Montevideo journey-to-work flows, 2016: A doubly constrained gravity model with random effects

Eugenia Riaño
Instituto de Estadística,
Facultad de Ciencias Económicas y de Administración,
UdelaR, Uruguay
E-mail: eugenia@iesta.edu.uy

Antonio Rey
Instituto de Estadística,
Facultad de Ciencias Económicas y de Administración,
UdelaR, Uruguay
E-mail: arey@iesta.edu.uy

Martín Hansz
Instituto de Estadística,
Facultad de Ciencias Económicas y de Administración,
UdelaR, Uruguay
E-mail: martinhansz@gmail.com

Fernando Massa
Instituto de Estadística,
Facultad de Ciencias Económicas y de Administración,
UdelaR, Uruguay
E-mail: fmassa@iesta.edu.uy

In this study, a doubly constrained gravity model was estimated using generalised linear mixed-effects models for journey-to-work flows in Montevideo, Uruguay. Under the mixed-models framework, Poisson and negative binomial regression models were estimated, finding a better fit for the last distribution. The negative binomial distribution used in the model specification improves the parameter estimation by up to 15%. The results are compared with a generalised linear model (GLM) specification, showing that considering regions as fixed effects are insufficient to account for dependence among flows. Modelling spatial autocorrelation by including random effects enables compliance with the no-correlation assumption between residuals while still considering different scenarios for spatial autocorrelation in flows: at the origin, destination, and both. Considering regions of origin and destination as random effects could be a solution for the spatial autocorrelation in flows, and in some cases more complex modelling strategies can be avoided.

Keywords:
mixed models,
flow data,
spatial autocorrelation

Introduction

Transportation modelling is a powerful tool for analysing origin–destination (OD) flows, especially in urban areas. These models are frequently used to analyse the impacts of interventions on transport systems, mainly on infrastructure and changes in regulatory policies, and to analyse changes in demand. The basic structure for transportation planning is based on a four-step model (Ortúzar–Willumsen 2011). According to this model, trips can be explained by four components: trip generation, trip distribution in a system of traffic zones, modal split, and trip assignment.

This study focuses on the initial stages of the model (trip generation and distribution), analysing the case of journey-to-work flows in the city of Montevideo, Uruguay, using data collected by the Mobility Survey of the Metropolitan Area of Montevideo (Mauttone–Hernández 2017) and considering the municipalities as traffic zones.

Gravity models have become the workhorse for modelling flows in an OD matrix. Some researchers, such as Black (1992) and Chun (2008), have shown the existence of spatial network autocorrelation in flow systems for several study cases. The authors highlighted the importance of including spatial autocorrelation in flow data modelling. Nevertheless, in many cases, spatial autocorrelation was not considered in the model strategy.

In recent years, approaches to handle spatial dependence in gravity models have been mainly based on two modelling strategies. The first, is the spatial econometric approach proposed by LeSage–Pace (2008). A linear relationship is assumed between OD flows and explanatory variables through a log-normal transformation. An endogenous term (called spatial lag) is added to the model to obtain a specification similar to that of a SAR model. The analysis estimates the parameters associated with the spatial lag by measuring the strength of the spatial autocorrelation between the OD flows. OD flows are not spatial entities per se, and the spatial weight matrix used in the model must reflect the association of the OD flows based on the neighbourhood matrix of the regions of the OD matrix. There are several ways to define the spatial weight matrix for OD flows using the Kronecker product, by obtaining a conformable matrix according to the OD flow dimension.

The second alternative is the spatial filtering approach proposed by Tiefelsdorf–Griffith (2007). Regarding linear regression models, spatial autocorrelation in model residuals arises from exogenous autocorrelated variables, which are not included in the model. The main objective is to find a set of proxy variables for the omitted variables that ensure that the model residuals behave like white noise. The answer is in the eigenvectors of a transformed spatial weight matrix for OD flows. A selected subset of eigenvectors is incorporated into the model from a search algorithm similar to a stepwise variable selection such that the model residuals become white noise.

Some of the most relevant publications found in the literature review on journey-to-work flows, considering spatial autocorrelation, belong to Griffith–Jones, (1980) and Griffith (2009). In the first study, the authors estimate a constrained gravity model incorporating spatial autocorrelation using spatial filters for 24 urban areas in Canada. In this case, the discussion focuses on the importance of considering autocorrelation in estimating the model for interurban mobility of work travel. The second study analyses journey-to-work flows for 469 German districts. A generalised linear model (GLM) with Poisson distribution is estimated, incorporating constraints and spatial filtering; a better fit is found in the model that includes spatial variables.

Additionally, Farmer (2011) developed a spatial interaction model for intercity work trips to identify regions associated with labour markets in Ireland, observing that the model adjusted with a negative binomial distribution demonstrates better performance than that fitted with a Poisson distribution. The results show that the spatial structure of the OD matrix, unemployment, household size, and education are significant when explaining journey-to-work flows. These cases highlight the importance of spatial autocorrelation modelling in the analysis. From an inferential perspective, the correct specification of the correlation structure between flows plays a significant role because the bias of the parameter estimates and the precision with which these estimates are carried out depends on it (Hawkins et al. 2007, Mets et al. 2017).

