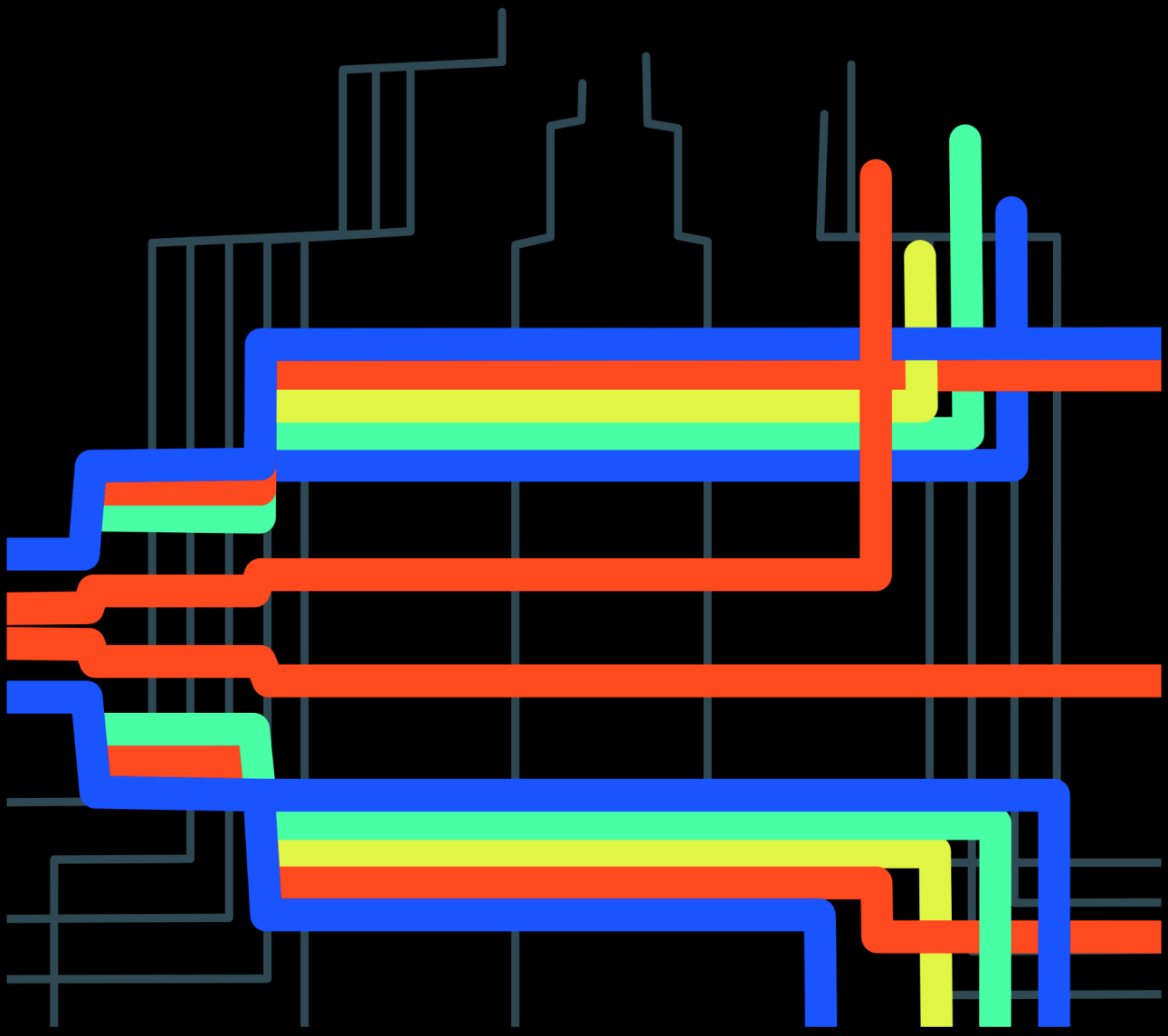
https://revistas.ucr.ac.cr/index.php/ingenieria/index
www.ucr.ac.cr / ISSN: 2215-2652

Ingeniería

Revista de la Universidad de Costa Rica

JULIO/DICIEMBRE 2021 - VOLUMEN 31 (2)



EDITORIAL UCR

DOI 10.15517/ri.v31i2.45018
Ingeniería 31(2): 80-97, Julio-diciembre, 2021. ISSN: 2215-2652. San José, Costa Rica

Fare inequities in public transit system in Costa Rica

Desigualdades tarifarias en el sistema de transporte público en Costa Rica

Cristhian Santiago Quirós-Calderón
Lic. Lecturer, Escuela de Ingeniería Civil, Universidad de Costa Rica, San José, Costa Rica
Email: cristhian.quiros@ucr.ac.cr
ORCID: 0000-0002-4172-6989

Jonathan Agüero-Valverde, Ph.D. (Corresponding autor)
Professor, Programa de Investigación en Desarrollo Urbano Sostenible, Escuela de Ingeniería Civil, Universidad de Costa Rica (ProDUS-UCR), San José, Costa Rica
Email: jonathan.aguero@ucr.ac.cr
ORCID: 0000-0002-9096-9274

Recibido: *8 de diciembre 2020* ***Aceptado:*** *11 de mayo 2021*

Abstract

Problems in transit fare equity affect the daily commute of specific groups that depend mostly on public transportation. Previous studies showed that some routes present operational characteristics that increased the price charged to the users. To address this issue, a methodology to identify the routes that have fares much higher than expected, after considering operational parameters, is developed. This paper presents a methodology implemented to evaluate fare inequities in public transport networks. The case study is the bus public transport network in Costa Rica. The evaluation is performed using fare per kilometer as independent variable and operational variables, such as route length, monthly ridership and vehicle occupancy by using cluster analysis and Bayesian multilevel modelling. The results indicate that random coefficients models perform better than independent models for clustered data. Furthermore, the routes with higher differences between observed and estimated (i.e., expected) fares are the ones to be addressed first in individual audits, because these are the routes who charge higher operational costs into the fare, increasing inequity among the population.

Keywords:
Bayesian models, Cluster, Evaluation, Fare, Multilevel Modeling.

