A Model of Deadheading Trips and Pick-Up Locations for Ride-Hailing Service Vehicles

Gopindra S. Nair
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: gopindra@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

Irfan Batur
Arizona State University
School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005, USA
Tel: 480-727-3613; Email: ibatur@asu.edu

Ram M. Pendyala
Arizona State University
School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005, USA
Tel: 480-727-4587; Email: ram.pendyala@asu.edu

William H.K. Lam
The Hong Kong Polytechnic University
Department of Civil and Environmental Engineering
9/F, Block Z, 181 Chatham Road South, Hung Hom
Kowloon, Hong Kong
Tel: +852-2766-6045; Email: william.lam@polyu.edu.hk
ABSTRACT
The mode share of ride-hailing services has been growing steadily in recent years and this trend is expected to continue. Ride-hailing services generate two types of trips – passenger hauling trips and deadheading trips. Passenger hauling trips are the trips made while transporting passengers between places. Virtually all other trips made by a ride-hailing vehicle when there are no passengers in the vehicle are called deadheading trips or empty trips. Trips between the drop-off location of one passenger and the pick-up location of the next passenger could comprise a substantial share of total travel by ride-hailing vehicles, both in terms of number of trips and miles of travel. This paper aims to model these deadheading trips at the disaggregate level of individual trips. Passenger trip data published by the ride-hailing company Ride Austin is used to impute deadheading trips. The pick-up locations of passengers are then modeled using a nonlinear-in-parameter multinomial logit framework, essentially capturing the deadheading that takes place from the drop-off one passenger to the pick-up of the next passenger. The model is sensitive to socio-demographic characteristics, as well as employment opportunities and built environment characteristics of the study area. The model results shed light on the characteristics of deadheading trips at different locations and at different time periods in a day. The paper concludes with a discussion of how transportation planners and ride-hailing companies may utilize knowledge about deadheading to enact policies and pricing schemes that reduce deadheading.

Keywords: Ride-hailing, TNC, deadheading, empty trips, Ride Austin, location choice model
1 INTRODUCTION
The past decade has seen dramatic growth in the use of ride-hailing services around the world. Ride-hailing is a mobile platform-based service that facilitates the real time matching of drivers who are willing to provide a ride and potential passengers based on the spatial proximity between the potential “matches”. Companies that provide this service, such as Uber, Lyft, DiDi, and Ola, are commonly referred to as Ride-hailing companies or Transportation Network Companies (TNCs). The revenue generated by ride-hailing services in the USA was $15.6 billion in 2018 and this figure expected to grow to $26.3 billion by 2023 (Statista, 2019). The rapid growth in the use of ride-hailing is a testament to its convenience and practicality.

Ride-hailing is clearly serving a market that seeks convenience, real-time vehicle tracking capability, and door-to-door service. Ride-hailing provides this door-to-door mobility on demand at a cost lower than that of traditional taxi services and can facilitate shared (pooled) travel in a seamless fashion for individuals who would like to share travel costs and reduce their carbon footprint. The appeal of these considerations is obviously not lost on individuals, as clearly evidenced by the increasing popularity and use of ride-hailing services across cities in the world. In addition to the increasing adoption of these services among the public, ride-hailing is also viewed as a potential means to recover vast areas of urban land currently dedicated to parking (Henao and Marshall, 2018). This is a significant consideration, especially because parking facilities can take up to a third of overall land area in the downtown areas of some cities (Ben-Joseph, 2012). The rise of ride-hailing in cities all over the world has captured the attention of urban planners and policy makers in view of such benefits that could accrue.

Despite the potential benefits, there are concerns that arise in the context of the increasing adoption of ride-hailing services. One of the key concerns relates to the impact of ride-hailing on vehicle miles of travel (VMT) and peak-period congestion. Ride-hailing could contribute to increases in VMT and/or traffic congestion for a number of reasons. First, there is a general reluctance to use pooled versions of ride-hailing, at least in the United States (see Lavieri and Bhat, 2019). Also, if any increased VMT is disproportionately positioned during the peak periods, it can lead to increased traffic congestion. Second, ride-hailing may contribute to growth in VMT by serving as a substitute for public transportation travel (mode shift), as well as by inducing new trips that would not have been made in the absence of ride-hailing services (Hall et al., 2018; Henao and Marshall, 2018). However, perhaps the most pressing concerns is the increased VMT due to deadheading trips or empty trips. These are the trips that are made by ride-hailing vehicles when there are no passengers in the vehicle.\(^1\) While it behooves ride-hailing companies to match drivers and riders in ways that reduce deadheading time and distance (to minimize waiting time and maintain cost affordability), it is estimated that between 35% and 50% of the total distance traveled by ride-hailing vehicles is lost to deadheading (Cramer and Krueger, 2016; Henao and Marshall, 2018; Komanduri et al., 2018). This has led some to argue that ride-hailing is at least partially responsible for worsening traffic congestion in a number of cities (LeBlanc, 2018).

\(^1\) In this paper, the focus is on “fully” deadheading trips or truly empty trips; that is, trips with no passengers whatsoever. One can also examine deadheading portions of non-empty trips, in which there may be some passengers already in the ride-hailing vehicle, but the ride-hailing vehicle travels (with no benefit to existing passengers in terms of time or movement toward destinations) to pick-up an additional new passenger in a pooled/shared-ride mode of operation. However, given the very low use of ride-hailing in a pooled mode, at least in the U.S., and the inherent complexity in identifying deadheading portions of non-empty ride-hailing trips, the analysis (or inclusion) of this component of deadheading is left to a future research effort. Besides, as discussed later, the data used in this study precludes any shared/pooled ride-hailing, thus rendering the examination of deadheading portions of non-empty trips impractical.
Even though deadheading trips constitute a significant share of the distance traveled by ride-hailing vehicles, most of the research on ride-hailing has focused on passenger trips (i.e., trips involving transportation of passengers from pick-up locations to drop-off locations). Prior research has explored various aspects of ride-hailing passenger trips, including trip frequencies (Cooper et al., 2018), trip purposes (Dias et al., 2019), and the competitive or complementary effects of ride-hailing services relative to other travel modes (Henao and Marshall, 2018; Hall et al., 2018). The socio-demographic and behavioral characteristics of ride-hailing passengers have also been studied rather extensively (e.g., Clewlow and Mishra, 2017; Dias et al., 2017; Lavieri and Bhat, 2019). However, research on the nature of deadheading trips has been much more limited, primarily due to the lack of disaggregate trip-level data that would enable the detailed analysis of such trips.

Recently, a ride-hailing company operating in Austin, called Ride Austin, released anonymized data on all ride-hailing passenger trips that were served over a ten month period between 2016 and 2017 (Ride Austin, 2017). The availability of this data presents an opportunity to impute and analyze deadheading trips at a disaggregate trip-level. The objective of this paper is to gain a deep understanding of the nature of deadheading trips, i.e., trips made in search of a new passenger after dropping off any previous passenger(s). In this study, ride-hailing passenger trips data released by Ride Austin is used to impute the deadheading trips. Then, a model is developed for predicting the location where the next passenger will be picked up when the origin of the deadheading trip (or the drop-off location of the previous passenger) is known.

It is certainly likely that ride-hailing companies have in-house algorithms for predicting locations of passenger demand in real-time (so that “new” trip requests are served quickly and efficiently at prices that balance passenger demand and driver supply). While ride-hailing companies are likely to develop models to optimize business operations and revenue, the purpose of the model presented in this paper is to inform the spatio-temporal distribution of deadheading trips for urban and regional transportation planning. The model presented in this paper uses location specific attributes of the deadheading trip’s origin as well as attributes of the potential destinations to predict the subsequent passenger pick-up location. These location specific attributes include employment opportunities, built environment attributes, and socio-demographic characteristics. A nonlinear-in-parameters multinomial logit model of pick-up location is estimated and the model results are used to gain insights on the factors that affect the propensity of passenger pick-ups at different locations and the distance that a ride-hailing driver travels in order to find a new passenger.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the past studies on ride-hailing. Section 3 presents a description of the data used in this study, while Section 4 explains the modeling methodology. Section 5 presents model estimation results. Section 6 presents a discussion of the potential planning and forecasting applications of the models developed in this study. Finally, Section 7 offers concluding thoughts.

2 LITERATURE REVIEW
Despite the relative youth of ride-hailing services (a mode that is less than 10 years old), it has undoubtedly proven to be quite disruptive and transformed mobility in urban areas around the world. As a consequence, interest in the research community on the use and operations of this mode of transportation has been intense and the body of literature on ride-hailing services is large and growing rapidly. As such, it is not possible to include a comprehensive literature review on the topic of ride-hailing within the scope of this paper. The literature review presented in this
section is intended to summarize evidence to date and help position this paper relative to the body of evidence to date.

Deadheading trips may be viewed as a corollary of passenger trips. After a passenger is served, a deadheading trip is very likely to take place as the ride-hailing driver proceeds to the next pick-up location. As such, deadheading trips cannot be fully understood unless the ride-hailing passenger trips that generate deadheading trips are understood in the first place. Many studies in the literature document the characteristics of ride-hailing trips and ride-hailing service users through a variety of surveys and secondary data collection efforts. One of the most comprehensive surveys on ride-hailing usage in the United States was conducted by Clewlow and Mishra (2017). They conducted a survey of ride-hailing users across seven major cities – Boston, Chicago, Los Angeles, New York, San Francisco, Seattle, and Washington D.C. Other studies that involved surveys of ride-hailing usage include Henao and Marshall (2018), Lavieri and Bhat (2019), and Rayle et al. (2016). Rayle et al. (2016) surveyed ride-hailing users in San Francisco, while Lavieri and Bhat (2019) surveyed commuters who used ride-hailing in the Dallas-Fort Worth metropolitan area. Henao and Marshall (2018) signed up to drive for ride-hailing companies, Uber and Lyft, in the Denver area and surveyed the passengers that rode in their vehicles. By serving as drivers, Henao and Marshall (2018) were also able to estimate the distance traveled while deadheading.

