CrossMark

Measuring transit service impacts on vehicle ownership and use

Yangwen Liu · Cinzia Cirillo

Abstract In this paper we measure the impact of public transportation on household vehicle ownership and use. Advanced econometric models are estimated on household travel survey data and on geographic data. In particular, data from the 2009 US National Household Travel Survey is merged with geographic information obtained from the General Transit Feed Specification source. The integration of variables specific to the spatial and temporal coverage of the transit service allows the analysis of different policy scenarios. Results obtained for the Washington DC Metropolitan Area indicate that enhanced transit services reduce the number of private vehicles and vehicle miles traveled. Effects are more marked when bus services are improved and on car use. The study is important for all Metropolitan Regions that are dealing with the problem of congestion, high levels of greenhouse gas emissions and that are planning to invest in more efficient and accessible public transportation services.

Keywords Transit service index · Metropolitan area · GTFS · NHTS · Vehicle ownership · Discrete–continuous model

1 Introduction

An article published in the press by Addison (2010) shows that Americans scrapped 14 million cars in 2009, while they bought only 10.5 million new ones. The 2009 drop was the first large decline in vehicle ownership registered in the past 50 years.

Y. Liu · C. Cirillo (✉)
Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742, USA
e-mail: ccirillo@umd.edu

Y. Liu
e-mail: aliceliu@umd.edu

Although the recession probably played a major role, this decline might be also due to the introduction of smart growth policies and the consequent increase in urban density (Bhat and Guo 2007; Ewing and Cervero 2001, 2010), the adoption of employer commute and flex-work programs (Kingham et al. 2001), the expansion of car sharing (Cervero and Tsai 2004; Cervero et al. 2007), the introduction of the Car Allowance Rebate System (Gallup 1989), colloquially known as "Cash for Clunkers", and improved rail connectivity and inter-modality (Cullinane 1992; 2002)). Addison also reported that the increase in the use of public transit is one of the top ten reasons for the drop in car ownership especially in large metropolitan areas (Addison 2010; Cullinane 2002). In February 2013, President Barack Obama fleshed out plans to invest in public transportation and repair the nation's aging infrastructure. In fact, the administration has invested in more than 350 miles of new rail and bus rapid transit, 45,621 buses, and 5,545 railcars (American Public Transportation Association 2013).

The effects of transit service level on car ownership has been examined in a number of national studies in the US (Deka 2002; Kim and Kim 2004; Tal et al. 2010) UK (Goodwin 1993); Australia (Hensher 1998), Canada (Bunt and Joyce 1998), The Netherlands (Kitamura 1989), Germany (Bratzel 1999), and China (Li et al. 2010). More specifically, Kitamura (1989) investigated the causal relation between car ownership and transit use on data obtained from the 1984 Dutch National Mobility Panel survey. The results show that car use determines transit use, and that transit use does not determine car use. Nevertheless, the current situation is very different from the 80s, when the "car boom" was taking place, the number of households with access to one or more cars was limited, and the fuel price was relatively low. Arrington and Cervero (2008) studied the effect of transit-oriented development (TOD) in four US metropolitan areas (Bay Area, Charlotte, North Carolina, and St. Louis, Missouri), and found that TOD housing produced considerably less traffic than what is generated by conventional development. Arrington and Sloop (2010) found that in the Washington DC metropolitan area, among the five mid- to high-rise apartment projects near Metrorail stations outside Washington, DC, vehicle trip generation rates were more than 60 percent below that predicted by the ITE report. Bunt and Joyce (1998) conducted a household survey to test the effectiveness of Vancouver's SkyTrain and its effect on car ownership patterns near the rapid transit stations. Statistics from the survey show that the average car ownership is much lower for households located near SkyTrain stations. Cullinane (2002) found that good public transport can deter car ownership based on an attitudinal survey in Hong Kong, where public transport is plentiful and cheap and car use is low. Deka (2002) applied regression models to examine the relationship between transit availability and auto ownership with travel survey data from Los Angeles. The conclusion is that significant improvements will be needed in transit services to bring a slight decrease in auto ownership among the general population. Kim and Kim (2004) developed econometric models to predict the effect of accessibility to public transit on automobile ownership and miles driven. Important findings in their analysis are: (1) the number of licensed drivers is the primary determinant of the number of automobiles owned, (2) the presence of children is not a significant factor in automobile ownership and Vehicle Miles

Traveled (VMT), and (3) VMT is affected more by transit in multi-vehicle households than in one-vehicle households. Haas et al. (2010) examine real-world potential to use transit and transit-oriented development as an emissions reduction strategy in three different future development scenarios for the Chicago metropolitan area.

In conclusion, recent studies provide evidence that good public transportation might encourage people to reduce vehicle ownership and use. However, very few studies use advanced quantitative methods to investigate the relationship between public transit service and vehicle ownership and use. Other difficulties include collecting geographic data and quantifying the transit service level. Moreover, many metropolitan areas are interested in improving public transportation in order to reduce traffic congestion and in providing more efficient transportation systems (Arrington and Cervero 2008; Arrington and Sloop 2010; WMATA 2012a; Rood 1998). Therefore, it is crucial to explore the impact of public transportation on vehicle ownership and use with advanced methods and accurate data based on geographic information systems.

Different measurements for transit level of service are found in the literature. The Local Index of Transit Availability (LITA) (Rood (1998)) measures the transit service intensity of an area with transit data and census data (demographic information). Depending on the data availability, LITA scores can be computed for any area unit. Transit Capacity and Quality of Service Manual (TCQSM) (TRB 2003) also uses transit data and census data but incorporates a service coverage measure to assess transit accessibility. TCQSM offers a comprehensive guide for infrastructure enhancements specific to public transportation systems. The Time-of-Day Tool (Polzin et al. (2002)) provides the relative value of transit service accessibility for each time period and requires data on temporal distribution of travel demand in addition to transit and census data.

