SPATIAL VARIATION IN SHARED RIDE-HAIL TRIPS & FACTORS CONTRIBUTING TO SHARING

Matthew D. Dean
NSF Graduate Research Fellow
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
mattdean@utexas.edu

Kara M. Kockelman, Ph.D., P.E.[1]
Dewitt Greer Centennial Professor in Transportation Engineering
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin– 6.9 E. Cockrell Jr. Hall
Austin, TX 78712-1076
kkockelm@mail.utexas.edu

Journal of Transport Geography 91, 102944, 2021.

Word Count: 6,000 words + 6 tables = 7,500 words (+ 6 figures)

[1] Corresponding author

ABSTRACT

As ride-hailing becomes a more prevalent mode in cities and the effect of this increased demand on fixed infrastructure manifests itself, authorities are seeking data-sharing agreements with transportation network companies (TNCs) or trusted third-parties to understand the spatial and temporal variation in demand of this emergent mode to enact sound regulations. Although rider-specific demographic information is unavailable to researchers due to privacy concerns, publicly available data of masked trips, like Chicago's transportation network provider data, in combination with census tract-level built environment and socioeconomic variables can yield valuable insights into trip-making, particularly the factors contributing to sharing. Effective sharing of TNC trips will increase vehicle occupancy, thereby reducing overall passenger miles traveled and improving traffic conditions and per-passenger mile emissions. This study uses spatial econometrics to understand and predict sharing based on built environment, socioeconomic, location, time, and trip factors in Chicago, Illinois. Areas with a high density of retail and entertainment jobs, multi-modal infrastructure, and a young population have a strong positive effect on shared TNC trip origin counts. Moreover, regions that have longer TNC trips tend to have higher rates of sharing, but the same cannot be said for longer work commutes or higher trip fares. Lastly, the ratio of shared TNC trips to total trips declined across the average tract in 2019 and incentivizing shared trips during periods of high leisure trips (i.e., weekends) and in areas with well-educated residents may help to reverse this trend.

Keywords: ride-hailing, sharing, spatial econometrics, spatial regression

BACKGROUND

In the last decade, the growth of on-demand ride-hailing (also referred to as ridesourcing or transportation network company (TNC) services), that matches drivers with passengers using smartphone apps and advanced routing algorithms, has led to spillover effects across the urban transportation system. TNCs are known to have a considerable impact on regionwide emissions, congestion, transit ridership, household vehicle ownership decisions, and trip behavior. Increasing average vehicle occupancy is critical in reducing congestion and per-passenger mile emissions, and the choice for riders to share their TNC trip[2] could lower their impacts (Shaheen and Cohen, 2019; Sperling, 2018). Yet, research guiding policy-making decisions on shared TNC rides is still in its infancy even though TNCs in the U.S. launched shared ride services in 2014.

Numerous studies document the differences of shared mobility services (Shaheen and Cohen, 2019; Chan and Shaheen, 2012), the factors influencing their adoption and demand (Alemi et al., 2018; Dias et al., 2017; Lavieri and Bhat, 2019; Yu and Peng, 2019), and studies answering how ride-hailing impacts traditional transit ridership (Clewlow and Mishra, 2017; Schwieterman and Smith, 2018), but very few have examined shared trips exclusively. A reasonable explanation is that shared rides form a small percentage of ride-hail trips, about 19% of trips in Hangzhou, China, 20% in Chengdu, China, and up to 22% of passenger miles traveled in California (Chen et al., 2018, p. 18; Zheng et al., 2019, p. 147; CARB, 2019, p. 1). Additionally, TNC trips constitute a fraction of a household's transportation budget. In the 2017 National Household Travel Survey, only 8% of respondents took a ride-hail trip in the previous month (FHWA, 2019). Furthermore, few datasets track if a shared trip was successful (i.e., when passengers were matched with other unrelated passengers). Still, a larger barrier to this field of research is the aversion of TNCs to

[2] Given multiple definitions in the literature, "sharing", where passengers are willing to travel with strangers to split the fare at the cost of additional travel time (in-vehicle and out-of-vehicle), is equivalent to pooling, ridesplitting, and dynamic ride-sharing (DRS).

share trip data without significant data aggregation measures due to privacy concerns. Naturally, only a few studies have investigated trends in shared TNC rides using surveys and real-world trip data – most leveraging DiDi Chuxing data.

Literature Review

Chen et al. (2017) employed Boosting ensemble learning, a machine learning technique that combines multiple base models, to create a strong classification model that predicts sharing of potential ride-hail customers based on ride-hail features such as trip length, fare, and trip time reliability. They reported better ride classification using this technique than logistic regression, support vector machine, and naïve Bayes classification. Chen et al. (2018) extended this preliminary analysis to real-world DiDi Express and Hitch ridesplitting data, and a post-ride questionnaire to study the reported mode shift to shared TNC rides and estimate vehicle miles traveled (VMT) savings. They estimated a savings of slightly more than 36,000 VMT per day (which seems insignificant for a city of 10 million). Their methodology assumes that a passenger would have ordered a more expensive, non-shared DiDi service had the sharing option been unavailable or outside their time budget. However, a survey from the Union of Concerned Scientists (2020) found that 36% of shared trips would have been taken by transit, active transportation, or foregone had ride-hail not been available, 15% by taxi, 21% by carpool, and 24% by driving alone. Additionally, the study used the Hitch ride-sharing service, which is more analogous to carpooling since users can schedule a ride in advance, confounding the on-demand shared TNC focus of this study.

Li et al. (2019) analyzed a one-month sample of trip data from DiDi Express in Chengdu, China, and estimated only 6.2% of trips are successfully shared and that 90% of these trips are between two unrelated parties. As a result, shared trips had a 22% savings in total vehicle service hours compared to single-party trips. The authors developed a regression analysis using density, development, and diversity of land use factors to tease out both the correlation between the spatial demand for TNCs and the built environment and shared TNC trip delays and the built environment. High density and more economic development of a census tract has a strong positive effect on the count of TNC rides, perhaps because of increased opportunities and better road infrastructure, but increases in diversity of land use are negatively associated with TNC trips. Furthermore, proximity to bus stops is positively correlated with delays in shared rides. Diversity of land use at the pickup and drop-off area has the same correlation, perhaps because of wayfinding issues.

Zheng et al. (2019) surveyed recent DiDi Express and Hitch customers in Hangzhou, China, to investigate the influence of TNC sharing on vehicle usage and ownership. Interestingly, sharing had the effect of removing 3.6% of the region’s vehicles from the road, daily, but the survey did not ask respondents about their willingness to use their personal automobile had ridesharing not been available. The authors also estimated that more than a fifth of respondents postponed their plan to purchase a new vehicle after the emergence of shared ride-hail services, particularly for households currently without a vehicle.