Some studies have attempted to derive insights on ride-hailing behavior from larger scale travel surveys (as opposed to targeted ride-hailing user surveys). Examples of such studies include: Dias et al. (2017), who studied ride-hailing usage using the 2017 Puget Sound regional household travel survey; Alemi et al. (2018), who used the 2015 California millennials survey; and Young and Farber (2019), who studied ride-hailing usage in Toronto. Because ride-hailing companies have not released trip-level ride-hailing data, a few studies, such as those by Cooper et al. (2018) and Kooti et al. (2017), have attempted to collect this data using novel data collection techniques. Cooper et al. (2018) obtained data on the movement of deadheading ride-hailing vehicles in the city of San Francisco directly through the Application Programming Interfaces (API) of Uber and Lyft. They repeatedly queried the Uber and Lyft servers for the locations of vehicles that are available for hire around 200 synthetically generated clients spread around the city. Using this approach, they obtained the coordinate traces of deadheading Uber and Lyft vehicles in the city for a duration of 40 days. Kooti et al. (2017) extracted data on Uber’s ride-hailing trips from Yahoo’s email servers. They access the receipts sent by Uber to passengers as well as the reports sent by Uber to drivers and combine these reports with data on the owners of the email accounts to understand the socio-demographic profile of Uber passengers and drivers.

Ride Austin is the first ride-hailing company to release data on individual passenger trips. Ride Austin is a non-profit that began operating in the city of Austin, Texas on June 16, 2016 when other companies, notably Uber and Lyft, were forced to cease operations because they were unable to abide by the city’s regulations (Kelly, 2016). By the beginning of 2017, Ride Austin held around one-third of the ride-hailing market share in the city. Since the release of the trip-level data, a number of studies utilizing this data have been conducted. Komanduri et al. (2018) analyzed the data to uncover aggregate ride-hailing characteristics such as average trip lengths and trip frequencies by time of day. Dias et al. (2019) and Lavieri et al. (2018) combined the trip data with census and land use datasets to develop models of ride-hailing demand. Dias et al. (2019) imputed the purposes of ride-hailing trips for frequent ride-hailing users based on land use characteristics of trip destinations. Wenzel et al. (2019) used the dataset to evaluate the impact of ride-hailing usage on vehicle miles of travel (VMT) and energy consumption, while considering both passenger trips and deadheading trips.
Virtually all of the studies to date show that ride-hailing users tend to be younger, more educated, employed, higher income, and belong to households of smaller size. They also tend to exhibit lower levels of vehicle ownership, although the direction of causality remains a question worthy of research. That is, it is unclear if vehicle availability influence ride-hailing use, or if ride-hailing adoption leads to shedding of vehicles. More pertinent to this study are past research efforts aimed at understanding spatio-temporal patterns of ride-hailing trips. Time-of-day distributions of ride-hailing demand are found to vary across cities. Komanduri et al. (2018) found that the traditional time-of-day distribution of a peak in the morning and a peak in the afternoon does not hold for ride-hailing demand in Austin. However, there are other studies that have observed a more traditional temporal distribution pattern for ride-hailing trips (Cooper et al., 2018; Kooti et al., 2017). Lavieri and Bhat (2019) observed that ride-hailing demand peaks in the afternoon and is the lowest at night in the Dallas-Fort Worth area. However, their study sample is exclusively comprised of commuters and hence not representative of the overall demand for ride-hailing services. Studies by Lavieri et al. (2018) and Komanduri et al. (2018), both of which used the Ride Austin dataset, found ride-hailing demand to be higher on weekends than on weekdays. Rayle et al. (2016) also observed that almost one-half of all ride-hailing trips occurred on Friday and Saturday in San Francisco. As expected, these studies have found that ride-hailing trips to/from work were more likely to occur in the morning and evening periods while trips for recreational purposes were more likely to occur in the evening and night periods. For the Ride Austin dataset, Wenzel et al. (2019) found that a significant share of early morning weekday trips (5-7 AM) are bound for the airport. They also found that the share of trips destined to recreational areas increases in the evenings and on weekends.

In terms of the spatial distribution of ride-hailing demand, it appears that there is a concentration in denser urban areas. Clewlow and Mishra (2017) found that 29% of individuals in urban areas had used ride-hailing, while only 7% of individuals in suburban areas had used it. Both Dias et al. (2019) and Lavieri et al. (2018) found that ride-hailing demand is high in dense residential neighborhoods in urban areas. Rayle et al. (2016) and Cooper et al. (2018) also report similar findings for San Francisco. Clewlow and Mishra (2017) identified parking and avoiding driving under the influence to be the two main reasons why individuals choose ride-hailing over other modes. This may imply that areas with a high concentration of entertainment establishments and/or lower parking availability will see higher ride-hailing demand. Rayle et al. (2016) reported that 67% of the ride-hailing trips were made for social and leisure purposes (this includes trips to bars and restaurants). Also, Dias et al. (2019) and Rayle et al. (2016) reported that around 4-5% of ride-hailing trips are airport-bound trips, which is disproportionately large when compared to the total share of airport trips. Overall, it appears that areas with leisure establishments and airports are likely to see high levels of ride-hailing demand.

A few studies have attempted to address the issue of measuring and quantifying deadheading directly. Cramer and Krueger (2016) compared the deadheading durations of ride-hailing and traditional taxis in the cities of Boston, Los Angeles, New York, San Francisco and Seattle. They measured the capacity utilization as the proportion of time for which there is a passenger in the ride-hailing vehicle to the total duration of operation of the vehicle. They were able to obtain the city-wide percentages for capacity utilization directly from Uber. The same metric was computed for regular taxis using data from a wide range of sources. In virtually all cities, the capacity utilization for ride-hailing was higher than that for traditional taxis, with the difference being the smallest in New York. The capacity utilization for ride-hailing ranged from
43.6% in Seattle to 54.3% in San Francisco. They estimated the percentage of distance traveled while deadheading to be 35.8% in Los Angeles and 44.8% in Seattle.

Henao and Marshall (2018) tried to estimate the share of deadheading distance based on their own driving data when serving as drivers for Uber and Lyft. They report 40.8% of their total distance traveled (VMT) being lost to deadheading. On the other hand, Komanduri et al. (2018) used the Ride Austin dataset to impute deadheading trips. They assumed that a deadheading trip occurs whenever the time gap between consecutive passenger trips made by a driver is less than 30 minutes. The deadheading distances were calculated as the straight-line distances between the origins and destinations. If the time gap between passenger trips was more than 30 minutes, it was assumed that drivers would not deadhead for the entire time gap. Instead, they assumed that two deadheading trips with distances of two miles each and durations of five minutes each would have occurred within that interval. The first deadheading trip would occur immediately after the drop-off of one passenger and the next deadheading trip would occur immediately before the subsequent passenger pick-up. Based on these assumptions, they estimated the percentage of deadheading miles traveled by ride-hailing vehicles to be 37%. In a more recent study, Wenzel et al. (2019) also used a similar approach to identify deadheading trips. Unlike Komanduri et al. (2018), Wenzel et al. (2019) assumed that continuous deadheading occurs between consecutive passenger trips that are less than 60 minutes apart. They also used a correction factor for converting straight-line distances to network distances. They estimated the length of an average deadheading trip that occurs between consecutive passenger trips to be 55% of the average passenger trip length. Additionally, they imputed the location of residences of drivers based on the spatial median of their first pick-up and last drop-off locations from every shift. A new ride-hailing shift is assumed to have begun if a passenger is picked up eight hours after the previous passenger is dropped off. Using the imputed locations of driver residences, the deadheading distances at the beginning and ending of shifts were also computed. Using these assumptions, the overall percentage of deadheading distance was estimated to be 45%.

The approach adopted in this paper to identify deadheading trips is similar to that used by Komanduri et al. (2018) and Wenzel et al. (2019). In this paper, only the deadheading trips made for repositioning vehicles between drop-off and pick-up locations of consecutive passengers are considered. While other studies have largely estimated deadheading miles at an aggregate scale as a percentage of the total distance traveled (by ride-hailing vehicles), this paper makes an important contribution by modeling deadheading at the disaggregate level of individual trips. This provides a deeper understanding of the variations in deadheading patterns across time and space at a fine spatio-temporal scale, while also identifying the key factors that contribute to deadheading variations. To build a model of deadheading (i.e., pick-up location for next passenger) with a rich specification, the ride-hailing trip data needs to be fused with secondary data sources including socio-demographic data, network travel times and distances, built environment data, and employment data. Through the execution of a novel data integration process, it is possible to develop a model of deadheading (next passenger pick-up location) with a rich specification of explanatory attributes. The insights from such a model would prove invaluable in predicting locations with high potential for deadheading (hotspots) and devising countermeasures to alleviate the adverse effects of deadheading mileage. To capture the variety of purposes for which ride-hailing trips are made at different times in a day, variables that measure employment opportunities at the potential destinations and identifiers for special generators such as the Austin Bergstrom International Airport (ABIA) and the main campus of the University of Texas at Austin, are
included in the model. Separate models are then estimated for different time periods of a day, thus providing an understanding of temporal variation in deadheading across the region.

