

DOCUMENTO DE TRABAJO DEL BID N° IDB-WP-01352

Subsidios al transporte público en América Latina desde una perspectiva de eficiencia: aplicación a Bogotá, Colombia

Julián Gómez Gélvez
Carlos Mojica

Banco Interamericano de Desarrollo
División de Transporte

Junio 2022

Subsidios al transporte público en América Latina desde una perspectiva de eficiencia: aplicación a Bogotá, Colombia

Julián Gómez Gélvez
Carlos Mojica

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo
Gómez-Gélvez, Julián A.

Subsidios al transporte público en América Latina desde una perspectiva de eficiencia:
aplicación a Bogotá, Colombia / Julián A. Gómez Gélvez, Carlos H. Mojica.

p. cm. — (Documento de trabajo del BID ; 1352)

Incluye referencias bibliográficas.

1. Transportation-Colombia-Finance. 2. Subsidies-Colombia. 3. Urban transportation-Colombia-Finance. 4. Local transit-Colombia-Finance. I. Mojica, Carlos. II. Banco Interamericano de Desarrollo. División de Transporte. III. Título. IV. Serie. IDB-WP-1352

Códigos JEL: H20, L91, R0

Palabras clave: subsidios, transporte público, eficiencia, Bogotá, Colombia

<http://www.iadb.org>

Copyright © 2022 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Después de un proceso de revisión por pares, y con el consentimiento previo y por escrito del BID, una versión revisada de esta obra podrá reproducirse en cualquier revista académica, incluyendo aquellas referenciadas por la Asociación Americana de Economía a través de EconLit, siempre y cuando se otorgue el reconocimiento respectivo al BID, y el autor o autores no obtengan ingresos de la publicación. Por lo tanto, la restricción a obtener ingresos de dicha publicación sólo se extenderá al autor o autores de la publicación. Con respecto a dicha restricción, en caso de cualquier inconsistencia entre la licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas y estas declaraciones, prevalecerán estas últimas.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



Subsidios al Transporte Público en América Latina desde una Perspectiva de Eficiencia: Aplicación a Bogotá, Colombia

Julián A. Gómez Gélvez Carlos H. Mojica

19 de mayo de 2022

Resumen

Este estudio busca evaluar el nivel de subsidio al transporte público que puede ser justificado desde una perspectiva de eficiencia en ciudades latinoamericanas. En una aplicación específica para Bogotá, Colombia, encontramos que altos niveles de subsidio solo se pueden justificar si la agencia de transporte público responde a aumentos en la demanda con aumentos casi proporcionales en la oferta de servicios. De lo contrario, el costo generado por un mayor hacinamiento en los buses supera los beneficios por reducción de tiempos de espera (efecto Mohring) y externalidades (congestión es la principal externalidad). Adicionalmente, una revisión de literatura identifica áreas en las que son necesarios estudios empíricos adicionales en la región para mejorar la calidad de los análisis de tarificación óptima en transporte. Estas áreas incluyen: 1) el efecto marginal de vehículos en congestión, 2) la elasticidad cruzada de las tarifas de transporte público al uso de vehículos particulares; y 3) el efecto de los precios de transporte en el mercado laboral.

Contenido

1. Introducción	3
2. Modelo económico	5
2.1. Demanda	5
2.2. Características de los modos de transporte	7
2.3. Externalidades por polución	9
2.4. Costos operacionales y variables de decisión de la agencia de transporte público	10
2.5. Subsidio óptimo	11
3. Revisión de literatura	14
3.1. Estudios sobre subsidios óptimos al transporte público	14
3.2. Elasticidades al precio propio y elasticidades cruzadas	15
3.3. Efecto marginal en congestión	18
3.4. El costo marginal de los recursos públicos	19
4. Aplicación a Bogotá, Colombia	21
4.1. Parámetros base	22
4.1.1. Demanda y oferta	22
4.1.2. Costos operacionales y tarifas	25
4.1.3. Tiempos de viaje, costo generalizado y elasticidad de la demanda . .	27
4.1.4. Externalidades	29
4.1.5. Elasticidades del tiempo de espera y del hacinamiento	31
4.2. Resultados	33
4.2.1. Subsidios y tarifas óptimas como función de la elasticidad de la oferta a la demanda	33
4.2.2. Componentes de la tarifa óptima	35
4.2.3. Sensibilidad a parámetros clave	36
5. Conclusiones	40
Referencias	43
A. Condiciones de primer orden	46
B. Subsidio óptimo	47
C. Elasticidad de la demanda al costo generalizado	52

1. Introducción

Varias ciudades de América Latina y el Caribe (ALC) no cubren los costos operacionales de sus sistemas de transporte público con los ingresos provenientes del cobro de tarifas (Rivas, Suárez-Alemán, y Serebrisky, 2019a), por lo que en muchos casos existe un nivel significativo de subsidio. Uno de los objetivos principales de estos subsidios es mejorar la asequibilidad al transporte público de personas de bajos ingresos y otros grupos vulnerables, por lo que la mayoría son implementados como subsidios a la demanda y focalizados a dichos grupos (Rivas, Suárez-Alemán, y Serebrisky, 2019b).

Sin embargo, existen también argumentos que respaldan los subsidios al transporte público desde una perspectiva de eficiencia. Estos argumentos se pueden agrupar en dos categorías principales (Parry y Small, 2009). En primer lugar, la provisión de transporte público presenta generalmente importantes economías de escala. Estas economías de escala pueden surgir tanto de la existencia de costos operacionales fijos, como el mantenimiento de estaciones, como del efecto positivo en reducción de tiempos de espera que conlleva la expansión de la oferta de transporte público (conocido como el efecto Mohring (Mohring, 1972)).¹ En segundo lugar, tarifas menores para el transporte público pueden desincentivar el uso de carros particulares y reducir sus impactos en externalidades negativas como congestión, polución y accidentalidad. Este argumento asume que estas externalidades no son internalizadas por los usuarios de carros particulares mediante, por ejemplo, cobros por congestión, por lo que es un segundo mejor argumento. Estos argumentos de eficiencia respaldan subsidios generales a todos los usuarios de transporte público, no solo a grupos vulnerables específicos.

El objetivo principal de este documento es evaluar el nivel de subsidio al transporte público que puede ser justificado desde una perspectiva de eficiencia en ciudades de ALC, así como el grado en el que las diferentes categorías de argumentos (economías de escala y externalidades) contribuyen a este nivel óptimo.

Este documento contiene tres secciones principales. La primera sección presenta un modelo económico que permite expresar el nivel de subsidio óptimo como la suma de varios componentes que reflejan los diferentes argumentos de eficiencia. Este modelo clarifica la forma como opera cada argumento y resalta los parámetros clave que determinan el tamaño de cada componente. Tomamos como base el modelo utilizado por Parry y Small (2009) para evaluar los subsidios al transporte público en Londres, Los Angeles y Washington D.C. Sin embargo, ajustamos este modelo con el fin de cuantificar directamente el efecto del nivel de

¹Los costos de infraestructura también se pueden considerar como una fuente de economías de escala en los sistemas de transporte público. Sin embargo, nuestro análisis no pretende evaluar la conveniencia de nuevas inversiones en infraestructura. Por esta razón, tenemos en cuenta solo costos operacionales, en línea con la literatura sobre subsidios al transporte público.

hacinamiento en los buses en el valor del tiempo de viaje e incluir su efecto en los tiempos de espera. Estos ajustes son importantes para aplicar el modelo a ciudades latinoamericanas, donde el nivel de hacinamiento es usualmente más alto.

La segunda sección presenta una revisión de literatura sobre estudios empíricos que cuantifican algunos parámetros clave para ciudades en ALC. Esta revisión nos permite identificar áreas en las que son necesarios estudios empíricos adicionales para la región. Para cuantificar los beneficios potenciales por reducción de externalidades, son necesarios estudios empíricos que identifiquen el impacto marginal de vehículos en congestión, así como estudios que estimen la elasticidad cruzada del uso de vehículos particulares a las tarifas de transporte público. Adicionalmente, estudios que permitan conocer mejor la relación entre el transporte y el mercado laboral son necesarios para evaluar en que medida el costo marginal de los recursos públicos es un argumento de peso en contra de altos subsidios al transporte público.

La tercera sección realiza un análisis detallado para el transporte público de Bogotá (BRT y buses convencionales). Encontramos que altos niveles de subsidio solo se pueden justificar en términos de eficiencia siempre que la agencia de transporte público pueda responder a aumentos en la demanda con aumentos casi proporcionales en la oferta de servicios. De lo contrario, el costo generado por un mayor hacinamiento en los buses supera los beneficios por reducción de tiempos de espera (efecto Mohring) y externalidades. Estos resultados sugieren también que el actual nivel de oferta es muy bajo, lo que genera costos excesivos para los usuarios por incomodidad durante el viaje.²

Encontramos también que la congestión es de lejos la principal externalidad que generan los usuarios de vehículos particulares y buses convencionales en Bogotá. Aunque el efecto marginal en congestión de un usuario de buses particulares es menor al de un usuario de vehículos particulares, el efecto neto en externalidades que resulta de aumentar la demanda de buses convencionales a través de reducciones de tarifa es negativo. Esto se debe a que solo un porcentaje bajo del aumento en demanda corresponde a usuarios que migran de vehículos particulares a buses convencionales. Aunque este porcentaje también es bajo para el sistema BRT TransMilenio, el efecto neto en este caso es positivo, debido a que el uso de carriles exclusivos limita el efecto en congestión.

²El modelo económico que utilizamos asume que la agencia de transporte público mantiene constante el tamaño de los buses y el número de rutas, y ajusta el número de bus-kilómetros a la demanda de acuerdo a una elasticidad constante. Un análisis que considere un ajuste óptimo en estas variables de oferta se encuentra fuera del alcance de este estudio, debido a que cambios significativos en la oferta requerirían inversiones de infraestructura y harían que varios de los parámetros marginales en los que se basa el modelo no fuesen confiables. En este sentido, nuestro análisis se enfoca en una optimización tarifaria de corto plazo que, sin restar importancia a las consideraciones de largo plazo, consideramos relevante para las agencias de transporte público de la región.

2. Modelo económico

El modelo económico que presentamos en esta sección representa los principales aspectos de la demanda y oferta de un sistema de transporte público que determinan su nivel óptimo de subsidio. Nos basamos en el modelo utilizado por Parry y Small (2009) para estimar el subsidio óptimo en Londres, Los Angeles y Washington D.C. Sin embargo, ajustamos este modelo con el fin de cuantificar directamente el efecto del nivel de hacinamiento en los buses en el valor del tiempo de viaje e incluir su efecto en los tiempos de espera. Estos ajustes son importantes para aplicar el modelo a ciudades latinoamericanas, donde el nivel de hacinamiento es usualmente más alto. La Sección 4 aplica el modelo de esta sección a los sistemas de transporte público de Bogotá, Colombia, por lo que algunos detalles del modelo se desarrollaron teniendo en mente esta aplicación.

El modelo asume que la tarifa de transporte público es igual para todos los usuarios, de forma que los subsidios no son focalizados hacia ningún grupo poblacional específico.³ Adicionalmente, el modelo considera solo costos operacionales, no costos de infraestructura. Debido a que no pretendemos evaluar la conveniencia de nuevas inversiones en infraestructura, estos costos son en gran medida costos hundidos. Finalmente, aunque esta sección hace referencia a un sistema de buses, el modelo se puede aplicar también a sistemas de trenes.

2.1. Demanda

Además de considerar la demanda de transporte público, es importante modelar también la demanda de viajes en carros particulares, debido a que uno de los principales argumentos para reducir las tarifas de transporte público es su efecto potencial en el uso de carros particulares. El modelo utiliza un marco conceptual de agente representativo, en el que este agente decide el número de kilómetros de viaje en transporte público (buses) y carro particular, así como el nivel de consumo de un bien general, con el objetivo de maximizar su utilidad. La distancia promedio por viaje se considera constante, por lo que cambios en el número de kilómetros en bus y en carro provienen de cambios en el número de viajes. La utilidad depende directamente de estos tres niveles de consumo (bus, carro y bien general), así como del tiempo de viaje en cada modo, el nivel de hacinamiento (*crowding*) en los buses y el nivel de polución causado por ambos modos. Las preferencias del agente representativo

³Los argumentos de eficiencia usualmente respaldan subsidios más altos a los usuarios de mayores ingresos. Esto se debe a que los usuarios de mayores ingresos tienen mayor elasticidad a la tarifa y utilizan en mayor medida vehículos particulares como modo alterno. Por lo tanto, al imponer una tarifa igual para todos los usuarios, sacrificamos eficiencia en favor de redistribución.

están dadas por la siguiente función de utilidad:

$$U(X, M^{bus}, M^{car}, T^{bus}, W^{bus}, A^{bus}, C^{bus}, T^{car}, Z^{pol}) \quad (1)$$

donde X representa el nivel de consumo del bien general, M^{bus} el número de kilómetros de viaje en bus y M^{car} el número de kilómetros en carro. La utilidad es creciente con respecto a estas tres variables, lo que refleja el beneficio que derivan las personas de viajar o consumir otros bienes. T^{bus} , W^{bus} y A^{bus} representan los tiempos totales en el vehículo, en espera y en acceso al bus respectivamente, mientras C^{bus} representa el nivel de hacinamiento agregado experimentado en el bus (pasajeros/m²), sin diferenciar entre pasajeros parados y sentados. T^{car} denota el tiempo en carro.⁴ La utilidad es decreciente con respecto a estas cinco variables, lo que refleja su inconveniencia para los viajeros. Los tiempos totales resultan de multiplicar el número de kilómetros por el respectivo valor unitario por kilómetro:

$$T^{bus} = t^{bus} M^{bus}; \quad W^{bus} = w^{bus} M^{bus}; \quad A^{bus} = a^{bus} M^{bus}; \quad T^{car} = t^{car} M^{car} \quad (2)$$

t^{bus} y t^{car} representan los tiempos de viaje en el vehículo por kilómetro (inverso de la velocidad promedio de buses y carros), mientras w^{bus} y a^{bus} denotan los tiempos de espera y acceso por kilómetro. Finalmente, Z^{pol} expresa el nivel de polución generado por buses y carros, el cual reduce la utilidad.

Aunque las velocidades (t^{bus} , t^{car}), el tiempo de espera por kilómetro (w^{bus}) y los niveles de hacinamiento (C^{bus}) y polución (Z^{pol}) dependen de los kilómetros de viaje en bus y carro, el agente representativo decide sus niveles de consumo y viajes tomando estos valores como fijos. Esta determinación refleja el hecho de que las decisiones son tomadas en realidad por una multitud de personas que no tienen en cuenta el efecto de sus decisiones sobre la utilidad de otros individuos y cuyo impacto en estas variables es marginal.

La restricción de ingreso del agente representativo está dada por:

$$I - Sub = X + p^{bus} M^{bus} + (p^{car} + \tau^{car}) \frac{M^{car}}{O^{car}} \quad (3)$$

donde I y Sub denotan el nivel de ingreso y la cantidad de subsidio necesaria para la operación del transporte público. El precio del bien de consumo general está normalizado a uno. p^{bus} denota la tarifa por kilómetro del bus y p^{car} el costo monetario operacional por kilómetro del carro. τ^{car} representa cargos adicionales al uso de carro particular (\$/km), que pueden provenir de impuestos a la gasolina, cobros por congestión u otras formas de cobro al uso de carros particulares. Como se verá más adelante, estos cargos adicionales tienden a internalizar

⁴Asumimos que los tiempos de espera, acceso y el nivel de hacinamiento para el carro son despreciables.

las externalidades negativas asociadas al uso del carro. o^{car} denota la ocupación promedio de los carros (ver Ecuación 9 más adelante), por lo que M^{car}/o^{car} determina el número de kilómetros vehiculares de los carros.

El agente representativo elige X , M^{bus} y M^{car} con el objetivo de maximizar su utilidad (1), sujeto a la restricción de ingreso (3) y tomando t^{bus} , w^{bus} , a^{bus} , C^{bus} , t^{car} , Z^{pol} , I , Sub , p^{bus} , p^{car} , τ^{car} y o^{car} como valores fijos. Las condiciones de primer orden de este proceso de maximización llevan a las siguientes expresiones (ver derivación en el Apéndice A):

$$\frac{U_{M^{bus}}}{U_X} = p^{bus} + \rho^{T^{bus}}(C^{bus})t^{bus} + \rho^{W^{bus}}w^{bus} + \rho^{A^{bus}}a^{bus} \quad (4)$$

$$\frac{U_{M^{car}}}{U_X} = \frac{p^{car} + \tau^{car}}{o^{car}} + \rho^{T^{car}}t^{car} \quad (5)$$

donde U_X , $U_{M^{bus}}$ y $U_{M^{car}}$ denotan las derivadas parciales de la función de utilidad respecto a las variables de decisión, por lo que $\frac{U_{M^{bus}}}{U_X}$ y $\frac{U_{M^{car}}}{U_X}$ representan las utilidades marginales de incrementar los kilómetros de viaje en bus y en carro expresadas en términos monetarios. Los valores ρ se definen como:

$$\rho^{T^{bus}}(C^{bus}) = -\frac{U_{T^{bus}}(C^{bus})}{U_X}; \quad \rho^{W^{bus}} = -\frac{U_{W^{bus}}}{U_X}; \quad \rho^{A^{bus}} = -\frac{U_{A^{bus}}}{U_X}; \quad \rho^{T^{car}} = -\frac{U_{T^{car}}}{U_X} \quad (6)$$

y representan la valoración monetaria de los tiempos en el vehículo, espera y acceso (\$/hora en todos los casos). Los valores de los tiempos de espera y acceso al bus ($\rho^{W^{bus}}$ y $\rho^{A^{bus}}$) y del tiempo de viaje en carro ($\rho^{T^{car}}$) se tomarán como constantes en la aplicación del modelo. Sin embargo, el valor del tiempo de viaje en el bus ($\rho^{T^{bus}}$) depende del nivel de hacinamiento.⁵

Las Ecuaciones 4 y 5 muestran que el agente representativo iguala la utilidad marginal de viajar en bus y en carro a sus respectivos costos generalizados (lado derecho), los cuales incluyen la tarifa o costo operacional y la valoración monetaria de los tiempos de viaje.

