A MODEL OF RIDESOURCING DEMAND GENERATION AND DISTRIBUTION

Patrícia S. Lavieri
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St, Stop C1761, Austin, TX 78712, USA
Tel: 512-471-4535; Email: lavieri@utexas.edu

Felipe F. Dias
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St, Stop C1761, Austin, TX 78712, USA
Tel: 512-471-4535; Email: fdias@utexas.edu

Natalia Ruiz Juri
The University of Texas at Austin
Center for Transportation Research, Network Modeling Center
3925 W. Braker Ln, Stop D9300, Austin, TX 78759, USA
Tel: 512-232-3099; Email: nruizjuri@mail.utexas.edu

James Kuhr
The University of Texas at Austin
Center for Transportation Research, Network Modeling Center
3925 W. Braker Ln, Stop D9300, Austin, TX 78759, USA
Tel: 512-232-3115; Email: jkuhr@utexas.edu

Chandra R. Bhat (corresponding author)
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St, Stop C1761, Austin, TX 78712, USA
Tel: 512-471-4535; Email: bhat@mail.utexas.edu
and
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Lavieri, Dias, Ruiz Juri, Kuhr, and Bhat

ABSTRACT
Ridesourcing has had an exponential growth in the past years, yet its impacts on individual travel are unclear and have not been adequately examined. Recently, an Austin-based ridesourcing company released a large dataset containing disaggregate trip-level information. In this research, we use this new dataset in tandem with several publicly available data sources to estimate two models: a spatially lagged multivariate count model, which is used to describe how many trips are generated in a specific zone on both weekdays and weekend days, and a fractional split model, which helps us identify the characteristics of zones that attract ridesourcing trips. Our results show spatial dependence in ridesourcing trips among proximally located zones, as well as correlation between weekday and weekend day trips originating in a zone. Another interesting finding is the identification of a possible substitution effect between ridesourcing and transit use for weekday trips. Moreover, our results suggest that different income segments in the population may use ridesourcing for different activity purposes. From a travel behavior researcher perspective, the results in this paper identify aggregate area-level variables impacting ridesourcing, which can guide future efforts to better understand the demand for ridesourcing as well as the demand for autonomous and connected vehicles.

1. INTRODUCTION

The concept of ridesourcing, in which a loose collection of drivers offers transportation services to the wider public based on coordination through a centralized dispatch (generally a mobile application), has introduced a new form of modern mobility. This type of service is coordinated by Transportation Network Companies (TNC), such as Uber and Lyft, and has experienced exponential growth in the past years. For instance, it took Uber, the largest TNC, six years to reach the one billion trip milestone in 2015, but only six additional months to reach the two billion milestone. One year after that, the company exceeded 5 billion trips (*1*). Such a substantial growth reflects the convenience that ridesourcing offers to users by being a reliable, lower cost (compared to traditional taxi services), on-demand and door-to-door transportation service that is requested and paid through smartphone apps. Additionally, these services are also able to easily reach a large geographically dispersed market, because they are less regulated than taxi companies and rely on drivers' personal vehicles.

Even as ridesourcing has gained considerable traction and is widely prevalent today in most urban areas, its impacts on individual travel are unclear and have not been adequately examined. This includes limited knowledge of which travel modes are being substituted by ridesourcing, what its potential impacts on vehicle ownership are, how ridesourcing may affect peak and off-peak travel patterns, and whether its convenience induces more or less travel. A main reason for the lack of studies on ridesourcing impacts is the scarcity of publicly available data. However, to fill this void, some researchers have resorted to specialized user surveys or large-scale household travel surveys that collect limited information on ridesourcing preferences. For instance, Rayle et al. (*2*) present results on an exploratory survey conducted among ridesourcing users (n=321) in San Francisco, and observe that, while ridesourcing is being used in lieu of taxis in many instances, at least one-half of the users replaced trips by modes other than taxi, including public transit and driving. Clewlow and Mishra (*3*) studied the impact of ridesourcing on vehicle ownership, conducting a survey in seven major U.S. Cities. In their study, 9% of the respondents indicated that they had made a decision to dispose one or more household vehicles because of ridesourcing availability. The remaining 91% reported that they had not changed their vehicle ownership, with 16% indicating that they had no vehicles to begin with. Dias et al. (*4*) studied determinants of the frequency of both ridesourcing and car-sharing trips, using a survey undertaken by the Puget Sound Regional Council (PSRC) and found that vehicle ownership impacts the these two modes differently. Households with vehicles are less likely to use car-sharing services; however, households with one or more vehicles and residing in a high-density location are more likely than their peers residing in low density areas to use both ridesourcing and car-sharing services. Finally, Lavieri et al. (*5*), using the same PSRC survey, analyzed the determinants of ridesourcing adoption (as one model component of a larger model system), and found that living in high residential density areas and having a tech-savvy life-style were important contributing factors that encouraged ridesourcing adoption. Overall, the transportation literature on ridesourcing is in its infancy, and studies based on actual trip records are virtually non-existent.

