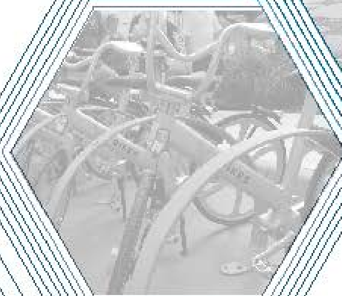
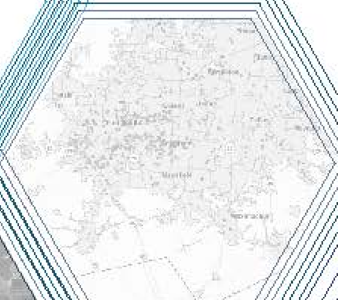




An Empirical Bayes Approach to Quantifying the Impact of Transportation Network Companies (TNCs) on VMT

Yunkyung Choi
Daniel G. Defoe
Gavin Van Steenburgh
Subhrajit Guhathakurta
Anurag Pande

FINAL REPORT

Center for Transportation Equity, Decisions & Dollars

AN EMPIRICAL BAYES APPROACH TO QUANTIFYING THE IMPACT OF TRANSPORTATION NETWORK COMPANIES (TNCS) ON VMT

FINAL PROJECT REPORT

By:
Yunkyung Choi
Georgia Institute of Technology

Daniel G. Defoe
California Polytechnic State University, San Luis Obispo

Gavin Van Steenburgh
California Polytechnic State University, San Luis Obispo

Subhrajit Guhathakurta
Georgia Institute of Technology

Anurag Pande
California Polytechnic State University, San Luis Obispo

Sponsorship:
CTEDD

For:
Center for Transportation, Equity, Decisions and Dollars **(CTEDD)**
USDOT University Transportation Center
The University of Texas at Arlington
601 W.Nedderman Dr. Suite 103
Arlington TX 76019-0108 United States
Phone: 817-272-5138 | Email: C-Tedd@uta.edu

Acknowledgments

This work was supported by a grant from the Center for Transportation Equity, Decisions, and Dollars (CTEDD) funded by U.S. Department of Transportation Research and Innovative Technology Administration (OST-R) and housed at The University of Texas at Arlington. We are also thankful to San Luis Obispo Council of Government staff, Mr. Daniel Audelo in particular, for providing part of the data used herein.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Center for Transportation, Equity, Decisions and Dollars (CTEDD), the U.S. Government and matching sponsor assume no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No. CTEDD 018-02	**2. Government Accession No.**	**3. Recipient's Catalog No.**

4. Title and Subtitle An Empirical Bayes Approach to Quantifying the Impact of Transportation Network Companies (TNCs) on VMT	**5. Report Date** January 2020 **6. Performing Organization Code**
7. Author(s): Yunkyung Choi, Daniel G. Defoe, Gavin Van Steenburgh, Subhrajit Guhathakurta, Anurag Pande	**8. Performing Organization Report No.** CTEDD 018-02
9. Performing Organization Name and Address Center for Transportation, Equity, Decisions and Dollars (CTEDD) USDOT University Transportation Center The University of Texas at Arlington 601 W.Nedderman Dr. Suite 103 Arlington TX 76019-0108 United States	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** 69A3551747134
12. Sponsoring Organization Name and Address United States of America Department of Transportation Research and Innovative Technology Administration	**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code**

15. Supplementary Notes

16. Abstract

Assessing the impacts of new and disruptive technologies on automobile usage and the modal split is emerging as a key issue for transportation planners and policymakers. This study offers a new approach to quantifying the impact of transportation network companies (TNCs) such as Uber and Lyft on vehicle-miles traveled (VMT). The approach is based on a simple idea from counterfactual theory, which is to compare VMT estimates after the TNCs introduction to a region to what the VMT would have been without the TNCs. The latter of the two is a counterfactual, and therefore more difficult to estimate. The study develops and demonstrates the Empirical Bayes (EB) measurement model for obtaining the counterfactual VMT estimates.

The EB method is widely used and accepted for traffic safety assessment. The approach proposed for estimating VMT changes is analogous to the quasi-experimental EB procedure for estimating crash reduction if a particular traffic safety treatment is applied to a roadway location. In this study, we reinterpret the traffic safety treatment as being akin to the introduction of TNCs and the estimation of crash reduction as analogous to the resulting change in VMT. This study develops an EB measurement model for the VMT in Atlanta and San Louis Obispo regions as a proof-of-concept. Counterfactual VMT estimate is obtained by combining two VMT estimates from 1) the cross-sectional models that estimated using data from comparative peer regions to Atlanta and San Luis Obispo regions and 2) the time-series models based on longitudinal data from Atlanta and San Luis Obispo regions. We measure the difference between the counterfactual VMT estimate and the current VMT estimate as an indicator of TNC impact. We find that the VMT estimate in a counterfactual scenario without TNCs is lower than the current VMT estimate over the period between 2012 and 2017. Also, the counterfactual VMT estimate shows a lower average annual growth rate compared to the current VMT estimate over the same period. The findings provide insight into the need to better integrate TNCs into the existing transportation system so that they don't increase VMT. We expect the approach to be useful in other research such as estimating effects of connected and automated vehicles (CAV) introduction on VMT in the future.

17. Key Words Vehicle miles of travel, VMT estimation, TNCs, Empirical Bayes, Counterfactual theory	**18. Distribution Statement**

19. Security Classification (of this report) Unclassified.	**20. Security Classification (of this page)** Unclassified.	**21. No. of Pages** 39	**22. Price**

Table of Contents

List of Tables

List of Figures

Abstract

Assessing the impacts of new and disruptive technologies on automobile usage and the modal split is emerging as a key issue for transportation planners and policymakers. This study offers a new approach to quantifying the impact of transportation network companies (TNCs) such as Uber and Lyft on vehicle-miles traveled (VMT). The approach is based on a simple idea from counterfactual theory, which is to compare VMT estimates after the TNCs introduction to a region to what the VMT would have been without the TNCs. The latter of the two is a counterfactual, and therefore more difficult to estimate. The study develops and demonstrates the Empirical Bayes (EB) measurement model for obtaining a counterfactual VMT estimate.

The EB method is widely used and accepted for traffic safety assessment. The approach proposed for estimating VMT changes is analogous to the quasi-experimental EB procedure for estimating crash reduction if a particular traffic safety treatment is applied to a roadway location. In this study, we reinterpret the traffic safety treatment as being akin to the introduction of TNCs and the estimation of crash reduction as analogous to the resulting change in VMT. This study develops an EB measurement model for the VMT in Atlanta and San Louis Obispo regions as a proof-of-concept. Counterfactual VMT estimate is obtained by combining two VMT estimates from 1) the cross-sectional models that estimated using data from comparative peer regions to Atlanta and San Luis Obispo regions and 2) the time-series models based on longitudinal data from Atlanta and San Luis Obispo regions. We measure the difference between the counterfactual VMT estimate and the current VMT estimate as an indicator of TNC impact. We find that the VMT estimates in a counterfactual scenario without TNCs are lower than the current VMT estimates over the period between 2012 and 2017. Also, the counterfactual VMT estimate shows a lower average annual growth rate compared to the current VMT estimate over the same period. The findings emphasize the need to better integrate TNCs into the existing public transit system to reduce their VMT impact. We expect the approach to be useful in other research such as estimating effects of connected and automated vehicles (CAV) introduction on VMT in the future.

Chapter I: Introduction

1.1 Background

Ridesharing connects passengers with owner-operator drivers through a smartphone application, offering a service that is more convenient than taxis and public transit (Brazil and Kirk, 2016). Transportation network companies (TNCs) such as Uber and Lyft have grown rapidly in recent years. Their ridesharing services have had a profound impact on reshaping urban transportation systems and have rendered the traditional modes (i.e. car, taxi, transit, walk, and bike) less competitive. Despite its significance, only a few U.S. cities can account for the impacts of ride-sourcing on traffic volumes and congestion levels at this point (DuPuis, Martin, & Rainwater, 2015). Also, existing research has produced conflicting results about the impact of TNCs on the transportation system due to a lack of available data.