2.2. Características de los modos de transporte

La velocidad de buses y carros depende del flujo vehicular total debido a la congestión del tráfico. Adicionalmente, la velocidad operacional de los buses depende de los tiempos de parada, los cuales incluyen los procesos de desaceleración y aceleración, apertura y cierre de puertas, y subida y bajada de pasajeros. Expresamos entonces el tiempo de viaje en el

⁵Parry y Small (2009) consideran el costo del hacinamiento como un término adicional en la Ecuación 4. Sin embargo, la literatura reciente sobre hacinamiento en transporte público identifica su costo como un multiplicador del valor del tiempo en el vehículo (Horcher y Tirachini, 2021), por lo que ajustamos el modelo para coincidir con esta aproximación.

vehículo por kilómetro para bus y carro como:

$$\begin{aligned} t^{car} &= t^{car}(V^{car} + \alpha V^{bus}) \\ t^{bus} &= t^{bus}(V^{car} + \alpha V^{bus}) + \theta o^{bus} \end{aligned} \quad (7)$$

donde $t^{car}(\cdot)$ y $t^{bus}(\cdot)$ son funciones que determinan los tiempos de viaje por kilómetro para carros y buses (incluyendo los tiempos de desaceleración/aceleración y apertura/cierre de puertas, pero no los tiempos de subida y bajada de pasajeros) como función del tráfico vehicular total ($V = V^{car} + \alpha V^{bus}$). El tráfico vehicular total incluye los kilómetros vehiculares de los carros (V^{car}) y de los buses (V^{bus}), teniendo en cuenta que los buses contribuyen en mayor medida a la congestión (factor de congestión $\alpha > 1$) debido a su tamaño y a la necesidad de hacer paradas. o^{bus} representa el nivel de ocupación (número de pasajeros) promedio de los buses, mientras θ denota el tiempo de parada por pasajero (por subida y bajada) dividido por la longitud promedio de viaje. θo^{bus} refleja entonces el tiempo por subida y bajada de pasajeros de los buses por kilómetro.

La ocupación promedio de los buses (variable endógena en el modelo) corresponde a la razón entre el total de kilómetros de pasajeros y kilómetros vehiculares:

$$o^{bus} = \frac{M^{bus}}{V^{bus}} \quad (8)$$

donde el total de kilómetros vehiculares de bus es una variable de decisión de la agencia de transporte público.

La ocupación promedio de los carros se define de forma equivalente, pero se considera constante (variable exógena).⁶ Existe entonces una relación lineal entre el número de kilómetros de pasajeros y kilómetros vehiculares para los carros:

$$o^{car} = \frac{M^{car}}{V^{car}} \quad (9)$$

La razón entre el nivel de ocupación de los buses y su capacidad (n^{bus}) determina el factor de carga de los buses ($0 < l^{bus} < 1$):

$$l^{bus} = \frac{o^{bus}}{n^{bus}} \quad (10)$$

La capacidad de los buses se considera constante. El nivel de hacinamiento experimentado por los usuarios de transporte público es una función creciente del factor de carga de los

⁶Este supuesto implica que no consideramos el impacto que cambios en la tarifa de transporte público pueden tener en decisiones de compartir carro (*carpooling*).

buses:

$$C^{bus} = C^{bus}(l^{bus}) \quad (11)$$

Esta función tiene en cuenta que el nivel de ocupación promedio de los buses no es igual al nivel de ocupación promedio experimentado por los pasajeros, ya que un bus con mayor ocupación tiene más pasajeros (Soza-Parra, Raveau, Muñoz, y Cats, 2019).

El tiempo de espera por los buses depende de la frecuencia promedio de servicio (f^{bus} en buses/hora) y del factor de carga (debido a la probabilidad de que los usuarios no puedan abordar el primer bus que pasa luego de su arribo a la estación).⁷ La frecuencia está dada por la razón entre el número de kilómetros vehiculares y el número de kilómetros de rutas de bus (D^{bus}) multiplicado por la duración del periodo de tiempo en análisis (h , por ejemplo la hora pico de la mañana):

$$w^{bus} = w^{bus}(f^{bus}, l^{bus}); \quad f^{bus} = \frac{V^{bus}}{hD^{bus}} \quad (12)$$

donde $w^{bus}(f^{bus}, l^{bus})$ es una función decreciente con respecto a f^{bus} (mayores frecuencias reducen los tiempos de espera) y creciente respecto a l^{bus} (mayor ocupación disminuye la probabilidad de poder abordar el primer bus). El número de kilómetros de rutas de bus se considera constante, por lo que incrementos en los kilómetros vehiculares de bus aumentan la frecuencia de servicio.

Finalmente, el tiempo de acceso al bus depende de la densidad de las rutas de bus, por lo que se puede expresar como una función decreciente de los kilómetros de rutas de bus:

$$a^{bus} = a^{bus}(D^{bus}) \quad (13)$$

2.3. Externalidades por polución

El nivel de polución depende de los kilómetros vehiculares de buses y carros:

$$Z^{pol} = z_{bus}^{pol} V^{bus} + z_{car}^{pol} V^{car} \quad (14)$$

donde z_{bus}^{pol} y z_{car}^{pol} son factores de emisión por kilómetro vehicular (gr/km). Esta aproximación permite tener en cuenta diversos contaminantes (ej: material particulado y CO₂).

⁷Parry y Small (2009) definen el tiempo de espera como función únicamente de la frecuencia. El efecto del factor de carga en los tiempos de espera es importante especialmente para sistemas o periodos de tiempo con alta demanda.

2.4. Costos operacionales y variables de decisión de la agencia de transporte público

Expresamos el costo operacional de los buses (CO) como la suma de un costo fijo (F) y un costo variable que depende del número de kilómetros recorridos por los buses (V^{bus}):

$$CO = F + KV^{bus} \quad (15)$$

donde K determina el costo por kilómetro (\$/km).

Las variables de decisión de la agencia de transporte público son la tarifa por kilómetro (p^{bus}) y el número de kilómetros vehiculares de los buses (V^{bus}). Como mencionamos anteriormente, la capacidad de los buses (n^{bus}) y los kilómetros de ruta (D^{bus}) se toman como constantes. La restricción presupuestal de la agencia es:

$$Sub + p^{bus}M^{bus} + \tau^{car}V^{car} = CO \quad (16)$$

Esta restricción muestra que los costos operacionales son cubiertos por los ingresos por cobro de tarifas ($p^{bus}M^{bus}$) y los cargos al uso de carros particulares ($\tau^{car}V^{car}$). Si estos ingresos no son suficientes, el faltante debe ser cubierto mediante subsidios (Sub).

Debido a que las características del transporte público dependen principalmente de los kilómetros vehiculares de los buses (V^{bus}), la forma como la agencia de transporte público ajusta esta oferta ante cambios en la demanda (M^{bus}) determina en gran medida el impacto en bienestar que causa un cambio en la tarifa. Es posible asumir que la agencia ajusta los kilómetros vehiculares de forma óptima, balanceando los efectos en tiempos de viaje, hacinamiento, externalidades y costos de operación. Sin embargo, esta suposición puede resultar demasiado exigente en términos de la capacidad de planeación requerida por parte de la agencia. En su lugar, asumimos, siguiendo a Parry y Small (2009), que la agencia responde a cambios porcentuales en la demanda con cambios porcentuales proporcionales en los kilómetros vehiculares de acuerdo a una elasticidad constante ε_V . Dados unos niveles de oferta y demanda iniciales V_{ini}^{bus} y M_{ini}^{bus} , podemos entonces expresar los kilómetros vehiculares de los buses como una función de los kilómetros de pasajeros:

$$V^{bus} = V_{ini}^{bus} \left(\frac{M^{bus}}{M_{ini}^{bus}} \right)^{\varepsilon_V} \quad (17)$$

Esta aproximación a cómo la agencia ajusta la oferta de kilómetros vehiculares otorga bastante flexibilidad y nos permite evaluar el impacto de diversos niveles de ajuste en el nivel óptimo de subsidios. Por ejemplo, si el sistema de transporte público presenta restricciones

de capacidad, podemos tomar $\varepsilon_V = 0$, de forma que la oferta de kilómetros permanece constante. Dadas algunas aproximaciones, Parry y Small (2009) muestran que un ajuste óptimo de los kilómetros vehiculares por parte de la agencia implica $\varepsilon_V = 0,67$, por lo que este valor se puede tomar como punto de referencia.

La única variable de decisión directa de la agencia de transporte público es entonces la tarifa. La agencia determina la tarifa con el objetivo de maximizar el bienestar general, que en este caso está dado por la utilidad del agente representativo (1). Las condiciones de primer orden del proceso de maximización del agente representativo (3, 4 y 5) determinan los niveles de consumo del bien general (X) y de demanda de buses y carros particulares (M^{bus} y M^{car}) como función de la tarifa. Adicionalmente, la restricción de presupuesto (16) determina el nivel de subsidio necesario para el transporte público.

2.5. Subsidio óptimo

El Apéndice B muestra que el nivel de subsidio óptimo, definido como la diferencia entre el costo promedio de operación por kilómetro de pasajero (CO/M^{bus}) y la tarifa óptima por kilómetro (p^{bus}) se puede expresar como:

$$\begin{aligned} \frac{CO}{M^{bus}} - p^{bus} = & \underbrace{\left(\frac{CO}{M^{bus}} - CM_{oferta} \right)}_{\text{Diferencia costo promedio y marginal}} + \underbrace{(BM_{espera_frec})}_{\text{Efecto Mohring}} - \underbrace{(CM_{espera_hacin} + CM_{viaje_hacin})}_{\text{Costo hacinamiento}} \\ & + \underbrace{(CM_{ext}^{car} m_{bus}^{car} - CM_{ext}^{bus})}_{\text{Externalidades netas}} \quad (18) \end{aligned}$$

Cada uno de los cuatro términos en paréntesis en el lado derecho de la expresión anterior refleja un argumento que soporta o demerita la existencia de subsidios al transporte público.

El primer término en paréntesis determina la diferencia entre el costo operacional promedio (CO/M^{bus}) y el costo marginal operacional que causa un incremento en la demanda de transporte público debido al ajuste en la oferta de kilómetros de bus (CM_{oferta}). Este costo marginal se calcula como:

$$CM_{oferta} = K \frac{\varepsilon_V}{o^{bus}} \quad (19)$$

K denota el costo por kilómetro vehicular, mientras ε_V/o^{bus} determina el número de kilómetros vehiculares causados por un kilómetro adicional de pasajero.

Los costos fijos determinan en gran medida la magnitud de este primer término, debido a que estos costos afectan el costo operacional total (CO) pero no el costo marginal (CM_{oferta}). Adicionalmente, la magnitud de este término aumenta en la medida en que la oferta de

kilómetros se mantiene constante frente a cambios en la demanda (ε_V bajo), pues en este caso un aumento en la demanda de pasajeros no genera un aumento de los costos operacionales.

El segundo término en paréntesis refleja la magnitud del efecto Mohring. BM_{espera_frec} determina el beneficio marginal por reducción en los tiempos de espera debido al aumento de la frecuencia. El tercer término en paréntesis identifica los costos por hacinamiento. CM_{espera_hacin} refleja el costo marginal por aumento en los tiempos de espera debido a un mayor hacinamiento, lo que incrementa la probabilidad de que los pasajeros tengan que esperar más de un bus para poder abordar. CM_{viaje_hacin} representa el costo marginal por mayor incomodidad durante el viaje debido también a un mayor hacinamiento. Estos efectos marginales se calculan como:

$$BM_{espera_frec} = \rho^{W^{bus}} \varepsilon_V w^{bus} \eta_{wf} \quad (20)$$

$$CM_{espera_amon} = \rho^{W^{bus}} (1 - \varepsilon_V) w^{bus} \eta_{wl} \quad (21)$$

$$CM_{viaje_amon} = \rho^{C^{bus}} (1 - \varepsilon_V) C^{bus} t^{bus} \eta_C \quad (22)$$

donde η_{wf} y η_{wl} denotan las elasticidades del tiempo de espera con respecto a la frecuencia y al factor de carga respectivamente, y η_C denota la elasticidad del nivel de hacinamiento con respecto al factor de carga. Adicionalmente, $\rho^{C^{bus}}$ determina el aumento en el valor del tiempo de viaje en bus que causa un incremento en el nivel de hacinamiento (ej: 1 pasajero/m² adicional aumenta el valor del tiempo de viaje en \$50/minuto).

La Ecuación 18 muestra que el beneficio por el efecto Mohring puede ser contrarrestado por los efectos negativos de mayores niveles de hacinamiento si los aumentos en demanda no son acompañados por un aumento proporcional en la oferta ($\varepsilon_V < 1$). La magnitud de estos efectos negativos es probablemente mayor en sistemas que cuentan ya con altos niveles de hacinamiento, como es el caso en varias ciudades latinoamericanas. Adicionalmente, estos efectos negativos resultan más relevantes en la medida en que la agencia de transporte público responde a aumentos en la demanda con incrementos bajos en la oferta de kilómetros ($1 - \varepsilon_V$ alto), debido a que en este caso reducciones en la tarifa llevan a incrementos significativos en el factor de carga.

Finalmente, el cuarto término en paréntesis muestra el efecto neto en externalidades. CM_{ext}^{car} representa el costo marginal por externalidades (congestión y polución) causado por un aumento en los kilómetros de usuarios de carros particulares, mientras CM_{ext}^{bus} representa este mismo costo para buses. Estos costos marginales se calculan como:

$$CM_{ext}^{car} = \left(\rho^{Z^{pol}} z_{car}^{pol} + \rho^{T^{car}} M^{car} \frac{dt^{car}}{dV} + \rho^{T^{bus}} M^{bus} \frac{dt^{bus}}{dV} - \tau^{car} \right) \frac{1}{o^{car}} \quad (23)$$

$$CM_{ext}^{bus} = \left(\rho^{Z^{pol}} z_{bus}^{pol} + \rho^{T^{car}} M^{car} \alpha \frac{dt^{car}}{dV} + \rho^{T^{bus}} M^{bus} \alpha \frac{dt^{bus}}{dV} \right) \frac{\varepsilon_V}{o^{bus}} + \rho^{T^{bus}} (1 - \varepsilon_V) o^{bus} \theta \quad (24)$$

donde $\rho^{Z^{pol}}$ representa la valoración monetaria de cambios en el nivel de polución (\$/gr), y dt^{bus}/dV y dt^{car}/dV determinan cómo aumentos en el tráfico vehicular total incrementan los tiempos de viaje por kilómetro para buses y carros.

El primer término del lado derecho de la Ecuación 23 refleja el costo marginal por polución para carros. Los siguientes dos términos determinan el costo externo por congestión, que incluye mayores tiempos de viaje para usuarios de carros y buses. Si existen cargos al uso de carros particulares (τ^{car}), estos tienden a disminuir la magnitud del costo marginal por externalidades para carros, ya que solo se debe considerar la porción de este costo que no es internalizada mediante cargos. Por ejemplo, si existe un cobro por congestión a los carros particulares de magnitud óptima (es decir, igual al costo marginal por congestión), solo se debe tener en cuenta las externalidades por polución.

La Ecuación 24 es muy similar a la Ecuación 23, pero incluye un termino adicional que refleja cómo aumentos en la ocupación de los buses conllevan incrementos en los tiempos de parada.

Finalmente, el parámetro m_{bus}^{car} en la Ecuación 18 representa la fracción del aumento en la demanda de bus debido a reducciones en la tarifa que proviene de usuarios del carro particular:

$$m_{bus}^{car} = - \frac{\partial M^{car} / \partial p^{bus}}{\partial M^{bus} / \partial p^{bus}} \quad (25)$$

Por ejemplo, si reducciones en la tarifa de bus no logran atraer usuarios del carro, $m_{bus}^{car} = 0$. En el otro extremo, si al reducir la tarifa de transporte público todos los usuarios adicionales son anteriores usuarios del carro, $m_{bus}^{car} = 1$.

El efecto neto en externalidades considera entonces que tanto los buses como los carros particulares afectan los niveles de congestión y polución, y que solo una fracción del aumento de usuarios de transporte público proviene de carros particulares. La reducción de externalidades como argumento para justificar subsidios al transporte público gana fuerza en la medida en que cambios en la tarifa de transporte público logran atraer usuarios del carro, y en la medida en que las externalidades causadas por los buses resulten menores a las causadas por los carros al ser evaluadas por pasajero transportado.

La Ecuación 18 refleja entonces los argumentos principales que respaldan o demeritan subsidios al transporte público desde una perspectiva de eficiencia: costos fijos, efecto Mohring, hacinamiento y externalidades. Esta ecuación muestra cómo la magnitud de cada uno de estos argumentos depende de diversos parámetros clave. La siguiente sección presenta una revisión de literatura sobre estudios que cuantifican algunos de estos parámetros para

ciudades en América Latina.

3. Revisión de literatura

3.1. Estudios sobre subsidios óptimos al transporte público

Parry y Small (2009) estimaron los niveles de subsidio óptimos para Londres, Washington D.C. y Los Angeles (año base 2002), basados en los argumentos de eficiencia detallados en la sección anterior. La Tabla 1 presenta los principales resultados de este estudio. Aunque el nivel de subsidio óptimo varía de acuerdo al tipo de sistema (tren o bus) y al periodo del día (pico o valle), en la mayoría de los casos es superior al 50 % y en más de la mitad es superior al 90 %. La proporción de este subsidio que obedece a cada uno de los componentes de la Ecuación 18 varía significativamente en cada caso. El efecto Mohring es más importante en las horas valle (en las que la oferta de servicios es menor), mientras que las externalidades son más relevantes en las horas pico (congestión es la mayor externalidad). Adicionalmente, el efecto Mohring es mayor para los sistemas de buses debido a que estos sistemas tienen usualmente menores frecuencias, y el componente de externalidades es mayor para los sistemas de trenes debido a que estos sistemas cuentan con carriles exclusivos (por lo que no aportan a la congestión).⁸

Para América Latina, Parry y Timilsina (2010) y Basso y Silva (2014) estimaron los niveles de subsidio óptimos para Ciudad de México y Santiago de Chile respectivamente, así como los cargos apropiados al uso del carro particular. Parry y Timilsina (2010) encontraron que la tarifa óptima para el bus equivalía a cerca del 50 % del costo marginal de operación y era muy cercana a la tarifa vigente en ese momento, mientras que la tarifa óptima del tren era negativa, lo que implica un nivel de subsidio óptimo mayor a 100 %.⁹ Basso y Silva (2014) estimaron un nivel de subsidio óptimo para el sistema de buses de 55 %, cercano al 35 % vigente en ese momento. Sin embargo, si otras políticas de transporte urbano, como cobros por congestión al uso de carros particulares o la implementación de carriles exclusivos para los buses, fuesen implementadas, no habría razón para otorgar subsidios a los buses.¹⁰

⁸El cuarto componente de la Tabla 1 (interacción transporte público) corresponde a interacciones entre sistemas de transporte público (bus-tren) y entre los periodos pico y valle de un mismo sistema. Este componente surge debido a que el estudio calcula la tarifa óptima para cada caso por separado (ej: bus en hora pico), por lo que la tarifa del otro sistema o del otro periodo no es óptima. Este componente no está incluido en la Ecuación 18 debido a que este modelo considera un solo sistema (bus) y un solo periodo de tiempo.

