MODELING FREQUENCY OF RURAL DEMAND RESPONSE TRANSIT TRIPS

Zohora Sultana[a], Sabyasachee Mishra[a*], Christopher R. Cherry[b], Mihalis M. Golias[a], Saman Tabrizizadeh Jeffers[b]

[a] Department of Civil Engineering, University of Memphis, 3815 Central Avenue, Memphis, TN 38152, United States

[b] Department of Civil & Environmental Engineering, The University of Tennessee, Knoxville, TN 37996, United States

ABSTRACT

Captive riders do not have many travel choices to meet the travel needs as fixed route transit services are not generally provided in rural areas. In many states, demand response transit (DRT) services are provided to meet such needs. However, state public agencies face the dilemma of whether to increase or decrease the service availability for on-call services. To enhance decision making of identifying what the causal factors related to DRT trips, the authors present a set of econometric models by integrating a sample DRT data with other explanatory variables such as land use, socio-economic, and demographic characteristics. Seven count data models including Poisson, Negative Binomial, Zero-inflated Poisson, Zero-inflated Negative Binomial (ZINB), Hurdle Poisson, Hurdle Negative Binomial, and ZINB Mixed Effect were developed to understand the factors that affect DRT trips. The ZINB Mixed Effect model that combines a zero-inflated negative binomial model with random effect was found to provide the best fit. A number of factors showed significant relationship with DRT trip frequency including distance, population density, elderly population, average income, and others. Further, the elasticity effects of these different factors were computed to quantify the magnitude of their impact on DRT. The proposed model can be helpful for transit agencies to predict the frequency of DRT trips and to provide adequate services in rural areas.

Author Keywords: Captive riders; Count models; Socio-economic and demographic characteristics, Demand response transit (DRT) services.

[*]Corresponding author.: Tel.: +1 901 678 5043, fax.: +1 901 678 3026

E-mail address: zsultan@memphis.edu (Zohora Sultana), smishra3@memphis.edu (Sabyasachee Mishra), cherry@utk.edu (Christopher R. Cherry), mgkolias@memphis.edu (Mihalis M. Golias), stabrizi@vols.utk.edu (Saman Tabrizizadeh),

INTRODUCTION

In rural areas, captive riders need to travel towards (sub)urban areas for financial, health, shopping and other needs. Such travel needs are crucial and unavoidable in most cases. This is because sub(urban) areas contain major public and private facilities for personal and professional services. With greater transit needs and fewer travel choices per capita, public transit is an important mode of transportation for rural residents who do not own or operate a car, albeit they do not have immediate access to private transportation or they are bound to use public transportation in order to meet their travel needs. In rural areas, travel demand density is lower and more dispersed, diminishing the effectiveness of traditional forms of fixed route bus-based public transport systems. Because of the low population density and dispersed origins and destinations, rural transit services usually have a very low fare box recovery rate, which results in abandonment of fixed route public transports after short period of operation. Alternatively, demand response transit (DRT) systems in rural areas can be more cost-effective by reducing frequencies and providing smaller vehicles. DRT service can adapt the changes in demand by either shifting its timetable and/or altering its route. The fare charged is very low or free depending on passenger socioeconomic characteristics and the route being served.

In many places, DRT remains an effective service which may only be available for specific groups of users like the elderly and/or mobility impaired. However, certainly there are other user groups who need DRT for non-discretionary trips. Davison et al. (2014) showed DRT as the most cost-effective way of ensuring the transit of rural communities without a conventional bus service. Enoch et al. (2006) evaluated DRT service in rural areas, but did not focus on demand distribution. This indicates a need of study to identify what destinations are essential for rural residents, as well as how the frequently rural residents will access these services.

In this research effort, the objectives are to identify transit demand in rural areas, exploring socio-economic and demographics patterns on DRT, and develop a method to assist state Departments of Transportation (DOTs) and transit providers to identify where transit connections and investments should be made. In this context, the authors have assembled a comprehensive dataset for analyzing DRT trip frequency, and developed count models to explore the effects of potential factors on DRT trip frequency. The rest of the paper is structured as follows. The next section presents literature review of rural transit mobility access. The third section discusses data requirement found in the literature and briefly introduces the reader to the data sources used in this research. The fourth section presents methodology for modeling DRT trips. The fifth section presents the model estimation results along with measures of fit, elasticity effects, and model validation results. The last section concludes the paper with summary of findings and proposes scope for future research.

LITERATURE REVIEW

Over the decades, DRT has developed as one of the most effective methods to provide transportation services to captive riders in rural areas. Many relevant studies in this area examine the effectiveness of DRT and explore the social and economic factors affecting transit trips. Bakker (1999) explained paratransit (DRT) as a "transportation option that falls between private car and conventional public bus services. It is usually considered to be an option only for less developed countries and for niches like elderly and disabled people". Ambrosino et al. (2004) described DRT as an "intermediate form of transport, somewhere between the bus and taxi, which covers a wide range of transport services, ranging from less formal community transport through to area-wide service networks". Wang and Winter (2010) showed that DRT has the potential to solve the challenges of the public transportation in low density urban areas. Braun and Winter (2009) have demonstrated that the collaborative transport can effectively solve classical transport planning problems in real-time. Ad-hoc DRT does not have pre-defined schedules and flexible routes but provides point-to-point transportation by reacting on demand in real-time. The fare of DRT is usually very low compared to taxis as it offers shared forms of transport and in some cases government subsidized costs. On the other hand, DRT has a long list of failure cases around the world. Enoch et al. (2006) listed several cases of DRT failed projects along with lessons that each provides. Their findings on the cause of DRT projects failure is that DRT often were not realistically designed with a full understanding of the demand of serving area and proper future plan. In many places, DRT cost is subsided by government considering this as a service for captive and low-income travelers and performance metrics should focus on the effectiveness as a social service. The cost of DRT implementation can be seven to ten times more expensive compared to fixed-route transit service. The cost of DRT implementation depends on service depends on the size of vehicles used, hours of operation, character and density of the service area, ridership levels, and to paying third-party contractors to provide the service (Goodwill, and Carapella 2008; TCRP 2004).

Paratransit microsimulation patron accessibility analysis tool has been developed by LaMondia and Bhat by combining paratransit trip data with census data to explore variables associated with paratransit trips in Brownsville, Texas. They conclude that paratransit trips are higher in census block groups with larger population, older populations, larger households, and close proximity to fixed route transit (LaMondia & Bhat, 2010)**.** Simple regression models have been developed for estimating ridership based on service characteristics of DRT service providers and demographic characteristics for rural demand response transit service (Mattson, 2017). They explored potential service characteristics as geographic coverage, span of services, fares, reservation requirements, and demographic characteristics as percentage of the population comprised of older adults or people without access to a vehicle etc. Multilevel models were developed to examine the effects of DRT supply-oriented factors and socio-economic attributes to estimate the demand for DRT services. The models predict that DRT users are higher in areas with higher levels of poverty, lower car ownership, lower population density, lower proportion of people working from home (Wang et al., 2014). Lerman et al. (1980) identified that vehicle ownership is negatively associated with service coverage of DRT. From a study of DRT services in Belgium, it was found that female, retired, homebound persons, and students are dominant users of DRT (Mageean & Nelson, 2003). Female and retired persons are identified as more than 50% of the users of DRT services from another study of DRT services in Tyne and Wear in the UK (Nelson & Phonphitakchai, 2012). Yang and Cherry (2017) studied the rural transit rider characteristics of Deviated Fixed Route Transit (DFRT) and DRT services for the State of Tennessee.