This study aims to develop a gravity model that accounts for the dynamics of OD flows corresponding to work trips in Montevideo city, modelling the spatial autocorrelation between flows using generalised mixed models.

The main innovation in this research is the application of generalised mixed models, which are used to solve the spatial autocorrelation problem in gravity models. Griffith–Fischer (2016) propose using random effects to incorporate constraints in the model, but say nothing about solving the spatial autocorrelation problem in this approach. In this study, we show how the regions of origin and destination could be used as random effects and deal with spatial autocorrelation by accomplishing the requirements of the residuals for valid inferences.

Additionally, from a computational point of view, mixed models can be estimated using standard packages in most statistical software. The implementation of spatial econometric models and spatial filtering is still in development for applications to flow data, and can be time-consuming.

The mixed model provides a simple framework for model estimation when estimating constrained models and can resolve the spatial autocorrelation problem.

This study also quantifies the effect of wrongly specifying the probability distribution corresponding to OD flows, and the impact of excluding spatial autocorrelation. Because of the leading role played by the impedance function in these models, the change in the precision with which they are estimated under different specifications, characterised by the previously mentioned methodological decisions, is evaluated.

Methodological approach

Origin–destination notation

Gravity models are widely used in transportation and are applied to other areas, such as migration, foreign trade, and tourism. They are called *gravity models* analogous to Newton's concept of gravity. In its general form, the model is represented as:

$$T_{ij} = A_i \, O_i \, B_j \, D_j \, f\,(d_{ij}) \qquad (1)$$

where T_{ij} is the number of trips between zones *i* and *j;* O_i and D_j represent the origin and destination characteristics reflecting propulsiveness or attractiveness of trips in regions *i* and *j,* and $f\,(d_{ij})$ is the impedance or deterrence function that expresses the cost of travelling between zones *i* and *j* (Griffith–Fischer 2016). The balancing factors A_i and B_j are used to incorporate constraints into the model, so that trip productions are distributed to match the trip attraction distribution and reflect the underlying travel impedance (Mc Nally 2007).

The values of the OD matrix are usually presented in a format similar to that in Table 1, where the marginal totals are also presented ($T_{i.}$ and $T_{.j}$), corresponding to both the origin and the destination.

Table 1

Representation of the origin–destination matrix

		Destination				
		1	2	…	K	Total
Origin	1	T_{11}	T_{12}	…	T_{1K}	$T_{1.}$
	2	T_{21}	T_{22}		T_{2K}	$T_{2.}$
	…	…		…	…	…
	K	T_{K1}	T_{K2}	…	T_{KK}	$T_{K.}$
	Total	$T_{.1}$	$T_{.2}$		$T_{.K}$	$T_{..}$

Constrained gravity model

From equation (1), it is possible to define three subclasses of models:

- Origin constrained
- Destination constrained
- Doubly constrained

This classification obeys the specification of the terms included in A_i and B_j, from which the model reproduces the marginal values of the OD matrix at the origin, destination, or both. The use of constraints depends on the object of the analysis. For instance, if a new industrial centre is developed in a specific area of the city, knowing the number of new jobs that will be generated, a model can be estimated to predict flows from other regions to the region where the new industrial centre is located (Griffith–Fischer 2016). In this type of analysis, a destination-constrained model is the most suitable for estimation. According to Farmer (2011), if the objective of the model is to make predictions in the OD matrix, the most appropriate specification is double-constrained, adjusting both marginals of the OD matrix.

In all these cases, the simplest way to define the factors A_i and B_i is by including dummy variables. In the origin-constrained type, the effect of each zone of the origin is modelled; in the destination-constrained type, the effect of each zone of the destination is modelled, and in the doubly constrained type, both sets of effects are modelled. In this way, following Griffith–Fischer (2016), the predictions of the model always reproduce the total flows in the origin, destination, or both, according to the specified constraints.

Fixed versus random effects

Griffith–Fischer (2016) stated that constrained gravity models can be estimated with fixed or random effects for OD regions. Although several assumptions differentiate these models, the main difference lies in the dependency structure between the flows. If regions are considered fixed effects, the model belongs to the family of GLMs. This assumption can be excessively restrictive because flows with a common end (either at the origin or destination) may present similar characteristics. One way to increase the flexibility of the probabilistic structure of the model is to specify the effects of the regions using random variables. Consequently, it is possible to account for the (possible) covariance structure between flows. This point may be of special importance because the correct characterisation of the covariance parameters can lead to a more precise evaluation of the regression parameters.

Spatial dependence

In cases where the OD flows present spatial autocorrelation, and the model fails to account for this, parameter estimates and their variances may present bias (LeSage–Pace 2009). Consequently, it is convenient to fit the model within a mixed-effects framework. Thus, if there is a spatial autocorrelation structure (either at the origin or destination), it would be captured by the covariance parameters of the random effects.