⁹Este estudio no tuvo en cuenta el efecto Mohring, por lo que el nivel de subsidio responde principalmente al manejo de externalidades. Los autores asumieron que el número de kilómetros vehiculares de buses y trenes cambian de forma proporcional con la demanda ($\varepsilon_V = 1$ en la Ecuación 18).

¹⁰Basso y Silva (2014) consideran el costo marginal de los recursos públicos. Este costo disminuye el nivel de subsidio óptimo. Por ejemplo, si este costo aumenta de 1.15 a 1.4, el subsidio óptimo disminuye de 55 % a

Tabla 1: Subsidio óptimo como porcentaje de costos operacionales (Parry y Small, 2009).

	Washington D.C.				Los Angeles			
	Tren		Bus		Tren		Bus	
	Pico	Valle	Pico	Valle	Pico	Valle	Pico	Valle
Subsidio óptimo (%)	>90	88	46	>90	>90	89	74	>90
Porcentaje debido a:								
Dif. costo promedio y marginal	43	54	37	39	38	49	42	35
Efecto Mohring neto	10	41	46	59	5	34	22	55
Externalidades netas	41	8	19	-1	51	13	37	4
Interacción transporte público	7	-3	-2	3	6	3	0	6
	Londres							
	Tren		Bus					
	Pico	Valle	Pico	Valle				
Subsidio óptimo (%)	>90	78	>90	>90				
Porcentaje debido a:								
Dif. costo promedio y marginal	31	56	35	33				
Efecto Mohring neto	0	21	15	61				
Externalidades netas	56	27	42	-11				
Interacción transporte público	13	-4	9	17				

Ambos estudios encuentran que los beneficios de imponer cargos apropiados al uso del carro son significativamente superiores a los beneficios que se pueden lograr a través de subsidios al transporte público, por lo que estos subsidios son una política de segundo mejor frente a cobros al uso del carro.

Los resultados de Parry y Timilsina (2010) y Basso y Silva (2014), en comparación con los resultados obtenidos por Parry y Small (2009), sugieren que los niveles de subsidios que pueden ser justificados mediante argumentos de eficiencia en ciudades latinoamericanas son cercanos a los que se justifican en ciudades de países desarrollados.

3.2. Elasticidades al precio propio y elasticidades cruzadas

La Sección 2 muestra que el nivel de subsidio óptimo depende tanto de la elasticidad de la demanda de transporte público a su propio precio, como de la elasticidad cruzada del uso del carro al precio del transporte público. En especial, una demanda de transporte público

15 %. Sin embargo, Basso y Silva (2014) no consideran el efecto positivo que los subsidios transporte público pueden tener en el mercado laboral, lo que puede contrarrestar este costo (ver Sección 3.4). Parry y Small (2009) consideran un costo marginal de los recursos públicos de 1.15 y concluyen que aunque este costo reduce el tamaño de los subsidios, el nivel de subsidio óptimo sigue siendo significativo.

bastante elástica a su propio precio aumenta el tamaño del efecto Mohring, mientras que una elasticidad cruzada alta del uso del carro aumenta la magnitud de los beneficios por externalidades.

Numerosos estudios han estimado la elasticidad del transporte público a su propio precio. A nivel internacional, la elasticidad de la demanda de buses suele estar alrededor de -0.4 en el corto plazo (1-2 años) y -1 en el largo plazo, mientras la elasticidad para sistemas de trenes varía entre -0.3 en el corto plazo a -0.6 en el largo plazo (Horcher y Tirachini, 2021). Adicionalmente, esta elasticidad tiende a ser mayor en horas valle (cuando la proporción de viajes con motivos obligatorios es menor) y para usuarios de mayores ingresos.

Para América Latina, de Grange, Gonzalez, Muñoz, y Troncoso (2013), Davis (2020) y Guzman, Beltran, Bonilla, y Cardona (2021) estimaron esta elasticidad con base en información detallada proveniente de sistemas de recaudo centralizado y tarjetas inteligentes (Tabla 2). de Grange y cols. (2013) estimaron las elasticidades para metro y buses en Santiago de Chile. Sus estimaciones varían entre -0.20 (hora valle del metro) y -0.58 (hora pico del metro). Este estudio también estimó las elasticidades cruzadas entre los dos modos de transporte público (0.13 y 0.23 entre bus y metro en horas valle y pico respectivamente) y entre diferentes periodos del día (0.16 entre los periodos pico y valle del metro). Este tipo de elasticidades es importante cuando se consideran simultáneamente tarifas óptimas para varios modos de transporte y periodos de tiempo.

Tabla 2: Elasticidad de la demanda de transporte público a la tarifa para ciudades latinoamericanas.

Estudio	Ciudad	Sistema	Elasticidad
de Grange y cols. (2013)	Santiago	Bus - Hora pico	-0.28
		Bus - Hora valle	-0.35
		Metro - Hora pico	-0.58
		Metro - Hora valle	-0.20
Davis (2020)	Ciudad de Mexico	Metro	-0.25
	Guadalajara	Tren ligero	-0.32
	Monterrey	Metro	-0.23
Guzman y cols. (2021)	Bogotá	BRT	-0.24

Guzman y cols. (2021) encontraron una elasticidad promedio para el sistema BRT de Bogotá (TransMilenio) de -0.24. Sin embargo, esta elasticidad es mucho menor (cercana a cero) para los usuarios que viven en la periferia de la ciudad cuyo nivel de ingreso es bajo. Finalmente, (Davis, 2020) utilizó cambios significativos en las tarifas de metro en Ciudad de México, Guadalajara y Monterrey para estimar sus elasticidades. Los resultados muestran elasticidades entre -0.23 y -0.32 para estas tres ciudades.

Los resultados de estos estudios sugieren que la elasticidad del transporte público a su propio precio tiende a ser menor en América Latina en comparación con el promedio internacional, que se basa mayoritariamente en estudios de países desarrollados. Esta diferencia se debe probablemente al menor nivel de ingreso y motorización de los usuarios de transporte público en América Latina, lo que convierte a la mayoría en usuarios cautivos del transporte público.

La elasticidad cruzada del uso del carro al precio del transporte público ha sido objeto de muchos menos estudios empíricos. Debido a que esta elasticidad, en conjunto con la elasticidad de la demanda de transporte público a su propio precio, define la fracción del aumento en la demanda de transporte público debido a reducciones tarifarias que proviene del carro (ver Ecuación 25),¹¹ varios estudios realizan suposiciones directamente sobre este parámetro. Por ejemplo, Parry y Small (2009) asumieron que en las ciudades estadounidenses (Los Angeles y Washington D.C.) entre el 60 % y el 85 % del incremento en el uso del transporte público proviene de usuarios del carro. En ciudades con menores niveles de motorización y uso del carro, como las ciudades latinoamericanas, es de esperar que esta fracción sea menor. Por ejemplo, Parry y Small (2009) asumieron que esta fracción se encontraba entre 40 % y 50 % para Londres, y Parry y Timilsina (2010) asumieron 35 % para Ciudad de México.

Teóricamente, existe una relación directa entre la elasticidad del uso del carro al precio del transporte público y la elasticidad del uso del transporte público al precio del carro,¹² por lo que estimaciones de esta segunda elasticidad son muy útiles. Por ejemplo, Börjesson, Fung, y Proost (2017) utilizaron datos sobre cómo cambió el uso del transporte público debido a la introducción de cobros por congestión en Estocolmo para estimar la elasticidad del uso del carro al precio de transporte público. Sus resultados muestran que esta elasticidad era 0.31 y 0.15 para horas pico y valle respectivamente. Estas elasticidades implican que la fracción de viajes que provienen del carro es mucho menor a lo asumido por Parry y Small (2009), lo que es coherente con el hecho de que el transporte público representa la mayoría (~80 %) de los viajes en el corredor analizado para Estocolmo. De forma similar, Basso y Silva (2014) encontraron que los cargos por congestión al uso del carro aumentan el uso del transporte

¹¹La Ecuación 25 muestra que esta fracción es igual a la razón entre las derivadas del uso del carro y el transporte público al precio del transporte público. Teniendo en cuenta la relación entre estas derivadas y las elasticidades, esta fracción se puede expresar como

$$m_{bus}^{car} = -\frac{\varepsilon_{bus}^{car} M^{bus}}{\varepsilon_{bus}^{bus} M^{car}}$$

donde ε_{bus}^{car} denota la elasticidad del carro al precio del transporte público, y ε_{bus}^{bus} la elasticidad propia del transporte público.

¹²Específicamente, la simetría de la matriz de Slutsky dicta que la derivada del uso del carro al precio del transporte público es igual a la derivada del uso del transporte público al precio del carro.

público mucho más en Londres que en Santiago.

La fracción del aumento en la demanda de transporte público debido a reducciones tarifarias que proviene del carro es un parámetro clave para determinar el nivel de subsidio óptimo al transporte público, así como para evaluar el efecto de cargos adicionales al uso del carro. Por lo tanto, estudios empíricos que estimen este parámetro para ciudades latinoamericanas permitirían mejorar sustancialmente la calidad de las evaluaciones de política tarifaria de transporte en la región.

3.3. Efecto marginal en congestión

La congestión es usualmente la principal externalidad del flujo vehicular en ciudades grandes, especialmente durante horas pico. El tamaño de esta externalidad depende en gran medida de la fuerza con que aumentos en el flujo vehicular disminuyen la velocidad promedio de viaje y aumentan los tiempos de recorrido.¹³ Los estudios de tarificación óptima para el transporte generalmente aproximan la relación entre el flujo vehicular y el tiempo promedio de recorrido mediante el uso de una forma funcional específica. La forma funcional más común se conoce como función BPR (*Bureau of Public Roads*). De acuerdo a esta forma funcional, el tiempo de recorrido (t) depende del flujo vehicular (V) de la siguiente forma:

$$t = t_f \left(1 + \alpha \left(\frac{V}{C} \right)^\beta \right) \quad (26)$$

donde t_f representa el tiempo de recorrido bajo condiciones de flujo libre, C mide la capacidad de la red vehicular y α y β son parámetros con valores comúnmente cercanos a $\alpha = 0,15$ y $\beta = 4$. Bajo esta forma funcional, el efecto marginal de aumentos en el flujo vehicular en el tiempo de recorrido es proporcional a la demora promedio $((t - t_f)/V)$. Específicamente, el efecto marginal es igual a la demora promedio multiplicada por β . Esta relación permite estimar fácilmente el efecto marginal del flujo vehicular con base en observaciones de los tiempos de recorrido actuales. Parry y Small (2009), Parry y Timilsina (2010) y Basso y Silva (2014) aproximan los efectos de congestión con base en la función BPR.

Algunos estudios asumen una relación lineal entre el flujo vehicular y la velocidad promedio. Esta suposición permite determinar completamente esta relación utilizando solamente información sobre los niveles de flujo y velocidad actuales y sobre la velocidad promedio en condiciones de flujo libre. Esta es la aproximación utilizada por Lopez-Ghio, Bocarejo, y Blanco (2018) en su estudio sobre cargos por congestión para Santiago, Ciudad de México y Bogotá.

¹³Este efecto corresponde a las derivadas dt^{bus}/dV y dt^{car}/dV en las Ecuaciones 23 y 24.

Aunque las anteriores aproximaciones permiten estimar el efecto marginal del flujo vehicular en la congestión con base en pocos datos, la mayor disponibilidad de información detallada sobre flujos vehiculares y velocidades en los últimos años ha permitido realizar estudios empíricos específicos para diferentes ciudades. Mangrum y Molnar (2018) utilizaron información detallada de velocidad para viajes en taxi e imágenes aéreas para estimar el efecto marginal de un taxi en la velocidad promedio en Nueva York. Su estrategia de identificación se basa en el aumento del número de taxis causado por un cambio en la regulación. Yang, Purevjav, y Li (2020) utilizaron información de más de 1,500 estaciones de monitoreo de tráfico en Beijing para estimar el costo marginal de la congestión. Su estrategia de identificación se basa en el cambio exógeno en el flujo vehicular causado por la política de restricción vehicular según el número de placa. Kreindler (2020) utilizó datos de GPS de viajes individuales y tiempos de viaje de Google Maps para estimar el efecto marginal en congestión en Bangalore. Su estrategia de identificación se basa en los cambios naturales en la demanda de viajes que ocurren durante el día en la ciudad. Akbar y Duranton (2017) y Gomez-Gelvez (2021) utilizan esta misma estrategia empírica, en conjunto con datos de la encuesta de movilidad de hogares y Google Maps, para estimar la relación entre tráfico vehicular y tiempos de recorrido en Bogotá.

El gran aumento en la disponibilidad de información sobre flujos vehiculares y velocidades en los últimos años sugiere que no se justifica continuar utilizando aproximaciones como la función BPR para estimar las externalidades por congestión. Dada la importancia de estas externalidades para los estudios de tarificación óptima del transporte, estimaciones empíricas como las anteriores para ciudades latinoamericanas permitirían fortalecer la base empírica de estos estudios en la región.

3.4. El costo marginal de los recursos públicos

El modelo presentado en la Sección 2 asume que los recursos públicos necesarios para cubrir un subsidio al transporte público son recolectados a través de cobros fijos a los ciudadanos (*lump-sum taxation*). Esto implica que la consecución de estos recursos no conlleva ningún costo en términos de eficiencia. Sin embargo, los gobiernos usualmente consiguen recursos públicos a través de impuestos a bienes o servicios específicos, lo que genera distorsiones en estos mercados y conlleva pérdidas de eficiencia. El mercado laboral es generalmente uno de los principales afectados a través de los impuestos al ingreso.

Este costo en mercados diferentes al de transporte debe ser tenido en cuenta en evaluaciones del nivel de subsidio óptimo de sistemas de transporte público, debido a que su efecto no es despreciable. Basso y Silva (2014) encuentran, por ejemplo, que si el costo marginal

de los recursos públicos aumenta de 1.15 a 1.4,¹⁴ el nivel de subsidio óptimo para buses en Santiago disminuye de 55 % a 15 %.

La literatura sobre impuestos ambientales muestra, sin embargo, que puede existir un efecto adicional importante que contrarresta el costo marginal de los recursos públicos. Este efecto se da en la medida en que el impuesto o subsidio bajo análisis genere cambios directos en la demanda u oferta del mercado distorsionado para conseguir recursos públicos. Por ejemplo, si el principal mercado distorsionado es el mercado laboral y el subsidio al transporte público genera un aumento en la participación en el mercado laboral (mayor oferta laboral), no es claro *a priori* si un aumento en el subsidio aumentaría o disminuiría la distorsión en el mercado laboral causada por impuestos al ingreso. Parry y Bento (2001) encuentran, por ejemplo, que este efecto contrarresta completamente el costo marginal de recursos públicos en el caso de cobros por congestión a viajes con motivos laborales, por lo que el cobro por congestión óptimo es igual al costo marginal externo por congestión (impuesto Pigouviano).¹⁵

Tener en cuenta el costo marginal de los recursos públicos al evaluar subsidios al transporte público, sin considerar el potencial efecto directo generado por la interacción entre el mercado de transporte y el mercado distorsionado para conseguir recursos públicos, puede entonces producir una visión artificialmente negativa de la conveniencia de subsidios. El parámetro relevante en este caso es el efecto cruzado que los subsidios al transporte público pueden generar en el mercado laboral. Subsidios en las horas pico pueden, por ejemplo, aumentar la oferta laboral en mayor medida que subsidios en horas valle, debido a que las horas pico contienen una mayor proporción de viajes con motivos laborales.

Para América Latina, es importante tener en cuenta el potencial impacto en el mercado laboral informal, ya que este constituye una porción muy significativa del mercado laboral total (Tirachini y Proost, 2021). Rodríguez, Gallego, Martínez, Montoya, y Peralta-Quiros (2016) evaluaron el efecto en el mercado laboral que generaron subsidios focalizados a los usuarios de transporte público con menores ingresos en Bogotá. Ellos encontraron que aunque el subsidio no tuvo un efecto significativo en la participación en el mercado laboral, sí aumentó el nivel de ingreso de los trabajadores informales. Zárate (2020) encontró que la construcción de nuevas líneas de metro en Ciudad de México disminuyó los niveles de informalidad en cuatro puntos porcentuales en las áreas afectadas.

En su estudio sobre el costo marginal de recursos públicos en los impuestos y subsidios

¹⁴Un costo marginal de recursos públicos igual a 1.15 implica que por cada peso adicional recolectado hay un costo de \$0.15 en términos de eficiencia debido a distorsiones en los mercados afectados.

¹⁵En el caso de cobros por congestión, el costo marginal de los recursos públicos determina un beneficio adicional del cobro, ya que este genera una fuente de recursos públicos que se puede utilizar para disminuir los impuestos que distorsionan el mercado laboral. Por otra parte, cobros por congestión a viajes con motivos laborales disminuyen la oferta laboral.

al transporte en Santiago, Tirachini y Proost (2021) asumieron que cualquier incremento en el costo del transporte (público o privado) disminuye la oferta laboral. Sin embargo, esta relación no es automática, debido a que los impuestos o subsidios al transporte afectan tanto viajes con motivos laborales como no laborales. Por ejemplo, West y Williams (2007) estimaron el efecto cruzado entre impuestos a la gasolina y la oferta laboral en Estados Unidos. Ellos encontraron que la gasolina es un sustituto de la oferta laboral, por lo que la interacción entre los mercados de transporte y laboral soporta un impuesto a la gasolina *mayor* al que dicta el costo asociado a sus externalidades. Estudios empíricos como este son necesarios para conocer mejor la relación entre el transporte y el mercado laboral en América Latina.

4. Aplicación a Bogotá, Colombia

Esta sección aplica el modelo económico de la Sección 2 al transporte público en Bogotá, Colombia. Tomamos como año de referencia el 2019, antes del comienzo de la pandemia generada por el COVID-19. Bogotá cuenta con dos sistemas de transporte público principales: un sistema BRT (*Bus Rapid Transit*) conocido como TransMilenio y un sistema de buses convencionales. Estos dos sistemas conforman el Sistema Integrado de Transporte Público (SITP). TransMilenio corresponde a la operación troncal del SITP, mientras los buses convencionales conforman la operación zonal (organizada en 13 zonas de operación).