This paper contributes to the literature by modeling and analyzing the demand for ridesourcing based on an open source database released by RideAustin, a nonprofit TNC in Austin, Texas. Using six months of detailed trip data, including trip origin and destination location and corresponding time stamps, we develop a two-step modeling framework to investigate the generation and distribution of daily ridesourcing trips at the traffic analysis zone (TAZ) level (the RideAustin data set does not provide user information and corresponding socio-

demographic characteristics; therefore, our analysis is undertaken at the zonal level and relies on zonal demographics to infer ridesourcing demand characteristics). We use a spatial bivariate count model to analyze ridesourcing trip generation and inform our understanding of the characteristics of the demand for this service. The use of a spatial analysis technique is important because spatial dependencies in TAZ-level trip generation are likely to exist. Thus, for example, it is possible that individuals in close proximity (say in neighboring TAZs) will be influenced by each other's ridesourcing propensity through social interactions, leading to a lagged endogenous variable effect. In our analysis, we consider this type of spatial dependence through the use of a spatial lag model. Subsequently, we apply a fractional split distribution model to identify zonal characteristics that attract ridesourcing trips and to examine how far individuals are willing to travel by this mode. Examples of explanatory variables used in our analysis are zonal distributions of income, gender, race, age, population and employment density. We also consider transit supply characteristics and land use information regarding presence of parks and universities.

The remainder of this paper is organized as follows. The next section provides a description of the multiple data sources used in our analysis and of characteristics of the area of our analysis. Section 3 describes the modeling frameworks. Section 4 presents the results. The final section summarizes our findings.

2. DATA

Several public data sets were compiled to undertake the analysis. The primary data source originated from RideAustin, a TNC operating in Austin, Texas. RideAustin entered the Austin ridesourcing market in 2016, shortly after Uber and Lyft shut down their operations in the city due to disputes over local regulations. The RideAustin data (*6*) provides trip-level information, including the location of trip origins and destinations, total trip length, and corresponding fare. To protect their clients' privacy, RideAustin added noise to the locations of the pickups and drop-offs. The dataset contains a total of 1,494,125 trips that occurred between June 4th, 2016 and April 13th, 2017. Since ridership during the first few months was limited, our analysis only includes data from August 2016 through January 2017. Based on information provided by the Austin Department of Transportation, we estimate that, during that semester, RideAustin was responsible for one third of Austin's ridesourcing market share, suggesting high representativeness of the data. The trip information is supplemented using TAZ-level demographic data obtained from the Capital Area Metropolitan Planning Organization (CAMPO) website and planning toolkit, and the most recent Census estimates. GTFS (General Transit Feed Specification) data is used to estimate the characteristics of the transit system (*7*).

2.1 Data Preparation

Raw data, including trip origins and destinations, transit availability, land use, and demographics were mapped to the TAZs defined by CAMPO using GIS software. Given the sparseness of origins and destinations in the outskirts of the city, we chose to focus this study on trips that originate in Central Austin, in the region delimited by Highway 183 to the east and north, Highway 290 to the south, and Texas State Highway Loop 1 (MoPac) to the north. There are a total of 458 TAZs in the area of analysis. The Austin-Bergstrom International Airport is outside the area of interest, but it attracts a large number of trips, so it was modeled as a special external zone in the fractional split model; a second dummy TAZ was used as the destination of all the trips that end outside the study area.

The trip data processing involved calculating the average number of daily trips per origin, and the corresponding average daily split by destination. Separate values were computed for weekdays and weekends. Demographic variables by TAZ were computed using data from the most recent Census, while land-use variables were obtained from the CAMPO planning model. GTFS data was aggregated to generate metrics of transit accessibility, including transit stops per zone, and the average frequency of buses per stop for weekdays and weekends.

2.2 Data Description

This research models the average daily count of ridesourcing trips for weekdays and weekends. The spatial unit of analysis is a TAZ. Table 1 presents the descriptive statistics for all the variables used in the model, and the year when the corresponding data was collected.

The analysis of descriptive statistics shows a large dispersion in the number of trips generated per zone. Figure 1 illustrates the spatial distribution of trips on an average week and weekend day. There is a clear concentration of trips in central and denser areas on both types of days. During weekdays, trips are more concentrated in specific zones that contain universities, parks, or active night life. On average there are almost four times more trips generated on a weekend day than on a weekday. These observations are consistent with Rayle et al.'s (*2*) results, which suggest, as in San Francisco, that ridesourcing in Austin too is used more for social and leisure activities than work-related activities. Indeed, Hampshire et al. (*8*) recently conducted an online survey in Austin and identified the same pattern. On a related note, the average cost of a ridesourcing trip in our sample is US$12.77.

The analysis of transit supply variables suggests that the distribution of transit in Austin is rather heterogeneous. The frequency of bus service averages at 3.12 per hour during weekdays (an average headway of about 20 minutes) and averages 1.64 per hour on weekend days (an average headway of about 37 minutes).

In terms of socio-demographics, there is again a large variation across zones in population density and employment density. The race/ethnicity and education variables indicate a predominantly white and highly educated population. There is a good distribution of individuals in the 18-60 age range. Households are small in size (average of less than 2 individuals), have a mean income of $48,000, and have high vehicle ownership rates (more than half of the sample has at least two vehicles per household). Finally, three variables considered in our model, but not presented in the table, are binary variables representing the presence of parks in a zone, the presence of The University of Texas (UT) territory in a zone, and an indicator of whether a zone is a central business district.