Assessing the impacts of such new and disruptive technologies on automobile usage and the modal split has emerged as a key issue for transportation planners and policymakers. VMT per capita is one of the primary performance indicators for investigating the impacts of changes in policy on the transportation system (Ewing et al., 2014; Cervero and Murakami, 2010). The growth of VMT has both positive and negative externalities. One may look at the increase in VMT as a measure of economic growth and higher mobility. On the other hand, it may also be an indicator of increased traffic congestion, crashes, and emissions resulting from urban sprawl. In California, SB 743 mandates that VMT or similar measures be used as the main criteria for assessing the environmental impacts of the new development with interest in other states (Handy and Borarnet, 2014). Regardless of how VMT is used as a proxy, it is undoubtedly a key indicator of transportation system performance (Ewing et al, 2014). Therefore, VMT is the measure of choice for this study on quantifying the impact of TNCs on the transportation system performance.

The primary challenge in quantifying the impact of TNCs (or any other policy/technological change or ‘treatment’) is that several factors besides the ‘treatment’ affect VMT. Due to several confounding factors affecting VMT per capita for a region, it is often difficult to estimate VMT accurately. Thus, it is important to develop an appropriate model that accounts for multivariate causal relationships among multiple variables affecting VMT. Also, the lack of consistent and reliable VMT estimates may lead to difficulties in developing transportation planning and policy that is suitable for the context of interest. Thus, the application of an appropriate framework that provides reliable estimates of VMT is critical for evaluating the impact of TNCs.

1.2 Study objectives and scope

This study aims to assess the impacts of TNCs on automobile usage by offering a new approach. To be specific, this study focuses on formulating a quasi-experimental Empirical Bayesian (EB) framework to quantify the impact of the introduction of TNCs on VMT changes. The approach of this study adopts counterfactual theory, which compares VMT estimates after the TNCs introduction to a region to what the VMT would have been without the TNCs. The latter of the two is the counterfactual, and therefore more difficult to estimate. To estimate counterfactual VMT estimate without TNCs, this study develops and demonstrates the quasi-experimental before-after analysis using the EB approach, which is widely used in traffic safety estimation.

This study deems the problem at hand akin to accurately assessing roadway safety improvement at a given roadway location after an engineering treatment has been applied. The approach proposed for estimating VMT changes is the quasi-experimental EB procedure for comparing crash reduction on a roadway location where a particular traffic safety treatment is applied. In this study, we reinterpret the traffic safety treatment as being akin to the introduction of TNCs and the estimation of crash reduction as analogous to the resulting change in VMT. This study develops an EB measurement model for the VMT in Atlanta and San Louis Obispo regions as a proof-of-concept. Counterfactual VMT estimate is obtained by combining two VMT estimates from 1) the cross-sectional models that estimated using data from comparative peer regions to Atlanta and San Luis Obispo regions and 2) the time-series models based on longitudinal data from Atlanta and San Luis Obispo regions. We postulate that combining evidence from other regions in the country to the longitudinal trends from Atlanta and San Luis Obispo regions, respectively, can provide a more robust estimate of the counterfactual VMT. We measure the difference between the counterfactual VMT estimate and the current VMT estimate as an indicator of TNC impact. The proposed approach will provide a better understanding of emerging technologies and their impact on VMT changes and will equip the local and regional agencies to craft policy responses that yield the most benefits for the communities.

Chapter II: Literature Review

This literature review covers three groups of studies related to the factors affecting VMT, the existing methods for estimating VMT, and the effects of ridesharing on the transportation system. From these studies, we analyzed VMT changes associated with the introduction of TNCs.

2.1 Factors affecting VMT

Travel behavior and demand modeling research have found major factors that affect regional VMT. Among various factors, VMT is frequently related to land use, fuel price, highway capacity, and transit access. The first factor that is significantly correlated with VMT is land use and built environment (Ewing and Cervero, 2010; Cervero and Murakami, 2010; Heres-Del-Valle and Niemeier, 2011; Salon et al., 2012; Zhang et al., 2012; McMullen and Eckstein, 2013; Ding et al., 2014; Garceau, 2015; Lu et al., 2016; Choi and Zhang, 2017). It is often characterized by '5D' variables indicating density, diversity, design, destination accessibility, and distance to transit. Ewing and Cervero (2010) conducted a meta-analysis on more than 200 studies of the built environment and travel to uncover an elasticity of explanatory variables in travel outcome including VMT. Among 5D variables, they found that destination accessibility and design metrics expressed in the street network and intersection density are strongly associated with VMT. Population density shows a relatively weak association with VMT, indicating that density plays an intermediate variable in the relationship between other D variables and VMT.

Numerous studies suggest that fuel price is a major factor that is correlated with VMT (Espey, 1998; Graham and Glaister, 2004; Goodwin et al., 2004; Small and Dender, 2007; Brons et al., 2008). In previous studies, fuel price showed a negative association with VMT by having elasticities ranging from -0.58 to -0.10. This trend persisted in both the short and long term. Also, Small and Dender (2007) found that fuel efficiency causes additional travel which increases VMT. Highway capacity is another factor that induces traffic (Cervero, 2002; Ewing and Cervero, 2008; Handy and Boarnet, 2014). In many cases, highway investments tend to contribute to decentralization and low-density development, resulting in increases in VMT. The elasticity of highway capacity regarding VMT, in the long run, was estimated at between 0.63 and 0.73 (Cervero, 2002).

The influence of transit on VMT has been largely discussed over the decades. New transit developments such as transit-oriented development (TOD) is believed to contribute to reducing household vehicle travel due to its compact and mixed land use, and pedestrian-friendly environment around transit facilities. Existing studies found that transit use is statistically significant and negatively related to VMT per capita (McMullen and Eckstein, 2013). Also, people who live near transit tend to own fewer vehicles, drive smaller distances, and use more non-motorized modes of travel compared to the residents of non-transit accessible areas (Boarnet and Crane, 2001; Arrington and Cervero, 2008; Hale, 2014; Nasri and Zhang, 2014; Ewing and Hamidi, 2014; Langlois et al., 2015; Laham and Noland, 2017; Park et al., 2018).

2.2 VMT estimation

There exist different methods to estimate regional VMT. Among various estimation methods, existing studies have identified two major approaches: the use of road network traffic counts and the use of non-traffic data sources (Kumapley and Fricker, 1996; Fricker and Kumapley, 2002; Liu and Kaiser, 2006). The first approach in VMT estimation is the use of traffic count data provided by the FHWA's Highway Performance Monitoring System (HPMS). This approach uses annual average daily traffic (AADT) which is calculated as the annual total volume of traffic passing a road segment in both directions divided by 365 days. VMT is calculated as the sum of all products of AADT and length for each road segment within the desired aggregation level (e.g. functional system, urbanized area).

Another approach uses non-traffic data that are indirect predictors of VMT such as the number of households, household income, and the number of vehicles in a household. This approach typically employs various statistical techniques such as regression model (Newmark and Haas, 2015; Leard, Linn, and Munnings, 2016; Kim et al, 2016), structural equation modeling (Cervero and Murakami, 2010; Ewing et al., 2014), multivariate time-series analysis (Choo et al., 2005; Soltani-Sobh, Heaslip, and Bosworth, 2016), and shortest path model based on road networks (Wang, Bai, and Bao, 2011). Due to such variations, the decision to choose a particular VMT estimation method largely depends on data availability and data structure.

2.3 Ridesharing effects

Since its introduction, ridesharing services have disrupted and reshaped urban transportation systems. Due to its rapid growth, previous studies have attempted to investigate the effects of ridesharing on various topics including VMT (Rodier, Alemi, and Smith, 2016; Henao and Marshall, 2019), traffic congestion (Dupuis, Martin, and Rainwater, 2015; Li, Hong, and Zhang, 2016; Erhardt et al., 2019), the shift in modes of transportation (Wallsten, 2015; Hoffman, Ipeirotis, and Sundararajan, 2016, Hall, Palsson, and Price, 2018), DUI and fatal vehicle crashes (Greenwood and Wattal, 2015; Martin-Buck, 2017; Wang, Ardakani, and Schneider, 2017; Dills and Mulholland, 2018), and environmental benefits (Caulfield, 2009; Greenblatt and Shaheen, 2015; Yu et al., 2017).

To date, however, existing research has produced conflicting results due to a lack of available data. Henao and Marshall (2019) found that the efficiency rate of ride-sourcing is low compared to other modes of transportation, which results in increases in VMT and marginal change in travel behavior. Erhardt et al. (2019) also found that TNCs are the biggest contributor to growing traffic congestion in San Francisco between 2010 and 2016. On the other hand, Rodier, Alemi, and Smith (2016) suggested that relatively large reductions in VMT about 11 percent to 19 percent are possible from moderate and high participation levels of ridesharing. Hall, Palsson, and Price (2018) also found that ride-sourcing complements the use of transit, increasing ridership by five percent, particularly more in larger cities. Such inconsistent results indicate the need for further investigation of the effects of ridesharing.

