



DYNA

ISSN: 0012-7353

ISSN: 2346-2183

Universidad Nacional de Colombia

Cantillo-García, Victor; Guzman, Luis A.; Arellana, Julian
Socioeconomic strata as proxy variable for household income in
transportation research. Evaluation for Bogotá, Medellín, Cali and Barranquilla
DYNA, vol. 86, no. 211, 2019, October-December, pp. 258-267
Universidad Nacional de Colombia

DOI: <https://doi.org/10.15446/dyna.v86n211.81821>

Available in: <https://www.redalyc.org/articulo.oa?id=49663345030>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UNEN

Scientific Information System Redalyc

Network of Scientific Journals from Latin America and the Caribbean, Spain and Portugal

Project academic non-profit, developed under the open access initiative



DYNA
ISSN: 0012-7353
ISSN: 2346-2183
Universidad Nacional de Colombia

Socioeconomic strata as proxy variable for household income in transportation research. Evaluation for Bogotá, Medellín, Cali and Barranquilla

Cantillo-García, Victor; Guzman, Luis A.; Arellana, Julian

Socioeconomic strata as proxy variable for household income in transportation research. Evaluation for Bogotá, Medellín, Cali and Barranquilla

DYNA, vol. 86, no. 211, 2019

Universidad Nacional de Colombia

Available in: <http://www.redalyc.org/articulo.oa?id=49663345030>

DOI: 10.15446/dyna.v86n211.81821

Socioeconomic strata as proxy variable for household income in transportation research. Evaluation for Bogotá, Medellín, Cali and Barranquilla

Estrato socioeconómico como variable sustituta del ingreso del hogar en la investigación de transporte. Evaluación para Bogotá, Medellín, Cali y Barranquilla

Victor Cantillo-García ^a va.cantillo@uniandes.edu.co
Universidad de los Andes, Colombia

Luis A. Guzman ^a la.guzman@uniandes.edu.co
Universidad de los Andes, Colombia

Julian Arellana ^b jarellana@uninorte.edu.co
Universidad del Norte, Colombia

DYNA, vol. 86, no. 211, 2019

Universidad Nacional de Colombia

Received: 21 August 2019

Revised document received: 24 October 2019

Accepted: 13 November 2019

DOI: 10.15446/dyna.v86n211.81821

CC BY-NC-ND

Abstract: In Colombia, a socioeconomic stratification (SES) system of households based on the physical characteristics of the dwelling and its surroundings is implemented in order to set subsidies to public services. Many studies use this classification as a proxy variable for the household income, mainly because of the ease to retrieve data. This investigation proposes an evaluation of the SES as an explanatory variable for the household income. Ordered logit regressions were estimated for the cities of Bogotá, Medellín, Cali and Barranquilla, where the dependent variable corresponds to the SES. Even though there is a positive correlation between SES and household income, the relation depends on certain characteristics of the household as its size, vehicle availability and percentage of workers and students. The results were validated by comparing trip generation models. By including the significant attributes from the ordinal regressions, trip generation models with income and SES had similar fit, which suggests that using the SES as a proxy variable for the income, might be appropriate as long as their correspondence is checked according to household characteristics.

Keywords: socioeconomic strata, household income, Colombia, transport models, ordered logit.

Resumen: En Colombia se ha desarrollado un sistema de estratificación de viviendas basado en las características físicas del inmueble y su entorno, con el fin de establecer tarifas diferenciales a servicios públicos. Diversos estudios de transporte usan esta clasificación como variable sustituta del ingreso del hogar debido a la facilidad relativa para obtener dicha información. Este artículo realiza una evaluación del estrato como variable explicativa del ingreso del hogar desde la perspectiva de estudios de transporte. Se estimaron modelos de regresión logit ordinal para Bogotá, Medellín, Cali y Barranquilla, usando el ingreso como variable dependiente y el estrato del hogar como variable independiente. Se encontró que a pesar de que existe una correlación positiva entre el estrato y los ingresos del hogar, esta relación depende de ciertas características del hogar como su tamaño, la disponibilidad de vehículo y la proporción de trabajadores y estudiantes. Estos resultados se validaron mediante la comparación de modelos de generación de viajes por hogar. Al incluir los atributos significativos de las regresiones ordinales, se obtienen ajustes similares para modelos de generación de viajes con ingreso y estrato, lo que indica que usar el estrato como sustituto del ingreso puede ser adecuado,

siempre y cuando se controle previamente la correspondencia entre las variables de las características del hogar.

Palabras clave: estratos socioeconómicos, ingresos del hogar, Colombia, modelos de transporte, logit ordinal.

1. Introduction

Income or payment capacity is essential in transportation research, especially in travel behavior analysis. This variable is one of the main factors that determine the individual choices of travelers, such as trip rates and the probability of choosing a determined transportation mode or route [1]. In addition, population segmentation in socioeconomic categories is relevant for the evaluation of externalities, costs and transport benefits, as well as social impact and equity studies. In the context of Colombian transportation research, due to the difficulty of collecting data and the absence of information on income, it is a common practice to use an official classification of households-socioeconomic strata (SES)-as a proxy variable for the income.

The socioeconomic stratification system valid in Colombia was created in 1994, classifying households into six different categories numbered 1 to 6. Strata 1 households correspond to those of lesser quality and strata 6 to the best conditions. This classification was developed to assign subsidies for public service through differential rates between high and low strata households, and between residential and commercial users. The concept of differential tariffs was created in 1968 by Decree 3069, establishing that public services rates should consider the payment capacity of different population segments. However, it was not possible to define a unified criteria to determine the payment capacity of the users until 1991. Many valuations could be observed throughout the territory during this period, some based on cadastral criteria using statistical sampling frames, and some methodologies established by the same companies that provide public services using different approaches according to the area [2].

With the promulgation of the Political Constitution of Colombia in 1991, and the compliance with article 365, which orders the state to ensure the efficient provision of public services, the SES system was established for household classification for the assignment of public services tariffs. Initially, the Administrative Department of Statistics (DANE) would supply the classification methodologies to the municipalities, oversight that was then assumed by the National Planning Department (DNP) in 1992, and that finally returned to DANE in 2004 through Resolution 392.