Hou et al. (2020) explored the socioeconomic, spatiotemporal, and trip-specific factors that influence one’s willingness to share. Their study binned Chicago TNC trips between November 2018 and April 2019 by census tract origin-destination pairs, time of day (in five discrete categories), weekend, airport trip pick-up, and airport trip drop-off indicator variables and used bins with at least 100 trips. This study used an ordinary least squares (OLS) model and a decision-tree-based ensemble machine learning algorithm called XGBoost to predict sharing. They found that increases in trip distance and duration and the price differential with shared services between locations (all else constant) increase a tract's ratio of shared trips. Additionally, trips to and from airports representing time-constrained trips lead to decreases in sharing. Similarly to Chen et al. (2017), the machine learning technique outperformed their OLS specification.

In this study, we explore the effect of built environment, socioeconomic, spatial-temporal, and trip factors on the demand for shared TNC trips by first approaching this topic with spatial econometric techniques. This is novel as previous studies have not modeled the spatial demand of shared TNC trips as observed in studies of other modes and services (Yu and Peng, 2019 (TNCs); Reck et al., 2020 (e-scooters); Becker et al., 2017 (car-sharing); Ma et al., 2018 (transit)). The results of this analysis can inform policy to incentivize sharing in the short-term and promote lasting land use changes that encourage sharing behavior in transportation modes. The remaining sections of this paper are organized as follows – explanation of the data sources and summary statistics; a review of spatial econometric models; implementation of the spatial econometric models and concluding remarks.

DATA DESCRIPTION

Since November 2018, the City of Chicago has maintained an online portal of ride-hail trips provided by licensed TNC vendors, totaling 152 million records (as of May 28, 2020). To maintain the privacy of sensitive transportation data in the public domain, Chicago masks individual records temporally, spatially, and financially. Trip start and end timestamps are rounded to the nearest 15 minutes, pickup and drop-off information is aggregated to the census tract level, or suppressed in some cases, and fares are rounded to the nearest $2.50 (Levy, 2020). Trips were obtained for the entire year of 2019 to account for temporal variations in TNC demand, such as seasonal or holiday effects. A random non-replacing sampling approach for each month scaled the total number of trips down by a factor of 10. Raw trip records were scrubbed to remove erroneous trip records such as those that have incomplete pickup/drop-off location data, temporal data, trip distance, and trip duration resulting in semi-truncated data beyond what is provided by Chicago. The full list of measures is given below:

1) Duration is greater than 60 seconds
2) Distance is greater than 0.1 miles
3) Fare is not $0 and is less than $100
4) Pickup/dropoff tract is not blank & both are in Chicago
5) Trip Start Timestamp is after 12/31/2018 11:59:59 PM & before 01/01/2020 12:00:00 AM

As an example, in January 2019 (before sampling) there were 5.87 million trips taken in Chicago meeting these criteria. Almost 23% of all trips were taken with a shared service, lower than the 25.8% that Hou et al. (2020) found over five months (November 2018 – March 2019) with the same dataset. Of those rides taken with the option to share, 75.0% were successful, defined as two or more consecutive unrelated parties (of unknown size) sharing a ride during a period when there were always passengers in the vehicle[3].

Socioeconomic data were obtained from the U.S. Census Bureau's American Community Survey (ACS) 2014—2018 (5-Year Estimate) followed by 2013–2017 (5-Year Estimate) if the latest data was not available for census tracts. Similarly, built environment data were obtained from the EPA Smart Location Database for the Chicago metropolitan region, and data was matched by census tract.

[3] This study examined shared TNC trips (as a service type) to estimate the spatial demand for this mode.

The majority of TNC trips are short in duration and distance. A histogram shown in Figure 1a reveals an exponential decline in trip duration roughly after the 10-minute mark. Figure 1b depicts an exponential decline of trip distance after two miles. Additionally, shared services are more likely to be taken midweek than during the weekend, with a lower proportion of sharing occurring on Saturdays (Figure 2). In addition to temporal variation, there is also significant spatial distribution in the rate of sharing. While there is high variation in the rate of shared trips by pick-up census tract (Figure 3a) for the January 2019 subset, the total number of average daily shared trips is largest in the central business district (CBD) (Figure 3b). Unsurprisingly, areas with higher trip generators have the highest number of shared trips, but the clustering of tracts with high and low ratios of sharing implies a strong positive spatial dependence on the rate of shared trips.

A descriptive statistics table for the average socioeconomic, built environment and trip factors by census tract for the January 2019 subset is shown below in Table 1. There are pockets of extreme wealth and poverty, as measured by median annual household income, owner-occupied median house value, and percent unemployment. The average tract is majority-minority (54.01% nonwhite) and young (32.31% of the population aged 18-34). Nearly a third of Chicago's population aged 25 years and above has a bachelor's degree with some tracts reaching nearly 100%. With an average tract population density of over 19,000 people per square mile, the availability of other transportation modes associated with high density areas can increase the rate of voluntary zero-vehicle households that would otherwise be negatively correlated with low incomes. The socioeconomic variation across tracts is also a function of the built environment – years of (dis)investment, zoning laws, and suburban sprawl has created pockets of high population and employment density (in retail and entertainment industries), multi-modal infrastructure, and transit. These socioeconomic and built environment factors, as suggested in the literature, may help to explain the observable count of shared TNC trips.