3. ANALYTIC FRAMEWORK

In this paper, we develop a two-step procedure to analyze ridesourcing trip generation and distribution between TAZs on weekdays and weekends. In the trip generation analysis, the average number of trips generated at each TAZ on an average weekday is modeled jointly with the average number of trips generated on an average weekend day. We utilize a spatial bivariate count model that takes into consideration the spatial dependence between TAZs as well as the correlation between the two types of days. Accounting for spatial dependency in trip generation models, as discussed earlier, is important because one can expect neighboring zones to present similar travel demand patterns, especially when considering that the delimitation of zone borders is usually made based on the transportation network and does not necessarily reflect differences in demand patterns. Of equal importance is to recognize that ridesourcing trip generation rates

may vary between weekdays and weekends, but that common TAZ-level unobserved factors may influence the counts on both types of days. For instance, a zone with very limited parking (an unobserved variable in our analysis) is likely to be associated with high ridesourcing trip generation rates, both on weekdays and weekend days.

For the trip distribution analysis, we develop a fractional split distribution model that analyzes the fractions of trips originating from a zone that terminate in each destination zone. This analysis provides an understanding of factors that "pull" ridesourcing trips toward a zone. The next two sections provide an overview of both the generation and distribution models.

3.1 Spatial Multivariate Count Model
The spatial multivariate count model used in this paper is based on Bhat et al. (*9*). There are two components to this model. The first part is the recasting of the basic count model and the second part is the spatial dependency formulation.

3.1.1 Count Model Recasting
The framework used here is based on a recasting of the basic count model as a special case of a generalized ordered-response (GOR) model, as proposed by Castro, Paleti, and Bhat, 2012 (*10*). In this approach, the count is viewed as a result of a latent demand generation propensity that gets mapped into the observed trip counts through thresholds that are themselves functions of exogenous variables. This approach offers the advantage of accommodating over-dispersion and excess zeros, which is useful when modeling zones that do not produce any trips (for example, open areas) and zones that produce unusually high numbers of trips (for example, zones that have high density of bars and active night life).

Let q ($q = 1, 2, \ldots, Q$) be the index for the territorial unit of analysis (a "TAZ" in the current paper) and let s ($s = 1, 2, \ldots, S$) be the index for day-type (weekday or weekend day in the current paper). Let y_{qs} be the index for the count of trips generated in a day-type s in a TAZ q, and let m_{qs} be the actual observed count of trips in the day-type s in the TAZ q. Next, consider that there is a TAZ-specific demand function that represents the propensity for trip generation on day-type s. This propensity is not directly observed, and so may be represented by a latent (unobserved to the analyst) variable y_{qs}^*. Then, in the generalized ordered response (GOR) notation, the latent propensity y_{qs}^* is written as a function of a $(K \times 1)$-vector of observed covariates $\boldsymbol{x}_q$ (excluding the constant) as:

$$y_{qs}^* = \boldsymbol{\beta}_s' \boldsymbol{x}_q + \eta_{qs}, \quad y_{qs} = m_{qs} \implies \psi_{qs,m_{qs}-1} < y_{qs}^* < \psi_{qs,m_{qs}}. \tag{1}$$

In the above specification, $\boldsymbol{\beta}_s$ is a $(K \times 1)$-vector whose elements capture the effects of the $\boldsymbol{x}_q$ variable vector on latent demand propensity y_{qs}^*. Finally, η_{qs} captures TAZ-specific unobserved factors that increase or decrease the latent propensity for generating trips in a week or weekend day. The thresholds in Equation (1) take the following form:

$$\psi_{qs,m_{qs}} = \Phi^{-1}\left(e^{-\lambda_{qs}} \sum_{l=0}^{m_{qs}} \frac{\lambda_{qs}^l}{l!} \right) + \alpha_{s,m_{qs}}, \quad \lambda_{qs} = e^{\gamma_s' \boldsymbol{h}_q}, \alpha_{s,0} = 0, \alpha_{s,m_{qs}} = \alpha_{s,L_s} \text{ if } m_{qs} > L_s, \tag{2}$$

where Φ^{-1} is the inverse function of the univariate cumulative standard normal, $\psi_{qs,-1} = -\infty \; \forall q, \; s$, $\alpha_{s,0} = 0 \; \forall s$ (this restriction is needed for identification, given the

parameterization of the thresholds), $\boldsymbol{h}_q$ is a vector of exogenous variables (including a constant) associated with TAZ q (there can be common variables in $\boldsymbol{h}_q$ and $\boldsymbol{x}_q$), $\boldsymbol{\gamma}_s$ is a corresponding coefficient vector to be estimated for day-type s, and L_s is a pre-defined count level that is determined based on empirical testing and on the context under consideration. Note that thresholds are impacted by the TAZ characteristics so that the translation from the trip generation propensity into an actual number of trips generated may vary between two zones (based on the exogenous variables $\boldsymbol{h}_q$) even if they have exactly the same $\boldsymbol{x}_q$. As in the typical ordered-response framework, the values of $\alpha_{s,m_{qs}}$ should be such that the ordering condition on the thresholds $(-\infty < \psi_{qs,0} < \psi_{qs,1} < \psi_{qs,2} < ....)$ is satisfied. The presence of the $\alpha_{s,m_{qs}}$ terms provides flexibility to accommodate high or low probability masses for specific count outcomes without the need for cumbersome treatment using hurdle or zero-inflated mechanisms (see Castro, Paleti, and Bhat (*10*)). If these terms are set to zero, and all elements of the vector $\boldsymbol{\beta}_s$ are also set to zero, the result is the traditional Poisson count model (*10*).