Chapter III: Methodology

3.1 Research design

The framework developed to quantify the impact of the introduction of TNCs on VMT changes is based on a simple idea from counterfactual theory, which considers possible alternatives to historical events and assesses the consequences of those changes. This study uses the counterfactual theory to compare VMT estimates after the introduction of the TNCs in a region to what the VMT would have been without the TNCs. The latter is the counterfactual VMT, and it is often difficult to estimate. To estimate counterfactual VMT, the research team employs a quasi-experimental Empirical Bayes (EB) approach. The EB method has been widely applied to observational before-and-after studies in traffic safety, such as estimating crash reduction on a roadway location of where traffic safety treatment is applied. We reinterpret the traffic safety treatment as being akin to the introduction of TNCs and the estimation of crash reduction as analogous to the resulting VMT changes.

According to Hauer (1997), the EB method can be a better alternative to a simple before-after comparison for evaluating the effect of a treatment. Contrary to other traditional before-and-after study methods, the EB method estimates the safety of a subject entity based on two separate sources of information: 1) the historical safety performance of a subject entity and 2) the expected safety performance for similar entities. By combining these two sets of evidence, the EB method increases the precision of estimation and corrects for the regression-to-mean bias (Hauer, Harwood, Council, & Griffith, 2002; Zhou, Zhao, Hsu, & Huang, 2013; Yu & Abdel-Aty, 2014). The idea behind the EB approach may be better explained by the following analogy from traffic safety: Consider a new teen driver, Sean, in Arlington, TX, who had no crashes during his first year of driving. Also, assume that on average new teen drivers have 0.10 accidents in their second year of driving. It would not be appropriate to claim that Sean is expected to have zero accidents based on his available historical record only in the coming year. Similarly, it would also be inappropriate to conclude that the number of expected crashes for him in the coming year would be 0.10 by totally disregarding his accident record (Hauer, Harwood, Council, & Griffith, 2002). A more appropriate quasi-experimental approach would be to statistically combine the two sets of evidence and treat the information from the reference group as the prior. Compared to the fully Bayesian approach, the EB approach attempts to learn an appropriate prior distribution from ongoing statistical experience, rather than rely on an assumption.

Using the theoretical framework of the EB approach, this study improves the accuracy of the VMT estimates for the counterfactual scenario—what would have happened to the VMT had the treatment not been applied—by combining the evidence from reference regions and longitudinal VMT estimates from the subject region. Thus, this study conducts 1) cross-sectional analysis from reference sites and 2) longitudinal analysis from the subject site to provide a complete picture of the counterfactual VMT estimates assuming TNCs do not exist. The joint use of the two VMT estimates from a cross-sectional analysis and longitudinal analysis is implemented using a weighted average. Hence, the VMT estimates from the methodology in this study can be understood as a 'weighted combination' of the estimates from a region's characteristics as well as the estimates from the reference group data. In essence, the VMT estimates from cross-sectional analysis form the EB prior estimate that is updated using the longitudinal VMT trends for the subject region. The weight assigned to each of the estimates depends on the variance of the respective estimates, and the goodness of fit of each analytical model is used to determine the weight. Figure 1 describes the EB measurements of counterfactual VMT that is a weighted combination of two estimates from cross-sectional data from reference regions and longitudinal data from the subject region. The weighted estimate of the counterfactual VMT (A) is compared to a current estimate of the VMT (B) to quantify the effect of a treatment. In other words, this study assumes the difference between the counterfactual VMT estimate and the current VMT estimates as an indicator of TNC impact.

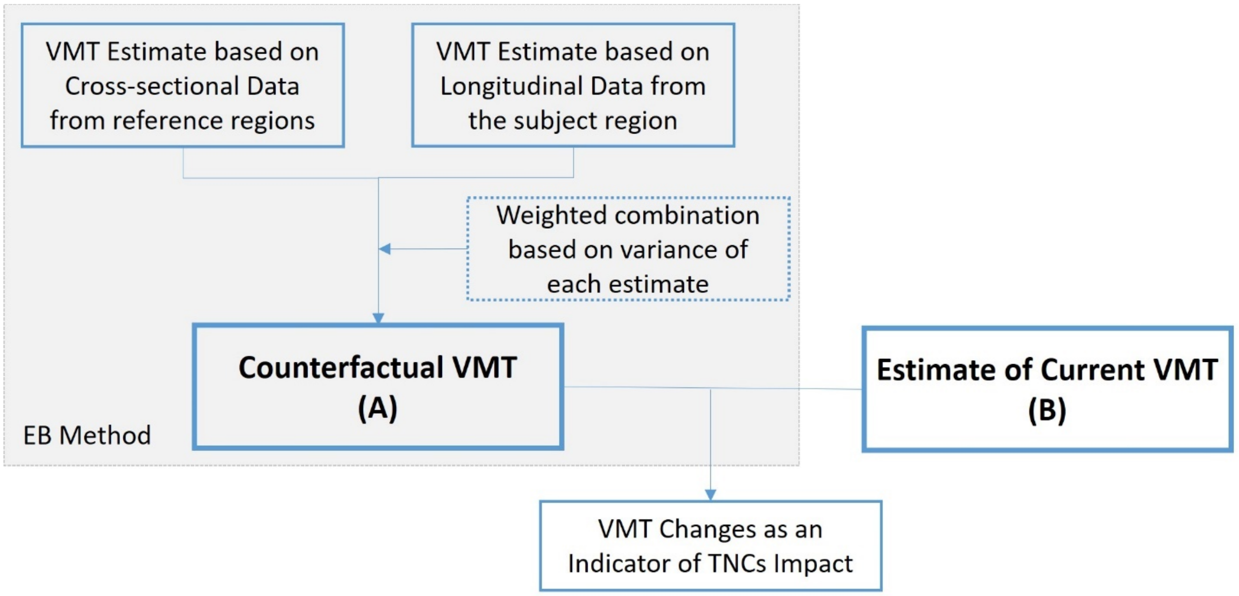


Figure 1. EB-based VMT estimation framework

This study applies the EB-based framework to quantify the impact of TNCs specifically in the Atlanta metropolitan statistical area and the San Luis Obispo metropolitan statistical area. For each region, the research team conducted a before-and-after assessment focusing on the change in daily VMT per capita first in 2011 when TNC services were not spread out and again in 2017 when the TNC service coverage was dramatically expanded. Since TNC was introduced in only five cities—San Francisco, New York, Seattle, Chicago, Washington D.C.—at the end of 2011, our counterfactual VMT estimates for Atlanta and San Luis Obispo regions in 2011 were both zero. In the process of estimating a counterfactual scenario where TNCs do not exist for the period between 2012 and 2017, we calibrated the model with 2011 conditions. This means that when the model is run for each year's inputs, it represents a counterfactual case with no TNCs. After estimating all counterfactual estimates for each year, we compared the counterfactual scenarios to the current conditions to evaluate the impact of TNCs on VMT by testing the following hypotheses:

1) If TNCs do not affect VMT, the gap between the counterfactual VMT estimate and the current VMT estimate should be negligible.

2) If TNCs generate VMT, then the current VMT estimate should be larger than the counterfactual VMT estimate.

3) If TNCs decrease VMT, then the current VMT estimate should be smaller than the counterfactual VMT estimate.

3.2 Method of analysis

3.2.1 VMT estimation for a counterfactual scenario

The main goal of this study is to estimate the VMT for a counterfactual case where TNCs do not exist. To achieve this goal, this study developed a cross-sectional model to capture the long-run relationships between land use and transportation based on data from the reference regions. Structural Equation Modeling (SEM) approach by Ewing et al. (2014) provides a framework to estimate the cross-sectional model. SEM is a statistical technique for representing multivariate causal relationships that involve multiple endogenous and exogenous variables (McDonald and Ho, 2002; Grace, 2006). As a modeling tool, SEM has been widely used in a range of fields including social, behavioral, and natural sciences. The use of SEM in travel behavior studies has also increased because it captures the complexity of the relationships between land use patterns, travel activities, and other factors (Golog, 2001). Thus, this study constructed a cross-sectional SEM model from reference regions to estimate the effects of land use, highway capacity, fuel price, and transit service on VMT in 2011 when TNC activities were negligible.