The spatial autocorrelation of the residuals is computed using the Moran's I Index:

$$MI = \frac{n}{\sum_i \sum_j w_{ij}^{OD}} \frac{\sum_i \sum_j w_{ij}^{OD} \, (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \tag{2}$$

where n is the number of spatial units indexed by i and j. w_{ij}^{OD} are the elements of the spatial weight matrix which is defined as (LeSage–Pace 2009):

$$W^{OD} = W^{O} \otimes W^{D} \tag{3}$$

where

$$W^{O} = W \otimes I_n \tag{4}$$

and

$$W^{D} = I_n \otimes W \tag{5}$$

The W matrix contains the neighbourhood structure of municipalities, ⊗ is the Kronecker product, and I_n is an identity matrix of dimension n x n, where n is the number of municipalities in our case. Consequently, W^{OD}, W^{O} and W^{D} have dimensions N x N, with $N = n^2$.

In this case, W is defined by contiguity (Bivand et al. 2013):

$$w_{ij} = \begin{cases} 1 \; if \; polygon \; shares \; boundaries \\ 0 \; if \; not \end{cases} \tag{6}$$

where w_{ij} is a generic element of the W matrix.

Distributional assumptions

Owing to the multiplicative specification of the model, the first approach to estimate it is to consider the logarithm of the flows T_{ij} adopting a linear specification. Additionally, it is usually assumed that model errors are normally distributed, with zero mean and constant variance (Anderson 1979, Sen–Smith 1995). This specification is the simplest for modelling gravity models. However, this distributional assumption is quite restrictive, as model residuals are usually abnormal and tend to present heteroskedasticity. Furthermore, this specification has the disadvantage of handling empty cells in the OD matrix (Burger et al. 2009).

Gravitational models with a Poisson specification help mitigate these limitations (Flowerdew–Aitkin 1982, Santos Silva–Tenreyro 2006). This probability distribution is more suitable than the normal distribution for modelling counting variables and admits zeroes as possible values in the response variable. However, because the Poisson distribution is characterised by equidispersion, the model fit may be suboptimal if the flows demonstrate overdispersion. This limitation can be overcome by using a negative binomial distribution (Lambert 1992, Greene 1994). The procedure to test overdispersion estimates the scale parameter based on deviance residuals, comparing it with a chi-squared distribution, as shown by Gelman–Hill (2006). In all these cases, the model estimation is usually performed through maximum likelihood.

Using the parameterisation of Cameron–Trivedi (2013) for the negative binomial count model, the gravitational model of mixed effects used in this study is as follows:

$$\begin{gathered} T_{ij} \sim NB(\mu_{ij}, \theta) \\ log(\mu_{ij}) = \beta Intra_{ij} + o_i + d_j + \alpha dist_{ij} \\ o_i \sim N(0, \sigma_o^2) \\ d_j \sim N(0, \sigma_d^2) \end{gathered} \tag{7}$$

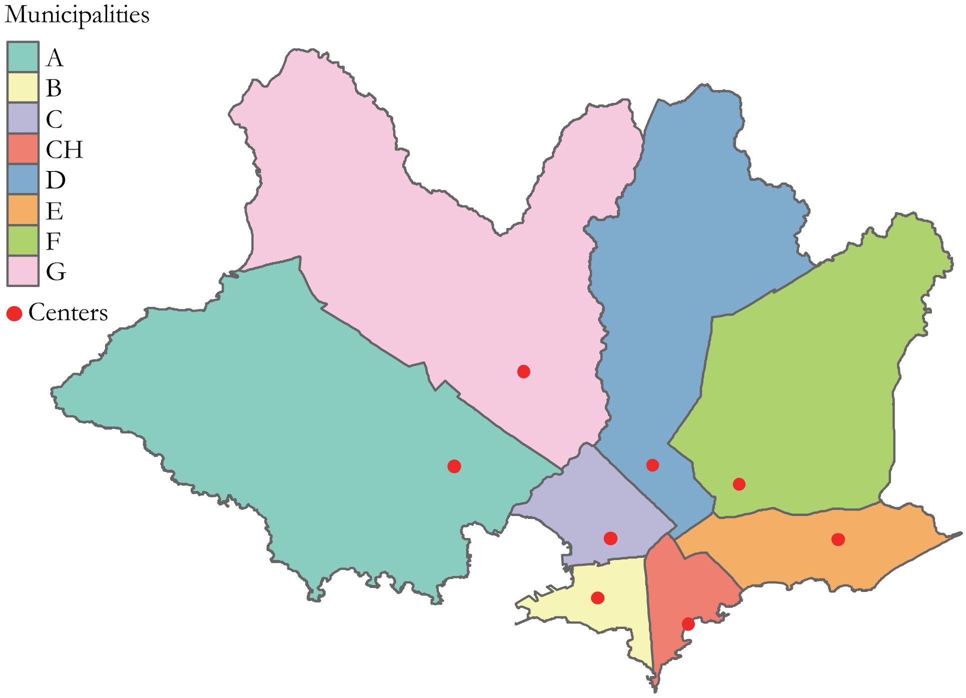
which is a generalised linear mixed-effects model, including random effects (o_i y d_j) for the origin and destination parameters, an intra-regional fixed effect (a dummy variable $Intra_{ij}$ for the diagonal elements of the OD matrix), and $dist_{ij}$ corresponding to a log-linear specification of the impedance function between zones *i* and *j*.