The study found that DFRT and DRT passengers are likely to be female, of minority races, of low personal/household income and low or zero car ownership, etc.

TCRP (1995) and TCRP (2004) found the elderly, mobility limited, and those on low incomes as potential markets for DRT in rural areas. Enoch et al. (2006) found target markets for DRT: people who cannot access public transport, people without personal transport, unemployed people, single pension households, individuals with a limiting long-term illness, ethnic minority households, and people aged 14-19 years. The various demographic characteristics of the population and trip characteristics affect the travel decisions. This research intends to review all those identified potential variables affecting DRT ridership, identify new variables and build a comprehensive set of statistical models to predict future traffic trends of DRT.

DATA

Consistent with previous findings (Mattson, 2017; Mageean & Nelson, 2003; Wang and Winter 2010), we categorized the independent variables into four groups: socio-economic, demographic, service characteristics, and land use characteristics. The socio-economic and demographic data includes age, gender split, vehicle ownership, household size and structure, and household income. The service characteristics include distance and time between zones. The land use characteristics include various types of land uses the origin and destinations of DRT trips. The hypotheses in selection of these variables are to assess the determinates of DRT trips. While it is expected that some variables such as lower income, higher age variables will positively associated with DRT trips; other variable such as larger distance between origin-destinations, and higher income will be inversely related to DRT trips. The goal of this research is to obtain the relationships of all four types of variables with DRT trips.

DRT services provided in the state of Tennessee are considered as the case study in this paper. All DRT trip occurrences for the year 2012 were collected from Tennessee Department of Transportation (TDOT). The characteristics of the DRT service is such that all the residents irrespective of income are eligible. Riders need to make a call for reserving a future ride trip. Services are available for Americans with Disability Act (ADA) as well. Services are provided both within and outside of the county. Prior reservation, at least five days in advance is needed for scheduling trips outside the county. Non-ADA travelers requiring special assistance for medical related trips were also available in the DRT service provided. The data set did not consider paratransit services provided by urban transit agencies.

Each trip record includes trip attributes such as origin and destination ZIP Code, County, and trip purpose. The data is provided by TDOT is at the ZIP Code level to maintain anonymity of the traveler. For each DRT trip corresponding demographic data was collected from American Community Survey (ACS) for each of the ZIP Codes in Tennessee. The demographic characteristics include age, gender split, vehicle ownership, household size and structure, household income etc. Combining socio-economic data for each of the ZIP Codes from ACS 2011 with DRT trip data, a comprehensive data set was developed. Further, service variables such as distance and travel time between ZIP Codes are determined using shortest path method and added to the dataset. All the trips from a specific origin ZIP Code to a destination ZIP Code have been accumulated to find total trip count for that pair. The final dataset contains number of trips between two ZIP Codes, the origin and destination ZIP Codes along with *DRT trip* features, *socio-economic and demographic* characteristics, and *level of service* measures.

Demographic data attributes of ZIP Codes are expected to be correlated to each other. To avoid multicollinearity problem, a correlation matrix is calculated consisting all continuous independent variables and one of the two highly correlated variables were dropped for inclusion in the final model dataset. There are total 640 ZIP Codes in the state of Tennessee which creates a total of 409,600 (640*640) origin-destination pairs. The number of *trips produced* from each ZIP Code to another ZIP Code is almost equal to the number of *trips attracted* by each ZIP Code from another ZIP Code. This is because almost all the trips reported in the travel diaries are round trips. Hence only production trips are considered for our study and a total of 205,120 (640*641/2) observations were found. By eliminating missing data for DRT trips, and eliminating zip codes where DRT service are not available, 185,500 records were kept for further analysis. 148,454 observations (80%) are used in model training and 37,046 (20%) observations are used for model validation. Figure 1 shows proportion of all trip purposes of DRT. The highest proportion of trips was for medical purposes (52.37 %). Hence, medical trip is the most important cause of making demand response trip in rural areas of Tennessee. Second largest trip purpose was for work related activities (employment, work, and customer home) which combines to 15.17 % of total travels. Other significant causes of trip request were recreation (3.62 %), senior center (3.37 %), and shopping (3.97 %).

<<Figure 1 Here>>

Figure 2 shows trip production for each ZIP code in six quintile levels (0, 1-4, 5-71, 72-346, 347-1099, 1100 and above). It is clear that most of the smaller cities closer to big cities are the main source of demand-response traffic generation. As example, Columbia city (ZIP Code 38401) is smaller sub-urban city, which is 44 miles away from Nashville, and produced highest number of trips in the whole state (13.10% of total trips). The second highest (6.31%) trip generating region was Tullahoma city (ZIP Code 37388) which is around 74 miles away from Nashville. Another significant trip generating area is Shelbyville city (ZIP Code 37160) which is 57 miles from Nashville. These information give insight of selecting covariates which may influence DRT trips.

<<Figure 2 Here>>

The potential independent variables are selected and presented in Table 1.

<<Table 1 Here>>

METHODOLOGY

Count or frequency models are usually considered as a parametric model where the model parameters are estimated from count observations. The parameters of the underlying distribution are specified as a function of different covariates to capture their influence on count dependent variable. Count variable has non-negative integer value which implies that a log-linear model is better fit for a count variable. A linear regression model generally produces negative predicted outcomes and there is a substantial problem of heteroscedasticity. Another advantage of using the log-linear specification is that, with count data, the effects of predictors are often multiplicative rather than additive. That is, one typically observes small effects for small counts, and large effects for large counts. If the effect is in fact proportional to the count, working in the log scale leads to a much simpler model. Poisson model is a good choice in this case.

The Poisson and Negative Binomial (NB) models are the two most commonly used parametric model in the literature for count data modelling (Washington et al., 2010). The Poisson model has a restrictive assumption of equi-dispersion property i.e., the expected mean is equal to the variance. The NB model overcomes that assumption, which makes it suitable for cases when there is over-dispersion or under-dispersion in the count data being modeled. Another aspect of considerable importance while modeling count data is over-representation of zeroes beyond the probability mass implied by the standard count models – a property referred to as the excess zeroes problem. Several variants of standard models including the zero-inflated count models, hurdle count models, and zero inflated mixed effect models were developed to address the excess zeroes problem (Fang et al., 2014; Gurmu, 1998; Hu et al., 2011; Hur et al., 2002; Moghimbeigi et al., 2008; Yang et al., 2016; Yau & Lee, 2001; Yau et al., 2003).

