vista por muchos como “... un comportamiento no ético pero sin un daño evidente o una víctima prominente” (Ruedy et al., 2013). De esta forma, mejoras en el nivel de servicio de Transantiago podrían contribuir a aumentar la

valoración de los usuarios con el sistema de transporte público y lograr que los pasajeros experimenten emociones negativas al evadir.

2.1.2 Efecto de la exposición a la evasión

En un mundo en el cual es frecuente encontrarse con la deshonestidad, es importante saber si la exposición a comportamientos no éticos puede incrementar o disminuir la disposición de los individuos a actuar conforme a sus principios éticos. Gino et al. (2009) sugieren que el comportamiento no ético de otros individuos puede influenciar a quienes lo observan mediante (al menos) tres mecanismos, que varían en importancia según el contexto en que la acción se desarrolle. En primer lugar, la exposición al comportamiento no ético puede derivar en la subestimación de la probabilidad de ser descubierto (e.g. multado por un inspector), lo cual tiende a aumentar la predisposición del individuo a cometer un acto deshonesto. En segundo lugar, si la gente observa a alguien actuando de forma deshonesto (e.g., no pagando el pasaje), la “prominencia” (del inglés *saliency*) del acto tiende a aumentar, induciéndolos a prestar más atención a sus propios estándares de honestidad y, por consiguiente, disminuyendo su predisposición a actuar en contra de los principios morales. Finalmente, el “mecanismo de las normas sociales” sugiere que la influencia de la exposición al comportamiento no ético depende del nivel de identificación que sienta un individuo con quienes cometen el acto ilícito. De esta forma, si la persona se identifica con un individuo que comete un acto ilícito, la probabilidad de participar de este comportamiento no ético aumenta; en cambio, si no se siente identificada, la probabilidad de cometer el acto no ético disminuye.

La literatura coincide en que la exposición de los pasajeros a la evasión aumenta la propensión a evadir (Reddy et al., 2011). Estudios recientes han mostrado que bajo ciertas condiciones se observa un aumento en la evasión cuando los pasajeros observan a otras personas evadiendo (Buccioli et al., 2013). Sin embargo, el impacto de los mecanismos de contagio social entre los evasores aún no ha sido estudiado con profundidad y no existen modelos que permitan cuantificar este efecto en las cifras globales de evasión del sistema de transporte público.

2.1.3 Anonimato

El anonimato es un componente clave dentro de los mecanismos de influencia grupal que inciden en la conducta individual. Generalmente, las personas son sensibles a las presiones sociales y se sienten más cómodos en el anonimato y actuando al interior de un grupo (Gino et al., 2009). Cuando la gente actúa en contextos colectivos, puede experimentar un proceso de “desindividualización”, en el cual los estándares individuales son reemplazados por los del grupo, derivando en un aumento de la susceptibilidad a los procesos de influencia grupal (Zimbardo y Leippe, 1991). De esta forma, aumenta la imitación de las conductas observadas en otros y se gatillan diversos mecanismos de contagio social (Hatfield et al., 1994), que tienen una influencia mayor cuando miembros del mismo grupo son quienes participan en el acto ilícito (discutido en Sección 2.1.2).

Durante los periodos de mayor afluencia de pasajeros en el sistema de transporte público y en las líneas de buses con mayores niveles de ocupación, se suele registrar un incremento significativo de las tasas de evasión (Reddy et al., 2011). En este sentido, el anonimato podría aumentar la disposición de algunos pasajeros a evadir. Así, desde una mirada de política pública, sería posible diseñar campañas con un efecto disuasivo para cambiar la percepción de la evasión e incrementar los “costos sociales” de evadir; vinculados con la vergüenza, el miedo o la culpa (Torres-Montoya, 2014), pero su efectividad dependerá de las motivaciones de cada evasor para no pagar el pasaje y de la incidencia de otros mecanismos de influencia social que estén condicionando su conducta.

2.2 Políticas actuales para controlar la evasión

En esta sección se presenta una descripción de algunos mecanismos de control de evasión utilizados tanto en Transantiago como en otros sistemas de transporte público en el mundo. Además, a partir de las conclusiones obtenidas de diversos estudios de campo desarrollados en la literatura, se evaluará las ventajas y desventajas de cada una de estas estrategias. A pesar que estos estudios se han desarrollado mayoritariamente en servicios

de Metro y ciudades con realidades locales muy diferentes (en relación a las características de los pasajeros, la operación del transporte público, entre otros), gran parte de las conclusiones también son aplicables para los sistemas de buses.

2.2.1 Inspección

Se han propuesto dos diseños principales para las políticas de inspección: la asignación de inspectores fijos en los paraderos y la fiscalización de pasajeros de forma aleatoria a lo largo del recorrido del servicio de buses. En Londres, se utilizan estas dos estrategias de inspección y la autoridad de transporte de la ciudad está encargada de llevar a cabo la labor de fiscalización. El primer grupo corresponde a inspectores uniformados que realizan la fiscalización en los buses y su función principal es controlar el pago de tarifas. Normalmente tienen un perfil cercano a la gente, y están capacitados para entregar información completa del servicio (sobre recorridos, frecuencias y tarifas, entre otros). El segundo grupo corresponde a inspectores sin uniforme y caracterizados por una presencia autoritaria (físicamente), que realizan fiscalización sorpresiva dentro de los buses.

En sistemas de transporte donde las estrategias de fiscalización han sido exitosas (como el caso de Londres), se observa que los inspectores cuentan con muchas atribuciones (Castillo, 2010). Por ejemplo, tienen facultades legales para aplicar multas a quienes evadan el pasaje, pueden solicitar documentación para verificar la identidad del evasor, requisar documentos y controlar otro tipo de faltas como el rayado o daño a la infraestructura de los buses. Bajo este diseño institucional, no es imprescindible la presencia policial y la figura del inspector cobra más relevancia como autoridad (Bonfanti y Wagenknecht, 2010; Torres-Montoya, 2014), activando mecanismos psicológicos que desincentivan la conducta ilícita en los pasajeros.

El sistema de recolección de multas y el monto de las mismas también determina la costo-efectividad del sistema de fiscalización. Sin embargo, no es claro cuál debe ser el balance óptimo entre la tasa de inspección del pasaje y el monto de la multa para reducir la evasión al máximo, ni tampoco cuál es el valor mínimo deseable (Clarke et al., 2010).

La costo-efectividad de las estrategias de fiscalización también está determinada por el horario en que se lleva a cabo la fiscalización (Lee, 2011). En el sistema de transporte público conviven diferentes tipos de evasores, y existen pequeños grupos que no cambian su comportamiento ante cambios de la severidad de las penas (e.g. monto de la multa) y la frecuencia de la fiscalización (Bijleveld, 2007). Las agencias de transporte público han observado que la presencia de estos grupos varía en cada periodo horario (Killias et al., 2009), y esto también se ha vinculado con el cambio sistemático de la evasión a lo largo del día.

A pesar que parece evidente que el aumento de las multas incrementa la auto-sustentabilidad del sistema de fiscalización, la literatura ha demostrado que existen otros factores que pueden llegar a ser más importantes que la severidad de la pena. Killias et al. (2009) sugieren que un factor determinante es que los pasajeros tengan certeza de que ser sorprendidos evadiendo el pasaje tendrá consecuencias. Bootheway (2009) plantea la dificultad de definir el monto de la multa, indicando que la probabilidad de que los evasores paguen es inversamente proporcional al costo de la penalidad. La literatura coincide en que todos los factores discutidos deben ser considerados dentro de las políticas de fiscalización. En la Sección 2.3, se presentarán diversos modelos matemáticos formulados en la literatura para responder estas interrogantes y encontrar los niveles óptimos de inspección en el sistema.

2.2.2 Zonas de prepago

Las zonas de prepago (ZP) son paraderos en que el pasajero debe pagar su pasaje antes de subir al bus, permitiendo que el abordaje sea más expedito (Tirachini, 2013). Las ZP están normalmente ubicadas en lugares con altos flujos de pasajeros, altos niveles de evasión, y donde pasan servicios con baja regularidad y bajo cumplimiento de las frecuencias definidas por las agencias de transporte público. Las ZP normalmente operan en horarios de alta demanda de pasajeros y cuentan con validadores electrónicos ubicados en el área de acceso a los buses. Para controlar el pago del pasaje, los paraderos suelen tener inspectores ubicados cerca de los validadores electrónicos y cuentan con barreras metálicas que impiden que los pasajeros eludan la inspección.

El mecanismo principal por el cual las ZP generan beneficios operacionales, es la disminución de los tiempos de subida de los pasajeros que utilizan cualquiera de las líneas de buses que se detienen en este paradero (Tirachini, 2013). Por la reducción de los tiempos de detención de los buses en el paradero, disminuye el tiempo de ciclo del servicio, y por lo tanto el operador requiere una flota de buses más pequeña para cumplir con el mismo total de kilómetros recorridos por la flota antes de la implementación de la ZP. En este escenario, una alternativa del operador es ofrecer la misma frecuencia que había antes de la implementación de la ZP y reducir sus costos operacionales (por la reducción del tamaño de la flota). Otra alternativa del operador es mantener la flota de buses previa a la implementación de la ZP y generar un aumento de la frecuencia ofrecida a los usuarios. Así, el nivel de servicio entregado a los pasajeros mejoraría, por la disminución de los tiempos de espera de los pasajeros, la reducción del nivel de ocupación de los buses y la disminución del tiempo promedio de viaje de todos quienes abordan el servicio en algún paradero ubicado antes de la ZP. Finalmente, la mejora del nivel de servicio podría generar un aumento de la demanda del servicio de buses (en el mediano y largo plazo) y una disminución significativa de los niveles de evasión en todos los paraderos del recorrido (Sección 2.1.1).

Otra fuente de beneficios para el operador que no está asociada a la disminución de los tiempos de subida, es la reducción potencial de la evasión por el efecto de la inspección en el paradero donde la ZP es implementada. Sin embargo, no existen metodologías que permitan calcular de forma confiable los beneficios potenciales originados por la reducción de evasión en las ZP. Normalmente, se asume que la tasa de evasión es nula debido a la presencia de los fiscalizadores (Sección 4.3); sin embargo, hasta ahora no se han realizado estudios que estimen cuál es la reducción real de la evasión en estos paraderos. Por otro lado, una proporción de la reducción en la tasa de evasión debido a la ZP, podría estar explicada por evasores que caminan a paraderos cercanos para eludir la inspección. De hecho, durante el horario de operación de las ZP, personal de Transantiago ha constado en terreno que muchos pasajeros caminan a paraderos

colindantes, ingresan por los costados de las barreras metálicas de la ZP o hacen caso omiso a los inspectores en el paradero.

En el sistema Transmilenio en la ciudad de Bogotá, Colombia, se han implementado medidas más extremas para impedir que los evasores eludan la inspección y las barreras que rodean a los paraderos. En la estructura de zonas de prepago en Bogotá se observa que las puertas de entrada de los buses calzan justo con el sistema de puertas en los andenes, por lo cual no hay espacio físico por el cual se pueda acceder al bus sin haber entrado a la estación y pagado la tarifa. No obstante, algunos pasajeros siguen evadiendo y este comportamiento continúa siendo causa de muchos accidentes. En algunos casos para evadir se deben atravesar pistas de automóviles y de buses (Castillo, 2010).

2.2.3 Barreras físicas

Las barreras físicas son implementadas normalmente en sistemas de Metro (Reddy et al., 2011) y su principal finalidad es dificultar o imposibilitar que los pasajeros utilicen el sistema sin pagar el pasaje. Sin embargo, en la literatura no existe consenso sobre la costo-efectividad de esta medida. Algunas agencias han observado en terreno que el costo de la infraestructura y personal necesario para implementar esta estrategia, sobrepasa los beneficios obtenidos por la reducción de evasión (Clarke, 1993; Weidner, 1996). Otros autores argumentan que las estrategias de inspección son mucho más costo-efectivas (Correa et al., 2014). Por otra parte, en algunos sistemas donde no se utilizan barreras físicas (conocidos como “honor systems”), la implementación de esta medida ha sido muy impopular entre los pasajeros (Weidner, 1996).

En Transantiago, uno de los operadores más grandes ha optado por implementar torniquetes en la puerta delantera de los buses para poner una dificultad adicional a los evasores. Sin embargo, el efecto de las barreras físicas en sistemas de buses no ha sido revisada en la literatura y, en comparación al Metro, la implementación podría tener dificultades adicionales que podrían hacer menos rentable esta estrategia. En particular, las barreras físicas en las puertas de los buses aumentan los tiempos de subida de los pasajeros y, por consiguiente, podría deteriorarse el nivel de servicio y aumentar los

costos de operación de los operadores (por el aumento de los tiempos de ciclo). Para mitigar este efecto, Transmilenio ha integrado las barreras físicas (como torniquetes) a las zonas de prepago. Esto facilita la labor de fiscalización de los inspectores que custodian las zonas de prepago y evita el aumento de los tiempos de subida de los pasajeros.

2.2.4 Estrategia comunicacional

La costo-efectividad de las campañas comunicacionales en los niveles de evasión no ha sido probada en la literatura. Esto requeriría aislar muchos factores que cambian sistemáticamente a lo largo del tiempo y que inciden significativamente en la evasión tales como el precio del pasaje, las tasas de inspección y el nivel de servicio. No obstante, la comunicación efectiva es un elemento clave para cambiar la percepción de los usuarios y transmitir la idea de que pagar el pasaje es un medio para garantizar la auto-sustentabilidad financiera del sistema (Reddy et al., 2011; Torres-Montoya, 2014). Las campañas de comunicación deben transformar los sentimientos de irritación, indignación y desconfianza que pueden inducir a algunos ciudadanos a no pagar el pasaje, a sentimientos de apreciación, satisfacción y justicia (Sección 2.1.1). Las estrategias legales y los programas de incentivos económicos pueden tener una gran eficacia cuando van acompañados por mejoras en la percepción del sistema (Torres-Montoya, 2014).

El éxito de la estrategia comunicacional también depende de la credibilidad que los ciudadanos otorguen a las instituciones que las diseñan. Esto requiere que las autoridades sean asertivas y procuren que el mensaje de las campañas sea diseñado asumiendo las responsabilidades mediáticas (Reddy et al., 2011). Para apoyar la prevención de la evasión, Transmilenio ha fomentado la creación de Comités de Seguridad Ciudadana, compuestos por juntas de vecinos, agentes de tránsito (inspectores), policías y representantes de establecimientos educacionales. Además, la autoridad ha generado programas preventivos de la evasión, mediante visitas a los colegios y charlas educativas a la población (Castillo, 2010).

2.2.5 Rol del conductor

En la literatura se recomienda que las agencias generen condiciones laborales adecuadas para que los conductores puedan cumplir su rol fiscalizador, procurando que exista un contexto situacional que facilite su interacción con los pasajeros (Bonfanti y Wagenknecht, 2010; Torres-Montoya, 2014). Es importante que exista suficiente proximidad para que se genere contacto visual con los pasajeros (Blass, 1999) y que los conductores cuenten con facultades que les permitan ser considerados una autoridad. Las políticas públicas deben estar dirigidas a proteger la integridad física y psicológica de los conductores, especialmente en sistemas de transporte público donde la relación entre conductores y pasajeros es conflictiva. En estos casos, los conductores sienten que ponen su integridad física en peligro al tomar el rol fiscalizador (Bonfanti y Wagenknecht, 2010). Los conductores podrían necesitar incentivos económicos para motivar su rol fiscalizador, lo cual probablemente requeriría cambios en los contratos establecidos entre las agencias y los conductores (Torres-Montoya, 2014). Sin embargo, las modificaciones a los contratos de los conductores deberían ser revisadas con cuidado, para evitar impactos indeseados en la operación del sistema (Díaz et al., 2006).

2.3 Enfoques matemáticos para modelar la evasión en el transporte público

Las políticas para enfrentar la evasión en la mayoría de los sistemas de transporte público en el mundo, han estado históricamente focalizadas en aumentar la inspección y la recaudación de multas a los infractores (Killias et al., 2009). De la misma forma, gran parte de los estudios desarrollados en la literatura se han focalizado en el perfeccionamiento de los mecanismos de inspección mediante la formulación de modelos de optimización para asignar inspectores a lo largo del día y en las zonas donde la fiscalización es más costo-efectiva. Se han propuesto tres enfoques para encontrar los niveles óptimos de inspección (Barabino et al., 2013):

- 1) Asociar la probabilidad de que el evasor sea capturado y el monto de la multa
- 2) Maximizar las utilidades de las agencias de transporte público
- 3) Teoría de juegos

El primer enfoque ha sido desarrollado por Polinsky y Shavell (1979), Kooreman (1993) y Bootheway (2009). Polinsky y Shavell (1979) asumen que todos los individuos son idénticos, excepto por la diferencias en los beneficios derivados por viajar sin pagar el pasaje. Kooreman (1993) agrupa a los pasajeros de acuerdo a su actitud frente al riesgo y su elección por evadir el pasaje, distinguiendo entre pasajeros neutrales al riesgo, aversos al riesgo y amantes del riesgo. Bootheway (2009) investiga el problema relacionado con la definición de una multa óptima, destacando que este valor depende de la elasticidad de la demanda al monto de la multa. Este autor también sugiere que si la multa por evadir sube, algunos usuarios podrían decidir evadir siempre la tarifa y no pagar la multa a futuro si son capturados.

El enfoque de maximizar las utilidades de las agencias de transporte público fue iniciado por Boyd et al. (1989). Estos autores usaron un modelo en el cual se definía una relación entre las utilidades obtenidas por las agencias y el nivel de inspección, expresada como una función del número de inspectores en servicio. Para el caso donde existe neutralidad al riesgo de los potenciales evasores, los autores empleaban dos funciones que representaban la distribución de las probabilidades subjetivas (percibida por los pasajeros) y objetivas (tasa de inspección real) de las probabilidades de ser inspeccionado, con el objetivo de estimar el número óptimo de inspectores. En esta misma línea, Thorlacius et al. (2010) presentan un modelo de optimización de programación entera para maximizar los beneficios netos de un equipo de inspectores, utilizando como función objetivo la maximización de las diferencias entre los ingresos por las multas recaudadas y los gastos incurridos para obtener esos ingresos. Finalmente, en base a información histórica de los periodos del día en que las multas fueron cursadas y los horarios de trabajo de los inspectores en el sistema de transporte público de Copenhague, Dinamarca, este modelo permitió generar una nueva asignación de inspectores a lo largo de los periodos del día que maximizaba las multas recaudadas.

El enfoque de teoría de juegos fue iniciado por Avenhaus (2004). Este autor abordó el problema del nivel óptimo de inspección, simulando la situación donde un inspector valida si un pasajero respeta la norma de pagar el pasaje. Posteriormente, Borndörfer et

al. (2012) presentaron un modelo de teoría de juegos que estudia, desde un punto de vista teórico, el problema de asignar inspectores a lo largo de una red de transporte. El problema se resuelve mediante un modelo de programación mixta, usando como función objetivo la disminución del número de evasores. A diferencia de trabajos anteriores (Thorlacius et al., 2010), este modelo incluye elementos espaciales y características de la topología de la red de transporte, y relaja el supuesto que la tasa de evasión no varía con el nivel de inspección. A pesar que el modelo no fue formulado directamente para el caso de evasión en el transporte público, este estudio ha sido relevante para el desarrollo de trabajos posteriores, que mediante teoría de juegos, sí han abordado este problema particular (Correa et al., 2014).

Un estudio reciente plantea una gran crítica a todos estos enfoques, principalmente por la falta de evidencia empírica para respaldar las teorías desarrolladas (Barabino et al., 2013). Los autores destacan que los estudios previos se focalizan en modelar el comportamiento de los evasores pero no consideran otros problemas que enfrentan las agencias de transporte público (por ejemplo, no se incluyen todos los costos asociados a la inspección). Por otra parte, se cuestiona la idea convencional de “caja negra”, en la cual se asume que las agencias de transporte público aceptan de forma pasiva la aplicación de los modelos. Según los autores, esto explicaría por qué estos trabajos no han sido aplicados por las agencias en la práctica.

En este contexto, la contribución principal de Barabino et al. (2013) consiste en formular un modelo para encontrar el nivel de inspección óptima en la red de transporte, que integra la teoría microeconómica del productor para la estimación de los beneficios (segundo enfoque) y la teoría microeconómica del consumidor para modelar las elecciones de los usuarios en cuanto a evadir o no evadir el pasaje (primer enfoque). El modelo es estimado y testeado utilizando datos de un operador en el área metropolitana de Cagliari, Italia, y el estudio detalla claramente como el modelo puede ser calibrado y mejorado a partir del conocimiento que tiene cada agencia sobre la operación del sistema de transporte público que administra.

Dentro de la línea de estudios dirigidos a modelar el comportamiento individual del evasor, hay un trabajo reciente desarrollado por Buccioli et al. (2013), que incluye variables relacionadas con las características sociodemográficas de los pasajeros. Este estudio explora el alto nivel de evasión (43%) en el sistema de transporte público en Reggio Emilia, Italia, mediante la estimación de un modelo de elección discreta (*probit* binario). La variable dependiente fue definida como una variable binaria que tomaba el valor uno si se observaba que la persona no pagaba el pasaje y cero si no, y el conjunto de variables explicativas incluía el género, edad y el nivel de educación de los pasajeros, entre otros. El estudio encontró que los individuos más jóvenes, los hombres, los inmigrantes no europeos, los viajeros ocasionales y los desempleados tenían mayor propensión a evadir (*ceteris paribus*). Por otra parte, los resultados revelaron que viajar con familiares disminuía la probabilidad de evadir el pasaje, mientras que viajar con amigos no tenía incidencia en evadir (excepto en el caso de estudiantes de colegio).