The concept of socioeconomic stratification has been developed through a technical discussion about which criteria should be taken for its classification and convenience [3]. Since the beginning, through Decree 969 of 1991, it was established that an adequate indicator for the payment capacity of users was a composite of factors related to the physical characteristics of the house and its surroundings. This criteria is still active. According to DANE [2], the current methodology

for SES employs information from the house and its surroundings for classification based on data from the official cadastral system of Colombia, including variables like legal status of the property, registry, ownership, geometry, physical characteristics, econometric models for property valuation and zoning of homogeneous physical and economic areas. With this information, qualification procedures for a weighted hierarchy by location are executed, from which the strata is defined taking into account the average characteristics of the households.

As reviewed, the methodology for SES of households is based on the physical characteristics of the house and its surroundings; it does not directly consider economic attributes of the inhabitants. Despite this fact, there is a notion that the SES represents household income, based on the assumption that the physical conditions of the house are an indicator of the payment capacity of the inhabitants, even though the SES is a dwelling classification. Considering this situation, some public authorities have made evaluations of the viability of the SES as a subsidies allocation tool [4, 5].

Despite the problems that the use of SES as a proxy variable for income can generate, this practice continues to be implemented due to minimal availability of income information. On the one hand, there are complete and reliable official databases for SES, given its relevance in the estimation of public service rates. Besides, the strata is an easy variable to collect in the field. In general, it has been observed that people are not willing to provide information concerning their income on surveys, mainly due to security issues, while providing the SES is a practice that does not generate rejection. Thus, there have been cases such as the last official mobility survey in Barranquilla, hired by city hall in 2012, which did not inquire about income, just the household SES.

Given this background, this research proposes an evaluation of SES as a proxy variable for household income. The analysis was made for the cities of Bogotá, Medellín, Cali and Barranquilla, by estimating ordered logit regression models that include additional variables such as household size, availability of vehicles and proportion of workers and students. The approach aims to establish the relation between SES and income for different segments of the population from a transport engineering perspective. The results were then validated and applied in a trip generation model.

The document contains five sections. The first corresponds to the current introduction and the second includes a literature review about the use of SES as an indicator of household income, with an emphasis on transportation research. The third chapter explains the methodology and the fourth includes the results. In the last section, the main conclusions are summarized.

2. Literature review

The literature review has two sections. First, we discuss studies that evaluated the viability of SES as an indicator of income or payment

capacity; while the second section reviews the use of SES and income in transportation research.

2.1. SES as indicator for household income

Some studies developed by public entities have evaluated the pertinence of SES as an indicator for household income, mainly from the perspective of the instrument as a tool to target subsidies and social programs [3-6]. In the report describing the current methodology for SES, DANE [2] made a correlation analysis and performed a multiple correspondence analysis to establish a relation between the strata and the physical status of the dwellings according to cadastral information and the payment capacity of the household. As expected, they found that there is a correlation between these characteristics.

DNP [4] evaluated the efficiency of the strata in the allocation of subsidies by comparing it with different indicators of payment capacity. The results indicated that although the strata is adequate to identify households with lower payment capacity, it has led to population growth in the subsidized zones by customers who can afford to pay the full rate; so the authors suggested evaluation of allocation alternatives, for example, use of the Identification System of Potential Beneficiaries of Social Programs (SISBEN). The planning office of Bogotá [5] carried out a similar evaluation through experimental economic methodologies based on the dictator game methodology. They determined that SES tends to concentrate the population in the second and third strata, reducing the redistribution ability of the territory, so they proposed a multidimensional targeting index.

A study was developed to evaluate the SES in Bogotá in 1999 [7]. In the process, the authors estimated a logit regression model to predict the strata from the income using the quality-of-life survey data of DANE. Even though a positive correlation between the variables was evident, in low-income households, the probability of belonging to the second and third strata was always higher compared to the first, while, in the highest strata, the probability of belonging to the fourth strata was always higher compared to the fifth and sixth. In the middle-income ranges, the results were relatively congruent, with a greater possibility of belonging to the fourth and fifth strata.

In the academic literature, the connection between strata and income has been moderately addressed. In 2006, Sepulveda et al. [6] analyzed the SES limits from the point of view of cadastral information as a reflection of urban socioeconomic differentiation. The author began by arguing that the strata classification methodology has important limitations in four main aspects: some input variables are defined in a subjective way, there is no weighting of variables by discriminating power, the criteria are vulnerable to political pressures and there are deficiencies in the flow of information between the institution involved in the process. In addition, based on a descriptive statistical analysis from the Bogotá multipurpose survey of 2011, they point out that the current SES does not

rigorously classify households according to their living conditions. They also performed a multivariate analysis to observe the correlations between a set of socioeconomic indicators and a series of cadastral variables, concluding that stratification fails to classify households adequately in terms of their socioeconomic characteristics.

In this way, Alzate [3] conducted a review of the SES system in Colombia and the methodologies applied for classification since its conception, focusing on the utility of the tool as an instrument for allocating public expenditures. The author indicated that the evaluations carried out to date tend to state that there is a correspondence between income and SES based on how well the latter manages to target subsidies on the poorest. However, this is not completely true, since the stratification is based on the physical characteristics of the house and its environment, so when classifying households by income quintiles and comparing with their strata, the correspondence is not evident. In addition, the relation of strata and income tends to be stronger in urban areas compared to rural areas.

Finally, Bogliacino, Jimenez and Reyes [8] carried out an experiment based on the trust game and the dictator game with a sample of 1000 participants in Bogotá in order to identify the incidence of SES on household segregation. In the exercise, the sample correlation between SES and payment capacity, measured based on the household expenditure and the difficulty of facing an unexpected expense in a limited time, stood out. The correlation was determined through a χ^2 test, with a significant statistic value of 0.44, although some measurement error should be considered.

2.2. SES, income and transportation

In the field of transportation and urban studies in Colombia, few studies have used both income and SES as indicators of payment capacity, while many use SES as a proxy. Along this line, Bocarejo and Oviedo [9] calculated accessibility in different areas of Bogotá in order to evaluate inequities in accessibility to the territory. They formulated a cost function based on the monetary travel cost by mode, location and SES with an accessibility component corresponding to the ratio between monthly expenditures on transportation and individual income, finding that individual accessibility does not depend entirely on SES. In this case, joint use of location and SES have advantages in the estimation of travel costs, since the spatial information for SES is available from different official sources.

In Medellín, Marquet, Bedoya and Miralles-Gausch [10] analyzed local accessibility inequalities and willingness to walk, finding that lower-strata individuals had higher probability of taking a walking trip and a lower probability of taking a short trip. This means that this segment of the population walks more and has longer trips. The authors suggested that recent transport investments made in the city, such as the Metrocable system, has proven to be a useful tool for the integration of neighborhoods

with difficult access, but Medellín is still a highly unequal city, facing huge challenges in mobility and accessibility.