3 DATA DESCRIPTION
The primary dataset used in this study is the dataset on ride-hailing passenger trips released by Ride Austin (2017). Deadheading trips were inferred and imputed based on the passenger trips in this dataset. The ride-hailing data was supplemented with data from several other sources. The Smart Location Database (SLD) was used to obtain data on socio-demographic, built environment, and other characteristics of census block groups in the study area (US EPA, 2014). Some of the variables in the SLD were updated using more recent figures obtained from the 2016 American Community Survey (ACS) dataset. The network distances (skims) used in this study are based on the zoning and network information of the study area as used by the Capital Area Metropolitan Planning Organization (CAMPO) in their travel demand forecasting models. CAMPO is responsible for transportation planning in the city of Austin. The zoning and network information were acquired for the counties of Burnet, Bastrop, Caldwell, Hays, Travis and Williamson. The six counties comprise a total of 2102 traffic analysis zones (TAZs).

3.1 Ride Austin Passenger Trips Dataset
The Ride Austin dataset contains the anonymized records of every passenger trip made using the Ride Austin application between early-June 2016 and mid-April 2017. There are 1,494,125 trip records in this dataset. Each record in the dataset includes the coordinates of the origin and destination, the trip start and end times, and identifiers for the driver and passenger. The coordinates in the dataset are censored to three decimal places to protect passenger privacy (if the coordinates were only truncated to three decimal places and not rounded, this corresponds to a maximum radial error in distance measurement of approximately 147 meters). The service did not allow for the sharing of rides between strangers. In other words, drivers would receive new ride requests only when there were no passengers in the vehicle. This is in contrast to the ride-sharing services provided by some of the competing TNCs (e.g., Uber Pool and Lyft Shared) where new passengers may be picked up enroute to the destination of a passenger already in the vehicle. The lack of a ride-sharing option ensures that each passenger trip in the dataset is associated with only one passenger account.

The number of records in the initial months of the operation of Ride Austin are very few in number. To better capture ride-hailing usage patterns in steady state, only data for the period of October 2016 to mid-April 2017 (corresponding to 195 days of data) is used in this study. Since it is computationally demanding to model locations as a continuous measure in the context of travel demand forecasting, all locations were mapped to their corresponding TAZs. A map of the average number of daily ride-hailing passenger trips generated by each TAZ (on a per square mile basis to account for differential TAZ sizes) is shown in Figure 1. In terms of density of origin of passenger trips, the zones adjacent to the University of Texas and the Austin downtown area zones had the highest ride-hailing trip origins per square mile, consistent with students availing of ride-hailing services for their trips as well as patrons of bars and other entertainment places in the Austin downtown. In terms of the total number of trip origins, the zone containing the Austin Bergstrom International Airport (ABIA) generated the greatest number of ride-hailing passenger trips. This TAZ was also the destination for the greatest number of ride-hailing trips (this gets a little masked in Figure 1 because the zone containing the airport is a large one, leading to a density of origins that is less than that for the University and Austin downtown areas). In particular,
4.1% of all ride-hailing trips originated in the airport TAZ, and 6% of all ride-hailing trips ended there.

Figure 1  Average Number of Daily Ride-Hailing Passenger Trips Generated by Each TAZ (Per Square Mile)

The average number of ride-hailing passenger trips generated by hour of the day and day of the week are shown in Figure 2. The demand for ride-hailing trips in Austin does not follow the usual time-of-day distributions of travel demand with peaks in the morning and the afternoon, a finding also reported by Komanduri et al. (2018). In Austin, the highest frequency of ride-hailing trips occurs on Friday and Saturday late night periods (and correspondingly Saturday and Sunday early morning periods). Since the purpose and frequency of ride-hailing trips vary by time-of-day, four separate models are estimated for the time periods of 7AM to 10AM, 10AM to 4PM and 4PM to 7PM on weekdays, and 10PM to 1AM the next day on Friday and Saturday. In the remainder of the paper, these time periods are referenced as AM Peak, Mid-day, PM Peak and Weekend
Night respectively. The AM peak, mid-day and PM peak periods were selected for modeling because these are the time periods often considered in travel demand forecasting models. Additionally, the weekend night period is considered separately because of the exceptionally high rate of utilization of ride-hailing services in this period.

![Figure 2: Average Number of Ride-Hailing Passenger Trips by Time-of-Day and Day-of-Week](image)

3.2 Location Attribute Databases

Various socio-demographic, employment, built environment, and transportation network-related attributes of locations in Austin were obtained from the Smart Location Database (SLD), CAMPO travel model databases, and the American Community Survey (ACS) dataset. The SLD and the ACS datasets provide attribute values at the census block group level, while the CAMPO dataset provides attribute values at the TAZ level. Because the census block group boundaries did not match the TAZ boundaries exactly, GIS overlay techniques and weighted aggregation and

---

2 To be precise, a deadheading trip is assigned to a particular time period if its destination time-stamp is within the time period. For example, if a deadheading trip’s destination time-stamp is between 7AM and 10AM on a weekday, it is designated as an AM peak deadhead trip (that is, if a passenger is picked up between 7AM and 10AM, the previous deadheading trip leading up to the passenger pick-up is assigned to the AM peak).
allocation methods were employed to aggregate SLD and ACS data from the census block level to the TAZ level.

Descriptive statistics for TAZ attributes constructed from the SLD, ACS, and CAMPO datasets are presented in Table 1.

### Table 1 Descriptive Statistics for TAZ Attributes

<table>
<thead>
<tr>
<th></th>
<th>Year</th>
<th>Source</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 18 years</td>
<td>2016</td>
<td>ACS</td>
<td>228.64</td>
<td>335.46</td>
</tr>
<tr>
<td>18 ≤ Age &lt; 35 years</td>
<td>2016</td>
<td>ACS</td>
<td>258.05</td>
<td>391.29</td>
</tr>
<tr>
<td>35 ≤ Age &lt; 65 years</td>
<td>2016</td>
<td>ACS</td>
<td>366.13</td>
<td>480.04</td>
</tr>
<tr>
<td>65 years ≤ Age</td>
<td>2016</td>
<td>ACS</td>
<td>92.50</td>
<td>157.49</td>
</tr>
<tr>
<td>Caucasian</td>
<td>2016</td>
<td>ACS</td>
<td>746.85</td>
<td>924.32</td>
</tr>
<tr>
<td>African American</td>
<td>2016</td>
<td>ACS</td>
<td>67.99</td>
<td>147.72</td>
</tr>
<tr>
<td>Other race</td>
<td>2016</td>
<td>ACS</td>
<td>130.49</td>
<td>224.31</td>
</tr>
<tr>
<td>Employed</td>
<td>2016</td>
<td>ACS</td>
<td>481.85</td>
<td>619.38</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2016</td>
<td>ACS</td>
<td>945.33</td>
<td>1184.76</td>
</tr>
<tr>
<td><strong>No. of Households (HH)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 vehicle</td>
<td>2010</td>
<td>SLD</td>
<td>17.51</td>
<td>37.77</td>
</tr>
<tr>
<td>1 vehicle</td>
<td>2010</td>
<td>SLD</td>
<td>122.18</td>
<td>173.98</td>
</tr>
<tr>
<td>2+ vehicle</td>
<td>2010</td>
<td>SLD</td>
<td>205.60</td>
<td>264.44</td>
</tr>
<tr>
<td>Income &lt; $50,000</td>
<td>2016</td>
<td>ACS</td>
<td>130.46</td>
<td>181.60</td>
</tr>
<tr>
<td>$50,000 ≤ Income &lt; $100,000</td>
<td>2016</td>
<td>ACS</td>
<td>108.43</td>
<td>146.94</td>
</tr>
<tr>
<td>$100,000 ≤ Income &lt; $150,000</td>
<td>2016</td>
<td>ACS</td>
<td>55.49</td>
<td>88.47</td>
</tr>
<tr>
<td>$150,000 ≤ Income</td>
<td>2016</td>
<td>ACS</td>
<td>49.27</td>
<td>96.50</td>
</tr>
<tr>
<td>HH Size = 1</td>
<td>2016</td>
<td>ACS</td>
<td>95.84</td>
<td>137.73</td>
</tr>
<tr>
<td>HH Size = 2</td>
<td>2016</td>
<td>ACS</td>
<td>115.28</td>
<td>150.20</td>
</tr>
<tr>
<td>HH Size = 3</td>
<td>2016</td>
<td>ACS</td>
<td>53.36</td>
<td>76.37</td>
</tr>
<tr>
<td>HH Size = 4+</td>
<td>2016</td>
<td>ACS</td>
<td>79.15</td>
<td>118.54</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2016</td>
<td>ACS</td>
<td>371.35</td>
<td>474.32</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>2015</td>
<td>CAMPO</td>
<td>93.76</td>
<td>351.73</td>
</tr>
<tr>
<td>Retail</td>
<td>2015</td>
<td>CAMPO</td>
<td>100.55</td>
<td>218.89</td>
</tr>
<tr>
<td>Service</td>
<td>2015</td>
<td>CAMPO</td>
<td>219.77</td>
<td>503.08</td>
</tr>
<tr>
<td>Education (K-12)</td>
<td>2015</td>
<td>CAMPO</td>
<td>23.11</td>
<td>56.83</td>
</tr>
<tr>
<td>Education (Higher)</td>
<td>2015</td>
<td>CAMPO</td>
<td>19.04</td>
<td>271.25</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2015</td>
<td>CAMPO</td>
<td>456.24</td>
<td>839.37</td>
</tr>
<tr>
<td><strong>Area (mi^2)</strong></td>
<td>2015</td>
<td>CAMPO</td>
<td>2.52</td>
<td>4.79</td>
</tr>
<tr>
<td><strong>Hourly frequency of transit per square mile in PM peak period</strong></td>
<td>2012</td>
<td>SLD</td>
<td>283.09</td>
<td>771.40</td>
</tr>
</tbody>
</table>

There is some inconsistency in the year corresponding to different datasets. This is especially so for the SLD data on the number of vehicles (from 2010) and the hourly frequency of transit per square mile in the PM peak (from 2012). However, it is unlikely that there were
substantial changes in these attributes at the zonal level within a span of six years or so, and thus
the use of these attribute values as a means to explain the 2016-2017 ride-hailing travel patterns
was considered acceptable. As shown in Table 1, the number of employment opportunities in a
TAZ is disaggregated into four categories – basic, retail, service, and education, based on the
Standard Industrial Classification (SIC) code of employment industry (US DOL, 2019). This
disaggregation is used to recognize the varying nature of ride-hailing trip demands based on
activity purpose and time-of-day (see, for example, Lavieri and Bhat, 2019).