TransMilenio opera en carriles segregados, con buses de alta capacidad y con recolección tarifaria en las estaciones mediante tarjetas inteligentes. TransMilenio cuenta también con un importante sistema de rutas alimentadoras, que operan en tráfico mixto y conectan las principales estaciones del sistema troncal con zonas residenciales de alta densidad. El uso de las rutas alimentadoras es gratuito. Los usuarios solo deben pagar al transbordar a servicios troncales. Debido a que las demandas de las rutas troncales y alimentadoras se encuentran altamente relacionadas (la gran mayoría de los usuarios de las rutas alimentadoras transbordan hacia o desde servicios troncales), incluimos el sistema alimentador como parte de TransMilenio. Adicionalmente, incluimos como parte de TransMilenio un número bajo de rutas duales que transitan tanto en carriles segregados como en tráfico mixto.¹⁶

El componente zonal del SITP (buses convencionales) opera en tráfico mixto con recolección tarifaria dentro del bus mediante tarjetas inteligentes. Aunque en 2019 buena parte del transporte público en buses convencionales en Bogotá operaba bajo el antiguo esquema empresarial afiliador (conocido actualmente como SITP Provisional), consideramos dentro

¹⁶La demanda de estas rutas duales conforma cerca del 2% de la demanda del SITP.

del componente zonal solo las rutas que operan bajo el nuevo sistema de concesiones.¹⁷ Esta decisión responde a la falta de información sobre el SITP Provisional y a que este sistema dejará de existir en los próximos años como resultado de la consolidación del SITP Zonal.

Para ambos sistemas (TransMilenio y SITP Zonal), consideramos dos periodos de tiempo. El periodo pico de un día hábil abarca seis horas, de 5am a 8am y de 4pm a 7pm. El periodo valle abarca ocho horas, de 8am a 4pm. Definimos estos periodos con base en el perfil horario de la demanda del SITP en un día hábil promedio. Para cada sistema y periodo, aplicamos el modelo descrito en la Sección 2 con el objetivo de estimar la tarifa y el nivel de subsidio óptimo. Las siguientes secciones presentan los parámetros base y los resultados del escenario principal, así como un análisis de sensibilidad de los resultados a varios parámetros clave.

4.1. Parámetros base

4.1.1. Demanda y oferta

La Tabla 3 presenta la información principal sobre demanda y oferta de los sistemas. En el lado de la demanda, estimamos el número de pasajeros por hora y la longitud promedio del viaje con base en los datos de la encuesta de movilidad 2019 de Bogotá (Secretaría Distrital de Movilidad, Bogotá D.C., 2019).¹⁸ El número de pasajeros por hora es aproximadamente 50 % mayor y la longitud del viaje 15 % mayor en el periodo pico para ambos sistemas. Como resultado, el número de pasajeros-kilómetro por hora es 75 % y 70 % mayor en el periodo pico para TransMilenio y el SITP Zonal respectivamente.

En el lado de la oferta, TransMilenio S.A. (agencia pública encargada de la planeación de ambos sistemas) nos facilitó el número de kilómetros recorridos en el año por las diferentes tipologías de buses en cada sistema. Adicionalmente, utilizamos información del horario de los despachos programados para las rutas de cada sistema.¹⁹ Asignamos el total anual de kilómetros a los periodos pico, valle y otros (días hábiles antes de 6am y después de 7pm,

¹⁷Estas concesiones son manejadas también por la agencia pública TransMilenio S.A.

¹⁸Las validaciones de tarjetas inteligentes al ingreso a la estación o bus permiten también estimar el número de pasajeros en ambos sistemas. Estos datos, utilizados por TransMilenio S.A. para estadísticas oficiales, son más precisos en cuanto permiten promediar mediciones exactas del número de validaciones durante todos los días hábiles. Sin embargo, la encuesta de movilidad nos permite incluir información sobre viajes que no hacen validación. Para TransMilenio, los viajes que utilizan solo rutas alimentadoras, sin transbordar a rutas troncales, no necesitan hacer validación ya que son gratis. Para ambos sistemas, es probable que la encuesta logre capturar parte de la evasión tarifaria. Según estimaciones de TransMilenio S.A., 15 % de los viajes en el componente troncal evadían la tarifa en 2019. El número de viajes en un día hábil promedio en TransMilenio según la encuesta de movilidad (sin tener en cuenta los viajes que solo utilizan rutas alimentadoras) es 5 % mayor al promedio de validaciones. Este resultado sugiere que la encuesta logra capturar solo parte de la evasión tarifaria. Para el SITP Zonal, el número de viajes según la encuesta de movilidad es 14 % mayor al número de validaciones (no existen estimaciones oficiales obre evasión en el SITP Zonal).

¹⁹Está información se encuentra disponible públicamente en el portal de datos abiertos www.datos.gov.co.

Tabla 3: Parámetros base - Demanda y oferta.

	TransMilenio - BRT		SITP Zonal - Bus	
	Pico	Valle	Pico	Valle
Pasajeros/hora	223,056	146,639	134,060	90,340
Longitud promedio del viaje (km)	13.9	12.0	10.8	9.4
Pas-km/hora	3,096,017	1,765,535	1,442,487	852,811
Bus-km/hora	38,759	33,034	53,822	46,823
Intervalo promedio (min)	4.7	4.7	11.1	12.4
Ocupación promedio (pas/bus)	79.9	53.4	26.8	18.2
Capacidad promedio (pas/bus)	146.4	146.4	54.7	54.7
Factor de carga (ocup./capac.)	0.55	0.37	0.49	0.33
Pas/m ² a capacidad	6.5	6.5	6.5	6.5
Coef. Var. ocupación	0.5	0.5	0.5	0.5
Hacinamiento promedio (pas/m ²)	4.43	2.97	3.98	2.71
	Vehículos particulares			
	Pico	Valle		
Pasajeros/hora	122,844	129,685		
Longitud promedio del viaje (km)	9.46	8.78		
Pas-km/hora	1,162,103	1,138,632		
Ocupación promedio (pas/carro)	1.54	1.54		

Notas: TransMilenio incluye rutas alimentadoras y duales. Periodo pico: 5-8am y 4-7pm. Periodo valle: 8am-4pm.

fin de semana y festivos) de acuerdo con el porcentaje de despachos que se realizan en cada periodo. Los resultados muestran que el número de bus-kilómetros por hora se incrementa en 17 % y 15 % en el periodo pico para TransMilenio y el SITP Zonal respectivamente.

Calculamos el intervalo promedio para cada sistema en cada periodo de tiempo con base en la programación de los despachos de las rutas de cada sistema. Debido a que no contamos con información de demanda específica para cada ruta, este promedio no es ponderado de acuerdo a la demanda de cada ruta. Aunque el número de bus-kilómetros por hora en TransMilenio es mayor en el periodo pico, el intervalo promedio es similar en ambos periodos. Esto se debe a que el aumento de la oferta de kilómetros en el periodo pico responde principalmente a la introducción de rutas expresas que solo operan en este periodo, no al aumento de frecuencias.

La ocupación promedio de los buses se calcula como la división del número de pasajeros-kilómetro entre el número de kilómetros de bus. Para ambos sistemas, la ocupación promedio es casi 50 % mayor en el periodo pico. La capacidad promedio del bus pondera la capacidad de los diferentes tipos de bus de acuerdo al número de kilómetros recorrido por cada tipo.²⁰ La relación entre la ocupación y la capacidad promedio determina el factor de carga en cada periodo.

Para determinar el hacinamiento promedio experimentado por los pasajeros (pas/m^2), debemos primero definir el hacinamiento que se obtiene cuando los buses van a capacidad. Para ambos sistemas, definimos este valor en $6.5 \text{ pas}/\text{m}^2$. Adicionalmente, debemos tener en cuenta que la ocupación promedio experimentada por los pasajeros es superior a la ocupación promedio de los buses, ya que los buses con mayor ocupación tienen más pasajeros. Esta diferencia depende de la variabilidad en la ocupación de los buses. Debido a que no tenemos datos para estimar esta variabilidad, asumimos un valor (relativamente conservador) de 0.5 para el coeficiente de variabilidad (CV) de la ocupación. Calculamos entonces el hacinamiento promedio experimentado por los pasajeros como el producto del hacinamiento a capacidad, el factor de carga y $(1 + CV^2)$.

Finalmente, la Tabla 3 muestra también la demanda de viajes en vehículos particulares con base en los datos de la encuesta de movilidad. A diferencia de la demanda de transporte público, esta demanda no presenta diferencias importantes entre horas pico y valle. De hecho, el número de pasajeros por hora es un poco superior durante la hora valle.

²⁰TransMilenio utiliza buses biarticulados (250 pasajeros) y articulados (150 pasajeros) para las rutas troncales, y buses de 80 y 50 pasajeros para las rutas alimentadoras y duales. El SITP Zonal utiliza buses con capacidad para 80, 50, 40 y 19 pasajeros. La información disponible no nos permitió obtener capacidades promedio específicas para cada periodo.

4.1.2. Costos operacionales y tarifas

La Tabla 4 presenta información sobre costos operacionales y tarifas. El costo marginal por kilómetro de bus está basado en los contratos de Fase III del SITP (información facilitada por TransMilenio S.A.). Estos contratos incluyen un pago a las empresas operadoras privadas por kilómetro de bus, cuyo objetivo es cubrir costos variables como combustibles, salario de conductores y mantenimiento, y un pago mensual por bus, cuyo objetivo es cubrir la inversión en la compra del bus.²¹ Ambos pagos varían de acuerdo al tipo de bus.²² El costo marginal durante el periodo valle para ambos sistemas equivale al costo promedio por kilómetro de bus ponderado de acuerdo al número de kilómetros recorrido por cada tipo de bus. El costo marginal durante el periodo pico incluye adicionalmente el pago mensual por bus, multiplicado por el tamaño de la flota de cada tipología y dividido por el número promedio de kilómetros que recorre la flota mensualmente. De esta forma, el costo marginal en el periodo valle refleja un aumento en los costos variables, sin necesidad de aumentar el tamaño de la flota, mientras el costo marginal en el periodo pico incluye la necesidad de adquirir nueva flota. Esta diferencia hace que el costo marginal de aumentar la oferta de kilómetros en el periodo pico sea más del doble que en el periodo valle para ambos sistemas.²³

El costo promedio por kilómetro de bus añade al costo marginal costos fijos relacionados con el funcionamiento de TransMilenio S.A. como ente gestor, el sistema centralizado de recaudo y actividades a cargo del ente gestor como seguridad, atención al usuario, campañas de cultura ciudadana y aseo y mantenimiento de estaciones y paraderos. Asignamos estos costos a los dos sistemas y a los diferentes periodos de acuerdo al número de pasajeros y despachos, excepto por el costo de aseo y mantenimiento de estaciones, el cual es mucho mayor para el sistema troncal.²⁴ Estos costo fijos incrementan el costo por kilómetro de

²¹En 2019, los contratos de Fase I y II de TransMilenio se encontraban en proceso de renovación a la Fase IV. Los nuevos contratos separan las actividades de provisión y operación de flota en diferentes empresas privadas, con el objetivo de disminuir el riesgo financiero de las empresas operadoras. La Fase V del SITP, que busca consolidar el SITP Zonal en la ciudad y acabar con el SITP Provisional, también separará la provisión y operación de flota.

²²El pago por kilómetro no depende del periodo de operación, por lo que no consideramos potenciales diferencias entre periodos en los costos variables. Estas diferencias pueden surgir principalmente por diferencias entre periodos en el salario de conductores y por la necesidad de garantizar jornadas laborales completas a los conductores, lo que hace que el costo marginal de conductores sea menor en periodos valle.

²³Los estudios de costos operacionales de transporte público usualmente asignan el costo de compra de la flota al periodo pico. Sin embargo, es importante resaltar que este supuesto es válido solo en la medida en que la vida útil del bus esté determinada por un número de años. Si la vida útil del bus está determinada por un número de kilómetros, los costos de compra deberían ser asignados a los diferentes periodos en proporción al número de kilómetros recorridos, lo que igualaría el costo marginal entre periodos pico y valle. La vida útil de los buses del SITP está determinada por un número de años, por lo que asignamos el costo de compra al periodo pico.

²⁴Estimamos el costo del sistema de recaudo con base en el pago anual realizado por TransMilenio S.A. al operador de recaudo (operador único para los dos sistemas). Tomamos el valor anual de los demás costos del

Tabla 4: Parámetros base - Costos operacionales y tarifas.

	TransMilenio - BRT		SITP Zonal - Bus	
	Pico	Valle	Pico	Valle
Costo marginal por bus-km	\$12,686	\$5,788	\$7,770	\$3,240
Costo promedio por bus-km	\$15,118	\$7,728	\$8,515	\$3,854
Costo promedio por pas-km	\$189	\$145	\$318	\$212
Costo promedio por pasajero	\$2,627	\$1,741	\$3,419	\$1,998
Tarifa promedio por pasajero	\$2,001	\$1,874	\$1,930	\$1,943
Tarifa promedio por pas-km	\$144	\$156	\$179	\$206
Subsidio	23.8 %	-7.7 %	43.5 %	2.7 %

Notas: Valores monetarios en pesos colombianos de 2019. La tasa de cambio osciló en 2019 alrededor de \$1 USD = \$3,200 COL. TransMilenio incluye rutas alimentadoras y duales. Periodo pico: 5-8am y 4-7pm. Periodo valle: 8am-4pm.

TransMilenio en cerca de 20 % en el periodo pico y 33 % en el periodo valle, y el costo por kilómetro del SITP Zonal en 10 % y 20 % en los periodos pico y valle respectivamente.

El costo promedio por pasajero-kilómetro resulta de dividir el costo por kilómetro de bus entre la ocupación promedio de los buses en cada periodo. El costo promedio por pasajero-kilómetro es 40 % menor para TransMilenio en horas pico en comparación con el SITP Zonal, y 30 % menor en el periodo valle. Este resultado muestra que TransMilenio es un sistema más eficiente en términos de costos operacionales. El costo promedio por pasajero resulta de multiplicar el costo por pasajero-kilómetro por la distancia promedio del viaje, y corresponde a la tarifa promedio por pasajero que se debe cobrar a los usuarios para que los sistemas sean autosostenibles (sin subsidio).

La estructura tarifaria de TransMilenio y el SITP Zonal se basa en una tarifa plana por viaje, con descuentos para hogares de bajos ingresos (25 %), adultos mayores (10 %) y discapacitados (40 %). La tarifa base de TransMilenio en 2019 era de \$2,400, y la del SITP Zonal de \$2,200. Los transbordos de TransMilenio al SITP Zonal eran gratis, mientras los transbordos en el sentido contrario tenían un costo de \$200. Los viajes en la rutas alimentadoras de TransMilenio (sin utilizar el sistema troncal) son gratis. La Tabla 4 muestra la tarifa promedio por pasajero para cada sistema y periodo teniendo en cuenta esta estructura tarifaria. Estas estimaciones se basan en los datos de la encuesta de movilidad 2019 de Bogotá.

informe de ejecución presupuestal de TransMilenio S.A. para 2019 (TransMilenio S.A., 2019). Asignamos los costos de recaudo, atención al usuario, seguridad y cultura ciudadana a los diferentes sistemas y periodos (pico, valle y otros) de acuerdo al número de pasajeros atendidos en cada periodo, y los costos de funcionamiento del ente gestor de acuerdo al número de despachos. Los costos de aseo y mantenimiento de estaciones y paraderos corresponden en su gran mayoría a la operación troncal de TransMilenio. No incluimos como parte de los costos fijos actividades de construcción de infraestructura ejecutadas por TransMilenio S.A.

La comparación entre el costo promedio y la tarifa promedio por pasajero nos permite establecer el nivel de subsidio (el porcentaje del costo no cubierto por la tarifa) de cada sistema en cada periodo de tiempo. El periodo pico tiene un subsidio significativo en ambos sistemas, siendo mayor para el SITP Zonal. En el periodo valle, el SITP Zonal tiene un subsidio bajo y TransMilenio uno negativo ($\text{tarifa} > \text{costo}$).²⁵

4.1.3. Tiempos de viaje, costo generalizado y elasticidad de la demanda

La Tabla 5 revela el tiempo promedio en cada etapa para los viajes en transporte público. Determinamos estos tiempos promedio con base en los datos de la encuesta de movilidad 2019 de Bogotá. Esta encuesta registra los tiempos de viaje de cada etapa de acuerdo a la percepción de los usuarios. Debido a que la literatura muestra que la percepción de los usuarios generalmente sobrestima los tiempos de espera, ajustamos los tiempos de espera promedio obtenidos de la encuesta dividiendo por un factor de 1.3 (Watkins, Ferris, Borning, Rutherford, y Layton, 2011; Fan, Guthrie, y Levinson, 2016).²⁶ Los resultados muestran que la duración de los viajes es muy similar entre TransMilenio y el SITP Zonal, y es cerca de 10 % mayor en hora pico para ambos sistemas.

Los valores del tiempo en cada etapa del viaje vienen del estudio realizado por Batarce y cols. (2015), los cuales ajustamos a pesos de 2019 con base en el crecimiento del salario mínimo. Este estudio utilizó información de preferencias reveladas y declaradas para estimar los valores del tiempo en TransMilenio, con un enfoque especial en el efecto del hacinamiento en el valor del tiempo de viaje en el bus. Los resultados muestran que este efecto es muy significativo. Desde una valoración base de \$88 por minuto cuando el bus contiene solo un pasajero por m^2 , el valor del tiempo se incrementa en \$52 por minuto por cada pasajero por m^2 adicional, por lo que resulta más de cuatro veces mayor cuando el bus se encuentra a capacidad ($6.5 \text{ pas}/\text{m}^2$). Utilizamos estos mismos valores del tiempo para el SITP Zonal. Finalmente, establecimos el valor del tiempo para los usuarios de vehículos particulares 50 %

²⁵De acuerdo a las cifras oficiales (TransMilenio S.A., 2019), el subsidio para TransMilenio y el SITP Zonal era alrededor de 11 % y 47 % respectivamente en 2019. A diferencia de los niveles de subsidio presentados en la Tabla 4, estos subsidios no incluyen los costos fijos y abarcan todos los periodos de tiempo, incluyendo noches y fines de semana.

²⁶Los tiempos promedio de espera tienen en cuenta que algunos usuarios transbordan entre buses una o más veces, por lo que enfrentan varios procesos de espera. Esta observación es especialmente relevante para TransMilenio, debido a que buena parte de los usuarios transbordan entre rutas alimentadoras y troncales o entre rutas troncales expresas y corrientes. Por ejemplo, en el periodo pico, el usuario promedio de TransMilenio enfrenta 1.5 procesos de espera y en promedio cada proceso toma 10.4 minutos (ajustado por percepción, es decir, el promedio obtenido de la encuesta es 1.3 veces mayor), por lo que el tiempo de espera promedio por viaje es 15.7 minutos. Un tiempo de espera promedio de 10.4 minutos es alto considerando que el intervalo promedio es 4.7 minutos (Tabla 3). Sin embargo, no podemos descartar que la variabilidad en el intervalo y la necesidad de esperar más de un bus para poder abordar lleven a este alto promedio por espera.