SES is used as a proxy for income evaluation and accessibility to the Bus Rapid Transit (BRT) of Cali [11], showing that 80% of the population was within a 15-min walk to the system. However, this accessibility varies with SES. Middle-range strata users (SES 3, 4 and 5) had higher walking accessibility to the bus rapid transit, compared to the lowest and highest SES-groups 1, 2, and 6.

Studies using household income as the analysis variable include the work of Guzman & Bocarejo [12], who analyzed the density distribution in the urban area of Bogotá, reviewing the implications for different segments of population and finding significant differences in travel times between income groups. Spatial differences between population and employment density lead to greater accessibility, shorter travel times and, on average, 150% more trips on a typical day for the high-income segments. The investigation did not use strata, but worked directly with the household income reported in the mobility survey from 2011. In addition, using data from the same survey, Guzman and Oviedo [13] addressed the issue of accessibility, affordability and equity from the perspective of subsidies to public transport.

In Latin America, many studies incorporating income can be found, in some cases, as an ordinal variable and as input in transportation research. Some examples include studies of departure time choice models [14], analysis of the relation between mobility and the built environment [15] or mode choice models [16, 17].

In contrast, studies addressing SES as an income indicator are more common in the Colombian context. Oviedo-Titheridge [18] studied social exclusion and accessibility in peripheral areas of Bogotá where low strata predominate. In this way, Teunissen, Sarmiento, Zuidgeest and Brussel [19] identified that Transmilenio, the BRT system of the city, offers equal access to all levels of SES, although alternative programs, like Sunday *ciclovía*, favor high strata. In the interaction between the transport and land price fields, Munoz-Raskin [20] determined through spatial hedonic models that properties located in areas of medium strata have more value in the market due to proximity to Transmilenio. Similar studies are evidenced in Cali, using SES as segmentation criteria [11,21].

The use of SES as an income indicator is also evidenced in other areas of transportation research. In demand modeling, is used as an input variable for trip generation models [22], in departure time and mode choice models [23] or in travel behavior analysis and other choice contexts, such as street crossing [24,25].

From the literature review, one could conclude that most of the studies use the SES as an indicator for household income in an implicit way, without formal justification. In this scenario, the current investigation proposes an evaluation to support that practice, with the purpose of providing a tool to validate its use.

3. Methodology

The methodology was divided into two parts: first, an ordered logit regression analysis was used to evaluate the pertinence of the use of SES as an explanatory variable for income and its correlation, incorporating some household characteristics. Then, the results from the first evaluation were validated by applying them in household daily trip generation models.

To carry out the evaluation of the SES as an explanatory variable for household income, we used data from mobility surveys of households in the four main cities of Colombia. This kind of survey collects information about trips made by the population and socioeconomic characteristics of users and their households. This information is generally used to develop transportation models, although the data are useful for other kinds of analysis.

The cities included in the investigation were Bogotá, with a mobility survey from 2015, Medellín with information from the origin-destination survey of 2017, Cali, with the mobility survey from 2015 and Barranquilla, with a household survey from 2017. From this information, we estimated ordered logit regression models in which the dependent variable was the household income and the independent variables were the SES and other characteristics of the household. This methodology was chosen because in all databases the income is found as an ordinal variable divided into non-uniform ranges. With this framework, it was possible to determine the level of explanation that the SES has over income, and if there is a correlation between them.

In an ordered logit regression model, level k of the observed response in the ordinal dependent variable Y is given by the relation between a linear estimator Y'_j , constructed from a set of explanatory variables x_i multiplied by the coefficients β_i to be estimated, plus a component of error ε_j , and some thresholds t_{k-1} that must also be estimated. This formulation is summarized as indicated in equations (1) and (2).

$$Y'_j = \sum_i \beta_i x_i + \varepsilon_j \quad (1)$$

$$Y = \begin{cases} 1 & \text{if } -\infty < Y' \leq \tau_1 \\ 2 & \text{if } \tau_1 < Y' \leq \tau_2 \\ \dots & \dots \\ k & \text{if } \tau_{k-1} < Y' \leq \infty \end{cases} \quad (2)$$

The probability of the response variable Y corresponding to a category smaller than k , is given by equation (3), where the logit function of such probability is equal to the threshold of the given category minus the linear estimator.

$$\text{logit}[P(Y \leq k)|x] = \tau_k - \sum_i \beta_i x_i \quad (3)$$

Note that the regression coefficients β_i correspond to the log-odds ratio of the independent variable against the dependent variable. Besides, the vector of coefficients β_i does not depend on the number of categories j , so one can assume that the connection between Y and x_i is independent of this value, so the log-odds ratio of probability is considered constant throughout all k points of response [26]. Hence, by estimating the values of the regression coefficients and their significance level, it is possible to identify the link between independent and dependent variables. The significance of the coefficients was evaluated through the Student's t-test, where the t-statistic is equal to the value of the coefficient divided by its standard deviation (equation 4). This value must be greater than 1.96, assuming a significance of 95%, to reject the null hypothesis that the coefficient is statistically equal to 0.

$$t = \beta_i / \delta_i > 1.96 \quad (4)$$

Two ordered logit regression models were specified. The first (M1) only contains the SES as an independent variable; while in the second model (M2), household size, two dichotomous variables that take the value of 1 if the household has a car/motorcycle available, and the proportion of workers and students in the household were included. We expected a correlation between these proportions and the income, so that as the number of workers increase, so does the income, while the opposite occurs with the number of students.

To compare the goodness of fit of the two models, we estimated the statistic ρ^2 , which relate the model quality to the null model which does not include explanatory variables, using the log-likelihood as shown in equation (5). The value of ρ^2 is between 0 and 1, and while the statistic is bigger, the model has a better fit.

$$\rho^2 = 1 - l(M) / l(M0) \quad (5)$$

Likelihood ratio tests were applied to compare the M1 and M2 models for each city. In this test, the log-likelihood ratio was estimated according to equation (6), which distributes Chi^2 with degrees of freedom equal to the difference in the degrees of freedom of the models. This statistic allowed us to evaluate the null hypothesis that the simple model (M1) is statically equal to the complex model (M2), considering that M2 is a more general version of M1.

$$LR = -2 * [l(M1) - l(M2)] \quad (6)$$

In the surveys from each city, the income was present as an ordinal non-uniform variable. Thus, it was necessary to aggregate it into three comparable levels, as indicated in Table 1.

