Modelling shared e-scooters: A spatial regression approach

Daniel J. Reck[a*], Sergio Guidon[ab], Kay W. Axhausen[a]

[a] *Institute for Transport Planning and Systems (IVT), ETH Zürich, Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland.*
[b] *Institute of Science, Technology and Policy (ISTP), ETH Zürich, Universitätstrasse 41, 8092 Zürich, Switzerland.*
[*] *Corresponding author (reckd@ethz.ch).*

Abstract

Shared e-scooters have appeared quickly and in large quantities, yet little is known about their use. In this study, we explore spatial drivers of demand for shared e-scooter trips in Louisville (KY). We estimate a generalized linear mixed model with conditionally autoregressive random effects using 15 months of booking data, points of interests from Open Street Maps and US census data. We find that population density, the presence of bikeways and university campuses have the strongest positive effect on shared e-scooter trip destination counts. We find a significant, yet less substantial positive effect of bus stops suggesting some first/last mile use and hypothesize tourists to be an overlooked, yet important segment in shared e-scooter demand.

Word Count (below line, excluding references): 2711 (+ 2 figures and 2 tables)

Keywords: shared e-scooters, micromobility, transport demand modelling, spatial regression

Introduction

Shared e-scooters have surprised many with their sudden and plentiful appearance. In 2018, just one year after their introduction, ridership in the US alone already surpassed 38M (NACTO 2019). Despite the popularity these numbers indicate, shared e-scooters have sparked heated debates between citizens, municipal governments and suppliers about road and curb use, safety and social equity.

Research to guide policy-making, however, is still in its infancy. This holds particularly true for spatial aspects of shared e-scooter trips: Where, how and why are they being used? Providing rigorous answers to these questions can support transport planning and regulation in various ways, such as informing the extensions of bikeway networks, identifying suitable locations for parking corrals and predicting future demand.

In this paper, we first review the extant literature on the spatio-temporal use of shared e-scooters. We then analyze spatial drivers of demand using a Negative Binomial-distributed generalized linear mixed model (GLMM) with a random effect following a conditional autoregressive (CAR) correlation model on e-scooter trip destination count data in Louisville (KY). We close with a discussion of our findings in the context of the broader literature on bike sharing to identify similarities and differences.

Our contributions are twofold. First, we identify spatial drivers of demand using a dataset that has not yet been used before, thus offering lessons on the generalizability of results in comparison to previous studies. Second, we estimate and compare several Negative Binomial-distributed non-spatial and spatial generalized linear (mixed) models. This is novel as most previous papers modeling spatial demand of shared e-scooters either focus on descriptive analyses (Espinoza *et al.*, 2020; McKenzie, 2019), use (non-spatial) linear regression models (Bai and Jiao, 2020; Hawa *et al.*, 2020) or spatial linear regression models assuming normally distributed residuals (Arnell *et al.*, 2020; Caspi *et al.,* 2020; Zuniga-Garcia and Machemehl, 2020) – an assumption that does not hold for (non-negative) count data.

Literature review

Several authors have started to analyze spatial drivers of shared e-scooter demand using a variety of methods. In this Section, we review these contributions grouped by method. First, we summarize studies giving descriptive overviews only. We proceed with studies using (non-spatial) linear regression models and finally summarize studies using spatial linear regression models.

McKenzie (2019) analyzed the spatio-temporal use of shared e-scooters in Washington, D.C. Using 3½ months of trip data accessed at a 5-min temporal resolution from the openly accessible API, he found shared e-scooter trips to exhibit a mid-day peak and a (slight) morning peak. He further analyzed trip starts by land use type finding that ~41% of all trips originated in areas of recreational or public land use, ~36% in areas of commercial land use and ~23% in areas of residential land use. He concluded by reiterating Noland's (2019) hypothesis that a substantial share of e-scooter trips may be of recreational use. Espinoza *et al.* (2020) used data accessed at a 10-min temporal resolution from Bird in the city of Atlanta (GA). They created buffers around origins and destinations of e-scooter trips and counted points of interests (POIs) within those buffers. Interestingly and in contrast to McKenzie (2019), they found POIs associated with their 'business' category (corresponding to the Google Maps API categories Accounting, Banks, Business, Car Rental, Embassy, Insurance Agency, Lawyer, Local Government Office, Real Estate, School) to appear most frequent near trip origins and destinations. Parking, food (Bakery, Cafe, Restaurant, Supermarket) and recreation (Aquarium, Bar, Casino, Library, Museum, Park, Place of Worship, Stadium) POIs also appeared frequently near trip origins and destinations while public transit stops only showed low counts. One limitation of this study is the missing link between the number of POIs in a specific category and their *relative* impact on e-scooter trips. This is addressed by the studies using regression models surveyed next.