Tabla 5: Parámetros base - Tiempos de viaje, costo generalizado y elasticidad de la demanda.

	TransMilenio - BRT		SITP Zonal - Bus	
	Pico	Valle	Pico	Valle
<i>Tiempo promedio por viaje (min)</i>				
Acceso y egreso	15.2	15.9	15.5	15.7
Espera*	15.7	12.7	15.2	14.7
En el bus	57.4	51.0	57.5	50.4
Total	88.2	79.6	88.2	80.8
<i>Valor del tiempo (\$/min)</i>				
Acceso y egreso	\$344	\$344	\$344	\$344
Espera	\$198	\$198	\$198	\$198
En el bus (a 1 pas/m ²)	\$88	\$88	\$88	\$88
En el bus (por pas/m ² adicional)	\$52	\$52	\$52	\$52
En el bus (según hacinamiento)	\$264	\$189	\$241	\$175
En carro	\$131	\$131	\$131	\$131
Costo generalizado	\$25,501	\$19,489	\$24,143	\$19,102
% tarifa en costo generalizado	7.8 %	9.6 %	8.0 %	10.2 %
<i>Elasticidad de la demanda</i>				
A la tarifa	-0.166	-0.178	-0.166	-0.178
Al costo generalizado	-2.11	-1.85	-2.07	-1.75

* Ajustamos por percepción los tiempos de espera reportados en la encuesta de movilidad dividiendo por un factor de 1.3.

Notas: Valores monetarios en pesos colombianos de 2019. La tasa de cambio osciló en 2019 alrededor de \$1 USD = \$3,200 COL. TransMilenio incluye rutas alimentadoras y duales. Periodo pico: 5-8am y 4-7pm. Periodo valle: 8am-4pm.

superior al valor del tiempo en el bus a un pasajero por m^2 , teniendo en cuenta que su nivel de ingreso es superior al de los usuarios de transporte público.

Los tiempos promedio y los valores del tiempo nos permiten estimar el costo generalizado promedio de los viajes en transporte público, el cual añade a la tarifa la valoración total de los tiempos de viaje. Los resultados revelan que la tarifa representa solo alrededor de 8 % del costo generalizado en hora pico, y 10 % en hora valle.

Guzman y cols. (2021) estimaron la elasticidad de la demanda de TransMilenio a la tarifa, e identificaron valores específicos (aunque similares) para los periodos pico y valle.²⁷ Para simplificar el modelo de demanda definido por las Ecuaciones 3, 4 y 5, asumimos que la relación entre la demanda y el costo generalizado tiene una elasticidad constante. Estimamos esta elasticidad dividiendo la elasticidad a la tarifa entre el porcentaje que la tarifa representa del costo generalizado (el Apéndice C analiza la relación entre las elasticidades a la tarifa y al costo generalizado). Para el SITP Zonal, asumimos las mismas elasticidades a la tarifa en ambos periodos de tiempo.²⁸

4.1.4. Externalidades

La Tabla 6 presenta los principales parámetros relacionados con las externalidades causadas por buses y carros. Tomamos los factores de emisión de material particulado y CO_2 para buses y carros del inventario de emisiones 2018 realizado por la Secretaría Distrital de Ambiente de Bogotá (Secretaría Distrital de Ambiente, Bogotá D.C., 2020), y promediamos estos factores de acuerdo a los kilómetros recorridos por cada tipo de bus.²⁹ Valoramos las emisiones de material particulado en \$1,189/gr de acuerdo a los resultados obtenidos por el Plan Decenal de Descontaminación del Aire para Bogotá (Secretaría Distrital de Ambiente,

²⁷La elasticidad para la hora valle que presentamos en la Tabla 5 es el resultado de promediar las elasticidades que los autores obtienen para los tres grupos geográficos analizados (centro, intermedio y periferia) de acuerdo al número de usuarios en cada grupo (Guzman y cols., 2021, Tabla 6). La elasticidad para la hora pico promedia adicionalmente los periodos pico de la mañana y la tarde con igual peso.

²⁸La evidencia internacional muestra que la elasticidad de los sistemas de buses convencionales es usualmente mayor a la de los sistemas de transporte público masivo (Horchler y Tirachini, 2021). Sin embargo, el ingreso promedio de los usuarios del SITP Zonal es un poco menor al de los usuarios de TransMilenio, lo que puede contrarrestar este efecto, ya que personas con menores ingresos tienen usualmente menores elasticidades, como Guzman y cols. (2021) encuentran para Bogotá.

²⁹Estos promedios tienen en cuenta que en 2019 TransMilenio contaba con cerca de 280 buses híbridos (diesel-eléctricos) en rutas duales y alimentadoras, mientras el SITP Zonal contaba con cerca de 56 buses híbridos. TransMilenio inició en el segundo semestre de 2019 la renovación de la flota de Fases I y II (buses articulados y biarticulados). Más de la mitad de los buses de la nueva flota utilizan gas natural. Sin embargo, no incluimos esta nueva flota en el análisis debido a que la mayor parte de la renovación ocurrió después de finalizar el 2019. De forma similar, no incluimos los buses eléctricos que comenzaron a operar en el SITP Zonal en 2021. Estas nuevas tecnologías seguramente reducirán los factores de emisión promedio que se muestran en la Tabla 6 para TransMilenio y el SITP Zonal. Para vehículos particulares, promediamos los factores de emisión de automóviles, camperos y camionetas. Los factores de emisión para buses y vehículos particulares no incluyen emisiones por generación o transporte de energía, solo emisiones en carretera (*tailpipe emissions*).

Bogotá D.C. y Universidad de los Andes, 2010),³⁰ y las emisiones de CO₂ en \$67,500/ton de acuerdo a las estimaciones del Gobierno de los Estados Unidos (Interagency Working Group on Social Cost of Greenhouse Gases, 2021).³¹

Tabla 6: Parámetros base - Externalidades.

	TransMilenio - BRT		SITP Zonal - Bus	
	Pico	Valle	Pico	Valle
<i>Factores de emisión (gr/bus-km)</i>				
Material particulado	0.120	0.120	0.184	0.184
CO ₂	1,367	1,367	1,299	1,299
Tiempo de parada por pasajero (seg)	1.5	1.5	5	5
	Vehículos particulares			
	Pico	Valle		
<i>Factores de emisión (gr/carro-km)</i>				
Material particulado	0.023	0.023		
CO ₂	400	400		
Impuesto a la gasolina (\$/km)	\$39	\$39		

Notas: Valores monetarios en pesos colombianos de 2019. La tasa de cambio osciló en 2019 alrededor de \$1 USD = \$3,200 COL. Los factores de emisión no incluyen emisiones por generación o transporte de energía, solo emisiones en carretera (*tailpipe emissions*). TransMilenio incluye rutas alimentadoras y duales. Periodo pico: 5-8am y 4-7pm. Periodo valle: 8am-4pm. Los tiempos de parada incluyen abordaje y salida.

Para las externalidades por congestión, utilizamos los resultados obtenidos por Gomez-Gelvez (2021) para Bogotá. Estos resultados indican que 10,000 vehículos particulares adicionales en las calles aumentan los tiempos de viaje en 0.24 minutos por kilómetro.³² Aplicamos este parámetro para estimar las externalidades por congestión que cambios en los flujos de vehículos particulares y buses del SITP Zonal generan en los usuarios de transporte público y privado.³³ Al igual que Parry y Small (2009) para Londres, asumimos que los buses del

³⁰El plan decenal contempló acciones que reducirían las emisiones de material particulado en el periodo 2008-2020 en 19,400 toneladas, con beneficios en salud de \$15.8 billones (pesos de 2009). Estos resultados sugieren un valor de \$814/gr. Actualizamos este valor a pesos de 2019 utilizando la variación del Índice de Precios al Consumidor (IPC).

³¹Tomamos un valor de referencia de USD \$50/ton. Expresamos este valor en pesos colombianos utilizando una tasa de cambio ajustada por paridad del poder adquisitivo de \$1,350 pesos por dólar.

³²Este efecto corresponde al parámetro obtenido por Gomez-Gelvez para el segundo rango más alto de flujo vehicular (Gomez-Gelvez, 2021, Tabla 2 - Columna 4). El rango más alto aplica solo para la hora pico de la mañana (6-7am). Como se observa en la Tabla 3, el flujo de vehículos particulares durante los periodos pico y valle es muy similar, por lo que utilizamos el mismo parámetro de congestión marginal para ambos periodos.

³³Además del efecto en los usuarios de vehículos particulares y del SITP Zonal, tenemos en cuenta el efecto en los usuarios del SITP Provisional, taxis y aplicaciones móviles tipo Uber. Aunque estos usuarios no son

SITP Zonal equivalen en promedio a cinco vehículos particulares en términos de su efecto en congestión (α en la Ecuación 24).³⁴ Debido a que las rutas troncales de TransMilenio operan en carriles exclusivos, asumimos también que su operación no afecta ni es afectada por los niveles de congestión en la ciudad.³⁵

Los tiempos de parada por pasajero (Tabla 6) para TransMilenio se basan en Scoria (2005), y para el SITP Zonal se basan en las recomendaciones del *Transit Capacity and Quality of Service Manual* (Transit Cooperative Research Program, 2013).³⁶ El parámetro θ en la Ecuación 7 resulta al dividir estos tiempos de espera por la distancia promedio de los viajes.

Colombia impone impuestos significativos a la producción y consumo de gasolina. En 2019, el valor total de estos impuestos era \$1,944 por galón.³⁷ El valor por kilómetro presentado en la Tabla 6 asume un rendimiento de 50 km/gal para vehículos particulares. Este valor corresponde al parámetro τ^{car} introducido en la Ecuación 3.

Guzman y cols. (2021) no encontraron una relación estadísticamente significativa entre el precio de la gasolina y la demanda de TransMilenio. Debido a la relación teórica entre esta elasticidad y la elasticidad del uso de vehículos particulares a la tarifa de TransMilenio, este resultado sugiere que la fracción del incremento de demanda en TransMilenio debido a reducciones tarifarias que proviene de usuarios de vehículos particulares es baja. Para los resultados del escenario base, asumimos que esta fracción es 10 % para cada sistema en ambos periodos de tiempo. El análisis de sensibilidad muestra cómo cambian los resultados cuando esta fracción varía.

4.1.5. Elasticidades del tiempo de espera y del hacinamiento

Las Ecuaciones 20, 21 y 22 expresan los componentes del efecto Mohring neto como función de las elasticidades del tiempo de espera a la frecuencia y al factor de carga, y del parte del modelo descrito en la Sección 2, sí son afectados por cambios en el nivel de congestión.

³⁴Parry y Small (2009) asumen que este parámetro es 5 para Londres y 4 para Washington D.C. y Los Angeles. Este parámetro es naturalmente superior a la relación de tamaño entre buses y vehículos particulares debido a la necesidad de los buses de realizar paradas, por lo que depende de factores como los tiempos de parada y la ubicación de los paraderos. En la Sección 4.2.3 presentamos cómo varía el subsidio óptimo del SITP Zonal cuando este parámetro de equivalencia es 4.

³⁵Esta suposición ignora los efectos que se pueden generar entre el flujo vehicular de TransMilenio y el flujo vehicular general debido a bloqueos en las intersecciones. Ignoramos también la congestión interna que se puede generar en los carriles exclusivos de TransMilenio.

³⁶Estos tiempos de parada incluyen abordaje y salida. Scoria (2005) muestra que el tiempo de parada por pasajero para TransMilenio es cerca de un segundo. Tomamos 1.5 segundos por pasajero teniendo en cuenta que un usuario promedio realiza 1.5 etapas en cada viaje. Para el SITP Zonal, el tiempo por pasajero incluye 2.75 segundos por abordaje con tarjeta inteligente, 0.5 segundos por abordaje a desnivel y 1.75 por salida por puerta trasera (Transit Cooperative Research Program, 2013, Capítulo 6 - Tabla 6-4).

³⁷Este total incluye el impuesto nacional (\$526/gal para gasolina corriente), el impuesto al carbono (\$148/gal) y la sobretasa territorial (25 % del precio de referencia \$5,079/gal).

hacinamiento al factor de carga (η_{wf} , η_{wl} y η_C). Cuando el intervalo promedio entre buses es pequeño (<15 minutos), es común asumir que los pasajeros llegan a las estaciones de forma aleatoria (Parry y Small, 2009). En este caso, la elasticidad del tiempo de espera a la frecuencia es -1.³⁸ Dado que el intervalo promedio para TransMilenio y para el SITP Zonal es menor a 15 minutos en ambos periodos de tiempo (Tabla 3), tomamos $\eta_{wf} = -1$ para todos los casos.

La elasticidad del tiempo de espera al factor de carga es mucho más difícil de aproximar de forma teórica.³⁹ Para estimar esta elasticidad para TransMilenio, sacamos provecho del hecho que el intervalo promedio en el periodo pico es muy similar al del periodo valle. Sin embargo, el tiempo promedio de espera por etapa es 19.6 % mayor en el periodo pico.⁴⁰ Este aumento se debe principalmente a que el factor de carga de los buses es 49.5 % mayor en el periodo pico (Tabla 3), lo que hace que un mayor porcentaje de usuarios no pueda abordar el primer bus que pasa luego de su arribo a la estación.⁴¹ Estos cambios porcentuales entre periodos sugieren una elasticidad promedio del tiempo de espera al factor de carga de 0.4. Debido a que esta elasticidad es probablemente mayor cuando el factor de carga es más alto, utilizamos $\eta_{wl} = 0,5$ para el periodo pico de TransMilenio y $\eta_{wl} = 0,3$ para el periodo valle. De forma similar, utilizamos $\eta_{wl} = 0,4$ y $\eta_{wl} = 0,2$ para los periodos pico y valle del SITP Zonal teniendo en cuenta que su factor de carga es menor que el de TransMilenio.

Finalmente, la relación entre el hacinamiento promedio experimentado por los pasajeros y el factor de carga está mediada solo por el hacinamiento a capacidad y el coeficiente de variabilidad de la ocupación de los buses (ver Sección 4.1.1). Asumimos que este coeficiente permanece constante ante cambios en la ocupación promedio, por lo que el hacinamiento

³⁸Teniendo en cuenta la variabilidad del intervalo entre buses, el tiempo de espera promedio experimentado por pasajeros que llegan de forma aleatoria a las estaciones es $\frac{h}{2}(1 + CV_h^2)$, donde h es el intervalo promedio y CV_h es su coeficiente de variabilidad. La elasticidad del tiempo de espera a la frecuencia (inverso del intervalo) es entonces -1, asumiendo que el coeficiente de variabilidad del intervalo se mantiene constante ante cambios en la frecuencia. Esta elasticidad es -1 incluso si algunos pasajeros no logran abordar el primer bus que pasa luego de su arribo a la estación, ya que en este caso su tiempo de espera se incrementa en h , por lo que este incremento también es inversamente proporcional a la frecuencia.

El uso de aplicaciones tecnológicas que permiten conocer en tiempo real la ubicación de los buses podría invalidar la suposición de llegadas aleatorias incluso para intervalos promedio bajos. Sin embargo, de acuerdo a los resultados de la encuesta de movilidad 2019 de Bogotá, el uso de aplicaciones tecnológicas para planear el viaje es muy bajo (alrededor de 5 %).

³⁹Oldfield y Bly (1988) y Cominetti y Correa (2001) ofrecen análisis teóricos de esta relación.

⁴⁰El tiempo promedio de espera por etapa (ajustado por percepción) es 8.7 minutos en el periodo valle y 10.4 minutos en el periodo pico. Los tiempos promedio de espera presentados en la Tabla 5 son por viaje, por lo que tienen en cuenta que un viaje promedio tiene 1.5 etapas en el periodo pico y 1.46 en el periodo valle.

⁴¹Aunque el intervalo promedio es igual para ambos periodos, la operación de rutas expresas adicionales en el periodo pico puede disminuir el tiempo promedio de espera si algunos usuarios eligen entre diferentes rutas con base en el primer bus que llega a su estación. Este efecto implicaría que estamos subestimando la elasticidad del tiempo de espera al factor de carga.

experimentado por los pasajeros resulta proporcional el factor de carga. Utilizamos entonces $\eta_C = 1$ para los dos sistemas en ambos periodos.

4.2. Resultados

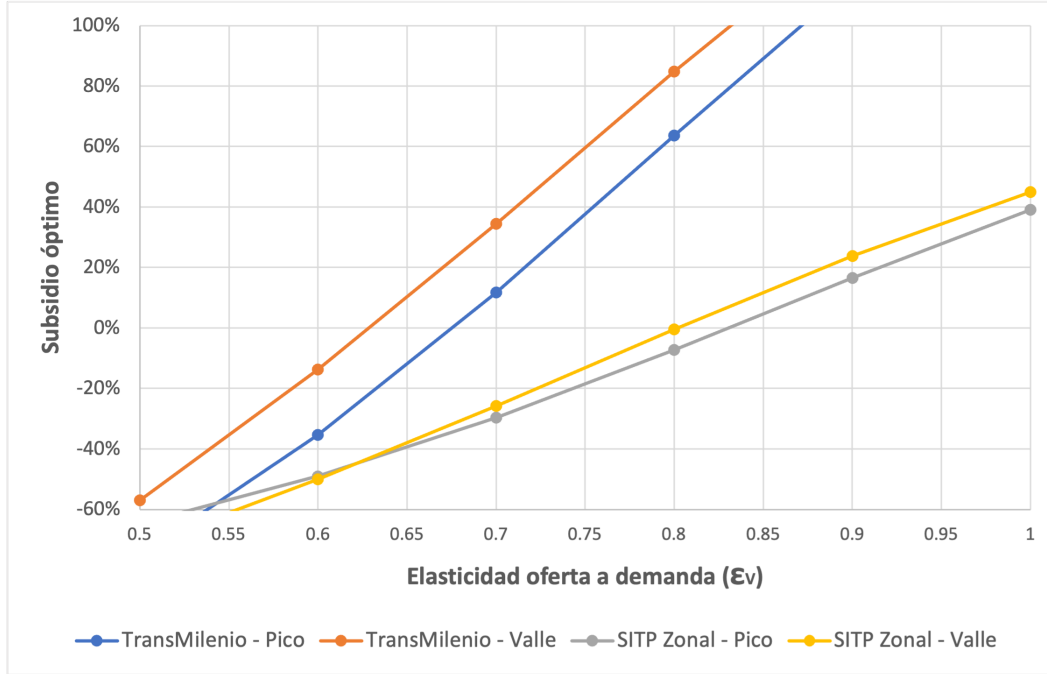
La sección anterior cuantificó todos los parámetros necesarios para aplicar el modelo de la Sección 2 a los sistemas de transporte público de Bogotá, excepto por la elasticidad de la oferta de kilómetros de bus a la demanda de transporte público (ε_V en la Ecuación 17). Esta elasticidad juega un papel crucial en el nivel de subsidio óptimo de los sistemas, debido a que afecta de forma directa el costo marginal operacional (Ecuación 19), todos los componentes del efecto Mohring neto (Ecuaciones 20 a 22) y las externalidades generadas por usuarios de transporte público (Ecuación 24). Presentamos entonces en primer lugar la forma como el subsidio y la tarifa óptimas en los dos sistemas varían en función de ε_V . Luego, revelamos los resultados en detalle dada una elasticidad específica para cada sistema ($\varepsilon_V = 0,7$ para TransMilenio y $\varepsilon_V = 0,9$ para el SITP Zonal). Estos resultados muestran la contribución al subsidio óptimo de cada uno de los argumentos que respaldan o demeritan subsidios al transporte público desde una perspectiva de eficiencia. Finalmente, realizamos un análisis de sensibilidad del subsidio óptimo en cada sistema a dos parámetros clave: la elasticidad de la demanda de transporte público a su tarifa y el porcentaje del aumento en demanda de transporte público que proviene de vehículos particulares. Para el SITP Zonal, analizamos adicionalmente la sensibilidad al número de vehículos particulares que equivalen a un bus en términos de congestión (α en la Ecuación 24).