In this paper, we aim to investigate the effects of improved public transportation services on household vehicle ownership and use. Specifically, we apply the method proposed by (Liu et al. (2014)), which simultaneously predicts the vehicle ownership and VMT. The case study is conducted for the Washington DC. Metropolitan Area, which is a mix of urban and suburban areas with a relatively good public transportation system for which further improvements are foreseen. The information used for model estimation was obtained from different sources. The 2009 NHTS data with geographic reference (US Census Tract level) was kindly provided by the Federal Highway Administration (FHWA), US DOT, while the General Transit Feed Specification data was obtained from WMATA.

We adopt an approach similar to the one proposed Keller (2012) to measure transit service. The method also follows the TCQSM manual and takes into account both spatial and temporal characteristics of the transit system. The indicator is based on the percent service coverage area, the average service headway and the service duration.

The remainder of this paper is organized as follows: the next section provides details on the data sources used for this study. Section 3 describes the data process which makes use of geographic information systems and data mining techniques. The estimation techniques based on discrete–continuous models, adopted for

household vehicle ownership and use, is explained in Sect. 4. Section 5 presents estimation results and policy analyses. Finally, Sect. 6 summarizes the main findings and gives perspectives for future research.

2 Data description and sources

2.1 2009 National Household Travel Survey Data

The primary data source used in this study is the 2009 NHTS (FHWA 2009). The analysis is restricted to the area of Washington DC. Metropolitan Area, for which 1,420 complete observations are available. Household social economic characteristics and information on each household vehicle, including year, make, model, and estimates of annual miles traveled, are the main variables extracted from the original dataset. Most importantly, the household location is geographically referenced on the US Census Tract level, which enables us to measure the transit service quality around each household's location in the sample.

Table 1 lists the basic statistics relative to the household sample. For the Washington D.C. Metropolitan area, the average vehicle ownership per household is 1.87, which is lower than the national average of 2.08 cars per household. The percentage of households without a car is 7.28 %, higher than the national average of 4.8 %. Average household income increases for households having up to two cars, but remains stable for households with 3 or 3+ cars. The number of cars per household is highly associated with the number of adults and number of drivers in the family. More than half of the households who do not have a car do not own a house. The land use variables, such as urban size, population density and housing density, greatly influence the household car ownership decisions. The households with more cars are generally located in more rural areas. In the Washington DC. Metropolitan area, the average age of the household head is around 55 years; households with zero or one car have older household heads. The average education level in this area is about college or Bachelor's degree, while households without a car have much lower education level. The average annual mileage traveled by a household is more than 20,000 miles per year. The mileage traveled increases accordingly with the number of cars in the household.

2.2 General transit feed specification data

The GTFS (https://developers.google.com/transit/gtfs/), which was originally developed by Google and Portland TriMet, defines a common data format for public transportation schedules and the associated geographic information. The GTFS is an open format and it is composed of a series of text files; each file contains a particular aspect of the transit service: stops, routes, trips and other schedule data.

The GTFS data for the Washington D.C. Metropolitan area is obtained from the WMATA. The database consists of the following files:

- Agency—contains the transit agency id, name and website.

Table 1 Data statistics

	Number of cars					Statistics				
	0	1	2	3	4	Min	Max	Median	Mean	sd
Car ownership	7.28 %	26.72 %	43.49 %	17.03 %	5.48 %		4	2	1.87	0.96
Household income level	6.47	10.98	14.55	15.29	15.99	1	18	16	13.21	5.30
Number of adults	1.31	1.40	1.99	2.21	2.76	1	5	2	1.86	0.66
Number of workers	0.53	0.69	1.16	1.44	1.63	0	4	1	1.06	0.83
Number of drivers	0.75	1.26	1.98	2.32	2.83	0	5	2	1.80	0.77
Owned house	0.45	0.75	0.91	0.97	0.98	0	1	1	0.85	0.36
Urban size	4.94	4.25	3.60	2.95	2.55	1	6	5	3.70	2.29
Age of household head	59.13	60.51	53.12	52.16	53.00	18	95	54	55.36	14.91
Female household head	0.78	0.64	0.52	0.55	0.44	0	1	1	0.57	0.49
Education household head	2.78	3.40	3.67	3.52	3.33	1	5	4	3.49	1.21
Housing units per square mile	8,657	4,187	1,525	858	632	50	30,000	750	2,593	4,740
Population per square mile	13489	7652	3648	2327	1941	50	30,000	3,000	5,116	6,820
Workers per square mile	2870	1899	1042	640	453	25	5,000	350	1,303	1,660
Annual miles traveled	0	10361	23890	35781	48728	0	20,864	19,173	21,922	17,986

sd standard deviation

Fig. 1 Data Structure of GTFS
Data
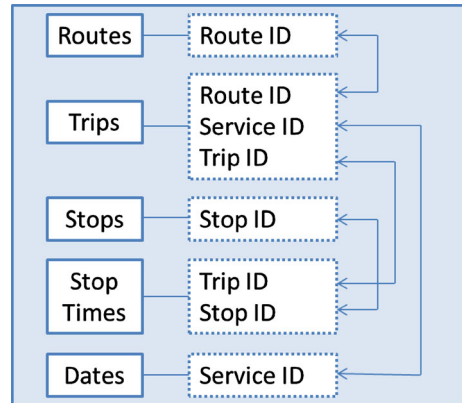
- Stops—individual locations where vehicles pick up or drop off passengers. The data contains information on stop id, stop name, latitude and longitude and stop location.
- Transit Routes—a route is a group of trips that are displayed to riders as a single service. The data contains information of route id, route name, route type (i.e., subway, rail and bus), etc.
- Trips for each route—a trip is a sequence of two or more stops that occurs at a specific time. The data contain information on the trip id, trip name, trip head sign, and the corresponding route id and service id.
- Stop times—times that a vehicle arrives at and departs from individual stops for each trip.
- Calendar dates—specify when service starts and ends, as well as days of the week when the service is available. The data contains information on the service id and service dates.
- Shapes—rules for drawing lines on a map to represent a transit organization's routes.