Data

The mobility survey of the metropolitan area of Montevideo was conducted in the second half of 2016. A total of 2,230 households were surveyed, of which 1,069 corresponded to Montevideo and the remaining to the entire metropolitan area. All trips made between 4 a.m. on the previous day and 4 a.m. on the day of the interview were surveyed, establishing the reason for the trip, time and place of departure and arrival, and mode of travel (e.g., buses, private vehicles, and taxis). The surveys were conducted during the week, not considering trips made on Saturdays and Sundays. This analysis is limited to Montevideo city since the sample size in the metropolitan area is small in relation to the number of regions that should be considered to capture its heterogeneity.

Figure 1

Municipalities of Montevideo



The municipalities of Montevideo are presented in Figure 1. In many cases, the geographic centroids coincide with rural areas, with a low population density, mainly in outlying municipalities. To overcome this problem, centres are defined based on the ‚centralities' of the city (GIS – Intendencia de Montevideo 2012). Centralities coincide with areas of high demographic density and commercial concentration in municipalities. Cases of municipalities with two or more ‚centralities' are addressed using an intermediate point (barycenter) between these centres.

As a measure of distance, we use the average travel time (in minutes) between municipalities' centres using public and private transport. The travel time using public transport is calculated using the application ‚Como IR' (Intendencia de Montevideo 2018), and the travel time by car is calculated using the Google Maps website (Google n.d.). The calculations were performed at the same time of the day, without significant traffic variations. The matrix of the average travel time (in minutes) between municipalities is shown in Table 2.

Table 2

Origin–destination matrix of average travel time in Montevideo, 2016

(minutes)

Destination / Origin	A	B	C	CH	D	E	F	G
A	0	34	23	51	33	76	53	29
B	34	0	29	17	29	40	38	49
C	23	29	0	28	15	51	31	26
CH	51	17	28	0	34	43	38	55
D	33	29	15	34	0	38	20	42
E	76	40	51	43	38	0	37	74
F	53	38	31	38	20	37	0	59
G	29	49	26	55	42	74	59	0

Table 3 displays the OD matrix corresponding to the eight municipalities in Montevideo. Given the size of the municipalities, dummy variables are incorporated into the flows corresponding to the diagonal of the OD matrix to account for the effect of intra-municipality flows.

Table 3

Origin–destination matrix for work trips in Montevideo, 2016

Destination / Origin	A	B	C	CH	D	E	F	G	Total
A	60	51	27	39	9	7	6	12	211
B	1	69	10	19	2	5	3	2	111
C	2	28	25	20	1	2	2	3	83
CH	1	50	4	36	3	6	0	3	103
D	3	35	17	39	50	21	17	8	190
E	0	51	11	27	3	18	4	2	116
F	11	37	18	22	15	17	41	3	164
G	10	22	10	6	4	5	3	31	91
Total	88	343	122	208	87	81	76	64	1,069

Modelling strategy

The modelling strategy included the estimation of the generalised linear mixed-effects gravity model with a negative binomial distribution (GLMM-NB), as stated in equation (7), doubly constrained, including random effects at the origin and destination. For validation, three additional model specifications are estimated.

GLMM with Poisson distribution (GLMM-P).

$$T_{ij} \sim Poisson(\mu_{ij}); \log(\mu_{ij}) = o_i + d_j + \alpha dist_{ij} \; ; \; o_i \sim N(0, \sigma_o^2) \; ; \; d_j \sim N(0, \sigma_d^2) \quad (8)$$

Unconstrained GLM with negative binomial distribution (GLM-NB).

$$T_{ij} \sim NB(\mu_{ij}, \theta); \log(\mu_{ij}) = \delta + \alpha dist_{ij} \quad (9)$$

Unconstrained GLM with Poisson distribution (GLM-P).

$$T_{ij} \sim Poisson(\mu_{ij}); \log(\mu_{ij}) = \delta + \alpha dist_{ij} \quad (10)$$

The choice of these alternatives reflects the fact that the Poisson model is simpler than the negative binomial if the equidispersion assumption is fulfilled. The evaluation of this assumption is conducted on the residuals of these models using the procedure described by Gelman–Hill (2006). Additionally, the random effects are included as balancing factors and reflect their ability to capture the autocorrelation between flows, not collected by the regression effects. Finally, the precision of the estimates of fixed effects is evaluated by comparing their standard deviations. Parameter significance is carried out using the likelihood ratio test statistic, and the corresponding p-values are obtained using asymptotic theory. All estimations and tests are performed using R software (R Core Team 2019). Libraries *MASS* (Venables–Ripley 2002), *lme4* (Bates et al. 2015), *performance* (Lüdecke et al. 2020) and *spdep* (Bivand et al. 2013) are used to estimate GLM and GLMM and to conduct the overdispersion and spatial autocorrelation tests.