Bai and Jiao (2020) analyzed e-scooter booking data from Austin (TX) and Minneapolis (MN) using a (non-spatial) negative binomial regression model on spatially aggregated trip data. They found the CBD and university campuses to be hotspots in both cities while temporal usage patterns differed (rides per weekday in Austin showed a peak on Saturdays while they were more evenly distributed in Minneapolis). Hawa *et al.* (2020) analyzed e-scooter data from Washington, D.C. using a (non-spatial) linear regression model on hourly counts of spatially aggregated data. They also found proximity to the CBD to be an important predictor of demand for shared e-scooters, while it was also positively correlated with higher population densities and bikeways. Despite yielding first insights, these analyses have the methodological shortcoming of not accounting for spatial autocorrelation. The independence condition of explanatory variables is likely violated due to the existence of spatial clusters (i.e., employment centers, shopping centers, residential areas) or spatial correlation of unobserved effects, which suggest the use of spatial regression models as employed in the studies surveyed next.

Arnell *et al.* (2020) analyzed e-scooter trip origin counts aggregated by spatial bins (500m diameter) from Nashville (TN) and San Diego (CA) using a spatial lag regression model. They found the most important predictor of trip starts to be rebalancing points (or e-scooter supply). With increasing distance from the CBD, origin counts in Nashville decreased (San Diego showed the opposite, yet a much weaker and less significant effect) and transit stops had a positive influence on origin counts in Nashville (San Diego, again, showed the opposite, yet a non-significant effect). Caspi *et al.* (2020) analyzed e-scooter trip data from Austin (TX) using a spatial lag regression model on spatially aggregated count data. Methodologically, they removed most cells with zero counts, added one to each dependent variable and took the natural logarithm of the value to approximate normally distributed residuals. They found most trips to be conducted in central Austin and to be associated with areas of denser employment and bicycle infrastructure. Finally, Zuniga-Garcia and Machemehl (2020) used the same dataset from Austin (TX) to apply a spatial error regression model on e-scooter trip origins and destinations. They found the University of Texas at Austin to be the strongest and most significant spatial driver of demand (both for origins and destinations, weekdays and weekends). Population density also had a positive and significant influence on e-scooter stops and origins as did employment density (yet with a much smaller coefficient). Most transit-related variables (no. of boardings and alightings, stop density, bus frequency) had a significant and negative, yet not substantial effect on e-scooter trip stops and origins.

The latter three studies (Arnell *et al.,* 2020; Caspi *et al.*, 2020; Zuniga-Garcia and Machemehl, 2020) all employ spatial lag / error linear regression models on count data. Yet, one of the main assumptions of this type of models is the Normal distribution of residuals which does not hold for count data as it is non-negative. Transforming the counts (Arnell *et al.*, 2020; Caspi *et al.*, 2020) is one way to address this limitation. Another way, which has not been explored until now to the knowledge of the authors yet appears promising, is to use the family of generalized linear (mixed) models which allows for more flexibility on the distributional assumptions of the residuals, i.e. a Poisson or Negative Binomial distribution. Using such a model, a transformation of the dependent variable becomes unnecessary.

In the following, we introduce our data, specify and estimate a conditionally autoregressive (CAR) generalized linear mixed model (GLMM) using a Negative Binomial (NB) distribution.

Data

We use 15 months (Aug/2018 – Oct/2019) of openly accessible shared e-scooter trip data from Louisville (KY) (Louisville Metro Government, 2019). Four e-scooter companies are operating within a 68 mi^2 service area: Bird (since Aug/2018), Lime (since Nov/2018), Bolt (since Jul/2019) and Spin (since Aug/2019).