Resumen

Los problemas en la equidad de las tarifas de transporte público afectan el viaje diario de grupos específicos que dependen exclusivamente de éste. Estudios anteriores mostraron que algunas rutas presentan características operativas que incrementaron la tarifa cobrada a los usuarios. Para abordar este problema, se desarrolla una metodología para identificar las rutas que tienen tarifas mucho más altas de lo esperado, luego de considerar los parámetros operativos. Este artículo presenta la metodología implementada para evaluar las inequidades tarifarias en la red de transporte público. El caso de estudio es la red de transporte público de autobuses en Costa Rica. La evaluación se realiza utilizando la tarifa por kilómetro como variable independiente y variables operativas como la longitud de la ruta, el número de pasajeros mensuales y la ocupación de vehículos, mediante el uso de análisis de conglomerados y modelos bayesianos multinivel. Los resultados indican que los modelos de coeficientes aleatorios funcionan mejor que los modelos independientes para datos agrupados. Además, las rutas con mayores diferencias entre las tarifas observadas y estimadas (es decir, esperadas) son las que deben abordarse primero en las auditorías individuales, porque estas son las rutas que cobran mayores costos operativos en la tarifa, aumentando la inequidad entre la población.

Palabras clave:

Conglomerado, Evaluación, Modelado multinivel, Modelos Bayesianos, Tarifa.

1. INTRODUCTION

User complaints regarding public transit fare must be addressed as a primary issue in transport policy. When these problems affect the daily commute of specific groups that depend mostly on public transit, response is needed from the competent authorities to prevent inequity. Fare inequities can be the result of flaws in the operational characteristics of the transit network, which affect the fare charged for the service.

Equity can be referred as the distribution of impacts (benefits and costs) and whether that distribution is considered fair and appropriate [1]. Equity in bus public transportation fares is difficult to accomplish when control and information are insufficient, particularly when the area of service is extensive and presents high operational route diversity. The bus public transportation network in Costa Rica is regulated nationwide and consists of nearly 750 routes not separated in regions or urban areas; therefore, due to the regulation structure, there is a considerable lack of information and control from the local authorities.

Previous studies in Costa Rica [2] – [4] showed that some routes present operational characteristics that affect the price charged to the users, increasing inequity for vulnerable population.

The public transit system in Costa Rica is private-operated and the fare is stablished by the Public Service Regulator Authority. The fare model is complex and each private operator request revisions of the fare individually, which has resulted in significant differences on the fare by kilometer between routes that are otherwise similar. To address this issue, a methodology to identify the routes that have fares much higher than expected, after considering operational parameters, is proposed. Solving these differences can contribute to increase equity and efficiency in the bus public transportation system.

Clearly, to compare the fare between different transit routes, these should have similar characteristics. Route grouping by similar characteristics, using existing information, allows the identification of outliers within groups. Multilevel or hierarchical modeling was chosen as the method for the group evaluation, since it allows to control for the natural correlation existing between routes of the same group [5] while also taking advantage of the nested data structure to improve model estimation [6]. More importantly, non-hierarchical models are usually inappropriate for hierarchical data because, with few parameters, they generally cannot fit accurately large datasets as the one use in this study, whereas with many parameters, they tend to overfit the data [5].

To identify inequities in the transit routes, the difference between the observed fare per kilometer and what is expected in similar routes (after controlling for several operational characteristics) is proposed. The routes that have higher differences between observed and expected fare per kilometer are the ones that charge higher operational costs into the fare, even after controlling for different operational characteristics. Each category or group sets its own value for comparison, according to the estimates. The larger outliers within each group are the ones to be addressed first as they impose the most expensive fares after controlling for operational

characteristics. This procedure is analogous to the one use to identify sites with excess crash frequency in highway safety [7]. As in the case of highway safety, the routes with the largest fare excess are the ones expected to have the highest probability of reduction.

A Bayesian approach is chosen because it provides the flexibility to model complex correlation structures as the ones included in multilevel models. Further, the Bayesian approach allows to easily compare different modeling approaches within the same framework. Bayesian models have been gaining popularity in several fields as the approach of choice when modeling multiple levels and incorporating random effects or complicated dependence structures [8], [9].

The purpose of this study is to propose a method to identify transit routes with significantly higher fares compared to similar routes by applying Bayesian modeling and creating different route groups or clusters. These differences could represent inequities or inefficiencies in the transit system that should be studied further.

This paper is organized as follows. First, a literature review regarding equity, multilevel modeling, and cluster analysis is presented. Then, the methodology used for the research is described, followed by the presentation of the data, which includes information about the background of the bus transit network in Costa Rica, the group creation and the parameters used for the models. Finally, the results, which conduct a comparison of goodness-of-fit and estimates between generalized linear models and random coefficients models, are presented and discussed; followed by the conclusions and recommendations for future research.

2. LITERATURE REVIEW

Equity has been widely explored regarding transit policies. It has important political implications, and it needs to be considered so that the whole transportation system [10] and the individual bus system revenue can be improved [11]. The proposed methodology doesn't involve the formulation of policies but seeks to identify the routes in which equity can be achieved while improving efficiency [12].

The proposed model estimates the fare-per-kilometer variable with covariates such as ridership, vehicle occupancy and route length. Other studies have also considered the effects of operational characteristics of the bus public network on fare structure. Ling [13] estimated, given the fares and ridership in the flat fare system, how the total ridership, operator revenue, passenger-km, and consumer surplus would change if the fare structure changes to fare differential system. Fairness and equity, increased revenue and increased ridership were identified as some reasons for promoting differentiated fares.

More recently, Liu et al [14] analyzed the effects of differential fare strategies on social welfare. The authors found that all differential fare strategies produced higher social welfare than the flat fare. Similarly, Tang et al [15] proposed an optimization of bus line fares using elastic demand. The authors maximize social welfare by proposing several operational strategies. The proposed methodology seeks to improve these and other aspects by identifying operational parameters that contribute to inefficiency within groups.

In order to avoid the biases due to analyzing systems with different sizes and operating environments, Karlaftis and McCarthy [16] used cluster analysis to classify 256 transit systems into homogenous groups. The 742 routes used in this analysis required a similar grouping in order to improve model performance. For regression of clustered data, random coefficients models are usually considered, as they control for between-cluster variation [17].

As mentioned above, the route classification allows a multilevel approach by analyzing two levels of data: route-level and category-level. Multilevel modeling has been applied to transportation in different areas of study, but none have explored the effect of operational parameters in fare estimation. Cervero and Kang [18] analyzed the impact in land-use changes and land values due to the upgrade of BRT services in Seoul Korea. The analysis was performed with multilevel modeling, in which the first level corresponded to the parcels and the second level corresponded to the neighborhood groups. Similar individual and neighborhood levels were used by Yavuz et al [19] and Paez and Mercado [20]. Yavuz et al [19] performed a multilevel approach to analyze how perceptions of bus and train safety in Chicago, Illinois vary as a function of person-level characteristics and neighborhood-level characteristics. Paez and Mercado [21] determined individual and neighborhood characteristics that affected the distance traveled and the variability of these factors on each mode type using multilevel analysis. Wang et al. [21] analyzed a Demand Responsive Transport (DRT) System in Manchester; they explored the relationship between a range of socioeconomic variables and service area factors and the demand for DRT using a linear multilevel model. In this model the first level was the census track and the second level the service area.