The data structure of GTFS is presented in Fig. 1.

2.3 The 2010 census TIGER/line shapefiles

The US Census TIGER/Line shapefiles contain the geographic extent and boundaries of both legal and statistical entities (http://www.census.gov/geo/maps-data/data/tiger-line.html). The 2010 data on Census Tract level is obtained for the State of Maryland, Virginia and District of Columbia because the main data source (NHTS data) was collected in 2009 and was geo-referenced on Census Tract level.

2.4 Vehicle characteristics

Vehicle characteristics, including the information on vehicle price (sale price or price of each new or used), fuel efficiency, seating, engine, etc., are obtained from a

secondary data source, the Consumer Reports (CR). The CR provides vehicle specification data on models tested within the past 10 years starting from 2009, and having up to four models per year.

3 Data geo-processing and data integration

3.1 Spatial measurements of transit service

This paper follows the TCQSM manual recommendations to calibrate the coverage of public transportation services. In particular, a service buffer is created for each area surrounding a station to derive the area of usage for potential transit users. The TCQSM (TRB 2003) suggests a 0.25-mile buffer around bus stops and a 0.5-mile buffer around rail stations. These buffers are based on willingness to travel studies; buffers based on these distance ranges tend to represent between 75 and 80 percent of all walking trips to a transit stop.

The GTFS data is firstly converted from.txt files to shapefiles for both transit stations and routes, and then projected in ArcGIS along with Census TIGER files. The buffer zones for the bus stops with a 0.25-mile radius and metro routes with a 0.5-mile radius are then created. The overlapping buffers are dissolved to eliminate double counting. The coverage area is joined to the census tract zone and the percentage of coverage is computed:

$$Percentage\ of\ Coverage = Area\ of\ the\ Buffer\ Zone\ within\ the\ Census\ Tract/$$
$$Area\ of\ the\ Census\ Tract$$

The process is repeated for each stop/route and for each census tract zone.

The final variables that are produced in this process include (1) percentage of bus stops coverage, (2) percentage of metro routes coverage, (3) total length of bus routes, (4) total length of metro routes, and (5) total number of bus stops. All the variables above are calibrated for each census tract in the Washington D.C. Metropolitan area.

3.2 Temporal measurements of bus service

The data related to transit timetable in the GTFS files is utilized to calibrate the temporal measurements of bus services. Firstly, the GTFS files are merged together with the key IDs (see Fig. 1). The merged data has information on the bus arriving time for each route and stop for an entire day (24 h). Then for each stop and for each route, the bus service duration and average headway is computed with data mining techniques. Finally, the average duration and headway are aggregated for each census tract zone.

3.3 Transit service index (TSI)

The indicator proposed to measure transit service is based on the percent service coverage area, the average service headway and the service duration. This approach

Table 2 Sample calibration of TSI

Zone ID	% coverage	Service headway (h)	Service duration (h)	TSI
11001009204	100 %	0.44	7.83	17.96
24033801309	6.68 %	0.34	4.03	0.80
51177020104	0	0	0	0
11001001402	49.55 %	0.38	14.69	19.19

is similar to the one proposed by Keller (2012) and at the same time follows the TCQSM manual by taking into account both spatial and temporal characteristics of the transit system.

It is calculated with the percent service coverage area, the average service headway and the service duration. For each census tract zone we have:

$$TSI = \frac{percent\ service\ coverage\ area}{average\ service\ headway} \times service\ duration$$

Table 2 presents some examples of TSI calibration from real data. In this paper, TSI is calculated for the bus service only. The reason for not including metro service is because the time schedule of metro subways in the DC area is comparatively rigid and does not create variation among different census tract zones. Instead, the percent service coverage area is created as the measurement of transit service.

3.4 Data integration and final database

The final database consists of three components: NHTS data, GIS output and vehicle characteristics. As shown in Fig. 2, the data sets are linked by key ID. Specifically, the 2009 NHTS data includes household socio-economic information, such as household income, household size, number of drivers, number of workers, and land use characteristics around the household location, such as the residential density of the census tract zone, the urbanization level, etc. The GIS output includes data on bus stop coverage percentage, metro route coverage percentage, total length of bus routes, total length of metro routes, total number of bus stops, transit service index, average bus headway, and average bus service duration for each census tract zone. The vehicle characteristic data includes purchase price, operating cost, fuel economy, seating, performance, and other specifications for each vehicle type.

4 The integrated discrete–continuous model

This paper aims to analyze the impact of transit service on household vehicle ownership decisions and predicts the vehicle ownership and VMT simultaneously. This is a typical discrete–continuous problem; here we make use of the integrated discrete–continuous model proposed by Liu et al. (2014) in order to estimate these joint decisions. This section presents the methodology and model development.
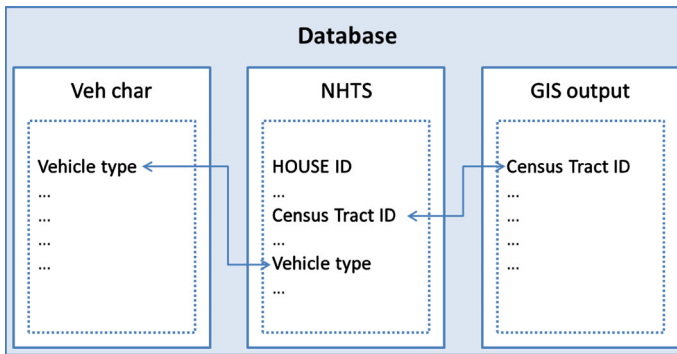
Fig. 2 Data structure of the final database