4 METHODOLOGY
The Ride Austin dataset provides information on passenger trips, including the time stamp and
location of the beginning and end of every passenger trip. Deadheading trips occur between the
drop-off and pick-up of consecutive passenger trips. Unless the drop-off location of one passenger
coincides with the pick-up location for the next passenger, there will be some non-zero distance
associated with deadheading. The driver cannot pick up a new passenger while transporting
another passenger because a ride-sharing option was not available in the period during which this
data was collected.

This section offers a detailed description of the procedure used to impute deadheading trips
(Section 4.1), the econometric framework for modeling destination of deadheading trips (Section
4.2), the procedures used to construct the choice set of possible destinations for model estimation
(Section 4.3), and the considerations involved in designing a valid specification for which all
coefficients can be identified (Section 4.4).

4.1 Imputing Deadheading Trips
Using the unique driver identifier (ID) and the start and end times of passenger trips, it is possible
to assemble the roster of passenger trips served by each driver over the 195-day period covered by
the data. Between the drop-off of one passenger and the pick-up of the next passenger, the
following may happen:

1. The driver begins searching for a new passenger immediately after dropping off the
previous passenger. This process may involve the relocation of the ride-hailing vehicle to
a location where the driver feels that passenger pick-ups are more likely, either based on
past experience or because of surge pricing (surge pricing is a pricing scheme enacted by
TNCs to manage ride-hailing demand and driver supply). When a passenger pick-up
request is received, the driver proceeds to the pick-up location of the passenger. In this
scenario, the travel between the drop-off location of one passenger and the pick-up location
of the next passenger constitutes a single deadheading trip. This represents the first
category of deadheading trips.

2. The driver may take a break from ride-hailing to perform other activities (such as eating
out) or going home to rest. For example, consider a driver who drops off a passenger at
location A, then travels to eat meal at location B, then travels home to location C, and then
starts another round of ride-hailing by picking up a passenger at location D. In this case,
what constitutes deadheading is somewhat ambiguous. For sure, the trip from location B
to C (returning home to rest) is not a deadheading trip (in the context of the definition used
in this paper). One may consider the trip from A to B to be a deadhead, but it is difficult to
say for sure because location B is chosen by the driver voluntarily. The only trip in this
chain that may be considered a deadhead trip with certainty is the trip from C to D, as the
A driver starts a new chain of ride-hailing trips. The deadheading associated with these kinds of complex patterns (where the driver participates in personal activities and ceases service between ride-hailing trips) correspond to a second category of deadheading trips.

The Ride Austin data does not provide a clear way of distinguishing between these two types of deadheading categories. This is because the data only has the pick-up and drop-off time-stamps and locations of the ride-hailing trips, and no information on the ride-hailing vehicle location between passenger trips. Thus, in the second category above, the data would only show the time-stamp of the drop-off at location A and the time-stamp of the pick-up at location D. Assuming that all the travel between A to D constitutes a deadhead would be incorrect. It is necessary to identify only “true” ride-hailing deadhead trips, but this is not possible based on the information available in the Ride Austin data (for the second category noted above). Therefore, a rule-based approach was adopted to infer deadheading trips. The rule is based on the duration of the deadhead time window (that is, the time window between consecutive passenger trips of the ride-hailing vehicle).

The time spent by drivers in searching for a new passenger can vary widely based on the ride-hailing market. Based on anecdotal evidence (Uber Drivers Forum, 2017), the passenger search duration is in the order of minutes in highly populated cities such as Dallas-Fort Worth and Las Vegas while it can be more than an hour in small towns. Figure 3 shows the cumulative probability distribution of the time duration between the start of a passenger trip and the end of the previous passenger trip for different periods of the day. This duration is generally longer for the weekday AM peak and shortest for the weekend night period. This is consistent with the high level of ride-hailing demand during the weekend night period, which naturally leads to more frequent passenger pick-ups. The elbowing of the cumulative distribution function near the one hour mark seems to indicate that in most cases the duration of time spent by drivers in searching for new passengers would be less than an hour (the same assessment is also made by Wenzel et al., 2019). Over a full day, the time gap between trips is less than one hour in 72% of the cases. If the time gap is more than an hour (between trips), it is more likely that the driver stopped serving passengers and made trips to perform other activities. Deadhead travel segments corresponding to time gaps in excess of one hour are therefore ignored in this study (due to the difficulty in parsing out the “true” deadheading portion). Once the deadheading trips are identified based on the criterion above, the origin and destination of a deadheading trip will respectively be the destination of one passenger trip and the origin of the subsequent passenger trip.

Based on the one-hour rule noted above, a total of 46505, 109287, 88882, and 191849 deadheading trips were identified in the AM Peak, Mid-day, PM Peak and Weekend Night periods, respectively. The average shortest path distance between the origins and destinations of deadheading trips was found to be 2.54 miles, while the average shortest path distance of passenger trips in the dataset is 4.53 miles. An estimate of the deadheading distance as a percentage of the total distance traveled by ride-hailing vehicles is about 36%. This value may be considered a lower bound for two reasons. First, while the passenger trips are likely to be direct from origin to destination, the deadhead trips may not be as direct since ride-hailing drivers may cruise around while waiting for the next demand request. Second, this study designates trips to be “true” deadhead only if their time durations were lower than one hour. In reality, some of the deadheads of one hour or more may also be true deadhead trips.
Figure 3 Distribution of the Time Gap Between Ride-hailing Trips by Time-of-Day

4.2 Econometric Framework

This study employs a random utility maximization framework to predict the destination TAZ of a deadheading trip, given the origin TAZ of the deadheading trip. Each TAZ has an intrinsic propensity (or utility) associated with being the destination of a deadheading trip. This may also be considered as a measure of the propensity of obtaining the next passenger pick-up in a TAZ location. TAZs generally cover a wide area and may comprise several elemental locations, each associated with a utility of being the destination. The utility of a TAZ, which is a measure of the propensity of obtaining the next pick-up from within the TAZ, may be viewed as an aggregate of the utilities of all elemental locations within the TAZ.

Let $D_j$ be the number of deadheading destination points within TAZ $j$, or alternatively, the number of elemental passenger pick-up points within TAZ $j$. If $D_j$ is relatively large (as is usually the case with passenger pick-up points within a zone), and assuming that the propensity distribution of passenger pick-ups at different points within the zone are about the same and are independent of one another, the TAZ-level utility function of a destination TAZ $j$ for a deadheading trip $n$ originating in TAZ $i$ may be written as (see Bhat et al., 1998):

$$U_{nij} = \beta' x_{ij} + \eta \log(D_j) + \xi_{nij},$$

where $x_{ij}$ is an independent variable vector that includes variables related to the impedance between zones $i$ and $j$, the non-size characteristics of TAZ $j$, and interactions between impedance measures, characteristics of TAZ $i$, and characteristics of TAZ $j$. $\xi_{nij}$ is a random term assumed to be distributed IID Gumbel across deadheading trips, zonal deadheading origins, and zonal deadheading destinations. $\beta$ is a parameter vector and $\eta$ is a scalar parameter to be estimated. In Equation (1), the log transformation for the size term is essential because it guarantees that if two destination zones (with identical non-size zone attributes) are merged into one, the probability of choosing the combined zone is exactly the sum of the probabilities of choosing the original two zones when $\eta=1$. $D_j$ is not easily quantifiable. However, $D_j$ may be represented by a set of proxy observable size variables such as employment in zone $j$, population of zone $j$, and land area of zone...
Let \( z_j \) represent a vector of proxy size variables for zone \( j \) and let \( \delta \) be a corresponding vector reflecting the contribution of the proxy size variables to the actual zone size \( D_j \). Then, Equation (1) may be rewritten as:

\[
U_{nij} = \beta' x_{ij} + \eta \log(\delta' z_j) + \xi_{nij} = V_{ij} + \xi_{nij}.
\]  

Equation (2)