The studies mentioned above applied multilevel modeling to population or land distribution parameters as parcels, neighborhoods, census tracks or service areas. This research applied multilevel modelling to the transportation system itself, in which the reference level is the whole network, while subnetworks represent the subsequent levels. Ma and Lebacque [22] applied multilevel modeling using similar reference and subsequent levels to study system optimal routing for public transit systems.

Generalized linear models and random coefficients models have been applied to clustered observations. Similar approaches were used by Laird and Ware [23] for longitudinal data and by Aguero-Valverde and Jovanis [24] for crash frequency models. Laird and Ware [23] explored a general family of two-stage random-effects models in two examples from an epidemiological study of the health effects of air pollution. Aguero-Valverde and Jovanis [24] implemented multilevel models in road segments of different functional types by using a full Bayes hierarchical approach to analyze spatial correlation in road crash models in Pennsylvania and Washington. The research concluded that random effects significantly improved the precision, particularly for small sample sizes and low sample means.

3. METHODS

The models are estimated using a full Bayes hierarchical approach in order to estimate not only the parameters but also the estimated excess fare of each transit route and its statistical significance.

In Bayesian inference, the posterior distribution of the parameters of interest is estimated as the product of the prior distribution times the likelihood of the model up to a constant, based on the Bayes Theorem. For more details on Bayesian inference, interested readers can refer to Gelman et al [5], Congdon [6] or Koch [25].

At the first level of hierarchy, the logarithm of the fare per kilometer is assumed normally distributed:

$$\ln(FKM_{ik}) \sim N(\mu_{ik}, \tau) \tag{1}$$

where FKM_{ik} is the fare per kilometer for route i of group k, μ_{ik} is the expected value (i.e. the mean) for route i of group k and τ is the precision for the normal distribution (i.e. the inverse of the variance).

For this model, the second level of the hierarchy defines the mean:

$$\mu_{ik} = \beta_{0k} + \beta_{1k} * \ln(len_{ik}) + \beta_{2k} * \ln(rid_{ik}) + \beta_{3k} * \ln(voc_{ik}) \tag{2}$$

where β_{0k} is the constant term for route group k, β_{1k} is the coefficient for length for route group k, len_{ik} is the length of route i of group k, β_{2k} is the coefficient for monthly ridership for route group k, rid_{ik} is the average monthly number of passengers of route i for group k, β_{3k} is the coefficient for vehicle occupancy for route group k, voc_{ik} is the average vehicle occupancy for route i of group k.

This formulation is equivalent to a linear model:

$$\ln(FKM_{ik}) = \beta_{0k} + \beta_{1k} * \ln(len_{ik}) + \beta_{2k} * \ln(rid_{ik}) + \beta_{3k} * \ln(voc_{ik}) + e_i \tag{3}$$

where the error term is normally distributed:

$$e_i \sim N(0, \tau). \tag{4}$$

The hyper-prior for τ is supposed gamma:

$$\tau \sim gamma(0.01, 0.01). \tag{5}$$

The selection of the prior distribution for the betas differentiates between and independent and correlated model. The prior distributions in the case of the independent coefficient model are:

$$\beta_{jk} \sim N(0, 0.0001), \quad j=0,1,2,3 \text{ and } k=1,..,30. \tag{6}$$

The coefficients for each group of routes are completely independent and the model is almost equivalent to estimate a single model for each group. The only "shared" information among groups is the random variability between transit routes, since they share the same τ as shown in (5).

For the random coefficients model (i.e. correlated model), the coefficients change for each route group, but belong to the same normal distribution:

$$\beta_{jk} \sim N(0, \tau_j), \quad j=0,1,2,3 \text{ and } k=1,..,30. \tag{7}$$

The hyper-prior for each τ_j is supposed gamma:

$$\tau_j \sim gamma(0.01,0.01). \quad (8)$$

This correlated or random coefficients model allows for "shared" information among route groups through the precision parameter, which in turn, improves parameter estimation.

Finally, to identify inequities and inefficiencies in the transit routes the difference between the observed fare per kilometer and what is expected based on similar routes (i.e. belonging to the same group and controlling for the covariates) is estimated:

$$\Delta_{ik} = FKM_{ik} - \exp(\mu_{ik}) \quad (9)$$

Clearly , the highest deltas express the highest excess or inequities between transit routes that should, in theory, have very similar fares per kilometer. Here, another of the advantages of Fully Bayesian models is evident since the full posterior distribution of the deltas is known, which allows to explore the statistical significance of the excess fare.

Traditionally, two different goodness-of-fit measures are used for model comparison in a Bayesian framework: posterior mean deviance ($\bar{D}$) and Deviance Information Criterion (DIC). The posterior mean deviance can be seen as a Bayesian measure of fit or 'adequacy'. On the other hand, the deviance Information Criterion was proposed by Spiegelhalter et al [26] to account model complexity. The DIC is considered the Bayesian equivalent of the Akaike Information Criterion (AIC). DIC is defined as an estimate of fit plus twice the effective number of parameters.

$$\text{DIC} = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \quad (10)$$

where $D(\bar{\theta})$ is the deviance evaluated at $\bar{\theta}$, the posterior means of the parameters of interest, p_D is the effective number of parameters in the model, and $\bar{D}$ is the posterior mean of the deviance statistic $D(\bar{\theta})$. As with AIC, models with lower DIC values are preferred.

4. DATA

The data for the study was provided by ARESEP (Regulatory Authority of Public Services). Several details about public transport background in Costa Rica are needed to have a better understanding of the methods applied to the dataset of the bus routes. Urban development in Costa Rica is concentrated in the Central Valley, the natural barriers serve as borders for the Great Metropolitan Area (GAM), which main city is San José, the capital.

In terms of transport network, cities in Costa Rica can be hierarchized in four categories:

1. San José
2. Main districts of the provinces within the GAM: Heredia, Cartago and Alajuela
3. Main districts of the main cities outside the GAM
4. Main districts of the secondary cities outside the GAM

The transport network revolves around the capital; San José is the distribution center for all the country's regions, with about 33 % of the total routes connecting San José directly with other cities, within or outside the GAM. Each of the cities mentioned above present their own transport system, this represents about 65 % of the routes of the country. The last 2 % is composed of the routes that communicate main districts between themselves, without stopping in San José. As noted, the transport system in Costa Rica presents substantial size and operational differences according to its location.