Table 1
Income levels by city (millions of COP - nominal values)

| Level | Bogotá | Cali | Medellín | Barranquilla |
|--------|-----------|-----------|-----------|--------------|
| Low | <1.3 | <1.3 | <1.5 | <1.5 |
| Medium | 1.3 - 4.0 | 1.3 - 4.5 | 1.5 - 5.0 | 1.5 - 5.0 |
| High | >4.0 | >4.5 | >5.0 | >5.0 |

Source: The Authors.

The second stage of the evaluation consisted in the validation of the results obtained in the ordered logit analysis by comparing trip generation models per household using income and SES, with the attributes that were significant in the first analysis. For the models to be comparable, in those where the SES was included as an explanatory variable, the attribute was reclassified into three categories. In this way, strata 1 and 2 group into the low level, strata 3 and 4 into medium, and strata 5 and 6 into the higher level, so the classification coincides with income class.

The second analysis allowed us to implement and validate the findings of the ordered logit regression evaluation in a typical transportation model. For this purpose, we estimated ordinary linear regression models where the dependent variable corresponded to the number of household trips made in a typical day according to the information collected in the surveys, which has been found to be related to the payment capacity of the travelers [1]. We then compared these models by goodness of fit using the coefficient of determination R^2 and the significance of the estimators.

4. Results and discussion

The results chapter contains two sections according to the methodology applied. The first part refers to the estimation of the ordered regression models with their respective parameters of goodness of fit. The second includes the proposed generation models. In addition, Table 2 shows a description of each sample used by city. In all cities, the majority of households belonged to low-range SES and income, which is consistent with the results from the National Household Budget Survey 2016-2018 carried out by DANE. The survey determined that around 60.7% of Colombian households subsist with an income of less than two times the monthly minimum wage.

Table 2
Database description

| | | Bogotá | Cali | Medellin | Barranquilla |
|-------------------------|--------|--------|-------|----------|--------------|
| Number of observations | | 9,076 | 7,231 | 14,692 | 2,971 |
| Income Level | Low | 61.0% | 78.1% | 71.5% | 78.9% |
| | Medium | 34.4% | 20.0% | 25.9% | 19.8% |
| | High | 4.7% | 1.8% | 2.6% | 1.3% |
| SES | 1 | 13.4% | 21.6% | 13.0% | 47.7% |
| | 2 | 41.3% | 33.5% | 42.1% | 26.4% |
| | 3 | 35.1% | 31.1% | 29.5% | 15.0% |
| | 4 | 7.8% | 7.6% | 8.6% | 7.9% |
| | 5 | 1.5% | 4.4% | 5.6% | 2.1% |
| | 6 | 1.0% | 1.7% | 1.2% | 0.9% |
| Household size | Mean | 3.1 | 3.4 | 3.2 | 3.7 |
| | Min. | 1.0 | 1.0 | 1.0 | 1.0 |
| | Max. | 10.0 | 20.0 | 11.0 | 15.0 |
| Car availability | | 23.4% | 19.2% | 20.1% | 15.9% |
| Motorcycle availability | | 11.6% | 31.4% | 23.5% | 17.7% |
| Bicycle availability | | 0.1% | 22.0% | 4.1% | 11.4% |
| Average daily trips | Mean | 2.1 | 2.3 | 2.1 | 2.2 |
| | Min. | 1.0 | 1.0 | 0.6 | 1.0 |
| | Max. | 7.5 | 11.0 | 17.0 | 12.0 |
| Workers proportion | Mean | 0.48 | 0.42 | 0.44 | 0.40 |
| | Median | 0.50 | 0.40 | 0.40 | 0.33 |
| Students proportion | Mean | 0.13 | 0.17 | 0.18 | 0.21 |
| | Median | 0.00 | 0.00 | 0.00 | 0.20 |

Source: The Authors.

4.1. Ordered logit regression models

Tables 3 and 4 contain the results of the estimation of the ordered logit regression models M1 and M2, as well as the two thresholds of the three income categories for each city.

Table 3
M1 model results

| | | Bogotá | Cali | Medellin | Barranquilla |
|---|----------|---------------------|---------------------|---------------------|---------------------|
| Income Threshold | τ_1 | 1.815 (22.03)*** | 2.544 (26.15)*** | 2.446 (28.93)*** | 2.537 (24.91)*** |
| | τ_2 | 5.082 (48.5)*** | 5.773 (41.42)*** | 6.007 (55.08)*** | 5.937 (30.05)*** |
| SES | 2 | 0.767 (8.48)*** | 0.646 (5.64)*** | 0.706 (7.70)*** | 1.319 (9.95)*** |
| | 3 | 1.824 (20.37)*** | 1.372 (12.58)*** | 1.758 (19.44)*** | 1.721 (11.90)*** |
| | 4 | 3.681 (30.38)*** | 2.694 (20.94)*** | 2.930 (28.85)*** | 2.972 (17.97)*** |
| | 5 | 5.050 (25.21)*** | 3.401 (22.41)*** | 4.629 (37.70)*** | 3.482 (12.15)*** |
| | 6 | 5.585 (22.76)*** | 4.992 (21.87)*** | 5.884 (31.31)*** | 4.237 (9.47)*** |
| Coefficient (t-test) | | | | | |
| Significance levels: 90% (*), 95% (**), 99% (***) | | | | | |

Source: The Authors.