The initial number of e-scooter trips in the dataset was 434,582. Several data cleaning steps were necessary to exclude unrealistic or non-informative trips, such as trips with a distance of 0 or more than 25 miles, durations of 0 or more than 12 hours and average speeds of more than 30 mi/h. 351,514 trips remained.

We aggregated trip stops by US census blocks within the service area (5'942 blocks) and combined them with the latest-available block-level census information on population (2010) and employment (2015), and Open Street Maps (OSM) data on locations for bus stops, The University of Louisville, restaurants, hotels, stadiums and length of bikeways using QGIS. We further included the area (square miles) of each block as a control variable as census blocks substantially differ in size. Table 1 shows an overview of basic statistics for the dependent and independent variables.

Table 1
Summary of dependent and independent variables per US Census Block used in the regression models.

Variable	Unit	Min	1st Quartile	Median	Mean	3rd Quartile	Max
E-scooter trip stops	Count	0.000	1.000	3.000	57.980	20.000	18'203.000
Restaurants	Count	0.000	0.000	0.000	0.064	0.000	15.000
Bus stops	Count	0.000	0.000	0.000	0.183	0.000	7.000
University of Louisville	Count	0.000	0.000	0.000	0.013	0.000	2.000
Hotels	Count	0.000	0.000	0.000	0.007	0.000	5.000
Stadiums	Count	0.000	0.000	0.000	0.003	0.000	3.000
Population	Count (thousands)	0.000	0.004	0.029	0.045	0.057	2.269
Jobs	Count (thousands)	0.000	0.000	0.000	0.032	0.004	13.636
Bikeways	Miles	0.000	0.000	0.000	0.046	0.058	4.083
Area	Square miles	0.000	0.003	0.005	0.011	0.010	1.082

Figure 1 displays descriptive analyses of the dataset. Shared e-scooter trip starts show clear afternoon peaks both for weekdays and weekends (Figures 1a, 1b). Many trips are short. The median distance is ~0.8 mi (Figure 1c) and the median trip duration is 9 min (Figure 1d). The share of rentals per weekday reveals a peak on Saturdays (Figure 1e) and the distribution of number of e-scooter trip stops per block shows a high share (~24%) of zero observations (Figure 1f).

Figure 1
Descriptive analyses of shared e-scooter trips in Louisville (KY).

