# Technology choices in public transport planning: a classification framework

Paul Basnak[1*], Ricardo Giesen[1], Juan Carlos Muñoz[1].

[1]Departamento de Ingeniería de Transporte y Logística
Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna 4860, Macul, Santiago, Chile. +56(2) 2354 2000

## Author Contribution

Paul Basnak: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing – Original Draft, Visualization

Ricardo Giesen: Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition

Juan Carlos Muñoz: Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition

## Conflict of Interest:

The authors declare no conflict of interest

## Funding sources:

Scholarship for the corresponding author was provided by Cedeus (Centro de Desarrollo Urbano Sustentable, Santiago, Chile) and BRT Centre of Excellence (Pontificia Universidad Católica de Chile, Santiago, Chile).

# Technology choices in public transport planning: a classification framework


## Abstract

Choice of public transport technologies in cities is not straightforward: while the academy focuses on optimization models to determine which modes should a specific city have, policy makers rely on simple recommendations which are based on city population and income. We estimated six types of classification models that could allow for more precise recommendations yet are simple enough to be applied by the authorities. We considered typical variables as population and Gross Domestic Product of cities but also geographic and morphologic characteristics in a database of 400 cities from North and South America. Ordered Probit and Multinomial Logit models were the most accurate, with a success rate over 80% in the validation subset. Among the explanatory variables, city population and GDP per capita were as expected the most significant, but fare integration, car ownership and city shape were also relevant. Even if existent public transport modes in cities are not necessarily optimal, the classification models developed can give an insight for policy makers, in the sense that cities whose public transportation complexity cannot be explained by the models are more likely to have a suboptimal public transportation system.



*Keywords: public transport, transport modes, cities, classification models, ordered Probit*

*JEL codes: R40, R41, R58,*


## 1. Introduction

There is a growing gap between lecturers and decision makers concerning which public transport modes to recommend in cities. While the academic literature focuses on optimization models that incorporate the specific behavior of people, stakeholders taking strategic decisions usually rely on general recommendations (Chen, 2017) which are based on limited case studies, such as the Bus Rapid Transit guide (TRB, 2003) that recommends this technology for cities over 750,000. These recommendations consider the public transport modes that cities currently have but are usually poor in explanatory variables and use small datasets. That said, we propose to correlate the existent public transport modes with the characteristics of diverse cities through classification models. This will be useful to provide more specific recommendations to policy makers on the public transportation technologies to consider in their cities.

### 1.1. Problem and research gap

In this section, we will present the relevant literature on studies that recommend public transport modes based on the existent systems, focusing on the research gaps, such as ignored variables and limited datasets.

The recommendation of public transport modes in cities is usually divided in two groups of studies:

1) Mathematical models that compare costs of different transport options in a given network or line. Most of these models are focused on tactical parameters, such as the optimal frequency in a transit line (Mohring, 1972), optimal vehicle size (Oldfield & Bly, 1988) or optimal distance between stops (Medina et al., 2013). However, other studies are focused on a strategic analysis, which compares total costs in a network or line for different technologies, usually high-capacity modes as Metro, Light Rail Transit (LRT) or Bus Rapid Transit (BRT). Their results are usually expressed in demand thresholds, i.e. minimum levels that justify the implementation of a given mode.

2) General recommendations based on the performance of existent public transport systems in selected cities. These recommendations are usually very simple, setting minimum populations (thresholds) for cities to consider a given public transport mode as BRT, LRT or Metro, such as the TCRP 90 (Transportation Research Board, 2013) guideline.

**Table 1** below summarizes the minimum standards to justify different transit modes according to various studies. The first part of the table refers to optimization models corresponding to the first group of studies described above. The second part of the table refers to operational analysis and other general recommendations. It is worthwhile noting that recommendations might be expressed in different units, and therefore may not be comparable.

**Table 1 - Demand thresholds that justify the construction of public transport modes in optimization models & other data sources**

| ***Optimization models*** | | | | | |
|---|---|---|---|---|---|
| **Authors** | **Network type** | **Minimum thresholds to justify transit modes** | | | |
| | | **Unit** | **BRT** | **LRT** | **Metro** |
| *Moccia & Laporte (2016)* | *Single line* | *Users/hour (both directions, 70%-30% balance)* | *4,000* | *13,000* | *20,000* |
| *Tirachini et al. (2010)* | *Circle with radial lines* | *Users/day (total network demand)* | | | *3,200,000* |
| *Daganzo (2010)* | *Grid with trunk & feeder lines* | *Users/hour (total network demand)* | | *None (BRT has lower costs in all scenarios)* | |
| ***Other data sources*** | | | | | |
| **Authors & context** | | **Minimum thresholds to justify transit modes** | | | |
| | | **Unit** | **BRT** | **LRT** | **Metro** |
| *Pushkarev & Zupan (1980) – cost analysis in US cities* | | *Users/peak hour/direction (revised by Kain (1988))* | | *8,000 (level) to 27,000 (grade separated)* | *30,000 (above ground) to 58,000 (tunnel)* |
| *Hidalgo (2005) – cost analysis for a transit line in Colombia* | | *Users/peak hour/direction* | | | *40,000* |
| *Verma & Dhingra (2001) – operational analysis in Indian cities* | | *Users/peak hour/direction* | | *12,000* | *50,000* |
| *AECOM (2012) – segregated busway requirement in Canberra (Australia)* | | *Vehicles/hour* | *75* | | |
| *LAMTA (2012) – bus lanes requirement in Los Angeles (USA)* | | *Users/peak hour/direction* | *1,000* | | |

Cells in grey indicate no data available

Most of the relevant literature refers to high capacity modes, such as Metro, LRT or BRT. Wide differences are observed among different data sources, depending on the methodology and the national/regional context, which makes difficult to set generalized recommendations. Even in the most rigorous mathematical models, demand thresholds that define the superiority of implementing a given technology are highly sensitive to factors such as construction costs and users' value of time.

A simpler method, based on the characteristics of cities, is set in public transport planning guides, consultancy reports or handbooks. Most of these recommendations consist of suggesting the minimum population that a city should have to adopt a specific technology, particularly for high capacity modes such as BRT, LRT and Metro. These recommendations, summarized in **Table 2**, show lower limits of between 0.25 and 1 million inhabitants to build BRT, between 0.3 and 3 million for LRT and between 1 and 6 million for Metro.

In practice, national policy frameworks that sets criteria for developing transit systems in cities based on city characteristics are also quite simple:

- Chen (2017) reports that the Chinese government relies on 3 criteria ("population scale", "transport demand" and "economic development level") to fund the construction of new Metro and Light Rail systems. Chinese cities should require a minimum population of 3 million to be eligible for Metro funding, or a minimum of 1.5 million to apply for Light Rail.
- According to White (1979), planners in the Soviet Union considered "indispensable" the construction of Metro systems in cities over one million inhabitants. Several cities of the USSR built Metro lines between the 1960s and 1980s after exceeding that population, in every case with national State funds.
- Colombia has, as reported by Jiménez (2017) a national urban transport policy, in which bus systems that include BRT lines are planned by the national government – subject to specific demand studies – for cities over 600,000.