SEM involves a set of matrix equations in a model. Contrary to other modeling techniques, a variable in SEM can be the dependent variable in one equation and an independent variable in another equation. This attribute requires SEM to specify endogenous and exogenous variables in the system. In short, endogenous variables appear as dependent variables in at least one equation while exogenous variables are independent variables that are not influenced by other variables in a model. Although exogenous variables may be correlated with one another, the estimation of their association is not part of the model. The causal relations between variables are expressed by one-way arrows with regression coefficient values, and covariances or correlations between exogenous variables are depicted by double-headed arrows without a causal interpretation. An error and residual term for each endogenous variable are represented by a circle. On hypothesized causal relationships, SEM estimates both direct effects between an independent variable and a dependent variable and indirect effect between an independent variable and a mediator variable and between a mediator variable and a dependent variable. Then, SEM calculates the total effect by combining a direct and one or more indirect effects.

The development of SEM involves several steps, which include model specification, identification, parameter estimation, and model evaluation, and model modification (Hoyle 2011; Kline 2015). Model specification is based on a combination of theory, hypotheses being analyzed, and empirical results from previous studies. The hypothesized relationships among the variables in a model are visualized using a path diagram. The next step is to check for model identification whether the model is under-identified or over-identified. Once coefficients of parameters are estimated from the just-identified model, the model performance can be evaluated with statistical goodness-of-fit indices. Based on changes in model fit, researchers can adjust the model by adding, dropping, or changing variables and pathways. By following these steps, we constructed an SEM model using the maximum likelihood (ML) estimation method in STATA. The selected endogenous and exogenous variables were derived from previous empirical studies, and the multicollinearity among the independent variables and outliers in data were checked before estimating the model. We tested a large number of possible path combinations that logically capture causal relationships between land use, economic factors, transportation supply, and transportation demand. After several iterations, we finalized the model based on the goodness-of-fit measures.

Several goodness-of-fit tests can be used to evaluate the outcome of the SEM analysis. Among several indices, the Chi-squared test, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the standardized root mean square residual (SRMR) are commonly used (Kline, 2015). The Chi-squared test, RMSEA, and SRMR are categorized into absolute fit indices that evaluate how well the hypothesized model fits the sample data based on the discrepancy between the proposed model and the data. CFI is an incremental fit index that determines an optimal model out of a set of competing models. The Chi-squared test assesses the difference between observed and expected covariance matrices, but this statistic is sensitive to sample size. Compared to the Chi-squared test, other indices are less sensitive to sample size and distribution of the data. RMSEA evaluates how well the proposed model approximates the true model. An RMSEA value less than 0.05 indicates a good approximation of model, and a value between 0.05 and 0.08 is also an acceptable level to determine the adequacy of model fit (Steiger and Lind, 1980; Browne and Cudeck, 1992). RMSEA tends to reject true-population models when the sample size is smaller than 250, but CFI produces a better quality of estimation in the same condition (Hu and Bentler, 1998). CFI compares the hypothesized model with a baseline model with no restrictions, and a value greater than 0.90 is generally acceptable (Hu and Bentler, 1998). SRMR measures the standardized difference between the hypothesized and the observed covariance matrices. An SRMR value of less than 0.08 is considered a good fit (Hu and Bentler, 1998). To mitigate the limitations of various indices, Hu and Bentler (1999) recommends evaluating model fit based on a combination of two indices (e.g. the combination of RMSEA lower than 0.06 and SRMR lower than 0.09, the combination of CFI higher than 0.96 and SRMR lower than 0.09).

The accuracy of the EB method largely depends on how we select the reference regions that serve the prior estimate for VMT, but the definition of similarity between the reference region and the subject region remains somewhat ambiguous. This study employs the factor- and cluster-analysis techniques to identify homogeneous groups of regions based on the similarity of the attributes affecting VMTs that are frequently employed in previous studies. By conducting factor analysis, we derived three factors—density, socioeconomic, and transit-related—from multiple variables and conducted a k-means cluster analysis that produced a six-clusters solution for all metropolitan statistical areas (MSAs). The optimal number of clusters was determined based on the Elbow method that measures the point when the total within-cluster sum of square is minimized. After identifying clusters that include the Atlanta MSA and the San Luis Obispo MSA, respectively, we applied the SEM method to the two clusters for constructing the cross-sectional model. The cross-sectional SEM models for 2011 were used to estimate the predicted values of counterfactual VMT for each year between 2012 and 2017.

The output from the cross-sectional SEM-based approach forms the EB prior estimate that was updated using the longitudinal VMT trends for Atlanta and San Luis Obispo region. We employed time series analysis to predict counterfactual VMT estimates based on observed historical VMT data. Among several time series models, the autoregressive integrated moving average (ARIMA) is frequently used due to its flexibility to represent several varieties of times series with simplicity and the ability to model non-stationary time series using differencing (Box, Jenkins, Reinsel, and Ljung, 1976). The predicted VMT estimates from the ARIMA model was combined with the VMT estimates from the cross-sectional SEM model based on the respective variances of each estimate to obtain the counterfactual VMT estimates for the regions. We used the adjusted R-squared value of the ARIMA model to determine the weight of the two VMT estimates.

3.2.2 Current VMT estimation

The developed research framework involves the comparison between counterfactual VMT estimates and current estimates of VMT to quantify the impact on TNCs on VMT. Thus, this study formulated a framework for estimating current VMTs based on traffic-count-based methods. In traffic-count-based methods, annual average daily traffic (AADT), which is calculated as the annual total volume of traffic passing a road segment in both directions divided by 365 days, is frequently used for estimating VMT. We used actual traffic counts from the AADT dataset to calculate VMT for each MSA and adjusted the VMT estimates with a weight factor based on the Federal Highway Administration (FHWA)'s official highway statistics for statewide VMT. This framework allows researchers to estimate VMTs for any geographic boundaries (e.g. urbanized area, county, city) depending on research interests.

According to Highway Performance Monitoring System (HPMS) field manual (2016), daily vehicle miles of travel (DVMT) can be estimated as the sum of all products of AADT and length for each road segment. Based on the manual, we calculated VMT by multiplying the AADT with the road length and summarizing it by MSAs. As a result, VMT estimates for 353 MSAs were calculated. The DVMT at the MSA level can be expressed as follows:

$$\text{DVMT}_j = \sum \text{AADT}_{ij} \times \text{Roadway Length}_{ij}$$

where DVMT_j is the daily VMT estimate in MSA j, AADT_{ij} is the annual average daily traffic volume on a road segment i in MSA j, and roadway length_{ij} is a length in miles for a roadway segment i in MSA j.

The validation of VMT estimates was followed by two steps. First, we checked the discrepancy between the original HPMS's DVMT and the projected HPMS's DVMT at the state level. Since HPMS geospatial AADT dataset is in a shapefile format, the projection of data is required to calculate the length of road segments within the boundary of each MSA. In the ArcGIS program, we applied different projected coordinate systems to each state based on their location (e.g. State System NAD83 Georgia Lambert for Georgia, UTM NAD83 Zone 10 for California). The percentage difference between the original DVMT and the projected DVMT at the state level was used to estimate DVMT at the MSA level. Then, we adjusted DVMT estimates based on FHWA's official VMT statistics at the state level. This is because HPMS data may not completely cover all Federal-aid highways. To mitigate this limitation, FHWA recommends the estimation of local roadway VMT using statewide data, which is more statistically significant than local-level HPMS data. The discrepancy between the original HPMS's DVMT and the FHWA's official DVMT at the state level was used as a weight factor to adjust DVMT estimates at the MSA level. Figure 2 describes the process of validating VMT estimates for the MSA level.