Table 4
M2 model results

| | | Bogotá | Cali | Medellin | Barranquilla |
|---|----------|---------------------|----------------------|---------------------|---------------------|
| Income Threshold | τ_1 | 4.009 (31.4)*** | 3.758 (26.4)*** | 4.672 (38.6)*** | 3.562 (16.5)*** |
| | τ_2 | 7.703 (49.6)*** | 7.234 (39.6)*** | 8.645 (57.7)*** | 7.110 (24.9)*** |
| SES | 2 | 0.784 (8.27)*** | 0.586 (5.01)*** | 0.683 (7.22)*** | 1.225 (9.0)*** |
| | 3 | 1.865 (19.5)*** | 1.255 (11.1)*** | 1.687 (17.9)*** | 1.547 (10.2)*** |
| | 4 | 3.707 (28.7)*** | 2.492 (18.4)*** | 2.879 (26.5)*** | 2.734 (15.4)*** |
| | 5 | 4.911 (22.8)*** | 3.027 (18.9)*** | 4.293 (33.3)*** | 3.308 (11.0)*** |
| | 6 | 5.695 (21.2)*** | 4.404 (19.0)*** | 5.524 (27.7)*** | 4.096 (9.0)*** |
| Car availability | | 1.357 (23.37)*** | 1.357 (18.75)*** | 1.244 (23.98)*** | 0.832 (6.8)*** |
| Motorcycle availability | | 0.321 (4.39)*** | 0.168 (2.34)** | 0.279 (5.67)*** | 0.141 (1.04) |
| Household size | | 0.387 (20.1)*** | 0.205 (9.59)*** | 0.324 (19.78)*** | 0.134 (3.8)*** |
| Workers proportion | | 1.251 (14.5)*** | 0.727 (6.34)*** | 1.866 (22.73)*** | 1.116 (5.4)*** |
| Students proportion | | -0.166 (-1.25) | -0.554 (-3.14)*** | -0.113 (-0.99) | -0.315 (-1.23) |
| Coefficient (t-test) | | | | | |
| Significance levels: 90% (*), 95% (**), 99% (***) | | | | | |

Source: The Authors.

In all cases, the thresholds increase in value with the income level, and the SES coefficients are highly significant. In addition, the coefficients of strata 1 were less than one, with the exception of Barranquilla, which increases as the strata goes higher. This means that low SES households have a higher probability of belonging to a low-income level.

Regarding M2 models, the availability of a car or motorcycle in the household was significant and relevant in the prediction of income from SES, with exception of the motorcycle in Barranquilla, where the value of the t-test was 1.04, so it is not significant. The car coefficients were higher than those in Bogotá, Cali and Medellín, so a household with this kind of vehicle has a higher probability of having a higher income compared to one that does not have a car. In the case of the motorcycle, the value of the coefficient was less than one, so households with this attribute tend to have a lower income than those with a car.

As expected, the household size was also significant, as the probability of having low-income increases with the number of inhabitants. Furthermore, the proportion of workers was significant in all cities, except

Cali, with coefficient values higher than one. Instead, the percentage of students was only significant for Cali, with values even below zero.

Graphical results for the M1 models of Bogotá and Cali are shown in Figs. 1 and 2 as examples. In Bogotá, the probability distribution of having a certain income level given the SES had the expected trend, since the probability of a household with high income was higher for SES 5 and 6. In the case of middle income, the probability was greater in SES 3 and 4, and for the low-income levels, the probability was higher in SES 1 and 2. However, in the other cases, this correlation was not clear because these patterns were not clear. Fig. 2 shows that an SES 5 household in Cali was more likely to have a medium or low income, while in SES 6, the probability of having a middle income was much higher than the probability of having a high income level. These results suggest that to use SES to segment the population according to their income in Bogotá, it was appropriate to use three categories (low, medium and high); while in Cali, Medellín, and Barranquilla there was no clear difference between medium and high income levels, so it is recommended to use two categories (low and medium/high).

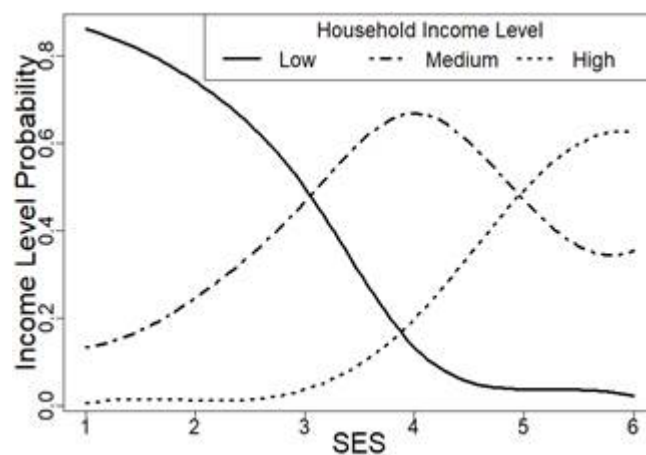


Figure 1

Probability of having a determined income level given the SES in Bogotá

Source: The Authors.

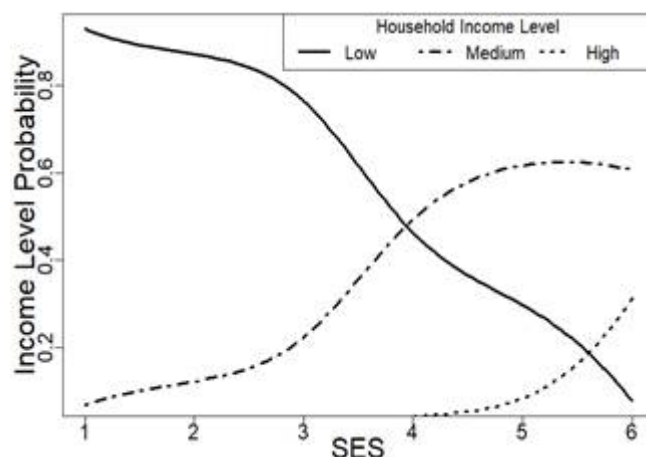


Figure 2

Probability of having a determined income level given the SES in Cali

Source: The Authors.

With the purpose of visualizing the difference in the interaction between SES and income depending on household characteristics, Figs. 3 and 4 show the probability distribution of having a determined income level given the SES in households with and without car for Bogotá and Cali. The probability changed when comparing two household of the same SES with and without car. Households with car had a higher probability of having a medium or high-income level and a lower probability of having low income.

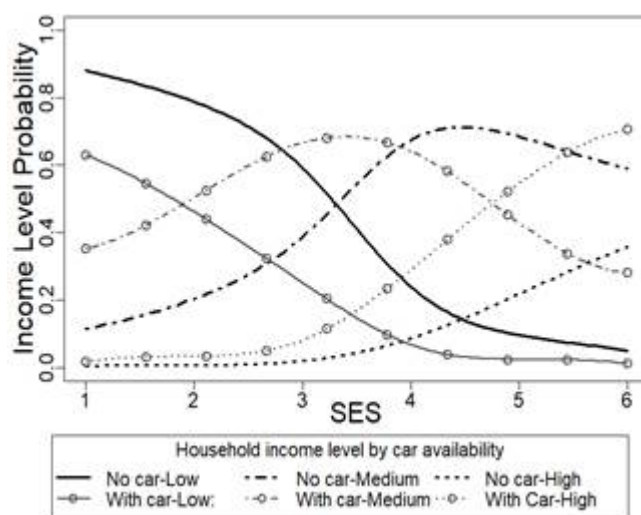


Figure 3

Probability of having a determined income level given the SES in households with and without car available in Bogotá

Source: The Authors.