Después de esta revisión bibliográfica, se puede constatar que la literatura en evasión y el foco de las agencias ha estado en el diseño de políticas orientadas a la inspección. A pesar que los estudios han ido enriqueciendo los modelos de comportamiento individual de los evasores, las variables incluidas en la función de utilidad han estado siempre vinculadas a elementos económicos, tales como la aversión al riesgo, el beneficio esperado de evadir y el costo de la multa. Esta concepción es útil para obtener soluciones analíticas para la formulación de modelos microeconómicos y ofrece una herramienta práctica para cuantificar los beneficios obtenidos por la implementación de estrategias de inspección. Sin embargo, el apego evidente a la tradición clásica del “homo-economicus” es peligrosa, porque limita la mirada del comportamiento individual de los evasores y omite la existencia de dispositivos psicológicos, afectivos y sociales que están presentes en la concepción del acto ilícito (Sección 2.1).

3 MODELOS DE REGRESION LINEAL Y DE CONTEO: ESPECIFICACION Y SUPUESTOS

Los modelos de regresión son una metodología ampliamente usada para modelar datos estadísticos en temas de transporte y dentro de otras áreas de alta relevancia en política pública (Osgood, 2000; Shields y Huang, 1995). Estos modelos permiten definir una relación analítica entre el valor esperado de una cierta variable, llamada variable dependiente (o explicada), y un conjunto de variables independientes (o explicativas). Cada variable independiente tiene asociado un ponderador conocido como parámetro estimado, que cuantifica el impacto relativo de la variable en el valor esperado de la variable dependiente. Además, para representar la desviación entre el valor observado y esperado de la variable dependiente se utiliza un término de error aleatorio. Dependiendo del tipo de relación funcional entre la variable dependiente y las variables explicativas, la distribución de probabilidades de los errores y los supuestos estadísticos que se definan, se derivan diferentes familias de modelos de regresión (Greene, 2007).

3.1 Modelos de regresión lineal múltiple (RLM)

Los modelos de regresión más utilizados son los modelos de regresión lineal múltiple (RLM). Los modelos RLM asumen una relación de dependencia lineal entre el valor esperado de la variable dependiente y el conjunto de parámetros estimados asociados a las variables explicativas (Greene, 2007):

$$Y = X\beta + \epsilon \quad (3.1)$$

donde Y es el vector de variables dependientes, X la matriz de variables independientes, ϵ el vector de errores aleatorios (o residuos no observados) y β el vector de parámetros estimados. Normalmente la ecuación de regresión (Ecuación (3.1)) contiene un intercepto asociado a la primera columna (de unos) de la matriz X .

La validez de los resultados obtenidos con los modelos RLM depende del cumplimiento de sus supuestos teóricos. Así, cuando la distribución de densidad de probabilidad (FDP) de la variable dependiente corresponde a un proceso de conteo, los supuestos de RLM suelen ser violados debido a la gran asimetría y curtosis de la distribución de los datos

(Cameron y Trivedi, 1998). En estos casos, la variable dependiente es discreta, está restringida a valores positivos y el valor esperado de la FDP condicional de la variable dependiente no corresponde necesariamente a una combinación lineal de los regresores (Cameron y Trivedi, 1998).

3.2 Modelos de regresión de conteo

En la literatura existe una amplia gama de modelos de regresión para trabajar con procesos de conteo y cada uno define un determinado conjunto de supuestos (Hilbe, 2011). Los modelos de regresión de conteo suelen ser estimados como modelos lineales generalizados (GLM, por el inglés *Generalized Linear Models*), empleando el método de máxima verosimilitud (MLE, por el inglés *Maximum-Likelihood-Estimation*). En cambio, los modelos RLM normalmente son estimados por mínimos cuadrados ordinarios (OLS, por el inglés *Ordinary Least Squares*). Aunque algunos autores consideran una desventaja que los modelos de regresión de conteo no tengan indicadores de ajuste robustos (como el coeficiente de determinación R^2), usar modelos de RLM en casos donde el fenómeno es un proceso de conteo puede generar problemas de inconsistencia y sesgo en los valores de los parámetros estimados (Osgood, 2000). Para estimar un modelo como GLM, se asume que la FDP de la variable dependiente pertenece a la familia exponencial (Normal, Binomial, Poisson, entre otras) y se define una función de enlace (por el inglés *link function*) que permite relacionar el valor esperado de la FDP de la variable dependiente con un conjunto de predictores lineales (Hilbe, 2011).

3.2.1 Regresión de Poisson

La regresión de Poisson, basada en la FDP de Poisson, es el método más utilizado para modelar datos obtenidos de un proceso de conteo (Hilbe, 2011). En la forma canónica del modelo de Poisson, la FDP (Ecuación (3.2)) y la función de enlace (Ecuación (3.3)), están dadas por las siguientes ecuaciones:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (3.2)$$

$$\ln(\lambda_i) = \sum_{k=0}^K \beta_k x_{i,k} \quad (3.3)$$

donde y_i representa el número de conteos medidos en la observación i , λ_i es el valor esperado de la FDP de Poisson para la observación i , β_k el parámetro estimado asociado a la k -ésima variable explicativa, y $x_{i,k}$ el valor que toma la k -ésima variable explicativa en la observación i . La forma funcional de la variable dependiente (logaritmo natural) permite limitar la variable dependiente a valores positivos y que cada coeficiente de regresión (en el caso de las variables categóricas o con forma funcional lineal) refleje las diferencias proporcionales entre los valores esperados de la FDP de Poisson producida por un cambio unitario en la variable explicativa correspondiente a cada parámetro (Osgood, 2000).

Una de las propiedades principales en los modelos de Poisson es que la varianza condicional de la variable de respuesta es igual a su media condicional. Sin embargo, esta condición no se cumple cuando la información presenta sobredispersión (del inglés *overdispersion*). En algunos casos la fuente de sobredispersión está asociada a una dependencia de la probabilidad de ocurrencia entre eventos de la distribución (Cameron y Trivedi, 1998). En el modelo de Poisson, este fenómeno es conocido como contagio positivo (del inglés *positive contagion*) puesto que la probabilidad de que un evento ocurra aumenta con la ocurrencia de un evento previo (Cameron y Trivedi, 1998). En otros casos, el exceso de varianza está explicado por una alta heterogeneidad de la variable dependiente, lo cual se conoce como contagio espurio o dependencia aparente (Cameron y Trivedi, 1998). Cualquiera sea la causa de la sobredispersión, esta suele causar problemas con la inferencia estadística en modelos de regresión de Poisson. Como los errores estándar son subestimados (más pequeños de lo que deberían ser), algunos parámetros estimados pueden aparentar ser significativos cuando en realidad no lo son (Hilbe, 2011).

3.2.2 Regresión binomial negativa

La regresión binomial negativa es una de las alternativas más utilizadas cuando hay sobredispersión en los datos, especialmente cuando no se conoce su causa exacta. A pesar que hay muchos métodos para resolver el problema de sobredispersión, diversos estudios han demostrado que la aplicación del modelo de regresión binomial negativa en la mayoría de los casos es satisfactoria (Hilbe, 2011). Sin embargo, como la forma canónica del modelo de regresión binomial negativa (NB-C) utiliza una función de enlace Logit, no se puede contrastar la existencia de dispersión positiva en el modelo de Poisson (Hilbe, 2011). Por esta razón, se utiliza una especificación¹ alternativa del modelo (NB2) que utiliza una FDP de la familia binomial negativa (Ecuación (3.4)), pero que al igual que el modelo de regresión de Poisson, tiene una función de enlace logarítmica (Ecuación (3.5)).

$$P(Y_i = y_i | \mu_i, \alpha) = \binom{y_i + \frac{1}{\alpha} - 1}{\frac{1}{\alpha} - 1} \left(\frac{1}{1 + \alpha\mu_i} \right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i} \right)^{y_i} \quad (3.4)$$

$$\ln(\mu_i) = \sum_{k=0}^K \beta_k x_{ik} \quad (3.5)$$

La Ecuación (3.4) representa la FDP binomial negativa en su versión estándar (NB2) condicionada en la media de los conteos observados (μ_i) en la observación i , y en el parámetro de heterogeneidad (α). Mientras que la Ecuación (3.5) presenta la función de enlace logarítmica del modelo de regresión NB2, en la cual β_k es el parámetro estimado asociado a la k -ésima variable explicativa, y x_{ik} el valor de la k -ésima variable

¹ El término especificación refiere a la forma funcional definida para la variable dependiente y el conjunto de variables explicativas incluidas en el modelo de regresión.

explicativa en la observación i . El parámetro de heterogeneidad (α) es una medida de sobredispersión del modelo de Poisson; cuando este parámetro es cero, el modelo NB2 es igual al modelo de Poisson, y a medida que el nivel de sobredispersión es mayor en el modelo de Poisson, el valor de α va aumentando (Hilbe, 2011).

3.3 Supuestos de los modelos

3.3.1 Ausencia de multicolinealidad

En las tres clases de modelos de regresión descritos previamente, es necesario testear la ausencia de altos niveles de colinealidad entre las variables explicativas. A pesar que la multicolinealidad no genera sesgos sistemáticos en la estimación, sí produce distorsiones en las varianzas estimadas para cada variable (Greene, 2007). Como los datos no experimentales nunca serán completamente ortogonales, siempre existirá algún grado de multicolinealidad y, dependiendo de su magnitud, esta afectará en menor o mayor medida las varianzas estimadas de cada parámetro (invalidando la inferencia estadística). En este sentido, el caso más común de presencia de multicolinealidad se produce cuando las variables están altamente, pero no perfectamente correlacionadas.

La literatura presenta diversos indicadores para detectar problemas de multicolinealidad (Greene, 2007), dentro de los cuales uno de los más usados es el indicador VIF (del inglés, *Variance Inflation Factor*). Este indicador se calcula para cada variable explicativa incluida en el modelo de regresión, y lo que esencialmente entrega es la varianza explicada de cada predictor como una función de la varianza de los otros predictores. En base a su definición matemática, la cota inferior del VIF es uno, y valores superiores a un cierto valor crítico son indicativos de un problema de multicolinealidad. Como no hay un consenso claro acerca del valor crítico del indicador, se utilizó el menor valor (cuatro) encontrado en la literatura (O'Brien, 2007).

Los indicadores VIF suelen ser utilizados para detectar problemas de multicolinealidad en modelos RLM pero también pueden ser usados para modelos de regresión de conteo. Como solo las variables explicativas del modelo son utilizadas para el cálculo de estos indicadores, si la especificación utilizada en el modelo RLM coincide con la

especificación de la función de enlace definida en modelo de regresión de conteo (NB2 o Poisson), el resultado del test es aplicable para ambos casos.

3.3.2 Normalidad y homocedasticidad

El supuesto base de los modelos RLM es que los errores aleatorios tengan una distribución Normal homocedástica (Greene, 2007). Esto garantiza que la distribución condicional de la variable dependiente sea Normal y que tenga varianza constante, dado un conjunto de valores tomado por los regresores (Ortúzar y Willumsen, 2011). El supuesto de normalidad puede ser contrastado estadísticamente utilizando los tests de Shapiro-Wilk y Lilliefors². La hipótesis nula de los dos test es que la muestra de residuos aleatorios proviene de una distribución Normal. El supuesto de homocedasticidad es típicamente contrastado mediante el test de Breusch-Pagan. Este test define como hipótesis nula la existencia de homocedasticidad en los errores aleatorios, verificando si su varianza es explicada por un subconjunto de las variables explicativas incluidas en la especificación del modelo RLM (Greene, 2007).

3.3.3 Ausencia de sobredispersión

Para seleccionar el modelo más apropiado entre Poisson y NB2, la literatura sugiere tres test estadísticos (Hilbe, 2011): i) Test de Wald, para contrastar la significancia estadística del parámetro de heterogeneidad α ; ii) Test de log-verosimilitud (del inglés *Log-Likelihood ratio test*), para comparar el ajuste estadístico de cada modelo y iii) Factor de dispersión de Pearson en el modelo de Poisson (valores mayores a uno sugieren la presencia de sobredispersión).

² Este test consiste en una adaptación del test de Kolmogorov-Smirnov, para el caso particular en que se prueba si la distribución de los residuos aleatorios es Normal (Greene, 2007).

El test de Wald sugiere que la muestra tiene una distribución de Poisson cuando se rechaza su hipótesis nula (α igual a 0). Como la estimación por GLM del modelo NB2 asume que α es un parámetro que solo toma valores positivos, para evaluar la significancia estadística del parámetro se utiliza un test t de una cola (con grados de libertad es igual al número de observaciones menos dos), en vez del test t de dos colas que es reportado erróneamente en algunos paquetes estadísticos (Cameron y Trivedi, 1998). Por otra parte, como el modelo Poisson está anidado en el NB2 (son idénticos cuando α es cero), los dos modelos pueden ser comparados usando un test de log-verosimilitud (test LR) con un grado de libertad (el modelo NB2 estima α como parámetro adicional). A diferencia de los dos test anteriores, el factor de dispersión de Pearson es solo un indicador de sobredispersión en el modelo de Poisson, dado que no existe un criterio estadístico para evaluar cuan lejos de uno el indicador debe estar para probar la existencia de sobredispersión (Hilbe, 2011).

4 FORMULACION, ESTIMACION Y SELECCION DEL MODELO

En este capítulo, se describe la metodología utilizada por el Programa Nacional de Fiscalización del Ministerio de Transporte y Telecomunicaciones (MTT) para recolectar los datos de evasión en terreno. Posteriormente, se presentan las variables explicativas y dependientes definidas en cada modelo de regresión y se realiza un análisis estadístico descriptivo de éstas. Luego, se explican los criterios para seleccionar el conjunto de variables explicativas incluidas en cada modelo, y a partir de lo anterior, se presentan los resultados de la estimación de algunos modelos de regresión lineal, regresión de Poisson y regresión binomial negativa. Finalmente, para seleccionar la mejor especificación, se revisa el cumplimiento de los supuestos de cada modelo de regresión, se evalúa el nivel de ajuste estadístico de los modelos y se compara su poder predictivo. Todos los modelos de regresión y test estadísticos fueron estimados con el programa computacional R Studio.

4.1 Generación del banco de datos

El Programa Nacional de Fiscalización recolecta mensualmente datos para estimar la tasa de evasión de cada operador de Transantiago (conocidas también como unidades de negocio) y del sistema completo. Cada mes se selecciona una muestra aleatoria y representativa del sistema de buses, en base a dos criterios de estratificación: las unidades de negocio (UN) del sistema y el número de pasajeros trasladado por los servicios de cada UN. El propósito del primer criterio es obtener un panel representativo de todas las UN, por lo cual se elige aleatoriamente la misma cantidad de servicios en cada UN. El segundo criterio se establece para garantizar que la muestra seleccionada sea lo más representativa posible de los servicios de cada UN. Para llevar esto a cabo, los servicios de cada UN son estratificados en base a la proporción de pasajeros que trasladan respecto al total de pasajeros trasladados en la UN respectiva. De esta forma, el conjunto total de servicios de cada UN se divide en tres estratos (o grupos); el estrato 1 incluye todos los servicios que mueven menos del 1.0% del total de pasajeros, el estrato 2 corresponde a los servicios que mueven entre un 1.0% y 5.0% y el estrato 3 está compuesto por todos los servicios que mueven más del 5.0% de los pasajeros en la UN

respectiva. Finalmente, en todos los estratos se selecciona la misma cantidad de servicios de buses.

Posteriormente, a cada servicio de buses elegido en el proceso anterior, se asignan equipos de fiscalizadores (denominados “pasajeros incógnitos”), encargados de realizar mediciones durante toda una jornada laboral en el servicio. Cada tanda de mediciones comienza en uno de los paraderos terminales (o “cabecales”) del servicio y finaliza en el siguiente paradero terminal. Una vez que el bus llega a un paradero terminal, el grupo inicia una nueva tanda de mediciones en una expedición del mismo servicio, pero que se dirige en el sentido opuesto del recorrido. Este proceso se hace de forma iterativa a lo largo del día, garantizando que existan mediciones de evasión para varios periodos del día y en los dos sentidos del recorrido (ida y vuelta).

Para efectuar las mediciones, cada fiscalizador se ubica en una de las puertas del bus, y registra de forma escrita (Figura A-1, Anexo A) la cantidad de pasajeros que entran/salen/evaden en cada paradero durante toda la duración del recorrido de una expedición (todas las variables medidas por los fiscalizadores son descritas en el Anexo A). Además, para garantizar que durante el proceso de recolección de datos no exista influencia en el comportamiento de los pasajeros y el conductor del bus, el Programa Nacional de Fiscalización instruye a los fiscalizadores para que el registro sea realizado de forma anónima. Posteriormente, digitadores traspasan las mediciones hechas en cada expedición a una planilla de datos, que contiene todos los registros de evasión efectuados durante el mes de medición.

4.2 Variable dependiente

En base a un modelo de regresión lineal múltiple y dos tipos de modelos de regresión de conteo (Poisson y binomial negativo), se testeó estadísticamente la influencia de diferentes variables en los niveles de evasión registrados en cada medición de la base de datos. Para utilizar la información de la forma más desagregada posible, cada observación del banco de datos de estimación corresponde a una medición de evasión en una puerta específica del bus y en un determinado paradero donde se detuvo la

expedición. Para estimar los modelos se seleccionó solo las mediciones recolectadas durante Octubre del año 2012. Este mes se consideró representativo de la operación del sistema debido a la baja cantidad de días festivos, el desarrollo normal de las clases en los establecimientos de enseñanza media, básica y universitaria, y la baja proporción de trabajadores que toma vacaciones.

En el modelo RLM la variable dependiente fue definida como el logaritmo de la suma entre la tasa de evasión (calculada como la división entre el número de evasores y el número pasajeros entrando por la puerta medida) y una constante igual a uno (sumada en el argumento de la función). En el Anexo C, se presentan los resultados de la estimación de tres especificaciones alternativas de modelos RLM y se argumenta la selección del mejor modelo RLM (presentado en el cuerpo de la tesis). En cada especificación se define una forma funcional diferente para la variable dependiente, pero en todos los casos se mantiene el nivel de agregación de las observaciones (mediciones por puerta) y se utiliza el mismo conjunto de variables explicativas. Por último, en los dos modelos de regresión de conteo (Poisson y NB2), la variable dependiente fue definida como la cantidad de personas que evaden por una puerta del bus, en un cierto paradero e instante del día.

4.3 Variables explicativas

A continuación, se presentan las variables explicativas utilizadas en la formulación de los modelos. Para facilitar el análisis, las variables fueron divididas en seis categorías: operación de las puertas, variables temporales, calidad del servicio, operación de la ruta, tipo de paradero y nivel socioeconómico (NSE). En el Anexo A se presenta una descripción de todas las variables disponibles en la base de datos.

4.3.1 Operación de las puertas

Naturalmente, la cantidad de evasiones observadas en una puerta aumenta con el número de personas que entra al bus. A medida que más gente entra por una puerta, aumenta la posibilidad de que se detecte evasión en esa puerta. En la literatura de los modelos de regresión de conteo, este efecto es conocido como “efecto de exposición”, y para

controlarlo, se agrega como variable explicativa el logaritmo natural de la “variable de exposición” (Hilbe, 2011). Tanto en el modelo de regresión de Poisson como en el NB2, la variable de exposición corresponde al número de pasajeros entrando por la puerta medida. Así, para capturar el efecto de exposición, la variable ***Log(Entradas)*** fue definida como el logaritmo del número de pasajeros que entra por la puerta donde la observación fue recolectada. En el modelo RLM, la información correspondiente al número de entradas de pasajeros por una puerta fue utilizada para definir la variable dependiente del modelo (logaritmo de la suma entre uno y la tasa de evasión) y, por consiguiente, esta variable no fue incluida en el conjunto de predictores.

Los operadores de Transantiago instruyen a los conductores de buses para controlar la operación de las puertas. A pesar que los pasajeros deberían entrar solo por la puerta delantera y salir por las puertas traseras, los datos recolectados indican que la norma no siempre se cumple (ver Tabla 4-8 más adelante). Para capturar el efecto del número de pasajeros saliendo por una puerta del bus en el número de evasores contabilizado en esa puerta, inicialmente se agregó la variable ***Salidas***, definida como el número de personas que salen por la puerta analizada. Sin embargo, la variable resultó estadísticamente significativa (con un nivel de confianza del 95%) solo cuando se incluyó como una variable muda, tomando el valor uno si salen pasajeros por la puerta, y cero en otro caso (ningún pasajero saliendo por esa puerta).

En Transantiago existen diferentes tipos y tamaños de buses operando. El rango va desde grandes buses articulados de cuatro puertas hasta buses pequeños de solo dos puertas. Para evaluar si el número de puertas influye en la evasión, se definió la variable ***Puertas*** como el número de puertas del bus. Por otra parte, para capturar la diferencia de la tasa de evasión promedio en cada puerta del bus (y no explicada por el resto de las variables), inicialmente se utilizaron tres variables mudas (como el máximo número de puertas de un bus es cuatro, se pudo incluir hasta tres variables). Posteriormente, en base a la significancia estadística de los parámetros y el nivel de ajuste estadístico obtenido con los modelos de regresión (bajo diferentes agrupaciones de variables), se incluyó la

variable ***Puerta Delantera***, que toma el valor uno si la observación fue tomada en la puerta delantera del bus, y cero en otro caso (segunda, tercera y cuarta puerta).