Considering modeling frequency of demand response trips, we analyze multiple models that showcases strengths in terms of performance. To achieve this, we begin modeling approach with basic poisson and negative models. To further model, excessive zeros we analyze some of the count data models that consider explicit considerations of excessive zeros by considering zero inflated poisson and negative binomial model, hurdle models, and mixed effect models. In this section, we briefly present specification of each model type for analyzing DRT trip frequency. A brief discussion of alternate modeling methods are follows.

Poisson Model:

In Poisson model, the probability of an event count y_i , given the vector of covariates X_i , is given by the Poisson distribution:

$$P(Y_i=y_i|X_i)= \frac{e^{-\lambda_i} \times \lambda_i^{y_i}}{y_i!}, y_i=0,1,2, \ldots\ldots \quad \text{Equation 1}$$

The mean parameter λ_i is a function of the vector of covariates.

$$E(y_i|X_i) = \lambda_i = exp(X_i'\beta) \quad \text{Equation 2}$$

Where, X_i' is the vector of exogenous variables and β is the corresponding vector of coefficients.

Negative Binomial Model

In the NB model, the probability of observing count outcome y_i conditional on the expected mean parameter λ and dispersion parameter $\theta>0$ is given by:

$$P(Y=y)=\left(\frac{\theta}{\theta+\lambda}\right)^{\theta}\times\frac{\Gamma(\theta+y)}{\Gamma(y+1)\Gamma(\theta)}\times\left(\frac{\lambda}{\theta+\lambda}\right)^{y} \qquad \text{Equation 3}$$

Where Γ is the gamma function defined as follows:

$$\Gamma(t)=\begin{cases}\int_{x=0}^{\infty}x^{t-1}e^{-x}dx & \text{for positive non-integer t}\\ (t-1)! & \text{for positive integer t}\end{cases} \qquad \text{Equation 4}$$

The variance of the NB model is $v=\lambda+\frac{\lambda^2}{\theta}$. Here, θ is an over-dispersion parameter and λ is the expected mean.

Zero-inflated Models

Zero-inflated count models provide a way of modeling the excess zeros in addition to allowing for over dispersion. In particular, for each observation, there are two possible data generation processes. For each observation, Process 1 is chosen with probability p_i and Process 2 with probability 1-p_i. Process 1 generates only zero counts, whereas Process 2, $P(Y_i=y_i|X_i)$, generates counts from either a Poisson or a NB model. In general:

$$y_i=\begin{cases}0 & with\ probability\ p_i\\ P(Y_i=y_i|X_i) & with\ probability\ 1-p_i\end{cases} \qquad \text{Equation 5}$$

The probability $P(Y_i=y_i|X_i)$ depends on the process where it is zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB). Zero-inflated model consists of binary logit model and counts models. Binary logit model is commonly used to predict a behavior's occurrence, but with ZIP /ZINB, the logistic regression part of the model predicts non-occurrence (i.e., it predicts the zeros). The count models predict how frequently the behavior occurred.

The expected count is function of the two processes. In this study, the expected trip count is defined as follows:

$$E(y_i) = p_i * 0 + (1 - p_i) * e^{\alpha_i} \qquad \text{Equation 6}$$

p_i is the predicted probability that trip count is zero, e^{α_i} is the expected trip count given it is not zero and it is modeled using Poisson/NB regression.

The probability whether the trip is not possible (zero part), p_i is modeled by a logistic regression. Its form is:

$$log\left(\frac{p_i}{1\text{-}p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \qquad \text{Equation 7}$$

β_i is the parameter that will be estimated and X_i is the feature of the ZIP Code, such as population density, household income, and trip distance. e^{α_i} is modeled using Poisson/NB regression. Its form is:

$$e^{\alpha_i} = e^{(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots)} \qquad \text{Equation 8}$$

α_i is the parameter that will be estimated and again, X_i is the feature of the ZIP Code.

Hurdle Models

In hurdle models, the count data generating process is controlled by Bernoulli probability that governs the binary outcome of whether a count variable has a zero or non-zero value. If the value is positive, the hurdle is crossed, and the conditional distribution of the non-zero outcome is governed by a Poisson/NB count data model. Hence Hurdle models can take shape of various count structures such as: Hurdle Poisson (HP) or Hurdle NB (HNB). In general, the hurdle model has two parts:

1. Zero count generating model.
2. Value (positive) generating model.

These two models are not considered to be the same. Hence, the difference from zero-inflated model is that the value generating part is not allowed to create zero outcomes. If the predicted variable $y_i>0$, the hurdle is crossed, the conditional distribution of the count value is governed by value generating model part. The zero-generating model can be considered as a logit model:

$$P(y_i = 0|X_i) = \frac{\exp(X_i'\gamma)}{1+\exp(X_i'\gamma)} = p_i \qquad \text{Equation 9}$$

The value generating part of the model has conditional probability of count value given that the number is greater than zero. If we consider that the value generating model is Poisson model:

$$P(y_i = j|y_i > 0, X_i) = \frac{P(y_i=j \ \& \ y_i>0|X_i)}{P(y_i>0|X_i)} = \frac{\exp(X_i'\beta)^\wedge j \ \mathrm{e}^{-(\mathrm{X}_j'\beta)}}{j![1-e^{-(X_i'\beta)}]}, j = 1,2, \ldots \qquad \text{Equation 10}$$

So, the expected value of y_i is

$$E[y_i|X_i] = p_i * 0 + (1 - p_i) * E[y_i|y_i > 0, X_i] \qquad \text{Equation 11}$$

If there is over-dispersion, the estimate of the parameters from HP will be biased and inconsistent. In that case, the NB is a good substitute as a value generating model. For a HNB model, a dependent variable Y_i (i=1, 2, …, n) has the distribution

$$Pr(Y_i = y_i) = \begin{cases} p_i, & y_i = 0, \\ (1-p_i)\frac{\Gamma(y_i+\theta^{-1})}{\Gamma(y_i+1)\Gamma(\theta^{-1})}\frac{(1+\theta\lambda_i)^{-\theta^{-1}-y_i}\theta^{y_i}\lambda_i^{y_i}}{1-(1+\theta\lambda_i)^{-\theta^{-1}}}, & y_i > 0, \end{cases} \qquad \text{Equation 12}$$

Where, θ (≥0) is dispersion parameter that is assumed not to be dependent on independent variables. p_i is a non-negative function that is modeled via logit link function,

$$logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \sum_{j=1}^{m} z_{ij}\delta_j \qquad \text{Equation 13}$$

Where, z_{ij} is i-th row of covariate matrix Z and δ_j are unknown m-dimensional column vector of parameters.