The values of the elements of the \( \delta \) vector should be greater than or equal to zero, because increasing any proxy size measure for a zone must increase the chances that the zone will be a destination zone for a deadhead trip. Also, for identification purposes, one of the elements of the \( \delta \) vector should be normalized to a specific value. Then, the coefficients on the other size variables provide the importance of these other size variables in attracting deadhead trips (i.e., passenger trip originations from the zone) relative to the normalized attribute. Finally, \( V_{ij} \) represents the systemic component of the utility function in Equation (2). The magnitude of the parameter \( \eta \) in Equation (2) characterizes the presence of common unobserved zonal attributes affecting the attractiveness of all elemental alternatives in a zone as deadhead destinations. For example, consider a uniformly elevated attractiveness of elemental destinations within a zone. This results in cannibalization of sorts in terms of passenger pick-ups (because prospective passengers may go to any elemental location within a zone, all of which are highly favorable; and going to one elemental location implies a drop in pick-up at another elemental location). This parameter is expected to lie between 0 and 1, with a value of 1 denoting no cannibalization effects and a value of zero essentially indicating such a high cannibalization level that the number of elemental locations within a zone (i.e., the zone size) does not matter. Another way to see this is to write the probability expression for a deadheading trip \( n \) with origin TAZ \( i \) and destination TAZ \( j \) as:

\[
P_{n(i)}(j) = \frac{e^{V_{ij}}}{\sum_{k \in \Omega} e^{V_{ik}}} = \frac{(D_j)^\eta e^{\beta' x_{ij}}}{\sum_{k \in \Omega} (D_k)^\eta e^{\beta' x_{ik}}} \frac{(\delta' z_j)^\eta e^{\beta' x_{ij}}}{\sum_{k \in \Omega} (\delta' z_k)^\eta e^{\beta' x_{ik}}},
\]  

Equation (3)

where \( \Omega \), is the set of all possible destination TAZs. As \( \eta \) decreases from the value of 1, an increase in the size of a zone \( j \), \( D_j \), has less and less of an effect on the probability of choice of alternative \( j \). The model in Equation (3) is a nonlinear-in-parameters multinomial logit model (NPMNL) because of the presence of the size effects represented in the \((\delta' z_j)^\eta\) component. The estimation of the model is accomplished by using the maximum likelihood method in the GAUSS matrix programming language.

### 4.3 Choice Set Formation

In Equation (3), one could assume that the choice set \( \Omega \) consists of all TAZs in the study area (2102 TAZs). However, in reality, only a subset of all possible TAZs will represent the potential destination choice set for any deadhead trip. This is because the behavioral and operational process determining deadhead trips is very likely to limit deadhead distance, as customers seek low wait times, and drivers and ride-hailing companies seek to minimize non-revenue time (miles). Ignoring this aspect of choice set determination will lead to inconsistent parameter estimates (Bhat, 2015). The issue then becomes one of determining an appropriate behavioral rule for the destination choice set generation process. In the Ride Austin dataset, 99.3% of all deadheading trips are less than 15 miles in length. Therefore, 15 miles is used as the threshold distance and the very small number of deadheading trips that had a distance greater than 15 miles were removed. Specifically,
only those TAZs that are within a distance of 15 miles (shortest network distance) from the origin TAZ are candidate destination zones for the deadhead trip. Across all deadheading trips in the sample, applying this choice set rule resulted in an average of 865.41 zones in the choice set, with a minimum of 12 and a maximum of 949 zones.

If the choice set has a large number of alternatives, there will be several TAZs that have a near zero probability of being selected. This creates issues with convergence when using a maximum likelihood framework for estimating the values of $\beta$, $\delta$, and $\eta$. To avoid this issue, only a sample of all possible TAZs are included in the choice set. For a trip $n$ between origin $i$ and destination $j$, the choice set $\Omega_n$ is generated by including the TAZ $j$ and 29 other TAZs chosen at random without replacement from within 15 miles (shortest network distance) of the origin TAZ. Since $\Omega_n$ is generated based on the destination TAZ, the conditional probability of selecting a TAZ from $\Omega_n$ has to account for this bias in the sampling of alternatives. The conditional probability of selecting a TAZ $j$ from $\Omega_n$ will be

$$P_{n(i)}(j \mid \Omega_n) = \frac{e^{y_{ij} + \log(P(\Omega_n \mid i, j))}}{\sum_{k \in \Omega_n} e^{y_{ik} + \log(P(\Omega_n \mid i, k))}},$$

(4)

where $P(\Omega_n \mid i, j)$ is the probability that $\Omega_n$ is the sampled choice set given an origin $i$ and destination $j$. Because the TAZs in $\Omega_n$ other than the chosen TAZ $j$ are chosen randomly from within a threshold distance of $i$, it follows that:

$$P(\Omega_n \mid i, j) = P(\Omega_n \mid i, k) \forall j, k \in \Omega_n$$

(5)

Substituting Equation (5) in Equation (4) results in the following:

$$P_{n(i)}(j \mid \Omega_n) = \frac{e^{y_{ij}}}{\sum_{k \in \Omega_n} e^{y_{ik}}}.$$

(6)

This means that when $\Omega_n$ is generated in the manner described above, the effect of bias in the sampling of alternatives can effectively be ignored. McFadden (1978) has shown that maximizing the conditional likelihood function based on the conditional probability given in Equation (6) will yield consistent estimates. The reader is referred to Ben-Akiva et al. (1984) for the derivation of Equation (4) and for details on other methods of sampling location alternatives.

### 4.4 Specification

Model specification was guided by the variables that were found to be significant for predicting ride-hailing demand in earlier studies and the attributes that were available through the ACS dataset and SLD. In addition, the model specification included a measure of impedance between the origin and destination and a measure of the spatial accessibility of the destination.

#### 4.4.1 Impedance measure

Using the road network file obtained from CAMPO, the shortest network distance between each pair of TAZs was computed. Since the shortest distance from a TAZ $i$ to itself would be zero, an intrazonal impedance value was calculated as:
\[ d_{ii} = \frac{128}{45\pi} \sqrt{\frac{A_i}{\pi}}, \]  

where \( A_i \) is the area of the TAZ. \( d_{ii} \) would be the expected distance between two randomly selected points in a circular disc with the same area as that of the TAZ. The impedance measure between TAZs \( i \) and \( j \) may be any monotonic transformation of the distance between the TAZs, \( d_{ij} \). The final specification uses the square root of the network distance as the impedance measure. This measure was selected after systematically comparing the performance of different impedance measures formed from the monotonic transformations of \( d_{ij} \) such as \( \ln(d_{ij}) \), \( d_{ij}^2 \), and \( d_{ij} \) itself. The impedance measures were evaluated based on the log-likelihood at convergence of the specification provided in Equation (8), estimated on a sample dataset with 5000 observations from each of the four time periods:

\[ U_{ij} = \beta_c(d_{ij}) + \beta_u u_i c(d_{ij}) + \beta_r r_i c(d_{ij}) + \beta_n n_{ij} + \beta_s s_{ij}, \]  

where \( c(d_{ij}) \) is the impedance between TAZs \( i \) and \( j \), \( u_i \) is an indicator of the origin \( i \) being urban and \( r_i \) is an indicator of the origin \( i \) being rural (TAZs can be urban, suburban, or rural), \( n_{ij} \) is an indicator to denote \( i \) and \( j \) being neighbors (takes a value of 1 if the boundary of \( i \) touches the boundary of \( j \)), and \( s_{ij} \) indicates the equality of \( i \) and \( j \) (takes a value of 1 when intrazonal trips are considered). The impedance measure is introduced into the size independent component of the specification (the \( x_{ij} \) vector in Equation (2)).

**4.4.2 Accessibility measure**

The retail and service accessibility of the destination zone is included in the model specification. This measure indicates the degree to which a zone is close to locations that have retail and service opportunities. After accounting for the number of retail and service locations within the subject zone, the coefficient of the accessibility measure captures the effect of the retail and service opportunities in nearby zones. A negative coefficient on this measure would imply that the retail and service opportunities in nearby zones have a competing effect on the retail and service locations within the zone. On the other hand, a positive coefficient suggests that the agglomeration of retail and service opportunities enhances the attractiveness of individual locations. The effect of agglomeration of shopping locations was explored previously by Bhat et al. (1998), who find that the effect of having other shopping locations nearby is primarily competitive. The retail and service accessibility of a zone \( i \) is computed using the Hansen-type accessibility measure:

\[ M_i = \frac{1}{N} \sum_{j=1}^{N} \frac{R_j + S_j}{d_{ij}^\alpha}, \]  

where \( N \) is the total number of TAZs, \( R_j \) and \( S_j \) are respectively the number of retail and service employment opportunities in TAZ \( j \), and \( \alpha \) is a constant parameter. The value of \( \alpha \) is set at 1.4; this value was obtained by systematically comparing several specifications with the accessibility measures calculated using different values of \( \alpha \) ranging from 0.6 to 3.
4.4.3 Conditions for identification and validity of the likelihood function

As mentioned in Section 4.2, one of the coefficients of the variables used in size term ($\delta$) must be normalized to a constant (a value of one is adopted here) to allow for the identification of all coefficients. The number of retail employment opportunities in a TAZ was used for normalization.

Another issue to be considered when building the specification is that the size term must always be strictly positive. It is possible that there are TAZs for which $z_j$ is a zero vector (it cannot be negative because of the nature of variables in $z_j$). This would make the utility of those TAZs undefined. To avoid this issue, area of the TAZ is retained as one of the variables in the size term even when it is insignificant. The presence of this variable ensures that the size term is positive.