A more comprehensive analysis to set a reference for the construction of metro lines was performed by Loo & Cheng (2010). The authors gathered the main characteristics of 57 cities in 20 countries at the time they built their first metro line. The population (5 million) and per capita GDP averages (11,400 US dollars) are proposed as "useful benchmarks". Loo & Cheng (2010) identified missing factors in these benchmarks such as environmental, political and technological considerations, as well as the existence of "wide regional variations" between cities in Europe – with an average of 2.9 million inhabitants when building their first Metro lines – and cities of poorer continents which required 8 million (Africa) and 5.7 million (Asia). An opportunity to study the influence of the income and other variables on the thresholds arises from these findings.

An intermediate method between optimization models and study cases was explored by Verma & Dhingra (2001). Based on an operational analysis for Indian cities, they found that metro is suitable over 50,000 passengers/hour/direction (pphpd), while LRT is recommended over 12,000 pphpd. The authors estimated transit demand for three types of cities (linear, semi-circular and circular) with 3 activity structures (mononuclear, polynuclear uniform and polynuclear non-uniform) through simulation models. By comparing estimated transit ridership with the demand thresholds, the recommendation of public transport modes depends on city population, shape and activity structure: for instance, Metro is recommended for every linear city above 3 million, but for circular cities above 6 million only.

**Table 2** below summarizes city population thresholds recommended in diverse studies:

**Table 2 - City population thresholds recommended for new public transport modes**

| Authors | Population thresholds (million) | | | | Comments |
|---|---|---|---|---|---|
| | bus | BRT | LRT | Metro | |
| *Chen (2017)* | | | *1.5* | *3* | *Report on Chinese Govt. criteria* |
| *Lehner (1975)* | | | *0.3-0.8* | *1* | *Based on western European cities* |
| *Verma & Dhingra (2001)* | | | *1.5 - 3* | *3 - 6* | *Dependent on city shape + activity structure (Indian cities)* |
| *Transportation Research Board – TCRP 90 (2003)* | | *0.75* | | | *Guidelines for US and Canadian cities* |
| *Levinson et al. (1975)* | *0.025 – 0.05* | *0.25 – 0.3 ("min"), 1 ("greatest potential")* | | | ** referred to as "express bus" on freeways + busways* |
| *Loo & Cheng (2010)* | | | | *5* | *"Wide regional variations" observed* |

Cells in grey indicate no data available

As we have seen, most of the studies base their recommendations on the population of cities, while some authors (Verma & Dhingra, 2001) also take into account city structure and others (Loo & Cheng, 2010) warn about "regional variations" related with GDP. Establishing simple criteria based on known variables for local authorities is undoubtedly attractive. However, a large difference is observed in the population thresholds recommended by different authors that use data from different countries, which suggests that the variability between regions of different characteristics should be considered.

Babalik-Sutcliffe (2002) performed an *ex-post* analysis of existing urban rail systems based on performance measures such as operating costs per passenger, impact on mode share, and the comparison between existing and forecast demand to derive what factors define the success or failure of a given system. The author uses data from eight cities of similar income and population located in the United States, Canada and Europe, which built Metro or LRT lines between 1980 and 1995. Among the factors that identify the success or failure of a given project, Babalik-Sutcliffe (2002) identifies the urban form, population density, predominance of the CBD (Central Business District) and fare integration. These variables can be added to the population of cities in order to provide more precise recommendations.

It is also known that other variables are relevant in travel patterns and particularly in public transport use, such as the road network (Crane, 2000) and car ownership (Paulley et al., 2006).

We propose to estimate classification models that allow to correlate the main characteristics of a given city with the modes of public transport that the city has, incorporating the relevant variables mentioned in previous paragraphs. As Babalik-Sutcliffe (2002) explains, the presence of a given public transport mode in a city does not necessarily imply that it should exist: in any case, classification models based on a large sample of cities will allow to set more precise recommendations than those provided by the planning manuals. Next, we will explain the criteria used to classify cities by their urban public transport system.

## 1.2 How to describe a public transport system? A 5-tier classification

Most of the relevant literature is focused on the drivers that trigger high-capacity transport mode investments, such as Metro or BRT. Large urban areas are therefore the bulk of these studies, while small and medium size cities are usually ignored. In this work, we will characterize the public transport system of a given city by the technology that provides more capacity, by proposing a 5-tier scale based on the right-of-way classification proposed by Vuchic (2005).

Vuchic (2005) divides public transportation modes in 3 classes, according to the right-of-way (ROW) of their alignment:

- ROW “C”, corresponding to mixed traffic conditions. This includes standard bus services, trolleybuses and trams, with a typical average commercial speed below 20 km/h and maximum capacity below 7,000 pphpd (Deen & Pratt, 1992).

- ROW "B", which corresponds to a partially separated right-of-way, in which public transport operates on segregated lanes but with level crossings, such as BRT and LRT lines, with commercial speeds of 20 to 40 km/h (O'Flaherty et al., 1997) and usual maximum capacities between 5,000 and 30,000 pphpd.

- ROW "A", consisting of a full segregated right-of-way, which may be present in metro lines, regional trains or high capacity BRT corridors with dedicated infrastructure and separate-grade intersections. Typical maximum capacities for these systems range between 10,000 and 70,000 pphpd, with average speeds above 25 km/h (Vuchic, 2005).

Based on this classification, we propose to divide the urban public transport systems into five classes. The three superior categories are associated with the highest capacity mode according to Vuchic (2005), while the two remaining classes represent towns without public transport or with demand-responsive transit only.

- Type I systems, where mobility is carried out only in private modes (such as cars, motorcycles, non-motorized transport).

- Type II systems, in which public transport is carried out exclusively through demand-responsive services such as taxis or minivans. Point deviation systems - vehicles that respond to orders within an area but must pass through some sites on a mandatory basis (Potts et al., 2010) - are also included in this category.

- Type III systems, in which the mode of greatest capacity corresponds to ROW "C", regardless of the capacity of the runners. Examples of these systems include cities whose public transport is provided through conventional buses such as Oxford (England), trams such as Belgrade (Serbia) or trolleybuses such as Salzburg (Austria).

- Type IV systems, in which the highest capacity mode corresponds to ROW “B”, usually in the form of BRT - as in Quito (Ecuador) - or LRT lines as in Porto (Portugal). Although the concept of BRT implies diverse standards of buses, corridors and infrastructure, a minimum standard of 3 kilometers of exclusive track with specific design for buses (ITDP, 2014) was considered.