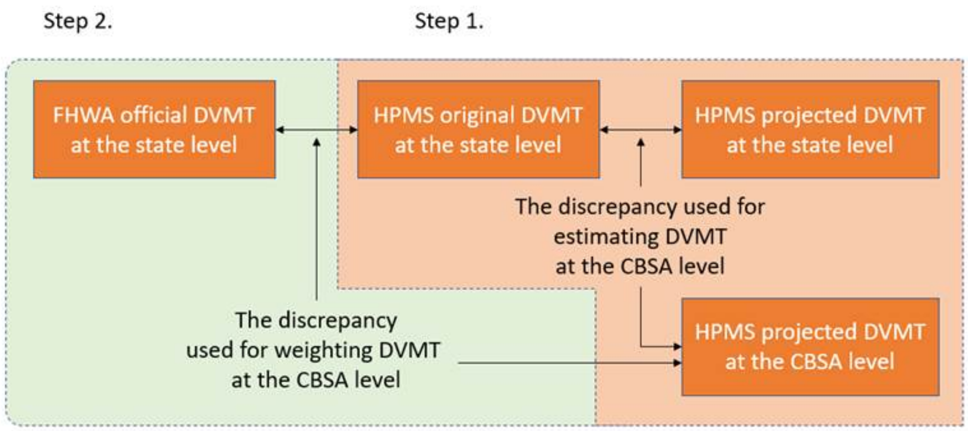


Figure 2. HPMS VMT calculation and validation process for the MSA level

3.3 Data and Variables

The research team collected data from several primary sources for cross-sectional and longitudinal analyses. To construct a cross-sectional model, we developed several criteria for identifying comparative peer regions to Atlanta and San Luis Obispo regions. The selection criteria are based on multiple statistics including Census, American Community Survey (ACS), and National Transit Database for the attributes affecting VMTs. Table 1 defines the variable names that are employed in the cross-sectional SEM model, along with their definitions and data sources. The SEM model includes three types of variables: outcome, exogenous, and endogenous variables. The key outcome variable is VMT per capita at the MSA level and it is estimated using AADT for road segments from HPMS geospatial data provided by FHWA. Since the boundaries of MSAs have changed over time, we selected 358 MSAs that remained unchanged between 2011 to 2017. Exogenous variables include population, per capita income, average fuel price, railroad density, primary road density, and other roads density. Endogenous explanatory variables are population density, percentage of commuting mode by automobile, transit service frequency, and annual passenger miles per capita. All variables in the SEM model were transformed by taking natural logarithms to have a normal distribution.

Table 1. Variables included in the cross-sectional model for 2011

Variable	Definition	Source
Outcome variable		
VMT	Natural log of daily VMT per capita	FHWA HPMS
Exogenous variables		
Population	Natural log of population (in thousands)	ACS
Per capita income	Natural log of per capita income	ACS
Average fuel price	Natural log of average metropolitan fuel price	AAA.com
Rail density	Natural log of railroad density	U.S. Census TIGER
Primary road density	Natural log of primary road density per square mile	U.S. Census TIGER
Other roads density	Natural log of other roads density per square mile	U.S. Census TIGER
Endogenous variables		
Population density	Natural log of gross population density	ACS
Commuting mode by auto	Natural log of percentage of commuting mode by auto	ACS
Transit service frequency	Natural log of transit service frequency	National Transit Database
Passenger miles per capita	Natural log of annual transit passenger miles per capita	National Transit Database

For the longitudinal analysis, we obtained historical VMT statistics from the Georgia Department of Transportation (GDOT) and the San Luis Obispo Council of Governments. Historical VMT statistics span between 1984 and 2017 for the Atlanta MSA, and 1996 and 2017 for the San Luis Obispo MSA. The Office of Transportation Data in GDOT has published mileage by route and road system report 445 since 1984, and this report contains VMT that is calculated by multiplying the AADT by the section length for each section and summarizing it at the County level. Within the Atlanta MSA, there are 28 counties including Barrow, Bartow, Butts, Carroll, Cherokee, Clayton, Cobb, Coweta, Dawson, DeKalb, Douglas, Fayette, Forsyth, Fulton, Gwinnett, Haralson, Heard, Henry, Jasper, Lamar, Meriwether, Newton, Paulding, Pickens, Pike, Rockdale, Spalding, and Walton. These county-level VMT statistics were aggregated to the Atlanta MSA.

Chapter IV: Results and Discussions

4.1 Identification of reference regions

To identify comparative peer regions, we employed both factor- and cluster-analysis techniques. Because there are multiple variables, we conducted the factor analysis to reduce the variables into a set of factors by applying the principal components method. We used an eigenvalue greater than one as the threshold to determine the number of factors, which resulted in the following three factors: density and transit, socioeconomic, and highway capacity factors. The density and transit factor was extracted based on population density, vehicle or train revenue miles, and passenger car revenue miles. The socioeconomic factor includes per capita income, average fuel price, and share of commuting mode by automobile. The highway capacity factor is defined as the roadway density by the type of road system. Table 2 presents the results of the factor analysis on nine variables. Factor loadings smaller than 0.3 were suppressed.

Table 2. The factor analysis results

Variables	Density and transit factor[a]	Socioeconomic factor[a]	Highway capacity factor[a]
Population density	0.628		
Vehicle or Train revenue miles	0.922		
Passenger car revenue miles	0.945		
Per capita income		0.691	
Fuel Price		0.810	
Commuting mode by Auto		-0.663	
Primary road per square mile			0.807
Secondary road per square mile			0.836
Local road per square mile			0.834

[a] Extraction method: Principal component analysis. Rotation method: varimax with Kaiser normalization.

With three factor scores, we conducted the cluster analysis to classify all MSAs into relative groups that share similar attributes. Table 3 summarizes the results of the cluster analysis to MSAs. When we increase the number of clusters from four to six, the clusters that include Atlanta and the San Luis Obispo regions, respectively, present a relatively constant number of MSAs within the cluster. The optimal number of clusters is based on the Elbow method, which suggests a six-cluster solution, and this solution results in Atlanta being grouped with 79 other comparable MSAs in cluster type 3 and San Luis Obispo being grouped with 66 other comparable MSAs in cluster type 2. Regardless of the number of clusters, New York MSA is identified as an outlier due to its unique attributes on land use and the built environment.

Table 3. The cluster analysis results for all MSAs

Cluster type	Number of clusters			
	3	4	5	6
1	93	161	6	1
2	221	83	67	67
3	1	1	1	80
4		70	160	2
5			81	6
6				159
Total	315	315	315	315

*Orange colored cells represent clusters that include the Atlanta MSA; Green colored cells represent clusters that include the San Luis Obispo MSA.

4.2 VMT estimation for the EB prior estimate

4.2.1 Cross-sectional results

We developed a cross-sectional SEM model to capture the long-run relationship between transportation and land use based on data from the reference regions. The estimated SEM model involves factors that influence VMT per capita in 2011 at the MSA level. Figure 3 plots the finalized path diagram for the SEM model that includes three types of variables: 1) key outcome variable, which is daily VMT per capita, 2) exogenous explanatory variables including rail density, population, primary road density, other road density, per capita income, and fuel price, and 3) endogenous explanatory variables including transit service frequency, passenger miles per capita, population density, and the share of commuting mode by automobile. The causal pathways between variables are represented as one-way arrows and covariances or correlations between exogenous variables are expressed as double-headed arrows. An error and residual term for each endogenous variable and outcome variable are symbolized by a circle. All variables in the SEM model were transformed by taking natural logarithms to have a normal distribution. Using the constructed SEM model, we conducted a cross-sectional analysis separately for all MSAs, cluster with the Atlanta MSA, and cluster with the San Luis Obispo MSA. After excluding outliers, we obtained a total of 302 MSAs and identified that 78 and 40 MSAs are comparable to the Atlanta MSA and the San Luis Obispo MSA, respectively.