5 EMPIRICAL RESULTS

Deadheading trips were identified for the time periods of AM peak, mid-day, PM peak, and weekend night using the approach described in Section 4.1. Two-thirds of the trips were used for estimation, while the other one-third were set aside for validation. The destination TAZs for each time period were modeled separately. Section 5.1 presents the estimation results for the NPMNL model and inferences are drawn from the estimated coefficients. Section 5.2 presents a comparison of the goodness-of-fit between the NPMNL model and a simple Multinomial Logit (MNL) model.

5.1 NPMNL Model Estimation

For the estimation of the NPMNL model, all of the variables provided in Table 1 and the variables discussed in Section 4.4 were considered. Additionally, indicator variables denoting the area type of TAZs (urban, suburban, or rural), whether an urban TAZ belongs to the Central Business District (CBD), and TAZs corresponding to the airport (ABIA) and the University of Texas campus were included in the specification. The final specification was chosen based on statistical significance and logical intuitiveness of the estimation results. Table 2 shows the coefficients estimated using the NPMNL model for different time periods in a day.

5.1.1 Influence of impedance

The coefficients for impedance have the expected negative signs indicating that passenger pick-ups are more likely to be made closer to the origin of the deadheading trip (i.e., drop-off location of prior passenger). There is also variation in the effect of impedance based on the area type of the origin of the deadheading trip. Deadheading trips that originate in urban areas are likely to be shorter than those that originate in suburban or rural areas. This is likely due to higher demand for ride-hailing in denser urban areas, leading to shorter distances between drop-off and pick-up locations.

The effect of impedance on deadheading trips originating within CBD zones in comparison to the effect on trips originating in urban areas outside the CBD is not consistent across time periods. Deadheading trips originating in the CBD are shorter in the PM peak period, but longer in the mid-day and weekend night period. This may be due to a larger number of work-based ride-hailing trips occurring in the PM Peak period, consistent with what Lavieri and Bhat (2019) found.
Table 2  Estimation Results for NPMNL Model of Deadheading Trip Destination (Next Passenger Pick-up Location)

<table>
<thead>
<tr>
<th>Variable</th>
<th>AM Peak</th>
<th>Mid-day</th>
<th>PM Peak</th>
<th>Weekend Night</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>t-stat</td>
<td>Coef</td>
<td>t-stat</td>
</tr>
<tr>
<td>TAZ Size Independent Attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impedance (mi&lt;sup&gt;½&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance&lt;sup&gt;½&lt;/sup&gt;</td>
<td>-2.0708</td>
<td>-61.20</td>
<td>-1.7573</td>
<td>-82.88</td>
</tr>
<tr>
<td>Distance&lt;sup&gt;½&lt;/sup&gt; x Origin is rural&lt;sup&gt;a&lt;/sup&gt;</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Distance&lt;sup&gt;½&lt;/sup&gt; x Origin is urban&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.2942</td>
<td>-8.13</td>
<td>-0.4334</td>
<td>-18.87</td>
</tr>
<tr>
<td>Distance&lt;sup&gt;½&lt;/sup&gt; x Origin is in CBD</td>
<td>--</td>
<td>--</td>
<td>0.1259</td>
<td>7.26</td>
</tr>
<tr>
<td>Destination same as origin</td>
<td>-0.5011</td>
<td>-9.79</td>
<td>0.2265</td>
<td>7.70</td>
</tr>
<tr>
<td>Destination is a neighbor of origin</td>
<td>-0.1632</td>
<td>-6.27</td>
<td>0.2133</td>
<td>13.35</td>
</tr>
<tr>
<td>Attributes of destination TAZ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-1.0599</td>
<td>-4.35</td>
<td>-0.7384</td>
<td>-5.95</td>
</tr>
<tr>
<td>Urban&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.7446</td>
<td>20.36</td>
<td>0.4618</td>
<td>18.96</td>
</tr>
<tr>
<td>Central business district (CBD)</td>
<td>-1.1016</td>
<td>-28.95</td>
<td>-0.5859</td>
<td>-25.15</td>
</tr>
<tr>
<td>Presence of ABIA</td>
<td>1.4020</td>
<td>12.41</td>
<td>3.6209</td>
<td>71.67</td>
</tr>
<tr>
<td>Presence of UT main campus</td>
<td>-0.8824</td>
<td>-15.74</td>
<td>0.3393</td>
<td>12.30</td>
</tr>
<tr>
<td>Retail and Service Accessibility (/10&lt;sup&gt;6&lt;/sup&gt;)</td>
<td>5.1236</td>
<td>18.52</td>
<td>4.5318</td>
<td>27.30</td>
</tr>
<tr>
<td>Transit frequency at PM peak (mi&lt;sup&gt;-2&lt;/sup&gt;h&lt;sup&gt;-1&lt;/sup&gt;) (/1000)</td>
<td>--</td>
<td>--</td>
<td>0.1119</td>
<td>3.54</td>
</tr>
<tr>
<td>TAZ Size Attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of employment opportunities (/1000)</td>
<td>1.0000</td>
<td>(fixed)</td>
<td>1.0000</td>
<td>(fixed)</td>
</tr>
<tr>
<td>Retail employment</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Basic employment</td>
<td>0.4081</td>
<td>8.24</td>
<td>0.2254</td>
<td>23.19</td>
</tr>
<tr>
<td>Service employment</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Education employment</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>No. of Households (HH) by Type (/1000)</td>
<td>11.5970</td>
<td>8.96</td>
<td>3.2838</td>
<td>20.08</td>
</tr>
<tr>
<td>Annual HH Income ≥ $150,000</td>
<td>--</td>
<td>--</td>
<td>0.1298</td>
<td>7.90</td>
</tr>
<tr>
<td>HH without vehicles</td>
<td>9.4261</td>
<td>8.93</td>
<td>1.2848</td>
<td>13.10</td>
</tr>
<tr>
<td>Single person HH</td>
<td>3.6952</td>
<td>9.80</td>
<td>0.8146</td>
<td>24.33</td>
</tr>
<tr>
<td>No. of Persons by Type (/1000)</td>
<td>0.0010</td>
<td>0.02</td>
<td>0.0001</td>
<td>0.01</td>
</tr>
<tr>
<td>Age between 18 and 35 years</td>
<td>0.7743</td>
<td>25.08</td>
<td>0.7519</td>
<td>38.77</td>
</tr>
</tbody>
</table>

<sup>a</sup> Base area type is Suburban
<sup>b</sup> t-stats for coefficient being different from 1

-- Not statistically significantly different from zero at the 95% level of confidence and removed from the specification
for the Dallas-Fort Worth region. Since the CBD has a relatively high employment density, the higher demand for ride-hailing in the CBD region in the PM Peak period will result in lower deadheading distances. The effect of impedance is further nuanced by the intrazonal and neighboring zone indicators. These variables help capture the possible non-linearities in the effect of impedance on the utility of a TAZ. These indicators depict differential effects by time-of-day. It appears that deadheading trips in the peak periods (which are more likely to be associated with work-based passenger trips) are less likely to be within the same zone or in neighboring zones. On the other hand, mid-day and weekend night deadheading trips are more likely to have origins and destinations within the same vicinity of one another. Further exploration of ride-hailing passenger trip patterns and characteristics would provide additional insights on the reasons for these findings.

5.1.2 Influence of TAZ attributes independent of TAZ size
Model estimation results show that passenger pick-ups are most likely to occur in urban areas that are not part of the CBD, and least likely to occur in CBD and rural areas. Urban areas (outside CBD) and suburban areas are the most likely markets for ride-hailing passengers – these are highly populated areas, have lower levels of transit service, and involve trip distances that are generally too long to walk. Higher ride-hailing demand from urban areas because of higher population and employment concentrations was also reported by Clewlow and Mishra (2017), and Lavieri et al. (2018). Although the effective coefficient for the destination being within the CBD (this would be the sum of coefficients of the destination being urban and the destination belonging to the CBD) is lower than that of a TAZ in the suburban area, it cannot be concluded that suburban areas attract more trips than the CBD during these periods. This is because there is a large spatial separation between CBDs and suburban areas, and the penalty for considering farther TAZs is more for deadheading trips originating near the CBD (where the area type is mostly urban) than for deadheading trips originating near suburban areas.

There is a higher propensity for passenger pick-up at the University of Texas campus during the PM peak and mid-day periods than in the AM peak and weekend night periods, presumably due to a higher level of activity and presence of students during those periods. Also, university students, younger individuals, and educated adults have been shown to adopt ride-hailing services at a higher rate than other socio-economic groups (Rayle et al., 2016). Another area that generates a large share of ride-hailing demand is the airport (ABIA). Previous studies have already established that ride-hailing is a popular mode for airport access/egress trips (Lavieiri et al., 2018; Rayle et al., 2016).

Locations with higher retail and service accessibility are associated with a higher number of passenger pick-ups, as expected. The clustering of retail and service opportunities leads to a larger ride-hailing market demand in these locations. Another reason for this finding may be that higher retail and service accessibility renders it easier for ride-hailing drivers to pick-up passengers from nearby TAZs. Ride-hailing drivers may prefer to cruise in search of passengers in and around TAZs that have high accessibility.

The hourly frequency of transit per square mile of the TAZ (during PM peak) is used as a proxy for the transit connectivity of TAZs. Except for the AM peak period, all other periods depict a higher likelihood of passenger pick-ups in TAZs that are better served by transit. To some degree, this may be indicative of a competitive substitution effect between ride-hailing and transit, as has been reported in previous studies (Dias et al., 2017; Schaller, 2018). TAZs with better transit service are also likely to be locations with more attractions, employment and population concentrations, and destination opportunities. Therefore these locations are likely to see higher
levels of ride-hailing demand, thus increasing the probability that they will be destinations of deadheading trips.