3.1.2 Spatial Component
The model adopted in this study considers spatial dependence across TAZs in observed covariates $\boldsymbol{x}_q$ vector as well as in the unmeasurable terms η_{qs}. To conserve on space, we refer the reader to Bhat et al. (*9*) for a complete explanation and formulation of the spatial structure of the model. In general terms, we have the usual distance-based spatial weight matrix (**W**), which indicates whether a pair of TAZ should be considered spatially dependent $(w_{qq'} > 0)$. The trip generation propensity from a zone is influenced by exogenous variables specific to that zone, and the trip generation propensity from neighboring zones based on an autoregressive coefficient represented by δ_s. δ_s may be positive or negative $(-1 < \delta_s < 1 \; \forall \; s)$. In our model we adopt weight matrices based on functions of the distance between the center points of two TAZs. Since we are analyzing the central area of Austin, we consider that spatial dependence may occur between zones that are up to 3 kilometers apart. The final equation of the multivariate count model incorporating spatial dependence for a specific zone for a specific type of day is:

$$y_{qs}^* = \delta_s \sum_{q'=1}^{Q} w_{qq'} y_{q's}^* + \boldsymbol{b}_s' \, \boldsymbol{x}_q + \eta_{qs}, \quad y_{qs} = m_{qs} \implies \psi_{qs,m_{qs}-1} < y_{qs}^* < \psi_{qs,m_{qs}}. \tag{3}$$

We consider the joint nature of the demand propensities across day-types for each TAZ q by allowing the elements η_{qs} to be correlated across the two day types (s=1, 2) for each TAZ q. A final important point to be noted here is that the spatial dependency in counts is generated through spatial "spillover" effects and spatial error correlation effects in the latent ridesourcing demand propensity, not through the localized TAZ-specific characteristics that impact the thresholds in the count model.

3.2 Fractional Split Distribution Model
To estimate the fractional split model, we use a quasi-likelihood estimation approach (see Sivakumar and Bhat (*11*); Gourieroux et al. (*12*)). Let f_{qj} be the fraction of the total trips that

originate in zone q that terminate in zone j, such that $\sum_{j=1}^{J} f_{qj} = 1$. We can write the log-likelihood for the qj^{th} zone pairing as follows:

$$LL_{qj} = f_{qj} \log\left[G\left(\boldsymbol{\mu}, z_{qj}\right)\right] + \left(1 - f_{qj}\right)\log\left[1 - G\left(\boldsymbol{\mu}, z_{qj}\right)\right] \tag{4}$$

where $\boldsymbol{\mu}$ is a vector of coefficients to be estimated and z_{qj} is a vector of exogenous variables with characteristics of zones q and j. The function $G(.)$ takes the following logit functional form:

$$G\left(\boldsymbol{\mu}, z_{qj}\right) = \frac{1}{1 + e^{-\mu' z_{qj}}} \tag{5}$$

4. RESULTS

In this section we present and discuss the results for the trip generation and distribution models. We considered all the variables presented in Table 1 in our analysis. Several variables and functional forms (including logarithmic transformations) were tested to arrive at the final specification. The model estimation process was guided by prior research, intuitiveness, and parsimony considerations. A few variables that were only marginally statistically significant (i.e., not significant at the 0.05 level of significance) were retained in the final model specification because of their intuitive effects and potential to guide future research efforts.

4.1 Trip Generation by Day-Type

As discussed in Section 3.1.1, the count variable is viewed as a result of a latent demand generation propensity that gets mapped into the observed trip counts through thresholds that are themselves also functions of exogenous variables. The first half of Table 2 presents the results for the demand generation propensity. In the first row of results for the weekday trips model, we observe a positive effect of the variable representing the presence of The University of Texas (UT) in the zone. This positive effect indicates that UT is an area with a high intensity of ridesourcing trip origins during a typical weekday, presumably a combination of activity opportunities in the UT area and because students are a segment of the population more likely to use ridesourcing than other population segments. The effect of transit supply in Table 2 indicates the expected negative effect, suggesting that ridesourcing decreases as transit service improves. Another perspective is that ridesourcing tends to get used more in areas with relatively poor transit service. Areas with higher residential density and activity intensity, not surprisingly, have more originating ridesourcing trips, a finding supported by earlier studies (see *3*, *4*, and *5*). Interestingly, on weekdays, more ridesourcing trips are generated from high activity intensity zones than from high population density zones, while the reverse holds on weekend days. This suggests that ridesourcing is more used after an out-of-home activity on weekdays, and more used from home as an individual leaves home for an out-of-home activity on weekend days.