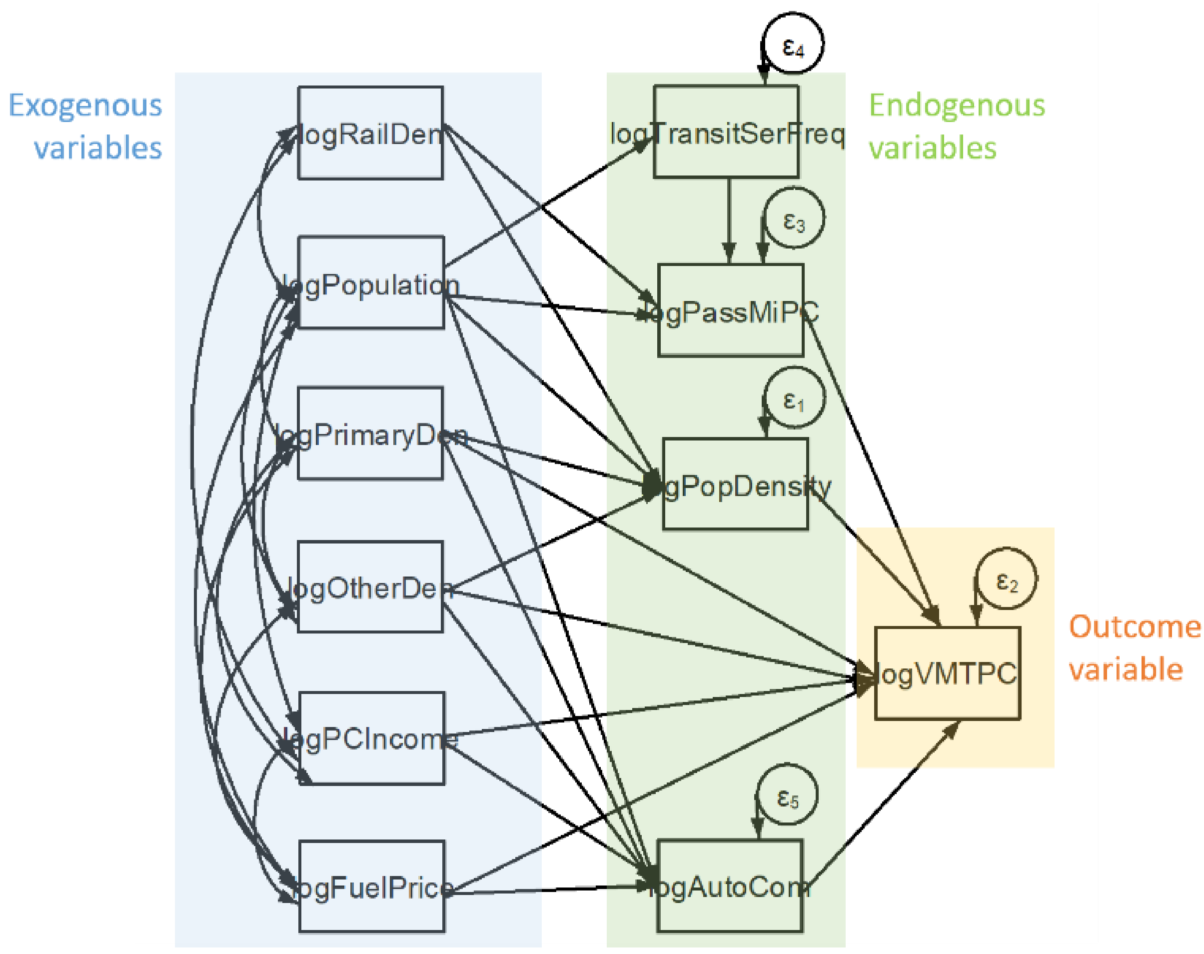


Figure 3. Path diagram for the cross-sectional SEM analysis

From the results of the cross-sectional SEM model for the entire 302 MSAs, we found that most of the causal relationships shown in the path diagram are statistically significant. Although a few paths are not statistically significant, their causal relations are theoretically significant. Table 4 presents the direct effects of one variable on another from all 302 MSAs. The regression coefficients in Table 4 indicate the elasticities of individual variables that directly affect VMT per capita, holding other factors constant. The results show that fuel price, primary road density, and per capita income are primary and exogenous drivers of VMT. These three path coefficients are statistically significant at the 0.05 probability level. Low fuel price is also strongly associated with high VMT per capita since lower fuel costs lead to more driving. According to the SEM model, a one percent decrease in fuel price is associated with a 0.61 percent increase in VMT per capita. Primary roads are generally limited-access highways within the interstate highway system or under state management. Thus, the direct influence of primary road density on VMT per capita implies that areas with more highway capacity induce more driving. The direct elasticity of 0.029 indicates that a one percent increase in primary road density leads to a 0.03 percent increase in VMT per capita. The positive association between per capita income and VMT per capita indicates that people tend to travel more by private vehicle as per capita income rises. When per capita income increases by one percent, we would expect VMT per capita to change by 0.22 percent.

Population density, the share of commuting mode by automobile, and passenger miles per capita are endogenous predictors of VMT. The direct effect of population density on VMT per capita is negative because trip destinations become closer as density rises. A one percent increases in population density is associated with a 0.04 percent decline in VMT per capita. The positive association of the share of commuting mode by automobile on VMT per capita is high, as expected. The direct elasticity of 1.020 implies that we would expect a 1.02 percent increase in VMT per capita when the share of automobile-commuting increases by one percent. Transit passenger miles per capita has a negative association with VMT per capita, although the effect is quite small. If we change transit passenger miles per capita by one percent, VMT per capita would decrease by 0.004 percent.

Table 5 presents the total effects of different variables on VMT per capita, accounting for both direct and indirect pathways for the model of all MSAs. The significant positive direct effect of per capita income on VMT per capita becomes weaker due to its negative indirect effect. Thus, a one percent increase in per capita income is associated with a 0.11 percent increase in VMT per capita. On the other hand, the direct effects of fuel price and primary road density become stronger with indirect effects. As a result, fuel price and primary road density present total elasticity of -0.739 and 0.032, respectively. While population density, the share of commuting mode by automobile, and passenger miles per capita only have direct effects on VMT per capita, population, rail density, and transit service frequency only have indirect effects on VMT per capita.

The use of the goodness-of-fit measures is recommended by previous studies, so we measured each model's goodness-of-fit using RMSEA, CFI, and SRMR with their corresponding cut-off values as follows: RMSEA below 0.08, CFI above 0.90, and SRMR below 0.08. The bottom of Table 5 presents the summary statistics of the SEM model for all 302 MSAs. The goodness-of-fil measures in Table 5 indicate that the model has a good model fit at RMSEA value of 0.08, CFI value of 0.958, and SRMR value of 0.062.

Table 4. Path coefficients estimates for direct effects in 2011 (all MSAs)

Variables			Coef.	P>|z|
log VMT Per Capita	←	log Population Density	-0.036	0.004
	←	log Commuting by Auto	1.020	0.000
	←	log Passenger Mile	-0.004	0.026
	←	log Fuel Price	-0.611	0.000
	←	log Primary Road Density	0.029	0.000
	←	log Other Road Density	-0.006	0.193
	←	log Per Capita Income	0.221	0.003
log Population Density	←	log Population	0.466	0.000
	←	log Primary Road Density	0.014	0.500
	←	log Other Road Density	0.014	0.392
	←	log Rail Density	0.527	0.000
log Commuting Mode by Auto	←	log Fuel Price	-0.126	0.000
	←	log Primary Road Density	0.003	0.007
	←	log Other Road Density	-0.002	0.105
	←	log Per Capita Income	-0.107	0.000
	←	log Population	-0.003	0.216
log Passenger Mile Per Capita	←	log Transit Service Frequency	70.672	0.000
	←	log Rail Density	0.175	0.564
	←	log Population	1.676	0.000
log Transit Service Frequency	←	log Population	0.011	0.000

Table 5. Direct, indirect, and total effects of variables on VMT per capita for all MSAs

Variables	Direct Effects	Indirect Effects	Total Effects
Population	0.000	-0.029	-0.029
Per capita income	0.221	-0.110	0.111
Average fuel price	-0.611	-0.128	-0.739
Rail density	0.000	-0.020	-0.020
Primary road density	0.029	0.003	0.032
Other roads density	-0.006	-0.002	-0.009
Population density	-0.036	0.000	-0.036
Commuting mode by auto	1.020	0.000	1.020
Transit service frequency	0.000	-0.281	-0.281
Annual passenger miles per capita	-0.004	0.000	-0.004
Summary statistics			
N	302		
χ^2	58.626		
RMSEA (< 0.05-0.08)	0.080		
CFI (> 0.90-0.95)	0.958		
SRMR (< 0.06-0.08)	0.062		

* RMSEA: root mean square error of approximation; CFI: comparative fit index; SRMR; standardized root mean square residual

The SEM model for the cluster that includes the Atlanta MSA is constructed using the hypothesized causal relationships shown in Figure 3. After removing 2 outliers, we used 78 comparable MSAs as the basis for our cross-sectional analysis. Similar to the results of the SEM model for all MSAs, we found that most of the causal relationships shown in the path diagram are statistically significant in the SEM model for the Atlanta type of cluster. Some variables become statistically insignificant in the SEM model for the Atlanta type of cluster but the general trend is very similar. Table 6 presents the elasticities of individual variables on VMT per capita, holding other factors constant. The results show that population density, passenger miles per capita, fuel price, and per capita income are statistically significant while the share of commuting mode by automobile and highway capacity become statistically insignificant. In short, fuel price and per capita income are primary and exogenous drivers of VMT, and population density and passenger miles per capita are endogenous predictors of VMT.