4.2.1. Subsidios y tarifas óptimas como función de la elasticidad de la oferta a la demanda

La Figura 1 muestra cómo varía el nivel de subsidio óptimo en cada sistema, definido como el porcentaje de los costos operacionales que no son cubiertos por ingresos tarifarios, en función de la elasticidad de la oferta de kilómetros de bus a la demanda de transporte público (ε_V). Para ambos sistemas, el subsidio aumenta a medida que esta elasticidad crece. Este efecto se debe a que la magnitud del efecto Mohring y del costo por hacinamiento cambia sustancialmente en función de la elasticidad. Cuando la elasticidad es cercana a uno (la oferta responde de forma proporcional a la demanda), el efecto Mohring es un argumento importante a favor de los subsidios. En este caso, el beneficio por reducción de tiempos de espera es significativo (Ecuación 20), mientras que los costos generados por un mayor nivel de hacinamiento (Ecuaciones 21 y 22) son bajos debido a que este se mantiene constante. Sin embargo, a medida que la elasticidad disminuye, la magnitud del beneficio en tiempos de

espera decrece mientras el tamaño de los costos por hacinamiento aumenta.⁴² El efecto de la elasticidad en el nivel de subsidio óptimo es menor para el SITP Zonal (pendiente menor en la Figura 1). Esta diferencia se debe a que una mayor elasticidad aumenta también las externalidades generadas por los usuarios de transporte público (Ecuación 24), las cuales reducen el subsidio óptimo y son mayores para el SITP Zonal que para TransMilenio.

Figura 1: Subsidio óptimo como función de ε_V .



Notas: El subsidio se mide como el porcentaje de los costos operacionales que no son cubiertos por ingresos tarifarios.

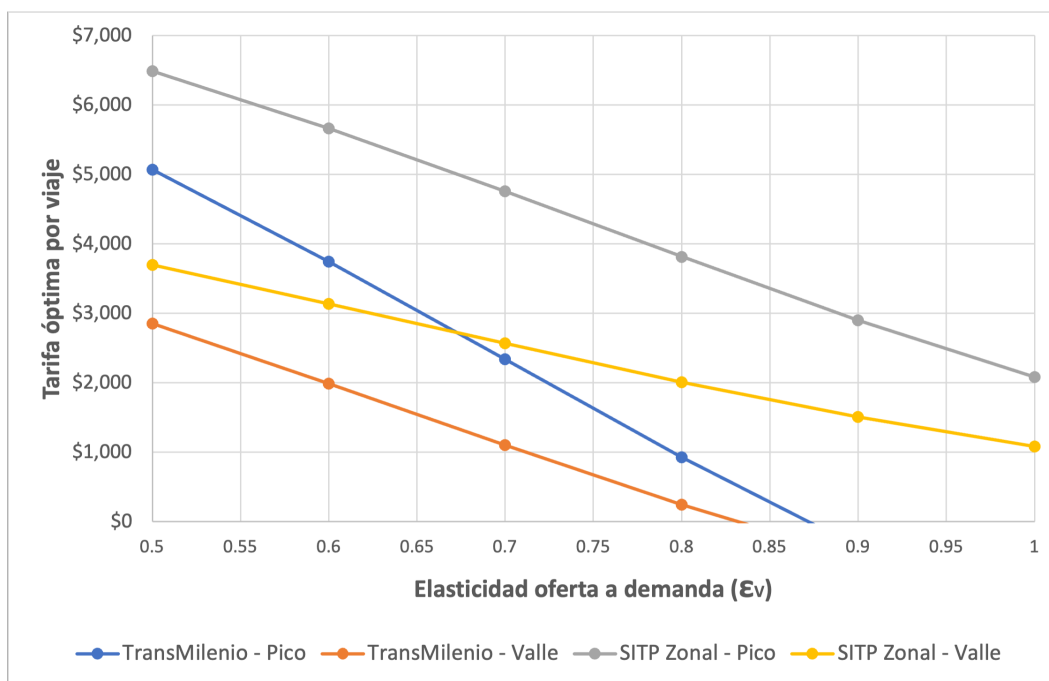
La Figura 1 muestra que un subsidio positivo para TransMilenio solo se justifica cuando la elasticidad de la oferta a la demanda es mayor a 0.65. Este umbral es 0.8 para el SITP Zonal. En general, el subsidio óptimo para TransMilenio es mayor que para el SITP Zonal debido a que el beneficio por reducción de externalidades es mucho mayor para TransMilenio (ver Sección 4.2.2).

El subsidio óptimo, expresado como porcentaje de los costos operacionales, es similar entre periodos pico y valle para ambos sistemas. Sin embargo, la tarifa óptima es significativamente menor para los periodos valle debido a que los costos operacionales son menores (no incluyen el costo de compra de vehículos). La Figura 2 muestra las tarifas óptimas por viaje para los dos sistemas en ambos periodos de tiempo. La tarifa promedio por viaje era cerca de

⁴²Para TransMilenio, los costos por hacinamiento superan el efecto Mohring cuando la elasticidad es menor a 0.85 en el periodo pico o 0.8 en el periodo valle. Para el SITP Zonal, estos umbrales son 0.8 y 0.7 para los periodos pico y valle respectivamente.

\$2,000 en 2019 en todos los casos (Tabla 4). Este nivel de tarifa, o incluso niveles menores, se puede justificar en todos los casos siempre que la elasticidad de la oferta a la demanda sea suficientemente alta. Para TransMilenio, la tarifa óptima resulta negativa (subsidio mayor a 100 %) cuando la elasticidad es mayor a 0.85.

Figura 2: Tarifa óptima por viaje como función de ε_V .



4.2.2. Componentes de la tarifa óptima

El objetivo de esta sección es presentar en detalle los factores que llevan a los niveles de tarifa y subsidio óptimos presentados en la sección anterior. Estos niveles varían significativamente de acuerdo a la elasticidad con la que TransMilenio S.A. ajusta la oferta de kilómetros de bus a cambios en la demanda, como mostramos en la sección anterior. En esta sección, tomamos como valores de referencia $\varepsilon_V = 0,7$ para TransMilenio y $\varepsilon_V = 0,9$ para el SITP Zonal. Estas elasticidades llevan a niveles de subsidio generales cercanos a los que ambos sistemas tenían en 2019. Adicionalmente, una menor elasticidad para TransMilenio reconoce que este sistema puede enfrentar límites de capacidad en ciertas estaciones al aumentar su oferta, especialmente en periodos pico.⁴³

La Tabla 7 muestra los resultados para ambos sistemas dadas estas elasticidades. El costo operacional marginal por pasajero-kilómetro es menor para TransMilenio debido a su

⁴³Parry y Small (2009) utilizan $\varepsilon_V = 0,67$ como valor de referencia, ya que esta elasticidad es la forma óptima como la agencia de transporte público debe responder a cambios en la demanda dadas algunas simplificaciones.

mayor eficiencia (mayor número de pasajeros por bus) y a que la oferta reacciona en menor medida ante cambios en la demanda (menor ε_V , ver Ecuación 19). El efecto Mohring es mayor para el SITP Zonal debido a que este sistema tenía en 2019 menores frecuencias. El costo por hacinamiento es mayor para TransMilenio debido a que su factor de carga en 2019 era superior. Como resultado, el costo por hacinamiento supera el beneficio por el efecto Mohring para TransMilenio, mientras que lo contrario sucede en el SITP Zonal.

La cuantificación de las externalidades muestra que la congestión es de lejos la principal externalidad, especialmente durante el periodo pico. El aporte de los usuarios de TransMilenio a las externalidades es relativamente bajo, debido a que el uso de carriles exclusivos limita su impacto en congestión. Como resultado, el beneficio neto por externalidades es positivo para TransMilenio, incluso teniendo en cuenta que solo 10 % de su demanda adicional proviene de vehículos particulares. El SITP Zonal tiene por el contrario un impacto importante en congestión. Aunque este impacto es menor por pasajero-kilómetro que el de los vehículos particulares, el efecto neto en externalidades es negativo debido a que solo 10 % de los usuarios adicionales usaban vehículos particulares.⁴⁴

Las externalidades por contaminación (material particulado y CO_2) juegan un papel menor en el subsidio total. Aunque todos los buses fuesen eléctricos (cero emisiones), esto tendría solo un impacto marginal en el subsidio final.

Para TransMilenio, el costo por hacinamiento supera los beneficios por efecto Mohring y reducción de externalidades, por lo que la tarifa óptima es superior al costo operacional marginal. Sin embargo, la tarifa no es superior al costo promedio, que incluye los costos fijos, por lo que el nivel de subsidio es positivo.

Para el SITP Zonal, el beneficio del efecto Mohring compensa el costo por hacinamiento y aumento de externalidades, por lo que la tarifa óptima es muy cercana al costo operacional marginal. El nivel de subsidio equivale entonces a la diferencia entre el costo promedio y el costo marginal causada principalmente por la existencia de costos fijos.

4.2.3. Sensibilidad a parámetros clave

Esta sección muestra cómo los niveles de subsidio óptimos varían ante cambios en dos parámetros clave: la elasticidad de la demanda de transporte público a la tarifa y la proporción del incremento en demanda que proviene de vehículos particulares. Adicionalmente, exploramos el impacto de cambios en la equivalencia de buses del SITP Zonal a vehículos particulares en términos de congestión. En todos los casos, mantenemos las elasticidades

⁴⁴Los resultados de Parry y Timilsina (2010) resaltan también la importancia del efecto en congestión de los buses tradicionales en comparación con sistemas que cuentan con carriles exclusivos (metro en el caso de Ciudad de México, BRT en Bogotá).

Tabla 7: Componentes de la tarifa óptima (\$/pas-km).

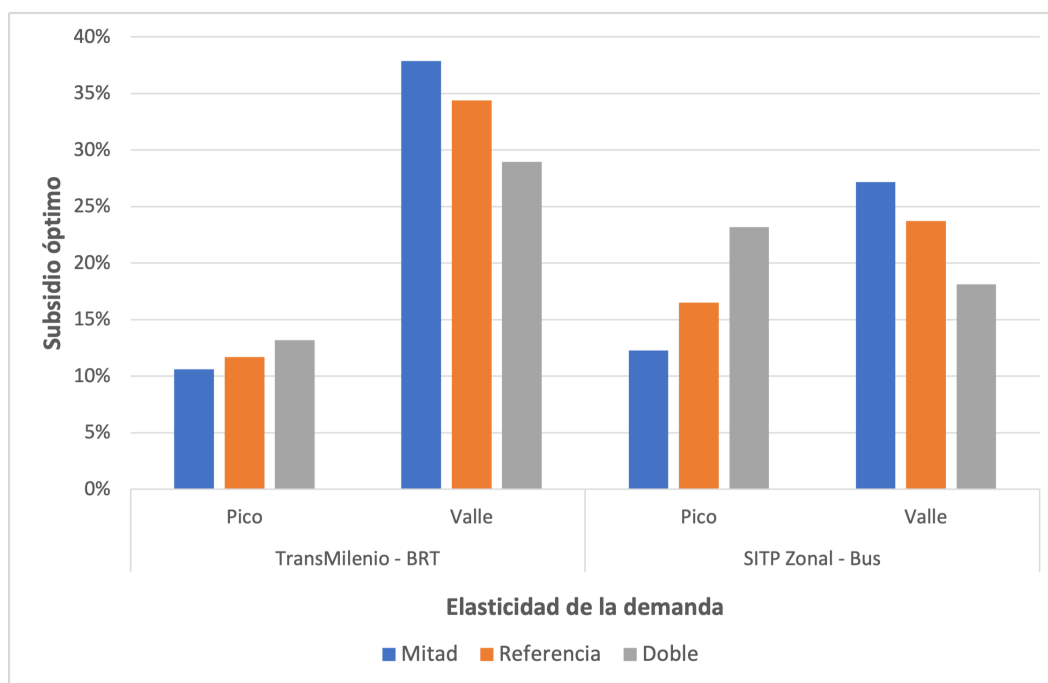
	TransMilenio - BRT		SITP Zonal - Bus	
	Pico	Valle	Pico	Valle
Elasticidad oferta a demanda (ε_V)	0.7	0.7	0.9	0.9
Costo operacional marginal (CM)	\$112.0	\$74.3	\$263.2	\$159.4
Beneficio efecto Mohring (BM):	\$158.7	\$140.9	\$271.7	\$267.7
Costo hacinamiento (CH):				
Tiempo de espera	\$34.0	\$18.1	\$12.1	\$5.9
Hacinamiento en viaje	\$281.3	\$198.3	\$108.1	\$75.0
Total	\$315.3	\$216.4	\$120.2	\$80.9
Externalidades bus (E_b):				
Material particulado	\$0.86	\$1.25	\$5.07	\$7.36
CO ₂	\$0.81	\$1.18	\$2.97	\$4.31
Congestión	-	-	\$252.9	\$240.5
Tiempos de parada	\$11.3	\$6.5	\$4.9	\$2.8
Total	\$12.9	\$9.0	\$265.9	\$255.0
Externalidades carro (E_c):				
Material particulado	\$12.2	\$12.2	\$12.2	\$12.2
CO ₂	\$17.5	\$17.5	\$17.5	\$17.5
Congestión	\$1,123	\$664	\$1,077	\$676
Impuesto a la gasolina	-\$25.2	-\$25.2	-\$25.2	-\$25.2
Total	\$1,128	\$668	\$1,082	\$681
Beneficio por externalidades ($BE=0.1 \cdot E_c - E_b$)	\$99.8	\$57.9	-\$157.7	-\$186.9
Tarifa óptima (\$/pas-km) ($=CM-BM+CH-BE$)	\$169	\$92	\$269	\$160
Tarifa óptima por viaje (TO)	\$2,342	\$1,106	\$2,899	\$1,507
Costo promedio (\$/pas-km)	\$191	\$140	\$323	\$209
Costo promedio por viaje (CP)	\$2,653	\$1,686	\$3,473	\$1,976
Subsidio óptimo ($=(CP-TO)/CP$)	11.7 %	34.4 %	16.5 %	23.7 %

Notas: Valores monetarios en pesos colombianos de 2019. La tasa de cambio osciló en 2019 alrededor de \$1 USD = \$3,200 COL. TransMilenio incluye rutas alimentadoras y duales. Periodo pico: 5-8am y 4-7pm. Periodo valle: 8am-4pm.

de la oferta a la demanda utilizadas en la sección anterior ($\varepsilon_V = 0,7$ para TransMilenio y $\varepsilon_V = 0,9$ para el SITP Zonal).

La Figura 3 revela la sensibilidad del subsidio óptimo a la elasticidad de la demanda. Además de los valores de referencia (Tabla 5), presentamos los resultados cuando la elasticidad es el doble o la mitad de estos valores.⁴⁵ Una mayor elasticidad aumenta el nivel de subsidio en los periodos pico, en los que la tarifa óptima es mayor a la tarifa promedio de 2019, y lo disminuye en los periodos valle, en los que la tarifa óptima es menor a la de 2019. En todos los casos, el subsidio óptimo sigue siendo positivo.

Figura 3: Subsidio óptimo como función de la elasticidad de la demanda.



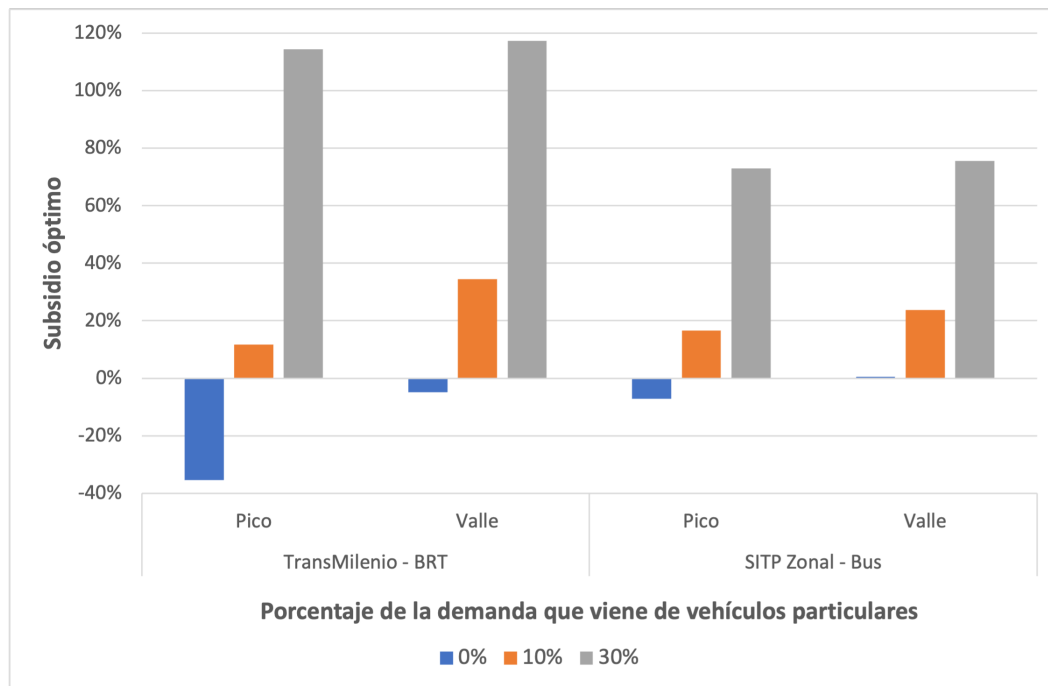
Notas: El subsidio se mide como el porcentaje de los costos operacionales que no son cubiertos por ingresos tarifarios.