In terms of population characteristics, zones with higher proportions of white population present a lower propensity to generate ridesourcing trips, both during the weekday and the weekend day. The white population is historically associated with a higher use of the "drive alone" mode (*13*, *14*) than other segments of the population, which may explain the negative signs on ridesourcing use. The proportion of young adults (18 to 29 years of age) in the zone contributes to an increase in the propensity of ridesourcing trips. This result corroborates findings from previous studies (*2*, *3*, *8*). The effect of the median household income of the zone is interesting. It shows that wealthier areas are associated with an increase in the weekday ridesourcing trips, but a decrease in weekend ridesourcing trips. The literature often suggests that

ridesourcing users are in the high income segments of the population (*3*, *4*). Our results are not inconsistent with the previous literature, but suggest that there is heterogeneity in the income effect based on day of the week. Perhaps high income individuals "buy" time on weekdays through ridesourcing (because they can work or relax rather than drive), while low income individuals gravitate toward ridesourcing on weekend days because of relatively poor transit levels of service. Another possible explanation could be that higher income individuals conduct more social and recreational activities during the week (compared to lower income segments) and use ridesourcing to access these activities. Finally, as expected higher rates of vehicle ownership are associated with a decrease in the generation of ridesourcing trips, a result also observed in Dias et al. (*4*) and Lavieri et al. (*5*).

The second half of Table 2 presents the threshold results. The elements of the α vector do not have any substantive interpretations, but play the very important role of accommodating high or low probability masses for specific outcomes The elements in the γ vector are presented next in Table 2. The constants within the γ vector do not have any particular interpretation. For the other variables, a positive coefficient shifts all the thresholds toward the left of the demand intensity scale (see Castro, Paleti and Bhat (*10*) for a detailed discussion), which has the effect of reducing the probability of the zero trip count. A negative coefficient, on the other hand, shifts all thresholds toward the right of the generation propensity scale, which has the effect of increasing the probability of the zero count. We observe that the proportion of the male population in a zone has opposite effects on weekdays and weekends. Zones that have a higher male population proportion are more likely (than zones with a higher female population proportion) to have non-zero ridesourcing trips during the weekday and zero ridesourcing trips during the weekend days. Also, zones with high vehicle ownership rates are more likely to have zero ridesourcing trips generated in a weekday, while zones that have parks are less likely to have zero trips generated in a weekend. Both results are expected, since having vehicles available in the household reduces the necessity of seeking alternative modes, while parks are associated with recreational activities that are more prevalent on weekends.

Finally, at the bottom of the table we present the cross-correlation between weekdays and weekend days as well as the spatial autoregressive parameter. For the spatial correlation between zones we tested two different weight matrices, one based on the inverse of the distance between the centroid of the zones and another based on the inverse of the squared distance. The best model fit was obtained with the first one. The results confirm ours hypothesis that the number of trips that a zone generates in a weekday is positively associated with the number of trips generated on a weekend day. Additionally, the number of trips in a zone is influenced by observed and unobserved factors of the neighboring zones.

4.2 Trip Distribution by Day-Type
The University of Texas has a positive effect on trip attractions during the weekday, suggesting that people might be using ridesourcing to access the campus area at significantly higher rates than other zones. This effect seems to disappear, however, on weekends. This is very likely due to the reduced number of activities on campus during weekends, which results in less people visiting (and traveling to) the area. While zones located in the central business district (CBD) do not attract ridesourcing trips more so than zones in other areas of town, the demand to such CBD zones does decrease on weekends, likely due to the lack of activities during that period. The coefficients related to the airport and external zones are somewhat difficult to interpret directly

since zones that fell in these categories had no associated data besides the trip cost. Therefore, many effects are entangled and cannot be immediately interpreted.

The proportion of retail employment, regardless of the day of the week, positively impacts trip attractions. Retail employment may be viewed as a proxy for opportunities for out-of-home activities: people will tend to travel more to places that have activities they want to partake in, such as shopping, and this appears to have a direct positive association with ridesourcing destination points. As expected, there is a positive influence of population density on trip attraction, representing return-home trips. Curiously, though, this effect is not statistically significant during the weekend. This could be a simple reflection of the higher number of out-of-home activities pursued during the weekend days. Thus, even if there are more ridesourcing return-home trips on weekend days than on weekdays, the proportion of such trips (as a fraction of total trips) may be lower on weekend days.

As explained previously, the negative effect of the proportion of the white population in a zone on ridesourcing trips generated from the zone may be a reflection of an intrinsic dislike for non-private travel (that is, a generic private auto-inclination). Similarly, the results in Table 3 indicate that, as the proportion of households with two or more vehicles in a zone increases, the "attractiveness" of the zone as a terminating point for ridesourcing trips decreases.

The average monetary cost of the trip plays a significant role in the trip distribution process for both weekends and weekdays. Throughout the estimation process, both distance and cost were used, but these two variables were too strongly correlated for both of them to be statistically significant. Therefore, given that the cost variable successfully explained most of the variance of these two variables, the distance variable was omitted from the estimation. The final variable is a pure size effect.

5. CONCLUSIONS

This paper has undertaken an analysis of the demand for ridesourcing trips in the city of Austin, Texas. Based on data provided by a non-profit TNC that entered Austin's market after the exit of Uber and Lyft, we develop two models that analyze characteristics of the generation and distribution of ridesourcing trips at a TAZ level. Several public data sets were compiled to complete the analysis, including TAZ-level demographic data obtained from the Capital Area Metropolitan Planning Organization, the most recent Census estimates, and GTFS available from the state of Texas website. The use of open source data is in its early stages and this paper provides a first glimpse of the potential that these data sources have in informing transportation models.