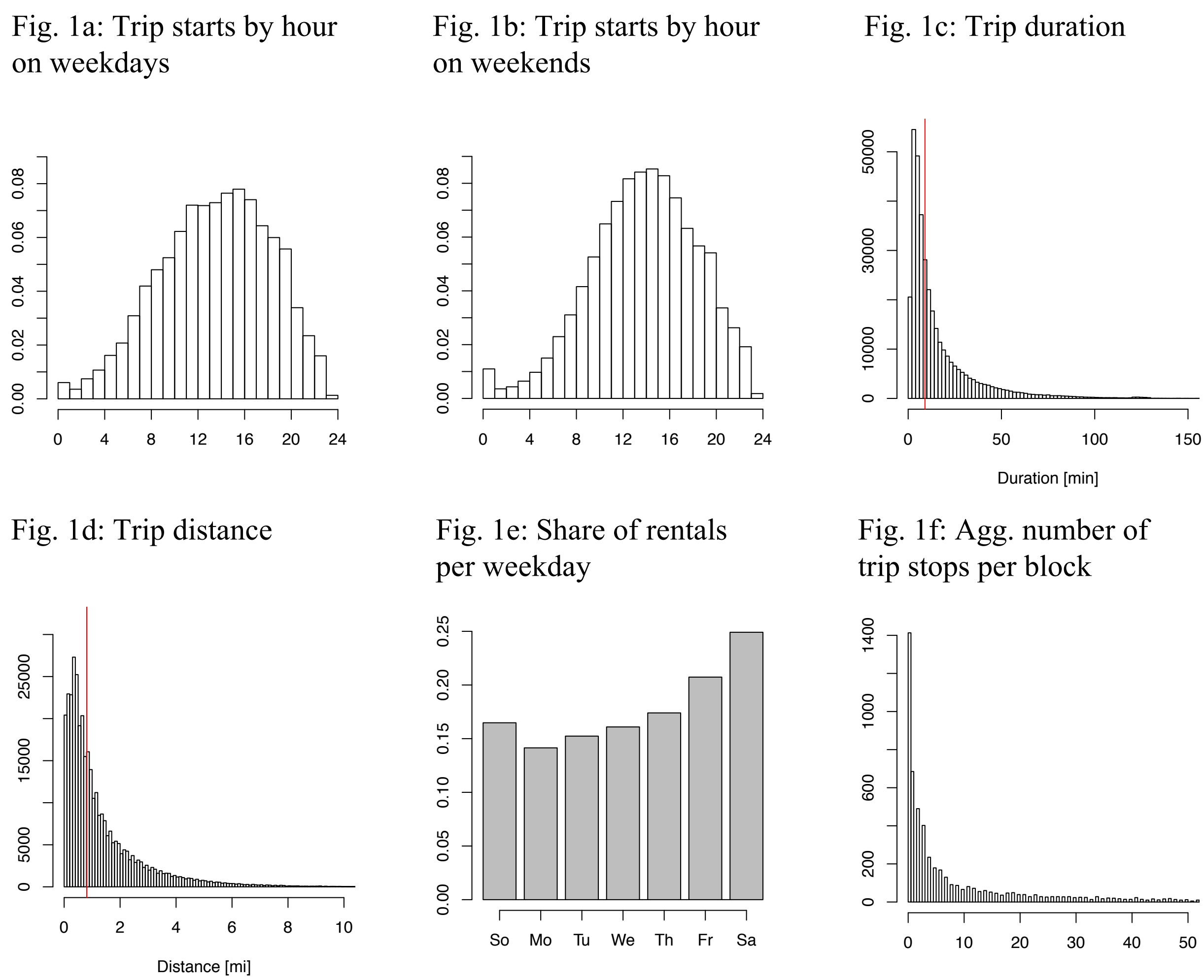
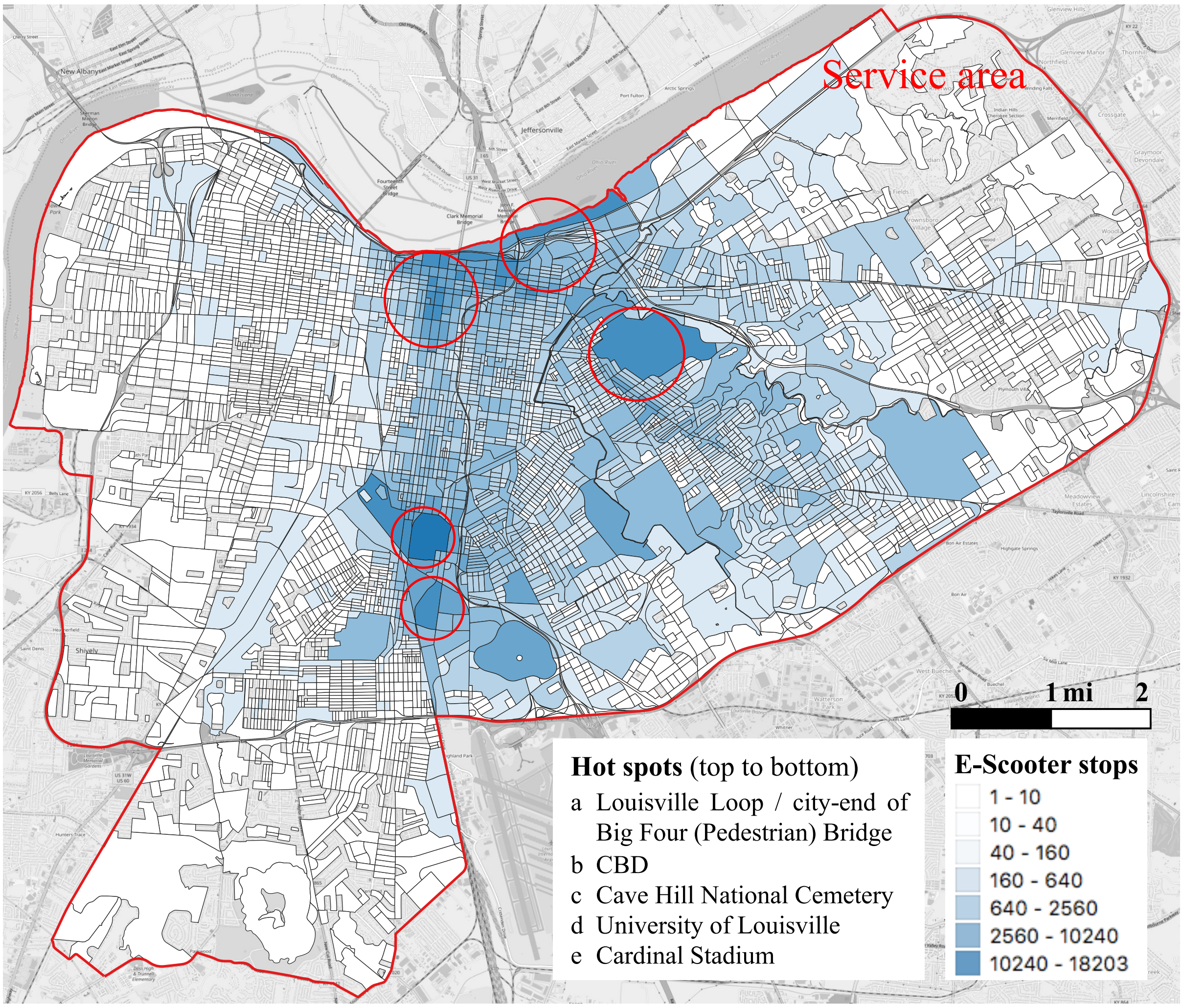