Zero-inflated Negative Binomial Mixed Effect Model

Zero-inflated negative binomial mixed effect models (ZINB Mixed Effect) were developed to address over-dispersed count data with excess number of zeros (Fang et al., 2014; Moghimbeigi et al., 2008; Yau et al., 2003). This mixed model contains extra parameters to model the probability of excess zero values and the variability in non-zero values, allowing for repeated measures incorporating independent random effects for these two parts. ZINB Mixed Effect model can be expressed as follows:

$$\log(\lambda_{ij}) = X_{ij}\beta + a_i \quad \text{Equation 14}$$

$$logit(p_{ij}) = Z_{ij}\gamma + b_i \quad \text{Equation 15}$$

Here, X_{ij} represents the matrix of covariates and β is their respective regression coefficient for the negative binomial part, Z_{ij} represents the covariate matrix and the respective vector of regression coefficient γ for the logistic part, a_i and b_i are the random intercepts and they follow normal distribution with mean zero. For simplicity those intercepts are assumed to be independent. This assumption is also used in the literature of ZINB/ZIP with random effects (Fang et al., 2014; Hur et al., 2002; Yau & Lee, 2001).

RESULTS AND DISCUSSION

In this section, results of count data models for DRT trip frequency is presented followed by goodness-of-fit measures, elasticity estimation and model validation.

Estimation Results

A comparison of the estimation results of seven count data models: Poisson (Model 1), NB (Model 2), ZIP (Model 3), ZINB (Model 4), HP (Model 5), HNB (Model 6) and ZINB Mixed Effect (Model 7) is presented in Table 2. The statistically significant explanatory variables along with their estimated coefficients and t-statistics (in parenthesis) for each of the developed models are shown in Table 2. The log-likelihood value at convergence, the Bayesian Information Criterion value (BIC), and the total number of observations are also included for each model.

Poisson regression is one of the most basic count regression models. The explicit assumption used for Poisson model is that the mean and variance of count variable are statistically equal. Given that there is no a priori reason for the mean and variance in any practical context to be equal, the use of a NB distribution for Model 2, 4, 6, and 7 is an important empirical generalization over the Poisson distribution. The NB model is considered as a generalization of Poisson model since it has the same mean structure as Poisson regression and it has an extra parameter (θ) to model over-dispersion. If the conditional distribution of the outcome variable is over-dispersed, the confidence intervals for the NB regression are likely to be narrower as compared to that of a Poisson model. In the NB model, the dispersion parameter properly captures the difference between mean and variance. However, the NB model needs to be further examined to model DRT trip frequency due to the presence of excessive zeros in this dataset.

Zero-inflated models (Model 3 through 7) accounts for presence of excess zeros in the trip frequency. The distribution of dependent variable (Trips) is extremely skewed because of excess number of zero trip occurrences (97.78% of origin-destination pairs) in trip count data. Zero inflated and hurdle models are good candidates for this data which address over-dispersion effectively. The difference between mean and variance is still high even if all zero trips occurrences are not taken into account which has standard deviation (678.03) that is much higher than mean (73.51). The results from ZINB model demonstrate that we can indeed reject the hypothesis that the trip generation process is Poisson, since $\log(\theta)$ =-2.470 with p-value < 0.0001, and thus the variance of the process is much larger than the mean. The estimate of significant positive intercept in logistic model part proves that there is excess number of zeros in the data.

On the other hand, the Hurdle model in case of zero estimation is different from zero-inflated models. The sign of estimated parameters in the hurdle model is not opposite for value-estimation and zero-estimation parts because these two processes are independent and likely to follow similar effect over trip count. In addition, to achieve inter-ZIP Code trip variability, which is not captured well by covariates, the origin and destination ZIP Codes are introduced in the zero-estimation part of the model as random effect parameters. The estimated standard deviations of those random effects are significantly large. Another random effect variable incorporated in the value estimate part of the ZINB Mixed Effect model is, the rural urban commuting area (RUCA), which indicates the type of ZIP Code area based on the size and direction of the primary commuting flows. The estimated standard deviation of this random effect variable indicates that the trip count has variability across different types of ZIP Codes. The estimated standard deviations (σ) of the random effects are presented in Table 2.

The variables that have significant effect on DRT trip frequency includes *Origin ZIP Code Population Density, Distance Between Two ZIP Codes, Population Aged 14 Years Or Less, Population Aged 65 Years Or Over, The Number Of Disabled People, Household Median Income , Homeowner Vacancy Rate , Average Household Size, The Number Of Renter Occupied Housing Unit, Dominant Sex (Male), Proportion Of White People, The Number Of Wholesale Trade Establishments In Destination ZIP, and The Number Of Retail Trade Establishments In Destination ZIP Code.* The estimated parameter signs are similar across the models which means the effect of variables are consistent. The results indicate that lower *Population Density* is likely to increase the overall trip count. The similar relationship between this variable and DRT demand is also found in the demand model developed by Wang et al.(2014). This is intuitive because of unavailability of demand response service in an urban area where the population density is higher and lower the density means the ZIP Code area is in rural area. It is more likely to have fixed route public transportation services in an urban area. Moreover, people living in higher population density areas can coordinate with others to make a trip. The *Distance Between ZIP Codes* has an opposite effect over trip count. The results indicate that with increasing distance the likelihood of occurrence of a DRT trip decreases. This is intuitive because DRT serves trips that are relatively short and not supporting inter-city type services that tend to cover long distance.

Trip count is likely to decrease with the increase of *Younger Population (age 14 or less)* in the origin ZIP Code area. The presence of children of less than 14 years reduces DRT trips as parents are typically not elderly and may own a car in such households. On the other hand, *Older Age Group (age 65 or over)* population has positive association on trip count. This finding is consistent with the most previous studies (Mattson, 2017; C. Wang et al., 2014). They likely rely on DRT for medical services and increasingly, the baby boomer generation is “aging in place”. Moreover, the aged population might not own a car or be unable to drive. A similar effect observed from *Disabled Population* in the origin ZIP Code where the trip count increases with the disabled population size increase. This result is consistent with the research conducted by Mattson (2017). The disabled population tends to be most captive to transit services and may need additional medical services. The variable *Household Median Income* has a negative impact on trip count because people like to get their own vehicle when they have higher income level. This finding coincides with the research conducted by Yang and Cherry (2017). The variables showing positive association with DRT trip frequency shows the need for providing service in rural areas where fixed transit service is not provided.