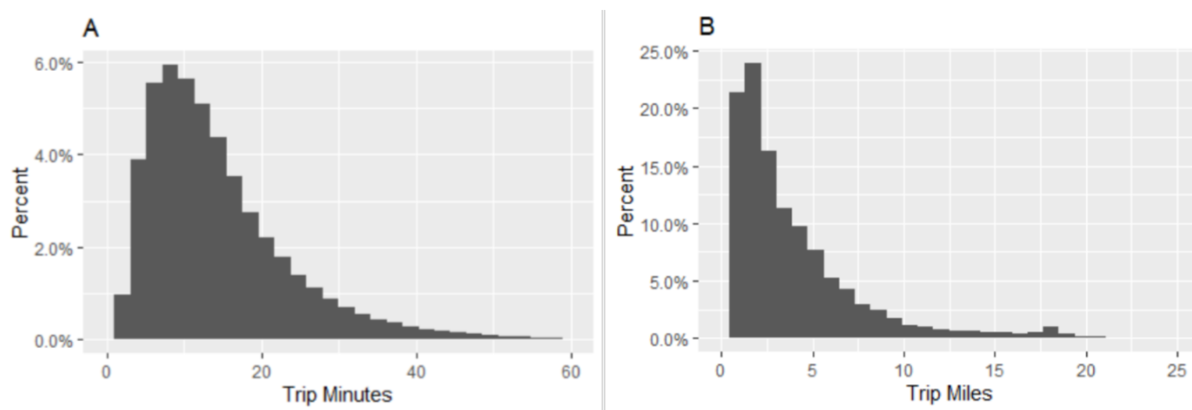

FIGURE 1 Histogram of trip durations and trip distance for a January 2019 subset

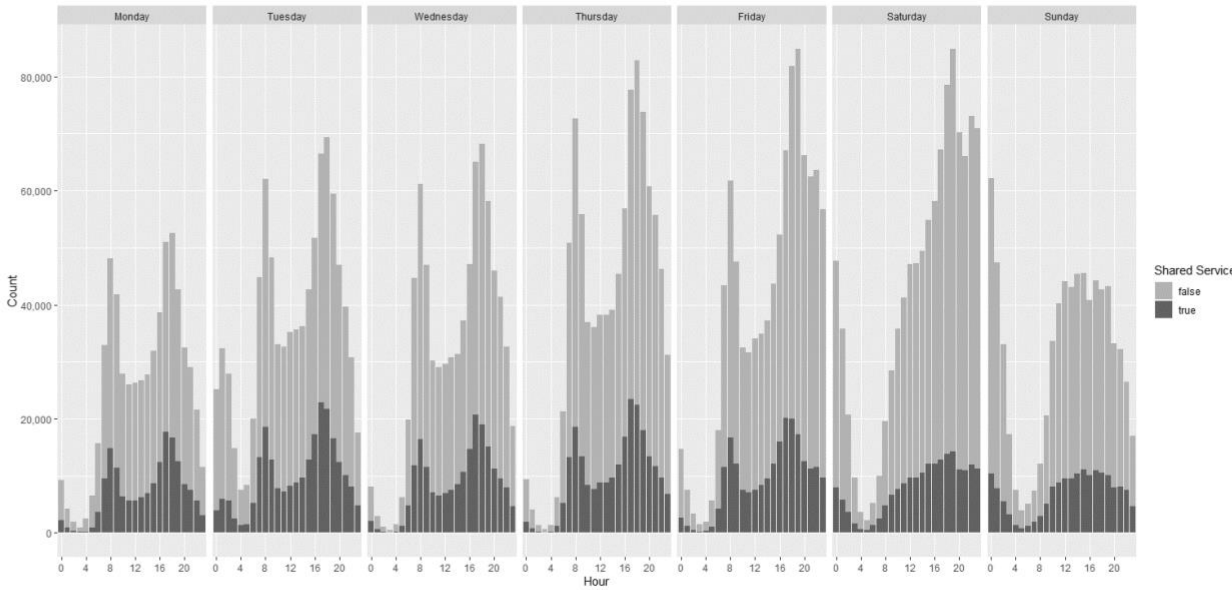

FIGURE 2 Count of TNC trips by service type for a January 2019 subset

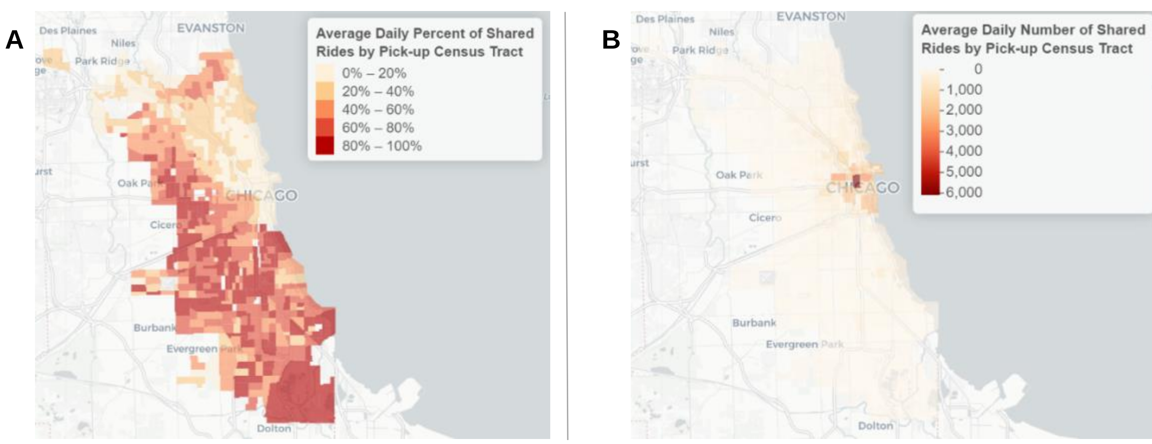

FIGURE 3 Average daily percent and total shared rides by pick-up for the January 2019 subset

TABLE 1 Descriptive Statistics by Census Tract for January 2019 Subset (N=772)

Variable	*Label*	*Min*	*Max*	*Mean*	*SD*	*Median*	*Source*
HHSIZE	Average Household Size	1.34	4.5	2.64	0.60	2.62	
HH_INCOME	Median Household Income (1,000$)	9.79	178.75	57.96	32.40	48.63	
HOUSE_VALUE	Median Value of Owner-Occupied Dwellings (1,000$)	53.80	868.40	266.57	151.08	227.20	ACS 2014-2018, ACS 2013-2017 (5-Yr Est.)
PCT_NONWHITE	Percent Nonwhite Population	3.30	100.00	54.01	32.31	48.62	
PCT_AGE18TO34	Percent Population Ages 18-34	10.23	86.60	29.86	11.63	26.51	
PCT_BACHELORS	Percent Population Ages 25 and Over with at least a bachelor's degree	0.51	94.95	36.19	26.19	28.63	

PCT_VEH0	Percent Zero-Vehicle Households	0.00	80.85	36.33	16.78	35.58	
POP_DENS	Population Density (1,000 people per square mile)	0.46	263.99	19.11	15.77	16.10	
PCT_UNEMPLOYED	Percent Unemployed	0.00	48.47	10.77	8.54	7.94	
COMMUTE	Average Work Commute (minutes)	17.00	53.00	35.64	5.54	36.00	
MILES	Trip Distance (miles)	2.11	18.04	5.50	1.74	5.26	
MINUTES	Trip Duration (minutes)	5.41	38.69	18.47	3.87	18.13	
FARE	Trip Fare ($)	2.50	22.50	9.75	1.93	9.32	
TOTALCOST	Total Trip Cost ($)	5.05	27.55	12.52	2.16	12.03	Chicago TNC Dataset for Jan. 2019
TRIPS_SHARED	Average Number of Consecutively Shared Trips	1.00	3.00	1.50	0.25	1.48	
SHARED	Fraction of Trips with Shared Service	0.00	1.00	0.38	0.16	0.38	
SHARED_CNT	Count of Shared Trips	0.00	6,080	161.04	373.48	55.00	
WEEKEND	Indicator for Weekend	0.00	1.00	0.28	0.11	0.28	
DENS_ENTER	Gross Entertainment Employment Density (jobs/acre)	0.00	90.01	1.70	6.73	0.24	
DENS_RET	Gross Retail Employment Density (jobs/acre)	0.00	55.77	1.07	3.36	0.29	
NTWKDENS_MM	Network Density of multi-modal links (miles/sq. mile)	0.00	21.08	3.25	2.45	2.93	EPA Smart Location Database (2014)
DIST_TR	Distance from population-weighted centroid to nearest transit stop (meters)	0.00	1,033.64	510.73	160.59	517.10	
FRQ_TR	Aggregate frequency of transit service per square mile	71.58	84,441.17	3,085.36	5,055.51	1,985.11	