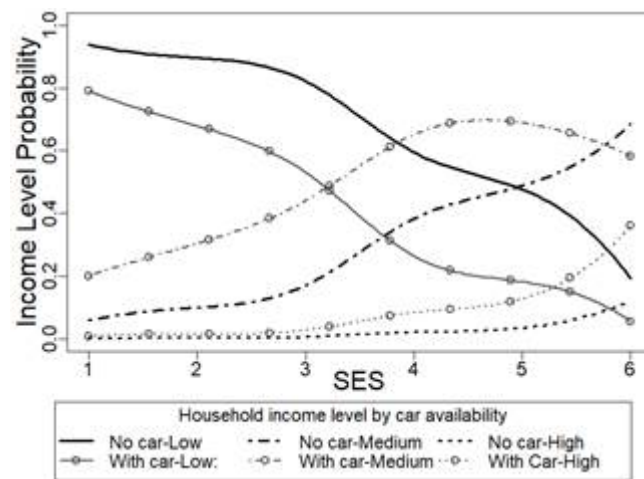


Figure 4

Probability of having a determined income level given the SES in households with and without car available in Cali

Source: The Authors.

Table 5 includes ρ^2 parameters to test the goodness of fit of the ordered logit models M1 and M2 by city. The statistic was higher for the more general models M2, meaning that they had a better fit in comparison with the simpler model M1 that only considered the SES as an explanatory variable for income, without introducing any more household attributes.

Table 5

ρ^2 Coefficients for ordered logit models by city

| Model | Bogotá | Cali | Medellín | Barranquilla |
|-------|--------|--------|----------|--------------|
| M1 | 0.1500 | 0.1526 | 0.1912 | 0.1569 |
| M2 | 0.2375 | 0.2139 | 0.2740 | 0.1893 |

Source: The Authors.

Moreover, Table 6 includes the likelihood ratio test between M1 and M2 models for each city. In all four cases, the test results indicate that the M2 model had a better fit compared to the M1 model, results that were in concordance with the ρ^2 parameter estimated.

Table 6

Likelihood ratio test for M1 and M2 models by city

| | | Bogotá | Cali | Medellín | Barranquilla |
|---------------------|----|----------|----------|----------|--------------|
| Log - Likelihood | M1 | -6267 | -3602.9 | -8148.8 | -1411.6 |
| | M2 | -5622.5 | -3341.9 | -7314 | -1359.2 |
| LR | | 1288.9 | 521.83 | 1669.7 | 104.84 |
| Prob. ($>\chi^2$) | | <2.2e-16 | <2.2e-16 | <2.2e-16 | <2.2e-16 |

Source: The Authors.

Most of the household characteristics included in the analysis, with the exception of the student proportion, were significant with a 95% level of confidence in the ordered logit regressions. A household with a high SES

(strata 5 and 6) does not necessarily have a higher income than one with a lower SES, a situation that was especially relevant in Cali, Medellín and Barranquilla. Therefore, when dividing the population by SES groups, precautions must be taken since the strata might not be representative of the household income. The results suggest that the relation between SES and strata is not direct, but that it depends on some household characteristics.

4.2. Household trip generation models

Two household daily trip generation models for each city were estimated. One considering the grouped SES (low, medium and high), and the other, the income levels as explanatory variables. In both cases, the household attributes considered in the ordered logit analysis were included as independent variables in order to evaluate whether similar results can be achieved in transportation models using SES and income through the incorporation of these attributes. The framework considered the use of a simple linear regression model to predict the number of trips per household. The estimation results, coefficients, t-tests and coefficients of determination R^2 can be reviewed in Table 7 for the SES models and Table 8 for income.

Table 7
Income trip generation models results

| | Bogotá | Cali | Medellin | Barranquilla |
|--------------------------------|----------------------|--------------------|--------------------|---------------------|
| Intercept | 0.052 (1.52) | -0.152 (-1.63) | 1.098 (19.)* | 0.020 (0.14) |
| Income Level | Medium | 0.342 (4.21)*** | 0.196 (4.6)*** | 0.153 (1.45) |
| | High | -0.005 (-0.08) | 0.667 (5.88)*** | 1.336 (3.68)*** |
| Car availability | 0.065 (2.34)** | 0.315 (3.75)*** | 0.261 (5.51)*** | 0.633 (5.44)*** |
| Motorcycle availability | 0.029 (0.85) | 0.237 (3.49)*** | 0.211 (5.11)*** | 0.227 (2.12)** |
| Household size | 2.085 (252)*** | 2.258 (106)*** | 1.654 (125)*** | 2.156 (79.16)*** |
| Workers proportion | -0.060 (-1.66)*** | 0.013 (0.12) | -0.034 (-0.54) | -0.135 (-0.80) |
| Students proportion | -0.004 (-0.07) | 0.429 (2.65)*** | 0.650 (7.51)*** | -0.150 (-0.76) |
| R² | 0.892 | 0.681 | 0.577 | 0.710 |

Coefficient (t-test)

Significance levels: 90% (*), 95% (**), 99% (***)

Source: The Authors.

Table 8
SES trip generation models results

| | Bogotá | Cali | Medellin | Barranquilla |
|--------------------------------|----------------------|----------------------|---------------------|--------------------|
| Intercept | 0.078 (2.14)** | -0.222 (-2.24)** | 0.969 (16.74)*** | -0.017 (-0.11) |
| SES | Medium (3-4) | -0.065 (-2.88)*** | 0.175 (2.65)*** | 0.228 (6.10)*** |
| | High (5-6) | 0.092 (1.26) | 0.230 (1.64) | 0.104 (1.02) |
| | | | 0.400 (5.17)*** | 0.619 (2.54)** |
| Car availability | 0.059 (2.20)** | 0.468 (5.63)*** | 0.265 (5.56)*** | 0.671 (5.76)*** |
| Motorcycle availability | 0.021 (0.63) | 0.230 (3.38)*** | 0.216 (5.24)*** | 0.232 (2.16)** |
| Household size | 2.081 (250)*** | 2.269 (107)*** | 1.671 (126.1)*** | 2.165 (78.6)*** |
| Workers proportion | -0.069 (-1.93)*** | 0.071 (0.64) | 0.028 (0.45) | -0.113 (-0.67) |
| Students proportion | -0.001 (-0.02) | 0.410 (2.53)** | 0.649 (7.49)*** | -0.188 (-0.94) |
| R² | 0.890 | 0.678 | 0.577 | 0.709 |

Coefficient (t-test)

Significance levels: 90% (*), 95% (**), 99% (***)

Source: The Authors.