- Type V systems, in which the highest capacity mode corresponds to ROW “A”, which corresponds to Metro systems such as in Moscow (Russia) or urban heavy rail lines (HRT) such as Johannesburg (South Africa). Some high-capacity full segregated BRT corridors, such as Troncal Sur in Bogotá (Hidalgo, 2013) also fit in this class.

The modal offer is usually cumulative: given that almost every city having a given public transport mode also has technologies that belong to inferior categories[1], the proposed classes can be considered as an accurate representation of the complexity of public transport systems, covering both small towns and large metropolitan areas.

Characteristics of the proposed classification are presented in **Table 3** below:

**Table 3 - Proposed classification for public transport systems in cities**

| Class | Highest capacity public transport modes | Typical ROW | Typical max capacity (pphpd) |
|---|---|---|---|
| *I* | *(no public transport, private modes only)* | | |
| *II* | *Demand-responsive (taxi)* | *"C" (mixed traffic)* | - |
| *III* | *Bus, tram, trolleybus* | *"C" (mixed traffic)* | *0 – 7,000* |
| *IV* | *BRT, LRT* | *"B" (partial priority)* | *5,000 – 30,000* |
| *V* | *Metro, HRT, BRT (high standard corridors)* | *"A" (full priority)* | *10,000 – 70,000* |

In this work, we propose to determine which factors explain the public transport modes chosen by each city using a dataset from 400 cities located in South and North America through classification models. The dependent variable will be represented by the public transport mode that offers most capacity in passengers/h using a I-V scale, in which "I" represents cities without public transport, "II" corresponds to cities that only have demand-responsive modes such as taxis, while "III", "IV" and "V" correspond to cities in which the highest capacity modes are normal buses or trams (III), light rail (LRT) or Bus Rapid Transit (BRT) lines (IV), and heavy rail/Metro lines (V). It is noteworthy that almost every city belonging to a given category has modes that belong to every inferior class, so that the scale proposed represents accurately the complexity of a public transport system.

## 2. Drivers explaining public transport modes in cities

Here we explain which types of variables influence on the public transport modes that different cities have. As many authors have stated, it is expected that more populated cities have more complex public transport systems, usually with Metro or BRT lines. Other factors, such as GDP and car ownership, may also have a significant impact on trip patterns in a given city, thus affecting public transport modes: as car ownership grows, the incentive to have an efficient public transport may decline, while a higher GDP may allow for more investments allocated to dedicated infrastructure. City shape and interaction between nearby cities can also be relevant: a linear city should imply longer trip distances which may justify more efficient modes compared with a circular city of similar area, and trip attraction produced by adjacent larger cities might have a similar effect. We will represent city shape through two indicators (compactness and attractiveness) that are defined in this chapter. Finally, road network and slopes may also affect the appeal of public transport compared with other modes.

### 2.1. Population, urban area & density

**Population** of cities is, as described above, the variable most frequently used to recommend public transport modes in cities. A larger population is correlated with higher public transport

---

[1] *In the sample analyzed, more than 95% of the cities belonging to a given class also have technologies associated with the lower categories.*

ridership (Taylor & Fink, 2003), which may justify public transport systems with higher capacity modes.

A larger **urban area** usually involves longer trips: by increasing the travel distance, faster technologies allow greater time savings, so that these modes can be more beneficial. On the other hand, modes with the highest average speed are those corresponding to ROW "A" and ROW "B" (Vuchic, 2005). In this sense, larger cities would be expected to have public transport systems of higher categories.

However, when considering the effect of population on the surface of cities - that is, when using **density** as independent variable - predicting the expected effect is not straightforward. On one hand, a greater density implies a smaller surface in cities with a given population, which is associated with shorter trip distances and therefore can discourage the construction of faster modes. On the other hand, a lower population density implies a lower concentration of trips, and therefore a lower efficiency of high capacity modes such as Metro and BRT, which need a minimum amount of trips to be cost-effective (see **Table 1**). In general, a higher urban population density is considered a favorable condition for the implementation of high capacity public transport modes (Babalik-Sutcliffe, 2002).

## 2.2. Socioeconomic factors

**Income** of users is one of the main variables that explain mode choice in transport (Train & McFadden, 1978). Since value of time tends to increase with income (Jara-Diaz, 2000) it is expected that more expensive but faster modes such as Metro are prioritized in wealthier cities over slower and cheaper technologies such as buses. Given that in many countries there is scarce information about average income of residents in their cities, national and regional *per capita* GDP were used as proxies for income.

Although greater **car ownership** is usually correlated with less use of public transport (Kitamura, 2009), cities that have similar car ownership may have significant differences in public transport trip share (McIntosh et al., 2014). That said, the presence of higher user costs in public transport is also correlated with the purchase of private vehicles (Dargay, 2002). Therefore, car ownership may be an endogenous variable for public transport classification models, because of simultaneity with the dependent variable (Train, 1980; Kitamura, 2009).

For this reason, **fuel prices** were also considered as a proxy for car ownership in our models. Fuel prices are correlated with the use of public transport (Currie & Phung, 2007) and with car purchases (Dargay, 2002), while public transport use does not have an apparent influence on fuel prices. When fuel prices increase, the incentive to acquire or use private cars diminishes, which should result in a greater demand for public transport and therefore in the use of more efficient modes in a city.

**Other economic and political variables** that were not analyzed in this work should be taken into account in further studies, such as parking costs (Taylor & Fink, 2003), financial costs for infrastructure project – which is especially relevant in the construction of metro lines – and the degree of political centralization in the funding of public transport projects.

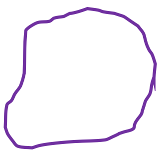
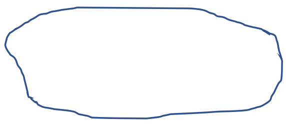
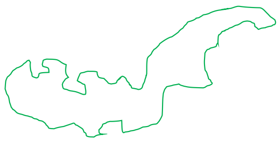
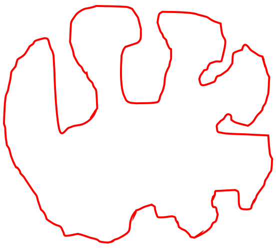
## 2.3. Geographic and morphologic variables

Shape of cities can be studied under different approaches, among which graph theory (Fielbaum et al., 2017) and spatial shape metrics (Huang et al., 2007) are widely known. In this work, we characterized city shape by simple indicators that are available for any urban area. That said, city shape is represented with 3 generalized indicators:

**Compactness**: this standardized shape factor, whose formula is $\frac{2\sqrt{\pi \, . \, area}}{perimeter}$ (Jiang, 2007), measures in practical terms how similar a shape is to a circle. The indicator takes values between 0 and 1, and the smaller the compactness the greater the average distance for areas of equal surface. More compact cities (where the urban form favors shorter travel distances and a greater dispersion of origins and destinations) are expected to have less high capacity corridors in their public transport networks, while in less compact cities trips tend to be longer and with greater spatial concentration, which favors the adoption of more efficient modes in terms of capacity and speed such as Metro or BRT.