4.3.2 Variables temporales

Transantiago define 12 periodos dentro de los días laborales (Tabla 4-1). A partir de esta clasificación de horarios y la hora en que fue tomada la medición, se utilizó un algoritmo para asignar el bloque horario correspondiente a cada observación. Posteriormente, los periodos del día fueron reagrupados en base al nivel de ajuste estadístico de los modelos y la significancia estadística de los parámetros. Como el banco de datos solo contiene mediciones realizadas entre las 6:30 y las 22:00, los periodos del día se reagruparon en cuatro bloques: Punta Mañana (6:30 - 8:29), Fuera Punta Mañana (8:30 - 12:29), Tarde (12:30 - 20:29) y Noche (20:30 - 21:29). Luego, en base a esta agrupación de periodos, se definieron tres variables mudas para capturar las diferencias promedio en los niveles de evasión de cada periodo del día (*ceteris paribus*); ***Fuera Punta Mañana***, ***Tarde*** y ***Noche***, que toman el valor uno si la observación fue tomada en ese horario y cero en otro caso (el horario punta mañana se dejó como referencia).

Tabla 4-1 Clasificación de los periodos del día en Transantiago

Horario	Inicio	Fin	Bloque
PRENOC1	0:00	0:59	-
NOC	1:00	5:29	-
TNOC	5:30	6:29	-
PMA	6:30	8:29	1
TPMA	8:30	9:29	2
FPMA	9:30	12:29	2
PMD	12:30	13:59	3
FPTA	14:00	17:29	3
PTA	17:30	20:29	3
TPTA	20:30	21:29	4
FPNOC	21:30	22:59	*
PRENOC2	23:00	23:59	-

Nota: Punta Mañana (1); Fuera Punta Mañana (2); Tarde (3); Noche (4).

(-) No hay mediciones tomadas en este periodo

(*) Las observaciones de este periodo fueron eliminadas (Sección 4.4)

Fuente: (DTPM, 2013a)

Como la base de datos fue recolectada en diferentes fechas entre Martes y Jueves de Octubre de 2012 (Sección 4.4), se pudo crear sólo dos variables mudas para capturar este efecto, *Miércoles* y *Jueves*, que toman el valor uno si la observación fue medida el día Miércoles o Jueves, respectivamente, y cero en otro caso (el día Martes fue definido como referencia). Además, como solo se usó información de un mes, no se pudo crear variables para capturar las diferencias en los niveles de evasión a lo largo de los meses y años de la operación de Transantiago (no explicadas por las demás variables incluidas en el modelo).

4.3.3 Calidad del servicio

La base de datos incluye un índice de ocupación de cinco categorías, dependiendo de la proporción de pasajeros sentados y parados en el bus (Tabla 4-2)..

Tabla 4-2 Categorías usadas para medir el índice de ocupación (IO) del bus

IO	Descripción
A	Bus completo, pasajeros casi “colgando”
B	Más de la mitad del pasillo con pasajeros de pie
C	Menos de la mitad del pasillo con pasajeros de pie
D	Más de la mitad de los asientos ocupados
E	Menos de la mitad de los asientos ocupados

Fuente: Programa Nacional de Fiscalización (2012)

Notar que esta clasificación no es perfecta dado que está sujeta a la percepción de los fiscalizadores. Para medir el efecto del nivel de ocupación del bus en el nivel de evasión se definió la variable muda *Baja Ocupación*, que toma el valor uno si el índice de ocupación es C, D, E y cero en otro caso (A y B). De esta forma, un bus tiene baja ocupación cuando menos de la mitad de su pasillo tiene pasajeros parados (independiente de cuántos pasajeros estén sentados) y alta ocupación cuando más de la mitad del pasillo tiene gente parada. La agregación de los índices de ocupación para crear la variable *Baja Ocupación* fue definida en base a la significancia estadística de los parámetros estimados y el nivel de ajuste estadístico obtenido con los modelos de regresión.

La frecuencia del servicio de buses percibida por los pasajeros también parece explicar los niveles de evasión registrados en cada paradero. Sin embargo, esta variable no estaba disponible en la base de datos proporcionada por el Programa Nacional de Fiscalización (Anexo A). Dado el alto nivel de cumplimiento de frecuencia de todos los operadores de Transantiago durante 2012 (superior al 90%), se asumió que la frecuencia teórica definida en el programa de operación era una buena *proxy* de la frecuencia real percibida por los usuarios (DTPM, 2013b). El programa de operación de Transantiago define el conjunto de paraderos que debe recorrer cada servicio de buses, y especifica la comuna y la coordenada de localización geográfica de cada paradero (DTPM, 2012). Además, para cada servicio de buses se establecen indicadores operacionales que varían según el periodo del día (Tabla 4-1), el día de la semana y el sentido del recorrido, tales como la frecuencia promedio (buses/hora), la capacidad de la ruta (buses/hora) y la velocidad promedio (km/hora). Luego, dependiendo de la media y la variabilidad de estos indicadores operacionales (e.g. regularidad de la frecuencia), los operadores pueden recibir penalizaciones económicas por parte de Transantiago.

Para vincular la frecuencia teórica definida en el plan operacional de Transantiago, con cada observación de la base de datos proporcionada por el Programa Nacional de Fiscalización, se utilizó la información correspondiente al día, hora y minuto en que fue hecho el registro y el sentido del servicio de buses que fue medido (ida o vuelta). Como el sentido del recorrido no es digitado en la base de datos de evasión (solo se registra el servicio de buses), se digitó manualmente, asociando la dirección del paradero de inicio que registraban los fiscalizadores en la planilla de medición (Figura A-1, Anexo A) con las direcciones de la lista de paraderos definidos en el programa de operación de Transantiago para cada servicio-sentido de buses (DTPM, 2012). Luego, se creó un algoritmo para asignar la frecuencia teórica a cada observación de la base de datos, según el periodo del día y el servicio-sentido en el que se hizo la medición de evasión.

Para evaluar el impacto de la frecuencia en el modelo de regresión, se probaron dos especificaciones alternativas; agregando la frecuencia como una variable lineal y como su recíproco. En el primer caso, la variable fue denominada ***Frecuencia*** (buses/hora), y

en el segundo caso, *Headway* (minutos). El recíproco de la frecuencia tiene un significado teórico dentro de la literatura de transporte, su valor está relacionado con el tiempo entre pasadas de buses (conocido como *headway*) y los tiempos de espera percibidos por los usuarios. Finalmente, la bondad de ajuste de los modelos fue muy superior con la variable *Headway* (Tabla E-1, Anexo E).

4.3.4 Operación de la ruta

El sistema de buses en Transantiago cubre 10 zonas y cada una está compuesta por diferentes municipalidades. La red de Transantiago fue inicialmente diseñada como un sistema de servicios troncales y alimentadores, donde el Metro es un elemento central de la red troncal. Los servicios troncales operan en grandes corredores de buses y permiten viajar entre diferentes zonas del sistema. En cambio, los servicios alimentadores operan solo dentro de una zona del sistema y su función principal es conectar a los usuarios con el sistema de buses troncales y las líneas de Metro (Muñoz et al., 2008). Los códigos de los servicios troncales están compuestos por un número de 3 dígitos, que identifica a un servicio particular de las líneas de buses 100, 200, 300, 400 y 500 de Transantiago. En los servicios alimentadores, el primer dígito representa la zona del sistema y los siguientes dos al servicio específico dentro de la zona.

Tabla 4-3 Características de la operación de las unidades de negocio de Transantiago

Unidad de Negocio	Empresa	Zonas	Vehículos	Servicios	Grupo ^a
1	Alsacia	100	715	35	1
2	Subus	200 / G	1.236	63	3
3	Vule	300 / E / H / I	1.241	83	1
4	Express	400 / D	1.261	59	2
5	Metbus	500 / J	859	49	2
6	Redbus	B / C	590	55	3
7	STP	F	396	30	4
Total	7	-	6298	374	-

^a Grupos definidos para crear las variables explicativas asociadas a los operadores de Transantiago.

Fuente: (DTPM, 2013b)

Desde Junio de 2012, los servicios de Transantiago han sido manejadas por siete operadores (o unidades de negocio); Alsacia, Subus, Vule, Express, Metbus, Redbus y STP. Inicialmente, se crearon seis variables mudas para controlar posibles diferencias promedio en la evasión entre las siete agencias, y que no estuvieran capturadas por las demás variables explicativas incluidas en los modelos. Posteriormente, en base a criterios de significancia estadística de los parámetros estimados y el nivel de ajuste de los modelos obtenido bajo distintas agrupaciones de las unidades de negocio, se definieron cuatro grupos de operadores. El primero incluye a Alsacia y Vule, el segundo a Express y Metbus, el tercero a Subus y Redbus y el cuarto solo al operador STP. Finalmente, se crearon tres variables mudas: ***Operadores I***, ***Operadores II*** y ***Operadores III***, que toman el valor uno si la observación fue tomada en el primer, segundo o tercer grupo de operadores respectivamente, y cero en otro caso (el operador STP fue definido como referencia).

Finalmente, para evaluar si existe algún efecto asociado a las características del tipo de servicio de buses, se definió la variable muda ***Alimentador***, que toma el valor uno si la observación fue tomada en un servicio alimentador, y cero en otro caso (servicio troncal).

4.3.5 Tipo de paradero

En la base de datos, los paraderos de buses son clasificados de dos formas; paraderos primarios (P) y zonas pagas (ZP). Inicialmente, a partir de la información sobre el tipo de paradero registrado por los fiscalizadores, se creó la variable muda ***Zona Paga***, que tomaba el valor uno si el paradero estaba ubicado en una ZP y cero si no. Esta variable buscaba medir el impacto de las ZP en los niveles de evasión de ese paradero (por efectos no considerados por el resto de las variables del modelo). Desafortunadamente, la evasión registrada por los fiscalizadores en estos paraderos no es medida correctamente. Como el pago del pasaje en las ZP es realizado antes que el bus llegue al paradero, los fiscalizadores no pueden detectar si alguien no pagó su pasaje antes de abordar el bus. Por esto, son instruidos por el Programa Nacional de Fiscalización para que registren cero evasiones en los paraderos con ZP.

Para comprobar si los fiscalizadores cumplían con el criterio establecido para medir evasión en paraderos con ZP, se realizó un análisis exhaustivo de la base de datos y se detectaron inconsistencias en los registros. Por ejemplo, en algunas observaciones medidas en zonas pagas, se reportaba erróneamente que nadie pagó, mientras que en otro casos se registraba algo de evasión (estas mediciones podrían haber sido hechas fuera del horario de operación de la ZP). Finalmente, considerando la gran cantidad de observaciones disponibles en la base de datos, se decidió no incluir las mediciones hechas en paraderos con ZP, y por consiguiente, no se incluyeron variables para capturar el efecto del tipo de paradero en los modelos de regresión.

4.3.6 Nivel socioeconómico (NSE)

La base de datos entregada por el Programa Nacional de Fiscalización no cuenta con información asociada al lugar donde fue tomada la observación, o del nivel socioeconómico de los individuos que suben al bus. Por esta razón, para probar si el NSE tiene alguna influencia en los niveles de evasión, se desarrollaron dos métodos para inferir el área geográfica (comuna o zona) donde está ubicado cada paradero. Posteriormente, a partir del nivel de ingreso promedio de los hogares en el área, se asoció una variable *proxy* del NSE a cada observación.

En el primer método, se intentó estimar el impacto del NSE en la evasión en base al nivel de ingreso de la zona de Transantiago donde fue tomada la observación, detectándose varios problemas que son discutidos en detalle en el Anexo D. En primer lugar, en el caso de los servicios troncales no existe información de la zona donde fue tomada la medición en la base de datos y, por ende, no se puede analizar los efectos del NSE en estos servicios. Además, las zonas de Transantiago son unidades geográficas muy agregadas, por lo cual no es posible caracterizar con precisión el área donde viven los individuos. Por otra parte, como todas las rutas alimentadores de una zona del sistema son operadas por una sola unidad de negocio (Tabla 4-3), al incluir variables para capturar el efecto del NSE en base a una estratificación por ingreso de las zonas del sistema, se genera una alta colinealidad con las variables asociadas al tipo de servicio (*Alimentador*) y los operadores.

Para corregir estos problemas, el segundo método utiliza un algoritmo para estimar la municipalidad donde cada observación fue recolectada. A diferencia del primer método, se determina una ubicación geográfica para cada observación, independiente de si esta fue medida en un servicio troncal o alimentador. El algoritmo utiliza información disponible en una base de datos pública correspondiente al programa operacional de Transantiago del segundo semestre de 2012 (DTPM, 2012), que contiene el listado de todos los servicios de buses del sistema y las características de cada uno de sus paraderos. En una primera etapa, se construye un vector para cada servicio-sentido con la proporción de paraderos que están ubicados en cada comuna. En una segunda etapa, cada vector es ordenado secuencialmente en base al sentido del servicio y cuál es la municipalidad que se visita primero. Finalmente, el algoritmo replica proporcionalmente el orden en el cual las municipalidades son visitadas por cada servicio de bus³ dentro de la lista completa de observaciones recolectadas por cada fiscalizador en una tanda determinada de mediciones (las observaciones recolectadas por cada fiscalizador son ordenadas secuencialmente según la hora de la medición). Por ejemplo, si la mitad de los paraderos de la ruta medida están ubicados en la municipalidad A, el algoritmo asume que la primera mitad de las observaciones de esa tanda de mediciones fue recolectada en esta municipalidad. Finalmente, el algoritmo es aplicado al conjunto completo de observaciones medidas en cada expedición, con lo cual cada observación queda asociada a alguna municipalidad (Tabla 4-4).

³ Los fiscalizadores solo registran observaciones en los paraderos donde el bus se detiene. Por lo tanto, no es posible asociar secuencialmente la lista de paraderos recorridos por un servicio de buses en la red real, con el orden en que fueron recolectadas las mediciones.

Tabla 4-4 Nivel de ingreso familiar promedio en comunas de Transantiago 2009

Comuna	Ingreso	Categoría de Ingreso	Zona
Cerrillos	670.134	II	I
Cerro Navia	456.527	I	J
Conchalí	581.024	I	B
El Bosque	611.632	II	G
Estación Central	742.263	II	I
Huechuraba	745.901	II	B
Independencia	822.374	II	B
La Cisterna	794.021	II	G
La Florida	818.682	II	E
La Granja	465.848	I	E
La Pintana	420.204	I	G
La Reina	2.276.947	III	D
Las Condes	3.245.018	IV	C
Lo Barnechea	3.242.710	IV	C
Lo Espejo	57.589	I	H
Lo Prado	617.056	II	J
Macúl	573.758	I	D
Maipú	973.887	III	I
Nuñoa	1.249.592	III	D
Pedro Aguirre Cerda	520.351	I	H
Peñalolén	792.344	II	D
Providencia	2.270.405	III	C
Pudahuel	774.621	II	J
Puente Alto	580.084	I	F
Quilicura	676.411	II	B
Quinta Normal	735.100	II	J
Recoleta	595.663	I	B
Renca	550.915	I	B
San Bernardo	567.397	I	G
San Joaquín	606.080	II	H
San Miguel	1.023.307	III	H
San Ramón	527.209	I	G
Santiago	1.217.787	III	A
Vitacura	3.065.842	IV	C

Fuente: (BCN, 2012)

Posteriormente, para capturar el efecto del NSE en los modelos de regresión, las municipalidades se dividieron en cuatro grupos según el nivel de ingreso mensual de los hogares en cada municipalidad (Tabla 4-4). Para definir los tramos de ingreso en cada

categoría, se utilizaron datos sobre el ingreso promedio por quintil de hogares disponible en la encuesta de Caracterización Socioeconómica Nacional (CASEN) realizada en el año 2009 (MDS, 2009). De acuerdo con esta encuesta, los ingresos promedio mensuales por quintil en el área urbana de Santiago son: \$184.920, \$382.597, \$596.623, \$937.731 y \$2.875.997, respectivamente (Tabla 4-5). Como el ingreso promedio de todas las municipalidades es siempre mayor a los primeros dos quintiles (Tabla 4-4), solo fue posible crear cuatro grupos de municipalidades. El primero corresponde a municipalidades con un ingreso familiar promedio menor a \$596.623, el segundo incluye a municipalidades con un ingreso familiar promedio entre \$596.623 y \$937.731, el tercero a las municipalidades con un ingreso familiar promedio entre \$937.732 y \$2.875.997, y el cuarto cubre a todas las municipalidades con un ingreso familiar promedio superior a \$2.875.997. Luego, se definieron tres variables mudas como *proxies* del NSE: **Ingreso Medio-Bajo**, **Ingreso Medio-Alto** y **Ingreso Alto**, que toman valor uno si la municipalidad donde las observaciones fueron recolectadas pertenece al segundo, tercero o cuarto grupo respectivamente, y cero en otro caso (el primer grupo de municipalidades fue definido como referencia).

Tabla 4-5 Características por quintil de ingreso área urbana Santiago 2009

Quintil	Ingreso	Población	Miembros por hogar
1	184.920	1.436.365	4,0
2	382.597	1.409.588	3,9
3	596.623	1.358.436	3,7
4	937.731	1.235.267	3,4
5	2.875.997	1.083.013	2,9

Fuente: (MDS, 2009)

Las diferencias entre las características socioeconómicas de los viajeros de cada municipalidad podrían explicar las diferencias en los niveles promedio de evasión entre los grupos de municipalidades. Dados los altos niveles de segregación residencial e ingreso en Santiago (Sabatini et al., 2009), el NSE de los individuos que viven en una municipalidad tiende a ser similar al ingreso promedio de esa municipalidad. No obstante, los individuos que se suben en un paradero no necesariamente viven en ese

lugar. Por esta razón, durante el horario punta mañana, los estimadores de las variables mudas asociadas al ingreso debieran capturar con mayor precisión el efecto del NSE de los pasajeros en la evasión (en este periodo los pasajeros salen de sus casas). En contraste, si el modelo de regresión es estimado con datos de todos los periodos del día, los parámetros estimados podrían ser menos precisos. En la Sección 5.1, se testearán estadísticamente estas hipótesis, estimando un nuevo modelo de regresión solo con los datos correspondientes al horario punta mañana y analizando el cambio de los parámetros asociados a las variables *proxy* del NSE.

4.4 Análisis estadístico descriptivo

La base de datos original contenía 22.259 observaciones tomadas durante cuatro días laborales en Octubre del año 2012; Jueves 18 de Octubre, Martes 23 de Octubre, Miércoles 24 de Octubre y Jueves 25 de Octubre. En total, 311 buses fueron monitoreados, de los cuáles 61,4% eran alimentadores y 38,6% troncales (equivalente al 4,9% de todos los buses que operaron en 2012). La muestra cubre 41 servicios de buses (10,9% del total de servicios de Transantiago), los siete operadores de Transantiago y las 10 zonas del sistema (Tabla 4-6). Durante los cuatro días de medición, se registraron 23.670 pasajeros abordando los buses y solo 18.726 validaron su pasaje; notar que en promedio para un día laboral el año 2012 se registraron 3.184.269 de transacciones (DTPM, 2013b).

Tabla 4-6 Subconjunto de servicios medidos de cada operador de Transantiago

Empresa	Alimentadores	Troncales
Alsacia	-	103, 105, 107, 117, 125
Subus	G01, G19, G22	201, 202, 203, 208, 228
Vule	E03, E05, H13, H15, I08, I09	301
Express	D08	405, 406, 423, 435
Metbus	J08, J19	502, 508, 514
Redbus	B16, B18, C01, C15, C18	-
STP	F01, F15, F17, F20, F22, F23	-

Fuente: Programa Nacional de Fiscalización (2012)

Para el análisis descriptivo de los datos, se utilizó una base de datos depurada con un total de 21.244 observaciones; en esta se eliminaron las observaciones duplicadas (570),

las observaciones con información faltante (176) y con otras inconsistencias básicas, tales como cuando el número de evasiones medida fue mayor que el número de personas entrando por una puerta (20) y las mediciones en que el número de la puerta registrada era más grande que el número total de puertas del bus (37). Además, no se incluyeron las observaciones medidas después de las 21:30 (6), ni las correspondientes a zonas pagas (206) (discutido en la Sección 4.3.5).

En la Tabla 4-7 se presenta la media, desviación estándar y el coeficiente de variación del conjunto de variables no categóricas.

Tabla 4-7 Estadísticas variables no categóricas, Octubre 2012

Variable	Media	Desviación Estándar	Coficiente de Variación
Evasión (pasajeros)	0,207	0,901	4,352
Entradas (pasajeros)	1,025	3,088	3,013
Salidas (pasajeros)	1,050	2,448	2,331
Puertas	2,982	0,638	4,674
Frecuencia (buses/hora)	9,263	4,377	0,473
Headway (minutos)	7,802	2,960	0,379

Fuente: Elaboración propia

En la Tabla 4-8 se presenta un cuadro de estadísticas descriptivas, tanto de las variables categóricas obtenidas a partir de la base de datos, como de las variables creadas con los algoritmos presentados en la Sección 4.3. Las columnas de la tabla representan para cada nivel de las variables categóricas, la proporción del total de las observaciones, la proporción del total de entradas de pasajeros, la proporción del total de las evasiones totales, y la tasa de evasión muestral.