5.1.3 Influence of TAZ attributes correlated with TAZ size

As mentioned in Section 4.4.3, the number of retail employment opportunities in a TAZ was set as the base against which the effects of other size related attributes are compared. As expected, passenger pick-ups are more likely to occur in TAZs with a higher number of retail employment and service employment. Shopping centers and eating and drinking establishments account for retail and service employment numbers to a substantial degree, suggesting that individuals summon ride-hailing services in these locations. Previous research has indicated that individuals use ride-hailing services to avoid driving under the influence and the hassle of finding and paying for parking (Clewlow and Mishra, 2017; Lahkar, 2018). Indeed, the coefficient for retail and service accessibility is much higher in the weekend night period, suggesting that popular drinking locations in Austin lie within close proximity to each other and depict a high propensity for passenger pick-ups. Passenger pick-ups are also more likely in high income neighborhoods, consistent with the notion that ride-hailing services are used more by higher income individuals.

The number of employment opportunities in education was found to be insignificant for predicting ride-hailing demand. However, this insignificance manifested only after including the indicator variable for the presence of the University of Texas at Austin main campus. This means that once the effect of the University of Texas at Austin main campus is captured, other locations with education employment opportunities do not contribute significantly to ride-hailing demand. In all of the time periods, ride-hailing demand (passenger pick-ups) is higher in locations with a greater presence of high-income households, and also in areas with a greater number of individuals aged between 18 and 35 years. This is consistent with previous studies that have repeatedly identified the demographic group of wealthy younger adults to be the most frequent users of ride-hailing services (Alemi et al., 2018; Clewlow and Mishra, 2017; Rayle et al., 2016). Interestingly the lack of vehicles in households contributed to ride-hailing demand more so in the mid-day and weekend night periods. This may be because the trips undertaken during the AM peak and PM peak periods are more likely to be associated with routine activities for which individuals without vehicles have already come to rely on other modes of transport. Lavieri et al. (2018) and Clewlow and Mishra (2017) also found that areas with higher vehicle ownership rates are associated with lower ride-hailing demand. The results also indicate that ride-hailing usage is likely to be higher in areas with a larger presence of single person households, consistent with previous findings (e.g., Henao and Marshall, 2018) that ride-hailing users are more likely to be young, single, and unmarried.

The coefficient for the log of the zonal size measure $\eta$ is estimated to be less than 1 indicating that there is some degree of cannibalization of prospective passengers by the different attraction locations within a zone. Indeed, earlier studies on shopping trips and work trips have also found competing effects between elemental locations within the same TAZ in attracting individuals (Bhat et al., 1998).

5.2 Measures of Fit

In this section, the goodness-of-fit of the NPMNL models is compared with that of simpler MNL models. All variables in the category of “TAZ Size independent attributes” in Table 2 are used in the MNL model specification. Since the MNL model does not allow for the nonlinear-in-parameter structure, the effect of TAZ size (in Equation (1)) must be captured by a single proxy variable for
size that is strictly positive. In the specification for the MNL model, the TAZ area is used for this purpose.

The metrics used for measuring the goodness-of-fit are log-likelihood, adjusted likelihood ratio index (ALRI), and average probability of correct prediction. To measure the model goodness-of-fit for the estimation datasets, the adjusted likelihood ratio index (ALRI; Windmeijer, 1995) is calculated as:

\[ \hat{\rho}^2 = 1 - \frac{L(\hat{\beta}) - Q}{L(c)}, \]

where, \( L(\hat{\beta}) \) is the likelihood of the model at convergence, \( L(c) \) is the likelihood at convergence of a naïve model, and \( Q \) is the difference in number of parameters between the naïve model and the estimated model. The naïve model refers to selecting a TAZ from the choice set at random, corresponding to a model where all estimated parameters are zero. Similarly, the predictive ALRI is calculated, with the likelihoods computed with respect to the validation sample instead of the estimation sample. The log-likelihoods and ALRIs are computed using the sampled choice sets that have 30 TAZs (including the destination). The average probability of correct prediction is computed using the full choice set containing all TAZs within 15 miles of the origin. The likelihood ratio test is used to statistically compare the goodness-of-fits of the NPMNL models and the MNL models for all time periods.

The goodness-of-fit statistics are presented in Table 3. An average probability of correct prediction in the order of 0.01 is quite reasonable considering that the average choice set size is 865.41 (note that a random TAZ selection would provide an average probability of correct prediction of only 00012; see third row of Table 3). The NPMNL model outperforms the MNL model in every metric for all time periods. The likelihood ratio test statistic between the NPMNL and the MNL models clearly illustrates the statistical superiority of the NPMNL model. Similar results are found in the validation sample, providing clear evidence that the superior data fit of the NPMNL in the estimation sample is not simply an artifact of overfitting.

6 IMPLICATIONS AND APPLICATIONS OF MODEL SYSTEM

The models developed in this paper provide key insights into the nature of deadheading trips and can be used to guide the development of policies related to ride-hailing services. Additionally, the modeling framework of this study can be incorporated in travel demand forecasting models, which do not currently take into account the impacts of deadheading trips. This section offers a discussion of the implications and applications of the model system developed in this study. Section 6.1 provides some of the direct implications of the model results presented in Section 5.1. A useful metric for policy makers and TNCs interested in reducing deadheading is the expected deadheading distance at various locations. Section 6.2 illustrates how this metric may be computed. Section 6.3 presents a discussion on how the number of deadheading trip interchanges may be computed from the model results.
Table 3  Goodness-of-Fit and Validation Statistics for the NPMNL and MNL Models

<table>
<thead>
<tr>
<th>Summary Statistic</th>
<th>AM Peak</th>
<th>Mid-day</th>
<th>PM Peak</th>
<th>Weekend</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimation Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. observations</td>
<td>31003</td>
<td>72858</td>
<td>59261</td>
<td>127899</td>
</tr>
<tr>
<td>Log-likelihood at zero</td>
<td>-105447.32</td>
<td>-247804.44</td>
<td>-201558.36</td>
<td>-435009.74</td>
</tr>
<tr>
<td>Average probability of correct prediction (random selection)</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0012</td>
</tr>
<tr>
<td><strong>NPMNL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of parameters</td>
<td>16</td>
<td>19</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Log-likelihood at convergence</td>
<td>-67846.04</td>
<td>-160071.21</td>
<td>-116372.01</td>
<td>-260577.59</td>
</tr>
<tr>
<td>ALRI</td>
<td>0.3564</td>
<td>0.3540</td>
<td>0.4225</td>
<td>0.4009</td>
</tr>
<tr>
<td>Average probability of correct prediction</td>
<td>0.0108</td>
<td>0.0191</td>
<td>0.0286</td>
<td>0.0143</td>
</tr>
<tr>
<td><strong>MNL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of parameters</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Log-likelihood at convergence</td>
<td>-70349.53</td>
<td>-165496.22</td>
<td>-120551.10</td>
<td>-273488.99</td>
</tr>
<tr>
<td>ALRI</td>
<td>0.3327</td>
<td>0.3321</td>
<td>0.4018</td>
<td>0.3713</td>
</tr>
<tr>
<td>Average probability of correct prediction</td>
<td>0.0093</td>
<td>0.0173</td>
<td>0.0257</td>
<td>0.0123</td>
</tr>
<tr>
<td><strong>Likelihood Ratio Test Statistic between MNL &amp; NPMNL</strong></td>
<td>5006.98</td>
<td>10850.01</td>
<td>8358.17</td>
<td>25822.81</td>
</tr>
<tr>
<td><strong>Validation Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. observations</td>
<td>15502</td>
<td>36429</td>
<td>29621</td>
<td>63950</td>
</tr>
<tr>
<td>Log-likelihood at zero</td>
<td>-52725.36</td>
<td>-123902.22</td>
<td>-100746.87</td>
<td>-217506.57</td>
</tr>
<tr>
<td>Average probability of correct prediction (random selection)</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0012</td>
</tr>
<tr>
<td><strong>NPMNL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of parameters</td>
<td>16</td>
<td>19</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Predictive Log-likelihood</td>
<td>-34071.74</td>
<td>-80592.04</td>
<td>-58115.99</td>
<td>-130487.14</td>
</tr>
<tr>
<td>Predictive ALRI</td>
<td>0.3535</td>
<td>0.3494</td>
<td>0.4230</td>
<td>0.4000</td>
</tr>
<tr>
<td>Average probability of correct prediction</td>
<td>0.0106</td>
<td>0.0195</td>
<td>0.0277</td>
<td>0.0142</td>
</tr>
<tr>
<td><strong>MNL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of parameters</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Predictive Log-likelihood</td>
<td>-35332.08</td>
<td>-83333.42</td>
<td>-60201.87</td>
<td>-136949.13</td>
</tr>
<tr>
<td>Predictive ALRI</td>
<td>0.3296</td>
<td>0.3273</td>
<td>0.4023</td>
<td>0.3703</td>
</tr>
<tr>
<td>Average probability of correct prediction</td>
<td>0.0092</td>
<td>0.0176</td>
<td>0.0248</td>
<td>0.0122</td>
</tr>
</tbody>
</table>
6.1 Planning and Policy Implications
The distance traveled per deadheading trip seems to be higher if the deadheading trip originates (or if the previous passenger was dropped off) in a suburban or rural area. This may be because it is relatively more difficult to find a new passenger in these areas and hence the driver needs to travel a farther distance for the next passenger pick-up. This may suggest that it would be prudent to discourage ride-hailing in suburban and rural areas so that the amount of deadheading mileage can be minimized. However, if policies that penalize or discourage ride-hailing services to operate in suburban and rural areas are implemented, the consequences of such policies need to be considered carefully. It is generally difficult to serve these areas with conventional transit (due to lower densities) and ride-hailing services constitute a convenient mobility option, especially for the transportation disadvantaged in these locations. The fact is that ride-hailing services fill an important mobility gap in areas that are not well served by other modes of travel. The question then arises as to how ride-hailing services in these areas can be deployed and priced in such a way that deadheading is minimized while fully realizing the mobility benefits that ride-hailing provide.