**Slenderness**: this indicator is a proxy of the ratio between major and minor semi axes of a given shape, through the expression $\frac{S_C}{S_F}$, where Sf is the surface of the city to be analyzed and Sc is the area of the minimum envolving circle (Baker & Cai, 1992). The effect of this indicator should be alike that of compactness: in general, more slender cities tend to be less compact, so that a growing slenderness should be related to the greater complexity of transport systems, as well as other factors. However, there are some differences between both indicators: as illustrated below, there are different possible combinations of slenderness and compactness.

**Table 4 – Slenderness and compactness of different shapes**

| **Low** slenderness<br>**High** compactness | **Low** slenderness<br>**High** compactness | **High** slenderness<br>**Low** compactness | **Low** slenderness<br>**Low** compactness |
|---|---|---|---|
|  |  |  |  |

**Attractiveness**: this indicator, whose mathematical expression for a city "i", based on Schneider (1959), is $A_i = max_j \left[\frac{Population_j}{(dist_{ij})^n}\right]$ where "n" is a positive number to calibrate, allows to identify the proximity of a city with other localities "j". It is presumed that cities that interact more with their neighbors (those with greater attractiveness) can have a more complex public transport system than similar isolated cities, since they have greater probability of having a joint transportation system.

**Street provision**, measured in km of street per square km, and **network density**, measured in intersections per square km, may also be relevant. These variables should have a similar effect to road length per capita, whose increase favors car use and discourages public transport share (McIntosh et al., 2014).

**Topography** of cities can also be relevant (Taylor & Fink, 2003). A mountainous topography can decrease the average speed of at-grade transport, both private and public, which may favor the implementation of technologies such as cable cars, used in cities such as La Paz (Bolivia), Medellín (Colombia) and Rio de Janeiro (Brazil). For the models, average and median slopes in city networks were estimated through a Google Maps® Api.

### 2.4. Public transport planning variables

**Fare integration** is important when analyzing the demand of a particular mode within a public transport system. Not only has integration been identified as one of the factors for the success of a public transport project (Babalik-Sutcliffe, 2002), but it has also been verified that the introduction of an integrated fare can double the demand of a previously existing Metro network (Muñoz, Batarce & Hidalgo, 2014). Given that it is reasonable for users to make more transfers to faster modes but with less coverage when transfer costs are lower, the presence of fare integration should favour the implementation of more efficient transport technologies.

In the models, the variable "fare integration" was defined as binary, where 1 corresponds to transfers for free or at reduced cost between the highest category mode and other modes in a given city, or among the different public transport routes in the case of type III cities. Meanwhile, 0 is assigned to cities that do not comply with this requirement, including those where transfers at reduced cost are limited to a particular station or service.

**Network integration**, understood as the existence of a spatially planned public transport network (Givoni & Banister, 2010), may also contribute to higher demand for high capacity public transport modes as experienced in the trunk-feeder Santiago de Chile network (Muñoz, Batarce & Hidalgo, 2014). Alike the fare integration variable, network integration was considered as binary, in which 1 is assigned to cities in which lower capacity mode routes are (re)designed as feeders for high capacity corridors.

It should be noted that the dummy specification for the integration variables may introduce simultaneity with the dependent variable, as only cities that have mass public transport modes (i.e. those belonging to classes III, IV & V) can have fare or network integration. However, proper instruments for such (qualitative) variables are yet to be found.

## 3. Model

This chapter is divided in three sections: first, we will explain the six types of model used (multinomial Logit, nested Logit, Linear Discriminant Analysis, Quadratic Discriminant Analysis, ordered Logit & ordered Probit) with their advantages and disadvantages. We will then characterize the data gathering process, focusing on the specific advantages, issues and limitations that arise from relying on diverse sources, and we will describe the dataset, which consists of 400 cities of 22 countries in North and South America, including all capital cities and every city with Metro, LRT and BRT systems. Finally, we will present the modeling process and analyze the results, emphasizing on the best functions for every type of model and comparing the pros and cons of each.

### 3.1. Methodology

The dependent variable to analyze is discrete in five categories. These may be considered as ordered classes, as new public transport modes with higher capacity are added in superior categories: in this sense, the classification represents an increasing complexity of public transport systems. Meanwhile, most of the explanatory variables are quantitative and continuous.

Models that relate the basic characteristics of cities with their public transport systems were proposed by Saidi (2016), who estimated the probability that a given city has circular Metro lines through a Multinomial Logit model. The model considered 94 cities, 13 of them with circular Metro lines, with four explanatory variables: city area, population, length of the Metro network and the age of the system, while GDP per capita was discarded for not being significant.

In addition to Multinomial Logit models, other functional forms such as ordered functions and discriminant analysis functions allow representing discrete variables which depend on continuous variables (James et al., 2014). In this paper, we estimated six types of models that can be grouped into three categories, which are briefly explained below:

- ***Logistic regression models (Multinomial and Nested Logit)***

Logistic regression models, typically used in disaggregated models such as mode choice models, are also useful to represent other phenomena where the dependent variable takes discrete values. The main advantage of the Logit models is their flexibility, allowing different specifications for the functions from which the probability of each category is estimated. In particular, nested models assume a correlation between some of the categories (Ortúzar & Willumsen, 2011), which can be adapted to decision-making: for example, given certain conditions for a city that only has bus lines, its authorities may decide to build a mass transit system with BRT (type IV) or Metro (type V) lines.

- ***Ordered models (ordered Logit & Probit)***

Ordered Logit and Probit models consist of comparing the value of a linear function of the independent variables with thresholds whose values, like the coefficients of the linear function, are determined by maximum likelihood. The dependent variable represents quantitative or qualitative categories that follow a logical order; in the Probit model, errors are distributed according to a normal distribution, whereas they do so according to a Gumbel distribution in the Logit model (Train, 2009). Since the linear function is unique for all categories, these models are easier to interpret than Multinomial and Nested Logit models, although they have less flexibility than these. As the dependent variable should increase as cities grow in population and income, these models could be adequate to represent the public transport systems of cities.

- ***Discriminant Analysis models (LDA, QDA)***

Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are Bayesian methods that assume normal distribution of the variables in the sample (James et al., 2014). Despite being more rigid than Logit models since they use the same variables in all categories, they allow to represent in a simple way the boundary between the different classes, which for two variables is reduced to a line in the case of the LDA and a parabola for the QDA.

Analysis in several simulated samples (Pohar et al., 2004) have yielded similar results for multinomial Logit models and LDA with more than two classes, even if the variables do not adopt normal distributions, so that none of these methods can be discarded *a priori*.