Tabla 4-8 Estadísticas variables categóricas, Octubre 2012

Variable	Proporción muestra (%)	Proporción entradas (%)	Proporción evasión (%)	Tasa de evasión (%)
Puerta de entrada				
Puerta delantera	37,2	96,7	87,9	18,4
Puerta trasera 2	33,1	1,4	5,4	76,0
Puerta trasera 3	24,8	1,3	4,9	74,3
Puerta trasera 4	5,0	0,5	1,8	72,2
Periodo del día				
PMA (6:30-8:29)	13,5	18,8	20,9	22,5
FPMA (8:30-9:29)	24,2	18,5	13,4	14,7
TPMA (9:30-12:29)	8,1	8,2	6,9	17,1
PMD (12:30-13:59)	6,5	4,6	4,5	20,0
FPTA (14:00-17:29)	24,9	26,7	29,8	22,5
PTA (17:30-20:29)	20,9	22,0	22,9	21,1
TPTA (20:30-21:29)	1,9	1,2	1,4	23,4
Día de la semana				
Martes	26,0	26,2	23,1	17,8
Miércoles	23,1	25,4	26,0	20,7
Jueves	51,0	48,4	51,0	21,3
Puertas				
2	21,3	23,1	19,4	17,0
3	59,3	55,3	56,4	20,6
4	19,4	21,7	24,2	22,6
Ocupación				
A (alta)	2,8	5,5	9,3	34,2
B (alta)	8,8	15,4	17,8	23,5
C (baja)	15,8	21,0	18,1	17,4
D (baja)	19,5	20,7	21,0	20,0
E (baja)	53,1	37,4	34,3	18,5
Tipo de servicio				
Alimentador	46,4	45,7	43,2	19,1
Troncal	53,6	54,3	56,8	21,1
Operador				
Alsacia	15,7	17,0	21,5	25,5
Subus	21,1	19,9	16,1	16,3
Vule	13,5	13,4	14,7	22,2
Express	13,3	13,3	13,2	20,0
Metbus	12,9	13,3	11,2	17,0
Redbus	10,2	12,7	8,8	13,9
STP	13,3	10,4	14,5	28,6
NSE municipalidad				
Ingreso bajo	37,4	35,0	42,1	24,3
Ingreso medio-bajo	36,0	36,0	34,5	19,4
Ingreso medio-alto	17,2	18,9	17,4	18,6
Ingreso alto	9,4	10,1	6,0	12,1

Fuente: Elaboración propia

4.5 Estimación de modelos de regresión para explicar la evasión

El banco de datos final para la estimación de los modelos (5.334 observaciones) no incluye las mediciones correspondientes a casos donde nadie entra por alguna puerta del bus (15.910). Como se explicó en la Sección 4.1, cada observación de la muestra representa un conteo de cuántos pasajeros entran/salen/evaden por una puerta en un cierto paradero. Por lo tanto, algunas observaciones podrían registrar que ningún pasajero entró pero que sí algunos pasajeros salieron por esa puerta (estas variables podría también ser cero si alguna persona entró o salió por otra puerta del bus). Sin embargo, en estos casos la evasión contabilizada siempre será igual a cero (no entran pasajeros). Por consiguiente, estas observaciones no aportan información útil para la estimación de los modelos.

Posteriormente, el conjunto de variables explicativas incluidas en cada modelo de regresión fue definida en base a los siguientes criterios (Ortúzar y Willumsen, 2011); i) la consistencia de los signos de los parámetros estimados; ii) la significancia estadística de los parámetros estimados (con un nivel de confianza del 95%); iii) el nivel de ajuste estadístico de los modelos. Finalmente, los modelos incluyeron el siguiente conjunto de predictores para explicar la evasión; número de pasajeros entrando⁴ y saliendo por una puerta del bus, número de puertas del bus, puerta de entrada de los pasajeros, nivel de ocupación del bus, *headway* del servicio de buses y periodo del día. En la Tabla 4-9 se presentan los resultados obtenidos con los modelos RLM, RP y NB2, utilizando el conjunto de variables explicativas y dependientes descritas previamente.

⁴ Excepto en el modelo RLM, puesto que el número de entradas se utilizó para definir la variable dependiente (Sección 4.2).

Tabla 4-9 Estimación modelos de regresión lineal, Poisson y binomial negativa

Variable (test-t)	Lineal (RLM)	Poisson (RP)	Binomial (NB2)	VIF
Operación puertas				
Log (entradas)	-	0,976 (59,3)	1,079 (42,7)	1,105
Salidas	0,044 (3,8)	0,292 (5,5)	0,221 (2,8)	1,384
Puertas	0,032 (6,3)	0,178 (6,6)	0,210 (5,6)	1,385
Puerta Delantera	-0,391 (-24,0)	-1,105 (-18,3)	-1,336 (-14,0)	1,520
Periodo del día				
Punta Mañana (08:30-12:30)	-0,022 (-2,3)	-0,231 (-4,6)	-0,166 (-2,4)	1,978
Tarde (12:30-20:30)	0,026 (2,9)	0,156 (3,8)	0,231 (3,9)	1,958
Noche (20:30-22:00)	0,064 (2,7)	0,343 (2,6)	0,499 (2,8)	1,102
Calidad servicio				
Baja Ocupación	-0,043 (-4,8)	-0,137 (-3,6)	-0,150 (-2,6)	1,185
Headway	0,004 (3,0)	0,016 (2,6)	0,026 (2,9)	1,572
Operador				
Operadores I	-0,043 (-4,4)	-0,265 (-5,5)	-0,271 (-4,0)	2,363
Operadores II	-0,075 (-7,4)	-0,528 (-10,2)	-0,567 (-7,8)	2,261
Operadores III	-0,083 (-8,2)	-0,613 (-11,6)	-0,661 (-8,9)	2,395
Intercepto	0,485 (16,8)	-0,748 (-5,4)	-0,871 (-4,4)	-
R ² Ajustado	0,176	-	-	-
Log-verosimilitud	-	-5.831,7	-5.477,8	-
AIC	-	11.689,4	10.981,5	-
Nº de observaciones	5.334	5.334	5.334	-

^a VIF: *Variance Inflation Factor* (Sección 3.3.1)

Fuente: Elaboración propia

En los tres modelos de regresión, las variables mudas que intentaban capturar el efecto del día de la semana (*Miércoles, Jueves*) y el tipo de servicio (*Alimentador*) no resultaron significativas al 95% de confianza (Tabla E-1, Anexo E). La baja significancia estadística de esos parámetros podría indicar que las otras variables del modelo están capturando correctamente las diferencias promedio de los niveles de evasión observadas entre los días Martes, Jueves y Viernes, y entre los servicios alimentadores y troncales (Tabla 4-8).

Si se utilizara información recolectada durante el fin de semana para estimar los modelos, las variables para capturar el efecto del día de la semana podrían ser

estadísticamente significativas. Durante el fin de semana, algunos factores que contribuyen a la evasión tales como los patrones de viaje de los pasajeros y los niveles de inspección (Buccioli et al., 2013), cambian significativamente respecto a los días laborales, y estos efectos no están siendo capturados por las demás variables explicativas incluidas en los modelos de regresión.

4.6 Cumplimiento de los supuestos de los modelos

La evasión en el transporte público ha sido un tema poco estudiado en la literatura, y pocos trabajos han desarrollado una aproximación econométrica para analizar el fenómeno (Capítulo 2). En este contexto, la inferencia estadística es fundamental, tanto para determinar un conjunto de variables que explique significativamente los niveles de evasión reportados en la base de datos, como para contrastar las hipótesis planteadas sobre los signos esperados de los parámetros estimados. No obstante, para que la inferencia estadística sea válida, se deben cumplir algunos supuestos base de los modelos de regresión (Sección 3.3).

Para validar el cumplimiento de los supuestos en los modelos de regresión lineal múltiple, regresión de Poisson y regresión binomial negativa, la literatura sugiere diferentes test estadísticos (Sección 3.3). En la Tabla 4-10 se presentan todos los test aplicados a las especificaciones de los modelos RLM, RP y NB2, utilizando como referencia un 99% de confianza.

Tabla 4-10 Test estadísticos para validar los supuestos de los modelos de regresión

Test	Estadístico	Valor de tabla
Shapiro-Wilk	0,00 ^a (W=0,84)	0,01
Lilliefors	0,23 ^b	0,01
Breusch-Pagan	197,18 (gdl=11)	26,22
Factor de dispersión de Pearson	1.31	1,00
Test de Wald ($\alpha = 0.66$)	14,64 (gdl=5332)	1,65
Test LR para sobredispersión	709,88 (gdl=1)	3,84

Nota: Los estadísticos fueron calculados con paquetes implementados en R Studio.

^a Valores p menores al nivel de significancia elegido (1%) rechazan la hipótesis de normalidad.

^b Abdi et al. (2009) presentan la metodología para calcular el estadístico en muestras grandes.

Fuente: Elaboración propia

En el modelo RLM, los resultados de los test de Shapiro-Wilk y Lilliefors rechazan la hipótesis nula de distribución Normal de los residuos. Por otra parte, el test de Breusch-Pagan rechaza la existencia de homocedasticidad en la distribución de los errores aleatorios (Greene, 2007). En la Tabla C-1, Anexo C, también se presentan los resultados de la estimación de los tests de normalidad y heterocedasticidad de las tres especificaciones alternativas de modelos de regresión lineal múltiple (Sección 4.3). En el modelo de Poisson, los resultados del factor de dispersión de Pearson, el test de Wald y el test LR, indican la presencia de sobredispersión en los datos (Tabla 4-10), y por consiguiente, también sugieren la aplicación de otro tipo de modelos (especificación RP, Tabla 4-9).

Con todo lo anterior, el único modelo remanente es el binomial negativo (NB2). En base a lo discutido en la Sección 3.3.3, para detectar problemas de multicolinealidad en este modelo se puede testear un modelo de regresión lineal múltiple formulado con la misma especificación definida en la función de enlace del modelo de regresión binomial negativa. En la Tabla 4-9 se observa que los indicadores VIF de todas las variables explicativa incluidas en el modelo NB2 son menores que cuatro, lo cual sugiere que no existe un problema de multicolinealidad. Además, notar que los indicadores VIF de todas las variables resultaron más cercanos a la cota inferior del indicador (uno) que al valor crítico (cuatro).

En conclusión, la violación del supuesto de normalidad en la distribución de los residuos aleatorios del modelo de regresión lineal múltiple (especificación RLM), la presencia de sobredispersión en el modelo de Poisson (especificación RP), y la ausencia de multicolinealidad en el modelo de regresión binomial negativa (especificación NB2), sugieren la aplicación del modelo NB2.

4.7 Comparación del poder predictivo de los modelos

A partir de lo sugerido en la literatura de modelos de regresión de conteo (Hilbe, 2011), se utilizó un enfoque de enumeración muestral para construir el histograma de las predicciones del modelo de regresión de Poisson (RP) y binomial negativo (NB2). El

procedimiento consiste en calcular el valor esperado de la función de densidad condicional asociada a cada observación de la base de datos, usando la función de enlace de cada modelo y los valores que toman los regresores en cada observación. De esta forma, cada medición de la base de datos corresponde a una realización de la distribución de probabilidades del número de evasiones, condicional al valor que tomen los regresores y los parámetros estimados. Posteriormente, se obtiene una matriz de probabilidades, en la cual cada columna representa la probabilidad de que ocurra un determinado número de evasiones en la puerta del bus (el número de columnas es igual al máximo número de evasiones registrado en la base de datos), y cada fila representa a una observación de la base de datos. Finalmente, se calcula el promedio de las probabilidades calculadas en cada columna de la matriz, y se obtiene un vector con el valor esperado de la probabilidad que ocurra un cierto número de conteos de evasión.

En cambio, en los modelos de regresión lineal múltiple (RLM), no hay una función de densidad de probabilidades asociada a cada observación, que permita estimar la probabilidad condicional de que ocurra un determinado valor en el proceso de conteo. Para resolver este problema, en esta tesis se presenta una metodología simple que permite obtener un vector de predicciones análogo al generado con los modelos de regresión de conteo (programado en R Studio). En la primera etapa, con la ecuación de regresión del modelo RLM se calculan predicciones puntuales de evasión para cada observación de la base de datos. Posteriormente, para obtener valores enteros, las predicciones son redondeadas al entero más cercano (cuando el valor obtenido es negativo, se asume que la predicción es igual a cero). Finalmente, se calcula la proporción de predicciones correspondiente a cada rango de valores que toma el conteo de evasión.

4.7.1 Gráficos de predicción

Para comparar gráficamente el nivel de ajuste estadístico de cada modelo, se construyeron curvas con sus predicciones. En la Figura 4-1 se presenta la distribución de valores observados y las curvas de predicción de los modelos de regresión RLM, RP y NB2. Como se observa, las predicciones del modelo RLM son cercanas a los valores

observados solo cuando el conteo de evasión es igual a cero. Sin embargo, el modelo presenta un gran sesgo prediciendo conteos de una evasión y un error constante cuando el conteo es mayor a dos evasiones. Respecto a los modelos de regresión de conteo, la predicción del modelo NB2 es más precisa que el modelo RLM para todo el rango de la distribución de valores observados, mientras que el modelo RP es superior a RLM solo cuando el rango de conteos de evasión es mayor a dos.

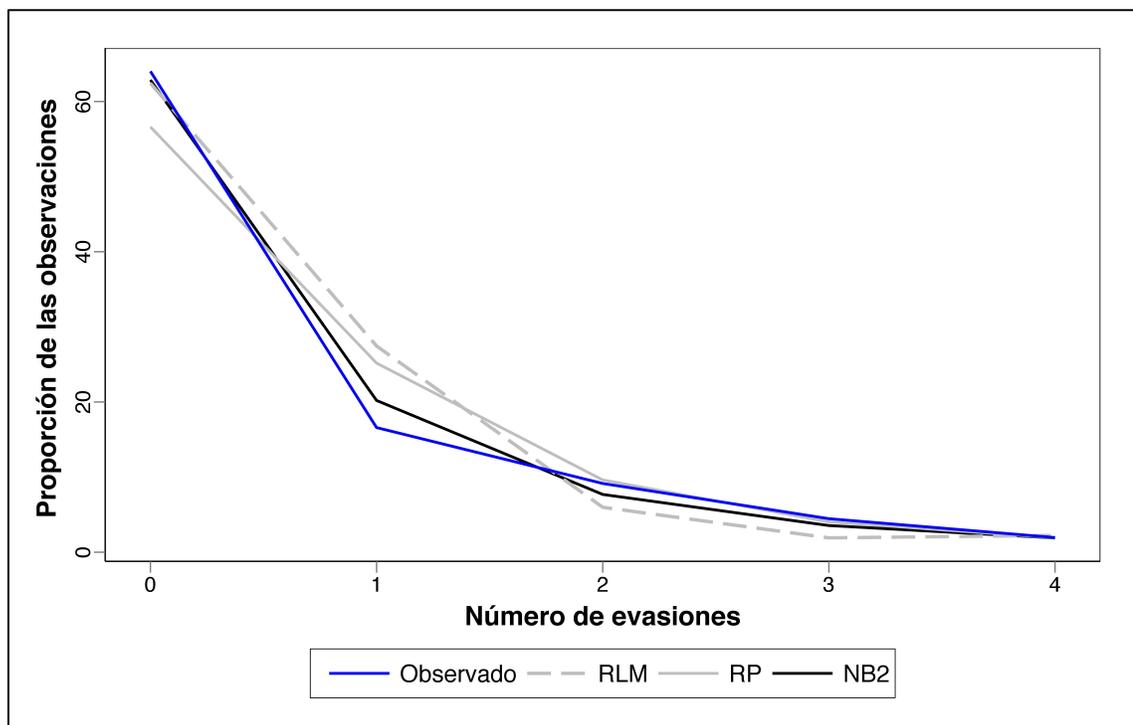


Figura 4-1 Valores observados y predicciones de los modelos RLM, RP y NB2

Fuente: Elaboración propia (gráfico obtenido con R Studio)

4.7.2 Indicador de bondad de ajuste

Para contrastar estadísticamente las diferencias entre las curvas de predicción de cada modelo de regresión y la curva de valores observados de evasión (Figura 4-1), se utilizó una adaptación de la prueba de bondad de ajuste chi-cuadrado. A pesar que la adaptación del test para modelos de regresión de conteo fue sugerida por Hilbe (2011), esta no está implementada en un paquete computacional. Por esta razón, se programó un código computacional en R Studio para aplicar el método propuesto por el autor. Además, el

método fue extendido para analizar el nivel de ajuste en los modelos de regresión lineal múltiple.

Tradicionalmente, el estadístico del test chi-cuadrado se define como (Wonnacott y Wonnacott, 1969):

$$\chi^2 = \sum_{i=1}^n \frac{(E_i - O_i)^2}{E_i} \quad (4.1)$$

compara con el valor de tabla de la distribución chi-cuadrado con $n - 1$ grados de libertad, y para un cierto nivel de significancia estadística α . La hipótesis nula del test (H_0) es que la distribución de los datos observados y pronosticados es la misma. De esta forma, H_0 se rechaza cuando el valor del estadístico calculado es mayor al valor de tabla de la distribución chi-cuadrado.

Para aplicar el test a las predicciones obtenidas con los modelos de regresión RLM, RP y NB2, los intervalos del test fueron definidos agrupando el rango de valores que toma el conteo de evasión. Para garantizar que cada grupo tuviera al menos un uno por ciento de las observaciones (~50 observaciones), se construyeron cinco intervalos para el test (equivalente a cuatro grados de libertad). En base a los intervalos definidos, se calculó la las predicciones obtenidas con los modelos de regresión (Figura 4-1), se calculó la (E_i) en cada intervalo i . En la Tabla 4-11, se presentan los resultados de los test de bondad de ajuste para los modelos RLM, RP y NB2 y la proporción de predicciones de los modelos correspondiente a cada intervalo definido para el test.

Tabla 4-11 Resultados del test chi-cuadrado para los modelos RLM, RP y NB2

Conteos	Valores Observados	Proporción Observada [%]	Proporciones Predichas [%]		
			RLM	RP	NB2
0	3415	64,02	62,43	56,62	62,87
1	885	16,59	27,45	25,17	20,20
2	488	9,15	5,98	9,62	7,69
3	238	4,46	1,91	4,04	3,55
Mayor a 3	308	5,78	2,23	4,55	5,69
$\chi^2=9,49$			55,82	63,70	15,91

Nota: El valor de tabla chi-cuadrado fue calculado con un 95% de confianza y cuatro grados de libertad

Fuente: Elaboración propia

Para todos los modelos, se rechaza la hipótesis nula que la distribución de los datos observados es equivalente a la distribución de las predicciones. Sin embargo, se observa una gran disminución del estadístico de χ^2 cuando se utiliza el modelo NB2. Los problemas de ajuste reflejados en los resultados del test pueden estar explicados por la naturaleza de los datos (alta heterogeneidad en el rango de conteo de evasión), el nivel de agregación de las observaciones, la omisión de variables relevantes y el tamaño de la muestra utilizada para estimar los modelos (Hilbe, 2011).

5 MEJORAS AL MODELO Y ANALISIS DE RESULTADOS

5.1 Inclusión del efecto del nivel socioeconómico (NSE) en el modelo

Tras evaluar el cumplimiento de los supuestos y el nivel de ajuste estadístico de diferentes especificaciones de modelos de regresión en el capítulo anterior, se concluyó que la regresión binomial negativa era la metodología econométrica más adecuada para modelar los datos de evasión. En este capítulo, se reestima el modelo base de regresión binomial negativa (modelo NB2) incluyendo nuevas variables para capturar los efectos del nivel socioeconómico (NSE) en la evasión (modelo NB2-I). Además, se valida la consistencia con los parámetros obtenidos en el modelo NB2 y se evalúa si existe una mejora significativa cuando se incluye el efecto del NSE. Posteriormente, para capturar con mayor precisión el efecto del NSE de los pasajeros en la evasión (discutido en Sección 4.3.6), el modelo NB2-I fue reestimado solo con los datos medidos en el periodo punta mañana (modelo NB2-I-AM). La Tabla 5-1 presenta los resultados de la estimación de los modelos NB2, NB2-I y NB2-I-AM.

Tabla 5-1 Mejoras al modelo de regresión binomial negativa (NB2)

Variable (test t)	NB2	NB2-I	NB2-I-AM
Operación puertas			
Log (entradas)	1,079 (42,7)	1,079 (42,8)	1,136 (19,5)
Salidas	0,221 (2,8)	0,215 (2,7)	0,040 (0,3)
Puertas	0,210 (5,6)	0,186 (4,9)	0,181 (2,3)
Puerta Delantera	-1,336 (-14,0)	-1,344 (-14,1)	-1,608 (-9,2)
Periodo del día			
Punta Mañana (08:30-12:30)	-0,166 (-2,4)	-0,140 (-2,0)	-
Tarde (12:30-20:30)	0,231 (3,9)	0,253 (4,3)	-
Noche (20:30-22:00)	0,499 (2,8)	0,547 (3,1)	-
Calidad servicio			
Baja Ocupación	-0,150 (-2,6)	-0,174 (-3,0)	-0,095 (-0,8)
Headway	0,026 (2,9)	0,018 (1,9)	-0,001 (-0,1)
Operador			
Operadores I (Alsacia, Vule)	-0,271 (-4,0)	-0,228 (-3,3)	-0,329 (-2,1)
Operadores II (Express, Metbus)	-0,567 (-7,8)	-0,445 (-5,9)	-0,482 (-2,8)
Operadores III (Subus, Redbus)	-0,661 (-8,9)	-0,608 (-8,1)	-0,541 (-3,0)

Nivel socioeconómico^a (NSE)

Ingreso Medio-Alto	-	-0,233 (-3,9)	-0,135 (-1,2)
Ingreso Alto	-	-0,366 (-4,1)	-0,684 (-2,0)
Intercepto	0,871 (-4,4)	-0,722 (-3,6)	-0,460 (-1,1)
Nº de observaciones	5.334	5.534	808
Log-verosimilitud	-5.477,8	-5.463,1	-945,4
AIC	10.981,5	10.958,2	1916,8

^a Los tramos de ingreso para agrupar las municipalidades fueron descritos en la Sección 4.3.6.