DEVELOPMENT OF A MODEL

A model to estimate the sharing decision in TNC trips as a function of socioeconomic, built environment, temporal, and trip-specific variables is proposed. The Y variable is either a proportion or count variable of shared trips with census tract data for all X variables. Two models are developed because although a higher count of shared trips may be associated with a higher rate of shared trips, Figure 3 dispels this general notion. As sharing is offered only in select markets where the density of trip generators/attractions exists to match riders to vehicles, associating built environment density variables may allow for a better predictive model (Yu and Peng, 2019). Due to privacy concerns, trip spatial data is aggregated to census tracts, which can introduce spatial autocorrelation (Wang et al., 2013). Thus, a model that supports a likely spatiotemporal phenomenon in sharing choice behavior should be conducted, especially since prior work in taxi trip models shows spatial autocorrelation exists (Correa et al., 2017)

The choice to split a ride can be reasonably estimated using multiple models, borrowed from studies on crime data, crashes, and the spread of diseases. According to Lord and Park (2013), a Poisson-Gamma (NB) can capture spatial data with a spatial autocorrelation term for each observation. The resulting non-negative mean of the model is $\lambda_i = e^{X_i'\beta+\varepsilon_i+\phi_i}$, where spatial autocorrelation, ϕ, is parameterized. A downside of these spatial interaction models (e.g., Conditional Auto-Regressive [CAR] Model) is that it cannot be estimated by Maximum Likelihood Estimation (MLE) but with more computationally intensive processes such as the Markov Chain Monte Carlo (MCMC) or Integrated Nested Laplace Approximation (INLA) (Khana et al., 2018). A different approach to expressing spatial dependencies in a regression model that yields similar results is to have explicit spatial variables as opposed to the above parameter approach (Levine et al., 2013). Examples of spatial relationship variables include the distance to the CBD, area of a zone (e.g., census tract or traffic analysis zone), or a distance-weighted value reflecting adjacency with zones (e.g., impedance). While Waldo Tobler's First Law of Geography holds true (i.e., near zones are more related than distant zones *ceteris paribus*), this approach requires zonal independence and uniformity intra-zone, which cannot necessarily be guaranteed. The best model is ultimately the one that captures the shared choice behavior in the simplest means possible.

One approach is to ignore temporal variation, at first, and test for spatial dependence of the data. This involves estimating the rate of sharing in TNC trips by ordinary least squares (OLS) regression while assuming there is no spatial autocorrelation, or rather there is no correlation in the rate of shared ride-hail trips in space (Cliff and Ord, 1973). Spatial autocorrelation has the general condition: $Cov(y_i, y_j) \neq 0 for\ i \neq j$ for all y observations at locations i and j, and has been simplified in econometric work to a spatial weights matrix which measures the interactions with neighboring census tracts (Ignacio Sarmiento-Barbieri, n.d.). Recalling Tobler's Law, neighbors can be defined to include tracts that are within a set distance (such as 30 minutes of travel time by car), nearest k neighbors, or a simple queens (side and edge included) and rooks (side only) rule spatial matrices. A test for spatial autocorrelation is Moran's I test (Moran, 1950), which is produced using a vector of OLS residuals and a row of standardized spatial weights matrix by means of standardizing the spatial autocovariance.

Since the results are relative to the modeler's choice of a spatial weights matrix, different results may apply to different inputs. If the Z-value from Moran's I test is statistically significant (at α = 0.05, for example), then the null hypothesis that sharing is randomly distributed across Chicago can be rejected, implying that there is significant spatial autocorrelation. If indeed the results from Moran's I test indicate a valid alternative hypothesis, two spatial lag models may sufficiently capture the effect of spatial dependence on sharing – the spatial autoregressive model (SAR) and the spatial error model (SER). SAR introduces the spatial weight matrix similarly to the endogenous variables like trip, socioeconomic, and built environment variables ($y = \rho W y + X\beta + \epsilon$)[4] whereas SEM incorporates the spatial weights matrix in the error term ($y = X\beta + \epsilon$, where $\epsilon = \lambda W \epsilon + u$) (Millo, 2014). An implication of these two approaches is that they yield two different conclusions for the OLS model if SAR or SEM is an appropriate model (i.e., a spatial residuals plot does not appear to show a presence of spatial autocorrelation). A SAR model

[4] Subscripts dropped for brevity

specification implies biased and inconsistent estimates, while the SEM model would imply unbiased but inefficient estimates of the OLS model.

To include the temporal distribution in ride-hailing behavior, including seasonal variations because of Chicago's climate, a regression on panel data is done. A basic linear panel model has both cross-sectional invariant variables and variables that vary over both cross-section, i, and time, t, and error terms ($y = \alpha + X_{1,t}\beta_1 + X_{2,it}\beta_2 + u_{it}$) (Srinivasan and Kockelman, 2002). Since this study samples trips by month for 2019 (t = 1 to 12), the number of cross-sections (census tracts) varies and results in an unbalanced panel (Figure 4). With this panel data, two model specifications were developed based on assumptions of correlation across census tracts with the regressors. The error term can be formulated as constant across census tracts (fixed effects) or assumed to be randomly distributed with a variance of σ_u^2 (random effects). Intuitively, random effects may be more aptly suited for this work, as it allows for variation in sharing behavior across census tracts and even months to follow expected behavior shown in Figures 3a and 3b; but the chosen specification should be determined using an appropriate statistical test, like Hausman (Greene, 2003, p. 301). The Hausman test (Hausman, 1978) determines if the underlying data supports random effects by assuming a null hypothesis of no correlation with the regressors (such that fixed effects is an inefficient estimator). If the resulting p-value of the test is statistically significant, the conclusion is that the random effects estimator is inconsistent, and thus fixed effects is a better model. Additional model specifications such as the time-fixed effects model may be necessary to capture variation in TNC trip variables (e.g., distance, duration, and fare) if correlation is found. Lagrange Multiplier for time effects can provide guidance on which model specification best explains the data.

A limitation of the previous approach is that it fails to capture any spatial interactions across the individual cross-sectional units (i.e., census tracts) and over time. Hence, a balanced spatial panel data model of 763 census tracts was developed for the count data with just eight tracts dropped across the City to estimate these spatial interactions, as suggested in Millo and Piras (2012). Both random and fixed effects models with an added spatial autoregressive error term are developed, and a spatial Hausman test (Mutl and Pfaffermayr, 2011) is conducted to determine which specification is more appropriate. The models are estimated with generalized moments, although maximum likelihood is also possible but at a cost of computational time. A random effects model with serially correlated remainder errors and a spatial autoregressive term is then used to compare with the previously estimated RE specification (see (Millo, 2014) for further model taxonomies and a commentary on their development). A generalized flowchart of these model specifications for both the ratio of shared trips and count of shared trips is shown in Figure 5.