## 3.2. Data Collection

Demographic, socioeconomic and morphological explanatory variables were obtained or calculated from 400 cities in North and South America, including all those with Type IV or V systems. Study area was selected as to consider different cities in population density, income and car ownership with available data.[2] Location of the cities is shown in the map below:

---

[2] *Reliable statistic information on other cities from developing countries, including many cities from Africa and Asia,*

**Figure 1 – Location of cities in database**


Polygons were obtained in Google Earth® as geographic data .kml files for the calculation of the surface of the cities and the shape factors - compactness and slenderness - considering that the area of the cities is limited to a continuous urban surface. Polygons were then analyzed in QGIS3® for the calculation of these variables.

to display geographic data

Calculation of distances between cities for estimating attractiveness was made through Google Maps®. Estimation of slopes (average, median, standard deviation) and network density in cities was performed in Python® with a Google® Api that extracts intersection coordinates (X,Y,Z) from Google Maps®.

On the other hand, the population of the cities, as well as the GDP per capita and the price of fuels, were obtained from official sources in each country, such as population censuses and Domestic Product estimates.

Once the data were obtained, the Logit models were estimated in Biogeme® (Bierlaire, 2003), and the rest in Stata 12® (StataCorp, 2011). The results are shown below.

---

*is difficult to obtain. That said, a broader model that covers cities from other continents would be welcome.*

## 3.3. Results

Descriptive statistics for the main variables are presented in Table 5. As shown, information was gathered for cities of diverse population, area, shape and wealth. Type III cities are the most common class, representing 36% of the sample.

**Table 5 – Database descriptive statistics**

| Variable | Definition | Mean | Min | Max |
|---|---|---|---|---|
| *Class (dependent)* | *5-tier classification of PT systems* | - | *I* | *V* |
| Population | Inhabitants | 977,692 | 300 | 20,850,000 |
| Area | City surface (km$^2$) | 295 | 0.15 | 6,108 |
| National GDP | National per capita GDP (US $) | 20,775 | 2,222 | 59,532 |
| Regional GDP | Subnational (state/province/region) per capita GDP (US $) | 21,275 | 2,169 | 104,893 |
| Density | Population density (inhabitants/km$^2$) | 5,147 | 524 | 23,522 |
| Attractiveness | Gravitational indicator representing inter-city connectivity | 237 | 0 | 4,148 |
| Compactness | Standardized (0-1) shape factor | 0.482 | 0.161 | 0.863 |
| Slenderness | Min. envolving circle/city ratio | 2.976 | 1.041 | 32.346 |
| Fuel price | Unleaded fuel price, national average (US$/l) | 0.99 | 0.01 | 1.70 |
| Street provision | Km of street / km$^2$ | 10.6 | 1 | 26.4 |
| Interchange density | Number of junctions / km$^2$ | 63.5 | 0 | 218.1 |
| Car ownership | Cars/1000 people, nationwide (city specific if data available) | 324 | 38 | 809 |
| Average slope | Average network slope (%) | 2.27 | 0.26 | 10.82 |
| Median slope | Median network slope (%) | 1.66 | 0.13 | 11.00 |
| Fare integration | (dummy) | 0.20 | 0 | 1 |
| Network integration | (dummy) | 0.11 | 0 | 1 |

We estimated more than 200 models corresponding to six function types: multinomial and nested Logit models with separation between classes I-II (node without mass public transport) and III-IV-V (node with public transport), single-function ordered Logit and Probit models, and LDA and QDA models.

First, the sample was randomly divided into a set of training data (75% of the dataset) to estimate the coefficients, and a test set (25%) used to compare the errors associated with each model. Excepting the nested Logit models that did not show significance in the separation between categories, all model types have a comparable adjustment to data.

According to the First Choice Recovery (FC), an error criterion representing the proportion of the cities in which the models assign the highest probability to their actual classification, QDA and MNL are the best models. However, QDA is more difficult to interpret and does not ensure that the coefficients have the proper sign for prediction. Other error indicators, such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are quite similar between multinomial and ordered models.

**Table 6 – Error indicators, all model types (test subset for preliminary model selection)**

| model | variables | FC | FC (III-V) | MAE | RMSE |
|---|---|---|---|---|---|
| **MNL** | **Population (log), per capita GDP (region), density (sqrt), attractiveness (log), compactness, fare integration, car ownership, avg. slopes** | **83.0%** | **82.1%** | **27.80%** | **37.13%** |
| LDA | Population (log), per capita GDP (region), density (sqrt), attractiveness (log), compactness, car ownership, avg. slopes | 77.0% | 82.1% | | |
| QDA | Population (log), per capita GDP (region), density (sqrt), attractiveness (log), compactness, car ownership, avg. slopes | 81.0% | 83.2% | | |
| OL | Population (log), per capita GDP (region), attractiveness (log), compactness, fare integration, car ownership, avg. Slopes, density (interacts with population) | 81.0% | 78.6% | 29.44% | 37.80% |
| **OP** | **Population (log), per capita GDP (region), attractiveness (log), compactness, fare integration, car ownership, avg. Slopes, density (*nteracts with population*)** | **81.0%** | **78.6%** | **29.36%** | **37.33%** |

Cells in grey indicate no data available

Given that the models used are intended to be applied by decision makers, simplicity is a key factor: thus, Multinomial Logit and ordered Probit were selected as the most appropriate models. Both assign the highest probability to the actual classification of cities in more than 80% of the sample.

In order to have a more efficient estimation of coefficients for both model types, 5-fold cross-validation with random quintiles was performed. Variable and error parameters for both models are presented in Table 7:

**Table 7 – General and variable parameters of selected models**

**Multinomial Logit (MNL)**

| | |
|---|---|
| *Number of variables* | *19* |
| *Number of significant variables (T-test> 95%)* | *15* |
| *Log likelihood* | *-161.33* |
| *Rho-squared* | *0.687* |
| *Rho-squared (adjusted)* | *0.650* |
| *Cross-Validation First Choice Error (%)* | *16.25* |
| *C-V Mean Average Error (%)* | *28.15* |
| *C-V Root Mean Square Error (%)* | *37.18* |

| ***Coefficient*** | ***Value*** | ***T-test*** | ***Significance*** |
|---|---|---|---|
| ASC_2 | *-14.4* | *-4.14* | *>99%* |
| ASC_3 | *-44* | *-6.73* | *>99%* |
| ASC_4 | *-78.2* | *-8.2* | *>99%* |
| ASC_5 | *-101* | *-9.07* | *>99%* |
| βPop2 (Ln) | *4.51* | *4.6* | *>99%* |
| βPop3 (Ln) | *11* | *7.55* | *>99%* |
| βPop4 (Ln) | *16.3* | *9.08* | *>99%* |
| βPop5 (Ln) | *20* | *9.96* | *>99%* |
| βGDP3 (Sqrt) | *0.0264* | *2.8* | *>99%* |
| βGDP4-5 (Sqrt) | *0.029* | *2.56* | *99%* |
| βAttract3(ln) | *0.475* | *1.09* | *83%* |
| βAttract4-5(ln) | *1.09* | *2.2* | *97%* |
| βCompac3 | *-4.41* | *-1.92* | *94%* |
| βCompac4 | *-9.37* | *-2.4* | *98%* |
| βCompac5 | *-13.3* | *-2.57* | *99%* |
| βDens2-3 (Sqrt) | -0.035 | -2.86 | >99% |
| βCarOwn1-2 | *$4.4.10^{-3}$* | *1.85* | *94%* |
| βSlope4-5 | *0.249* | *1.11* | *74%* |
| βFare_int3-5 | *1.61* | *2.47* | *99%* |