Fuente: Elaboración propia

Todos los parámetros en el modelo NB2-I son estadísticamente significativos al 95% de confianza, excepto por la variable *Headway* que es significativa solo al 94% de confianza. Una vez más, las variables para capturar el efecto del tipo de servicio de bus (*Alimentador*) y el día de la semana (*Miércoles* y *Jueves*) no resultaron estadísticamente significativas. Respecto a las variables para controlar por el NSE, solo *Ingreso Medio-Alto* e *Ingreso Alto* resultaron significativas al 95% de confianza, mientras que *Ingreso Medio-Bajo* fue significativa solo al 80% de confianza (ver Tabla E-2, Anexo E). Por lo tanto, la variable muda de referencia en el modelo NB2-I incluye las observaciones tomadas en municipalidades de ingreso bajo e ingreso medio-bajo (en la formulación propuesta en la Sección 4.3.6, el grupo de referencia de las variables mudas para controlar el efecto del NSE, solo incluía las observaciones recolectadas en municipalidades de ingreso bajo).

Los resultados indican que la bondad de ajuste del modelo NB2-I es significativamente mayor que la del modelo NB2. Un test LR entre NB2-I y NB2 con dos grados de libertad (las dos variables mudas agregadas para controlar el efecto del NSE), confirma (con un 99% de confianza) que el modelo NB2-I tiene un mayor poder explicativo ($23,374 > \chi_{2,99\%}^2$). Además, como la diferencia entre los indicadores AIC (del inglés *Akaike Information Criteria*) de los modelos NB2 y NB2-I es mayor que 10 ($\Delta = 23,370$), independiente del número de observaciones utilizadas para estimar los modelos, el modelo NB2-I es preferido (Hilbe, 2011). El indicador AIC es una de las

estadísticas más usadas para evaluar el nivel de ajuste estadístico en los modelos lineales generalizados. Mientras más bajo es su valor, mejor ajustado es el modelo (Hilbe, 2011).

5.2 Estimación del modelo para el periodo punta mañana

En el modelo estimado solo con los datos medidos durante el periodo punta mañana (NB2-I-AM), los parámetros de las variables *Log(Entradas)*, *Puertas* y *Puerta Delantera* y de los grupos de unidades de negocio (*Operadores I*, *Operadores II*, *Operadores III*) se mantienen estadísticamente significativos con un 95% de confianza (Tabla 5-1). Sin embargo, las variables *Baja Ocupación* e *Ingreso Medio-Alto* resultan significativas con un nivel de confianza menor al 60%, y las variables *Salidas* y *Headway* con uno menor al 20%. La disminución de la significancia estadística de los parámetros del modelo NB2-I-AM podría estar explicada por la reducción del número de observaciones para estimarlo (solo un 14,6% de las observaciones totales utilizadas para estimar el modelo NB2 ocurren durante el periodo punta mañana). Por otra parte, durante el periodo punta mañana, algunas variables tienen poca variabilidad (e.g la variable *Headway*) y son medidas de forma imprecisa (e.g la variable *Baja Ocupación*).

Los parámetros de las dos variables para capturar el efecto del NSE en la evasión (*Ingreso Medio-Alto* e *Ingreso Alto*), presentan signos y magnitudes consistentes con lo esperado (el valor absoluto de los parámetros es mayor a medida que el ingreso aumenta). Sin embargo, el parámetro estimado de la variable *Ingreso Medio-Alto* no resultó estadísticamente significativo. Esto indicaría que no hay diferencias significativas en los niveles promedio de evasión (*ceteris paribus*) entre los grupos de municipalidades de ingreso bajo, ingreso medio-bajo e ingreso medio-alto (este resultado difiere con el obtenido en el modelo NB2-I). Otro resultado interesante y que puede ser interpretado fácilmente en el contexto de Transantiago, es el aumento significativo en el valor absoluto del parámetro de la variable *Ingreso Alto* (comparado con el valor del parámetro obtenido para la misma variable en el modelo NB2-I). Durante este periodo muchas personas de ingreso bajo viajan a zonas de ingreso alto dentro de Santiago (normalmente por trabajo) y, por consiguiente, se esperaría que la evasión aumente en el grupo de municipalidades de altos ingresos (porque el nivel de

ingreso promedio de los pasajeros en sectores de ingreso alto disminuye). De esta forma, debido al aumento promedio de la diferencia en los niveles de evasión entre las municipalidades de ingreso alto y las demás municipalidades, es razonable que aumente el valor absoluto del parámetro de la variable *Ingreso Alto* (Tabla 5-1).

5.3 Análisis de sensibilidad

A partir de los resultados anteriores (Sección 5.1 y 5.2), se concluye que la estratificación basada en el promedio de ingreso familiar mensual de los hogares en cada municipalidad, es un criterio adecuado para definir las variables *proxy* que capturan el efecto del nivel del NSE en la evasión. La inclusión de estas variables en el modelo NB2-I incrementa significativamente el poder explicativo del modelo de regresión binomial negativa base (NB2). Además, los parámetros estimados de todas las variables explicativas que fueron incluidas en el modelo NB2, continúan siendo altamente significativos y con signos consistentes con lo esperado (Tabla 5-1). En el modelo estimado con los datos punta mañana (NB2-I-AM), los parámetros de las variables *Ingreso Medio-Alto* e *Ingreso Alto* también siguen presentando signos correctos, alta significancia estadística y magnitud consistentes con las hipótesis planteadas (al igual que los parámetros obtenidos para las dos variables en el modelo NB2-I). A pesar que se observan cambios en el valor y la significancia estadística de los parámetros de algunas variables (explicado principalmente por la reducción del número de observaciones usado para estimarlo), la mayoría de los signos y las magnitudes relativas entre los parámetros se mantienen igual.

Finalmente, el análisis de sensibilidad se realizó con el modelo NB2-I, puesto que explica la dinámica de la evasión para diferentes periodos del día (el modelo NB2-I-AM solo utiliza datos de un periodo), aumenta el poder explicativo del modelo NB2 y también presenta parámetros con alta significancia estadística, y que tienen signos y magnitudes consistentes con lo esperado. Para estudiar el impacto de cambios en el valor de las variables explicativas del modelo NB2-I, se calcularon las razones de tasas de incidencia (IRR, por la abreviación del inglés *Incidence Rate Ratio*), un indicador ampliamente utilizado para interpretar los resultados obtenidos con los modelos de

regresión de conteo (Hilbe, 2011). El indicador IRR representa cuántas veces la variable dependiente aumenta o disminuye con respecto a su valor esperado, cuando el valor de una variable explicativa sube en una unidad. Los cambios en la variable dependiente son calculados *ceteris paribus*, es decir, se evalúa el impacto de la variable explicativa manteniendo constante el valor del resto de las variables. Considerando que la función de enlace en los modelos de regresión de Poisson (Ecuación (3.3)) y binomial negativa (Ecuación (3.5)) definen una relación logarítmica entre la variable dependiente y el conjunto de variables explicativas, el indicador IRR de cada variable con forma lineal o de tipo categórica, es igual a la exponencial de su parámetro estimado (Anexo B).

Para facilitar la interpretación de los resultados obtenidos con los indicadores IRR, se calculó un indicador denominado factor de cambio (FC). El indicador FC representa el cambio porcentual en el valor esperado de conteos medidos en la variable dependiente, ante un cambio unitario de la k -ésima variable independiente. El indicador FC para cualquier variable explicativa k con forma lineal o de tipo categórica, es igual a $100 \cdot (IRR_k - 1)$ (revisar Anexo B). En cambio, para las variables con forma logarítmica (e.g. variable de exposición del modelo de conteo), el parámetro estimado (sin exponenciar) entrega directamente la elasticidad porcentual de la variable (i.e. variación porcentual de la variable dependiente ante un cambio de un uno por ciento en la variable independiente). Así, si el parámetro estimado de la variable es exactamente igual a uno, independiente del valor que tome la variable independiente, la tasa de ocurrencia de la variable dependiente permanecerá constante. En cambio, si el parámetro es mayor o menor que uno, la tasa de ocurrencia aumentará o disminuirá, respectivamente, a medida que el valor de la variable independiente aumente.

En la Tabla 5-2 se muestran los indicadores IRR de cada variable, el intervalo de confianza de los indicadores IRR (con un nivel de confianza del 95%) y el indicador FC calculado para cada variable incluida en el modelo de regresión NB2-I.

Tabla 5-2 Razón de tasas de incidencia y factores de cambio modelo NB2-I

Variable (t-test)	IRR	IC ^a	FC
Operación puertas			
Log (entradas)	-	-	1,08 ^b
Salidas	1,240	[1,062, 1,446]	23,985
Puertas	1,204	[1,117, 1,299]	20,431
Puerta Delantera	0,261	[0,216, 0,315]	-73,911
Periodo del día			
Punta Mañana (08:30-12:30)	0,869	[0,759, 0,996]	-13,067
Tarde (12:30-20:30)	1,288	[1,147, 1,447]	28,783
Noche (20:30-22:00)	1,729	[1,220, 2,431]	72,877
Calidad servicio			
Baja Ocupación	0,840	[0,750, 0,942]	-16,008
Headway	1,012	[1,000, 1,036]	1,775
Operador			
Operadores I (Alsacia, Vule)	0,797	[0,697, 0,911]	-20,349
Operadores II (Express, Metbus)	0,641	[0,552, 0,745]	-35,909
Operadores III (Subus, Redbus)	0,544	[0,470, 0,630]	-45,582
Nivel socioeconómico^c (NSE)			
Municipalidad Ingreso Medio-Alto	0,792	[0,703, 0,891]	-20,813
Municipalidad Ingreso Alto	0,693	[0,581, 0,825]	-30,663

^a Los intervalos de confianza de los indicadores IRR son calculados con el método Delta (Hilbe, 2011).

^b Cambio porcentual del número de evasiones ante un cambio unitario porcentual del número de entradas.

^c Los tramos de ingreso de cada grupo de municipalidades se describen en la Sección 4.3.6.

Fuente: Elaboración propia

Acorde a lo esperado, en el modelo NB2-I el parámetro estimado de la variable **Log(Entradas)** es significativamente mayor que uno (un test t de dos colas rechaza al 99% de confianza la hipótesis nula que el parámetro es igual a 1). En efecto, por la forma logarítmica de la variable de exposición (Sección 4.3.1), este resultado indica que a medida que entran más pasajeros por la puerta del bus, se produce un incremento de la tasa de evasión por esa puerta. Asimismo, la elasticidad porcentual de esta variable (igual a su parámetro estimado) indica que un aumento de uno por ciento en la cantidad de pasajeros entrando por una puerta, aumenta en 1,08 por ciento el valor esperado de la evasión medida en esa puerta. El aumento progresivo en la tasa de evasión podría estar

explicado por la dificultad de pagar el pasaje y el mayor anonimato (Sección 2.1.3) que ocurre cuando existe apelotonamiento en la entrada de pasajeros por una puerta del bus.

El número de personas saliendo por una puerta también es una variable con alta significancia estadística en el modelo. Como se discutió en la Sección 4.3, la variable ***Salida*** resultó significativa solo cuando fue agregada como una variable muda que toma el valor uno si alguien sale por la puerta, y cero en otro caso. Entonces, el indicador FC de esta variable indica que cuando un pasajero sale por la misma puerta que alguien entra, la evasión aumenta en un 24,0%, independientemente de si la observación fue medida en la puerta trasera o delantera. Naturalmente, en las puertas traseras se espera un aumento del nivel promedio de evasión cuando alguien sale por la puerta (i.e. la evasión en una puerta trasera solo puede ocurrir si un pasajero sale por la puerta). En cambio, en la puerta delantera el resultado no es evidente, ya que la evasión puede ocurrir independientemente de si alguien salió por esa puerta (donde los validadores “BIP” están ubicados). En este caso, el aumento de la evasión podría deberse al incremento del anonimato de los pasajeros que entran cuando un grupo de personas sale por la puerta delantera.

La tasa de evasión en una puerta también aumenta con el número de puertas del bus (Tabla 4-8). El indicador FC de la variable ***Puertas*** indica que la evasión esperada por una puerta del bus, aumentaría en un 20,4% si el bus tuviese una puerta adicional. Este resultado puede deberse a que cuando los buses tienen más puertas es más difícil para el conductor controlar la entrada de pasajeros. La disminución del contacto visual y la proximidad física del conductor con los pasajeros (por el mayor número de puertas), aumenta el anonimato, por lo cual se esperaría un aumento de la evasión (Sección 2.1.3).

La puerta de entrada elegida por los pasajeros es también una variable que explica significativamente los niveles de evasión. En la puerta delantera se espera una disminución de 73,9% en el nivel de evasión respecto a la puerta trasera (*ceteris paribus*). La reducción de la evasión en la puerta delantera, podría estar explicada por el aumento del contacto visual entre los pasajeros y el conductor del bus, y la menor dificultad de los pasajeros para pagar el pasaje (por la existencia del validador

electrónico). Además, se espera un aumento significativo de la evasión en la puerta trasera, puesto que existe un grupo de evasores, quienes independiente de las opciones que tengan de entrar por la puerta delantera, entran por las puertas traseras para evadir. En consecuencia, incluso si este grupo de pasajeros fuera forzado a entrar por la puerta delantera, probablemente seguirían evadiendo, y por consiguiente, hay que ser cuidadoso con las conclusiones derivadas a partir del efecto de cambios en esta variable sobre el número esperado de evasores.

Niveles altos de ocupación en los buses también parecen incrementar la evasión. El indicador FC de la variable ***Baja Ocupación*** indica que si un bus pasa de alta a baja ocupación (menos de la mitad del pasillo del bus con pasajeros parados), la evasión en cada puerta cae en un 16%. Este resultado puede deberse a que cuando la ocupación de los buses es baja, los pasajeros pueden pagar su pasaje con mayor facilidad por el uso del validador electrónico (ubicado en la puerta delantera), independientemente de la puerta por la que hayan entrado (las personas que entran la puerta trasera podrían ir a pagar después). Por otra parte, cuando la ocupación es alta, hay mayor anonimato, lo cual también promueve la disposición de los pasajeros a evadir (Sección 2.1.3). Otro factor relevante relacionado con el hacinamiento en los buses, es el empeoramiento en la percepción de la calidad del sistema de transporte público, lo cual también parece aumentar los niveles de evasión (Sección 2.1.1).

En servicios de buses con baja frecuencia, también se observan niveles mayores de evasión. La forma funcional que mejor capturó el efecto de la frecuencia del servicio de buses, fue el recíproco de esta variable. Como se discutió en la Sección 4.3.3, esta transformación funcional de la frecuencia es una *proxy* del *headway* (la brecha entre los intervalos entre pasadas de los buses) en la ruta. Bajo esta especificación, el indicador FC obtenido para la variable ***Headway*** indica que un aumento de un minuto en el *headway* promedio de una ruta de buses, incrementa en un 1,8% la evasión en cada puerta del bus. Como el modelo incluye una variable para controlar por el nivel de ocupación (***Baja Ocupación***), la variable ***Headway*** podría estar capturando principalmente el efecto de los tiempos de espera de los pasajeros en el paradero. Al

igual que con el hacimiento en los buses, tiempos de espera más largos deterioran la percepción del sistema de transporte público (mayores *headway*), lo cual podría explicar el aumento de los niveles de evasión (Torres-Montoya, 2014).

Los resultados también sugieren la existencia de diferencias significativas de los niveles de evasión observados a lo largo del día (*ceteris paribus*). Los valores de los parámetros estimados asociados a las variables incluidas para capturar el efecto del periodo del día en la evasión (Tabla 5-1), sugieren que; i) las tasas de evasión promedio durante la tarde son significativamente mayores que las reportadas durante la mañana ii) las mayores tasas de evasión se producen durante la noche; iii) Los niveles de evasión no son significativamente distintos entre los periodos de la tarde (punta medio día, fuera punta tarde y punta tarde). En cuanto a los dos primeros resultados, los indicadores FC de las variables **Tarde** y **Noche** indican que durante la tarde y la noche se espera un aumento de 28,9% y 72,9% en la evasión, respectivamente (comparado con el periodo punta mañana). El tercer resultado es interesante puesto que desafía la idea que las tasas de evasión son necesariamente mayores durante el horario punta. Aunque durante el periodo fuera punta de la mañana la evasión disminuye en un 13,1% respecto al periodo punta mañana (*ceteris paribus*), durante los periodos de la tarde no se observan diferencias significativas en los niveles de evasión. Las variables para controlar por los periodos del día, podrían estar capturando cambios sistemáticos de algunos factores a lo largo del día, que inciden en los niveles de evasión pero no están considerados en el modelo de regresión, tales como las características sociodemográficas de los pasajeros (e.g. rango etario y género), los propósitos de viaje y los niveles de inspección.

Los resultados también indican que existen diferencias significativas de los niveles de evasión entre las unidades de negocio de Transantiago, que no están explicadas por las demás variables incluidas en el modelo. Los indicadores FC obtenidos para estas variables indican que los operadores exhiben distintos niveles de evasión promedio (*ceteris paribus*); en particular, el primer grupo (Alsacia, Vule) presenta un 20,3% más de evasión que el grupo de control (STP), mientras que el segundo (Express, Metbus) y tercer grupo (Subus, Redbus) presentan un 35,9% y 45,6% más de evasión que el grupo

control (STP) respectivamente. En Transantiago, los operadores tienen la facultad de implementar diferentes estrategias para disminuir la evasión tales como: el uso de torniquetes en la puerta delantera y la asignación de inspectores supervisando la entrada de los pasajeros al bus por la puerta delantera (Sección 2.2). Por otra parte, la incidencia de los conductores de buses en el control de la evasión podría variar según los incentivos entregados por el operador. Como estos factores no están incluidos en el modelo de regresión, las variables mudas asociadas a cada grupo de operadores podrían estar capturando parcialmente las diferencias en los niveles de evasión observados entre ellos. La Figura 5-1 y la Figura 5-2 presentan el valor promedio y el intervalo de confianza (estimado con el método Delta) del indicador FC asociado a cada grupo de operadores, en el modelo NB2-I y NB2 respectivamente. Notar que en ambos modelos se observan diferencias significativas en los niveles de evasión de los tres grupos de operadores y éstas mantienen la misma tendencia (e.g. el grupo STP es el que presenta mayor evasión en promedio, y el grupo de Subus y Redbus es el que tiene menor evasión).

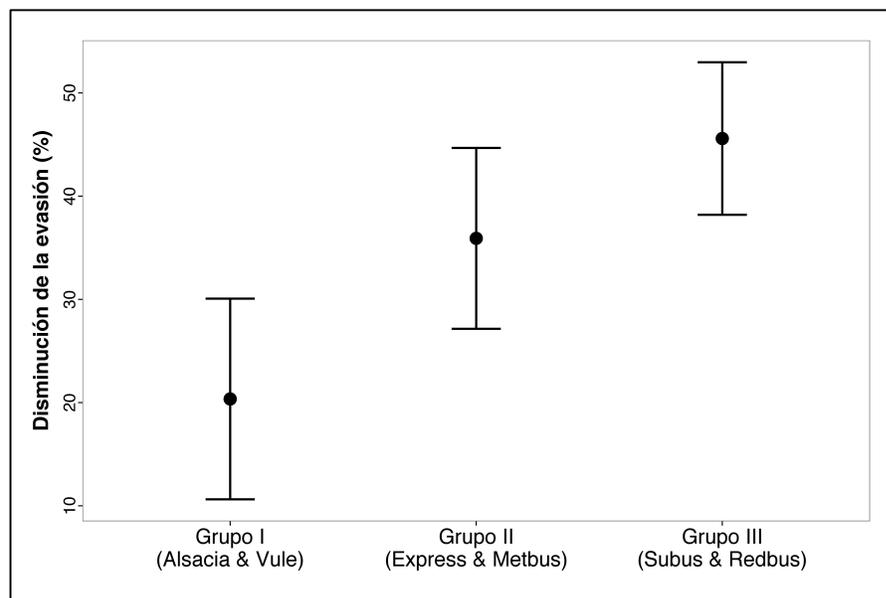


Figura 5-1 Factores de cambio según grupo de operadores (modelo NB2-I)

Fuente: Elaboración propia

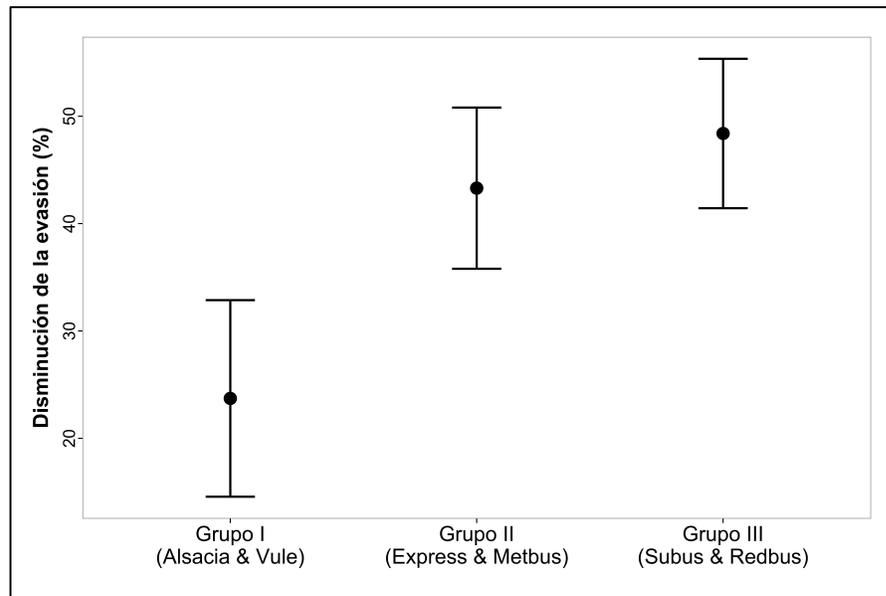


Figura 5-2 Factores de cambio según grupo de operadores (modelo NB2)

Fuente: Elaboración propia

El nivel de ingreso también es un factor importante para explicar el comportamiento de los evasores. Los indicadores FC de las variables **Ingreso Medio-Alto** e **Ingreso Alto** obtenidas en el modelo NB2-I, indican que en paraderos ubicados en municipalidades de ingreso medio, se espera que la evasión medida en cada puerta disminuya en 20,8% respecto a mediciones realizadas en municipalidades de ingresos medios-bajos y bajos (grupo de control), mientras que en municipalidades de ingreso alto se espera que la evasión disminuya 30,7% en promedio respecto al grupo control. Sin embargo, el nivel socioeconómico (NSE) del área de influencia de los paraderos no siempre está vinculado con el NSE de los individuos que suben al bus. Para capturar de forma más clara el efecto del NSE individual, el modelo NB2-I fue reestimado usando datos recolectados solo durante el periodo punta mañana (Sección 5.1). En este caso, los resultados también fueron consistentes con lo esperado, es decir, a medida que aumentaba el ingreso familiar mensual promedio de la comuna donde fue realizada la medición, la evasión disminuía significativamente (*ceteris paribus*). A continuación se presenta el valor promedio y el intervalo de confianza del indicador FC asociado a cada grupo de municipalidades, obtenidos con el modelo NB2-I (Figura 5-3) y NB2-I-AM (Figura 5-4).