SES and income coefficients resulted in similar levels of significance for both models in all cities. Furthermore, when reviewing the coefficient values and t-tests of household characteristics, it can be noted that these attributes were highly significant, especially household size and car and motorcycle availability, the last one with the exception of Bogotá. These results indicate that the number of trips made by a household in a typical day depended heavily on these attributes. On the other hand, the coefficients of determination of the models by city were practically the same, so the fitness of the models that use SES and those with income was similar.

According to the results, trip generation models can be estimated with similar results using SES or income, if the significant household attributes from the ordered logit analysis are added as explanatory variables. This means that by incorporating car and motorcycle availability, household size and the proportion of workers, similar predictions might be obtained using either SES or income.

5. Conclusion

This research provided useful information about the pertinence of using household SES as a proxy variable for income. This is a common practice in transportation research in Colombia due to the ease in obtaining information regarding SES in comparison with people's income. Nevertheless, the methodology of household classification in SES is based on the physical characteristics of the dwelling and its surroundings, and it does not directly consider the economic attributes of the residents. Keeping this in mind, the evaluation of the relation between income and strata for Colombian cities was justified.

The analysis focused in Bogotá, Cali, Medellín and Barranquilla, the four most important cities in Colombia. The work had two relevant components. In the first, we estimated ordered logit regression models to determine the predictive power of income from SES, considering

the following household characteristics: size, vehicle availability and proportion of workers and students. In the second, daily trip generation models by household were estimated using SES and income as explanatory variables, and adding the significant attributes from the first evaluation in order to test if similar results can be obtained when using SES as a proxy variable for income.

The main conclusion of the research is that there was a positive correlation between SES and household income. However, this correlation depends on specific household characteristics: household size, number of workers, and vehicle availability. Along this line, the probability of having a higher income increased in households with a car and decreased in those with a motorcycle. The correlation between SES and income was not completely clear because the probability of having a certain income level given that SES varies according to these attributes. For this reason, in order to obtain trustworthy results for income distribution given the SES, the inclusion of these characteristics is recommended when using the SES as a proxy variable for income.

Regarding the differences between the cities analyzed, Bogotá had the clearest correlation between SES and income. In other words, households in SES 1 and 2 were more likely to have low-income levels, those in SES 3 and 4 were more likely to have a middle-income range and high-strata households were more likely to have a high income. However, according to the results in Cali, Medellín and Barranquilla, the correlation was diffuse for these cities. This might be because a high strata household does not necessarily have an income higher than one in a lower strata. In such cities, households belonging to the high strata were more likely to have a medium income than a high income. In addition, homes in the middle SES, strata 3 or 4, had similar probabilities of having either low or middle incomes. From these results, it is possible to conclude that the correlation between SES and income is stronger in Bogotá compared to the other cities. However, the segmentation of low, medium and high SES to represent households within the same categories of income in different cities is not recommended.

Otherwise, the household daily trip generation models estimated using SES and income had similar fits when incorporating the significant attributes of the ordered logit regression analysis. This indicates that using the SES as a proxy variable for income might be viable if the correspondence between the variables is reviewed considering specific household characteristics. This situation is especially important when grouping the income or SES in categories with different levels.

Finally, future research should consider expanding the analysis to more cities, especially medium and small towns, and rural areas where the relationship between SES and income could be even more diffuse. On the other hand, it is pertinent to increase the spectrum of application of the results found, comparing results of models that include SES and income in other fields of transportation research, as is the case of route or mode choice exercises.

Acknowledgements

We would like to express our gratitude to the Bogotá Urban Planning Department (www.sdp.gov.co), for the funding and general support given for the development of this study through the agreement 369 of 2018.

References

- [1] Ortuzar, J.D. and Willumsen, L.G., *Modelling Transport*, 4th Ed., Oxford, England, 2011.
- [2] Departamento Administrativo Nacional de Estadística (DANE). *Metodología de estratificación socioeconómica urbana para servicios públicos domiciliarios. Enfoque conceptual*. Bogotá, Colombia, 2015.
- [3] Alzate, M.C., *La estratificación socioeconómica para el cobro de los servicios públicos domiciliarios en Colombia ¿Solidaridad o focalización?*, Bogotá, 2006.
- [4] Departamento Nacional de Planeación (DNP). *Evaluación de la estratificación socioeconómica como instrumento de clasificación de los usuarios y herramienta de asignación de subsidios a los servicios públicos domiciliarios*. Bogotá, Colombia, 2008.
- [5] Secretaría Distrital de Planeación. *La estratificación en Bogotá: Impacto social y alternativas para asignar subsidios*. Bogotá, Colombia, 2016.
- [6] Sepulveda, C., Lopez, D., Gallego, J.M., Acosta, J., Bernal, L., Bonilla, J. y Tejedor, F., *Los límites de la estratificación, en busca de alternativas*. En: Sepulveda, C., Lopez, D. y Gallego, J.M., Eds., 1st ed.. Editorial Universidad del Rosario, Bogotá, Colombia, 2014.
- [7] Econometría S.A., *Revisión de la metodología de estratificación socioeconómica de Santa Fe de Bogotá*. Bogotá, Colombia, 1999.
- [8] Bogliacino, F., Jimenez, L. y Reyes, D., *Identificar la incidencia de la estratificación socioeconómica urbana sobre la segregación de los hogares Bogotáños*. *Investigaciones y Productos CID*, 24, 2015.
- [9] Bocarejo, J.P. and Oviedo, D.R., *Transport accessibility and social inequities: a tool for identification of mobility needs and evaluation of transport investments*. *Journal of Transport Geography*, 24, pp. 142-154, 2012. DOI: 10.1016/j.jtrangeo.2011.12.004
- [10] Marquet, O., Bedoya, V.R. and Miralles-Guasch, C., *Local accessibility inequalities and willingness to walk in Latin-American cities: Findings from Medellín, Colombia*. *International Journal of Sustainable Transportation*, 11(3), pp. 186-196, 2016. DOI: 10.1080/15568318.2016.1230804
- [11] Delmelle, E.C. and Casas, I., *Evaluating the spatial equity of bus rapid transit-based accessibility patterns in a developing country: the case of Cali, Colombia*. *Transport Policy*, 20, pp. 36-46, 2012. DOI: 10.1016/j.tranpol.2011.12.001
- [12] Guzman, L.A. and Bocarejo, J.P., *Urban form and spatial urban equity in Bogotá, Colombia*. *Transportation Research Procedia*, 25, pp. 4491-4506, 2017. DOI: 10.1016/j.trpro.2017.05.345