**Ordered Probit (OP)**

| | |
|---|---|
| *Number of variables* | *8* |
| *Number of significant variables (T-test> 95%)* | *6* |
| *Log likelihood* | *-154.34* |
| *Pseudo R2 (Mc Fadden)* | *0.643* |
| *Pseudo R2 adjusted (Mc Fadden)* | *0.615* |
| *Cross-Validation First Choice Error (%)* | *20.00* |
| *C-V Mean Average Error (%)* | *28.72* |
| *C-V Root Mean Square Error (%)* | *38.11* |

| ***Coefficient*** | ***Value*** | ***T-test*** | ***Significance*** |
|---|---|---|---|
| βPop (Log) | *2.30* | *5.25* | *>99%* |
| βGDP (Ln) | *0.799* | *4.51* | *>99%* |
| βAttract | *$5.13.10^{-4}$* | *3.37* | *>99%* |
| βCompac | -2.91 | -3.33 | >99% |
| βDens_Pop (interaction) | 0.091 | 2.24 | 98% |
| βCarOwn | $-1.66.\ 10^{-3}$ | -2.16 | 97% |
| βSlope | 0.0422 | 0.71 | 52% |
| βFare_int | 0.352 | 1.37 | 83% |

According to both models, the most significant variables are population and per capita GDP per capita: cities with a higher population and income tend to have public transport systems with higher capacity modes.

Furthermore, more linear cities with steeper slopes tend to have more efficient modes. It is also observed that fare integration, as well as a higher population density, less car ownership and presence of nearby cities are related to more complex public transport systems. These variables have signs consistent with other studies, but their significance in the models is less.

Finally, other variables such as the price of fuel, slenderness and network integration were not significant in both functions and were therefore not included in the models.

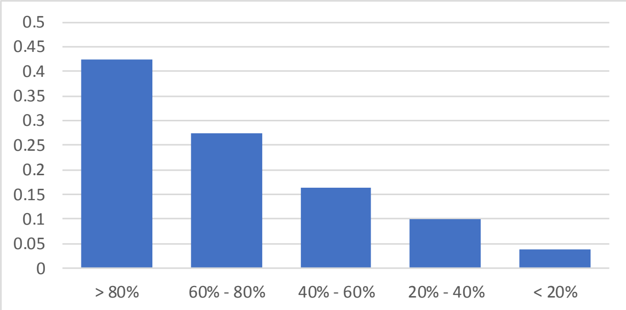
Although the MNL specification has a slightly better performance as shown in Table 7, the main advantage of the ordered Probit model is that a unique function ("Score") is applied for all cities, which makes results easier to comprehend. The class assigned by the model to a given city arises from comparing this Score with limit values (thresholds between classes) estimated by maximum likelihood and called “cuts”, whose values are shown in Table 9. Validation results from the ordered Probit model (cross-validation test subset) are shown in Table 8 below:

**Table 8 – Cross-Validation results for selected OP model**

**Comparison of predicted versus actual class (N=80)**

| OP model | model class | | | | |
|---|---|---|---|---|---|
| actual class | 1 | 2 | 3 | 4 | 5 |
| 1 | 7 | 0 | 0 | 0 | 0 |
| 2 | 2 | 10 | 1 | 0 | 0 |
| 3 | 0 | 2 | 37 | 2 | 0 |
| 4 | 0 | 0 | 4 | 7 | 2 |
| 5 | 0 | 0 | 0 | 3 | 3 |

**Proportion of cities by probability assigned to actual class (probability bands)**




The OP model assigns highest probability to the actual classification of cities in 80% of the test subset. In this subset, the Probit model assigns more than 80% probability to the actual classification of nearly half of the cities. In 20% of the cities, the model gives more than 50% probability to other classes, which indicates that these cities may actually have different public transport modes that most similar cities.

## 4. Discussion and conclusions

Results show that a simple Ordered Probit function can classify accurately the transport systems of cities in North and South America, with a success rate above 80%. Among the explanatory variables, city population and size are the most relevant as expected. That said, socioeconomic factors (regional per capita GDP and access to credit), geographic and morphologic variables (city size, represented by the compacity indicator, and interurban relation, represented by attractiveness) as well as fare integration are also relevant.

It is important to remember that the models aim to explain the existent PT modes in cities, which are not necessarily the ideal or recommended ones: in the city sample there might be cities requiring new public transport infrastructure, such as Metro or Bus Rapid Transit lines, and there might be cities that have more infrastructure than what they actually need. However, the classification models developed can give an insight for policy makers, in the sense that cities whose public transport complexity cannot be explained by the models are more likely to have a suboptimal public transport system.

### 4.1. Key findings

The selected model is an Ordered Probit function, which allows assigning a Score to each city, according to a linear combination of attributes that can be easily obtained:

$$Score = 2.30 * log_{10}(population) + 0.799 * ln(per\ capita\ reg.GDP) + 5.13.10^{-4} * attractiveness - 2.91 * compactness + 5.64.10^{-5} * log_{10}(population) * ln(density) - 1.66.10^{-3} * car\ ownership + 0.0422 * avg_{slope}(\%) + 0.352 * fare_{int}(dummy)$$

The proposed classification models make it possible to establish that the presence of certain public transport modes in a city depends not only on its population, but also on its residents' income, the form of the cities and the fare integration in their systems of transport.

Moreover, fare integration may also be relevant in providing an enhanced system efficiency, especially for cities whose Score is close to the threshold of the upper class. The influence of network integration, although the associated variable was not significant, should also be taken into account in future models. Considering a more detailed specification of integration in public transport systems is an interesting opportunity for further works.

When contrasting this Score with the thresholds estimated by the model, a class is assigned to a given city as shown in Table 9:[3]

**Table 9 – Score ranges and classes in Ordered Probit model**

| Score range | Assigned class (OP model) |
|---|---|
| 0 – 16.71 | I |
| 16.71 – 19.31 | II |
| 19.31 – 24.20 | III |
| 24.20 – 26.47 | IV |
| >26.47 | V |

To analyze the influence of the variables other than the population on the limits between the different classes, we compared the population thresholds that would require four cities with different characteristics to incorporate different transport modes, which is equivalent to find the population values in which the probability associated with a certain classification becomes greater than that of the lower level.

Four cities with different characteristics were selected:

- Asunción (Paraguay), a relatively compact city, with medium / low income, without large nearby cities and public transport without integration.