To perform the analysis adequately, 742 bus public transportation routes were classified in 30 groups according to four main operation characteristics: location, route structure, terrain and length. Location was the main parameter for the classification. The other three parameters depend on the route configuration. For route structure, according to Molinero and Sánchez [27] and Vuchic [28], routes were classified in radial or circular configuration. Almost 95 % of the routes in Costa Rica are radial. The terrain was classified according to the Central American Manual for Geometric Design of Highways and Streets [29] in Level – Rolling (0 % to 10 %) and Rolling – Mountainous (> 10%). Last, the routes were classified by length, in order to differentiate the ones that operate in the inner city, from the ones that reach the outer city or travel between cities. They were classified in short (0 km to 20 km), medium (20 km to 40 km), long (40 km to 125 km) and very long (more than 125 km).

Fig.1 shows the summary of the classification method described above. Location is the first parameter for grouping the routes, then each unit is assorted to a different category according to its characteristics from the subsequent three levels. For example, one of the final groups consists of the units of San José, with radial structure, level – rolling terrain and short length. Not every possible combination was defined as a group, each category was evaluated by size and relevance to decrease the total number of groups.

The classification parameters returned 30 categories from the 742 routes analyzed, some groups consist of less than 10 routes, in contrast with groups with more than 50 routes per category. Therefore, models such as random coefficients models, that controls for between-cluster variation, must be considered [17]. For more details about the classification of bus routes interested readers can check [4].

Once the categories were established, the variable fare per kilometer was selected to model inequities within each group. As a result of exploratory analysis and linear regression models, three independent variables were included in the model:

1. Route length: route length measured in kilometers.
2. Monthly ridership: total average passengers per month for each route.
3. Vehicle occupancy: monthly ridership divided by the total average number of trips per month.

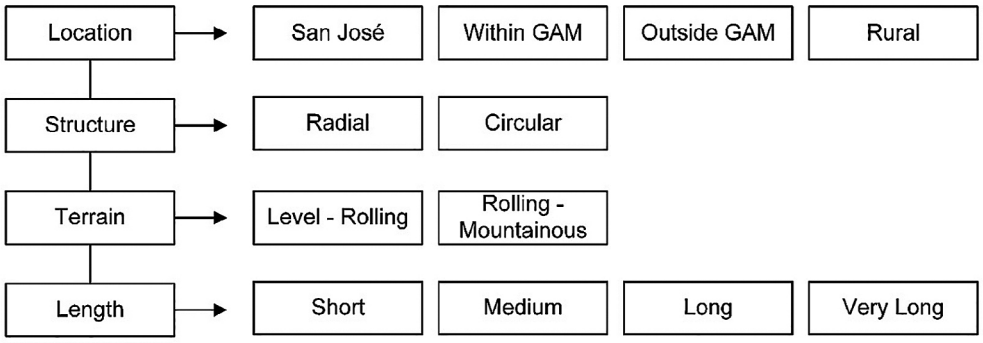


Fig.1. Route classification process

Several other variables can be included in the models to explain the differences in fare per kilometer. In the case of this study, these three operation variables were used as they were available for almost all the bus transit routes in Costa Rica. In addition to the variables used in the categorization of the bus routes, these variables should explain most of the variability in the data. Additional variability in the data should be represented by the excess fare per kilometer or delta, previously introduced in (9). Further analysis of deltas will reflect on the routes with the highest operational costs per user.

From the summary statistics in Table 1, it is observed that monthly ridership has high dispersion, since the standard deviations are higher than the mean frequencies for many of the groups and the total routes. The bus route length also presented high dispersion on the aggregated statistics but the dispersion inside each group is significantly reduced. This was expected since the route length was one of the criteria used to create the groups. The fare per kilometer and the vehicle occupancy presented small variance both at the aggregated level and at the group level.

TABLE 1
SUMMARY STATISTICS OF THE VARIABLES USED IN THE MODEL

Group	Number of routes	Fare per kilometer (₡/km)		route length (km)		Montly ridership		Vehicle Occupancy	
		mean	Std. Dev.	mean	Std. Dev.	mean	Std. Dev.	mean	Std. Dev.
1	5	25.47	11.52	9.25	2.61	142259	64189	41.78	18.51
2	90	20.38	5.62	13.86	3.44	117777	85528	53.19	16.36
3	70	14.07	2.52	26.52	9.57	118206	111803	67.10	18.80
4	17	23.13	8.61	12.43	5.89	25799	21750	32.53	10.87
5	12	24.28	5.45	12.85	3.49	38704	47842	52.25	11.52
6	14	19.88	6.91	36.15	16.55	30285	39705	65.14	20.64
7	58	23.61	9.57	12.05	3.99	61655	74363	48.31	18.62
8	33	13.88	3.87	32.19	11.24	44904	31212	69.62	20.27
9	17	27.62	12.04	14.30	9.20	23253	13090	45.28	16.68
10	9	21.96	7.16	13.88	4.48	42757	25989	39.70	12.10
11	18	13.02	2.82	45.76	10.90	115404	145771	71.54	16.85

12	14	19.54	5.84	70.64	14.79	33413	43441	67.50	14.54
13	17	11.07	1.43	366.66	226.71	15515	20447	68.36	21.17
14	43	10.37	3.71	325.13	151.63	15877	20653	66.05	22.75
15	22	11.33	3.11	163.33	80.24	17259	19551	68.01	25.27
16	9	38.04	14.00	8.59	3.88	28537	23044	33.24	16.43
17	18	33.48	15.80	10.98	4.16	16231	7832	35.59	14.44
18	19	21.96	6.44	27.49	5.28	11626	8234	59.05	17.52
19	15	18.72	6.44	63.12	14.96	15908	21831	69.79	47.64
20	20	24.66	10.58	13.08	4.35	28111	27769	39.72	15.21
21	28	15.52	3.45	29.61	6.55	26294	26560	50.01	24.07
22	39	13.44	3.80	70.59	17.59	14257	22573	71.57	20.42
23	14	9.95	2.96	180.74	81.79	11915	12030	59.28	16.18
24	22	29.53	10.02	13.23	3.52	15072	11445	50.23	27.80
25	25	18.38	7.68	25.13	6.05	14913	17236	44.72	15.14
26	19	15.34	7.42	67.75	26.50	4444	4441	56.91	17.48
27	12	26.11	10.25	11.71	3.69	19975	8077	50.15	48.37
28	17	18.83	11.39	31.40	5.49	14171	16371	60.97	29.94
29	43	14.90	7.28	82.73	46.07	9261	7553	57.98	18.33
30	3	17.99	1.85	61.00	30.99	7747	4227	55.60	18.64
TOTAL	742	18.54	9.33	62.68	105.02	47631	72206	56.97	23.58

5. RESULTS

Models were estimated using the open source software OpenBUGS [30]. OpenBUGS is a popular software used for Full Bayesian estimation in fields ranging from transportation to medicine. For more details on Bayesian estimation using OpenBUGS interested readers can check Lawson et al [31], Congdon [32] or Ntzoufras [33].