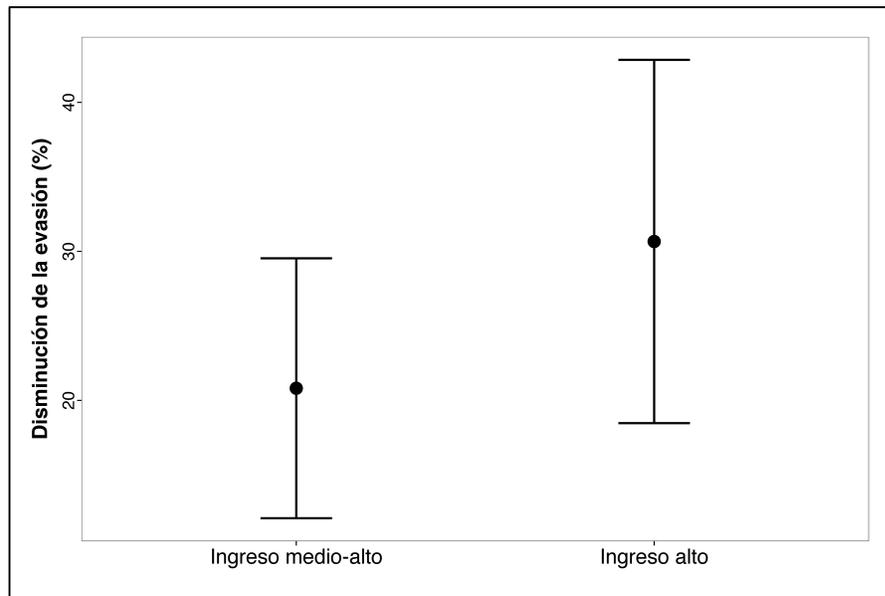


Figura 5-3 Factores de cambio según NSE municipalidad (modelo NB2-I)

Fuente: Elaboración propia

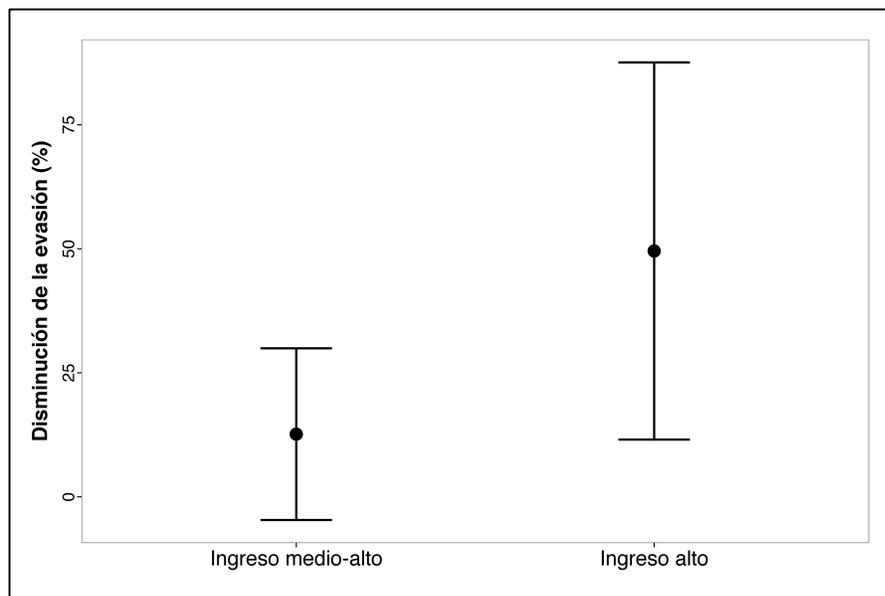


Figura 5-4 Factores de cambio según NSE municipalidad (modelo NB2-I-AM)

Fuente: Elaboración propia

6 PROPUESTAS DE POLITICA PUBLICA

Para reducir la evasión en el transporte público se proponen cinco estrategias, que fueron derivadas a partir de los resultados obtenidos con el modelo de regresión (NB2):

- i) Aumento del tamaño de la flota
- ii) Mejora de la regularidad de los intervalos entre pasadas de los buses
- iii) Implementación de zonas pagas
- iv) Cambio del sistema de pago en el bus
- v) Cambio del diseño de la flota (número de puertas o capacidad del bus).

A continuación se discute el impacto de cada estrategia sobre las variables explicativas incluidas en los modelos y su efectividad para reducir la evasión. Posteriormente, se plantean algunas recomendaciones relacionadas con el diseño de las políticas de inspección y subsidio a la tarifa. A diferencia de las cinco estrategias descritas inicialmente, las últimas dos políticas no pueden ser evaluadas con precisión, ya sea porque los modelos no incluyen variables para capturar sus efectos (i.e. relación entre la tasa de inspección y la evasión) o porque no es posible medir su impacto sobre las variables explicativas.

6.1 Tamaño de la flota

Una flota más numerosa de buses operando en una ruta, aumenta la frecuencia entre pasadas de los buses. En términos de las variables incluidas en el modelo, esta política debiera reducir los *headway*, el nivel de ocupación de los buses, el número de pasajeros entrando simultáneamente por una puerta, los casos en los cuales alguien entra por la misma puerta que alguien está saliendo (capturado por la variable **Salidas**), y la proporción de gente entrando por las puertas traseras (como se discutió en la Sección 5.3, algunas personas van a evadir incluso si tienen la oportunidad de entrar por la puerta delantera). Además, un aumento de la frecuencia tiene un impacto en todos los paraderos visitados por el servicio de buses, y por consiguiente, esta política podría generar una reducción importante en el nivel de evasión a lo largo de la ruta.

6.2 Regularidad

Mejorar la frecuencia de la línea incrementando el tamaño de la flota podría aumentar significativamente los costos operacionales de las agencias de transporte público. En contraste, la introducción de estrategias de control de flota (Delgado et al., 2009) para mejorar la regularidad de los *headway*, podría reducir la acumulación de pasajeros tanto en los buses como en los paraderos, sin la necesidad de agregar buses adicionales o contratar a más conductores. En este escenario, los resultados del modelo de regresión sugieren que la mejora de la regularidad de los buses debería reducir la evasión, puesto que disminuye el número de pasajeros entrando simultáneamente por una puerta del bus, el número de casos en que los pasajeros entran por la misma puerta que alguien sale, y la proporción de buses con alta ocupación (debido al mejor balance en el número de pasajeros por bus en la ruta). En un proyecto piloto realizado en Transantiago para mejorar la regularidad de los *headway* en algunas rutas de buses manejadas, se observó que las estrategias de control de flota lograron reducir la irregularidad de los *headway* y la evasión (Lizana et al., 2014).

6.3 Sistema de pago en el bus

Mejoras en el sistema de pago del pasaje dentro del bus mediante la implementación de validadores electrónicos en las puertas traseras podrían disminuir la evasión. La entrada por las puertas traseras ayuda a disminuir los tiempos de subida de los pasajeros en los paraderos (Stewart y El-Geneidy, 2014) y, por consiguiente, si se mantiene la misma flota, aumentaría la frecuencia del servicio. Como fue discutido, un aumento de la frecuencia puede generar una reducción significativa de la evasión en todos los paraderos de la ruta.

Los altos niveles de ocupación de los buses también inciden en que los pasajeros entren por las puertas traseras (donde no hay validadores electrónicos de tarjetas BIP). En algunos casos, la gente ingresa por las puertas traseras cuando el bus va muy lleno, incluso si no están intentando evadir. Los validadores electrónicos en las puertas traseras podría ayudar a reducir la evasión dentro de este grupo de evasores, pero esto no

aseguraría que todos los pasajeros entrando paguen su pasaje; en particular, muchos evasores oportunistas podrían tomar ventaja del anonimato en la puerta trasera para evadir (Torres-Montoya, 2014). Por otra parte, es importante evaluar la factibilidad técnica de implementar validadores electrónicos adicionales y el costo de su implementación. En Transantiago la tecnología actual solo permite tener dos validadores por bus, y en la actualidad los dos están ubicados en la puerta delantera del bus. En este contexto, uno de los validadores electrónicos de la puerta delantera podría trasladarse a la puerta trasera, y luego evaluar el cambio neto de los niveles de evasión en el servicio.

6.4 Zonas de prepago (ZP)

Las ZP podrían no ser el mejor método para reducir la evasión en el paradero donde son implementadas, debido a sus altos costos de infraestructura y operación, y al uso del espacio público. Sin embargo, al agilizar el proceso de subida de los pasajeros en el paradero, las ZP pueden mejorar la frecuencia del servicio (Sección 2.2.2), y por consiguiente, reducir significativamente la evasión en todos los paraderos del servicio. Además, al permitir que todos los pasajeros paguen su pasaje antes de entrar al bus, las ZP reducen la acumulación de personas en los paraderos de alta afluencia, lo cual también contribuye a disminuir la evasión.

En cuanto a los beneficios por la fiscalización, no es claro si los inspectores en las ZP son eficaces para reducir la evasión. Una importante reducción de la evasión en el paradero donde las ZP son implementadas podría estar explicada por evasores que caminan a un paradero cercano para evitar ser fiscalizados. En este caso, el nivel de evasión en el sistema sería el mismo y, por consiguiente, la asignación de inspectores no sería costo-efectiva. Por esta razón, la inspección en ZP debería ser evaluada cuidadosamente, especialmente cuando se espera que haya una alta proporción de evasores dispuestos a caminar a paraderos cercanos para evitar ser fiscalizados. Sería interesante desarrollar estudios ex-ante y ex-post para evaluar cambios en el flujo de pasajeros entre los paraderos cercanos a las ZP durante el periodo en que fueron implementadas. Esto ayudaría a determinar los beneficios reales que son obtenidos con

los inspectores en la ZP, y luego éstos podrían ser comparados con la efectividad de asignar inspectores que fiscalicen de forma aleatoria a lo largo del servicio.

6.5 Diseño del bus

Los resultados indican que la evasión es más alta en buses con mayor ocupación y con más puertas. Por lo tanto, la evasión podría ser reducida disminuyendo el número de puertas o aumentando la capacidad de los buses (i.e. buses de dos pisos). En ambos casos, es recomendable que los operadores hagan un análisis costo beneficio para ponderar el cambio neto en la evasión y los costos de la implementación de estas estrategias. La inversión para aumentar la capacidad de los buses podría ser muy alta, comparado con los beneficios obtenidos por la reducción de evasión. Por otra parte, una reducción del número de puertas de los buses podría incrementar significativamente los tiempos de subida de los pasajeros. En consecuencia, la disminución del número de puertas, podría aumentar el *headway* de la ruta (por el aumento del tiempo de subida de los pasajeros) y producir un aumento neto de la evasión.

6.6 Programación horaria de la inspección

Los resultados sugieren que no hay diferencias en promedio (*ceteris paribus*) en los niveles de evasión durante la tarde (el periodo de la tarde incluye el horario punta medio día, fuera de punta y punta tarde), y que en la tarde se espera un aumento de la evasión respecto a las mediciones realizadas durante la mañana (horario punta mañana y fuera de punta mañana) (Sección 5.3). Esto sugiere, entre otras cosas, que la inspección en servicios de buses que operan durante el horario fuera de punta tarde podría ser costo-efectiva, ya que durante este horario también podrían existir paraderos con altos niveles de evasión. Por otra parte, en horario fuera de punta se facilita la labor de inspección de los pasajeros (los buses están más vacíos) y se espera un menor impacto negativo en la operación del sistema. Cuando los inspectores suben al bus, el vehículo se detiene y, por consiguiente, la fiscalización en horario punta podría generar aumentos significativos de los tiempos de viaje de los pasajeros y en el nivel de congestión del sistema.

También podría ser conveniente aumentar los niveles de inspección durante la noche. Usando otra base de datos proporcionada por el Programa Nacional de Fiscalización del MTT con información sobre el nivel de inspección en el sistema (en el año 2010), se observó que los niveles de inspección variaban significativamente según el periodo del día. Según estos datos, el 35,7% del total de inspecciones fue durante la tarde, mientras que en los horarios fuera de punta mañana, punta mañana y noche, fue de 30,2%, 7,4% y 2,7% respectivamente. Esto podría explicar el aumento de la tasa de evasión durante la noche que fue capturada por el modelo de regresión (*ceteris paribus*). Otras agencias de transporte público en el mundo también han concluido que la inspección durante la tarde y noche puede ser más costo-efectiva que en otros periodos del día (Lee, 2011).

6.7 Subsidios

Los aumentos de tarifa afectan a gran parte de los usuarios de transporte público. Para la mayoría de las personas de menores ingresos el sistema de transporte público es la única forma de ir al trabajo o al estudio (SECTRA, 2006). Hasta 2006, el primer quintil de hogares en el área urbana de Santiago tenía un ingreso promedio mensual menor a \$165.553 y la composición familiar promedio era de 4,3 personas (Tabla 6-1).

Tabla 6-1 Características por quintil de ingreso área urbana Santiago 2006

Quintil	Ingreso	Población	Miembros por hogar
1	165.553	1.428.251	4,3
2	329.586	1.364.423	4,1
3	494.373	1.300.673	3,8
4	760.276	1.182.969	3,4
5	2.212.200	1.056.601	3,1

Fuente: (Observatorio Social, 2006)

Los resultados de la encuesta realizada por SECTRA (2006) indican que más del 29% de los usuarios de transporte público pertenecían a familias con un ingreso promedio menor a \$160.000. Por lo tanto, una familia tradicional con dos hijos, perteneciente al primer quintil de ingreso en Santiago, en la cual todos los miembros viajaran dos veces al día y 20 días al mes, gastaría en 2006 el 25% de su presupuesto mensual en transporte.

SECTRA (2006) reportó que los pasajeros de familias con ingreso familiar promedio menor a \$280.000, realizaban en promedio 1,3 viajes diarios en modos motorizados.

En este contexto, las altas tasas de evasión registradas en lugares de menores ingresos, podrían estar vinculadas, legítimamente, a la incapacidad de pagar el alto costo del pasaje en transporte público (relativo al nivel de ingreso) y podrían explicar parcialmente las diferencias en los niveles de evasión observados entre municipalidades. Desde un punto de vista de política pública, parecería necesario considerar subsidios para reducir el costo del pasaje, ya sea dirigidos exclusivamente a personas de menor ingreso o a todos los usuarios del sistema de transporte público (reducción del precio del pasaje). En contraste, estrategias de fiscalización en zonas de bajos ingresos podrían reducir marginalmente la tasa de evasión, pero al costo de limitar la accesibilidad de quienes evaden porque no pueden pagar el pasaje.

7 CONCLUSIONES

La evasión en transporte público es un fenómeno más complejo que simplemente considerar la existencia de pasajeros actuando como agentes racionales que buscan maximizar su utilidad. Para manejar el problema de evasión es necesario examinar el contexto social y el sistema de transporte donde tiene lugar. En esta tesis se presenta una metodología basada en modelos de regresión, estimados con información recolectada por el Programa Nacional de Fiscalización del MTT para medir la tasa de evasión mensual en Transantiago durante Octubre de 2012. A pesar que la inclusión de variables en los modelos estuvo limitada a la información disponible en la base de datos original, los resultados obtenidos fueron consistentes con lo esperado y entregan evidencia valiosa para la literatura existente sobre evasión en el transporte público. Después de probar diferentes especificaciones de modelos de regresión lineal múltiple, regresión de Poisson y regresión binomial negativa, se concluyó que la regresión binomial negativa era el método más adecuado para modelar los datos sobre evasión.

En base a todos los resultados obtenidos, se propusieron cinco métodos para disminuir la evasión, i) incrementar el tamaño de la flota de buses; ii) mejorar la regularidad de los *headway*, iii) implementar zonas pagas; iv) cambiar el sistema de pago dentro del bus; v) modificar el diseño de los buses (número de puertas o la capacidad del bus). Los resultados de la estimación de los modelos indican que la evasión aumenta si más pasajeros entran (o salen) por una puerta, más pasajeros entran por las puertas traseras, los buses tienen más puertas, los buses tienen mayor ocupación y mayores intervalos entre pasadas (*headway*). Además, controlando por estas variables (*ceteris paribus*), se encontró que la evasión era mayor durante la tarde y la noche, aunque no es claro si la tasa de evasión es mayor durante el horario punta. En cuanto a las variables asociadas al nivel socioeconómico (NSE) de los pasajeros, se encontró que la evasión en paraderos ubicados en áreas de mayor ingreso era significativamente menor que en áreas de menor ingreso. En el contexto de Transantiago, las altas tasas de evasión medidas en zonas de ingreso bajo podrían estar legítimamente asociadas a la incapacidad de pagar el pasaje. Desde una perspectiva de política pública, para enfrentar la evasión existe una necesidad

de considerar subsidios para reducir el costo de la tarifa, ya sea dirigidos a usuarios de menor ingreso o a toda la población.

En síntesis, el enfoque econométrico utilizado en la tesis permitió capturar correctamente la influencia conjunta de diferentes variables en los niveles de evasión. Los modelos estimados proveen una herramienta poderosa para predecir el impacto de diferentes políticas públicas en la evasión que modifiquen el nivel de algunas variables incluidas en éstos. La metodología también puede ser aplicada a otros sistemas de transporte público, aunque los valores de los parámetros estimados seguramente cambiarían y, por consiguiente, sería distinto el impacto de las políticas públicas dirigidas a reducir la evasión. Por ejemplo, en algunos sistemas de transporte público los niveles de ocupación de los buses podrían ser un problema mayor, por lo cual se esperaría que las políticas operacionales para reducir la ocupación fueran más eficaces para reducir la evasión.

En futuras investigaciones, se debería formular un modelo de costo-beneficio para evaluar la rentabilidad de cada estrategia para reducir la evasión. Dicho modelo debería ponderar al menos el aumento de la recolección de tarifas por la reducción de la evasión y el costo de implementar cada estrategia en el conjunto de rutas y paraderos dentro del horizonte de intervención. Además, se deberían formular nuevas especificaciones de los modelos de regresión, con mediciones más precisas que permitan capturar el efecto del *headway* de la ruta (p.ej. con datos GPS de los buses), su nivel de ocupación (número exacto de pasajeros en el bus) y el NSE de los individuos (datos GIS permitirían estimar el nivel de ingreso de los vecinos viviendo dentro de un área de influencia definida para cada paradero). Por otra parte, es necesario desarrollar un análisis longitudinal con toda la información disponible sobre evasión, integrando variables sobre la opinión pública, el precio del pasaje y las tasas de inspección, entre otros.

El efecto de algunas variables psicosociales también debería ser capturado en nuevas especificaciones de los modelos de regresión. Cuando la gente actúa en contextos colectivos, podría experimentar un efecto de “desindividualización”, que aumenta la susceptibilidad a procesos de influencia grupal (Zimbardo y Leippe, 1991). En este

contexto, la gente tiende a reemplazar sus estándares individuales por los del grupo, y por consiguiente, aumenta la imitación de conductas observadas en otros produciéndose un proceso de contagio social (Hatfield et al., 1994). En consecuencia, la percepción de que el resto de las personas está evadiendo podría realmente incrementar la evasión e incluso legitimarla. Para probar estas hipótesis, sería útil analizar si la fuente de sobredispersión detectada en el modelo de regresión de Poisson está vinculada a contagio verdadero más que una dependencia aparente entre las variables aleatorias de la distribución de probabilidades. La presencia de contagio verdadero apoyaría la hipótesis que la evasión aumenta cuando se observa a otros evadir, y esto sería consistente con los procesos de contagio social presentes en el comportamiento no ético (Gino et al., 2009). Por otra parte, modelos con datos longitudinales deberían permitir testear estadísticamente si existen efectos auto regresivos (i.e. efecto de la tasa de evasión en el periodo anterior), para verificar si la percepción de mayor evasión en el sistema produce una mayor predisposición a evadir en periodos posteriores.

Finalmente, se invita a que futuros investigadores interesados en trabajar en este tema, se ubiquen en una perspectiva interdisciplinaria, integrando metodologías de investigación cuantitativas y cualitativas (Clifton y Handy, 2001). Confiar exclusivamente en el análisis estadístico podría generar una mala comprensión de un problema complejo como la evasión. Aunque las metodologías utilizadas en esta tesis se basan principalmente en métodos cuantitativos, teorías desde la psicología sobre la conducta no ética (Gino et al., 2009; Karstedt y Farrall, 2006; Ruedy et al., 2013) y comportamiento grupal (Blass, 1999; Hatfield et al., 1994; Zimbardo y Leippe, 1991) resultaron esenciales para interpretar y respaldar los resultados de los modelos de regresión.