Our model provides important initial insights on characteristics of ridesourcing demand. Additionally, it identifies interesting heterogeneities between ridesourcing use on weekdays and weekend days. For example, in the context of university campuses, our results suggest that students may be the beneficiaries of the availability of ridesourcing services. This may be either because vehicle ownership rates among university students is lower (compared to working individuals), or because of restricted parking regulations and high parking fees in such areas. Moreover, bus frequencies seem to have a negative impact on the generation of ridesourcing trips during the week, suggesting a substitution effect between ridesourcing and transit use. Another interesting finding is that the effect of the median household income in a zone on trip generation is opposite for weekdays and weekend days, suggesting that different income segments in the population may use ridesourcing for different activity purposes. Overall, the estimated parameters of the multivariate count model can be used to forecast the number of new

ridesourcing trips in a TAZ in response to changing TAZ economics and demographics. The trip distribution model indicated that, as expected, the airport is a major ridesourcing trip attraction. This result leads to the question of whether only taxi and carpooling trips to the airport are being substituted, or if travelers who used to park at airports earlier are now opting for ridesourcing instead. A better understanding of this issue can help future parking planning at airports. Finally, the trip distribution model also provides evidence of the substantial use of weekday ridesourcing for returning home, which may suggest that ridesourcing is becoming integrated into the multi-modal use routine of individuals and/or is being used to avoid driving while impaired.

The results and methods used in this study can serve multiple purposes. First, from a travel behavior researcher perspective, we have identified aggregate-level variables that impact ridesourcing, and can guide efforts to better understand the demand for autonomous and connected vehicles in the future. Second, from a planner's perspective, we provide an analytic framework to develop predictive models of ridesourcing movements that can be accommodated in regional and planning network models. Finally, our results may also have relevance to operators in their understanding of travel demand, which can lead to better strategies to allocate drivers to rides, or to estimate optimal fleet size when entering a new market.

ACKNOWLEDGEMENTS

REFERENCES

1. Uber Blog. Uber Hits 5 Billion Rides Milestone. www.uber.com/en-SG/blog/uber-hits-5-billion-rides-milestone/. Accessed July 26, 2017.

2. Rayle, L., D. Dai, N. Chan, R. Cervero, and S. Shaheen. Just a Better Taxi? A Survey-Based Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco. *Transport Policy*, Vol. 45, 2016, pp. 168-178. https://doi.org/10.1016/j.tranpol.2015.10.004.

3. Clewlow, R. and G.S. Mishra. Shared Mobility: Current Adoption, Use, and Potential Impacts on Travel Behavior. Presented at the 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.

4. Dias, F.F., P.S. Lavieri, V.M. Garikapati, S. Astroza, R.M. Pendyala, and C.R. Bhat. A Behavioral Choice Model of the Use of Car-Sharing and Ride-Sourcing Services. *Transportation*, Vol. 44, No. 6, 2017, pp. 1307-1323. https://doi.org/10.1007/s11116-017-9797-8.

5. Lavieri, P.S., V.M. Garikapati, C.R. Bhat, R.M. Pendyala, S. Astroza, and F.F. Dias. Modeling Individual Preferences for Ownership and Sharing of Autonomous Vehicle Technologies. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2665, 2017, pp. 1-10. https://doi.org/10.3141/2665-01.

6. RideAustin. Data File and Code Book. Available from data.world: https://data.world/ride-austin/ride-austin-june-6-april-13. Accessed July 26, 2017.

7. Texas Government. Schedule Data for All Fixed Route MetroBus Lines and MetroRail. https://data.texas.gov/Transportation/GTFS/r4v4-vz24. Accessed July 26, 2017.

8. Hampshire, R.C., C. Simek, T. Fabusuyi, X. Di, and X. Chen. Measuring the Impact of an Unanticipated Suspension of Ride-sourcing in Austin, Texas. https://ssrn.com/abstract=2977969 or http://dx.doi.org/10.2139/ssrn.2977969. Accessed July 26, 2017.

9. Bhat, C.R., R. Paleti, and P. Singh. A Spatial Multivariate Count Model for Firm Location Decisions. *Journal of Regional Science*, Vol. 54, No. 3, 2014, pp. 462-502. https://doi.org/10.1111/jors.12101.

10. Castro, M., R. Paleti, and C.R. Bhat. A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections. *Transportation Research Part B*, Vol. 46, No. 1, 2012, pp. 253-272. https://doi.org/10.1016/j.trb.2011.09.007.

11. Sivakumar, A. and C. Bhat. Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1790, 2002, pp. 80-88. https://doi.org/10.3141/1790-10.

12. Gourieroux, C., A. Monfort, and A. Trognon. Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, Vol. 52, No. 3, 1984, pp. 681-700.

13. Giuliano, G. Travel, Location and Race/Ethnicity. *Transportation Research Part A*, Vol. 37, No. 4, 2003, pp. 351-372. https://doi.org/10.1016/S0965-8564(02)00020-4.