For the models, 1.000 iterations were discarded as burn-in. The following 175.000 iterations were used to obtain summary statistics of the posterior distribution of parameters. Convergence was assessed by visual inspection of the Markov chains for the parameters. Furthermore, the number of iterations was selected so that the Monte Carlo error for each parameter in the model would be less than 5 % of the value of the standard deviation of that parameter.

5.1 Independent and correlated or random parameters models

Table 2 presents the estimates and the standard deviation for each coefficient for the independent and correlated random coefficients model. The main interest is to determine which model performs better in terms of goodness-of-fit and coefficient significance.

The goodness-of-fit measures commonly used in full Bayesian statistics are presented in the table: the posterior mean of the deviance and the deviance information criterion (DIC)[26]. The

deviance is estimated in the same way for frequentist and Bayesian statistics, while the DIC is the Bayesian equivalent of the Akaike information criterion. As in the case of their frequentist counterparts, the deviance and the DIC quantify the relative goodness-of-fit of the models; therefore, they are useful for comparing models.

Regarding the results in Table 2, the estimates, in both the independent and the correlated models, corresponding to the independent variables have a negative effect over the fare per kilometer. As expected, when the route length (β_1), average occupancy (β_2) and monthly ridership (β_3) increase, the fare per kilometer decreases. According to (3) , the coefficients of the variables correspond to the constant elasticities since they relate the natural logarithm of the fare per kilometer with the natural logarithm of the covariates. Therefore, it is observed that route length has higher influence over the fare per kilometer than any other independent variable, since in general it has the highest coefficents.

Concerning the estimates significance, both models present a similar behavior. The random coefficients model performs slightly better with the route length estimates, which, as mentioned previously, have more influence in the general model. On the other hand, β_2 and β_3, are significant just in one more group in the independent model compared to the correlated or random parameters model.

The main difference between models can be observed at the estimates for the mean and standard deviation. The random coefficients model return values that are closer to the mean, as they consider the whole sample for modelling and not just each group sample. For the same reason, the standard deviation is also reduced, resulting in a better estimation of the fare per kilometer. In terms of the goodness-of-fit measures, the DIC for the random parameters model is 22 points lower, which means that the model is s significantly better [26].

From Table 2 it is observed that the estimates for groups 13, 23 and 30 are not significant. Even though, these categories have a small sample size (17, 14 and 3 respectively) the groups 1, 10 and 15 have also small sample sizes (5, 9, and 22 respectively) but they return one or more significant variables. The random coefficients model, as mentioned before, has a stronger influence in reducing the small sample groups mean and standard deviation.

The estimated kernel density of the posterior distribution of the parameters further shows the benefits of random coefficients models over univariate models. The estimated density of β_0 for the groups with the largest and the smallest sample size are presented in Fig. 2. The density in group 2, with a sample size of 90, shows almost no changes between the univariate and the multivariate model. On the other hand, the density in Group 30, with a sample size of 3, shows how by considering the population and not just the within-group variation sample, the multivariate model improves the model prediction compared to the univariate model, particularly for groups with small samples.