The direct effect of the fuel price is strongly associated with VMT per capita: a one percent decline in fuel price leads to a 1.4 percent increase in VMT per capita. The size of this effect is larger than the one in the SEM model for all MSAs. This trend indicates that MSAs in the Atlanta type cluster tend to be more sensitive to fuel price fluctuations compared to the total MSAs. Compared to other MSAs, the Atlanta type cluster has more suburban sprawl. Thus, the distance of driving or the number of trips may depend on the fuel price to a greater extent. The direct effect of per capita income on VMT per capita is positive and the size of coefficient (0.244) is similar to the one (0.221) in the SEM model for all MSAs. Thus, VMT per capita increases by 0.24 percent as per capita income increases by one percent. Similarly, the direct effects of population density (-0.125) and transit passenger miles per capita (-0.012) are larger than those in the model for all MSAs.

Table 6. Path coefficients estimates for direct effects in 2011 (Atlanta type cluster)

Variables			Coef.	P>|z|
log VMT Per Capita	←	log Population Density	-0.125	0.001
	←	log Commuting by Auto	-0.859	0.253
	←	log Passenger Mile	-0.012	0.001
	←	log Fuel Price	-1.339	0.000
	←	log Primary Road Density	0.021	0.180
	←	log Other Road Density	0.074	0.406
	←	log Per Capita Income	0.244	0.064
log Population Density	←	log Population	0.277	0.000
	←	log Primary Road Density	0.012	0.753
	←	log Other Road Density	1.038	0.000
	←	log Rail Density	0.073	0.231
log Commuting Mode by Auto	←	log Fuel Price	-0.188	0.000
	←	log Primary Road Density	0.008	0.001
	←	log Other Road Density	0.010	0.354
	←	log Per Capita Income	-0.055	0.003
	←	log Population	-0.007	0.001
log Passenger Miles Per Capita	←	log Transit Service Frequency	27.689	0.000
	←	log Rail Density	-0.236	0.731
	←	log Population	1.483	0.000
log Transit Service Frequency	←	log Population	0.008	0.197

Table 7 presents the total effects of individual variables on VMT per capita, consisting of both direct and indirect effects for the model of the Atlanta type cluster. While population density, the share of commuting mode by automobile, and passenger miles per capita have only direct effects on VMT per capita, population, rail density, and transit service frequency have only indirect effects on VMT per capita. Contrary to the SEM model for all MSAs, the positive direct effect of per capita income on VMT per capita becomes stronger due to its positive indirect effect. Thus, the total effect of per capita income on VMT per capita is 0.291 (= 0.244 + 0.047), indicating a 0.29 percent increase in VMT per capita due to a one percent increase in per capita. On the other hand, the negative direct effect of fuel price becomes less strong because of its positive indirect effect. As a result, the total effect of average fuel price is -1.178 (= -1.338 + 0.161).

The goodness-of-fit measures in Table 7 indicate that the SEM model for the Atlanta type cluster has a moderate model fit at RMSEA value of 0.089, CFI value of 0.937, and SRMR value of 0.114. Previous studies, including Hu and Bentler (1998), suggest that RMSEA tends to reject true-population models at small sample sizes below 250. Thus, the value of RMSEA (0.089), which is higher than the cut-off value (0.08), is a relatively moderate fit, considering the small sample size of the model (N=78).

Table 7. Direct, indirect, and total effects of variables on VMT per capita for Atlanta type cluster

Variables	Direct Effects	Indirect Effects	Total Effects
Population	0.000	-0.048	-0.048
Per capita income	0.244	0.047	0.291
Average fuel price	-1.339	0.161	-1.178
Rail density	0.000	-0.006	-0.006
Primary road density	0.021	-0.008	0.013
Other roads density	0.074	-0.138	-0.064
Population density	-0.125	0.000	-0.125
Commuting mode by auto	-0.859	0.000	-0.859
Transit service frequency	0.000	-0.321	-0.321
Annual passenger miles per capita	-0.012	0.000	-0.012
Summary statistics			
N	78		
χ^2	32.224		
RMSEA (< 0.05-0.08)	0.089		
CFI (> 0.90-0.95)	0.937		
SRMR (< 0.06-0.08)	0.114		

* RMSEA: root mean square error of approximation; CFI: comparative fit index; SRMR; standardized root mean square residual

For the San Luis Obispo type cluster, we conducted the cross-sectional analysis using the same SEM model. After excluding outliers and MSAs with missing data, we were left with only 40 MSAs out of 67 MSAs, which imposed significant limitations on our analysis. The goodness-of-fit measures—RMSEA, CFI, and SRMR—for this cluster do not meet the cut-off values (see Table 8). As a result, we used data from all 302 MSAs for the cross-sectional analysis to estimate the counterfactual VMT of the San Luis Obispo MSA.

Table 8. Path coefficients estimates for direct effects in 2011 (San Luis Obispo type cluster)

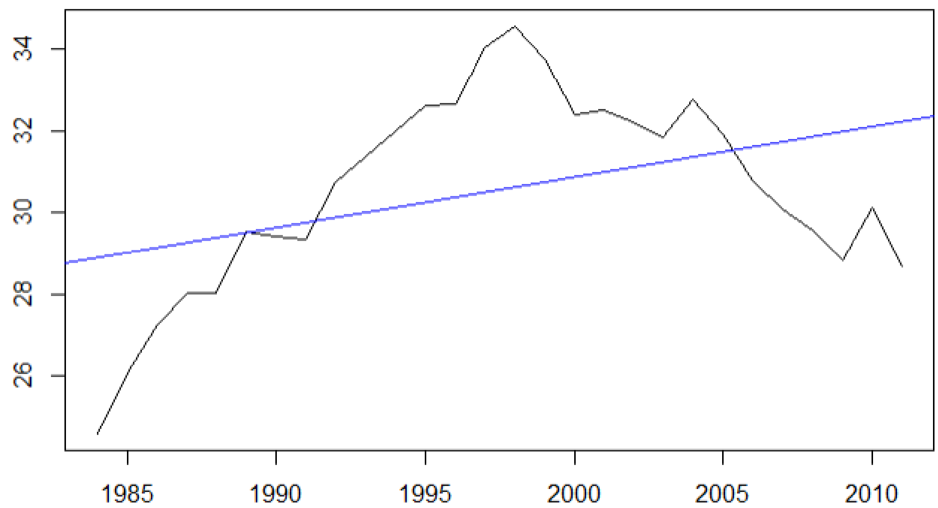
Variables			Coef.	P>|z|
log Population Density	←	log Population	0.386	0.000
	←	log Primary Road Density	0.017	0.736
	←	log Other Road Density	1.613	0.000
	←	log Rail Density	1.290	0.367
log VMT Per Capita	←	log Population Density	-0.280	0.000
	←	log Commuting by Auto	1.910	0.126
	←	log Passenger Mile	0.004	0.710
	←	log Fuel Price	-0.081	0.838
	←	log Primary Road Density	0.042	0.047
	←	log Other Road Density	0.238	0.045
	←	log Per Capita Income	0.613	0.008
log Commuting Mode by Auto	←	log Fuel Price	0.014	0.759
	←	log Primary Road Density	0.005	0.028
	←	log Other Road Density	0.009	0.439
	←	log Per Capita Income	-0.048	0.054
	←	log Population	-0.001	0.882
log Passenger Mile Per Capita	←	log Transit Service Frequency	27.784	0.000
	←	log Rail Density	4.631	0.424
	←	log Population	1.437	0.000
log Transit Service Frequency	←	log Population	-0.009	0.362
Summary statistics				
N	40			
χ^2	44.013			
RMSEA (< 0.05-0.08)	0.173			
CFI (> 0.90-0.95)	0.786			
SRMR (< 0.06-0.08)	0.157			

* RMSEA: root mean square error of approximation; CFI: comparative fit index; SRMR; standardized root mean square residual