The discrete problem concerns the forecast of the number of vehicles in a household (Y) using a set of predictors. Suppose there are $k + 1$ (0, 1, ..., k) vehicle ownership levels, the utility for each level consists of one observed part (systematic utility) and one unobserved part (error term):

$$U_0 = \epsilon_0$$
$$U_1 = X_1^T \beta_1 + \epsilon_1$$
$$U_2 = X_2^T \beta_2 + \epsilon_2$$
$$\ldots$$
$$U_k = X_k^T \beta_k + \epsilon_k$$

where, U_k is the utility of having k vehicles; X are the explanatory variables associated with the household, the vehicles, and the land use; β's are the corresponding parameters to be estimated and εs are the error terms.

In the unordered structure, the household is assumed to be rational and to choose the alternative of vehicle ownership level that maximizes its utility. In this case, we adopt a multinomial probit model for the vehicle holding decisions and, therefore, the error terms follow a multivariate normal distribution with full, unrestricted covariance matrix. The likelihood function can be expressed as follows:

$$P(Y = y | X, \beta, \Sigma) = \int_{\mathbb{R}^{k+1}} \mathbb{I}\left(X_y^T \beta_y + \epsilon_y > X_j^T \beta_j + \epsilon_j \ \ \forall j \neq y\right) \phi(\epsilon) d\epsilon$$

where, X is the $X_1, \ldots, X_k$, β is the $(\beta_1, \ldots \beta_k)$, ε is the $(\varepsilon_0, \ldots, \varepsilon_0)$, $\phi(\varepsilon)$ is the density of the normal distribution, Σ Covariance of error term.

The functional indicator $[\mathbb{I}(\cdot)]$ ensures that the observed choice is indeed the one with the biggest utility. The subscript y indicates the predictors and coefficients of the chosen alternative and the subscript j indicate the other alternatives.

Since only differences in utility matter, the choice probability can be equivalently expressed as (k)—dimensional integrals over the differences between the errors.

Suppose we differentiate against alternative y, the alternative for which we are calculating the probability. Define:

$$\tilde{\epsilon}_{jy} = \epsilon_j - \epsilon_y$$
$$\tilde{V}_{jy} = X_j^T \beta_j - X_y^T \beta_y$$
$$\tilde{\epsilon}_y = \tilde{\epsilon}_{1y}, \ldots, \tilde{\epsilon}_{ky}$$

Then

$$P(Y = y | X, \beta, \Sigma) = \int_{\mathbb{R}^k} \mathbb{I}(\tilde{V}_{jy} + \tilde{\epsilon}_{jy} < 0 \quad \forall j \neq y) \phi(\tilde{\epsilon}_y) d\tilde{\epsilon}_y$$

which is a (k)-dimensional integral over all possible values of the error differences. The probit has been normalized using the procedure proposed by Train (2009) to ensure that all parameters are identified. For more details on the normalization in the context of discrete–continuous models we refer to Liu et al. (2014).

Regression is adopted to model the continuous part of the model: the decision on the household vehicle mileage. In a regression, the dependent variable Y_{reg} is assumed to be a linear combination of a vector of predictors X_{reg} plus some error term:

$$Y_{reg} = X_{reg}^T \beta_{reg} + \epsilon_{reg} \quad \epsilon_{reg} \sim N(0, \sigma_{reg}^2)$$

Given β_{reg}, X_{reg} and σ^2, the likelihood of observing Y_{reg} is given by the normal density function:

$$P\left(y_{reg} | \beta_{reg}, X_{reg}^T, \sigma_{reg}^2\right) = \phi\left(y_{reg} | X_{reg}^T \beta_{reg}, \sigma_{reg}^2\right)$$

In order to jointly capture the correlation between the discrete and continuous parts, we allow the error term of the regression to be correlated with the error terms of the utilities in the probit. Therefore, the specifications of the observable part of the utilities and of the regression remain the same, but the error terms follow an "incremental" normal distribution:

$$\left(\tilde{\epsilon}_{1y}, \ldots, \tilde{\epsilon}_{ky}, \epsilon_{reg}\right) \sim MN(0, \Sigma_{k+1})$$

.

In another expression:

$$\begin{bmatrix} \tilde{\epsilon}_y \\ \epsilon_{reg} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \Sigma_y & \Sigma_{y,reg} \\ \Sigma_{reg,y} & \sigma_{reg}^2 \end{bmatrix}\right)$$

The probability of observing Y and Y_{reg} is the product of the probability of observing Y_{reg} ($P(Y_{reg})$) and the probability of observing Y given Y_{reg} ($P(Y|Y_{reg})$).

$$P\left(Y, Y_{reg}\right) = P\left(Y_{reg}\right) P\left(Y | Y_{reg}\right)$$

The conditional probability of probit is:

$$P\big(Y|Y_{reg}\big) = \int_{\mathbb{R}^k} \mathbb{I}\big(\tilde{V}_{jy} + \tilde{\epsilon}_{jy} < 0 \ \forall j \neq y\big)\varphi\big(\tilde{\epsilon}_y\big)d\tilde{\epsilon}_y$$

where $\varphi\big(\tilde{\epsilon}_y\big)$ is the density function of a multivariate distribution and

$$\tilde{\epsilon}_y \sim \mathcal{N}\left(0 + \frac{\Sigma_{y,reg}}{\sigma_{reg}^2}\big(\epsilon_{reg} - 0\big), \Sigma_y - \frac{\Sigma_{y,reg}\Sigma_{reg,y}}{\sigma_{reg}^2}\right)$$