La Figura 4 muestra que el porcentaje de demanda que proviene de vehículos particulares influye de forma significativa en el subsidio óptimo, debido al alto impacto de los vehículos particulares en la congestión. Si cambios en la tarifa de transporte público no afectan la demanda de vehículos particulares, los subsidios al transporte público no se justifican en ningún caso. Para TransMilenio, el costo por hacinamiento superaría el beneficio por efecto Mohring y los costos fijos. Para el SITP Zonal, el principal costo sería el de sus externalidades por congestión. Por el contrario, si el porcentaje de la demanda que viene de vehículos particulares es 30 % (tres veces el valor de referencia), el subsidio óptimo para TransMilenio

⁴⁵Cambios en la elasticidad de la demanda a la tarifa llevan a cambios proporcionales en su elasticidad al costo generalizado, que es el parámetro que asumimos constante en el modelo de demanda.

es mayor al 100 % (tarifa negativa) y el subsidio al SITP Zonal es más de tres veces mayor. Este resultado se debe principalmente a la importante externalidad por congestión causada por los vehículos particulares, por lo que la introducción de cobros por congestión disminuiría en gran medida el tamaño de los subsidios. Dado que muy pocos estudios empíricos identifican el porcentaje de demanda que proviene de vehículos particulares, estos resultados resaltan su importancia para la región.

Figura 4: Subsidio óptimo como función del porcentaje de demanda que viene de vehículos particulares.



Notas: El subsidio se mide como el porcentaje de los costos operacionales que no son cubiertos por ingresos tarifarios.

Finalmente, estudiamos también la sensibilidad del subsidio al SITP Zonal al factor de equivalencia entre buses y carros en términos de congestión. El valor de referencia en este caso determina que un bus del SITP Zonal equivale a cinco vehículos particulares. Si un bus equivale a solo cuatro vehículos particulares, el subsidio óptimo para el SITP Zonal aumenta de 16.5 % a 28.5 % en el periodo pico y de 23.7 % a 41.2 % en el periodo valle. Este parámetro incide significativamente en los resultados debido a que las externalidades por congestión juegan un papel crucial en el tamaño del subsidio óptimo. Análisis empíricos que estimen este parámetro resultarían entonces muy útiles para mejorar la confiabilidad de los estudios de tarificación óptima en transporte en la región.

5. Conclusiones

Parry y Small (2009) concluyeron que, para las áreas metropolitanas de Londres, Los Angeles y Washington D.C., los altos subsidios operacionales que reciben los sistemas de transporte público se pueden justificar con base en argumentos de eficiencia. Para los sistemas de transporte público de Bogotá (BRT y buses convencionales), encontramos que subsidios operacionales mayores a los que existían en 2019 (antes de la pandemia) solo se pueden justificar con base en argumentos de eficiencia siempre que TransMilenio S.A. (agencia de transporte público) logre responder a aumentos en la demanda con aumentos proporcionales en la oferta de servicios (kilómetros de bus). De lo contrario, el costo generado por un mayor hacinamiento en los buses sobrepasa los beneficios por reducción de tiempos de espera (efecto Mohring) y externalidades. El modelo económico que utilizamos cuantifica el costo que un mayor hacinamiento genera en incomodidad (mayor valor del tiempo de viaje) y en tiempos de espera (necesidad de esperar más de un bus para abordar). Estas consideraciones son importantes para las ciudades latinoamericanas, las cuales tienen usualmente altos niveles de uso de transporte público, y constituye una contribución a los estudios anteriores sobre subsidios óptimos al transporte público.

Estos resultados sugieren también que los niveles de oferta en los sistemas de transporte público de Bogotá en 2019 eran muy bajos, lo que genera costos excesivos para los usuarios por hacinamiento. El modelo económico que utilizamos asume que la agencia de transporte público mantiene constante el tamaño de los buses y el número de rutas, y responde a cambios en la demanda con cambios en el número de bus-kilómetros de acuerdo a una elasticidad constante. Si la agencia ajusta estas variables de forma óptima, el tamaño del subsidio probablemente cambiaría de forma significativa.⁴⁶ Sin embargo, este análisis se encuentra fuera del alcance de este estudio debido a que cambios significativos en la oferta (ej: aumento significativo del número de rutas o frecuencias) requerirían inversiones en infraestructura y harían que varios de los parámetros marginales que utiliza nuestro modelo (ej: elasticidad de la demanda) no fuesen confiables. En este sentido, nuestro análisis se enfoca en una optimización tarifaria de corto plazo que, sin restar importancia a las consideraciones de largo plazo, consideramos relevante para las agencias de transporte público de la región.

El efecto en congestión es la principal externalidad causada por los usuarios de vehículos particulares y buses convencionales en Bogotá.⁴⁷ Aunque el impacto marginal en congestión

⁴⁶No es claro si el tamaño del subsidio aumentaría o disminuiría. Aumentos en la oferta reducirían el costo por hacinamiento, pero también el beneficio por reducción en tiempos de espera.

⁴⁷Cabe recordar que Bogotá usualmente ocupa los primeros puestos en rankings internacionales sobre congestión (ver por ejemplo el *Global Traffic Scorecard* (Inrix, 2019) o Calatayud, Sánchez, Bedoya, Giráldez, y Márquez (2021) para América Latina).

de un usuario de buses convencionales es menor al de un usuario de vehículos particulares, el efecto neto en externalidades que resulta de aumentar la demanda de buses convencionales a través de reducciones en la tarifa es negativo. Esto se debe a que solo un porcentaje bajo del aumento en demanda corresponde a usuarios que migran de vehículos particulares a buses convencionales. Sin embargo, la evidencia empírica de este porcentaje es muy escasa, por lo que estudios adicionales serían muy útiles para la región. Aunque este porcentaje también es probablemente bajo para TransMilenio, el efecto neto en externalidades es positivo en este caso, debido a que el uso de carriles exclusivos limita el efecto en congestión de los buses articulados. Esta diferencia hace que se justifiquen subsidios mayores para TransMilenio en comparación con el SITP Zonal, y sugiere que la implementación de carriles exclusivos para el SITP Zonal aumentaría el tamaño del subsidio óptimo.⁴⁸ Adicionalmente, la importancia de las externalidades por congestión subraya la utilidad de estudios empíricos que estimen el efecto marginal de vehículos en congestión aprovechando el aumento reciente en fuentes de información sobre flujos vehiculares y velocidades.

Aunque la reducción de externalidades es un argumento que promueve la existencia de subsidios en TransMilenio, no debemos olvidar que este es un argumento de segundo mejor. Como estudios anteriores han documentado (Basso y Silva, 2014), el beneficio que genera un cobro por congestión adecuado a vehículos particulares es mucho mayor al que se puede lograr mediante subsidios óptimos al transporte público. Los impuestos a la gasolina que actualmente existen en Colombia solo logran internalizar las externalidades que los vehículos particulares generan por contaminación local (material particulado) y global (CO_2),⁴⁹ pero son completamente insuficientes para hacer frente a las externalidades por congestión. Bogotá inició en 2020 una política de cobro al uso de vehículos particulares de acuerdo a su número de placa, reemplazando su anterior política de prohibición. Aunque este esquema se encuentra aún lejos de un cobro por congestión adecuado, probablemente permitirá internalizar parte del impacto en congestión de los vehículos particulares, lo que reduciría el tamaño del subsidio que se puede justificar para TransMilenio y el SITP Zonal en términos de eficiencia.

Nuestro análisis para Bogotá no tiene en cuenta el costo en eficiencia que conlleva recaudar, mediante impuestos a otros mercados, los recursos públicos necesarios para cubrir el subsidio (conocido como el costo marginal de los recursos públicos). Aunque este costo es un argumento usual en contra de los subsidios al transporte público, la literatura económica reciente sobre impuestos ambientales muestra que puede existir un efecto positivo de los sub-

⁴⁸Bogotá inició en 2020 la implementación de carriles exclusivos para buses convencionales en algunas de las vías principales de la ciudad.

⁴⁹En 2019, el impuesto equivalía a \$39/km (Tabla 6), mientras las externalidades de vehículos particulares por material particulado y CO_2 suman \$29.7/km (Tabla 7). Las externalidades por congestión sobrepasan \$1,000/km en hora pico (Tabla 7).

sidios que contrarreste o incluso sobrepase este costo, por lo que nuestra omisión no genera un sesgo claro en favor de los subsidios. Este efecto positivo se da en la medida en que el subsidio al transporte público genere aumentos en la oferta o demanda del mercado distorsionado por los impuestos. Dado que uno de los principales mercados afectado por impuestos es el mercado laboral, son necesarios estudios empíricos que permitan conocer mejor la relación entre el transporte y el mercado laboral en ciudades latinoamericanas para juzgar la validez del costo marginal de los recursos públicos como argumento en contra de los subsidios.

El enfoque de este estudio en argumentos de eficiencia no busca restarle importancia a otros motivos que pueden justificar subsidios al transporte público. Por ejemplo, estos subsidios se pueden ver como una política redistributiva o que busca garantizar la asequibilidad al transporte público de personas con bajos ingresos y otros grupos vulnerables. Sin embargo, estos objetivos pueden resultar mejor atendidos a través de mecanismos externos al sector transporte, como sistemas generales de impuestos y transferencias monetarias (Serebrisky, Gómez-Lobo, Estupiñan, y Muñoz-Raskin, 2009).

Referencias

- Akbar, P. A., y Duranton, G. (2017). *Measuring the cost of congestion in a highly congested city: Bogotá* (Working paper n.º 2017/04). CAF Development Bank of Latin America.
- Basso, L. J., y Silva, H. E. (2014). Efficiency and substitutability of transit subsidies and other urban transport policies. *American Economic Journal: Economic Policy*, 6(4), 1-33.
- Batarce, M., Muñoz, J. C., de Dios Ortuzar, J., Raveau, S., Mojica, C., y Rios, R. A. (2015). *Evaluation of passenger comfort in bus rapid transit systems* (Technical Note n.º 770). Inter-American Development Bank.
- Börjesson, M., Fung, C. M., y Proost, S. (2017). Optimal prices and frequencies for buses in stockholm. *Economics of Transportation*, 9, 20-36.
- Calatayud, A., Sánchez, S., Bedoya, F., Giráldez, F., y Márquez, J. M. (2021). *Congestión urbana en américa latina y el caribe: características, costos y mitigación* (Monografía n.º 902). Banco Interamericano de Desarrollo.
- Cominetti, R., y Correa, J. (2001). Common-lines and passenger assignment in congested transit networks. *Transportation Science*, 35(3), 250-267.
- Davis, L. W. (2020, December). *Estimating the price elasticity of demand for subways: Evidence from mexico* (Working paper n.º 28244). National Bureau of Economic Research.
- de Grange, L., Gonzalez, F., Muñoz, J. C., y Troncoso, R. (2013). Aggregate estimation of the price elasticity of demand for public transport in integrated fare systems: The case of Transantiago. *Transport Policy*, 29, 178-185.
- Fan, Y., Guthrie, A., y Levinson, D. (2016). *Perception of waiting time at transit stops and stations* (Inf. Téc. n.º 16-01). Center for Transportation Studies, University of Minnesota.
- Gomez-Gelvez, J. A. (2021). *Market structure and congestion externalities: Theory and application to the ride-hailing industry* (Tesis Doctoral no publicada). University of Maryland, College Park - Department of Agricultural and Resource Economics.
- Guzman, L. A., Beltran, C., Bonilla, J., y Cardona, S. G. (2021). BRT fare elasticities from smartcard data: Spatial and time-of-day differences. *Transportation Research Part A*, 150, 335-348.
- Horcher, D., y Tirachini, A. (2021). A review of public transportation economics. *Economics of Transportation*, 25.
- Inrix. (2019). *Inrix global traffic scorecard*. Descargado de <https://inrix.com/scorecard/>
- Interagency Working Group on Social Cost of Greenhouse Gases. (2021). *Social Cost of*

- Carbon, Methane and Nitrous Oxide - Interim estimates under executive order 13990* (Technical support document). United States Government.
- Kreindler, G. E. (2020). *Peak-hour road congestion pricing: Experimental evidence and equilibrium implications*. (Available at <https://sites.google.com/site/gabrielkreindler>)
- Lopez-Ghio, R., Bocarejo, J. P., y Blanco, A. (2018). *Políticas de tarificación por congestión: Efectos potenciales y consideraciones para su implementación en Bogotá, Ciudad de México y Santiago* (Monografía n.º 603). Banco Interamericano de Desarrollo.
- Mangrum, D., y Molnar, A. (2018). *The marginal congestion of a taxi in New York City*. (Available at <https://www.danielmangrum.com>)
- Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *American Economic Review*, 62(4), 591-604.
- Oldfield, R. H., y Bly, P. H. (1988). An analytic investigation of optimal bus size. *Transportation Research Part B*, 22B(5), 319-337.
- Parry, I. W. H., y Bento, A. (2001). Revenue recycling and the welfare effects of road pricing. *The Scandinavian Journal of Economics*, 103(4), 645-671.
- Parry, I. W. H., y Small, K. A. (2009). Should urban transit subsidies be reduced? *American Economic Review*, 99(3), 700-724.
- Parry, I. W. H., y Timilsina, G. R. (2010). How should passenger travel in Mexico City be priced? *Journal of Urban Economics*, 68, 167-182.
- Rivas, M. E., Suárez-Alemán, A., y Serebrisky, T. (2019a). *Hechos estilizados de transporte urbano en América Latina y el Caribe* (Nota Técnica n.º 1640). Banco Interamericano de Desarrollo.
- Rivas, M. E., Suárez-Alemán, A., y Serebrisky, T. (2019b). *Políticas de transporte urbano en América Latina y el Caribe: dónde estamos, cómo llegamos aquí y hacia dónde vamos* (Monografía n.º 719). Banco Interamericano de Desarrollo.
- Rodríguez, C., Gallego, J. M., Martínez, D., Montoya, S., y Peralta-Quiros, T. (2016). Examining implementation and labor market outcomes of targeted transit subsidies. *Transportation Research Record: Journal of the Transportation Research Board*(2581), 9-17.
- Scorcio, H. M. (2005). *Mejoramiento de la función tiempo de parada para transmilenio* (Tesis de Master no publicada). Universidad de los Andes, Bogotá, Colombia.
- Secretaría Distrital de Ambiente, Bogotá D.C. (2020). *Inventario de emisiones de Bogotá* (Inf. Téc.). Secretaría Distrital de Ambiente, Bogotá D.C.
- Secretaría Distrital de Ambiente, Bogotá D.C. y Universidad de los Andes. (2010). *Plan decenal de descontaminación del aire para Bogotá* (Inf. Téc.).
- Secretaría Distrital de Movilidad, Bogotá D.C. (2019). *Encuesta de movilidad 2019*. Des-

- cargado de <https://www.simur.gov.co/encuestas-de-movilidad>
- Serebrisky, T., Gómez-Lobo, A., Estupiñan, N., y Muñoz-Raskin, R. (2009). Affordability and subsidies in public urban transport: What do we mean, what can be done? *Transport Reviews*, 29(6), 715-739.
- Soza-Parra, J., Raveau, S., Muñoz, J. C., y Cats, O. (2019). The underlying effect of public transport reliability on users' satisfaction. *Transportation Research Part A*, 126, 83-93.
- Tirachini, A., y Proost, S. (2021). Transport taxes and subsidies in developing countries: The effect of income inequality aversion. *Economics of Transportation*, 25.
- Transit Cooperative Research Program. (2013). *Transit capacity and quality of service manual - Third edition* (Inf. Téc.). Transportation Research Board.
- TransMilenio S.A. (2019). *Informe de gestión 2016-2019* (Inf. Téc.). TransMilenio S.A.
- Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., y Layton, D. (2011). Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A*, 45, 839-848.
- West, S. E., y Williams, R. C. (2007). Optimal taxation and cross-price effects on labor supply: Estimates of the optimal gas tax. *Journal of Public Economics*, 91, 593-617.
- Yang, J., Purevjav, A.-O., y Li, S. (2020). The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in Beijing. *American Economic Journal: Economic Policy*, 12(1), 418-453.
- Zárate, R. D. (2020, July). *Spatial misallocation, informality and transit improvements: Evidence from Mexico City*. (Available at <https://www.romandavidzarate.com>)

Apéndice

A. Condiciones de primer orden

El proceso de maximización del agente representativo se resume como:

$$\max_{X, M^{bus}, M^{car}} U(X, M^{bus}, M^{car}, T^{bus}, W^{bus}, A^{bus}, C^{bus}, T^{car}, Z^{pol}) \quad (27)$$

$$\text{Sujeto a: } I - Sub = X + p^{bus} M^{bus} + (p^{car} + \tau^{car}) \frac{M^{car}}{o^{car}} \quad (28)$$

Donde:

$$T^{bus} = t^{bus} M^{bus}; \quad W^{bus} = w^{bus} M^{bus}; \quad A^{bus} = a^{bus} M^{bus}; \quad T^{car} = t^{car} M^{car} \quad (29)$$

En este proceso, el agente representativo toma t^{bus} , w^{bus} , a^{bus} , C^{bus} , t^{car} , Z^{pol} , I , Sub , p^{bus} , p^{car} , τ^{car} y o^{car} como valores fijos. El lagrangiano asociado a este proceso de maximización es:

$$\begin{aligned} \mathcal{L} = & U(X, M^{bus}, M^{car}, T^{bus}, W^{bus}, A^{bus}, C^{bus}, T^{car}, Z^{pol}) \\ & + \lambda \left(I - Sub - X - p^{bus} M^{bus} - (p^{car} + \tau^{car}) \frac{M^{car}}{o^{car}} \right) \end{aligned} \quad (30)$$

Donde λ representa el multiplicador de lagrange asociado a la restricción de ingreso.