<<Table 2 Here>>

Homeowner Vacancy Rate in origin ZIP Code are likely to increase DRT trip in the sense that we have higher homeowner vacancy rate in rural area. The *Average Household Size* is also likely to increase the trip count. Number of *Renter Occupied Housing Unit In Origin ZIP Code* has positive impact over trip count. This is because the renter occupied people in rural area is less likely to own and operate a vehicle. The variable *Sex* indicates that women are the primary user of DRT service. If the Origin ZIP Code with higher number of female compared to male, it is more likely to induce demand for DRT trip. This similar relation is also observed in the DRT and DFRT study of Yang and Cherry (2017). In case of variable *Race,* white people are most likely to use DRT service in rural areas. This finding coincides with the research conducted by Wang et al.(2014). When destinations are based on *Retail Trade*, they are likely to attract more DRT trips as population from neighboring areas will likely to make trips for retail goods. However, *Whole Sale Trade* shows an inverse relationship with DRT frequency.

Model Selection and Statistical Fit

Two goodness-of-fit indices were used to evaluate the fitness of the model, log-likelihood, and BIC. Goodness-of-fit indices for the seven models are shown in Table 2. The log-likelihood value of ZINB Mixed Effect model is highest. To facilitate comparison across different models estimated in this study, BIC value was computed as: $-2\times LL+K\times LN(N)$, where K is the number of model parameters and N is the number of observations in the estimation sample. According to the BIC criterion, a model with lower BIC value is preferred over model with higher BIC value. The ZINB Mixed Effect model also had the lowest BIC value.

Elasticity Effects

In order to determine the magnitude of effects of the independent variables on DRT trip frequency, it is necessary to compute their corresponding elasticity effects. The elasticity effect represents the percentage change in the expected number of DRTs due to a unit percentage change in an explanatory variable. Table 3 presents the elasticity effect of the best performing ZINB Mixed Effect model. The elasticity parameter of population density indicates that doubling the log of population density in the origin ZIP Code will cause the expected trip counts to be decreased by 0.716%. Highest elasticity effect was observed on distance variable. It indicates that the trip generation will be decreased by 2.026% with one unit increase of the distance (in miles) between origin and destination ZIP Code. Other elasticity values can be interpreted similarly.

<<Table 3 Here>>

Figure 3 shows the effects of socio-economic variables on the frequency of DRT trip. The vertical axis shows percentage changes in frequency of DRT trip. The effects of 10, 20, and 30% increase in *Retail Trade, The Number Of Disabled People, Population Aged 14 Years Or Less, The Number Of Renter Occupied Housing Unit And Population Density* on DRT trip count are shown. As expected, the figure shows that the DRT trip frequency decreases with the *Population Density* and *Population Aged 14 Years* or less increases. On the other hand, DRT trip frequency increases with the increased number of *Disabled People*, and *Population Aged 14 Years Or Less, Renter Occupied Housing Unit, Retail Trade* as shown in Figure 3.

Model Validation

To test the predictive power of these models, a validation exercise was undertaken in which the predicted demand trip counts were compared with the observed counts in the data (Table 4). The dataset is divided into training set (80%) and test set (20%) by randomly taking data points. Absolute Percentage Difference (APD) between predicted and observed shares for each count outcome was computed. Next, Average Absolute Percentage Difference (AAPD) across all count outcomes was computed and used as a metric of predictive performance. Models with lower AAPD value are preferred over models with higher AAPD values. From Table 4, it is seen that ZINB Mixed Effect model better suited to capture dispersion in count data among all models for DRT trips in rural areas.

<<Table 4 Here>>

CONCLUSIONS

The contribution of this study is threefold. First, application of various count data models to analyze frequency of rural DRT trips. Second, determination of socio-economic, demographic, service, supply and demand characteristics' impact on rural DRT. Third, identifying important factors contributing to DRT trips to assist state Departments of Transportation (DOTs) and transit providers to identify where transit connections and investments should be made. The primary objective of this research was to develop a set of econometric models that can predict DRT trip frequency as a function of land use, socio-economic and demographic characteristics. We test these models on DRT trip data for rural areas in the state of Tennessee. To be specific, seven count data models; Poisson, Negative Binomial (NB), Zero-inflated Poisson (ZIP), Zero-inflated Negative Binomial (ZINB), Hurdle Poisson (HP), Hurdle Negative Binomial (HNB), and ZINB Mixed Effect were developed to determine the causal factors related to DRT trips. BIC and Log-likelihood was computed to compare different models. In addition, the predicted number of DRT trips was used for model validation. The ZINB Mixed Effect model performed better compared to all other models on model fit statistics and on the validation exercise. The significant contributing factors of DRT trip frequency are trip distance, population density, population aged 14 years or less, population aged 65 years or over, average household size, average income, retail and wholesale trade and others. The elasticity effects of all variables entered ZINB Mixed Effect model were also computed to understand clearly the impacts of those variables. The variables with the largest effect were trip distance, population aged 65 years or over, disabled population etc.

In terms of future research, characteristics of DRT service providers should be taken into account while developing models for better prediction of DRT trip frequency. These characteristics (i.e. reservation requirements, fare, days of operation per week etc.) may impact the trip count in their serving area. Inclusion of theses service characteristics information with demographic and land use data of ZIP Codes should provide better predictive outcome. In addition, if more attributes of the trip makers were available (Yang & Cherry, 2017), the models could have developed at a finer geographic level or even at individual level rather than ZIP Codes. The models can be strengthened if time-of-day travel information is available to predict DRT trips by various times of the day. Future research can also investigate the frequency of DRT trips in areas where fixed transit service is available versus areas with no fixed transit service.

ACKNOWLEDGEMENT

This research was funded by TDOT. The authors would like to thank TDOT for providing data and support during the research period. In addition, computational facilities at University of Memphis and University of Tennessee Knoxville are greatly acknowledged.

REFERENCES

Ambrosino, G., Nelson, J. D., & Romanazzo, M. (2004) Demand Responsive Transport Services: Towards the Flexible Mobility Agency, Italian National Agency for New Technologies. Energy and the Environment, Rom.

Bakker, P. (1999) Large scale demand responsive transit systems - A local suburban transport solution for the next millennium? Working Paper, AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management, Rotterdam, Netherlands.

Beyazit, E. (2011). Evaluating social justice in transport: Lessons to be learned from the capability approach. Transport Review, 31 (1), pp. 117-134.

Braun, M., Winter, S. (2009) Ad hoc solution of the multicommodity-flow-over-time problem. IEEE Transactions on Intelligent Transportation Systems, 10, 658–667.doi: 10.1109/TITS.2009.2026443.

Davison, L., Enoch, M., Ryley, T., Quddus, M., Wang, C. (2014) A survey of demand responsive transport in Great Britain. Transport Policy, 31, 47–54. doi: 10.1016/j.tranpol.2013.11.004

Enoch, M., Potter, S., Parkhurst, G., Smith, M. (2006) Why do demand responsive transport systems fail? Presented at the Transportation Research Board 85th Annual Meeting.

Fang, R., Wagner, B., Harris, J.K., Fillon, S.A. (2014) Application of zero-inflated negative binomial mixed model to human microbiota sequence data (No. e215v1). PeerJ PrePrints.