The integral has no closed form so we rely on simulation. The final simulated log-likelihood of the unordered discrete–continuous model is given by the following formula:

$$SLL\big(\beta, \beta_{reg}, \Sigma|Y, Y_{reg}, X, X_{reg}\big)$$
$$= \sum_{i=1}^{n} \log\left(\frac{1}{B}\left(\sum_{b=1}^{B} \mathbb{I}\big(\tilde{V}_{jy} + \tilde{\epsilon}_{jy}^{(b)} < 0 \ \forall j \neq y\big)\right) \times \phi(y_{i,reg}|X_{i,reg}^T\beta_{reg}, \sigma_{reg}^2)\right)$$

where $\tilde{\epsilon}_{jy}^{(b)}$ is a draw from a multivariate normal with mean $0 + \frac{\Sigma_{y,reg}}{\sigma_{reg}^2}\big(\epsilon_{reg} - 0\big)$ and variance $\Sigma_y - \frac{\Sigma_{y,reg}\Sigma_{reg,y}}{\sigma_{reg}^2}$. B is the number of draws in the ith probit simulation.

In this paper, simulations have been executed using 1000 pseudo Monte Carlo draws, while standard errors have been computed using Bootstrap re-sampling techniques. The model has been calibrated using code developed in R by the authors.

5 Empirical results

5.1 Estimation results

Table 3 presents the parameter estimates of the joint vehicle ownership and usage model. The model includes a logsum variable derived from the vehicle type and vintage model in Table 4. The logsum represents a feedback variable from the class/vintage models and reflects the interdependence of the household choice of how many vehicles to own with its choice of class and vintage for each car in the household (Train 1986). Variables that enter the vehicle type and vintage model are vehicle characteristics and household socio economics; in particular, vehicle price has been interacted with household income to capture non-linear effects. Results attest that increasing vehicle prices are always perceived as a disutility but higher income households are less sensitive to price than those in the bottom of the income scale. Larger cars and those with more luggage space are still preferred in the US. Households tend to own vehicle types for which more makes and models are available. The vehicle's miles per gallon (MPG) is not significant, while large differences in owned cars' MPG are found to have a positive effect on the utility of owning multiple cars.

Table 3 Estimation Results

Variable	Coefficient	Std. err
Dependent variable: number of cars		
Logsum (expected utility from vehicle type choice)	0.430	0.011
Alternative specific constant		
1 car	**−3.161**	**0.193**
2 cars	**−17.050**	**0.267**
3 cars	**−22.913**	**0.219**
4+ cars	**−27.934**	**0.178**
Household income level		
1 car	**−0.090**	**0.021**
2 cars	**0.446**	**0.053**
3 cars	**0.490**	**0.054**
4 + cars	**0.440**	**0.052**
Number of drivers		
1 car	−0.038	0.197
2 cars	**7.185**	**0.193**
3 cars	**7.982**	**0.193**
4+ cars	**7.791**	**0.183**
Gender of household head (female)		
1 car	0.089	0.199
2 cars	**−2.350**	**0.189**
3 cars	**−2.495**	**0.194**
4+ cars	**−2.637**	**0.159**
Urban size		
1 car	−0.049	0.048
2 cars	−0.071	0.115
3 cars	−0.153	0.113
4+ cars	−0.204	0.112
Residential density (census tract level)		
1 car	**0.051**	**0.014**
2 cars	**−0.524**	**0.135**
3 cars	**−0.704**	**0.150**
4+ cars	**−0.549**	**0.151**
TSI of bus		
1 car	**0.018**	**0.008**
2 cars	**−0.103**	**0.038**
3 cars	**−0.105**	**0.036**
4 + cars	**−0.116**	**0.039**
Percentage coverage of metro routes		
1 car	0.280	0.164
2 cars	**−2.212**	**0.267**
3 cars	**−1.756**	**0.296**
4+ cars	**−9.442**	**0.185**

Table 3 continued

Variable	Coefficient	Std. err
Dependent variable: miles (10 k)		
Constant	**1.470**	**0.121**
Household income level	**0.124**	**0.006**
Own home	**0.372**	**0.107**
Gender of household head (female)	−0.080	0.076
Residential density (census tract level)	**−0.055**	**0.012**
Driving cost ($ per mile)	**−4.823**	**0.294**
TSI of bus	**−0.025**	**0.004**
Percentage coverage of metro routes	**−0.324**	**0.159**
Log-likelihood at zero $LL\left(\hat{\beta}^0\right)$		−5880.231
Log-likelihood at convergence $LL\left(\hat{\beta}^1\right)$		−3260.811
Number of parameters		41
Number of observations		1420
Adjusted R^2		0.44

Variables that are significant at 95 % level or above are bolded

See Appendix for the models in the logsum calibration

The variables TSI, created to represent bus and metro coverage percentage, are significant and have a negative impact on household vehicle ownership and miles traveled. The variable <TSI of bus> is selected instead of other measures because it gives a more comprehensive representation with respect to both spatial and temporal bus service information. Metropolitan subways have time schedules that are comparatively rigid and for this reason we have decided to measure metro service using the percentage coverage only.

With good accessibility to bus and metro services, households tend to own fewer cars. The magnitudes of the coefficients increase with the number of cars owned by the household, indicating that the transit service level has greater impacts on multi-vehicle households. In particular, the coefficient of metro service coverage for 4-car households is significantly greater than the one obtained for other alternatives. However, caution should be taken when drawing these conclusions as households which use their car(s) less might have a tendency to reside close to public transportation facilities (self selection bias).