Las derivadas del lagrangiano con respecto a las tres variables de decisión (X , M^{bus} y M^{car}) y al multiplicador son:

$$\frac{\partial \mathcal{L}}{\partial X} = U_X - \lambda = 0 \quad (31)$$

$$\frac{\partial \mathcal{L}}{\partial M^{bus}} = U_{M^{bus}} + t^{bus} U_{T^{bus}}(C^{bus}) + w^{bus} U_{W^{bus}} + a^{bus} U_{A^{bus}} - \lambda p^{bus} = 0 \quad (32)$$

$$\frac{\partial \mathcal{L}}{\partial M^{car}} = U_{M^{car}} + t^{car} U_{T^{car}} - \lambda \frac{(p^{car} + \tau^{car})}{o^{car}} = 0 \quad (33)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = I - Sub - X - p^{bus} M^{bus} - (p^{car} + \tau^{car}) \frac{M^{car}}{o^{car}} = 0 \quad (34)$$

En este proceso, asumimos que el efecto del hacinamiento (C^{bus}) en la utilidad se refleja en la utilidad marginal del tiempo total de viaje en el bus ($U_{T^{bus}}$), por lo que expresamos esta utilidad marginal como función del hacinamiento.

Utilizamos ahora la primera condición ($U_X = \lambda$) para sacar el multiplicador de las si-

guientes dos condiciones, y expresamos estas condiciones como:

$$\frac{U_{M^{bus}}}{U_X} = p^{bus} - t^{bus} \frac{U_{T^{bus}}(C^{bus})}{U_X} - w^{bus} \frac{U_{W^{bus}}}{U_X} - a^{bus} \frac{U_{A^{bus}}}{U_X} \quad (35)$$

$$\frac{U_{M^{car}}}{U_X} = \frac{(p^{car} + \tau^{car})}{o^{car}} - t^{car} \frac{U_{T^{car}}}{U_X} \quad (36)$$

La relación entre la utilidad marginal de los tiempos en el vehículo, espera y acceso y la utilidad marginal del bien general (U_X) determina la valoración monetaria de estos tiempos. Estas valoraciones las consideramos constantes, excepto por el efecto del hacinamiento en el valor del tiempo de viaje en bus, y las expresamos como:

$$\rho^{T^{bus}}(C^{bus}) = -\frac{U_{T^{bus}}(C^{bus})}{U_X}; \quad \rho^{W^{bus}} = -\frac{U_{W^{bus}}}{U_X}; \quad \rho^{A^{bus}} = -\frac{U_{A^{bus}}}{U_X}; \quad \rho^{T^{car}} = -\frac{U_{T^{car}}}{U_X} \quad (37)$$

Las anteriores condiciones se pueden expresar entonces como:

$$\frac{U_{M^{bus}}}{U_X} = p^{bus} + \rho^{T^{bus}}(C^{bus})t^{bus} + \rho^{W^{bus}}w^{bus} + \rho^{A^{bus}}a^{bus} \quad (38)$$

$$\frac{U_{M^{car}}}{U_X} = \frac{(p^{car} + \tau^{car})}{o^{car}} + \rho^{T^{car}}t^{car} \quad (39)$$

Estas son las condiciones de primer orden que presentamos en la Sección 2.1. En conjunto con la restricción de ingreso, estas condiciones determinan las funciones de demanda del bien general y de los kilómetros en bus y en carro.

B. Subsidio óptimo

Como se mencionó al final de la Sección 2.4, la agencia de transporte público determina la tarifa de transporte público (p^{bus}) con el objetivo de maximizar la utilidad del agente representativo (ver Ecuación 1 de la Sección 2.1):

$$U(X, M^{bus}, M^{car}, T^{bus}, W^{bus}, A^{bus}, C^{bus}, T^{car}, Z^{pol}) \quad (40)$$

Aunque esta función de utilidad no depende directamente de la tarifa de transporte público, sí depende indirectamente de esta debido a que el nivel de consumo del bien general (X) y de demanda de buses y carros (M^{bus} y M^{car}) dependen de la tarifa a través de las condiciones de primer orden del proceso de maximización del agente representativo (el cual determina las funciones de demanda). Estas condiciones de primer orden son (ver Ecuaciones 3, 4 y 5

de la Sección 2.1 o el Apéndice A):

$$I - Sub = X + p^{bus} M^{bus} + (p^{car} + \tau^{car}) \frac{M^{car}}{o^{car}} \quad (41)$$

$$\frac{U_{M^{bus}}}{U_X} = p^{bus} + \rho^{T^{bus}} (C^{bus}) t^{bus} + \rho^{W^{bus}} w^{bus} + \rho^{A^{bus}} a^{bus} \quad (42)$$

$$\frac{U_{M^{car}}}{U_X} = \frac{p^{car} + \tau^{car}}{o^{car}} + \rho^{T^{car}} t^{car} \quad (43)$$

Adicionalmente, la restricción presupuestal de la agencia de transporte público determina el subsidio (Sub) necesario para cubrir los costos operacionales (CO) (ver Ecuación 16 de la Sección 2.4):

$$Sub + p^{bus} M^{bus} + \tau^{car} V^{car} = CO \quad (44)$$

donde los costos operacionales equivalen a la suma de los costos fijos y variables (ver Ecuación 15 de la Sección 2.4):

$$CO = F + KV^{bus} \quad (45)$$

Podemos utilizar las Ecuaciones 44 y 45 para simplificar la restricción de ingreso (41) y obtener:

$$I = F + KV^{bus} + X + p^{car} \frac{M^{car}}{o^{car}} \quad (46)$$

Esta restricción muestra que el ingreso total debe cubrir los costos operacionales del transporte público, el consumo del bien general y los costos operacionales del carro. Los ingresos públicos por tarifas al transporte público y cargos al uso del carro se cancelan en la anterior restricción debido a que representan transferencias monetarias sin ningún impacto en eficiencia. Sin embargo, la tarifa al transporte público y el cargo al uso del carro sí tienen efecto en las demandas de buses y carros y de consumo del bien general a través de su impacto en las Ecuaciones 42 y 43.

Para encontrar la tarifa óptima, derivamos la función de utilidad (40) con respecto a la tarifa, teniendo en cuenta que X , M^{bus} y M^{car} dependen de la tarifa, y que los tiempos de viaje, nivel de hacinamiento y externalidades (T^{bus} , W^{bus} , A^{bus} , C^{bus} , T^{car} , Z^{pol}) dependen de M^{bus} y M^{car} a través de las Ecuaciones 7 a 14 y 17 (esta última ecuación expresa el número de kilómetros vehiculares de los buses como función del número de kilómetros de pasajeros). Este proceso resulta en la siguiente expresión para el cambio en utilidad con respecto al

cambio en la tarifa:

$$\begin{aligned}
\frac{dU}{dp^{bus}} = & U_X \frac{\partial X}{\partial p^{bus}} + U_{M^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} + U_{M^{car}} \frac{\partial M^{car}}{\partial p^{bus}} \\
& + U_{T^{bus}} \left(t^{bus} \frac{\partial M^{bus}}{\partial p^{bus}} + M^{bus} \frac{dt^{bus}}{dV} \left(\frac{1}{o^{car}} \frac{\partial M^{car}}{\partial p^{bus}} + \alpha \frac{dV^{bus}}{dM^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} \right) + M^{bus} \theta \frac{do^{bus}}{dp^{bus}} \right) \\
& + U_{W^{bus}} \left(w^{bus} \frac{\partial M^{bus}}{\partial p^{bus}} + M^{bus} \left(\frac{dw^{bus}}{df^{bus}} \frac{1}{hD^{bus}} \frac{dV^{bus}}{dM^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} + \frac{dw^{bus}}{dl^{bus}} \frac{1}{n^{bus}} \frac{\partial o^{bus}}{\partial p^{bus}} \right) \right) \\
& + U_{A^{bus}} \left(a^{bus} \frac{\partial M^{bus}}{\partial p^{bus}} \right) + \frac{dU_{T^{bus}}}{dC^{bus}} M^{bus} t^{bus} \frac{dC^{bus}}{dl^{bus}} \frac{1}{n^{bus}} \frac{\partial o^{bus}}{\partial p^{bus}} \\
& + U_{T^{car}} \left(t^{car} \frac{\partial M^{car}}{\partial p^{bus}} + M^{car} \frac{dt^{car}}{dV} \left(\frac{1}{o^{car}} \frac{\partial M^{car}}{\partial p^{bus}} + \alpha \frac{dV^{bus}}{dM^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} \right) \right) \\
& + U_{Z^{pol}} \left(z_{car}^{pol} \frac{1}{o^{car}} \frac{\partial M^{car}}{\partial p^{bus}} + z_{bus}^{pol} \frac{dV^{bus}}{dM^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} \right) \quad (47)
\end{aligned}$$

Esta expresión se puede simplificar mediante los siguientes pasos. En primer lugar, dividimos por U_X ambos lados de la ecuación para expresar el cambio en utilidad en términos monetarios. En segundo lugar, usamos las Ecuaciones 42 y 43 para remplazar varios términos por la tarifa de bus (p^{bus}) y el costo monetario total del carro por kilómetro de pasajero ($(p^{car} + \tau^{car})/o^{car}$). En tercer lugar, la derivada de la Ecuación 46 con respecto a la tarifa lleva a la siguiente igualdad:

$$0 = \frac{\partial X}{\partial p^{bus}} + \frac{p^{car}}{o^{car}} \frac{\partial M^{car}}{p^{bus}} + K \frac{dV^{bus}}{dM^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} \quad (48)$$

Usamos esta igualdad para sacar $\partial X/\partial p^{bus}$ fuera de la Ecuación 47. En cuarto lugar, utilizamos la Ecuación 17 para expresar la derivada del número de kilómetros vehiculares de bus con respecto a la demanda como:

$$\frac{dV^{bus}}{dM^{bus}} = \frac{\varepsilon_V}{o^{bus}} \quad (49)$$

y la derivada de la ocupación promedio de los buses con respecto a la tarifa de bus como:

$$\frac{do^{bus}}{dp^{bus}} = \frac{(1 - \varepsilon_V)}{V^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} \quad (50)$$

En quinto lugar, expresamos algunos términos de la Ecuación 47 en función de las elasticidades del tiempo de espera con respecto a la frecuencia y al factor de carga (η_{wf} y η_{wl} respectivamente) y del nivel de hacinamiento con respecto al factor de carga (η_C), las cuales

se definen como:

$$\eta_{wf} = -\frac{f^{bus}}{w^{bus}} \frac{dw^{bus}}{df^{bus}}; \quad \eta_{wl} = \frac{l^{bus}}{w^{bus}} \frac{dw^{bus}}{dl^{bus}}; \quad \eta_C = \frac{l^{bus}}{C^{bus}} \frac{dC^{bus}}{dl^{bus}} \quad (51)$$

Adicionalmente, denotamos la valoración monetaria de la derivada de la utilidad marginal del tiempo de viaje en bus con respecto al nivel de hacinamiento como:

$$\rho^{C^{bus}} = -\frac{1}{U_X} \frac{dU_{T^{bus}}}{dC^{bus}} \quad (52)$$

Este término determina cómo cambia el valor del tiempo de viaje en bus ante cambios en el nivel de hacinamiento. Finalmente, denotamos la valoración monetaria de cambios en el nivel de polución como:

$$\rho^{Z^{pol}} = -\frac{U_{Z^{pol}}}{U_X} \quad (53)$$

Luego de realizar todas estas simplificaciones, el cambio en utilidad (en términos monetarios) con respecto al cambio en tarifa se expresa como:

$$\begin{aligned} \frac{1}{U_X} \frac{dU}{dp^{bus}} &= p^{bus} \frac{\partial M^{bus}}{\partial p^{bus}} - K \frac{\varepsilon_V}{o^{bus}} \frac{\partial M^{bus}}{\partial p^{bus}} \\ &+ \rho^{W^{bus}} \varepsilon_V w^{bus} \eta_{wf} \frac{\partial M^{bus}}{\partial p^{bus}} - \rho^{W^{bus}} (1 - \varepsilon_V) w^{bus} \eta_{wl} \frac{\partial M^{bus}}{\partial p^{bus}} - \rho^{C^{bus}} (1 - \varepsilon_V) C^{bus} t^{bus} \eta_C \frac{\partial M^{bus}}{\partial p^{bus}} \\ &- \left(\rho^{Z^{pol}} z_{car}^{pol} + \rho^{T^{car}} M^{car} \frac{dt^{car}}{dV} + \rho^{T^{bus}} M^{bus} \frac{dt^{bus}}{dV} - \tau^{car} \right) \frac{1}{o^{car}} \frac{\partial M^{car}}{\partial p^{bus}} \\ &- \left(\left(\rho^{Z^{pol}} z_{bus}^{pol} + \rho^{T^{car}} M^{car} \alpha \frac{dt^{car}}{dV} + \rho^{T^{bus}} M^{bus} \alpha \frac{dt^{bus}}{dV} \right) \frac{\varepsilon_V}{o^{bus}} + \rho^{T^{bus}} (1 - \varepsilon_V) o^{bus} \theta \right) \frac{\partial M^{bus}}{\partial p^{bus}} \end{aligned} \quad (54)$$

Utilizamos ahora las Ecuaciones 19 a 24 para simplificar la expresión anterior y expresamos la derivada de la demanda de bus con respecto a la tarifa de forma positiva ($-\partial M^{bus}/\partial p^{bus}$):

$$\begin{aligned} \frac{1}{U_X} \frac{dU}{dp^{bus}} &= (CM_{oferta} - p^{bus}) \left(-\frac{\partial M^{bus}}{\partial p^{bus}} \right) \\ &- (BM_{espera_frec} - CM_{espera_hacin} - CM_{viaje_hacin}) \left(-\frac{\partial M^{bus}}{\partial p^{bus}} \right) \\ &- CM_{ext}^{car} \left(\frac{\partial M^{car}}{\partial p^{bus}} \right) + CM_{ext}^{bus} \left(-\frac{\partial M^{bus}}{\partial p^{bus}} \right) \end{aligned} \quad (55)$$

Esta última expresión se puede utilizar para evaluar el efecto neto en bienestar de un

cambio en la tarifa actual de transporte público. Si, por ejemplo, la anterior expresión es positiva al nivel de tarifa actual, un aumento en la tarifa tendría un efecto neto positivo en el bienestar general. Si, por el contrario, la anterior expresión es negativa, reducciones en la tarifa aumentarían el bienestar.

Para encontrar la tarifa óptima, igualamos la expresión anterior a cero. Adicionalmente, dividimos por $-\partial M^{bus}/\partial p^{bus}$ y sumamos el costo promedio (CO/M^{bus}) en ambos lados de la ecuación para determinar el nivel de subsidio óptimo:

$$\frac{CO}{M^{bus}} - p^{bus} = \left(\frac{CO}{M^{bus}} - CM_{oferta} \right) + (BM_{espera_frec} - CM_{espera_hacin} - CM_{viaje_hacin}) + (CM_{ext}^{car} m_{bus}^{car} - CM_{ext}^{bus}) \quad (56)$$

Esta última ecuación es la que se presenta al inicio de la Sección 2.5. Es importante remarcar que esta ecuación por si sola no permite estimar el nivel de subsidio óptimo, ya que el lado derecho de la ecuación incluye dos variables endógenas al modelo: la demanda de transporte público (M^{bus}) y la demanda de carros particulares (M^{car}), así como otras variables que dependen de estos niveles de demanda (ej: tiempos de viaje). Esta ecuación se debe utilizar entonces en conjunto con el sistema de demanda (determinado por la Ecuaciones 41, 42 y 43) para obtener estimaciones de la tarifa o el nivel de subsidio óptimo. Este proceso se simplifica significativamente si se toman como constantes algunos efectos marginales en la Ecuación 56, como las elasticidades del tiempo de espera (η_{wf} y η_{wl} en las Ecuaciones 20 y 21) y las valoraciones de los tiempos de viaje. Adicionalmente, el sistema de demanda se puede simplificar considerando que la demanda responde al costo generalizado de viaje con una elasticidad constante. Sin embargo, es importante considerar al hacer estas simplificaciones que estos efectos marginales generalmente son medidos empíricamente para las condiciones actuales de tarifa y demanda, por lo que pueden variar significativamente bajo otras condiciones. En consecuencia, la confiabilidad del nivel óptimo de subsidio o tarifa estimado se reduce en cuanto este nivel se encuentre más lejos del nivel actual de subsidio o tarifa.

Para resolver numéricamente la Ecuación 56 en conjunto con la función de demanda (elasticidad constante al costo generalizado), programamos ambas ecuaciones en una hoja de Excel y encontramos la tarifa óptima por ensayo y error. En todos los casos, este proceso requirió solo algunas iteraciones.

C. Elasticidad de la demanda al costo generalizado

La relación entre la elasticidad de la demanda al costo generalizado y su elasticidad a la tarifa se puede analizar de la siguiente forma. Denotamos la demanda M , la tarifa p y el costo generalizado q . Las elasticidades al costo generalizado (ε_q) y a la tarifa (ε_p) se definen como:

$$\varepsilon_q = \frac{q}{M} \frac{dM}{dq}; \quad \varepsilon_p = \frac{p}{M} \frac{dM}{dp}$$

Las derivadas de la demanda a la tarifa y al costo generalizado se relacionan de la siguiente forma:

$$\frac{dM}{dp} = \frac{dM}{dq} \frac{dq}{dp}$$

La elasticidad al costo generalizado se puede expresar entonces como:

$$\varepsilon_q = \frac{\varepsilon_p}{\frac{p}{q} \frac{dq}{dp}}$$

La derivada del costo generalizado a la tarifa (dq/dp) incluye el efecto directo de la tarifa como parte del costo generalizado ($\partial q/\partial p = 1$), así como el efecto indirecto de la tarifa en los tiempos de viaje a través de su efecto en los niveles de demanda y oferta del sistema de transporte público. Este efecto indirecto puede ser positivo o negativo dependiendo de cómo los niveles de oferta y demanda afectan los tiempos de viaje, así como de la forma cómo la agencia de transporte público ajusta la oferta ante cambios en la demanda.

Guzman y cols. (2021) identificaron la elasticidad de la demanda de TransMilenio a la tarifa controlando por el nivel de oferta (utilizando el pago mensual de TransMilenio S.A. a los operadores como indicador de la oferta). Esto implica que el efecto indirecto de la tarifa en los tiempos de viaje fue parcialmente controlado. Para los resultados del escenario base, estimamos entonces la elasticidad al costo generalizado simplemente dividiendo la elasticidad a la tarifa entre la proporción del costo generalizado que representa la tarifa (p/q). Esta aproximación puede subestimar la elasticidad de la demanda al costo generalizado en la medida en que la demanda aumenta los tiempos de viaje, pues en este caso $dq/dp < 1$. Los análisis de sensibilidad muestran cómo varían los resultados ante cambios en la elasticidad de la demanda.