Results

We examine 64 flows corresponding to the work trips made between the 8 municipalities. The minimum and maximum were 0 and 69 trips, respectively, with an average of 16.703, and a standard deviation of 17.017. Regarding the distance, values were spread over the 14–69 min interval, with an average of 31.923 min and a standard deviation of 11.402.

Table 4 presents the estimates corresponding to the four models, and their standard deviations (in parentheses). Additionally, goodness-of-fit statistics, such as the Akaike and Bayesian information criteria, are presented. Based on the latter, the models specified with a negative binomial distribution present a better fit than those specified by the Poisson distribution. To confirm this, an overdispersion test was carried out using the deviance residues of the GLM-P and GLMM-P models. In both cases, it was concluded that the equidispersion assumption was too restrictive because the p-value of the test was less than 0.001. Regarding the effect of the impedance function, it is noteworthy that the Poisson specification proposes a value that is approximately 15% lower than that obtained by the NB specification. The distance parameter estimates are quite similar between GLM and GLMM specifications, but the standard deviation is 31.5% lower when using GLMM specifications.

Table 4

Parameters estimates and goodness of fit statistics

	GLM		GLMM	
	Poisson	negative binomial	Poisson	negative binomial
	Fixed effects			
Intercept	3.4203 (0.1114)	3.5725 (0.379)***	3.1140 (0.3099)***	3.1938 (0.3941)***
Distance	–0.0277 (0.0037)***	–0.0328 (0.0114)**	–0.0289 (0.0047)***	–0.0337 (0.0078)***
intra	0.2994 (0.1242)*	0.1471 (0.4968)	0.5361 (0.1542)**	0.5066 (0.276)°
	Dispersion parameter			
	–	1.2483	–	8.1237
	Random effects			
σ^2_{origin}	–	–	0.2226	0.2713
$\sigma^2_{destination}$	–	–	0.3780	0.4875
	Information criteria			
AIC	971.3831	481.8747	481.6463	431.4024
BIC	977.8597	490.5103	492.4407	444.3557

Notes: Likelihood ratio test p-values are shown using the significance codes where *** corresponds to a p-value minor or equal to 0.001, ** corresponds to a p-value in the interval (0.001, 0.01), * corresponds to a p-value in the interval (0.01, 0.05), and ° corresponds to a p-value in the interval (0.05, 0.1). The absence of significance codes implies p-values greater than 0.1.

The autocorrelation analysis of the residuals of the four models was conducted, considering the matrices defined by equations (3) to (5), to check whether spatial autocorrelation was removed from the model residuals. Table 5 presents the values of Moran's I statistic, as well as their corresponding p-values.

Table 5

Moran's I test of the study models

	GLM		GLMM	
	Poisson	negative binomial	Poisson	negative binomial
Origin	4.254 (<0.001)	4.388 (<0.001)	0.635 (0.263)	0.277 (0.391)
Destination	6.259 (<0.001)	6.109 (<0.001)	–1.065 (0.857)	–1.545 (0.939)
Origin–destination	3.142 (0.002)	3.617 (<0.001)	–0.081 (0.532)	–0.027 (0.511)

From the last two columns of Table 5, we note that the inclusion of random effects was sufficient to capture the autocorrelation between the flows of the OD matrix.