14. Smart, M.J. A Nationwide Look at the Immigrant Neighborhood Effect on Travel Mode Choice. *Transportation*, Vol. 42, No. 1, 2015, pp. 189-209. https://doi.org/10.1007/s11116-014-9543-4.

LIST OF FIGURES

FIGURE 1 Spatial distribution of trips on an average weekday (top) and on an average weekend day (bottom).
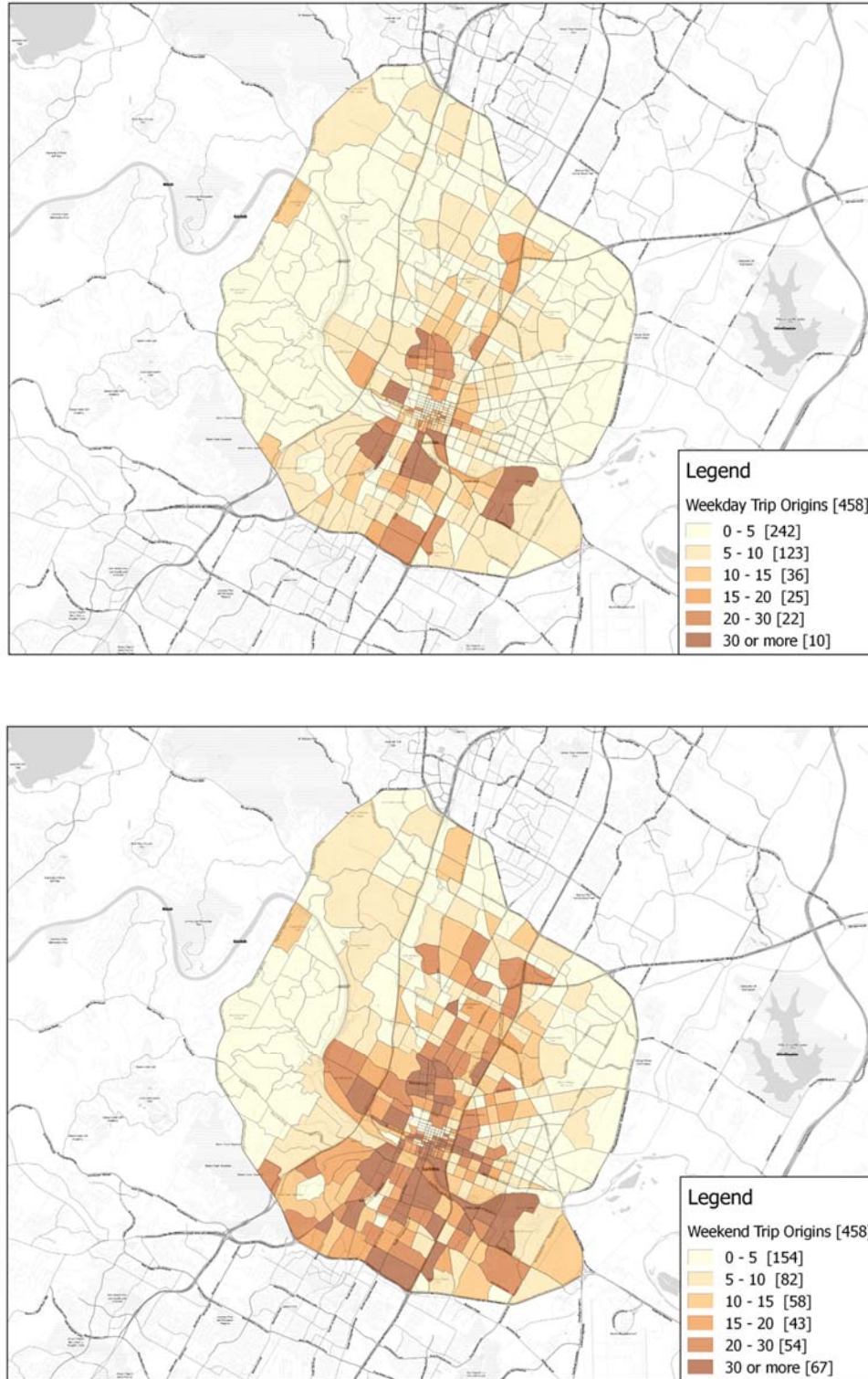

LIST OF TABLES

FIGURE 1 Spatial distribution of trips on an average weekday (top) and on an average weekend day (bottom).