Goodwill, J. and H. Carapella. "Creative Ways to Manage Paratransit Costs." USF Center for Urban Transportation Research. 2008. pp. 1-88. https://www.nctr.usf.edu/pdf/77606.pdf.

Gurmu, S. (1998) Generalized hurdle count data regression models. Economics Letters, 58, 263–268. doi: 10.1016/S0165-1765(97)00295-4.

Hu, M.C., Pavlicova, M., Nunes, E.V. (2011) Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: examples from an HIV-Risk Reduction Intervention Trial. The American Journal of Drug and Alcohol Abuse, 37, 367–375. doi:10.3109/00952990.2011.597280.

Hur, K., Hedeker, D., Henderson, W., Khuri, S., Daley, J. (2002) Modeling clustered count data with excess zeros in health care outcomes research. Health Services and Outcomes Research Methodology, 3, 5–20. doi: 10.1023/A:1021594923546.

Kim, C., Wang, S. (2015). Empirical examination of neighborhood context of individual travel behaviors Applied Geography, 60, pp. 230-239.

LaMondia, J.J., Bhat, C.R. (2010) Development of a paratransit microsimulation patron accessibility analysis tool for small and medium sized communities. Transportation Research Record: Journal of the Transportation Research Board, 2174, 29–38.

Lin, J., Long, L. (2008). What neighborhood are you in? Empirical findings of relationships between household travel and neighborhood characteristics. Transportation, 35, pp. ,739-758.

Lerman, S.R., Flusberg, M., Pecknold, W.M., Nestle, R.E., Wilson, N.H. (1980) A model system for forecasting patronage on demand responsive transportation systems. Transportation Research Part A: General, 14, 13–23. doi: 10.1016/0191-2607(80)90090-4.

Mageean, J., & Nelson, J. D. (2003) The evaluation of demand responsive transport services in Europe. Journal of Transport Geography, 11, 255–270.

Mattson, J. (2017) Estimating Ridership of Rural Demand-Response Transit Services for the General Public. Transportation Research Record: Journal of Transportation Research Board, 2647, pp. 127-133.

Moghimbeigi, A., Eshraghian, M.R., Mohammad, K., Mcardle, B. (2008) Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. Journal of Applied Statistics, 35, 1193–1202. doi: 10.1080/02664760802273203.

Nelson, J. D., & Phonphitakchai, T. (2012) An evaluation of the user characteristics of an open access DRT service. Research in Transportation Economics, 34, 54–65.

TCRP (1995) Workbook for Estimating Demand for Rural Passenger Transportation. Synthesis Report 3, Transit Cooperative Research Program (TCRP). Federal Transit Administration, Washington, DC.

TCRP (2004) Demand Responsive/ADA Traveler Response to Transportation System Changes, Synthesis Report 95, Transit Cooperative Research Program (TCRP). Federal Transit Administration, Washington, DC.

Vanasse Hangen Brustlin, Inc., LSC Transportation Consultants, Inc., and Erickson Consulting, LLC. (2013) Methods for Forecasting Demand and Quantifying Need for Rural Passenger

Transportation: Final Workbook. TCRP Report 161, Transit Cooperative Research Program, Washington, DC: Transportation Research Board.

Wang, C., Quddus, M., Enoch, M., Ryley, T., Davison, L. (2014) Multilevel modelling of Demand Responsive Transport (DRT) trips in Greater Manchester based on area-wide socio-economic data. Transportation, 41, 589–610.
doi: 10.1007/s11116-013-9506-1.

Wang, H., Winter, S. (2010) Utilizing taxi empty cruise time to solve the short distance trip problem. Presented at 17th ITS World Congress, Busan.

Wang, K., and Woo, M. (2018). The relationship between transit rich neighborhoods and transit ridership: Evidence from the decentralization of poverty. Applied Geography, 86, pp. 183-196.

Washington, S.P., Karlaftis, M.G., Mannering, F. (2010) Statistical and econometric methods for transportation data analysis. CRC press, NewYork.

Yang, H., Cherry, C.R. (2017) Use characteristics and demographics of rural transit riders: a case study in Tennessee. Transportation Planning and Technology, 40, 213-227. doi: 10.1080/03081060.2016.1266168.

Yang, H., Cherry, C.R., Zaretzki, R., Ryerson, M.S., Liu, X., Fu, Z. (2016) A GIS-based method to identify cost-effective routes for rural deviated fixed route transit. Journal of Advanced Transportation, 50, 1770-1784. doi: 10.1002/atr.1428.

Yau, K.K., Lee, A.H. (2001) Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. Statistics in medicine, 20, 2907-2920. doi: 10.1002/sim.860.

Yau, K. K., Wang, K., & Lee, A. H. (2003) Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. Biometrical Journal, 45, 437-452. doi: 10.1002/bimj.200390024.

TABLE 1: Summary Statistics of Model Variables

Variable	Min	Mean	Max	Standard deviation	Values
Destination ZIP Code Retail Trade related businesses	0	40.55	401	63.01	Continuous
Destination ZIP Code Wholesale Trade related businesses	0	11.94	275.00	24.70	Continuous
Distance between two ZIP Code (mi)	0	176.60	544.50	106.26	Continuous
Origin ZIP Code Population density (/mi^2)	4.72	475.10	17,840	1,235.69	Continuous
Origin ZIP Code Proportion of white population	0.02	0.859	1.00	0.19	Continuous
Origin ZIP Code Household median income	8,524	40,000	136,200	13,652.29	Continuous
Origin ZIP Code Average household Size	1.46	2.62	19.96	0.96	Continuous
Origin ZIP Code Homeowner vacancy rate	0	1.49	37.82	2.31	Continuous
Origin ZIP Code Renter occupied housing Unit	0	1,119	14,530	1,748.65	Continuous
Origin ZIP Code Disabled population	0	1,444	9,259	1,654.53	Continuous
Origin ZIP Code Population aged 14 years or less	0	1,781	16,800	2,399.16	Continuous
Origin ZIP Code Population aged 65 years or over	1	1,284	7,975	1,502.29	Continuous
Origin ZIP Code Household income 200K or more	0	93.93	4,448	242.51	Continuous
Rural urban commuting area type of Origin ZIP Code (1-Metropolitan, 2-Micropolitan, 3-Small town, 4-Rural)	1	2.24	4	1.16	Categorical
Rural urban commuting area type of Destination ZIP Code (1-Metropolitan, 2-Micropolitan, 3-Small town, 4-Rural)	1	1.84	4	1.10	Categorical
Dominant Sex of Origin ZIP Code (1-Male, 0-Female)	0	0.35	1	0.47	Categorical
Dominant Race of Origin ZIP Code (1-White, 0-Black)	0	0.92	1	0.26	Categorical