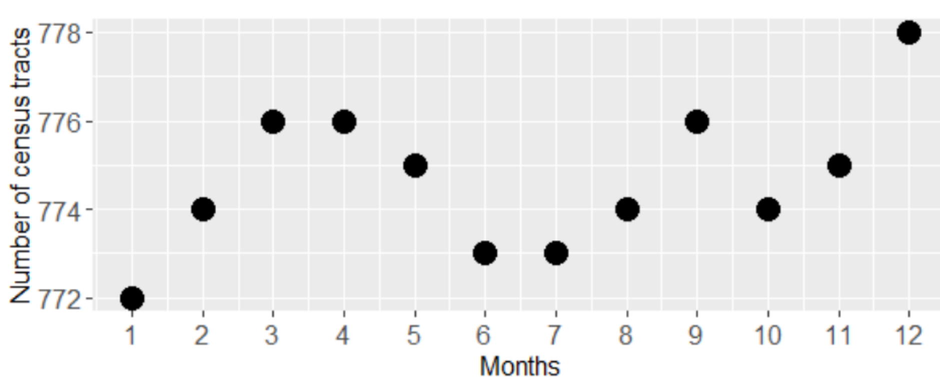

FIGURE 4 Number of census tracts in the unbalanced panel dataset for 2019 by month

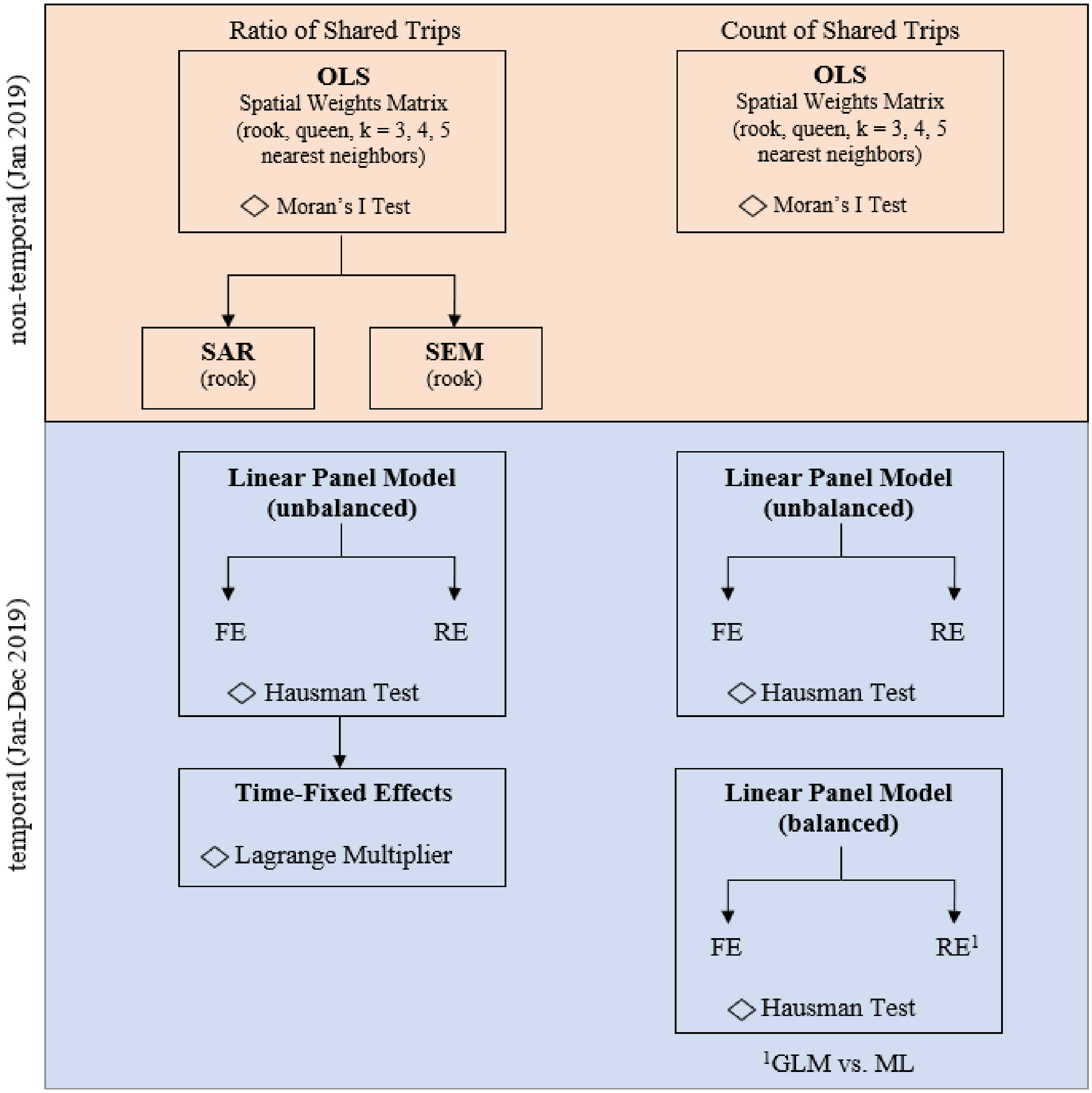

FIGURE 5 Generalized model specification flowchart

RESULTS

Ratio of Shared Trips

The final OLS specification for the ratio of shared trips is presented in Table 2 below, indicating that built environment variables are not beneficial in predicting the rate of sharing in a census tract. On the other hand, socioeconomic variables and trip length were statistically significant. Trip duration had the largest *t*-statistic; however, fare and distance could also be used but were not included because of multicollinearity. There were five different spatial weight matrices: two contiguity-based (queen and rook criterion), and three closest *k* neighbors (*k*=3, 4, 5). All spatial weight matrices yielded a p-value significantly lower than 0.05 with Moran's I test. The rook criterion had the lowest p-value and was chosen for the subsequent spatial regressive models. This indicates the null hypothesis can be rejected in all circumstances and that significant spatial autocorrelation between census tracts is evident in the OLS model. A subsequent Lagrange Multiplier diagnostic for spatial dependence confirmed this conclusion.

Two spatial regressive models, SAR and SEM, with the same independent variables as in the OLS model were estimated to test whether accounting for spatial autocorrelation improved the fit of the model. The last two columns of Table 2 present the results for these two spatial regressive models. The log-likelihood (LL) of the SEM model is lower than SAR, and a plot of residuals for SAR (Figure 6a) and SEM (Figure 6b) shows there is still some autocorrelation not accounted for in these models, predominately in Chicago's South Side and northwest leading to O'Hare International Airport. This suggests that residual autocorrelation is present and advanced spatial econometric models may be necessary. Nevertheless, the results from the SEM model of the OLS specification imply that census tracts with higher rates of sharing have longer TNC trips, which would allow for TNC to better match riders (hopefully avoiding poor matches) and would allow riders to split the higher fare. Areas with a high percentage of nonwhite residents and unemployed residents have a similar effect on increasing the proportion of shared trips to total trips. Moreover, these variables may indirectly account for wealth disparities that increase ridesplitting behavior. Additionally, areas with high average household sizes tend to have higher rates of sharing, which may suggest that individuals in larger households are more comfortable with sharing rides with strangers, an effect of intrahousehold sharing. Noticeably, tracts with a higher percentage of the population having a bachelor's degree tended to have lower sharing. Since this variable only considers the population of residents ages 25 and over, it fails to count current university students, who are more willing to share rides.