TABLE 1 Sample Descriptive Statistics (458 TAZ)

Variable	Min.	Max.	Mean	Std. Dev.
Outcomes [2016]				
Number of trips in a weekday	0.00	125.00	8.62	11.18
Number of trips in a weekend day	1.00	420.00	31.45	19.89
Transit Supply [2016]				
Number of bus stops	0.00	27.00	3.49	3.62
Frequency of buses in a weekday (bus per hour)	0.00	20.90	3.12	3.25
Frequency of buses in a weekend day (bus per hour)	0.00	13.70	1.64	1.66
Socio-Demographic Variables [2010]				
Population density (population per km^2)	0.00	59,257.67	4,603.24	6,167.92
Employment density (employment per km^2)	0.00	161,932.22	6,813.73	17,545.97
Employment density in retail sector (employment per km^2)	0.00	46,442.92	1,167.67	3,731.79
Race/Ethnicity Variables[2015]				
Proportion of White population	0.00	1.00	0.65	0.34
Proportion of Black and African American population	0.00	0.65	0.05	0.09
Proportion of Asian population	0.00	0.63	0.04	0.06
Proportion of other races	0.00	0.45	0.07	0.08
Educational Attainment Distribution [2015]				
Proportion of population 18 years and above with less than Associate degree	0.00	1.00	0.31	0.26
Proportion of population 18 years and above with Associate or Bachelor's degree or higher	0.00	0.67	0.30	0.19
Proportion of population 18 years and above with Graduate degree	0.00	0.64	0.20	0.15
Age Distribution [2015]				
Proportion of population aged 17 years and below	0.00	0.48	0.13	0.11
Proportion of population aged 18-29 years	0.00	0.99	0.22	0.20
Proportion of population aged 30-39 years	0.00	0.44	0.15	0.10
Proportion of population aged 40-59 years	0.00	0.49	0.20	0.13
Proportion of population aged 60 years and above	0.00	0.44	0.11	0.10
Median Household Size [2010]	0.00	4.00	1.76	1.00
Median Annual Household Income (US$) [2010]	0.00	248,200.00	48,812.00	48,049.00
Household Vehicle Ownership [2010]				
Proportion of household with zero vehicles	0.00	0.08	0.02	0.01
Proportion of household with one vehicle	0.00	0.83	0.37	0.25
Proportion of household with two or more vehicles	0.00	0.95	0.61	0.28
Distance between Centroids of Census Tracts (km)	0.10	19.08	5.75	3.31

TABLE 2 Estimation Results of the Trip Generation Count Model

Variables	Weekday		Weekend Day	
	Estimate	(t-stat)	Estimate	(t-stat)
Determinants of the latent demand generation propensity (y_{qs}^*)				
Special Land-Use				
If all or part of the zone is occupied by The University of Texas	1.271	(2.34)	--	--
Transit Supply				
Average frequency of buses in an average bus stop in the zone (bus per hour)	-0.023	(-1.75)	--	--
Residential Density				
Population density (Natural Logarithm of person per km^2)	0.436	(10.17)	0.753	(36.97)
Activity Intensity				
Retail employment density (Natural Logarithm of retail jobs per km^2)	0.524	(11.74)	0.492	(14.58)
Population Characteristics				
Proportion of White population	-3.644	(-10.18)	-0.488	(-11.13)
Proportion of population 18-29 years old	2.124	(5.64)	0.298	(1.94)
Proportion of population 30-49 years old	0.225	(1.95)	--	--
Median annual household income (divided by $10,000)	0.056	(3.77)	-0.073	(-3.58)
Proportion of households with 2 or more automobiles	--	--	-2.669	(-25.67)
Demand tipping points (threshold component)				
α_1	--	--	--	--
α_{10}	-0.240	(-2.96)	-0.350	(-2.66)
α_{20}	-0.474	(-2.55)	--	--
α_{25}	--	--	-0.661	(-5.48)
Determinants of the thresholds (γ vector elements)				
Constant	2.279	(50.92)	0.182	(6.39)
Population Characteristics				
Proportion of male population	0.366	(14.08)	-0.345	(-6.75)
Households Characteristics				
Proportion of households with 2 or more automobiles	-1.487	(-18.70)	--	--
Special Land-Use				
Presence of parks in the zone	--	--	0.312	(2.94)
Correlation between weekday and weekend	0.394 (9.53)			
Spatial Autoregressive Parameter (δ)	0.561 (29.74)			
Composite Marginal log-likelihood	-250,389.50			

Note: '--' means that the corresponding coefficient was not statistically significantly different from zero at the 90% level of confidence.

TABLE 3 Estimation Results of the Trip Distribution Split Model

Variables	Weekday		Weekend Day	
	Estimate	(t-stat)	Estimate	(t-stat)
Constant	-5.051	(-13.60)	-2.098	(-14.74)
Special Land-Use				
If all or part of the zone is occupied by The University of Texas	0.711	(2.85)	--	--
If the area is in the central business district	--	--	-0.673	(-4.37)
If the area contains the airport	4.142	(14.81)	--	--
If the area is outside the area of interest	3.119	(7.85)	--	--
Residential Density				
Population density (Natural Logarithm of person per km^2)	0.110	(2.35)	--	--
Activity Intensity				
Retail employment density (Natural Logarithm of retail jobs per km^2)	0.179	(3.33)	0.164	(3.34)
Population Characteristics				
Proportion of White population	--	--	-1.653	(-4.90)
Median annual household income (divided by $10,000)	--	--	0.069	(3.72)
Proportion of households with 2 or more automobiles	-0.851	(-2.36)	-1.723	(-5.03)
Proportion of males	--	--	2.089	(3.94)
Trip Characteristics				
Log of average cost between zones	-0.422	(-2.97)	-0.378	(-4.99)
Other Characteristics				
Area (km²)	0.367	(3.68)	0.218	(1.94)

Note: '--' means that the corresponding coefficient was not statistically significantly different from zero at the 90% level of confidence.