TABLE 2
INDEPENDENT AND RANDOM COEFFICIENTS MODEL ESTIMATES

Group	Independent				Random Coefficients Model			
	Estimate (standard deviation)							
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
1	10.650(4.067)	-0.689(0.920)	-0.244(0.603)	-0.436(0.635)	5.757(1.238)	-0.704(0.347)	-0.117(0.165)	-0.056(0.095)
2	5.110(0.422)	-0.704(0.111)	-0.045(0.100)	-0.011(0.036)	5.008(0.412)	-0.676(0.106)	-0.040(0.084)	-0.010(0.033)
3	4.794(0.583)	-0.323(0.126)	-0.266(0.131)	-0.001(0.036)	4.469(0.563)	-0.315(0.123)	-0.164(0.107)	-0.012(0.032)
4	4.169(0.898)	-0.702(0.171)	0.031(0.209)	0.049(0.097)	4.229(0.712)	-0.638(0.143)	0.005(0.128)	0.037(0.065)
5	5.638(1.255)	-0.279(0.310)	-0.288(0.456)	-0.064(0.087)	4.689(0.817)	-0.282(0.251)	-0.060(0.165)	-0.059(0.058)
6	6.600(1.152)	-0.402(0.234)	-0.283(0.228)	-0.116(0.085)	5.451(0.980)	-0.286(0.188)	-0.165(0.141)	-0.089(0.057)
7	5.398(0.381)	-0.680(0.110)	-0.076(0.092)	-0.034(0.029)	5.276(0.370)	-0.658(0.106)	-0.066(0.081)	-0.031(0.027)
8	4.473(0.732)	-0.233(0.145)	-0.235(0.124)	-0.010(0.040)	4.106(0.687)	-0.217(0.139)	-0.149(0.103)	-0.015(0.036)
9	4.043(1.313)	-0.602(0.119)	0.200(0.221)	-0.006(0.083)	4.332(0.860)	-0.545(0.112)	0.081(0.137)	-0.005(0.061)
10	5.824(1.569)	-0.879(0.350)	0.112(0.586)	-0.089(0.243)	5.092(0.963)	-0.683(0.253)	-0.022(0.166)	-0.021(0.085)
11	5.958(2.631)	-0.726(0.285)	-0.103(0.438)	-0.022(0.060)	4.390(1.366)	-0.514(0.241)	0.002(0.161)	0.009(0.038)
12	6.805(1.731)	-0.229(0.400)	-0.394(0.341)	-0.131(0.057)	5.029(1.363)	-0.161(0.291)	-0.100(0.161)	-0.105(0.048)
13	1.664(1.391)	0.055(0.152)	0.029(0.188)	0.033(0.089)	1.825(1.110)	0.045(0.124)	0.024(0.123)	0.024(0.061)
14	4.378(0.801)	-0.21(0.085)	-0.222(0.126)	0.002(0.040)	3.980(0.740)	-0.193(0.083)	-0.137(0.103)	-0.004(0.036)
15	3.184(1.084)	-0.043(0.122)	0.039(0.188)	-0.081(0.064)	2.983(0.906)	-0.031(0.118)	0.010(0.125)	-0.052(0.052)
16	3.554(1.780)	-0.529(0.322)	-0.200(0.286)	0.177(0.197)	4.425(0.910)	-0.577(0.207)	-0.049(0.133)	0.049(0.082)
17	8.172(1.533)	-0.816(0.168)	-0.109(0.179)	-0.259(0.138)	5.841(0.943)	-0.688(0.148)	-0.067(0.125)	-0.063(0.08)
18	7.456(1.443)	-0.491(0.336)	-0.121(0.226)	-0.253(0.100)	5.470(1.150)	-0.272(0.271)	-0.096(0.138)	-0.125(0.071)
19	4.855(1.588)	-0.189(0.300)	0.042(0.167)	-0.156(0.047)	4.267(1.231)	-0.109(0.249)	0.032(0.118)	-0.122(0.043)
20	5.428(1.160)	-0.806(0.150)	-0.137(0.180)	0.022(0.069)	4.912(0.821)	-0.739(0.145)	-0.064(0.121)	0.031(0.053)
21	4.640(0.996)	-0.386(0.245)	0.041(0.094)	-0.080(0.052)	4.108(0.886)	-0.277(0.219)	0.022(0.081)	-0.056(0.044)
22	5.619(0.951)	-0.433(0.171)	-0.190(0.130)	-0.049(0.039)	4.947(0.853)	-0.358(0.161)	-0.112(0.105)	-0.046(0.035)
23	3.346(2.103)	-0.128(0.257)	-0.147(0.279)	0.018(0.061)	2.530(1.450)	-0.057(0.216)	-0.026(0.147)	0.014(0.051)
24	4.177(0.752)	-0.510(0.202)	0.084(0.118)	0.014(0.064)	4.137(0.653)	-0.424(0.184)	0.058(0.096)	0.005(0.053)
25	6.351(0.759)	-0.794(0.187)	-0.249(0.179)	-0.008(0.048)	5.710(0.709)	-0.705(0.176)	-0.129(0.126)	-0.018(0.039)
26	12.28(1.316)	-0.989(0.189)	-0.984(0.211)	-0.205(0.090)	8.601(1.134)	-0.838(0.175)	-0.410(0.185)	-0.112(0.069)
27	7.539(2.050)	-0.329(0.337)	-0.174(0.121)	-0.300(0.243)	5.101(0.936)	-0.431(0.255)	-0.141(0.099)	-0.039(0.091)
28	6.077(1.546)	-0.044(0.370)	-0.068(0.177)	-0.320(0.078)	4.917(1.223)	0.052(0.294)	-0.116(0.124)	-0.207(0.065)
29	7.492(0.731)	-0.585(0.095)	-0.218(0.132)	-0.172(0.055)	6.687(0.665)	-0.564(0.093)	-0.141(0.108)	-0.126(0.049)
30	-88.810(82.580)	7.215(6.509)	19.820(18.230)	-1.726(1.755)	2.319(1.495)	0.063(0.204)	0.020(0.177)	0.027(0.090)
Goodness-of-fit measures								
	$\overline{D}$	$\hat{D}$	DIC	p_D	$\overline{D}$	$\hat{D}$	DIC	p_D
	123	2.79	243.3	120.2	131.4	41.42	221.4	89.98

Note: Gray cells indicate significance at 97.5% level.

The estimated kernel density of the posterior distribution of the parameters further shows the benefits of random coefficients models over univariate models. The estimated density of β_0 for the groups with the largest and the smallest sample size are presented in Fig. 2. The density in group 2, with a sample size of 90, shows almost no changes between the univariate and the multivariate model. On the other hand, the density in Group 30, with a sample size of 3, shows how by considering the population and not just the within-group variation sample, the multivariate model improves the model prediction compared to the univariate model, particularly for groups with small samples.

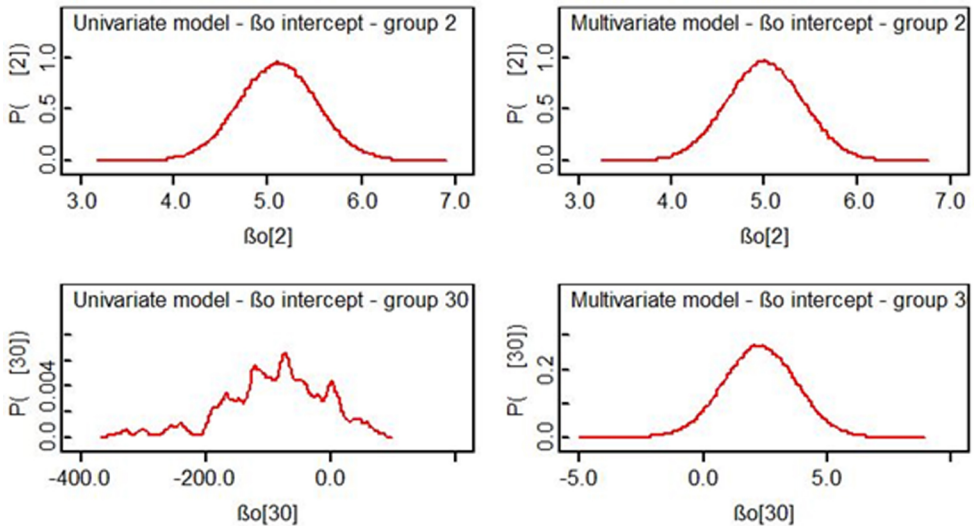


Fig. 2. Deltas ranking comparison between the univariate model and the multivariate model

5.2 Significant Deltas Obtained from the Independent and Random Coefficients Models

Bayesian modeling allows to estimate the significance of the deltas obtained after running both models. Table 3 shows the 5 % of the higher deltas obtained for all the population in the random coefficients model, and as seen, the independent model has seven not significant values, while the correlated model has three. A positive delta indicates a value of fare per kilometer higher than expected, these are the routes that have to be revised first in order to set an adequate fare in terms of operation, equity and the category in which the route is included.

The deltas are ranked from high to low and compared to the ranking obtained in the univariate model. Coincidentally, the highest 5 % deltas in the random parameters model are the same as in the independent model, though they are not in the same order of hierarchy. The delta ranked in the 4th position in the random parameters model serves as an example as how not considering the significance can mislead the interpretation of the results.

Once the outliers are identified, it is necessary to perform individual analysis to each route to identify which are the operational characteristics that have price implications for the served population. This method is intended to be used by the competent authorities as a first approach to perform

a diagnostic of the whole network, so that, the resources for specific audits can be focused to solve the most problematic routes in terms of equity and efficiency.