Figure 2 displays a map of Louisville with blocks colored by number of e-scooter trip stops. First hot spots (dark blue, top to bottom) such as the CBD, the Louisville Loop / city-end of the Big Four (Pedestrian) Bridge, Cave Hill National Cemetery, the University of Louisville and Cardinal Stadium can be observed.

Figure 2
Map of Louisville (KY) with census blocks colored by number of e-scooter trip stops and resulting hot spots.



Methods

The integrated and aggregated census block-level dataset includes a substantial number of blocks with 0 observed e-scooter trip stops (~24%) and overdispersion (var/mean ~2'310) which suggests using a Negative Binomial distribution instead of a Poisson distribution. Consequently, we first estimate a generalized linear model (GLM) and a generalized linear mixed model (GLMM) using the Negative Binomial distribution (for estimates and summary statistics, see Table 2) in R (packages MASS and lme4, respectively) using Maximum Likelihood and log link functions. While the GLM already indicates reasonable explanatory power (Naegelkerke R^2: 0.39), the model fit improves substantially by introducing random effects (AIC GLM: 44633, AIC GLMM: 44078). Yet, the model still exhibits a significant level of spatial autocorrelation (Moran's I statistic on GLMM residuals = 0.55, $p = 0.001$).

Spatial autocorrelation can be accounted for by including spatial lags or spatial errors. The rational for modeling spatial lags is the assumption of a diffusion process (i.e., events in one place increase the likelihood of similar events in neighboring places) while the rationale for modeling spatial errors is the assumption of spatial correlation in the error terms (which, in turn, is indicative for omitted spatial variables). As Arnell *et al.* (2020) have shown (and intuition suggests), e-scooter drop-off locations ("rebalancing points") are an important predictor of e-scooter trip origins. As vehicle IDs are not reported in the Louisville dataset and only realized trips are included, we cannot identify 'juicing-trips' (to reload the vehicles' batteries) and rebalancing points, and thus have to treat supply as part of the (spatially correlated) unobserved error. This suggests the use of a spatial error model.

Spatial error terms can be modeled using a conditional (CAR) or simultaneous (SAR) autoregressive correlation model. CAR-type models (originally introduced by Besag, 1974) account for local spatial autocorrelation (i.e., only the influence of direct neighbors), while SAR-type models account for global spatial autocorrelation. As it is reasonable to assume that users will park their e-scooter not much further than a census block from their final destination, we continue using a Negative Binomial-distributed GLMM with a random effect following a conditional autoregressive (CAR) correlation model ('Spatial GLMM') of the form

$$ln(y) = \alpha + \beta\,\mathrm{X} + u$$

Here, y denotes the number of e-scooter stops in the 15 months period per census block, $\beta\,\mathrm{X}$ the coefficients and vector of fixed effects as shown in Table 1, followed by the random effect u with a CAR-type covariance matrix of the form $\lambda(\mathbf{I} - \rho\,\mathbf{N})^{-1}$ where $\mathbf{N}$ is an adjacency matrix between the census blocks (i.e., a matrix with elements 1 if the blocks are adjacent and 0 otherwise). We estimate the model in R (package spaMM) using Maximum Likelihood and an ln link function.