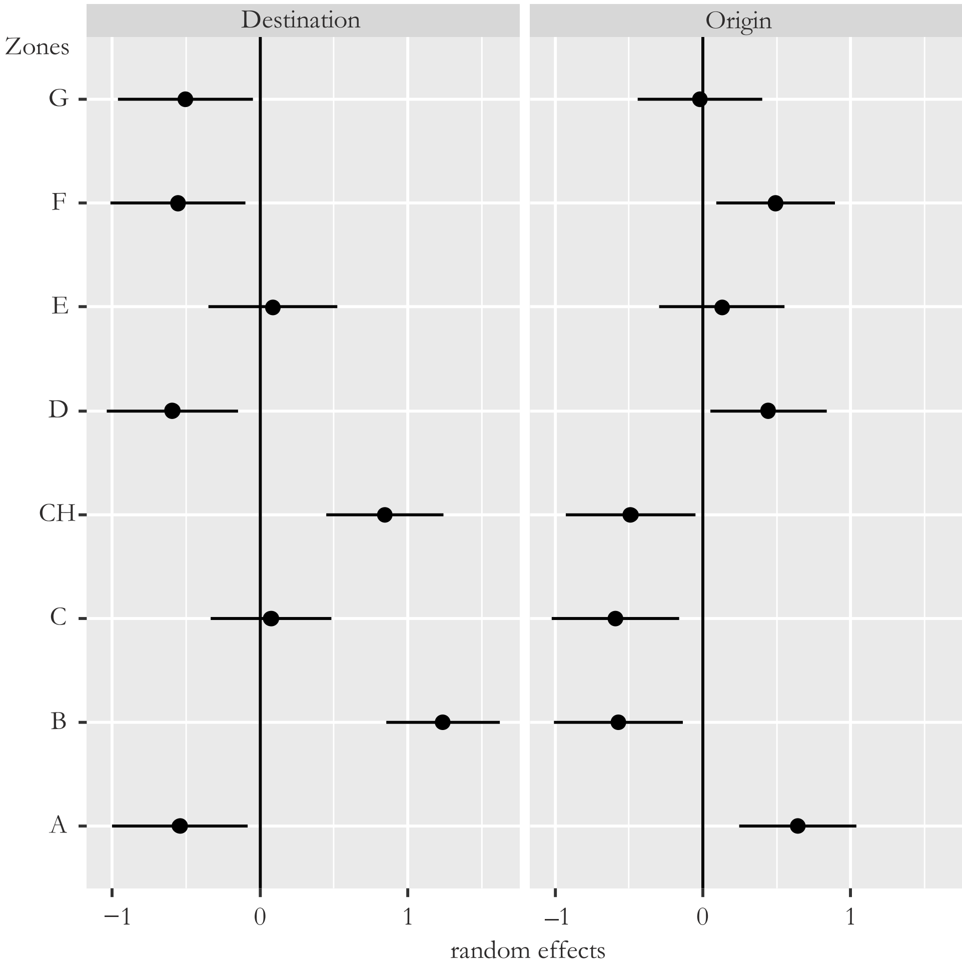
The GLMM-NB model was validated using the data collected by the autocorrelation and overdispersion tests. Furthermore, the inclusion of a scale parameter and the random effects contributed to a considerable decrease in the precision of the estimates corresponding to distance and intra-municipal trips. It is of special interest to determine how the effect of intra-municipality trips increases considerably when using the NB distribution and incorporating random effects.

Analysis of fixed-effect estimates corresponding to the GLMM-NB shows that longer trips are less frequent, and for each extra minute of travel, the number of trips decreased by 3.3%. Additionally, although the significance of the parameter associated with internal trips is barely above 0.05 (Table 4), internal trips are 65% more frequent than trips between municipalities.

Finally, Figure 2 presents the prediction of the random effects corresponding to the final model. The results shown in Figure 2 appropriately summarise the differences between the margins in Table 3, distinguishing municipalities as attractive or repealing. In the left panel, we analyse the flows according to their destinations. Municipalities B and CH are the most attractive, while municipalities A, D, F, and G attract trips below the flow average. This is reasonable because municipalities B and CH have higher concentrations of enterprises. Analysis of the trips according to their origin shows that municipalities A, D, and F could be considered *expelling*. In contrast, work trips with origins in municipalities B, C, and CH are less frequent. Municipalities A, D, and F are peripheral and have a low concentration of enterprises, which means that a large section of the workforce in the area has to move to other municipalities with greater job opportunities, such as municipalities B and CH. The effects in the origin and destination of municipality E are not significant, so it is impossible to define it as *attractive* or *expelling*. The same is true for municipality G at the origin and municipality C at the destination. These three cases present certain characteristics that can explain the results, as they are examples of *opposing* forces. Due to its location, Municipality C, is very attractive to outlying municipalities; it has greater job opportunities than those municipalities, but in turn, expels trips to municipalities B and CH. Municipality E has job opportunities similar to municipality C. However, its location makes it unattractive for the North West and West municipalities, attracting more trips from the nearest municipalities while expelling trips to municipalities B and CH, the most attractive municipalities in the city. Municipality G has characteristics similar to those of the other outlying municipalities, but it indicates a significant number of intra-municipality trips among the outlying municipalities.

Figure 2

Prediction of random effects in municipalities of Montevideo



Concluding remarks

A generalised linear mixed-effects model was adjusted to work travel between the municipalities of Montevideo, starting from a GLM specification. The model complexity was increased by altering the specification of the distribution of the response variable and accounting for the spatial autocorrelation within the analysis using random effects. It was observed that the presence of overdispersion and spatial autocorrelation caused biases in the estimates and in their precision. The inclusion of random effects allows the capture of spatial autocorrelation, tackling the problems in the literature and providing a simple analytical framework to summarise the main characteristics of municipalities.

Although the municipalities show spatial autocorrelation for journey-to-work flows, they could be considered large regions for trip analysis. The sample size of the Mobility Survey of the Metropolitan Area of Montevideo is too small to increase the number of regions, and the results could not capture trips at a local level. The size of the diagonal values of the OD matrix is large; therefore, we use an indicator variable to capture the effect of intra-regional flows. For other geographical units, spatial autocorrelation in model residuals may remain when random effects are used, in which case, it is necessary to complement this analysis with another approach.

The analysis reveals the importance of the choices made in developing a gravity model and the impact of these choices on the estimates of the parameters corresponding to the fixed effects. The spatial autocorrelation of the flows reported from the OD matrices can provide plausible explanatory power for incorporation into the analysis. The methodological decisions taken throughout the modelling should consider these potential effects.