[3] *Score is here defined as the point estimate of the ordered Probit function for a given city. Given that Probit model probabilities correspond to the integral of a normal distribution between class thresholds ("cuts"), point estimate may differ from maximum probably class, especially if sample size varies among different classes. In our database, in which type III cities are the most common class, several cities whose point estimate corresponds to Class II have higher probability assigned to Class III. That said, Score thresholds may be nevertheless considered as a simple, yet effective criteria for planning purposes.*

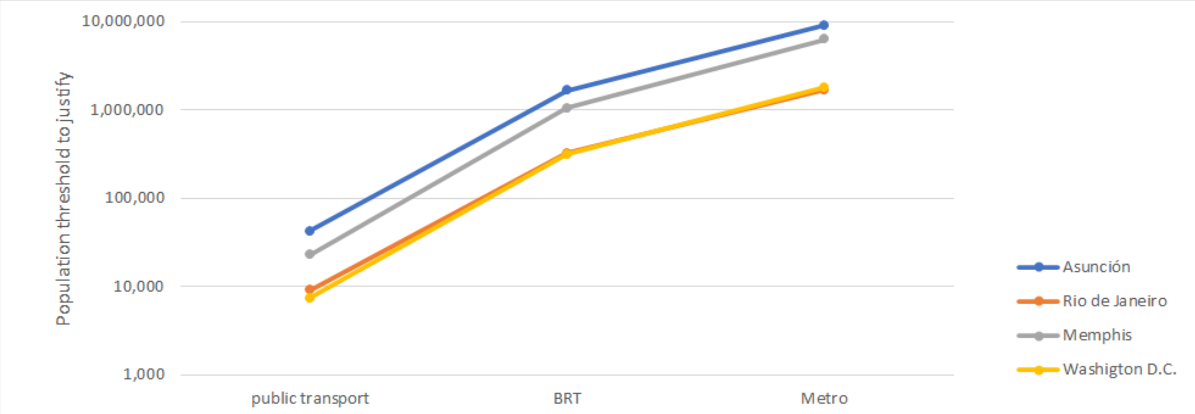
- Memphis (USA), a low-density city, relatively isolated, with representative characteristics of several North American cities.

- Rio de Janeiro (Brazil), a city of high density and complex shape, close to another large metropolitan area (Sao Paulo) and intermediate income.

- Washington (USA), a high-income city, with higher density than the average of North American cities and close to other large metropolitan areas (Philadelphia, New York).

Thresholds for the four city types are shown below:

**Figure 2 – Population thresholds to add public transport modes in different cities**




| Metropolitan Area | Population | Class (actual) | Probability assigned to actual class | population thresholds to justify (model) | | |
|---|---|---|---|---|---|---|
| | | | | public transport (III) | BRT (IV) | Metro (V) |
| Asunción | 2,130,000 | III | 37.4% | 42,658 | 1,671,788 | 9,167,355 |
| Rio de Janeiro | 11,900,000 | V | 99.6% | 9,090 | 324,403 | 1,703,345 |
| Memphis | 1,090,000 | III | 48.5% | 22,853 | 1,059,120 | 6,277,526 |
| Washington D.C. | 8,950,000 | V | 98.2% | 7,405 | 316,237 | 1,804,565 |

As can be seen, thresholds for cities alike Washington and Rio de Janeiro are much lower than those of cities that are compact, isolated and have a lower GDP, such as Asunción. This shows that, unlike what the simplest recommendations suggest, the form, density and GDP of cities are relevant when analyzing which modes of public transport they have.

Compared with other recommendations, guidelines for the existence of Metro (type V) that arise from our models are generally conservative, while the thresholds for implementing type IV modes are consistent with other findings.

The models are based on the existing modes of cities, which are not necessarily those with which a city should have. Next, we will discuss what are the implications of this.

### 4.2. Discussion on existent versus optimal modes

It is evident that public transport modes existing in a given city are not necessarily those that should exist. Several failures have been reported in the construction of transport projects, in which demand is much lower than was projected, such as the Miami metro (Babalik-Sutcliffe, 2002) and the BRT systems in Colombian cities apart from Bogotá (Jimenez, 2017). On the other hand, there are cities where the construction of more efficient modes than the existing has been discussed for decades.

It is possible to improve the developed models if the dependent variable is corrected in case the most efficient mode transports a number of passengers that is not compatible with its proper operation. On the other hand, optimization models representative of the cities of the sample can be applied, so that the dependent variable represents a recommended classification instead of the current classification of the transport systems of the cities.

In any case, the recommendations that emerge from the results of this study provide a more realistic guideline to transport planners compared to the general guidelines in place, given that the recommendation of modes of transport are not only dependent on the population of the cities, but also the structure of cities, their location and the income of their inhabitants.

### 4.3. Implications for policy makers

In this paper, classification models of public transport systems of 400 cities in the Americas were applied to identify which characteristics of cities determine the public transport technologies they have. Using a five-level classification, the selected ordered Probit model shows that the existent technologies depend not only on the population of the cities but also on other variables such as their form, fare integration between the different modes of transport and the GDP per capita. Among the independent variables in the selected model, fare integration is the one that can be changed in a short term: the other are related to urban planning and imply long-term policies.

Aggregate variables that can be obtained in any city were used. Thus, models are of general application, which is of interest for decision makers in localities where there are no strategic models or origin-destination surveys to make a specific analysis. In addition, the classification used not only provides a reference for the implementation of high capacity modes such as Metro or BRT, but also for the creation of a public transport system with buses, which is unusual in the literature.

Finally, it should be noted that the estimated Score can be useful for decision makers both for future projections and to justify interventions that improve the performance of existing modes. On the one hand, it is possible to estimate the future evolution of the Score of a city if it has projections of population, GDP and urban expansion plans, which allows estimating at what moment the characteristics of the city would be adequate to justify greater capacity modes. On the other hand, the models applied provide an argument under a simple approach to justify that certain interventions, such as increasing fare integration or promoting greater population density. This could help existing modes - particularly those that involve large investments in infrastructure - increase their demand to levels that are compatible with those originally projected.

## References:


AECOM (2012), Transit Lane Warrants Study. Consultancy Report for Roads ACT (Canberra, Australia).

Babalik-Sutcliffe, E. (2002) Urban rail systems: Analysis of the factors behind success, *Transport Reviews*, 22:4, 415-447

Baker, W. L., & Cai, Y. (1992). The role programs for multiscale analysis of landscape structure using the GRASS geographical information system. Landscape ecology, 7(4), 291-302.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. In Swiss Transport Research Conference.

Chen, C-L (2017). Modern Tram and Public Transport Integration in Chinese Cities – A case study of Suzhou. International Transport Forum, Discussion Paper 2017-21.

Crane, R. (2000). The influence of urban form on travel: an interpretative review. *Journal of Planning Literature*, 15-1: 3-23.