Ridership and origin-destination studies can be performed to determine the necessity of improving the route's design in terms of structure, configuration, frequency or schedule. The change of this parameter will imply a variation in the fare charged for the service and therefore modify the ridership and average occupancy. The route improvement must be accompanied by model optimization, seeking to conform categories with more homogeneous behavior.

Fig. 3 makes a comparison between both model rankings around the unitary diagonal, showing that the models present slight differences in the deltas ranking, as a considerable number of points are close to the diagonal; nevertheless, several values, specifically those of lower ranking for the univariate model are far from the unitary reference.

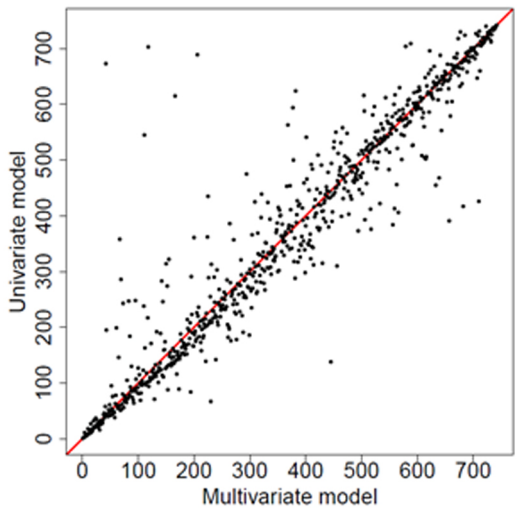


Fig. 3. Deltas ranking comparison between the univariate model and the multivariate model

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

- Univariate and multivariate models were estimated using a full Bayes hierarchical approach. The models estimated the effect of route length, average occupancy and monthly ridership in fare per kilometer, for 742 clustered routes of the bus network in Costa Rica. The estimates in both models have a negative effect over the fare per kilometer. An increase of the independent variables leads to a decrease in the fare per kilometer, being the route length the parameter with the higher influence overall.
- The random coefficients (correlated) model performed better than the univariate model in terms of goodness-of-fit statistics; the DIC for the multivariate model was 22 points lower, which means that the multivariate model is significantly better. The random coefficients model estimates usually returned smaller standard deviations, because it considers the whole sample for modelling and not just each group sample, resulting in a better estimation of the fare per kilometer.

TABLE 3
TOP 5 % OF THE DELTAS OBTAINED FROM THE INDEPENDENT

		Independent			Random Coefficient		
ID	**Group**	**Ranking**	**Estimate**	**sd**	**Ranking**	**Estimate**	**Sd**
delta[680]	28	1	27.63	2.49	1	29.99	1.94
delta[230]	7	2	25.00	7.04	2	27.05	6.33
delta[602]	24	3	21.29	3.00	3	21.21	2.67
delta[449]	17	11	14.35	12.32	4	20.32	10.96
delta[603]	24	4	20.69	2.64	5	20.28	2.34
delta[397]	14	5	19.43	0.93	6	19.63	0.87
delta[675]	27	8	15.68	3.65	7	18.61	2.08
delta[306]	9	6	16.96	6.57	8	18.04	6.30
delta[698]	29	7	16.88	1.10	9	17.55	0.95
delta[450]	17	18	12.06	4.21	10	16.44	2.70
delta[682]	28	9	15.52	1.73	11	15.69	1.56
delta[697]	29	17	12.07	2.15	12	14.21	1.80
delta[617]	24	14	12.36	4.42	13	14.06	3.83
delta[649]	26	10	15.45	1.26	14	14.05	1.32
delta[681]	28	27	9.068	4.21	15	13.81	3.06
delta[699]	29	12	12.74	1.04	16	13.15	0.92
delta[624]	25	13	12.42	1.42	17	12.66	1.22
delta[604]	24	15	12.25	2.14	18	12.22	1.99
delta[521]	21	19	11.72	1.67	19	11.93	1.56
delta[651]	26	32	6.896	2.67	20	11.47	1.81
delta[467]	18	22	10.51	2.55	21	11.34	2.04
delta[441]	16	20	10.97	3.51	22	10.65	3.36
delta[2]	1	36	2.653	6.66	23	10.52	3.76
delta[440]	16	37	-2.51	18.15	24	10.4	11.16
delta[207]	6	35	2.759	8.46	25	10.25	5.67
delta[452]	17	26	9.509	2.81	26	10.13	2.46
delta[233]	7	23	9.916	1.17	27	10.08	1.13
delta[345]	12	21	10.66	1.32	28	10.07	1.33
delta[701]	29	30	7.864	2.19	29	9.943	1.77
delta[702]	29	24	9.729	1.49	30	9.48	1.44
delta[670]	27	16	12.23	4.36	31	9.346	2.98
delta[605]	24	25	9.687	2.89	32	9.092	2.53
delta[668]	27	33	5.277	4.22	33	8.851	3.23
delta[210]	7	29	8.671	0.85	34	8.544	0.86
delta[703]	29	31	7.849	1.03	35	8.355	0.99
delta[488]	19	28	8.764	1.39	36	8.329	1.32
delta[650]	26	34	3.45	2.45	37	8.235	1.85

Note: Gray cells indicate significance at 97.5% level.

- The classification of the data in groups enables a better prediction of the within-group mean, standard deviation and outliers. Outliers allow to identify routes that behave significantly different from those in which they are grouped; therefore, individual analyses should be performed to determine the operation parameters that need to be adjusted to improve performance. Positive deltas, that show the difference between the observed values and the estimated ones, indicate a value of fare per kilometer higher than expected, hence, these are the routes that should be revised first in order to set an adequate fare in terms of efficiency and equity.
- Even though fare is just one of many parameters in a transit system, it has a direct effect on the users in terms of equity, and as stated by Ling [13], improving the fare structure can result in increasing ridership and revenue. With better data availability, similar analyses can be conducted in different bus transit networks as a tool for evaluating the efficiency of each system. Other parameters can be introduced as independent variables for determining specific efficiency aspects.

ACKNOWLEDGEMENTS

The Authors thank ARESEP (Regulatory Authority of Public Services) for providing data on bus public transportation routes.

REFERENCES

[1] T. Litman, *Evaluating transportation equity - guidance for incorporating distributional impacts in transportation planning.* Victoria Transport Policy Institute, Victoria, BC, 2016.

[2] Programa de Investigación en Desarrollo Urbano Sostenible. *Elaboración de auditorías de demanda y cálculo de indicadores y parámetros operativos del servicio de transporte remunerado de personas, modalidad autobús.* Universidad de Costa Rica, San José, Costa Rica, 2014.