Travel matrices play a central role in analysing travel flows in a city, and their correct specifications determine their potential as transport planning instruments. The results show that using random effects improves the variance of estimators, leading to a more accurate estimation of the gravity model parameter. This ensures that accounting with a reliable tool for transport policymakers will predict the flows between regions in an OD matrix.

The results obtained from this study provide evidence that space plays a role in determining a model's parameters. When the objective is to estimate a constrained gravity model, it can be handled in a generalised mixed model context.

We suggest five possible directions for future research. The first is the addition of explanatory variables to the model. This seems quite simple, but when dealing with aggregated variables for the geographical unit used in the model, the potential explanatory variables for a journey-to-work model (household income, number of companies, population with higher education) show a high correlation, which is a problem for model estimation. Additionally, it is important to examine whether spatial autocorrelation is associated with some of these variables and their assistance in removing autocorrelation from model residuals.

The second concerns the use of Moran's Index as an autocorrelation measure. Moran's index may not be appropriate when dealing with flows. Other alternatives, such as BiFlowLISA (Tao–Thill 2020) or the application of a dyadic approach (Bavaud et al. 2018), could improve the results for testing the spatial autocorrelation in model residuals.

The third is to explore the existence of separation effects between municipalities, using an extension of the model proposed by Szabó–Sipos (2020) within the mixed model framework. Still considering journey to work flows, it's possible to find resistances to travel to some places, beyond distances between municipalities. In this paper we make the assumption that journey-to-work flows structure is correctly described with a gravity model. For further research it would be interesting use another approaches to model travel behaviour. Varga et al. (2016) propose an improved radiation model, considering traveling cost, to analyse commuting patterns in Hungary. Yang et al. (2010) build a model incorporating spatial factors in the framework of discrete choice modelling. Both methodologies are suitable to test in a mixed model framework to handle spatial autocorrelation in flows.

Finally, sample weights are not considered in the model estimation. The gravity model estimation is based on the OD matrix cell estimates, and it cannot be weighted directly from a household's weights in the sample. Bootstrapping and resampling techniques could be an option to address this issue and improve the estimation of the parameters associated with the impedance function.

Acknowledgements

This research was supported by the National Agency for Research and Innovation (‚Fondo Sectorial de Investigación a partir de Datos') and by CSIC – UDELAR (I+D project ‚Calibración de modelos de interacción espacial para la modelización de la cantidad de Viajes en Montevideo').

REFERENCES

ANDERSON, J. E. (1979): A theoretical foundation for the gravity equation *The American Economic Review* 69(1): 106–116.

BATES, D.–MAECHLER, M.–BOLKER, B.–WALKER, S. (2015): Fitting linear mixed-effects models using lme4 *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01

BAVAUD, F.–KORDI, M.–KAISER, C. (2018): Flow autocorrelation: a dyadic approach *The Annals of Regional Science* 61 (1): 95–111. https://doi.org/10.1007/s00168-018-0860-y

BIVAND, R. S.–PEBESMA, E.–GOMEZ-RUBIO, V. (2013): *Applied spatial data analysis with* R, Second edition. Springer, New York.

BLACK, W. R. (1992): Network autocorrelation in transport network and flow systems *Geographical Analysis* 24 (3): 207–222. https://doi.org/10.1111/j.1538-4632.1992.tb00262.x

BURGER, M. J.–VAN OORT, F. G.–LINDERS, G. J. (2009): On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation *Spatial Economic Analysis* 4 (2): 167–190. https://doi.org/10.1080/17421770902834327

CAMERON, A. C.–TRIVEDI, P. (2013): *Regression analysis of count data* Cambridge University Press, New York.

CHUN, Y. (2008): Modeling network autocorrelation within migration flows by eigenvector spatial filtering *Journal of Geographical Systems* 10: 317–344. https://doi.org/10.1007/s10109-008-0068-2

FARMER, C. (2011): *Commuting flows and local labor markets: Spatial interaction modeling of travel-to-work* PhD thesis, National University of Ireland Maynooth, Ireland.

FLOWERDEW, R.–AITKIN, M. (1982): A method of fitting the gravity model based on the Poisson distribution *Journal of Regional Science* 22 (2): 191–202. https://doi.org/10.1111/j.1467-9787.1982.tb00744.x

GELMAN, A.–HILL, J. (2006): *Data analysis using regression and multilevel / hierarchical models* Cambridge University Press, Cambridge.

GRIFFITH, D. A.–JONES, K. G. (1980): Explorations into the relationship between spatial structure and spatial interaction *Environment and Planning A* 12 (2): 187–201. https://doi.org/10.1068/a120187

GRIFFITH, D. A. (2009): Modeling spatial autocorrelation in spatial interaction data: Empirical evidence from 2002 journey-to-work flows *Journal of Geographical Systems* 11 (2): 117–140. https://doi.org/10.1007/s10109-009-0082-z

GRIFFITH, D. A.–FISCHER, M. M. (2016): Constrained variants of the gravity model and spatial dependence: Model specification and estimation issues. In: PATUELLI, R.–ARBIA, G. (eds.): *Spatial Econometric Interaction Modelling* pp 37–66., Advances in Spatial Science, Springer International Publishing, Switzerland.