Esperamos que los resultados de esta tesis generen un impacto en la discusión sobre las políticas públicas que se debaten en la actualidad. La evasión observada es solo la punta del iceberg de un problema psicosocial mucho más profundo.

BIBLIOGRAFIA

Abdi, H., Edelman, B., Valentin, D. y Dowling, W.J. (2009) *Experimental Design and Analysis for Psychology*, Oxford University Press, Oxford.

Avenhaus, R. (2004) Applications of inspection games. *Mathematical Modelling and Analysis*, 9, 179-192.

Barabino, B., Salis, S. y Useli, B. (2013) A modified model to curb fare evasion and enforce compliance: empirical evidence and implications. *Transportation Research Part A: Policy and Practice*, 58, 29-39.

Bazerman, M.H. y Gino, F. (2012) Behavioral ethics: toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science*, 8, 85-104.

BCN (2012) Reportes Estadísticos Comunes 2012, Biblioteca del Congreso Nacional de Chile. <<http://reportescomunales.bcn.cl/2012/index.php/>> (recuperado 01.10.14).

Bijleveld, C. (2007) Fare dodging and the strong arm of the law: an experimental evaluation of two different penalty schemes for fare evasion. *Journal of Experimental Criminology*, 3, 183-199.

Blass, T. (1999) The Milgram paradigm after 35 years: some things we now know about obedience to authority. *Journal of Applied Social Psychology*, 29, 955-978.

Bonfanti, G. y Wagenknecht, T. (2010) Human factors reduce aggression and fare evasion. *Public Transport International*, 59, 28-32.

Boothway, G. (2009) On the optimality of fines when enforcement is risky. *ASBBS E-Journal*, 1, 33-39.

Borndörfer, R., Omont, B., Sagnol, G. y Swarat, E. (2012) A Stackelberg Game to Optimize the Distribution of Controls in Transportation Networks. En: *Game Theory for Networks*. pp. 224-235, Springer Berlin Heidelberg.

Boyd, C., Martini, C., Rickard, J. y Russell, A.R. (1989) Fare evasion and non-compliance. *Journal of Transport Economics and Policy*, 23, 189-197.

Buccioli, A., Landini, F. y Piovesan, M. (2013) Unethical behavior in the field: demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization*, 93, 248-257.

Cameron, A.C. y Trivedi, P.K. (1998) *Regression Analysis of Count data*. Primera Edición, Cambridge University Press, New York.

Castillo, P.C. (2010) *Benchmarking internacional: Experiencias en Fiscalización*, Gobierno de Chile, Santiago.

Clarke, R. V (1993) Fare evasion and automatic ticket collection on the London underground. *Crime Prevention Studies*, 1, 135-146.

Clarke, R. V, Contre, S. y Petrossian, G. (2010) Deterrence and fare evasion: Results of a natural experiment. *Security Journal*, 23, 5-17.

Clifton, K.J. y Handy, S.L. (2001) Qualitative methods in travel behaviour research. *International Conference on Transport Survey Quality and Innovation*. Kruger National Park, South Africa.

Collect Gfk (2010) Estudio Tracking Calidad de Servicio N° 2- Mar - Abril 2010. <[http://transantiago.cl/corporativo/descargas/Transantiago Marzo Abril 2010 final.pdf](http://transantiago.cl/corporativo/descargas/Transantiago_Marzo_Abril_2010_final.pdf)> (recuperado 01.10.14).

Correa, J.R., Harks, T., Kreuzen, V.J.C. y Matuschke, J. (2014) Fare Evasion in Transit Networks. *CoRR*, abs/1405.2.

Delgado, F., Muñoz, J.C., Giesen, R. y Cipriano, A. (2009) Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. *Transportation Research Record*, 2090, 59-67.

Díaz, G., Gómez-Lobo, A. y Velasco, A. (2006) Las micros en Santiago: de enemigo público a servicio público. En: Galetovic, A. (Ed.), *Santiago: ¿Dónde estamos y hacia dónde vamos?* Santiago, Chile: Centro de Estudios Públicos.

DTPM (2012) Programa de Operación Segundo Semestre 2012, Directorio de Transporte Metropolitano. <<http://www.dtpm.cl/index.php/plan-operacional-historico2>> (recuperado 01.10.14).

DTPM (2013a) Programa de Operación Segundo Semestre 2013, Directorio de Transporte Metropolitano. <<http://www.dtpm.cl/index.php/plan-operacional-historico2>> (recuperado 01.02.15).

DTPM (2013b) Informe de Gestión 2012, Directorio de Transporte Público Metropolitano. <[http://www.dtpm.cl/archivos/Informe de gestión 2012 4 de Septiembre.pdf](http://www.dtpm.cl/archivos/Informe_de_gestión_2012_4_de_Septiembre.pdf)> (recuperado 01.10.14).

DTPM (2014) Informe de Gestión 2013, Directorio de Transporte Público Metropolitano. <[http://www.dtpm.cl/archivos/Memoria_2013-Web_Final_\(1\).pdf](http://www.dtpm.cl/archivos/Memoria_2013-Web_Final_(1).pdf)> (recuperado 04.27.15).

EMOL (2013) Transantiago llega a su mayor desaprobación desde 2007, con 61 % de rechazo. <<http://www.elmercurio.com/blogs/2013/01/28/8825/Transantiago-llega-a-su-mayor-desaprobacion-desde-2007-con-61-de-rechazo.aspx>> (recuperado 01.10.14).

Gino, F., Ayal, S. y Ariely, D. (2009) Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychological Science*, 20, 393-398.

Greene, W.H. (2007) *Econometrics Analysis*. Séptima Edición, Pearson Education, Westford.

Hatfield, E., Cacioppo, J.T. y Rapson, R.L. (1994) *Emotional Contagion*, Cambridge University Press, Cambridge.

Hilbe, J. (2011) *Negative Binomial Regression*. Segunda Edición, Cambridge University Press, Cambridge.

Karstedt, S. y Farrall, S. (2006) The moral economy of everyday crime: markets, consumers and citizens. *British Journal of Criminology*, 46, 1011-1036.

Killias, M., Scheidegger, D. y Nordenson, P. (2009) The effects of increasing the certainty of punishment: a field experiment on public transportation. *European Journal of Criminology*, 6, 387-400.

King, G. (1989) Variance specification in event count models: from restrictive assumptions to a generalized estimator. *American Journal of Political Science*, 33, 762-784.

Kooreman, P. (1993) Fare evasion as a result of expected utility maximisation: some empirical support. *Journal of Transport Economics and Policy*, 27, 69-74.

Lee, J. (2011) Uncovering San Francisco, California, Muni's proof-of-payment patterns to help reduce fare evasion. *Transportation Research Record*, 2216, 75-84.

Lizana, P., Muñoz, J.C., Giesen, R. y Delgado, F. (2014) Bus control strategy application: case study of Santiago transit system. *Procedia Computer Science*, 32, 397-404.

MDS (2009) CASEN 2009, Ministerio de Desarrollo Social. <<http://www.ministeriodesarrollosocial.gob.cl/casen/Estadisticas/ingresos.html>> (recuperado 01.10.14).

Muñoz, J.C., Ortúzar, J. de D. y Gschwender, A. (2008) Transantiago: the fall and rise of a radical public transport intervention. En: Saleh, W., Sammer, G. (Eds.), *Travel demand management and road user pricing: Success, failure and feasibility*. pp. 151-172, Ashgate Publishing.

Myers, D.G. (2012) *Social Psychology*. 11.^a Edición, McGraw-Hill, Nueva York.

O'Brien, R.M. (2007) A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41, 673-690.

Observatorio Social (2006) Estadísticas Ingresos. <http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen-documentos.php?c=85&a=2006> (recuperado 01.10.14).

Ortúzar, J. de D. y Willumsen, L.G. (2011) *Modelling Transport*. Cuarta Edición, John Wiley & Sons, Chichester.

Osgood, D. (2000) Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology*, 16, 21-43.

Polinsky, M. y Shavell, S. (1979) The optimal tradeoff between the probability and magnitude of fines. *American Economic Review*, 69, 880-891.

Reddy, A., Kuhls, J. y Lu, A. (2011) Measuring and controlling subway fare evasion: improving safety and security at New York city transit authority. *Transportation Research Record*, 2216, 85-99.

Ruedy, N.E., Moore, C., Gino, F. y Schweitzer, M.E. (2013) The cheater's high: the unexpected affective benefits of unethical behavior. *Journal of Personality and Social Psychology*, 105, 531-548.

Sabatini, F., Wormald, G., Sierralta, C. y Peters, P.A. (2009) Residential segregation in Santiago: scale-related effects and trends, 1992-2002. En: *Urban Segregation and Governance in the Americas*. pp. 121-144, Palgrave Macmillan.

SECTRA (2006) Encuesta Origen Destino 2006, Secretaría de Planificación de Transporte. http://www.sectra.gob.cl/datos_e_informacion_espacial/gran_santiago/encuestas_movilidad.html (recuperado 01.10.14).

Shields, T.G. y Huang, C. (1995) Presidential vetoes: an event count model. *Political Research Quarterly*, 48, 559-572.

Stewart, C. y El-Geneidy, A. (2014) All aboard at all doors: route selection and running time savings estimation for multi-scenario all-door bus boarding. *93rd Annual Meeting of the Transportation Research Board*. Washington, D.C.

Thorlacius, P., Clausen, J. y Brygge, K. (2010) Scheduling of inspectors for ticket spot checking in urban rail transportation, DSB S-tog, Copenhagen. <http://orbit.dtu.dk/fedora/objects/orbit:60099/datastreams/file_5290566/content>

Tirachini, A. (2013) Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services. *Transportation Research Part C: Emerging Technologies*, 30, 239-256.

Torres-Montoya, M. (2014) Tackling fare evasion in Transantiago: an integrated approach. *93rd Annual Meeting of the Transportation Research Board*. Washington, D.C.

Weidner, R.R. (1996) Target-hardening at a New York city subway station: decreased fare evasion—at what price. *Crime Prevention Studies*, 6, 117-132.

Wonnacott, T.H. y Wonnacott, R.J. (1969) *Introductory Statistics*. Segunda Edición, John Wiley & Sons, Nueva York.

Zimbardo, P.G. (2008) *The Lucifer Effect, Understanding How Good People Turn Evil*, Random House, Nueva York.

Zimbardo, P.G. y Leippe, M.R. (1991) *The Psychology of Attitude Change and Social Influence.*, McGraw-Hill Book Company, Nueva York.

ANEXOS

ANEXO A: DESCRIPCIÓN DE LOS DATOS DISPONIBLES

A continuación se describen todas las variables disponibles en la base de datos de mediciones de evasión y la ficha utilizada por el Programa Nacional de Fiscalización para hacer la recolección de datos en terreno.

Tabla A-1 Descripción de variables asociadas a los digitadores y fiscalizadores

Variable	Descripción
FOLIO	Sigla del folio en que se registra la información
OT	Un número asignado a cada folio
CODIGO DIGITADOR	Un número asignado al digitador
TURNOS	Turno en que el digitador hizo el registro (AM-PM)
ZONA	Zona donde se digitaron los datos
FECHA INGRESO	Fecha de ingreso de los datos
NOMBRE FISCALIZADOR	Nombre del fiscalizador
CODIGO FISCALIZADOR	Número identificador de cada fiscalizador

Fuente: Programa Nacional de Fiscalización (2012)

Tabla A-2 Descripción de variables medidas por los fiscalizadores

Variable	Descripción
FECHA	Fecha en que se hizo el registro
HORA	Hora en que el fiscalizador hace el registro
MINUTOS	Minuto en que el fiscalizador hace el registro
PUERTA NÚMERO	Puerta en que el fiscalizador hace el registro
INGRESAN	Pasajeros que entran por la puerta medida
NO VALIDAN	Pasajeros que no pagan por la puerta medida
SALEN	Pasajeros que salen por la puerta medida
IO	Índice de ocupación del bus (A,B,C,D y E)
TP	Tipo de Paradero: Primario (P) o Zona Paga (ZP)
SERVICIO	Número identificador del servicio.
TIPO	Tipo de bus: Troncal (T) o Alimentador (A)
PLACA PATENTE	Patente del bus
NOMBRE DEL SERVICIO	Nombre del servicio del bus
NÚMERO DE PUERTAS	Número de puertas del bus (entre 2 y 4)
LUGAR INICIO	Lugar de inicio del recorrido
HORA INICIO	Hora de inicio del recorrido

Fuente: Programa Nacional de Fiscalización (2012)

Nº	Hora	Ingresan	No Validan	Salen	Indice		COMUNA
					IO	TP	
1	:						
2	:						
3	:						
4	:						
5	:						
6	:						
7	:						
8	:						
9	:						
10	:						
11	:						
12	:						
13	:						
14	:						
15	:						
16	:						
17	:						
18	:						
19	:						
20	:						
21	:						
22	:						
23	:						
24	:						
25	:						

	SUBSECRETARÍA DE TRANSPORTES PROGRAMA DE FISCALIZACIÓN		<i>Uso Interno del Programa de Fiscalización</i> Categoría: Nivel 3 Fecha: Enero 2015 Versión: 00
	FICHA DE TOMA DE DATOS DEL INDICE DE EVASION		
Servicio PPU Número de Puertas Puerta Número Lugar de Inicio	<input type="text"/> <input type="text"/> <input type="text"/> 2, 3 o 4 <input type="text"/> Puerta número 1, 2, 3 o 4 (donde la puerta 1 es la de adelante) <input type="text"/>	Por ejemplo: J01	FOLIO : <input type="text"/> FECHA : ___ / ___ / 2015 <small>IO: índice Ocupación del Bus (*) A (lleno), B (Medio), C (Vacío) TP: tipo de paradero (P Principal, Z zona paga)</small>

Figura A-1 Ficha de toma de datos usada por el Programa Nacional de Fiscalización

Fuente: Programa Nacional de Fiscalización (2015)

ANEXO B: RAZÓN DE TASAS DE INCIDENCIA Y FACTOR DE CAMBIO

El indicador IRR (del inglés *Incidence Rate Ratio*) representa cuántas veces la variable dependiente aumenta o disminuye con respecto a su valor esperado, cuando el valor de una variable explicativa sube en una unidad (Hilbe, 2011). Entonces, en el caso de las variables lineales, el indicador IRR para cualquier variable explicativa lineal se puede determinar con las primeras tres ecuaciones:

$$\ln(\lambda_1) = c + \beta x_1 \quad (\text{B.1})$$

$$\ln(\lambda_2) = c + \beta x_2 \quad (\text{B.2})$$

$$x_2 = x_1 + 1 \quad (\text{B.3})$$

x_1	:	Valor de la variable independiente en la situación inicial
x_2	:	Valor de la variable independiente en la situación final (aumento unitario)
λ_1	:	Tasa de evasión en la situación inicial
λ_2	:	Tasa de evasión en la situación final
β	:	Parámetro estimado asociado a la variable independiente
c	:	Constante

$$\text{IRR} = \frac{\lambda_2}{\lambda_1} = e^\beta \quad (\text{B.4})$$

Con el IRR se puede calcular un indicador denominado factor de cambio (FC). El indicador FC representa el cambio porcentual en el valor esperado de conteos medidos en la variable dependiente, ante un cambio unitario de una variable independiente. Entonces, el indicador de una variable explicativa k (con forma lineal) está dado por:

$$\text{FC}_k = 100 \left(\frac{\lambda_2 - \lambda_1}{\lambda_1} \right) = 100(\text{IRR}_k - 1) \quad (\text{B.5})$$

Es fácil notar que la Ecuación (B.5) también se puede utilizar para calcular los IRR y factores de cambio de variables mudas. En cambio, si los indicadores se calculan para variables categóricas de más de dos grupos, el valor de los indicadores será distinto según el grupo de referencia que se defina. Utilizando álgebra básica y sin pérdida de

generalidad (i.e. el análisis es válido, independiente del número de categorías), se puede plantear el siguiente sistema de ecuaciones para calcular el indicador IRR de variables categóricas creadas a partir de tres grupos (e.g. las dos variables mudas creadas para capturar las diferencias en la evasión entre los tres grupos de nivel socioeconómico):

$$x_1 + x_{1,2} + x_{1,3} = 1 \quad (\text{B.6})$$

$$\ln(\lambda) = c + \beta_2 x_{1,2} + \beta_3 x_{1,3} \quad (\text{B.7})$$

$$\ln(\lambda_{1,2}) = c + \beta_3 x_{1,3} \quad (\text{B.8})$$

$$\ln(\lambda_{1,3}) = c + \beta_2 x_{1,2} \quad (\text{B.9})$$

$$\ln(\lambda_{2,2}) = c + \beta_2 \quad (\text{B.10})$$

$$\ln(\lambda_{2,3}) = c + \beta_3 \quad (\text{B.11})$$

Donde

c	:	Constante
λ	:	Tasa de evasión en la situación inicial
$\lambda_{1,2}$:	Tasa de evasión cuando la variable muda asociada a la segunda categoría es igual a cero (situación inicial)
$\lambda_{1,3}$:	Tasa de evasión cuando la variable muda asociada a la tercera categoría es igual a cero (situación inicial)
$\lambda_{2,2}$:	Tasa de evasión cuando la variable muda asociada a la segunda categoría es igual a uno (situación final)
$\lambda_{2,3}$:	Tasa de evasión cuando la variable muda asociada a la segunda categoría es igual a uno (situación final)
x_1	:	Valor de la variable muda de referencia asociada a la primera categoría (igual a cero en la situación inicial y final)
$x_{1,2}$:	Valor de la variable muda asociada a la segunda categoría en la situación inicial (igual a cero)
$x_{1,3}$:	Valor de la variable muda asociada a la tercera categoría en la situación inicial (igual a cero)
β_2	:	Parámetro estimado de la variable muda asociada a la segunda categoría
β_3	:	Parámetro estimado de la variable muda asociada a la tercera categoría.

A partir de las ecuaciones anteriores, y usando la primera categoría como grupo de referencia, se deduce que los indicadores IRR para las variables mudas asociadas a la segunda y tercera categoría son iguales a:

$$\text{IRR}_2 = \frac{\lambda_{1,2}}{\lambda_{2,2}} = e^{\beta_2} \quad (\text{B.12})$$

$$\text{IRR}_3 = \frac{\lambda_{1,3}}{\lambda_{2,3}} = e^{\beta_3} \quad (\text{B.13})$$

Como se observa en la Ecuación (B.12) y (B.13), al dejar como base del análisis el grupo de referencia de las variables mudas, los indicadores IRR son iguales a los obtenidos con las variables lineales (Ecuación (B.4)). Por consiguiente, los indicadores FC en este caso pueden ser derivados directamente de la Ecuación (B.5). Cuando la variable explicativa tiene forma logarítmica, su indicador FC (Ecuación (B.18)) se puede derivar parametrizando su indicador IRR por un ponderador η que refleja cuántas veces aumenta la variable independiente (Ecuación (B.17)). Así, los indicadores IRR y los indicadores FC se pueden derivar a partir de las primeras tres ecuaciones:

$$\ln(\lambda_1) = c + \beta_k \log(x_{k,1}) \quad (\text{B.14})$$

$$\ln(\lambda_2) = c + \beta_k \log(x_{k,2}) \quad (\text{B.15})$$

$$x_{k,2} = \eta x_{k,1} \quad (\text{B.16})$$

$$\text{IRR} = \frac{\lambda_2}{\lambda_1} = \eta^{\beta_k} \quad (\text{B.17})$$

$$\text{FC}_k = 100 \left(\frac{\lambda_2 - \lambda_1}{\lambda_1} \right) = 100(\eta^{\beta_k} - 1) \quad (\text{B.18})$$

Notar que cuando η es igual a 1,01, el indicador FC es equivalente a la elasticidad porcentual de la variable dependiente con respecto a la variable independiente. Por lo tanto, si el parámetro estimado β_k es igual a uno, hay una tasa (o razón) constante entre la variable independiente y la dependiente. Mientras que si β_k es mayor o menor a uno, ante aumentos en la variable independiente, la tasa crece o disminuye, respectivamente.

ANEXO C: ESPECIFICACIONES ALTERNATIVAS DE LOS MODELOS DE REGRESIÓN LINEAL MÚLTIPLE

Para modelar con RLM información asimétrica y continua (e.g. los datos obtenidos de un proceso de conteo), la literatura sugiere utilizar una transformación logarítmica de la variable dependiente (Cameron y Trivedi, 1998). Esta transformación permite obtener errores homocedásticos y una distribución condicional más simétrica, y es particularmente convenientemente, en los casos donde la variable está distribuida log-normal (en este caso la nueva variable dependiente tiene distribución normal). Además, para evitar que el logaritmo natural se indefina cuando la tasa de eventos es nula, como solución estándar se suma una constante arbitraria (k) al argumento de la función logarítmica, y luego se estima $\ln(y + k)$ por OLS (Cameron y Trivedi, 1998). Sin embargo, algunos autores critican la arbitrariedad para definir el valor de la constante en el argumento del logaritmo y enfatizan que los modelos se desempeñan pobremente cuando se utiliza información proveniente de un proceso de conteo (King, 1989).

En base a lo anterior, en el modelo de regresión lineal múltiple presentado en el cuerpo del informe (modelo RLM), la variable dependiente fue definida como el logaritmo de la suma entre la tasa de evasión (calculada como la división entre el número de evasores y el número pasajeros entrando por la puerta medida) y una constante igual a uno (sumada en el argumento de la función). A pesar de utilizar esta transformación, la especificación no cumplió con los supuestos de los modelos de regresión lineal múltiple (Sección 4.6), y tuvo un desempeño muy inferior a los dos modelos de regresión de conteo (Sección 4.7). Para corregir este problema y evaluar si se producía una mejora, se formularon tres especificaciones alternativas de modelos de regresión lineal múltiple.