Coefficients of household income are positive and significant; the value of the coefficients is larger for households owning more cars. Households with higher income tend to own multiple cars and drive more, and the higher their income, the more likely that they will own more cars. Households with owned house are more likely to have higher mileage on their vehicles.

Households with more drivers own more vehicles and drive more often. The coefficients related to the number of drivers are significant except in the one-car alternative.

Table 4 Vehicle type models to calibrate the logsum in the discrete–continuous model

Variable	One-car households		Two-car households		Three-car households		Four-car households	
	Estimates	t statistic	Estimates	t statistic	Estimates	t statistic	Estimates	t statistic
Sum of shoulder room	0.0329	5.6	0.0483	19.9	0.0471	18.4	0.0678	9.5
Sum of luggage space	0.2162	7.3	0.1563	5.8	–	–	0.1032	4.7
Log (number of make/model in the class)	1.3820	28.4	0.8015	40.1	0.8241	38.5	1.182	16.3
Auto MPG	0.0026	0.3	–	–	–	–	–	–
Difference max MPG and min MPG	–	–	0.0158	3.0	0.0246	3.1	0.0220	1.1
Dummy at least one car is a foreign car	–0.4959	–7.6	–	–	–	–	–0.3361	–2.4
Dummy both foreign cars	–	–	1.0310	14.0	–	–	–	–
Dummy one old car	0.4515	8.3	0.4133	10.2	1.094	10.9	1.144	3.5
Dummy at least one SUV (Hhld 3 members)	2.0870	8.2	0.2794	3.0	–	–	–0.7771	–3.6
Dummy at least one pickup (Hhld 3 members)	–1.332	–1.3	0.2930	3.0	0.6034	5.1	0.5366	2.6
Dummy at least one van (Hhld 3 members)	–	–	1.1520	12.7	0.3519	3.0	–0.3945	–1.9
Dummy auto (Hhld 3 members)	1.1	4.7	–	–	–	–	–	–
Dummy both auto	–	–	–1.3330	–4.7	–	–	–	–
Dummy at least one sport car	–	–	–	–	–	–	0.4704	2.1
Purchase price (Hhld income <20 k)	–0.2037	–15.2	–	–	–	–	–	–
Purchase price (Hhld income = 20 – 40 k)	–0.0680	–8.5	–	–	–	–	–	–
Purchase price (Hhld income > 40 k)	–0.0069	–1.2	–	–	–	–	–	–
Purchase price (Hhld income < 45 k)	–		–0.1570	–28.2	–	–	–	–
Purchase price (Hhld income = 45 – 80 k)	–		–0.0730	–17.8	–	–	–	–
Purchase price (Hhld income > 80 k)	–		–0.0457	–7.7	–	–	–	–
Purchase price (Hhld income < 55 k)	–		–	–	–0.1688	–22.8	–	–
Purchase price (Hhld income = 55 – 100 k)	–		–	–	–0.0953	–18.1	–	–
Purchase price (Hhld income > 100 k)	–		–	–	–0.0289	–6.1	–	–

Table 4 continued

Variable	One-car households		Two-car households		Three-car households		Four-car households	
	Estimates	t statistic	Estimates	t statistic	Estimates	t statistic	Estimates	t statistic
Purchase price (Hhld income < 60 k)	–	–	–	–	–	–	-0.1841	-12.5
Purchase price (Hhld income = 60 – 100 k)	–	–	–	–	–	–	-0.1175	-11.2
Purchase price (Hhld income > 100 k)	–	–	–	–	–	–	-0.0678	-8.1
Number of observations	2995		4525		2015		545	
Log-likelihood at zero	-9118.34		-13776.46		-6134.71		-1659.26	
Log-likelihood at convergence	-7859.95		-10570.75		-3755.26		-788.66	
Adjusted R²	0.138		0.237		0.388		0.525	

Hhld household

In terms of the characteristics of the household head, the dummy variable "female household head" is significant except for the one-car household; the negative sign means that households with a female head tend to own fewer cars and drive less.

The coefficients of residential density are significant and negative (except for the one-car household), inferring that the households located in a more dense area have lower probability of owning more cars and of driving less.

The parameter of driving cost is negative and significant, indicating that higher operational cost induces households to drive less.

In addition to the coefficients of the variables, the covariance matrix between the discrete and the continuous independent variables is estimated and reported in Table 5. In particular, the bottom line of the matrix explains the correlation between the mileage travelled and the utility differences (with respect to the zero-car alternative) of the vehicle ownership alternatives. The positive numbers mean that higher mileage usage increases the utility of owning more cars; the magnitude of the correlation factors increases with the number of household vehicles. The negative value found for the correlation across mileage and zero-car alternative can be explained by the fact that zero miles or very low mileages further decrease the difference in utility of owning a car or not owning a car.

In order to investigate the significant role of the transit service attributes, the car ownership model is re-estimated without transit-related variables. A log-likelihood ratio test is conducted to test the significance of transit service variables in the vehicle ownership model:

H_0: Coefficients of transit service variables are not zero (full model).
H_1: Coefficients of transit variables are zero (reduced model).
Degree of freedom (DOF) = 10.

$$-2\left[LL\left(\hat{\beta}^1\right) - LL\left(\hat{\beta}^0\right)\right]$$
$$= -2[(-3349.812) - (-3260.811)]$$
$$= 178.002 > \chi^2_{10,0.05} = 25.188$$

The test statistic is much larger than the Chi square with $10°$ of freedom at the 95 percent confidence level. Therefore, we reject the hypothesis that the coefficients of transit service variables are zero and we conclude that the model could not be reduced. The testing result confirms again the significant role of transit service variables in vehicle ownership models.