HAWKINS, B.–DINIZ-FILH, J. A.–BINI, L.–DE MARCO, P.–BLACKBURN, T. (2007): Red herrings revisited: Spatial autocorrelation and parameter estimation in geographical ecology *Ecography* 30 (2): 375–384. https://doi.org/10.1111/j.0906-7590.2007.05117.x

LAMBERT, D. (1992): Zero-inflated Poisson regression with an application to defects in manufacturing *Technometrics* 34 (1): 1–14. https://doi.org/10.1080/00401706.1992.10485228

LESAGE, J. P.–PACE, R. K. (2008): Spatial econometric modeling of origin-destination flows *Journal of Regional Science* 48 (5): 941–967. https://doi.org/10.1111/j.1467-9787.2008.00573.x

LESAGE, J. P.–PACE, R. K. (2009): *Introduction to spatial econometrics* Taylor and Francis, Boca Raton, FL, London and New York.

MC NALLY, M. G. (2007): The four step mode.l In: HENSHER, D. A.–BUTTON, K. J. (eds.): *Handbook of Transport Modeling* pp. 35–53., Pergamon, Amsterdam.

METS, K. D.–ARMENTERAS, D.–DÁVALOS, L. (2017): Spatial autocorrelation reduces model precision and predictive power in deforestation analyzes *Ecosphere* 8 (5): e01824. https://doi.org/10.1002/ecs2.1824

ORTÚZAR, J.–WILLUMSEN, L. (2011): *Modeling transport* Wiley & Sons. New York.

R CORE TEAM (2019): *A language and environment for statistical* Computing, Vienna, Austria.

SANTOS SILVA, J. M.–TENREYRO, S. (2006): The log of gravity *The Review of Economics and Statistics* 88 (4): 641–658. https://doi.org/10.1162/rest.88.4.641

SEN, A.–SMITH, T. (1995): *Gravity models of spatial interaction behavior* Springer – Verlag, Berlin–Heidelberg.

SZABÓ, Z.–SIPOS, T. (2020): Separation effects in a microregion: traffic volume estimation between the settlements of Lake Velence *Regional Statistics* 10 (2): 186–205. https://doi.org/10.15196/RS100208

TAO, R.–THILL, J. C. (2020): BiFlowLISA: Measuring spatial association for bivariate flow data *Computers, Environment and Urban Systems* 83: 101519. https://doi.org/10.1016/j.compenvurbsys.2020.101519

TIEFELSDORF, M.–GRIFFITH, D. A. (2007): Semiparametric filtering of spatial autocorrelation: The eigenvector approach *Environment and Planning A* 39 (5): 1193–1221. https://doi.org/10.1068/a37378

VARGA, L.–TÓTH, G.–NÉDA, Z. (2016): An improved radiation model and its applicability for understanding commuting patterns in Hungary *Regional Statistics* 6 (2): 27–38. https://doi.org/10.15196/RS06202

VENABLES, W. N.–RIPLEY, B. D. (2002): *Modern applied statistics with S* Fourth Edition. Springer, New York.

YANG, M.–WANG, W.–CHEN, X.–WANG, W.–XU, R.–GU, T. (2010): Modeling destination choice behavior incorporating spatial factors, individual sociodemographics, and travel mode *Journal of Transportation Engineering* 136 (9): 800–810. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000140

INTERNET SOURCES

GIS – INTENDENCIA DE MONTEVIDEO (2012): https://geoweb.montevideo.gub.uy/geonetwork/srv/es/metadata.show?uuid=09a7b076-dea2-4f57-a143-ca9140337f05 (downloaded: April 2021).

GOOGLE (n.d.): *Google Maps directions for driving between Montevideo's municipalities centroids* (downloaded: November 2018).

GREENE, W. H. (1994): Accounting for excess zeros and sample selection in Poisson and negative binomial regression models *NYU Working Paper No. EC-94-10.* https://ssrn.com/abstract=1293115 (downloaded: April 2021).

INTENDENCIA DE MONTEVIDEO (2018): *„Cómo ir" application* https://montevideo.gub.uy/aplicacion/como-ir (downloaded: April 2021).

LÜDECKE, D.–MAKOWSKI, D.–WAGGONER, P.–PATIL, I. (2020): *Assessment of regression models performance* CRAN. https://easystats.github.io/performance/ (downloaded: April 2021).

MAUTTONE, A.–HERNÁNDEZ, D. (2017): *Mobility survey of the metropolitan area of Montevideo. Main results and indicators* (report). Montevideo: CAF, Municipality of Montevideo, Municipality of Canelones, Municipality of San José, Ministry of Transport and Public Works, University of the Republic, UNDP Uruguay. http://scioteca.caf.com/handle/123456789/1078 (downloaded: April 2021).