- [13] Guzman, L.A. and Oviedo, D., Accessibility, affordability and equity: Assessing “pro-poor” public transport subsidies in Bogotá. *Transport Policy*, 68 (June), pp. 37-51, 2018. DOI: 10.1016/j.tranpol.2018.04.012
- [14] Arellana, J., Daly, A., Hess, S., de Dios-Ortúzar, J. and Rizzi, L.I., Development of surveys for study of departure time choice: two-stage approach to efficient design. *Transportation Research Record*, 2303(1), pp. 9-18, 2012. DOI: 10.3141/2303-02
- [15] Larrañaga, A.M., Rizzi, L.I., Arellana, J., Strambi, O. and Cybis, H.B.B., The Influence of built environment and travel attitudes on walking: a case study of Porto Alegre, Brazil. *International Journal of Sustainable Transportation*, 10(4), pp. 332-342, 2016. DOI: 10.1080/15568318.2014.933986
- [16] Jara-Díaz, S.R. and Ortúzar, J. de D., Introducing the expenditure rate in the estimation of mode choice models. *Journal of Transport Economics and Policy*, 23(3), pp. 293-308, 1989.
- [17] Amador, F.J., González, R.M. and Ortúzar, J. de D., On confounding preference heterogeneity and income effect in discrete choice models. *Networks and Spatial Economics*, 8(2-3), pp. 97-108, 2008. DOI 10.1007/s11067-007-9043-6
- [18] Oviedo-Hernandez, D. and Titheridge, H., Mobilities of the periphery: informality, access and social exclusion in the urban fringe in Colombia. *Journal of Transport Geography*, 55, pp. 152-164, 2016. DOI: 10.1016/j.jtrangeo.2015.12.004
- [19] Teunissen, T., Sarmiento, O., Zuidgeest, M. and Brussel, M., Mapping equality in access: the case of Bogotá’s sustainable transportation initiatives. *International Journal of Sustainable Transportation*, 9(7), pp.457-467, 2015. DOI: 10.1080/15568318.2013.808388
- [20] Munoz-Raskin, R., Walking accessibility to bus rapid transit: does it affect property values?. The case of Bogotá, Colombia. *Transport Policy*, 17(2), pp. 72-84, 2010. DOI: 10.1016/j.tranpol.2009.11.002
- [21] Jaramillo, C., Lizárraga, C. and Grindlay, A.L., Spatial disparity in transport social needs and public transport provision in Santiago de Cali (Colombia). *Journal of Transport Geography*, 24, pp. 340-357, 2012. DOI: 10.1016/j.jtrangeo.2012.04.014
- [22] Márquez, L., Días, M., Ortiz, D. y Central, U., Transferibilidad geográfica de modelos de generación de viajes urbanos: comparación de modelos de regresión y tasas de viajes para algunas ciudades colombianas. *Ingeniería y Desarrollo*, 34(1), pp. 225-247, 2016.
- [23] Ramos, R., Cantillo, V., Arellana, J. and Sarmiento, I., From restricting the use of cars by license plate numbers to congestion charging: Analysis for Medellín, Colombia. *Transport Policy*, 60, pp. 119-130, 2017. DOI: 10.1016/j.tranpol.2017.09.012
- [24] Cantillo, V., Arellana, J. and Rolong, M., Modelling pedestrian crossing behaviour in urban roads: a latent variable approach. *Transportation Research Part F: Traffic Psychology and Behaviour*, 32, pp. 56-67, 2015. DOI: 10.1016/j.trf.2015.04.008
- [25] Delgado-Lindeman, M., Arellana, J. and Cantillo, V., Willingness to pay functions for emergency ambulance services. *Journal of Choice Modelling*, 30, pp. 28-37, 2019. DOI: 10.1016/j.jocm.2018.12.001

- [26] McCullagh, P., Regression models for ordinal data. Journal of the Royal Statistical Society, 26(2), pp. 211-252, 1964.

Notes

V. Cantillo-García, received the BSc. in Eng in Civil Engineering in 2014, from the Universidad del Norte, Barranquilla Colombia. And the BSc. Eng in Environmental and Land Planning Engineering in 2014 from the Politecnico di Milano, Milan, Italy. Currently, he is a research assistant at the Urban and Regional Sustainability Studies Group, Department of Civil Engineering, Universidad de los Andes, Colombia. ORCID: 0000-0003-3101-6825

L.A. Guzman, received the BSc. Eng in Civil Engineering in 2000 and a MSc. in Civil Engineering in 2003 from the Universidad de los Andes, Bogotá, Colombia. He also received a MSc. in Planning, Design and Sustainable Management of Civil Engineering Systems in 2008 and the PhD in Urban Transport Planning in 2011, both from Universidad Politecnica de Madrid, Spain. He is currently director of the Urban and Regional Sustainability Studies Group in the Department of Civil Engineering of the Universidad de los Andes, Colombia, leading projects and researches with a focus in transportation and urban planning. ORCID: 0000-0002-6487-7579

J. Arellana, received the Bs. Eng in Civil Engineering in 2004 from the Universidad del Norte, Barranquilla Colombia, and the MS in Engineering Science in 2012, and the PhD in Engineering Science in 2012, both from the Pontificia Universidad Catolica de Chile, Santiago, Chile. Currently, he is the director of Research and Graduate Studies for the College of Engineering at Universidad del Norte. He is also an associate professor of the same University. Julian has been the leader of different research and consulting projects related to transport and logistics in several countries. His research areas include non-motorized transport, transport planning, and advanced econometric modelling. ORCID: 0000-0001-7834-5541

How to cite: Cantillo-García, V, Guzman, L.A. and Arellana, J, Socioeconomic strata as proxy variable for household income in transportation research. Evaluation for Bogotá, Medellín, Cali and Barranquilla. DYNA, 86(211), pp. 258-267, October - December, 2019.