Results

The estimated spatial GLMM model as well as results for the non-spatial GLM and GLMM models are shown in Table 2. The spatial correlation structure of the random effect further improves the model fit from the non-spatial GLMM (AIC: 44'078) to the spatial GLMM (AIC: 38'981). In the following, we thus focus on the results of the spatial GLMM. It becomes apparent that the area of a census block has a strong influence (11.28) of the number of e-scooter stops in it (cf. Fig. 2), which was expected. Population size has the second strongest effect (2.29) on number of e-scooter stops, interestingly much stronger than the number of jobs (0.33) in a census block. This indicates that a substantial share of users drive shared e-scooters to their inner-city homes. The length of bikeways in/adjacent to a block has the third-strongest effect (1.43). Though bikeways may not be destinations per se, this does show that e-scooters are parked substantially more near where they are established. In terms of points of interests, the University of Louisville is the strongest attractor of shared e-scooters (0.98). Interestingly, hotels appear to be attractors (0.36), too, indicating that tourists may be form a substantive share of e-scooter users in Louisville. Restaurants and bus stops also show significant, yet less substantive positive effects on e-scooter trip stops (0.18 and 0.21, respectively).

Table 2

Estimations results for shared e-scooter trip stops. See Table 1 for a description of the variables.

		GLM			GLMM			Spatial GLMM		
Variable	Unit	Estimate	SE	z-value	Estimate	SE	z-value	Estimate	SE	t-value
(Intercept)		2.91	0.03	94.62	1.04	0.04	27.03	0.51	0.05	9.54
Restaurants	Count	0.93	0.06	15.86	0.70	0.07	10.52	0.18	0.05	3.50
Bus stops	Count	0.41	0.05	8.86	0.39	0.05	7.22	0.21	0.05	4.66
University of Louisville	Count	1.22	0.19	6.46	2.20	0.21	10.31	0.98	0.25	3.92
Hotels	Count	1.15	0.21	5.45	0.84	0.24	3.53	0.36	0.18	1.97
Stadiums	Count	0.97	0.33	2.90	1.42	0.38	3.78	0.19	0.28	0.70
Population	Count (thousands)	-0.16	0.36	-0.43	0.66	0.43	1.55	2.29	0.36	6.41
Jobs	Count (thousands)	1.98	0.09	20.95	0.75	0.11	7.02	0.33	0.08	4.00
Bikeways	Miles	5.21	0.26	20.14	2.82	0.30	9.44	1.43	0.24	5.84
Area	Square miles	1.20	1.08	1.11	-1.58	1.22	-1.30	11.28	1.10	10.26
ρ								0.12		
λ								2.25		
n		5'942			5'942			5'942		
AIC		44'633			44'078			38'981		
Naegelkerke R^2		0.39								
Marginal log-likelihood		-22'306			-22'027			-19'478		

Discussion

Our results largely confirm previous results on spatial drivers of shared e-scooter demand and thus add further evidence to their generalizability. For Louisville, we find that population and bikeways have particularly strong effects on shared e-scooter stops. This confirms previous findings for Washington D.C. (Hawa *et al.,* 2020) and Austin (TX) (Caspi *et al.,* 2020; Zuniga-Garcia and Machemehl, 2020). University campuses have also previously been found to have a substantially positive effect in Austin (TX) and Minneapolis (MN) (Bai and Jiao, 2020; Zuniga-Garcia and Machemehl, 2020). Interestingly, this was not true for Washington D.C. (Hawa *et al.,* 2020) which may have to do with its more diversified city center. We also find bus stops to have a significantly positive, yet less substantial effect on e-scooter stops than other POIs. Previous studies show mixed evidence for this relationship suggesting first/last mile usage with positive effects found for Nashville (TN) (Arnell *et al.*, 2020) and negative effects found for Austin (TX) (Zuniga-Garcia and Machemehl, 2020) and San Diego (CA) (Arnell *et al.*, 2020). Our results extend previous results by suggesting that tourism (i.e., hotels, restaurants) may be a driver of e-scooter demand. This appears plausible given the mobility demand of tourists and the fact that most e-scooter companies allow their users to rent e-scooters in different cities.