TABLE 2: Model Results

Variables	**Poisson (Model 1)**	**NB (Model 2)**	**ZIP (Model 3)**	**ZINB (Model 4)**	**HP (Model 5)**	**HNB (Model 6)**	**ZINB Mixed Effect (Model 7)**
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Value estimation part							
Intercept	6.003 (58.785)	-11.013 (-7.76)	2.486 (25.980)	1.142 (2.725)	-1.169 (-50.77)	-9.003 (-0.612)	0.512 (.223)
Distance between two ZIP Code	-0.112 (-939.448)		-0.090 (-762.450)	-0.031 (-35.488)	-0.087 (-725.68)	-0.031 (-32.717)	-0.030 (-39.610)
Origin ZIP Code Population density	-1.192 (-524.096)						
Log(Origin ZIP Code Population density)			-1.043 (-466.470)	-0.652 (-18.624)	-1.223 (-477.01)	-0.522 (-11.407)	-0.403 (-12.300)
Log(Origin ZIP Code Population aged 14 years or less)		-.665 (-7.674)		-0.387 (-5.784)	-0.215 (-30.940)	-0.329 (-3.936)	-0.346 (-5.960)
Proportion of white people in Origin ZIP Code		1.717 (8.156)					
Log(Origin ZIP Code Population aged 65 years or over)	0.317 (37.179)	0.713 (6.516)					
Log(Origin ZIP Code Household Median income)	-1.150 (-116.075)	-0.386 (-2.889)	-0.602 (-65.750)				
Destination ZIP Code Retail Trade related businesses	0.009 (459.281)						
Log (Destination ZIP Code Retail Trade related businesses)		1.439 (32.024)	0.727 (385.870)	0.731 (17.065)	0.823 (279.04)	0.843 (15.862)	0.562 (13.340)

Variables	Poisson (Model 1)	NB (Model 2)	ZIP (Model 3)	ZINB (Model 4)	HP (Model 5)	HNB (Model 6)	ZINB Mixed Effect (Model 7)
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Log (Destination ZIP Code Whole Sale Trade related businesses)		-0.698 (-13.724)		-0.336 (-6.533)	-0.180 (-63.780)	-0.415 (-6.160)	-0.004 (-.080)
Origin ZIP Code Homeowner vacancy rate	0.187 (185.638)		0.161 (165.060)				
Origin ZIP Code Average Household size	0.016 (2.134)		0.668 (113.060)				
Log(Origin ZIP Code Disabled population)	1.551 (172.618)	1.423 (17.111)	1.424 (491.550)	0.205 (2.441)	1.069 (151.3)	0.336 (3.065)	0.460 (5.720)
Log (Origin ZIP Code Renter occupied housing unit)				0.806 (10.546)	0.636 (96.42)	0.638 (6.416)	0.502 (7.230)
Origin ZIP Code Dominant Sex (male)	-0.344 (-68.014)						
Origin ZIP Code Dominant Race (white)	2.814 (114.842)						
Log (θ)		-5.149 (-50.902)		-2.47 (-63.796)		-11.690 (-0.796)	-1.437 (-46.477)
Random effects parameters:							
σ (RUCA type of Origin ZIP Code)							0.00013
σ (RUCA type of Destination ZIP Code)							0.502
Zero estimation part							
Intercept			4.592 (4.729)	4.841 (18.334)	-4.598 (-6.502)	-2.065 (-3.088)	9.662 (10.060)
Origin ZIP Code Average			0.533 (7.369)	0.205 (5.513)	-0.297 (-4.528)	-0.236 (-4.503)	0.438 (3.060)

Variables	**Poisson (Model 1)**	**NB (Model 2)**	**ZIP (Model 3)**	**ZINB (Model 4)**	**HP (Model 5)**	**HNB (Model 6)**	**ZINB Mixed Effect (Model 7)**
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Household size							
Destination ZIP Code Retail Trade related businesses			-0.284 (-19.866)				
Log(Destination ZIP Code Retail Trade related businesses)				-0.668 (-25.879)			-0.954 (-11.830)
Log(Origin ZIP Code Household Median income)			-0.217 (-2.321)		0.153 (2.375)	-0.197 (-3.128)	
Log(Origin ZIP Code Population aged 14 years or less)			0.522 (10.110)	0.343 (3.767)	-0.320 (-8.459)	-0.348 (-8.988)	1.143 (3.990)
Log(Origin ZIP Code Population aged 65 years or over)				-0.311 (-3.129)		0.530 (9.129)	-1.440 (-3.950)
Log(Origin ZIP Code Disabled Population)					0.274 (5.166)	0.196 (3.492)	
Log(Origin ZIP Code Renter occupied housing unit)			-0.781 (-14.048)	-0.730 (-13.171)	0.733 (15.619)	0.211 (6.857)	-0.656 (-3.170)
Origin ZIP Code Number of Households with income 200K or more			-0.001 (-5.404)				
Log(Origin ZIP Code Population density)			0.059 (2.717)		-0.289 (-13.949)		
Distance between two ZIP Codes				0.039 (37.368)	-0.034 (-59.382	-0.033 (-59.506)	0.053 (32.240)

Variables	**Poisson (Model 1)**	**NB (Model 2)**	**ZIP (Model 3)**	**ZINB (Model 4)**	**HP (Model 5)**	**HNB (Model 6)**	**ZINB Mixed Effect (Model 7)**
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Random effects parameters:							
σ (Origin ZIP Code)							2.636
σ (Destination ZIP Code)							2.509
Measures of fit							
Log-Likelihood at convergence	-613,048	-27,933	-36,8191	-22,421	-376,412	-23,222	-19,585
BIC	1,226,228	55,973	736,573	45,033	753,015	46,648	39,408
Number of observations	148,454	148,454	148,454	148,454	148,454	148,454	148,454
Number of parameters	11	9	16	16	16	17	20

TABLE 3: Elasticity Effects of the ZINB Mixed Effect Model

Variables	ZINB Mixed Effect
Value estimation part	
Log (Origin ZIP Code Population density)	-0.716
Log (Destination ZIP Code Retail Trade related businesses)	0.562
Log (Destination ZIP Code Whole Sale Trade related businesses)	-0.0024
Distance between two ZIP Code	-2.026
Log (Origin ZIP Code Population aged 14 years or less)	-0.869
Log (Origin ZIP Code Disabled population)	1.154
Log (Origin ZIP Code Renter occupied housing unit)	1.127
Zero estimation part	
Log (Origin ZIP Code Population aged 65 years or over)	-3.524
Log (Origin ZIP Code Renter occupied housing unit)	-1.437
Origin ZIP Code Average Household size	0.439
Log (Origin ZIP Code Population aged 14 years or less)	2.871
Distance between two ZIP Code	3.585
Log (Destination ZIP Code Retail Trade related businesses)	-0.955