En la primera formulación alternativa (modelo RLM1), la variable dependiente fue definida como la tasa de evasión (sin la función logarítmica). A diferencia del modelo RLM, el número de pasajeros entrando por la puerta del bus fue incluido dentro del conjunto de variables explicativas (*Entradas*). En la segunda formulación alternativa (modelo RLM2), se utilizó la misma forma funcional de la función de enlace definida en los modelos de regresión de conteo. Por lo tanto, la variable dependiente fue definida

como el logaritmo del número de evasiones (se sumó una constante igual a uno al argumento de la función logarítmica para que esta no se indefina con conteos de cero evasión), y la variable *Entradas* fue incluida con una forma funcional logarítmica dentro del conjunto de predictores. Finalmente, en la tercera formulación alternativa (modelo RLM3), la variable dependiente fue definida como el número de evasiones, y la variable *Entradas* fue incluida con una forma funcional lineal en el conjunto de variables explicativas. La Tabla C-1 presenta los resultados de la estimación del modelo RLM y de sus tres especificaciones alternativas (RLM1, RLM2 y RLM3).

Tabla C-1 Estimación modelos de regresión lineal múltiple (RLM)

Variable (test-t)	Log (Tasa ^a + 1) (RLM)	Tasa ^a (RLM1)	Log (Evasión + 1) (RLM2)	Evasión (RLM3)
Operación puertas				
Log (Entradas)	-	-	0,333 (41,2)	-
Entradas	-	-	-	0,167 (44,2)
Salidas	0,044 (3,8)	0,065 (4,1)	0,105 (4,0)	0,301 (4,1)
Puertas	0,032 (6,3)	0,043 (6,1)	0,064 (5,5)	0,158 (4,8)
Puerta Delantera	-0,391 (-24,0)	-0,586 (-26,3)	-0,595 (-16,0)	-1,436 (-13,7)
Periodo del día				
Punta Mañana	-0,022 (-2,3)	-0,025 (-2,0)	-0,057 (-2,7)	-0,158 (-2,6)
Tarde	0,026 (2,9)	0,037 (3,1)	0,046 (2,3)	0,085 (1,5)
Noche	0,064 (2,7)	0,098 (3,0)	0,126 (2,3)	0,210 (1,3)
Calidad servicio				
Baja Ocupación	-0,043 (-4,8)	-0,055 (-4,5)	-0,098 (-4,8)	-0,279 (-4,8)
Headway	0,004 (3,0)	0,005 (3,1)	0,009 (3,4)	0,011 (1,4)
Operador				
Operadores I	-0,043 (-4,4)	-0,062 (-4,7)	-0,092 (-4,1)	-0,217 (-3,5)
Operadores II	-0,075 (-7,4)	-0,104 (-7,4)	-0,179 (-7,7)	-0,437 (-6,7)
Operadores III	-0,083 (-8,2)	-0,117 (-8,4)	-0,177 (-7,7)	-0,493 (-7,6)
Intercepto	0,485 (16,8)	0,703 (17,8)	0,558 (8,5)	1,523 (8,2)
R ² Ajustado	0,176	0,196	0,317	0,331
Nº de observaciones	5.334	5.334	5.334	5.334
Shapiro-Wilk	0,00 (W=0,85)	0,00 (W=0,82)	0,00 (W=0,98)	0,00 (W=0,80)
Lilliefors	0,23	0,22	0,08	0,16
Breusch-Pagan	197,2 (gdl=11)	191,5 (gdl=11)	1621,9 (gdl=12)	1340,5(gdl=12)

Nota: Los indicadores R² ajustado de las tres especificaciones no son comparables

^a Razón entre el número de evasiones y entradas medidas en cada observación

Los resultados de los dos test de normalidad rechazan con un 99% de confianza la hipótesis nula de que los errores tienen distribución Normal, y el test de Breusch-Pagan rechaza la existencia de distribución homocedástica de los errores aleatorios en los cuatro modelos. Estos resultados confirman nuevamente que los modelos RLM no son una especificación adecuada para modelar datos obtenidos de un proceso de conteo.

Con la metodología propuesta en la Sección 4.7, también se construyó un gráfico para comparar las diferencias en el nivel de ajuste de cada modelo. A pesar que la forma funcional sugerida en la literatura para trabajar con datos de conteo (RLM) no cumple con los supuestos de homocedasticidad y heterocedasticidad, en la Figura C-1 se observa una mejora significativa del ajuste estadístico en comparación con las tres especificaciones alternativas (RLM1, RLM2 y RLM3).

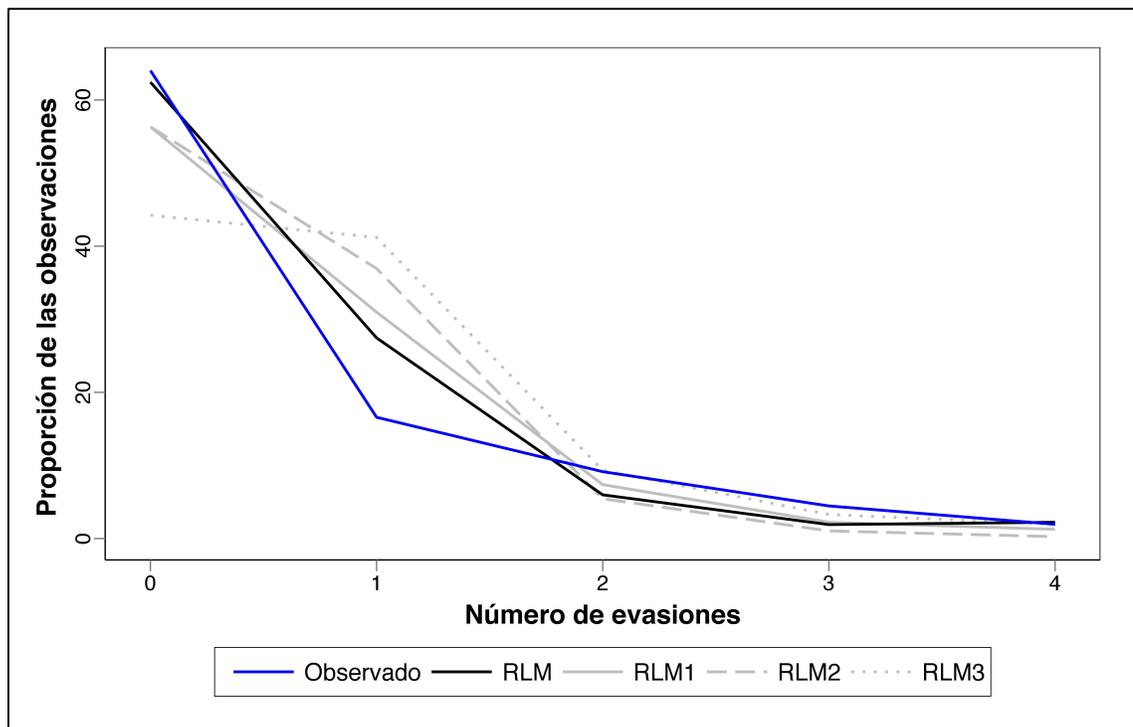


Figura C-1 Valores observados y predicciones modelos de regresión lineal múltiple

Fuente: Elaboración propia

ANEXO D: MODELO DE REGRESIÓN BINOMIAL NEGATIVA CON ESTRATIFICACIÓN POR ZONAS

Para capturar el efecto del nivel socioeconómico (NSE) en el modelo de regresión binomial negativa (NB2), se utilizaron dos métodos. En el primer método se asignó el nivel de ingreso en base a la zona recorrida por el servicio de buses donde fue tomada cada observación. En cambio, en el segundo método, se creó una variable categórica del de NSE según el ingreso familiar mensual promedio de la municipalidad donde la observación fue recolectada. A pesar que los resultados obtenidos con el segundo método fueron muy superiores, para estimar la comuna se requirió información adicional y la aplicación de un algoritmo computacional (Sección 4.3.6). En este sentido, la ventaja del primer enfoque es que puede ser utilizado directamente con la información disponible en la base de datos de evasión.

Para capturar el efecto del nivel socioeconómico (NSE), el primer método utiliza como referencia el la zona de Transantiago donde fue medida cada observación. Sin embargo, como la base de datos no cuenta con información explícita respecto la zona correspondiente a cada paradero, esta solo pudo ser inferida para las observaciones recolectadas en servicios alimentadores (primera letra del código identificador). En contraste, como los servicios troncales viajan entre varias zonas, el código identificador no permite identificar la zona específica donde cada observación fue recolectada (Sección 4.3.4).

Luego, a partir de los tramos de ingreso por quintil definidos en la encuesta CASEN 2009 (Tabla 4-4), las zonas fueron agrupadas en tres categorías de nivel socioeconómico. Como ninguna zona tiene un ingreso promedio familiar mensual menor que los dos primeros quintiles ni mayor que el quinto quintil, se pudo definir un máximo de tres tramos de ingreso (Tabla D-1). El nivel de ingreso de cada zona fue estimado como el promedio ponderado entre el ingreso de las comunas dentro de la zona y su población (el ingreso por zona no estaba disponible).

Tabla D-1 Promedio ponderado del nivel de ingreso en cada zona de Transantiago

Zona	Ingreso	Población	Categoría de Ingreso
A	1.172.100	159.919	III
B	650.979	119.084	II
C	2.717.412	152.055	III
D	1.251.329	145.776	III
E	865.221	259.665	II
F	593.448	757.721	I
G	558.692	167.280	I
H	587.011	83.303	I
I	877.275	354.404	II
J	570.521	145.336	I

Nota: Zona Ingreso Bajo (I); Zona Ingreso Medio (II); Zona Ingreso Alto (III)

Fuente: Elaboración Propia

Posteriormente, se definieron dos variables mudas como *proxies* del NSE en el modelo NB2: **Zona Ingreso Medio** (zonas B, E y I) y **Zona Ingreso Alto** (zonas A, D y C), las cuales toman el valor uno, si la zona donde las observaciones fueron recolectadas pertenecen a la segunda o tercera categoría de ingreso respectivamente (Tabla D-1), y cero en otro caso (las zonas G, J, H y F se dejaron como grupo de referencia). Además, para no excluir todas las observaciones recolectadas en servicios troncales en la estimación de los modelos, se creó una variable asociada al tipo de servicio (**Alimentador**). Esta variable cumple un rol muy importante; por un lado, captura las diferencias promedio en evasión entre los servicios de buses troncales y alimentadores no explicadas por las demás variables incluidas en el modelo, y por otro, permite que el valor de los parámetros de las dos variables asociadas al NSE de los clústeres de zonas que recorre cada servicio alimentador, sean estimados solo respecto al grupo base de los servicios alimentadores (zonas de ingreso bajo).

Finalmente, se estimó un nuevo modelo de regresión (modelo NB2-Z), incluyendo las dos variables de estratificación por zona (**Zona Ingreso Medio**, **Zona Ingreso Alto**) y del tipo de servicio (**Alimentador**). Luego, este modelo fue comparado con el modelo NB2-I y el modelo NB2 (sin estratificación por ingreso) para verificar cuál de las dos metodologías para capturar el efecto del NSE en la evasión es superior. En la Tabla D-2

se presentan los resultados del modelo NB2 (sin estratificación de ingreso), el modelo NB2-I y el modelo NB2-Z. Como se discutió previamente, el modelo NB2-I utiliza como criterio de estratificación el nivel socioeconómico de la municipalidad donde fue medida la observación (Sección 4.3.6), mientras que en el modelo NB2-Z, la estratificación es hecha en base el nivel promedio de ingreso de la zona de Transantiago donde fue recolectada cada medición (solo se captura el efecto en los servicios alimentadores, pero las observaciones de servicios troncales también son consideradas).

Tabla D-2 Estimación modelos de regresión binomial negativa con efecto ingreso

Variable (t-test)	NB2	NB2-I	NB2-Z
Operación puertas			
Log (entradas)	1,079 (42,7)	1,079 (42,8)	1.080 (42.7)
Salidas	0,221 (2,8)	0,215 (2,7)	0.217 (2.7)
Puertas	0,210 (5,6)	0,186 (4,9)	0.156 (3.3)
Puerta Delantera	-1,336 (-14,0)	-1,344 (-14,1)	-1.347 (-14.1)
Periodo del día			
Punta Mañana (08:30-12:30)	-0,166 (-2,4)	-0,140 (-2,0)	-0.152 (-2.2)
Tarde (12:30-20:30)	0,231 (3,9)	0,253 (4,3)	0.231 (3.9)
Noche (20:30-22:00)	0,499 (2,8)	0,547 (3,1)	0.508 (2.9)
Calidad servicio			
Baja Ocupación	-0,150 (-2,6)	-0,174 (-3,0)	-0.176 (-3.0)
Headway	0,026 (2,9)	0,018 (1,9)	0.026 (2.8)
Operación ruta			
Alimentador	-	-	-0.069 (-1.0)
Operadores I (Alsacia, Vule)	-0,271 (-4,0)	-0,228 (-3,3)	-0.303 (-3.9)
Operadores II (Express, Metbus)	-0,567 (-7,8)	-0,445 (-5,9)	-0.583 (-6.8)
Operadores III (Subus, Redbus)	-0,661 (-8,9)	-0,608 (-8,1)	-0.641 (-8.1)
Nivel socioeconómico^a (NSE)			
Municipalidad Ingreso Medio-Alto	-	-0,233 (-3,9)	-
Municipalidad Ingreso Alto	-	-0,366 (-4,1)	-
Zona Ingreso Medio (B,E,I)	-	-	-
Zona Ingreso Alto (A,D,C)	-	-	-0.208 (-2.0)
Intercepto			
Nº de observaciones	5.334	5.534	5.534
Log-verosimilitud	-5.477,8	-5.464,1	-5.474,5
AIC	10.981,5	10.958,2	10.978.9

^a Los tramos de ingreso para agrupar las municipalidades están descritos en la Sección 4.3.6.

En los tres modelos, todos los parámetros de las variables incluidas en el modelo base son estadísticamente significativos al 95%, excepto el parámetro de la variable *Headway* en el modelo NB2-I que fue significativo solo al 94% de confianza. En cuanto a las variables para controlar el efecto del NSE en el modelo NB2-Z, solo la variable *Zona Ingreso Alto* fue estadísticamente significativa, lo cual indicaría que la evasión en promedio en zonas de altos ingresos es menor que la evasión observada en zonas de ingreso bajo y medio. Como la variable *Zona Ingreso Medio* no fue significativa, los resultados sugieren que no existen diferencias significativas entre los servicios alimentadores que recorren zonas de bajos o medianos ingresos (zonas B, E, F, G, H, I y J). A pesar que la variable *Alimentador* en el modelo NB2-Z fue significativa solo al 65% de confianza, se mantuvo en el modelo para que el parámetro estimado de la variable *Zona Ingreso Alto* fuese estimado solo respecto al grupo de servicios alimentadores (ya discutido anteriormente). La baja significancia estadística de esta variable se debe principalmente a la alta correlación con las variables categóricas usadas para controlar el efecto de las unidades de negocio en la evasión (Sección 4.3.6).

Para evaluar si la inclusión de variables relacionadas con el NSE produce un aumento significativo en el poder explicativo del modelo base (NB2) y definir cuál de los métodos de estratificación por NSE es superior, se utilizaron dos indicadores estadísticos; el criterio de información de Akaike (AIC) y el test de log-verosimilitud (LR). Como la diferencia entre los indicadores AIC de NB2 y NB2-I es mayor a 10 ($\Delta AIC = 23.3$), NB2-I es preferido independiente del número de observaciones utilizadas para estimar los modelos (Hilbe, 2011). Por otro lado, los test LR entre NB2 y NB2-Z ($27.4 > \chi_{2,99\%}^2$), y entre NB2 y NB2-I ($20.7 > \chi_{2,99\%}^2$), indican que tanto el modelo NB2-I como el NB2-Z tienen un mejor ajuste que NB2 (Tabla D-2). Con estos resultados, se concluye que los modelos que incluyen variables de estratificación por NSE, ya sea por la municipalidad o la zona donde está ubicado el paradero, aumentan de forma significativa el poder explicativo del modelo base (NB2). Por otra parte, con la estratificación del NSE en base al nivel de ingreso de las municipalidades se obtiene un ajuste muy superior, comparado con la estratificación por zonas (modelo NB2-Z).

ANEXO E: VARIABLES NO SIGNIFICATIVAS EN LOS MODELOS DE REGRESIÓN

En la Tabla E-1 se muestran los resultados de los modelos RLM, RP y NB2 estimados incluyendo las variables de tipo de servicio (*Alimentador*), día de la semana (*Miércoles* y *Jueves*) y la frecuencia (en vez de *Headway*), denominados RLM-0, RP-0 y NB2-0, respectivamente. En los tres modelos, los parámetros de las cuatro variables no fueron significativos al 95% de confianza, excepto la variable *Miércoles* en el modelo RP.

Tabla E-1 Estimación modelos con variable de tipo de servicio, día y frecuencia

Variable (test-t)	RLM-0	RP-0	NB2-0
Operación puertas			
Log (entradas)	-	0,979 (59,0)	1,078 (42,6)
Salidas	0,045 (3,9)	0,291 (5,5)	0,234 (3,0)
Puertas	0,029 (4,6)	0,131 (4,0)	0,139 (3,0)
Puerta Delantera	-0,390 (-23,9)	-1,097 (-18,1)	-1,323 (-13,9)
VARIABLES TEMPORALES			
Punta Mañana (08:30-12:30)	-0,018 (-1,9)	-0,190 (-3,8)	-0,136 (-2,0)
Tarde (12:30-20:30)	0,029 (3,2)	0,187 (4,5)	0,250 (4,1)
Noche (20:30-22:00)	0,068 (2,8)	0,388 (2,9)	0,532 (3,0)
Miércoles	-0,006 (-0,7)	-0,093 (-2,1)	-0,082 (-1,4)
Jueves	-0,005 (-0,6)	-0,052 (-1,3)	-0,060 (-1,1)
Calidad servicio			
Baja Ocupación	-0,040 (-4,5)	-0,128 (-3,4)	-0,139 (-2,4)
Frecuencia ^a	-0,001 (-0,6)	0,004 (0,8)	-0,003 (-0,5)
Operación ruta			
Alimentador	0,008 (0,8)	0,012 (0,2)	-0,034 (-0,5)
Operadores I (Alsacia, Vule)	-0,038 (-3,4)	-0,251 (-4,4)	-0,275 (-3,4)
Operadores II (Express, Metbus)	-0,071 (-5,7)	-0,501 (-7,7)	-0,576 (-6,3)
Operadores III (Subus, Redbus)	-0,087 (-8,3)	-0,649 (-11,7)	-0,707 (-9,1)
Intercepto			
R ² Ajustado	0,175	-	-
Log-verosimilitud	-	-5.832,548	-5.480,121
AIC	-	11.697,090	10.992,240
Nº de observaciones	5.334	5.334	5.334

^a La variable *Headway* no fue incluida porque fue incluida la variable *Frecuencia* (Sección 4.3.3)

Fuente: Elaboración propia

Como se discutió en la Sección 5.1, la variable muda *Ingreso Medio-Bajo* no fue significativa en el modelo NB2-I. Por esta razón, el grupo de referencia de las variables para capturar el nivel socioeconómico (NSE) incluyó las observaciones medidas en municipalidades de ingreso bajo y medio-bajo. En la Tabla E-2, se presentan los resultados del modelo NB2-I-1 (incluye las tres variables categóricas para controlar el efecto del NSE) y del modelo NB2-I (solo incluye las dos variables del NSE que fueron estadísticamente significativas al 95% de confianza).

Tabla E-2 Estimación modelo NB2-I con todas las variables categóricas del NSE

Variable (test-t)	NB2-I	NB2-I-1
Operación puertas		
Log (entradas)	1,079 (42,8)	1,079 (42,8)
Salidas	0,215 (2,7)	0,218 (2,8)
Puertas	0,186 (4,9)	0,181 (4,7)
Puerta Delantera	-1,344 (-14,1)	-1,346 (-14,2)
Periodo del día		
Punta Mañana (08:30-12:30)	-0,140 (-2,0)	-0,134 (-1,9)
Tarde (12:30-20:30)	0,253 (4,3)	0,255 (4,3)
Noche (20:30-22:00)	0,547 (3,1)	0,547 (3,1)
Calidad servicio		
Baja Ocupación	-0,174 (-3,0)	-0,173 (-3,0)
Headway	0,018 (1,9)	0,016 (1,8)
Operación ruta		
	-	
Operadores I (Alsacia, Vule)	-0,228 (-3,3)	-0,197 (-2,7)
Operadores II (Express, Metbus)	-0,445 (-5,9)	-0,394 (-4,7)
Operadores III (Subus, Redbus)	-0,608 (-8,1)	-0,575 (-7,3)
Nivel socioeconómico (NSE)^a		
Municipalidad Ingreso Medio-Bajo	-0,233 (-3,9)	-0,076 (-1,4)
Municipalidad Ingreso Medio-Alto	-0,366 (-4,1)	0,281 (-4,1)
Municipalidad Ingreso Alto	-	-0,418 (-4,4)
Intercepto		
	-0,722 (-3,6)	-0,695 (-3,5)
Nº de observaciones	5.534	5.534
Log-verosimilitud	-5.464,1	-5.463,1
AIC	10.958,2	10.958,2

^a Los tramos de ingreso para agrupar las municipalidades se describen en la Sección 4.3.6

Fuente: Elaboración propia