TABLE 2 Final OLS, SAR, and SEM Models for the Ratio of Shared Trips in January 2019

	OLS Model		SAR Model		SEM Model	
	Coefficient	*t-stat*	*Coefficient*	*Z-value*	*Coefficient*	*Z-value*
CONSTANT	2.12E-01	4.53	1.24E-01	2.70	2.09E-01	4.33
MINUTES	4.01E-03	3.28	3.62E-03	3.09	4.90E-03	3.87
WEEKEND	-8.93E-02	-2.29	-4.21E-02	-1.09	-4.90E-02	-1.25
PCT_NONWHITE	1.67E-03	8.17	1.20E-03	5.78	1.70E-03	7.25
PCT_BACHELORS	-1.50E-03	-5.20	-9.95E-04	-3.52	-1.52E-03	-4.78
PCT_UNEMPLOYED	2.16E-03	3.05	1.67E-03	2.47	1.81E-03	2.47
HHSIZE	2.13E-02	2.30	1.21E-02	1.38	1.28E-02	1.27
	n	772	n	772	n	772
	R^2_{adj}	0.56	LL	668.20	LL	659.50
			AIC	-1318.4	AIC	-1301.0

An unbalanced panel dataset for the entire year contains 9,296 observations (ranging from 772 to 778 census tracts with data for 2 to 12 months, a result of the sampling approach to handling the large amount of raw data). The base OLS specification from above is utilized to test for fixed effects (FE) and random effects (RE). Under fixed effects, the model controls for possible biases in predicting shared trips within census tracts, a valid assumption given that the time-invariant variables (e.g., socioeconomic and built environment) are unique to each tract, even though they are correlated across space. However, a fixed effects model assumes that a census tract's error term and individual-specific constant will not be correlated with others. Given that sharing exhibits some spatial dependence, the assumptions of fixed effects may not be suitable, and thus a random effects model was also developed.

To understand which approach was more suitable for the panel data, the Hausman test was used with a null hypothesis that the individual effects are not correlated with the regressors in the model (hence random effects methods are preferred). The resulting p-value from this test is statistically significant, implying that the fixed effects model is the better choice. As the fixed effects model rejected all variables that were constant within the zones, leaving only average trip variables (e.g., trip fare, trip miles, and trip minutes), a time-fixed effects model accounting for variation across the months of the year was employed to predict the ratio of shared trips and to account for the decline in the ratio of shared trips (Table 3). The time-fixed effects model is a better choice for predicting the ratio of shared trips by tract, as determined by a Lagrange Multiplier test for time effects.

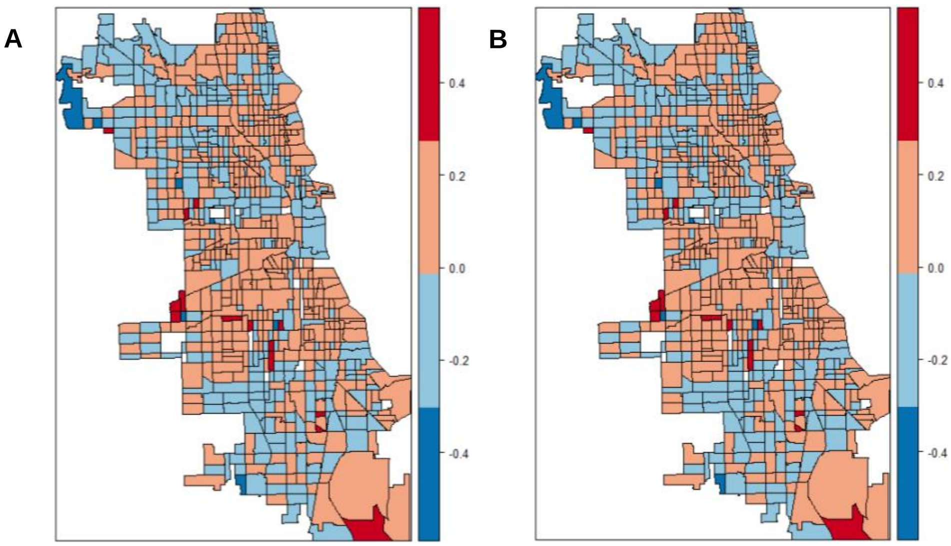

FIGURE 6 Plot of residuals from SAR and SEM models on January 2019 subset

TABLE 3 Fixed Effects and Time-Fixed Effects Model for Ratio of Shared Trips on 2019 Panel

	Fixed Effects		Time-Fixed Effects	
	Coefficient	*t-stat*	*Coefficient*	*t-stat*
MINUTES	5.40E-03	12.04	5.17E-03	13.92
January (base)				
February			-2.72E-02	-6.32
March			-3.51E-02	-8.15
April			-4.31E-02	-9.99
May			-6.48E-02	-14.95
June			-9.83E-02	-22.61
July			-1.18E-01	-27.27
August			-1.36E-01	-31.56
September			-1.52E-01	-35.24
October			-1.70E-01	-39.49
November			-1.75E-01	-40.66
December			-1.64E-01	-38.13
R^2_{adj}	-0.07		0.31	

Count of Shared Trips

The final OLS specification for the count of shared trips is presented in Table 4 below, indicating that built environment and socioeconomic variables can significantly explain the count of shared trips. Similarly to the ratio model, the lowest p-value from Moran's I test came from using the rook criterion spatial weight matrix and the null hypothesis could be rejected. As there is no direct equivalency of a SAR and SEM model for count data (see Glaser (2017) for a full discussion on spatial regressions of count data), a panel data approach was taken to accommodate temporal variation in trip.

TABLE 4 Final OLS Model for the Count of Shared Trips in January 2019

	Coefficient	*t-stat*
CONSTANT	3.73E+02	3.09
FARE	-1.10E+01	-1.91
WEEKEND	-2.91E+02	-2.95
HHSIZE	-7.47E+01	-4.01
POP_DENS (1,000/SQ MI)	-1.69E+00	-2.40
PCT_AGE18TO34	6.72E+00	6.09
HH_INCOME ($1,000)	1.50E+00	4.01
COMMUTE	-6.25E+00	-3.03
DENS_ENTER	1.32E+01	6.22
DENS_RET	1.60E+01	3.95
NTWKDENS_MM	3.11E+01	7.36
n	772	
R^2_{adj}	0.49	

When accounting for temporal variation of shared trip making in TNCs, the magnitude of the results changes, but the direction (i.e., positive and negative) relationship between each predictor and the outcome remains the same. An unbalanced panel dataset for the entire year was used to develop both a fixed effects and random effects model for the count of shared trips. The Hausman test was used with the same null hypothesis as before (i.e., individual effects are not correlated with the regressors in the model). The resulting p-value is not statistically significant, implying that the random effects model is the better choice (even though the model has a very poor adjusted R_2 value). Table 5 presents the results for the random effects model predicting the count of shared trips. The interpretation of the coefficients is difficult as it includes both within-tract effects due to changes in TNC behavior and between-tract effects (i.e., the average effect of X over Y when X changes over time and between tracts by one unit). For example, tracts with a higher density of retail and entertainment jobs per acre have a strong positive effect on shared trip-making as do those with a high density of linear miles of multimodal infrastructure. Interestingly, increased population density generally reduces the count of shared trips, but its possible interaction with the other job-oriented density terms is left unexplored.