Table 5 Covariance matrix

$$\hat{\Sigma} = \begin{pmatrix} 2.00 & -7.51 & -7.43 & -7.54 & -0.68 \\ -7.51 & 29.22 & 30.68 & 30.41 & 2.92 \\ -7.43 & 30.68 & 35.04 & 33.67 & 3.55 \\ -7.54 & 30.41 & 33.67 & 32.72 & 3.34 \\ -0.68 & 2.92 & 3.55 & 3.34 & 1.23 \end{pmatrix}$$

The integrated Discrete-Choice model has also been compared to a simpler model structure that includes a multinomial logit for the discrete part and a regression for the continuous part. We obtained the following values of the log-likelihood:

Log-likelihood (logit) $= -1244.062$.
Log-likelihood (regression) $= -2459.005$.

which gives a total value of the *Log-likelihood* $= -3703.067$ for the combined decisions.

This value is much lower than the value obtained with the DC model (*Log-likelihood* $= -3260.811$). We can, therefore, conclude that our integrated model system provides a much better goodness-of-fit when compared to established modeling techniques.

5.2 Policy analysis

The Washington Metropolitan area is developing a 30-year transit plan WMATA (2012a), which aims to provide a long term vision for future growth and to improve and expand transit service. The goal of the regional plan is to seek solutions such as making pedestrian and rail connections between lines, to bypass bottlenecks, to add new rail lines through the downtown core and to improve surface transit.

In 2012 WMATA (2012a) has announced an investment of $5 million to provide customers with better bus service. The project has now been completed and a limited-stop MetroExtra route is in operation; the new transit system offers a more frequent service, additional capacity, and expanded hours of operation.

On the other hand, the metrorail ridership is expected to top 1 million daily rides by 2040 and the system's core will be severely crowded (Johnson 2011). WMATA has been looking at long-term strategies for expanding transit. The Purple Line (WMATA 2012b), which is a 16-mile transit line that will connect the Red, Green and Orange lines of the metro system in the suburban area of Maryland, can be seen as part of the long term plan. Meanwhile, a Beltway Metro Line is under consideration.

Given the numerous investments foreseen for the public transportation system in the Washington DC Metropolitan area, it is worth to examine the impacts of improved transit services on household vehicle ownership and usage. In this paper, the model estimated in Sect. 4.1 is applied to evaluate different policy scenarios.

We first analyze the effects of improved bus services; in this hypothetical scenario every census tract zone has at least 50 % bus stop coverage (50 % of the census tract area has less than 0.25-mile walking distance to a bus stop), 15 min average headway and 6 peak hours duration (6:30–9:30 AM and 3:30–6:30 PM). In the improved metrorail service scenario, the core area of Washington Metropolitan area (urban size greater than 1 million) has at least 50 percent metro route coverage (50 % of the census tract area has less than 0.5-mile walking distance to a metro line).

The application results are presented in Table 6. The short-run impacts of improved transit service generally reduce both vehicle ownership and miles

Table 6 Policy analysis based on different improvement of the transit service

	Current	Improved bus service		Improved metro service	
		Predicted	%	Predicted	% change
0-car household	7.16 %	7.17 %	0.01	7.17 %	0.01
1-car household	23.06 %	26.05 %	2.99	24.60 %	1.55
2-car household	46.56 %	44.32 %	−2.25	44.84 %	−1.72
3-car household	17.82 %	17.19 %	−0.64	19.44 %	1.61
4-car household	5.40 %	5.28 %	−0.12	3.95 %	−1.45
Average vehicle ownership	1.91	1.87	−2.03	1.88	−1.49
Mileage	22231.70	20410.40	−8.19	21879.50	−1.58

traveled. The average vehicle ownership is reduced by 2 percent in the improved bus service scenario and 1.5 percent in the improved metro service scenario. The annual mileage traveled decreases by about 8 percent with improved bus service and 1.6 percent with improved metro service. Comparatively, the improved bus service has greater impacts on reducing both the vehicle ownership and the mileage traveled.

It should be noted here that the NHTS data has limited number of households in the DC and Maryland area due to the fact that neither of these regions are in the NHTS add-on program. The predictions provided could be more accurate with an increased number of observations available for model calibration.

6 Conclusions

The Washington Metropolitan area is a diverse region with both dense urban areas and suburban areas. This region is also served by a good public transportation system that will undergo several improvement plans in the short and long term. Given the increasing interest in transit investments from both federal and state governments, as well as the traffic concerns on the Beltway, it is important to understand and quantify the relation between public transportation service and household vehicle ownership and usage. In particular, this paper has analyzed the impact of improved bus and metro services on household ownership/use decisions in the Washington Metropolitan area.

This study proposes a methodology to integrate the household travel survey with geographic data. Specifically, the main data sources are the 2009 NHTS and the GTFS. Secondary data includes the 2010 Census TIGER shapefiles and vehicle characteristics from the *Consumer Reports*.

Both spatial and temporal measurements of transit service are created based on the GTFS data and geographic information data using data mining techniques. Transit service index is calculated with these measurements and then integrated together with the NHTS data, the GIS output data and the vehicle characteristics into one database referenced at the census tract level.

This paper jointly estimates the household decisions on vehicle ownership and usage; estimates are obtained for household social-demographic attributes, land-use characteristics, vehicle characteristics and transit service variables. The integrated model system provides much better goodness-of-fit statistics when compared to established modeling techniques. The model is then applied to policy scenarios that accounts for transit investments. The results obtained show that transit service generally reduces both vehicle ownership and miles traveled. The average vehicle ownership is reduced by 1.5–2.0 percent and the mileage decreases by about 1.6–8.0 percent, respectively, with improved bus service and with improved metro service.