Currie, G., & Phung, J. (2007). Transit Ridership, Auto Gas Prices, and World Events. Transportation Research Record: Journal of the Transportation Research Board, 1992(1), 3–10.

Daganzo, C. F. (2010). Structure of competitive transit networks. *Transportation Research Part B:Methodological*, 44(4), 434- 446.

Dargay, J. (2002). Determinants of car ownership in rural and urban areas. A pseudo-panel analysis. *Transportation Research Part E: Logistics and Transportation Review*, 38:351-366.

Deen, T. B., & Pratt, R. H. (1992). Evaluating Rapid Transit. *Public Transportation*, 293.

Fielbaum, A., Jara-Diaz, S., & Gschwender, A. (2017). A parametric description of cities for the normative analysis of transport systems. *Networks and Spatial Economics*, 17(2), 343-365.

Givoni, M., & Banister, D. (Eds.). (2010). Integrated Transport: from policy to practice. Routledge.

Hidalgo, D. (2005). Comparación de Alternativas de Transporte Público Masivo-Una Aproximación Conceptual. Revista de ingeniería, (21), 92-103.

Hidalgo, D., Pereira, L., Estupiñán, N., & Jiménez, P. L. (2013). TransMilenio BRT system in Bogota, high performance and positive impact–Main results of an ex-post evaluation. *Research in Transportation Economics*, 39(1), 133-138.

Huang, J., Lu, X., & Sellers, L. (2007). A global comparative analysis of urban form: Applying spatial metrics and remote sensing. *Landscape and Urban Planning* 82 (2007) 184–197

ITDP (2014). The BRT Standard.

James, G., Witten, D., & Hastie, T. (2014). An Introduction to Statistical Learning: With Applications in R. MIT, Cambridge: USA.

Jara-Díaz, S. R. (2000). Allocation and valuation of travel time savings. Handbooks in Transport, 1, 303-319.

Jiang, B. (2007). A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 384(2), 647-655.

Jiménez, D. (2017). Sistemas BRT en Colombia: Una aproximación a la evaluación de los factores asociados a la demanda que pueden generar bajo uso de los sistemas - caso de aplicación BRT Santiago de Cali. Universidad Nacional de Colombia: Bogotá.

Kitamura, R. (2009). A dynamic model system of household car ownership, trip generation, and modal split: model development and simulation experiment. *Transportation* (2009) 36:711–732

LAMTA (2012), Transit Service Policy, Los Angeles Metropolitan Transportation Authority

Lehner, F. (1975). *Light Rail and Rapid Transit*. Light Rail Special Report 161, 37-49.

Levinson, H.S. et al. (1975), Bus Use of Highways: Planning and Design Guidelines, National Highway Cooperative Research Program, Report 155, Transportation Research Board, Washington D.C.

Loo, B. P. Y., & Cheng, A. H. T. (2010). Are there useful yardsticks of population size and income level for building metro systems? Some worldwide evidence. *Cities*, 27(5), 299–306.

McIntosh, J., Trubka, R., Kenworthy, J., & Newman, P. (2014). The role of urban form and transit in city car dependence: Analysis of 26 global cities from 1960 to 2000. Transportation Research Part D: Transport and Environment, 33, 95-110.

Medina, M., Giesen, R., & Muñoz, J.C. (2013). Model for the Optimal Location of Bus Stops and Its Application to a Public Transport Corridor in Santiago, Chile. *Journal of the Transportation Research Board* 2352, 84-93.

Moccia, L., & Laporte, G. (2016). Improved models for technology choice in a transit corridor with fixed demand. *Transportation Research Part B: Methodological*, *83*, 245-270.

Mohring, H. (1972). Optimization and Scale Economies in Urban Bus Transportation. *The American Economic Review*, 62:4, 591-604.

Muñoz, J. C., Batarce, M., & Hidalgo, D. (2014). Transantiago, five years after its launch. *Research in Transportation Economics* 48 (2014) 184-193.

O'Flaherty, C., & Bell, M. G. (Eds.). (1997). *Transport planning and traffic engineering*. Elsevier.

Oldfield, R. & Bly, P. H. (1988). An analytic investigation of Optimal Bus Size. *Transportation Research*, 22B, 319-337.

Ortúzar, J. & Willumsen, L. G. (2011). *Modelling transport*. John Wiley y Sons, London: UK.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., ... & White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, 13(4), 295-306.

Pohar, M.; Blas, M.; & Turk, S. (2004). Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. Metodoloski zverzki, 1(1), 143-161.

Potts, J., Marshall, M., Crockett, E. & Washington, J. (2010). A guide for planning and operating flexible public transportation services. TCRP Report 140.

Pushkarev, B., & Zupan, J. (1980) Urban Rail in America: An Exploration of Criteria for Fixed-Guideway Transit, Regional Plan Association, New York, Report prepared for the Urban Mass Transit Administration, U.S. Department of Transportation, Report No. UMTA-NY- 06-0061-80-1, November 1980, pp. xviii and xix

Saidi, S. (2016). Long Term Planning and Modeling of Ring-Radial Urban Rail Transit Networks (Doctoral dissertation, University of Calgary).

Schneider, M. (1959). Gravity models and trip distribution theory. *Papers in Regional Science*, 5(1), 51-56.

StataCorp (2011). Announcing Stata Release 12. Retrieved from https://www.stata.com/news/statanews.26.2.pdf.

Taylor, B., & Fink, C. (2003). The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature. UCLA, Department of Urban Planning.

Tirachini, A., Hensher, D. A., & Jara-Díaz, S. R. (2010). Comparing operator and user costs of light rail, heavy rail and bus rapid transit over a radial public transport network. *Research in Transportation Economics*, 29(1), 231-242.

Train, K. (1980). A structured Logit model of auto ownership and mode choice. *The Review of Economic Studies*, Vol. 47, No. 2 (Jan., 1980), pp. 357-370

Train, K. (2009). *Discrete Choice Models with Simulation* (2nd Edition). Cambridge University Press, Cambridge: UK.

Train, K., & McFadden, D. (1978). The goods/leisure tradeoff and disaggregate work trip mode choice models. Transportation research, 12(5), 349-353.

Transportation Research Board (2003), Bus Rapid Transit – Volume 2, Implementation Guidelines, TCRP Report 90.

Verma, A., & Dhingra, S. L. (2001). Suitability of alternative systems for urban mass transport for Indian cities. *Trasporti Europei* 18:4-15.

Vuchic, V. R. (2005). Urban transit: operations, planning, and economics. Wiley: UK.

White, P. (1979). Planning in the urban transport systems of the Soviet Union: a policy analysis. *Transportation Research Part A*, 13A: 231-240.

# Technology choices in public transport planning: a classification framework

## CRediT author statement – author contributions

Paul Basnak: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing – Original Draft, Visualization

Ricardo Giesen: Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition

Juan Carlos Muñoz: Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition