Could the Built Environment Accurately Inform Us the Total Use of Cars in Our Region?

Predicting Total Use of Cars using Urban Forms Matrices in Washington DC

Purba, Denissa Sari Darmawi

dpurba2, 667758459

University of Illinois at Urbana-Champaign, US

ABSTRACT

The total number of car usage is an important input in designing transportation policy. However, providing an accurate total number of cars used on the roads remains scarce and expensive in its practice. Most transportation engineers and city planners are stuck in performing a computational expensive prediction model in estimating the total number of car usage. Ewing and Cervero in 2017 claim a potential strong relation between urban forms and travel behavior. They propose five potential urban form matrices. This paper studies the potential performance of the proposed urban forms matrices in predicting the total number of vehicles on the roads. The prediction model is constructed by comparing three linear regression models, i.e., Ordinary Least Squares (OLS), Ridge regression, and Lasso regression. Using Washington, DC, as the study region, we observe that all three models give reliable prediction performance. The Lasso model provides the best prediction performance with a 0.72 R^2 test value and 15.55 MSE in the testing sample. We find that the constructed prediction model shows that the urban form characteristics are informative to predict vehicle usage in Washington, DC.

1. INTRODUCTION

According to NHTS 2017, automobile dependency keeps increasing in the United States; almost 78% percent of people use private cars as a daily transportation mode to travel (NHTS, 2017). Automobile dependency consists of high per capita automobile travel, automobile-oriented land use patterns, and limited transport alternatives. It has many impacts on society's behavior and the economic activities within the urban. As the car-orientation increases, it also leads to the reduction of the viability of other travel modes. It leads to more dispersed land use and mobility intensive economic patterns that require more vehicle travel for access (Litman & Laube, 2002). This travel behavior has been believed as a primary cause of the long dilemma on sprawl (Barrington-Leigh & Millard-Ball, 2015). According to EPA in 2019, the gasoline cars on roads has contributed to the largest share of greenhouse gas emissions, 28.2 percent of 2018 greenhouse gas emissions (EPA, 2019).

One of the most enduring urban planning ideas is promoting compact and efficient cities that enforce people to drive less. Reduction use of cars would benefit our community in fighting some of the most vexing problems, e.g., sprawl, congestion, fuel-dependency, and climate changes (Ewing & Cervero, 2017). In its efforts to change automobile-oriented travel behaviors, many city planners often propose an urban redesign policy, yet, it remains low studies focusing on the interactions between urban forms conditions and cars travel behaviors (Manville, 2017). Several studies, including Boarnet and Crane (2001), Cao, Mokhtarian, and Handy (2009), and Cervero (2002), have demonstrated the interaction of built environments in influencing travel choices. However, past studies seem to undervalue the actual relation between urban forms and travel behavior. In most of the literature, the urban characteristics are limited to the land use determinations across the urban regions used to capture the effect of derived travel demands generated from a certain area. Those studies ignore transportation infrastructure and a variety of mode choices as part of the urban forms. On the other hand, some past studies also focus only on the relation of car-oriented behavior on the environmental characteristics (e.g., gas emissions and pollutants) rather than the urban forms of the regions (Ewing et al., 2008; Boarnet, 2006).

Ewing and Cervero (2017) argue that at least five necessary matrices (*Density, Diversity, Design, Destination Accessibility,* and *Distance to Transit*) measure the relations between urban built environments and the travel behaviors on every mode choice. The first measurement is *Density*, which captures the socio-demographic information of each region per unit of area. The variable of interest can be population, dwelling units, employment, building floor area, or something else. Population and employment are sometimes summed to compute an overall activity density per areal unit. The second measurement is *Diversity*, which measures the degree of land use variation in the given area. The land use characteristics include the macroscopic view of each functional space in the district (e.g., business districts, residential blocks, and campus area) and the microscopic view of each functional space inside the building (e.g., the total number of floors of the office building and total stories of apartments, number of garages, and size of parking lots).

Next, the third measurement is the *Design*, representing the availability of transportation infrastructure networks inside the given area. The measure includes the geometry of the roads and the number of commute highways. Design matrices could include sidewalk coverage; average building setbacks; average street widths; or numbers of pedestrian crossings, street trees, or other physical variables that differentiate pedestrian-oriented environments from auto-oriented ones. The fourth measurement is *Destination Accessibility*, representing the ease of access and mobility to the targeted destinations. In some studies, regional accessibility is simply the distance to the central business district. In others, it is the number of jobs or other attractions reachable within a given travel time. Baht et al (2000) , with the Texas Department of Transportation, published an extensive report on summarizing potential methods in determining the urban accessibility index. The last measurement is the *Distance to Transit*. Ewing and Cervero (2017) state that transit presence should be clustered separately from the infrastructure network to capture the competition or potential shift from automobile to transit uses.

This paper explores applying the proposed five urban forms matrices from Ewing and Cervero (2017) in predicting the total use of cars in a given region. The paper also demonstrates the data collections for each urban form category using the U.S Census Data 2019, Smart Location Database 2013, and real-time vehicle traffic data from TomTom Traffic API. The prediction model is constructed by comparing three linear regression models, i.e., Ordinary Least Squares (OLS), Ridge regression, and Lasso regression. For the model exercise and implementation, this paper uses census tract-level data and Washington, DC as the study region.

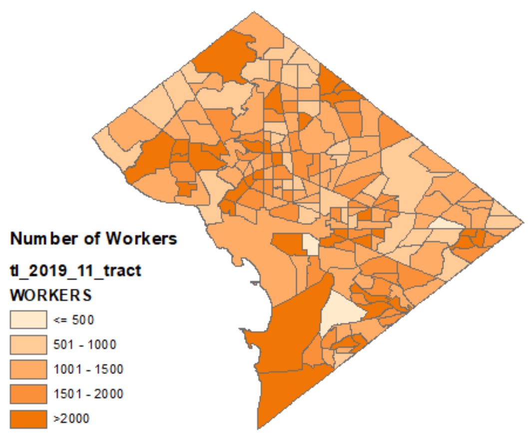
2. URBAN CHARACTERISTICS AND TRAVEL BEHAVIOR IN WASHINGTON, DC

This paper uses Washington, DC, as the study area for the data collections and prediction model exercise due to the dataset's availability. The following section describes the demographics and travel characteristics inside Washington, DC.

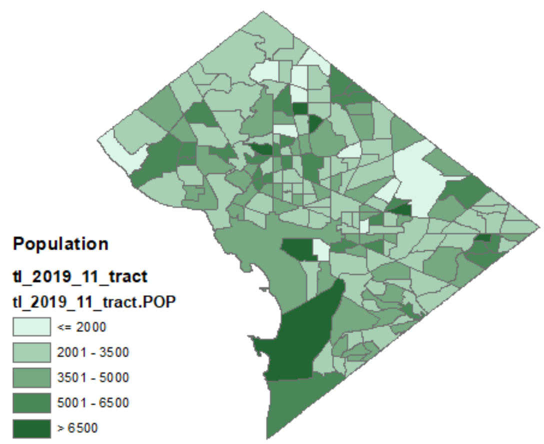
According to the 2019 Census, the District Columbia had a total population of 705,749 on July 1, 2019. The region has grown by 133,649 persons (or 23% percent) since 2000. Most of this growth occurred in the region's outer suburbs, which increased by 39.4 percent, and the outer ring jurisdictions increased by 26.2 percent. The Inner Suburbs and Central jurisdiction grew by 10.3 and 6.7 percent, respectively. The 2019 Census also showed that the total housing units inside the region are 26,936 units, with 41.1% owned property. According to DOES (2019), the number of employed District residents was unchanged at 381,900 in December 2018 to 381,900 in January 2019. The District's civilian labor force was up 400 from 403,500 in December 2018 to 403,900 in January 2019. The labor force participation rate was unchanged from 70.0 percent in December 2018 to 70.0 percent in January 2019. The total median household income is $82,602, which is higher than the average income in the United States. Figure 1(a) – 1(c) provides the distribution of the number of works, population, and household income across Washington, DC.

For the infrastructure, the mobility of citizens of Washington, DC is supported with ten major highways, three major international airports, and reliable Metrorail and Metrobus operated daily to the urban heart. Compared to the nation, the Washington, DC region relies less on driving and more on public transportation, influencing where people go, how they get there, and how long it takes. The total daily used cars are 31%, 47% lower than the national average (NHTS, 2017). Commuting in the Washington, DC area is also significantly less automobile dependency. Most of that difference is accounted for by the 44% higher share of transit riders. Nearly half of Washington, DC area workers commute by trains (24%) and buses (20%). Overall, more people opt to use alternative transit modes inside Washington, DC, than the national behavior. Figure 2 and Figure 3 show the commute time for workers using cars and the number of jobs with high accessibility to the transit point.

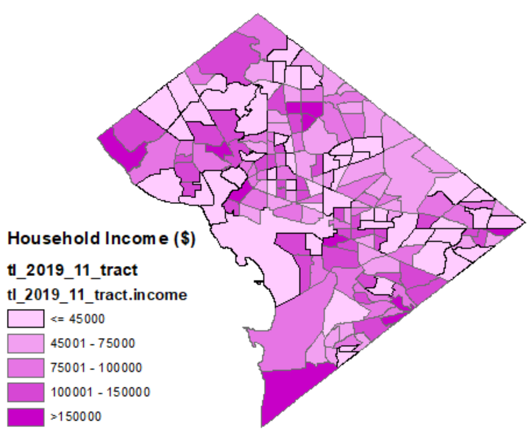
Additionally, Washington, DC is more active, with 41% saying they walk daily, compared to 22% nationally, and 5% biking daily, more than double the 2% national average (NHTS, 2017). Still, with the majority not walking or biking daily, the region seems to share similar built environment issues. In fact, of those who say they do not bike daily, the number of respondents who blame a lack of paths is significantly higher than in the nation. Additionally, compared to the nation, 8% more say they do not walk because of a lack of good sidewalks, revealing that a less car-dependent built environment is not necessarily more hospitable to pedestrians or cyclists as a result.



(a)



(b)



(c)

Figure 1 The distribution map of (a) number of workers; (b) Population; (c) household income in the Washington DC

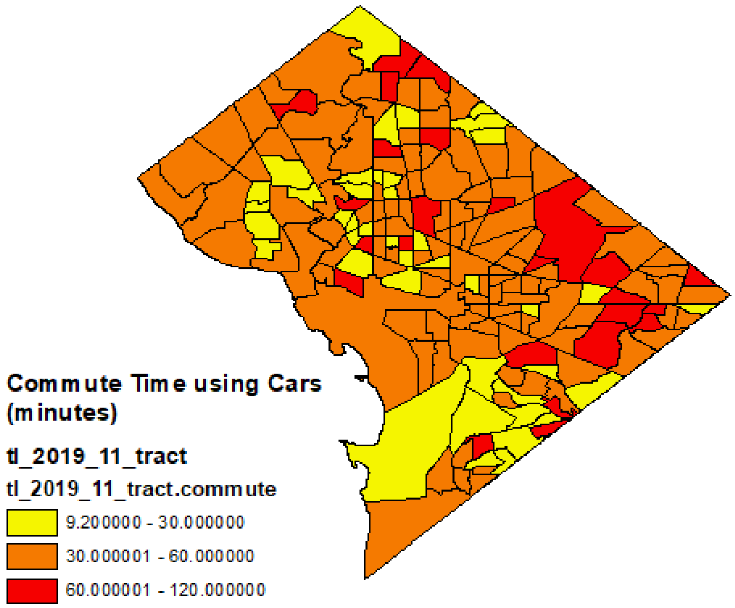


Figure 2 The distribution map of average commute time using cars in the Washington DC

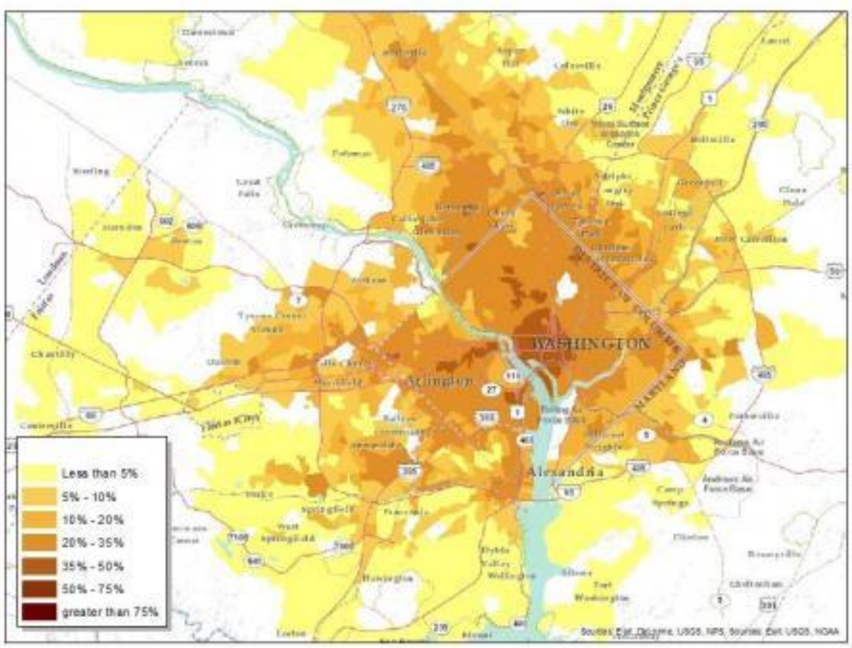


Figure 3 Percentage of all Jobs in the Metropolitan, D.C., Region That Are Accessible by Transit (Smart Location Database, 2013)

3. Data Collection and Preparations

The paper demonstrates the cross-sectional construction data for Washington, DC, with the Census Tract data cross-sections. For the selected study area, there are 179 census tract regions across Washington, DC. The selection of the area study is primarily based on limited data access for this study. Two types of the main dataset are constructed, i.e., the total vehicle data, as the predicted output, and the urban forms matrices dataset, as the set of predictors.

The total number of vehicles on the road is collected as a real-time vehicle counting on the Washington, DC road segments. The real-time traffic data is extracted using the TomTom Traffic API (TomTom, 2020). According to December 16, 2020, the service provides free access to real-time traffic data up to 1,000 queries. Due to this limited query access, the data collection was performed three times on Monday, November 23, 2020, between 6 am – 10 am, Tuesday, November 24, 2020, between 6 am – 10 am, and Wednesday, November 25, 2020, between 6 am-10 am. The data collection might not be perfect as we could not control the unobserved externality within each different day. In the ideal setting, it is important to control the traffic counting for the same day at the same time windows.

Using the TomTom Traffic API, we obtained the total number of vehicles on every road segment in Washington, DC. However, this study's targeted data structure is cross-sectional data that represents each census tract region. Hence, data manipulation is needed in the process. In this study, we conduct aggregation data from street maps to the census tract maps using ArcGIS. In each census tract, we aggregate the data by summing the number of vehicles of every road segment that falls inside the census tract polygon. For the road segments that are not inside the polygon but on the polygon side, we add the number of vehicles for each sided polygon. Through this strategy, we obtain the aggregated number of vehicles for every tract polygon in Washington, DC.

This paper uses an open-source database from the Smart Location Database (2013) and US Census Tract (2019a) for collecting the urban forms matrices data. The Smart Location Database is a nationwide geographic data resource for measuring location efficiency. It includes more than 90 attributes summarizing characteristics such as housing density, neighborhood design, destination accessibility, transit service, employment, and demographics. This database is primarily used to construct three matrices of urban forms, i.e., the *Design, Destination Accessibility,* and *Distance to Transit*. However, most attributes are available for every census block group in the United States. Hence, data aggregation is needed in the process. In this study, we conduct aggregation data from block maps to the census tract maps using ArcGIS. The data is aggregated according to the area of the block polygon and the targeted tract polygon. Next, we use the 2019 US Census Tract to collect the demographic data (e.g., population, employment, and residential) of Washington, DC, used for the *Density* matrix.

In limitation, this paper ignores the data collection of *Diversity* matric due to limited access. Recalling, *Diversity* measures the degree of land use variation in the given area. The land use characteristics include the macroscopic view of each functional space in the district (e.g., business districts, residential blocks, and campus area) and the microscopic view of each functional space inside the building (e.g., the total number of floors of the office building and total stories of apartments, number of garages, and size of parking lots). These data types could be collected using satellite data or maps API, yet, currently limited with access. Therefore, we only consider four urban matrices, i.e., *Density*, *Design*, *Destination Accessibility*, and *Distance to Transit*.

This paper also considers two categories of data as additional predictors that capture car-ownership and commute behavior inside Washington, DC. The car ownership and commute behavior categories are collected in US Census Tract (2019b) under the Commuting Flow Database. Therefore, in this study, we use 179 census tract data across Washington, DC, with four urban form matrices and two additional matrices as predictors to regress the total number of vehicles, which is 27 variables in total. Table 1 presents the statistical data for all 27 selected predictors.

4. Proposed Prediction Model and Model Selection Strategy

In this model, we propose a linear regression model to construct prediction of the total use number of vehicles in Washington, DC. Given a vector of predictor variables, X, and a total number of cars usage in the tract region, y, the prediction model is formulated as,

$$\hat{y}_i = \beta^T X_i + \varepsilon$$

We consider three linear regression models as comparison in this study, i.e., Ordinary Least Squares (OLS), Ridge regression, and Lasso regression. For the traditional linear regression, we define the loss functions using least squares as,

$$\arg\min \sum_{i=i}^{N} (y_i - \beta^T X_i)^2$$

The cost function for Ridge regression can be written as,

$$\arg\min \sum_{i=i}^{N} (y_i - \beta^T X_i)^2 + \lambda \sum_{j=0}^{p} \beta_j^2$$

The penalty term, λ, regularizes the coefficients such that if the coefficients take large values the optimization function is penalized. So, ridge regression shrinks the coefficients and it helps to reduce the model complexity and multi-collinearity. Ridge regression has two main benefits, i.e., adding a penalty term reduces overfitting and the penalty term guarantees the solution obtained.

Table 1 Summary statistics of independent variables (predictors)

Variable	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
Urban Forms Characteristic: Design, Destination Accessibility, Distance to Transit							
Network Density of Cars	1.23	2.85	0.00	0.00	0.00	0.71	16.18
Network Density of Multi-Mode	5.10	3.49	0.00	2.80	4.30	6.54	18.76
Network Density of Pedestrian	18.00	5.02	4.32	14.76	18.15	20.88	32.75
Intersection Density	117.11	50.61	26.46	83.64	114.42	144.49	304.18
Distance from population weighted centroid to nearest transit stop	201.72	110.38	0.00	140.23	182.67	250.32	561.76
Aggregate frequency of transit service within 0.25 miles during peak period	268.44	219.54	70.66	168.74	229.23	309.12	2298.33
Jobs within 45 minutes auto travel time	49085.91	18841.63	0.00	33629.09	48333.82	63056.75	97145.17
Jobs within 45-minute transit commute	27275.61	26983.61	0.00	11013.34	18378.51	34845.12	223643.81
Sociodemographic Characteristic: Density							
Population	3681.61	1404.59	65.00	2702.00	3379.00	4510.50	7923.00
Household	1544.95	724.58	3.00	1095.00	1366.00	1858.00	4538.00
Median Households Income	77046.31	44097.72	0.00	40417.00	74728.00	104734.50	235517.00
Household with 1 worker	702.15	414.91	0.00	425.50	623.00	865.50	2631.00
Household with 2 or more worker	464.47	280.58	3.00	242.00	426.00	603.50	1333.00
Workers	1486.62	627.98	101.00	1039.50	1400.00	1779.50	3598.00
Employment	3472.20	11367.64	0.00	204.00	596.00	1969.50	99063.00
Household with 2+ 18<	339.87	189.52	0.00	216.50	318.00	443.00	1153.00
Household with 65<	510.51	394.13	0.00	251.00	381.00	645.00	2213.00
Household with 1 65+	715.34	306.20	0.00	519.50	663.00	866.00	1648.00
Household with 2+ 65+	159.78	125.53	0.00	83.50	134.00	192.50	813.00
Car Ownership							
Household with no car	562.32	405.49	0.00	298.50	471.00	742.50	2255.00
Household with 1 car	675.44	376.78	0.00	429.00	587.00	844.00	2572.00
Household with 2 or more cars	307.19	230.73	0.00	159.50	240.00	391.50	1516.00
Commute Behavior							
Total worker commute with cars	42.63	15.53	9.20	30.80	42.10	53.15	81.30
Total worker commute by walks	34.77	11.37	0.00	28.00	35.00	43.40	63.00
Total worker work at home	5.67	4.08	0.00	2.80	5.30	8.00	19.10
Average commute time	31.00	6.02	0.00	27.55	31.10	34.90	48.90
Total Data	179	179	179	179	179	179	179
Total Training Data	90	90	90	90	90	90	90
Total Testing Data	89	89	89	89	89	89	89

The cost function for Lasso (least absolute shrinkage and selection operator) regression can be written as,

$$\arg\min \sum_{i=i}^{N} (y_i - \beta^T X_i)^2 + \lambda \sum_{j=0}^{p} |\beta_j|$$

The penalty term, λ, regularizes the coefficients such that if the coefficients take large values the optimization function is penalized. This type of regularization (L1) can lead to zero coefficients i.e. some of the features are completely neglected for the evaluation of output. Lasso regression not only helps in reducing over-fitting, but it can help in feature selection.

According to Mullainathan and Spiess (2017), we may use the prediction performance (e.g., R^2 and error) to the out sample data in selecting the best prediction model. However, Mullainathan and Spiess (2017) also remind that, as we choose model in term of its accuracy in predicting $\hat{y}$ outcome, this prediction model alone could not be used to determine the effect/relationship of $\hat{\beta}$. They argue that one obvious problem in making naïve inferences here is due to the lack of standard errors on the coefficients. The Ridge and Lasso regression in Machine Learning allow the prediction model to improve the accuracy in predicting the dependent variables by sacrificing the variance of independent variables. Hence, the result of selected prediction model in this paper will be used only in predicting the dependent variables without informing any possible inferences.

In this paper, we divide the data into two set of data (50%-50%), i.e., training data for exercising the prediction model and testing data (hold-out data) for measure the performance of the prediction model. We randomly select 90 cross-sectional data as training data and 89 cross-sectional data as testing data. All algorithm and regression formulations are coded in Jupyter using sklearn packages with cross-validation 5 folds (default).

5. Between Urban Forms and Total Car Usage

The prediction models result for OLS, Ridge regression, and Lasso regression are presented by Table 2. This prediction model is constructed using the training dataset (90 cross-sectional data) and tested with 89 cross-sectional testing data. The performance test for each prediction model is presented in Table 3. From the performance result, Lasso regression provide a better performance both in training dataset and testing dataset. According to our model selection strategy (see Section 4), the result shows that the best prediction model of the total number of cars usage in Washington, DC is Model 3 using Lasso regression, with 0.72 R^2 test value and 15.55 MSE.

Table 2 Prediction Models Result

Variable	**OLS Estimation**	**RIDGE Estimation**	**LASSO Estimation**
Urban Forms Characteristic: Design, Destination Accessibility, Transit			
Network Density of Cars	0.2353*	-0.018982	0.363*
	(0.003)	(0.211)	(0.003)
Network Density of Multi-Mode	0.049	0.068335	0.048792
	(0.109)	(0.183)	(0.183)
Network Density of Pedestrian	0.097*	0.085249	0.096811
	(0.007)	(0.191)	(0.190)
Intersection Density	-0.013	-0.010194	-0.013
	(0.012)	(0.012)	(0.012)
Distance from population weighted centroid to nearest transit stop	-0.002	-0.001872	-0.009*
	(0.005)	(0.005)	(0.012)
Aggregate frequency of transit service within 0.25 miles during peak period	-0.08769*	-0.008345*	-0.0765*
	(0.004)	(0.004)	(0.004)
Jobs within 45 minutes auto travel time	0.000	-0.000005	-
	(0.000)	0.00002	-
Jobs within 45-minute transit commute	-0.027*	-0.000031*	-0.003566
	(0.000)	(0.000)	(0.012)
Sociodemographic Characteristic: Density			
Population	-0.037*	-0.003393*	-0.0487*
	(0.006)	(0.001)	(0.183)
Household	0.001	0.000127	-
	(15007.110)	(15036.110)	-
Median Households Income	0.095*	0.09*	0.016513*
	(0.000)	(0.000)	(0.004)
Household with 1 worker	0.002	0.001827	0.1892
	(0.006)	(0.006)	(0.003)
Household with 2 or more worker	-0.005	-0.000226	-0.001371
	(0.007)	(0.008)	(0.046)
Workers	0.004	0.003362	0.004
	(0.002)	(0.002)	(0.002)
Employment	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Household with 2+ 18<	0.005	-0.005135	0.005
	(20.700)	(20.818)	(0.006)
Household with 65<	-0.003	0.010954	0.000
	(20.700)	(20.818)	-
Household with 1 65+	0.017	-0.010439	-
	(20.700)	(20.818)	-
Household with 2+ 65+	-0.008	-0.007351	-
	(20.700)	(20.818)	-
Car Ownership			
Household with no car	-0.007	-0.004131	-0.021*
	(5.431)	(5.431)	(0.00)
Household with 1 car	-0.002	-0.002	-0.457
	(0.005)	(0.005)	(0.046)
Household with 2 or more cars	0.010	0.010	0.0181*
	(0.213)	(0.213)	(0.00)
Commute Behavior			
Total worker commute with cars	0.112*	0.112*	0.112*
	(0.05)	(0.05)	(0.05)
Total worker commute by walks	-0.026	-0.042	-0.026
	(0.06)	(0.06)	(0.06)
Total worker work at home	-0.002	-0.03421	-
	(0.19)	(0.76)	-
Average commute time	0.089	0.089	0.089
	(0.16)	(0.16)	(0.16)

Table 3 Performance Comparison of the Prediction Models

Variable	OLS	RIDGE	LASSO
R^2 Train	0.749	0.748	0.775
MSE Train	20.962	21.044	12.559
R^2 Test	0.614	0.64	0.72
MSE Test	24.369	22.55	15.55

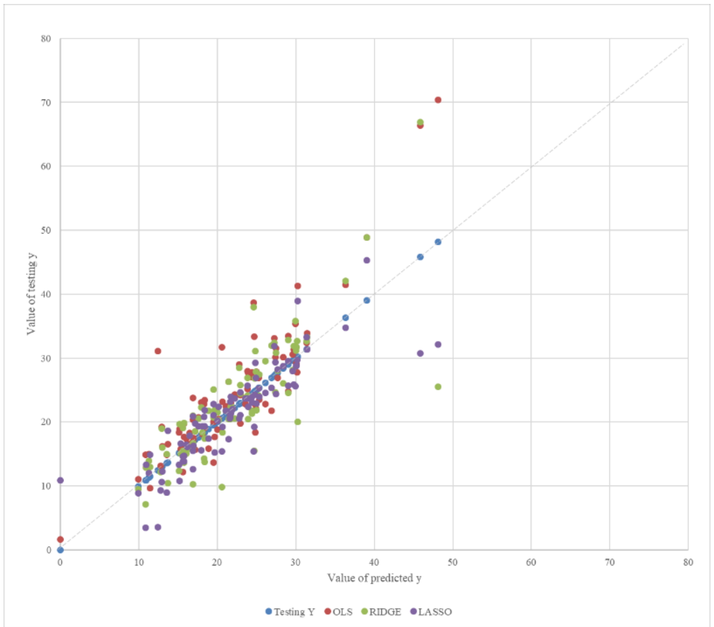


Figure 4 Comparison between testing value and predicted value of dependent variable for OLS, Ridge, and Lasso result

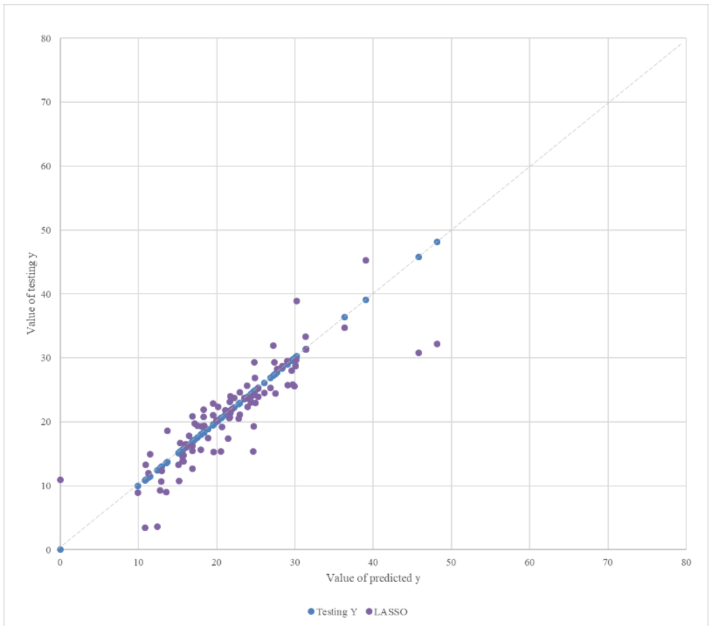


Figure 5 Comparison between testing value and predicted value of dependent variable for Lasso result (best prediction model)

The comparison between the out-sample testing data and the predicted results is illustrated in Figure 4 and Figure 5. In the Figure 4, we observed that Lasso regression predict better value of total number of cars usage compared to OLS model and Ridge model. We also observe that the error value in predicting higher number of cars (> 40 cars) increases in all three models in general, yet, Lasso regression performs better. We could observe more intensively the accuracy in predicting the number of car usage of Lasso regression in Figure 5.

Interestingly, the Lasso model indicates several dropping dependent variables from *Density* category and *Commute Behavior* category. According to Table 2, the total number of households in the tract area, the number of households with no elders (65+) and young age (<18) in the tract area, the number of households with elders (65+) in the tract area, and the total number of workers that work from home are not a good predictor according to Lasso regression. These drops have increased the accuracy of the prediction model about 17% from OLS model and 12.5% from Ridge regression. From the result, we also observe that all variables from *Design, Destination Accessibility, Distance to Transit,* and *Car Ownership* category is retained in the model as good predictors for in the Lasso regression. From these results, even though there are several drops, we still could conclude that the proposed five conceptual urban forms matrices (Ewing & Cervero, 2017) are capable to predict the total number of vehicle usage in Washington, DC with at-best performance 0.72 R^2 test value and 15.55 MSE in testing sample. Recalling the argument from to Mullainathan and Spiess (2017), the relation between urban forms and vehicles usage could not be inferred with this prediction models due to lack of standard error as bargain in aiming accurate prediction. Hence, the only relation that we could draw in these findings is that the constructed prediction model has proved that urban forms is potentially capable in predicting the vehicle usage in a region, specifically Washington, DC.

6. CONCLUSIONS

This paper explores the potential of urban forms characteristics, according to Ewing and Cervero (2017), in predicting the total usage of vehicles. The prediction model is constructed using three linear regression models, i.e., OLS, Ridge regression, and Lasso regression. In this paper, we choose Washington, DC, as a study area for the prediction exercise. According to the result, we observe that the Lasso model provides us a better prediction performance with a 0.72 R^2 test value and 15.55 MSE in the testing sample. In general, we conclude that all proposed urban forms categories and two additional categories, i.e., Car Ownership and Commute Behavior, can predict vehicle usage in Washington, DC. However, several drops of variables from the initial model show a performance improvement in the overall prediction model. The Lasso regression model could inform us of the potential drop variables. And, those removals show significant improvement to the prediction performances.

Regardless of the interesting findings of urban forms' capability to predict cars' total usage, there is room for improvements in advancing this study. First, it is important to notice that the prediction model is unique for the Washington, DC region. Earlier, we describe Washington, DC, as a high-density city with low car dependency. It is interesting to do the study comparisons with several cities with distinct urban structures, e.g., Jacksonville, FL, with low-density and high car dependency; and San Francisco, CA, with a low-density and low car dependency. Second, we need to include Diversity matrices as part of our urban forms to capture land use setting capability in predicting travel behavior. In this study, we exclude this category due to limited access. Next, it is also interesting to study urban forms' capability in predicting other travel modes' behavior, e.g., transit ridership, walking, cycling, and carpool ridership. And in the end, it is very important to advance this prediction study to the inference model to calculate the actual relationship between urban forms and the travel behavior in our regions. This inference information is crucial for city/transportation planners and engineers in providing better regulation/policy of urban design and improvement in mobility through integrated multi-modes.

REFERENCE

Athey, S., & Imbens, G. (2019). Machine Learning Methods Economists Should Know About. *Annual Review of Economics*, 685.

Baht, C., Handy, S., Kockelman, K., Mahmassani, H., Chen, Q., & Weston, L. (2000). *Urban Accessibility Indes: Literature Review.* Austin, TX: Texas Department of Transportation.

Barrington-Leigh, C., & Millard-Ball, A. (2015). A century of sprawl in the United States. *Proceedings of the National Academy of Sciences*, 8244–8249 | .

Boarnet, M. (2006). Planning's role in building healthy cities: An introduction to the special issue. *Journal of the American Planning Association*, 5-9.

Boarnet, M. G., & Crane, R. (2001). The influence of land use on travel behavior: Specification and estimation strategies. *Transportation Research A*, 823-845.

Cao, X., Mokhtarian, L., P., & Handy, S. L. (2009). The relationship between the built environment and nonwork travel: A case study of northern California. *Transportation Research Part A*, 548–559.

Cervero, R. (2002). Built environments and mode choice: Toward a normative framework. *Transportation Research D*, 265-284.

Department of Employment Services. (2019, March 11). *Department of Employment Services*. Retrieved December 16, 2020, from District of Columbia Unemployment Holds Steady at 5.4 Percent for January.

Ewing, R., & Cervero, R. (2017). Travel and the built environment: A Meta-analysis. *Journal of the American Planning Association*, 265-294.

Ewing, R., Bartholomew, K., Winkelman, S., Walters, J., & Chen, D. (2008). *Growing cooler: The evidence on urban development and climate change*. Washington, DC: Urban Land Institute.

Federal Highway Administration. (2017). *National Household Travel Survey*. Retrieved December 16, 2020, from National Household Travel Survey: https://nhts.ornl.gov/

Litman, T., & Laube, T. (2002). *Automobile Dependency and Economic Development* . Canada: Victoria Transport Policy Institute: Institute for Science and Technology Policy .

Manville, M. (2017). Travel and the built environment: Time for change. *Journal of the American Planning Association*, 29-32.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 87-106.

Pedregosa. (2011). Scikit-learn: Machine Learning in Python. *JMLR* 12, pp. 2825-2830.

TomTom. (2020). *TomTom Developer Portal*. Retrieved November 25, 2020, from TomTom Traffic API: https://developer.tomtom.com/traffic-api

United States Environmental Protection Agency. (2013). *Smart Location Mapping*. Retrieved Decemeber 16, 2020, from Smart Location Database: https://www.epa.gov/smartgrowth/smart-location-mapping

United States Environmental Protection Agency. (n.d.). *Greenhouse Gas Emissions: Sources of Greenhouse Gas Emissions*. Retrieved November 27, 2020, from United States Environmental Protection Agency: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#:~:text=Transportation%20(28.2%20percent%20of%202018,ships%2C%20trains%2C%20and%20planes.

US Census Bureau. (2019a). *US Census Data*. Retrieved December 16, 2020, from Population Profile in the United States: https://www.census.gov/data.html

US Census Bureau. (2019b). *US Census Data*. Retrieved December 16, 2020, from Commuting (Journey to Work): https://www.census.gov/topics/employment/commuting.html