TABLE 5 Final Panel Model with Random Effects for the Count of Shared Trips in 2019

	Coefficient	*Z-value*
CONSTANT	1.44E+02	1.91
WEEKEND	-2.93E+00	-3.81
HHSIZE	-4.23E+01	-3.12
POP_DENS (1,000/SQ MI)	-2.93E+00	-3.81

PCT_AGE18TO34	4.94E+00	6.71
HH_INCOME ($1,000)	6.87E-01	2.83
COMMUTE	-4.76E+00	-3.39
DENS_ENTER	9.99E+00	6.73
DENS_RET	9.17E+00	3.24
NTWKDENS_MM	2.17E+01	7.47
FRQ_TR	5.83E-03	2.26
HOLIDAY	2.42E+02	8.12
n	772-778	
R^2_{adj}	0.08	

Although the panel model in Table 5 considers between-tract effects, the spatial dependence of shared TNC trips is not captured. Hence, a balanced spatial panel data model of 763 census tracts followed to reflect these spatial interactions using a rook criterion contiguity spatial weights matrix. Both fixed effects and random effects were developed after adding a spatial autoregressive error term, but not an additional lag dependent variable. The spatial Hausman test comparing the two types resulted in an insignificant p-value of 0.178, revealing that the random effects model is preferred. Table 6 below presents the specification of this preferred random effects model by generalized moments (GM). This model is then compared to one developed through maximum likelihood (ML) with serial error correlation and a spatial lag dependent variable. The spatial autoregressive component, λ, for this new model type is statistically significant (t-stat of 26.07), but ϕ, the "ratio of the random effect's to the idiosyncratic error's variance," is about 0, meaning no random effect is truly present (Millo, 2014). The random effects assumption is removed, but the autoregressive serial correlation and spatial lag parameters are kept (Fixed Effects - ML), showing that the magnitude and sign of the coefficients largely do not change; but given the results of the Hausman test, the "Random Effects – GM" is the most appropriate spatial linear panel count model.

TABLE 6 Spatial Panel Model with Random and Fixed Effects for Count of Shared Trips

	Random Effects - GM		Random Effects - ML		Fixed Effects - ML	
	Coefficient	*t-stat*	*Coefficient*	*t-stat*	*Coefficient*	*t-stat*
CONSTANT	2.74E+01	0.42	3.69E+01	0.53	3.49E+01	0.51
FARE	-2.12E+00	-2.56	-6.89E-01	-2.59	-6.90E-01	-2.60
POP_DENS (1,000/SQ MI)	-3.29E+00	-4.25			-1.47E+00	-2.67
PCT_AGE18TO34	5.32E+00	7.28	5.11E+00	6.66	5.82E+00	7.20
HH_INCOME ($1,000)	8.23E-01	3.33				
COMMUTE	-4.45E+00	-3.09	-5.90E+00	-3.70	-5.64E+00	-3.55
DENS_RET	9.91E+00	3.52	1.18E+01	3.66	1.27E+01	3.95
DENS_ENTER	9.61E+00	6.40	1.03E+01	6.12	1.06E+01	6.34
NTWKDENS_MM	2.17E+01	7.45	2.32E+01	6.98	2.27E+01	6.82
FRQ_TR	8.34E-03	3.43				
HOLIDAY	1.95E+02	5.40				
	$\rho = 0.300$ σ_v^2 =9.62E+03		$\lambda = 0.328$ ϕ =1.00E-08 ψ = 9.82E-01		$\lambda = 0.329$ ψ = 9.82E-01	

CONCLUSION

The inclusion of density-specific and transit-oriented built environment variables in explaining the count of sharing was significant when examining the change in behavior both across time and space through a panel econometric approach. When isolating a specific period (e.g., January 2019) and investigating the spatial autocorrelation of the rate of shared trip making, only trip length and socioeconomic variables at the census tract level were significant predictors, both in SAR and SEM specifications. Considering the spatial correlation across neighboring tracts, a random effects model with a spatial autoregressive error term is a better fit than one with an additional spatial lag dependent term. Additionally, all count models appear to be best explained through random effects, while a time-fixed effects model best describes the ratio model for shared trips.

The results tend to indicate that longer TNC trips are a key indicator to the rate of sharing, while the cost of the ride has a negative relationship with the count of sharing by census tracts (although wealth disparities in areas of high unemployment and nonwhite residents may indirectly suppress the effects of increased fares on sharing). Operationalizing the fair incentive may be useful but requires studying sharing between origin-destination tracts and not pick-up tracts. Increasing average vehicle occupancy for long-distance trips can lead to sizeable decreases in vehicle emissions and on-road congestion and policymakers should further promote sharing to maximize these benefits. Areas with higher multimodal networks and employment for the retail and entertainment industries also exhibit higher counts of shared trips, while increased transit frequency leads to a significant but less substantial increase in shared rides, likely because transit competes for riders who would otherwise use a shared TNC service. Additionally, the share of riders taking a shared service (whether it was successfully shared or not) decreased throughout the year, necessitating further exploration to determine if TNC riders have become less willing to share rides, or if other factors are at play.

Tracts with a higher share of young people (ages 18 to 34) have a strong positive effect on the number of shared TNC trips but areas with a higher share of the population (ages 25 and above) with at least a bachelor’s degree have a strong negative relationship with the ratio of shared trips to total TNC trips. As Millennials (ages 24 to 39 in 2020) are more educated than previous generations, there is value in instilling the habit of sharing transportation to unlock benefits for years to come. This last part is critical in creating long term behavioral shifts that lead to higher usages of high occupancy vehicles – be it TNCs, transit, or shared autonomous vehicles (SAVs).

ACKNOWLEDGEMENTS

This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1610403. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: Dean, M.D.; data collection: Dean, M.D.; analysis and interpretation of results: Dean, M.D., and Kockelman, K.M.; draft manuscript preparation: Dean, M.D., and Kockelman, K.M. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

Alemi, F., Circella, G., Handy, S., Mokhtarian, P., 2018. What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California. Travel Behav. Soc. 13, 88–104. https://doi.org/10.1016/j.tbs.2018.06.002

Becker, H., Ciari, F., Axhausen, K.W., 2017. Modeling free-floating car-sharing use in Switzerland: A spatial regression and conditional logit approach. Transp. Res. Part C Emerg. Technol. 81, 286–299. https://doi.org/10.1016/j.trc.2017.06.008

CARB, 2019. SB 1014 Clean Miles Standard 2018 Base-year Emissions Inventory Report. California Air Resources Board.

Chan, N.D., Shaheen, S.A., 2012. Ridesharing in North America: Past, Present, and Future. Transp. Rev. 32, 93–112. https://doi.org/10.1080/01441647.2011.621557

Chen, X. (Michael), Zahiri, M., Zhang, S., 2017. Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. Transp. Res. Part C Emerg. Technol. 76, 51–70. https://doi.org/10.1016/j.trc.2016.12.018

Chen, Xiaowei, Zheng, H., Wang, Z., Chen, Xiqun, 2018. Exploring impacts of on-demand ridesplitting on mobility via real-world ridesourcing data and questionnaires. Transportation. https://doi.org/10.1007/s11116-018-9916-1

Clewlow, R.R., Mishra, G.S., 2017. Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States.

Cliff, A.D., Ord, J.K., 1973. Spatial autocorrelation, Monographs in spatial and environmental systems analysis. Pion, London.

Correa, D., Xie, K., Ozbay, K., 2017. Exploring the taxi and Uber demand in New York City: An empirical analysis and spatial modeling. Presented at the 96th Annual Meeting of the Transportation Research Board, Washington, D.C.

Dias, F.F., Lavieri, P.S., Garikapati, V.M., Astroza, S., Pendyala, R.M., Bhat, C.R., 2017. A behavioral choice model of the use of car-sharing and ride-sourcing services. Transportation 44, 1307–1323. https://doi.org/10.1007/s11116-017-9797-8

FHWA, 2019. FHWA NHTS Report: Changing Attitudes and Transportation Choices.

Glaser, S., 2017. A Review of Spatial Econometric Models for Count Data (Working Paper No. 19–2017). Faculty of Business, Economics, and Social Sciences, University of Hohenheim, Stuttgart.

Greene, W.H., 2003. Econometric analysis, 5th ed. Prentice Hall, Upper Saddle River, NJ.

Hausman, J.A., 1978. Specification Tests in Econometrics. Econometrica 46, 1251–1271. https://doi.org/10.2307/1913827

Hou, Y., Garikapati, V., Weigl, D., Henao, A., Moniot, M., Sperling, J., 2020. Factors Influencing Willingness to Pool in Ride-Hailing Trips. Transp. Res. Rec. J. Transp. Res. Board 036119812091588. https://doi.org/10.1177/0361198120915886

Ignacio Sarmiento-Barbieri, n.d. An Introduction to Spatial Econometrics in R [WWW Document]. URL http://www.econ.uiuc.edu/~lab/workshop/Spatial_in_R.html (accessed 3.24.20).

Khana, D., Rossen, L.M., Hedegaard, H., Warner, M., 2018. A BAYESIAN SPATIAL AND TEMPORAL MODELING APPROACH TO MAPPING GEOGRAPHIC VARIATION IN MORTALITY RATES FOR SUBNATIONAL AREAS WITH R-INLA. J. Data Sci. JDS 16, 147–182.

Lavieri, P.S., Bhat, C.R., 2019. Investigating objective and subjective factors influencing the adoption, frequency, and characteristics of ride-hailing trips. Transp. Res. Part C Emerg. Technol. 105, 100–125. https://doi.org/10.1016/j.trc.2019.05.037

Levine, N., Lord, D., Park, B.-J., Geedipally, S., Teng, H., Sheng, L., 2013. CrimeStat IV: Chapter 19 - Spatial Regression Modeling (Part V: Spatial Modeling II). Ned Levine & Associates and the National Institute of Justice, Washington, D.C.

Levy, J., 2020. Transportation Network Providers - Trips [WWW Document]. City Chic. Data Portal. URL https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p (accessed 3.22.20).

Li, W., Pu, Z., Li, Y., (Jeff) Ban, X., 2019. Characterization of ridesplitting based on observed data: A case study of Chengdu, China. Transp. Res. Part C Emerg. Technol. 100, 330–353. https://doi.org/10.1016/j.trc.2019.01.030

Lord, D., Park, B.-J., 2013. Appendix C - Negative Binomial Regression Models and Estimation Methods. Ned Levine & Associates and the National Institute of Justice, Washington, D.C.

Ma, X., Zhang, J., Ding, C., Wang, Y., 2018. A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. Comput. Environ. Urban Syst. 70, 113–124. https://doi.org/10.1016/j.compenvurbsys.2018.03.001

Millo, G., 2014. Maximum likelihood estimation of spatially and serially correlated panels with random effects. Comput. Stat. Data Anal. 71, 914–933. https://doi.org/10.1016/j.csda.2013.07.024

Millo, G., Piras, G., 2012. splm: Spatial Panel Data Models in R. J. Stat. Softw. 47.

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142

Mutl, J., Pfaffermayr, M., 2011. The Hausman test in a Cliff and Ord panel model. Econom. J. 14, 48–76. https://doi.org/10.1111/j.1368-423X.2010.00325.x

Reck, D.J., Guidon, S., Axhausen, K.W., 2020. Modelling shared e-scooters in Louisville, Kentucky: A spatial regression approach. Presented at the 99th Annual Meeting of the Transportation Research Board, Washington, D.C.

Schwieterman, J., Smith, C.S., 2018. Sharing the ride: A paired-trip analysis of UberPool and Chicago Transit Authority services in Chicago, Illinois. Res. Transp. Econ. 71, 9–16. https://doi.org/10.1016/j.retrec.2018.10.003

Shaheen, S., Cohen, A., 2019. Shared ride services in North America: definitions, impacts, and the future of pooling. Transp. Rev. 39, 427–442. https://doi.org/10.1080/01441647.2018.1497728

Sperling, D., 2018. Three Revolutions: Steering Automated, Shared, and Electric Vehicles to a Better Future. Island Press, Washington, D.C.

Srinivasan, S., Kockelman, K.M., 2002. The Impacts of Bypasses on Small- and Medium-Sized Communities: An Econometric Analysis. J. Transp. Stat. 5, 13.

Union of Concerned Scientists, 2020. Ride-Hailing’s Climate Risks: Steering a Growing Industry toward a Clean Transportation Future. Union of Concerned Scientists, Cambridge, MA.

Wang, Y., Kockelman, K.M., Wang, X. (Cara), 2013. Understanding spatial filtering for analysis of land use-transport data. J. Transp. Geogr. 31, 123–131. https://doi.org/10.1016/j.jtrangeo.2013.06.001

Yu, H., Peng, Z.-R., 2019. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. J. Transp. Geogr. 75, 147–163. https://doi.org/10.1016/j.jtrangeo.2019.01.004

Zheng, H., Chen, X., Chen, X.M., 2019. How Does On-Demand Ridesplitting Influence Vehicle Use and Purchase Willingness? A Case Study in Hangzhou, China. IEEE Intell. Transp. Syst. Mag. 11, 143–157. https://doi.org/10.1109/MITS.2019.2919503