The study conducted has a number of limitations that could be the object of future studies. The dataset considered has limited observations in the DC and Maryland regions given that they are not part of the 2009 NHTS add-on program. The model assumes linearity for income and density; this assumption can be relaxed if a larger sample will be available. The model predicts the total household mileage but it is not able to calculate the share for each car in the household. This modeling feature is particularly important when studying the use of more efficient cars but it is less relevant in this context and can be accommodated by allowing a more complex model structure. Furthermore, the effects of self selection bias on model estimates should be tested. The causality between public transport accessibility and owning/ using cars could be overestimated assuming that households which use less the car might have chosen to live in a neighborhood with better bus and metro services. Different indicators for public transportation coverage and quality of service could be created and their effects on car ownership compared. Finally, it would be interesting to apply the methodology proposed on geographical areas characterized by different density levels and served by a greater variety of public transportation services.

References

Addison J (2010) Ten reasons for drop in car ownership. Clean Fleet Report. (See http://www. cleanfleetreport.com/clean-fleet-articles/car-ownership-declines/)

American Public Transportation Association (2013) Obama proposes investments for public transportation. (See http://newsmanager.commpartners.com/aptapt/issues/2013-02-22/index.html)

Arrington GB, Cervero R (2008) Effects of TOD on housing, parking, and travel. Transit Cooperative Research Program, Report 128.) (See http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_128.pdf.)

Arrington GB, Sloop KI (2010) New transit cooperative research program research confirms transit-oriented developments produce fewer auto trips. ITE J 79(6):26–29

Bhat CR, Guo J (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transp Res Part B: Methodol 41(5):506–526

Bratzel S (1999) Conditions of success in sustainable urban transport policy— change in 'relatively successful' European cities. Transp Rev 19:177–190

Bunt P, Joyce P (1998) Car ownership patterns near rapid transit stations. Canadian Institute of Transportation Engineers. (See http://www.citebc.ca/Feb98_Ownership.html)

Cervero R, Tsai Y (2004) San Francisco City Car Share: second year travel demand and car ownership impacts. Transp Res Rec 1887:117–127

Cervero R, Golub A, Nee B (2007) City CarShare: longer-term travel demand and car ownership impacts. Transp Res Rec 1992:70–80

Cullinane S (1992) Attitudes towards the car in the UK: some implications for policies on congestion and the environment. Transp Res Part A 26(A):291–301

Cullinane S (2002) The Relationship between car ownership and public transport provision: a case study of Hong Kong. Transp Policy 9:29–39

Deka D (2002) Transit availability and automobile ownership: some policy implications. J Plan Educ Res 21:285–300

Ewing R, Cervero R (2001) Travel and the built environment: a synthesis. Transp Res Rec 1780:87–114

Ewing R, Cervero R (2010) Travel and the built environment: a meta-analysis. J Am Plan Assoc 76(3):265–294

Federal Highway Administration (2009) 2009 National Household Travel Survey. (See http://nhts.ornl.gov)

Gallup (1989) Transport Report, CQ926

Goodwin PB (1993) Car ownership and public transport use: revisiting the interaction. Transportation 20(1):21–33

Haas P, Miknaitis G, Cooper H, Young L, Benedict A (2010) Transit oriented development and the potential for VMT-related greenhouse gas emissions growth reduction. Report of the Center for Neighborhood Technology for the Center for Transit Oriented Development, pp. 1–64

Hensher DA (1998) The imbalance between car and public transport use in urban Australia: why does it exist? Transp Policy 5(4):193–204

Johnson M (2011) Metro planners contemplate system's second generation. (See, http://greatergreaterwashington.org/post/10965/metro-planners-contemplate-systems-second-generation/)

Keller A (2012) Creating a Transit Supply Index. Presented at the 12th Transport Chicago Conference

Kim HS, Kim E (2004) Effects of public transit on automobile ownership and use in households of the USA. Rev Urban Reg Dev Stud 16:245–262

Kingham S, Dickinson J, Copsey S (2001) Travelling to work: will people move out of their cars. Transp Policy 8(2):151–160

Kitamura R (1989) A causal analysis of car ownership and transit use. Transportation 16(2):155–173

Li J, Walker JL, Srinivasan S, Anderson WP (2010) Modeling private car ownership in China. Transp Res Rec 2193:76–84

Liu Y, Tremblay JM, Cirillo C (2014) An integrated model for discrete and continuous decisions with application to vehicle ownership, type and usage choices. Transp Res Part A 69:319–328

Polzin S, Pendyala R, Navari S (2002) Development of time-of-day–based transit accessibility analysis tool. Transp Res Rec 1799:35–41

Rood T (1998) The local index of transit availability: an implementation manual. Local Government Commission, Sacramento, California. (See http://www.lgc.org/)

Tal G, Handy S, Boarnet MG (2010) Draft policy brief on the impacts of transit access (distance to transit) based on a review of the empirical literature, for research on impacts of transportation and land use-related policies, California Air Resources Board (See http://arb.ca.gov/cc/sb375/policies/policies.htm)

Train K (1986) Qualitative choice analysis. Theory econometrics, and an application to Automobile Demand. MIT press, Cambridge

Train KE (2009) Discrete choice methods with simulation. Cambridge University Press, Cambridge, England

Transportation Research Board (2003) Transit Capacity and Quality of Service Manual, 2nd ed. TCRP Project 100. Washington, DC TRB, National Research Council. http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp100/part%203.pdf

Washington Metropolitan Area Transit Authority (2012) Web page launched to gather public input on 30-year transit plan. (See http://www.wmata.com/about_metro/news/PressReleaseDetail.cfm?ReleaseID=4753)

Washington Metropolitan Area Transit Authority (2012) Metro invests in 'better bus' service in DC, Maryland, Virginia. (See http://www.wmata.com/about_metro/news/PressReleaseDetail.cfm?ReleaseID=5233 and http://www.purplelinemd.com/en/)