TABLE 4: Model Validation Based on AAPD

Trips	Observed Count	Expected Count													
		Poisson		NB		ZIP		HP		HNB		ZINB		ZINB Mixed Effect	
		Count	**APD (%)**	**Count**	**APD (%)**	**Count**	**APD (%)**	**Count**	**APD (%)**	**Count**	**APD (%)**	**Count**	**APD (%)**	**Count**	**APD (%)**
0	36,250	34,743	4.15	25,625	29.31	34,991	3.47	35,158	3.01	33,378	7.92	33,221	8.35	34,829	3.92
1-10	511	143	72.01	219	57.14	163	68.10	132	74.16	259	49.32	241	52.83	187	63.41
11-100	202	54	59.00	59	61	61	69.80	55	72.77	84	58.41	109	46.04	158	21.78
>100	83	24	71.08	0	100	10	87.95	20	75.90	4	93.38	4	95.18	24	71.08
AAPD (%)			52.13		64.31		57.33		56.46		52.26		50.60		40.05

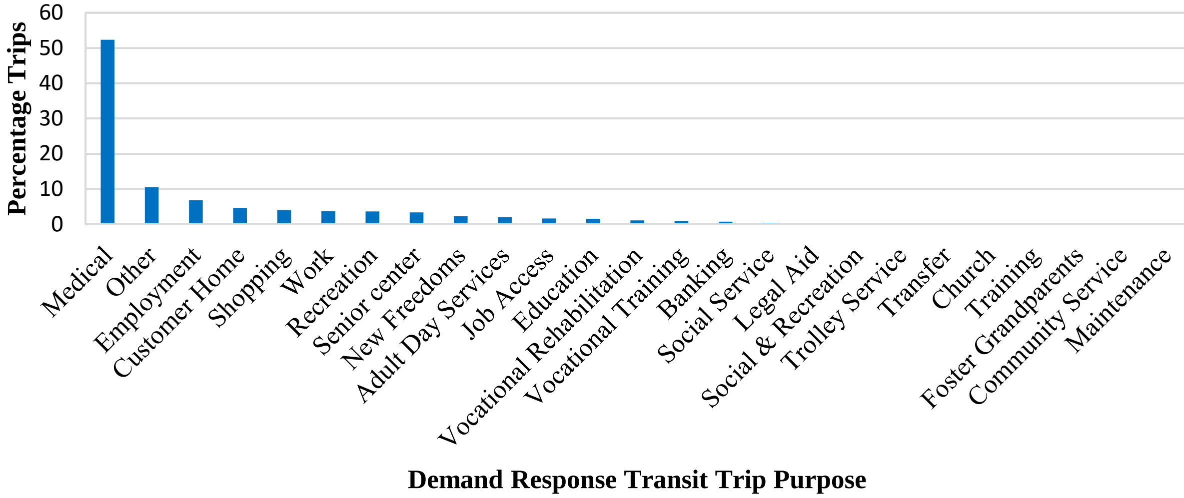

FIGURE 1: DRT trip purpose frequency

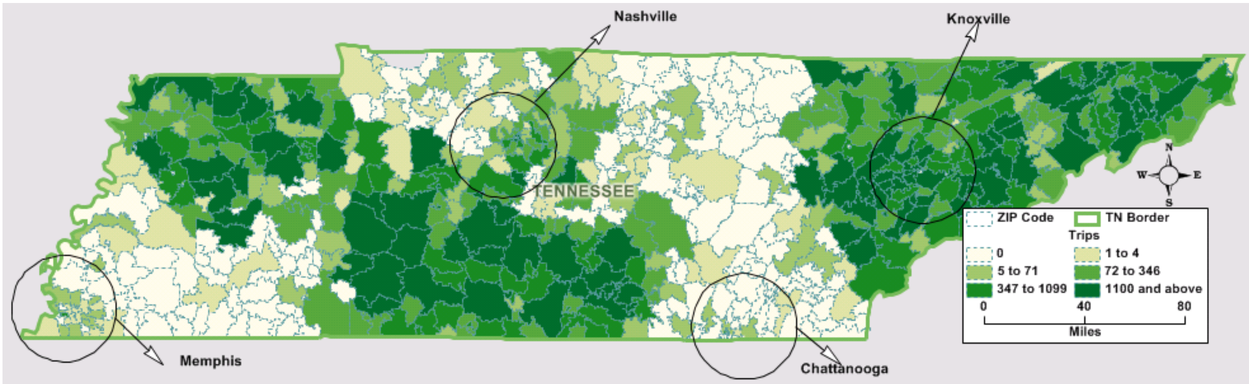

FIGURE 2: Trip production at ZIP code level

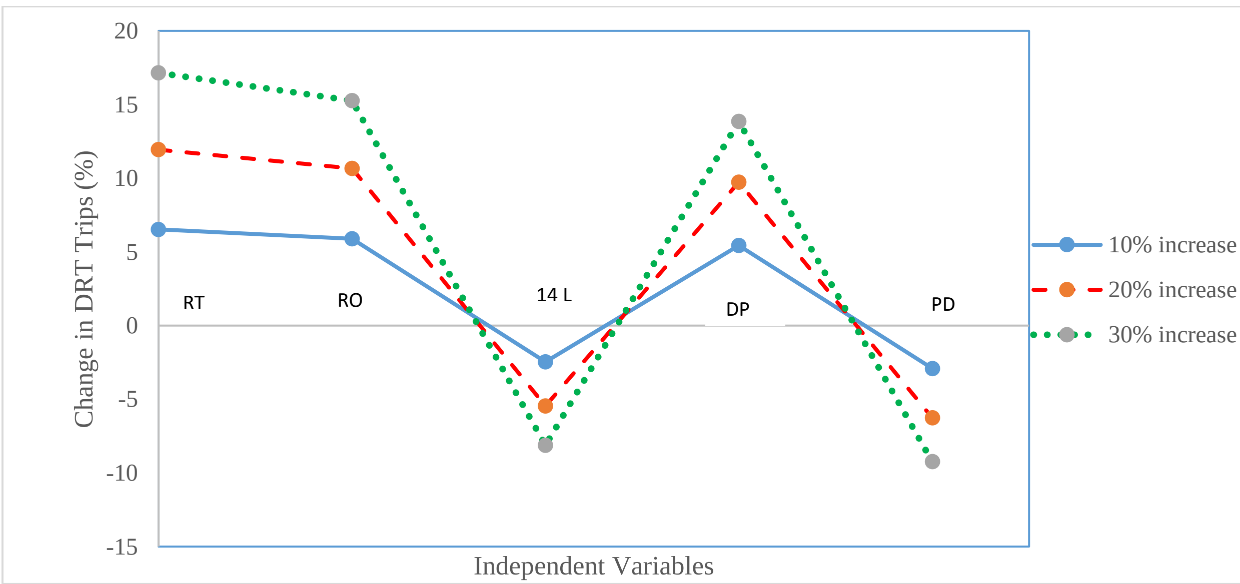

FIGURE 3: Changes in predicted DRT trips based on changes in independent variables

(Note: RT- Retail Trade; RO- Renter occupied; 14L- Population aged 14 years or less; DP- Disabled population and PD- Population density)