[3] Programa de Investigación en Desarrollo Urbano Sostenible. *Elaboración de auditorías de demanda y cálculo de indicadores y parámetros operativos del servicio de transporte remunerado de personas, modalidad autobús que brindan el servicio en los corredores San José* – Heredia y San José-Moravia. Universidad de Costa Rica, San José, Costa Rica, 2015.

[4] C.S. Quirós-Calderón and J. Aguero-Valverde, "Clasificación de las rutas de la red de transporte público modalidad autobús de Costa Rica", Revista Ingeniería, Vol. 28, no. 2, pp.: 112-136, 2018.

[5] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin., *Bayesian Data Analysis,* Chapman & Hall/ CRC Texts in Statistical Science, 2003.

[6] P. Congdon, Bayesian statistical modelling. Chichester, UK: John Wiley & Sons, 2007.

[7] AASHTO, *Highway Safety Manual*. Washington DC, USA. 2010.

[8] E. Hauer, "Identification of sites with promise", *Transportation Research Record,* 1542, pp. 54-60. , 1996

[9] S. Banerjee, B.P. Carlin, A.E. Gelfand, *Hierarchical modeling and analysis for spatial data.* Chapman & Hall/CRC, Florida, 2004.

[10] M. Garrett, and B. Taylor, "Reconsidering social equity in public transit", *Berkeley Planning Journal,* 13.1, pp. 6-27, 1999.

[11] C. Nuworsoo, A. Golub and E. Deakin, "Analyzing equity impacts of transit fare changes: Case study of Alameda–Contra Costa Transit, California", *Evaluation and Program Planning*, vol. 32, no 4, pp. 360-368, 2009. https://doi.org/10.1016/j.evalprogplan.2009.06.009.

[12] R. Cervero, "Flat versus differentiated transit pricing: what's a fair fare?", *Transportation*, vol. 10, no 3, pp. 211-232, 1981..

[13] J. H. Ling, "Transit fare differentials: A theoretical analysis", *Journal of advanced transportation*, vol. 32, no 3, pp. 297-314, 1998. http://dx.doi.org/10.1002/atr.5670320304.

[14] B.Z. Liu, Y.E. Ge, K. Cao, X Jiang et al., "Optimizing a desirable fare structure for a bus-subway corridor", *PLoS ONE*, vol 12, no 10, 2017. e0184815.

[15] C. Tang, A.A. Ceder, and Y.E. Ge, "Integrated optimization of bus line fare and operational strategies using elastic demand", *Journal of Advanced Transportation,* 2017. https://doi.org/10.1155/2017/7058789

[16] M. Karlaftis, and P. McCarthy, "Cost structures of public transit systems: a panel data analysis", *Transportation Research Part E: Logistics and Transportation Review*, vol. 38, no 1, pp. 1-18, 2002.. https://doi.org/10.1016/S1366-5545(01)00006-0.

[17] N.T. Longford, "Logistic regression with random coefficients", *Computational Statistics & Data Analysis*, vol. 17, no 1, pp. 1-15, 1994. https://doi.org/10.1016/0167-9473(92)00062-V.

[18] R. Cervero, and C. D. Kang. "Bus rapid transit impacts on land uses and land values in Seoul, Korea" *Transport Policy,* vol. 18, no 1, pp. 102-116, 2011. https://doi.org/10.1016/j.tranpol.2010.06.005.

[19] N. Yavuz, E. Welch and P. Sriraj, "Individual and neighborhood determinants of perceptions of bus and train safety in Chicago, Illinois: Application of hierarchical linear modeling", *Transportation Research Record: Journal of the Transportation Research Board*, no 2034, pp. 19-26, 2007. https://doi.org/10.3141/2034-03.

[20] A. Paez, and R. G. Mercado, "Mobility of Canadian Elderly: Multilevel Analysis of Distance Traveled in the Hamilton Census Metropolitan Area, Ontario, Canada". Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.

[21] C. Wang, M. Quddus, M. Enoch, T. Ryley and L. Davison, "Multilevel modelling of Demand Responsive Transport (DRT) trips in Greater Manchester based on area-wide socio-economic data", Transportation, vol. 41, no 3, pp. 589-610, 2014.. https://doi.org/10.1007/s11116-013-9506-1.

[22] T.Y. Ma, and J. P. Lebacque. "Dynamic system optimal routing in multimodal transit network." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2351, 2013, pp. 76-84, 2013.. https://doi.org/10.3141/2351-09.

[23] N.M. Laird, and J. H. Ware, "Random-effects models for longitudinal data", *Biometrics*, pp. 963-974, 1982. https://doi.org/10.2307/2529876.

[24] J. Aguero-Valverde, and P. Jovanis, “Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures”, *Transportation Research Record: Journal of the Transportation Research Board*, no. 2165, pp. 21-32, 2010. https://doi.org/10.3141/2165-03.

[25] K.R. Koch, *Introduction to Bayesian statistics*. Springer Science & Business Media, 2007.

[26] D.J. Spiegelhalter, N.G. Best, B.P. Carlin and A. Van Der Linde, “Bayesian measures of model complexity and fit”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 4, p. 583-639, 2002. https://doi.org/10.1111/1467-9868.00353.

[27] A. Molinero, and L. Sánchez Arellano, *Transporte público.* Universidad Autónoma del Estado de México, Ciudad de México, 1997.

[28] Vuchic, V., *Urban transit systems and technology.* John Wiley & Sons, Hoboken, NJ, 2005.

[29] Secretaría de Integración Económica Centroamericana (SIECA), *Manual Centroamericano de Normas para el Diseño Geométrico de las Carreteras Regionales con enfoque de Gestión de Riesgo y Seguridad Vial.* 3ª Edición. Ciudad de Guatemala, Guatemala, 2011.

[30] A. Thomas, B. O’Hara, U. Ligges and S. Sturtz. “Making BUGS open”, *R news*, vol. 6, pp. 12-17, 2006.

[31] A.B. Lawson, W. J. Browne and C. L. V. Rodeiro. *Disease mapping with WinBUGS and MLwiN.* John Wiley & Sons, Chichester, UK, 2003.

[32] P. Congdon, “The Basis for, and Advantages of, Bayesian Model Estimation via Repeated Sampling”, in *Applied Bayesian Modelling.* John Wiley & Sons, Chichester, UK, 2003, pp. 1-30.

[33] I. Ntzoufras, *Bayesian modeling using WinBUGS.* John Wiley & Sons, Hoboken, NJ, 2009.