The direction of the effects are further comparable to the effects observed in bicycle-sharing demand models. Previous literature on bicycle-sharing demand also reported positive effects of population, workplaces, proximity to central locations (such as university campuses and central business districts), restaurants and cycling infrastructure (Guidon *et al.,* 2019; Noland *et al.,* 2016; Shen *et al.,* 2018). While aggregate effects are usually consistent, the effect of population and workplaces can vary in disaggregate models for specific hours of the day or the weekend (Noland et al., 2016).

Conclusion

This paper reports on spatial drivers of shared e-scooter trip destinations in Louisville (KY). Our results largely confirm previous studies in that population density, the presence of bikeways and university campuses have the strongest positive effect on counts of shared e-scooter trip destinations. We find a significant, yet less substantial positive effect of bus stops suggesting some first/last mile use. Our results extend previous literature in that they suggest that tourists may be an overlooked, yet important segment in shared e-scooter demand.

We suggest future research to further explore the potential link between tourism and shared e-scooter demand, which can be done through targeted surveys or spatial regression models. For the latter, we see a particular need for comparative case studies estimating the same models on similar data for multiple cities.

References

Arnell, B.M., P. Noursalehi, E.M. Huntley and J. Zhao (2020) Shared Electric Scooters and Transportation Equity: A Cross-City Analysis. Paper presented at the *99th Annual Meeting of the Transportation Research Board*, Washington, January.

Bai, S. and J. Jiao (2020) Dockless E-Scooter Usage Patterns and Urban Built Environments: A Comparison Study of Austin, TX and Minneapolis, MN. Paper presented at the *99th Annual Meeting of the Transportation Research Board*, Washington, January.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological),* **36** (2) 192-225.

Caspi, O., M.J. Smart and R.B. Noland (2020) Spatial Associations in Dockless Shared e-Scooter Usage. Paper presented at the *99th Annual Meeting of the Transportation Research Board*, Washington, January.

Espinoza, W., M. Howard, J. Lane and P. van Hentenryck (2020) Shared E-Scooters: Business, Pleasure, or Transit. Paper presented at the *99th Annual Meeting of the Transportation Research Board*, Washington, January.

Guidon, S., H. Becker, H. Dediu and K.W. Axhausen (2019) Electric bicycle-sharing: a new competitor in the urban transportation market? An empirical analysis of transaction data. *Transportation Research Record*, **2673** (4) 15-26.

Hawa, L., B. Cui, L. Sun and A. El-Geneidy (2020) Scoot over: Determinants of shared electric scooter use in Washington D.C. Paper presented at the *99th Annual Meeting of the Transportation Research Board*, Washington, January.

Louisville Metro Government. *Open Data Platform*. Louisville, KY, 2019. https://data.louisvilleky.gov/dataset/dockless-vehicles. Accessed Nov. 25, 2019.

McKenzie, G. (2019) Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, D.C. *Journal of Transport Geography,* **78**, 19-28.

NACTO (2019) Shared Micromobility in the U.S.: 2018. *Technical Report*, National Association of City Transportation Officials, New York City, NY.

Noland, R.B. Trip patterns and revenue of shared e-scooters in Louisville, Kentucky. *Transport Findings,* 2019. April.

Noland, R.B., M.J. Smart and Z. Guo (2016) Bikeshare trip generation in New York City. *Transportation Research Part A: Policy and Practice,* **94**, 164-181.

Portland Bureau of Transportation. *E-Scooter Findings Report*. Portland, OR, 2018. https://www.portlandoregon.gov/transportation/article/709719. Accessed Sep. 25, 2019.

Shen, Y., X. Zhang and J. Zhao (2018) Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation,* **12** (9) 686-700.

Zuniga-Garcia, N. and R. Machemehl (2020) Dockless Electric Scooters and Transit Use in an Urban/University Environment. Paper presented at the *99th Annual Meeting of the Transportation Research Board*, Washington, January.