The model results suggest that ride-hailing caters to individuals who engage in social and entertainment activities in the weekend night period. The use of ride-hailing for these activities has merit because it leads to a reduction in the incidence of driving under the influence or using transit while intoxicated. In fact, several cities have witnessed a drop in driving under the influence after ride-hailing services became popular (Rabin, 2018; Richards, 2018). Also, ride-hailing demand is higher in zones well served by transit, particularly in the mid-day and PM peak periods when people are more likely to be engaging in out-of-home activities. Whether this is indicative of a competitive effect on transit usage remains to be determined, although recent evidence suggests that ride-hailing is likely to be substituting transit trips (Schaller, 2018), but complementing transit in some markets (Hall et al., 2018). However, when coupled with the finding that ride-hailing demand is higher in areas with higher retail and service accessibility and in urban areas, it is entirely possible that ride-hailing pick-ups are occurring in activity centers which have traditionally been better served by transit in the first place. In addition, it is likely that ride-hailing drivers searching specific areas for passengers can reduce waiting times and bolster ride-hailing demand in these areas.

It appears that ride-hailing is still primarily used by the niche demographics of young adults and wealthy households. If TNCs or urban planners wish to increase the use of ride-hailing in the city to reduce vehicle ownership, reduce need for parking, and promote Mobility-as-a-Service (MaaS), targeted efforts must be made to encourage the use of ride-hailing among other demographic groups. A reduction in space allocated to parking would allow cities to convert existing parking land to other valuable uses.

6.2 Computation of Expected Deadheading Distance
It is in the best interest of city planners, TNC operators, and society at large to minimize deadheading. The model developed in this paper can be used to compute expected deadheading distance for various locations. Once areas associated with higher expected deadheading distances are identified, policies and pricing schemes can be enacted to discourage individuals from making ride-hailing trips to these areas. However, the potential deadheading reductions associated with such strategies should be weighed against the potential adverse impacts on mobility enhancements that ride-hailing could provide such areas. The expected deadheading distance originating from a
TAZ can be computed as shown in Equation (11) using the probability mass function of TAZs being selected as the destination:

\[
E(d_i) = \sum_{j=1}^{N} d_{ij} P_{ij},
\]

(11)

where \( P_{ij} \) is the probability that a deadheading trip originating at TAZ \( i \) ends at TAZ \( j \) as computed by Equation (3) and \( d_{ij} \) is the network distance between the TAZs. The average deadheading distance in the PM peak period is mapped in Figure 4. To measure the accuracy of the expected deadheading distances predicted by the model, the predicted value is compared against the empirically calculated value of the expected deadheading distance for each TAZ. The expected deadheading distances are calculated empirically only for the 698 TAZs that have at least 10 deadheading trips originating within their boundaries. The mean absolute percentage error in the predicted expected deadheading distance across all of the TAZs is 20.75%, which is consistent with prediction accuracies that can be realized for models of the type presented in this paper. As expected, deadhead distances are smaller in higher density urban areas and larger in outlying suburban and rural areas.

6.3 Incorporation into Travel Demand Forecasting Models

It is relatively easy to forecast ride-hailing passenger trips in travel demand forecasting models. The mode choice model can be updated based on information gathered from travel surveys to reflect the use of ride-hailing options. The end points of passenger trips form the origins of deadheading trips. Once the origin of a deadheading trip is known, the destination of the trip can be determined based on the probability function given in Equation (3). If \( B_{ij} \) is the number of passenger trips that go from TAZ \( i \) to TAZ \( j \), then the number of deadheading trips \( M_{od} \) between an origin TAZ \( o \) and a destination TAZ \( d \) can be expressed as:

\[
M_{od} = \sum_i B_{io} P_{od},
\]

(12)

where \( P_{od} \) is the probability that a deadheading trip originating at TAZ \( o \) ends at TAZ \( d \) as computed by Equation (3). The above expression assumes that all ride-hailing vehicles that drop a passenger at TAZ \( o \) will make only one trip to pick up the next passenger. In other words, it is assumed that all deadheading trips belong to the first category as defined in Section 4.1. In reality, some drivers may end the ride-hailing session and return home or proceed to perform other personal activities. The deadheading trips resulting from the pursuit of these other personal activities are not modeled in this study. However, accounting for these deadheading trips in at least some manner is likely to be better than completely ignoring them.
7 CONCLUSIONS
The work in this paper is motivated by the rapidly increasing use of ride-hailing options in cities around the world. The use of ride-hailing services leads to empty vehicle mileage because ride-hailing service drivers (often) need to travel some finite distance to pick-up their next passenger after dropping off a prior passenger. These trips, referred to as deadheading trips, have important implications for vehicle miles of travel (VMT), traffic congestion, and carbon footprint of auto travel. Although there is considerable research dedicated to studying ride-hailing passenger trips and their characteristics, there is very little research on deadheading trips thus rendering it challenging to formulate strategies to reduce deadheading mileage and to account for such trips in travel demand forecasting models.

This paper presents a model for forecasting the destinations of deadheading trips. A dataset of deadheading trips was generated by imputing such trips from a dataset on ride-hailing passenger trips released by the TNC, Ride Austin. The model provides valuable insights on the factors that
affect deadheading trip patterns at different time periods in a day and at different locations. The model is sensitive to location specific characteristics related to the built environment, employment opportunities, and socio-demographic characteristics. The goodness-of-fit of the nonlinear-in-parameters multinomial logit (NPMNL) model developed in this paper is found to be significantly better than that of a simple multinomial logit (MNL) model. The paper presents a detailed discussion of possible applications of the model for transportation planning and travel demand forecasting.

Although this study makes a significant contribution to the understanding and modeling of ride-hailing deadheading trips, the following study limitations call for considerable future research in this domain. First, when computing deadheading distances, it was assumed that the deadheading occurs along the path with the shortest network distance between the origin and the destination. In reality, this would rarely be the case as the driver may cruise and meander in search of a new passenger. Deadheading trips are unlikely to occur along the path of lowest impedance, thus leading to an under-estimation of deadheading distances. Second, the model only accounts for deadheading trips that occur when the driver is searching for a new passenger after dropping off the previous passenger. There is another category of deadheading trips that involve the driver stopping the ride-hailing session and travelling to perform other activities. The dataset used in this study was not suitable for modeling this category of deadheading trips. Third, the model developed in this paper is not sensitive to the supply of ride-hailing vehicles. If there are more drivers searching/waiting for passengers, the average deadheading trip distance is likely to be higher. This effect is not explicitly captured in this study. Finally, the results in this paper do not apply to shared ride-hailing services where unrelated passengers can share rides. Because the Ride Austin platform did not provide this service at the time when the data was collected, this mode of ride-hailing could not be considered.

As noted earlier, most of the conventional travel demand forecasting models assume that all trips occur along the path with the lowest impedance. Since at least a part of the deadheading trip may be utilized to search for new passengers, the resulting path to the next pick-up location may not be along the path with the lowest impedance. The model presented in this paper only identifies the destination of the deadheading trip and not the path taken to reach that destination. Modeling the actual path taken in the search for new passengers would be a worthwhile endeavor for future research in terms of obtaining more accurate VMT implications of deadheading trips.

For the accurate modeling of travel demand, trips made while deadheading should not be ignored. A major constraint that inhibits the development and application of models of deadheading is the lack of data on individual ride-hailing trips. However, it appears that the tide is turning with ride-hailing companies beginning to release data. In 2018, DiDi – the TNC with the largest market share in China – made the data on all ride-hailing trips that occurred within a span of two months in the city of Chengdu, China, available to researchers. More recently, the city of Chicago passed an ordinance that requires TNCs to publish anonymized and disaggregate ride-hailing trip data every quarter. The first batch of this data was made available in April, 2019. If it is not possible to obtain the trip data directly from TNCs, researchers could conduct surveys of ride-hailing drivers to collect information on deadheading trips. Despite the limitations noted above, this paper presents one of the first attempts at understanding and modeling deadheading trips made by ride-hailing vehicles.

4 https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p
ACKNOWLEDGEMENT
This research was partially supported by the Data-Supported Transportation Operations and Planning (D-STOP) Center and the Center for Teaching Old Models New Tricks (TOMNET), both of which are Tier 1 University Transportation Centers sponsored by the US Department of Transportation. The work described in this paper was also supported by research grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (No. PolyU 152095/17E). The authors are grateful to Lisa Macias for her assistance in formatting the manuscript.

REFERENCES


Wenzel, T., Rames, C., Kontou, E. and Henao, A., 2019. Travel and energy implications of ridesourcing service in Austin, Texas. Transportation Research Part D, 70, 18-34.