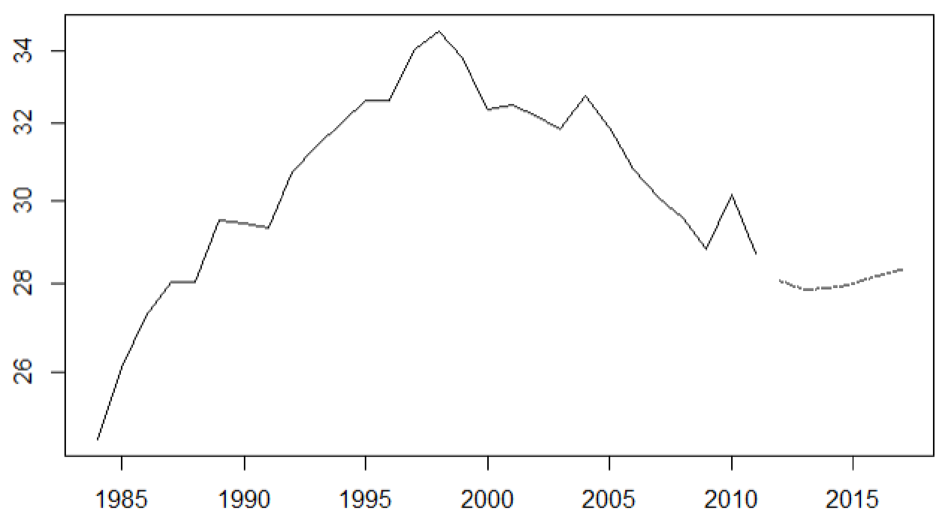
4.2.2 Longitudinal results

To estimate counterfactual VMTs for the Atlanta and San Luis Obispo MSAs, the output from the cross-sectional SEM-based approach forms the EB prior estimate that was updated using the longitudinal VMT trends for two MSAs. We employed the ARIMA model, which is frequently used in time-series analyses. The predicted VMT estimates from the ARIMA model was combined with the VMT estimates from the cross-sectional SEM model based on the respective variances of each estimate to obtain the counterfactual VMT estimates for the regions. We used the adjusted R-squared value of the ARIMA model to determine the weight of the two VMT estimates.

Figures 4 and 5 present historical and predicted VMT estimates for the Atlanta and San Luis Obispo MSAs. In the Atlanta MSA, daily VMT per capita grew by an average of 0.6 percent per year over the period between 1984 and 2011. In 1998, daily VMT per capita peaked at 34.56, but it has decreased since 1998. By 2011, VMT per capita was down to 28.70. The ARIMA model projected the expected VMT estimates for the year between 2012 and 2017 based on this observed VMT trend. The predicted values of VMT per capita in 2012 and 2017 were 28.07 and 28.33, respectively. This represents a growth of about 0.9 percent per year in the daily VMT per capita in the Atlanta MSA. The San Luis Obispo MSA, on the other hand, experienced growth in daily VMT per capita of about 0.24 percent per year over the period between 1996 and 2011. Despite this trend of persistent growth, daily VMT per capita has declined since 2007, which is when the daily VMT per capita peaked at 11.30. The result of the ARIMA model shows that VMT estimates for 2012 and 2017 are 10.62 and 10.77, respectively. This represents a growth of about 1.5 percent per year in daily VMT per capita in the San Luis Obispo MSA.

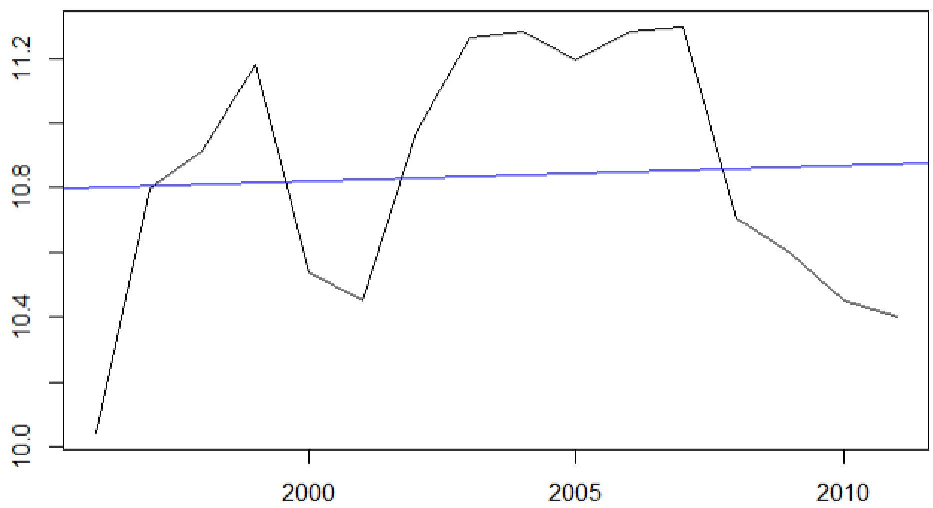


(a) Historical daily VMT per capita in the Atlanta MSA

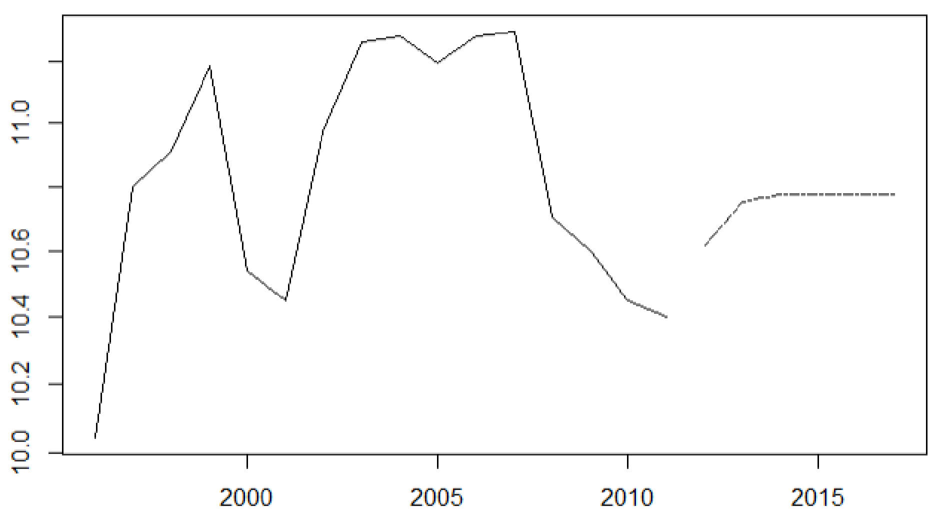


(b) Predicted daily VMT per capita in the Atlanta MSA

Figure 4. Historical and predicted daily VMT per capita in the Atlanta MSA



(a) Historical daily VMT per capita in the San Luis Obispo MSA



(b) Predicted daily VMT per capita in the San Luis Obispo MSA

Figure 5. Historical and predicted daily VMT per capita in the San Luis Obispo MSA

The results of the time-series analysis show an adjusted r-squared value of 0.157 for the Atlanta MSA and a value of 0.067 for the San Luis Obispo MSA. It should be noted that these values indicate, particularly for the San Luis Obispo region, that the time-series models don't explain the trend in VMT.

These goodness of fit values were used as a weight factor in combining the two VMT estimates: VMT estimates from the cross-sectional model and those from the longitudinal model. Table 9 presents the weighted combination of the two VMT estimates. To estimate the predicted values of counterfactual VMT for each year between 2012 and 2017, the cross-sectional SEM models for 2011 and the time-series models based on historical VMT trends were used. For the Atlanta MSA, the counterfactual VMT per capita of 25.89 in 2012 was calculated as follows: $25.48 \times (1 - 0.157) + 28.07 \times 0.157$. Similarly, the counterfactual VMT per capita for the San

Luis Obispo MSA was estimated using a weight factor of 0.067, which results in the counterfactual VMT per capita of 16.64 (= 17.08 × (1 – 0.067) + 10.62 × 0.067) in 2012. In a counterfactual scenario with no TNCs, the Atlanta MSA and the San Luis Obispo MSA were expected to have growth in VMT per capita of about 2.4 percent and 1.6 percent, respectively, over the period between 2012 and 2017.

Table 9. Counterfactual VMT estimates

Year	Atlanta			San Luis Obispo		
	VMT from cross-sectional analysis	VMT from longitudinal analysis	Counterfactual VMT	VMT from cross-sectional analysis	VMT from longitudinal analysis	Counterfactual VMT
2012	25.48	28.07	25.89	17.08	10.62	16.64
2013	25.37	27.87	25.76	17.10	10.75	16.67
2014	25.48	27.89	25.86	17.00	10.77	16.59
2015	25.52	28.01	25.91	16.78	10.77	16.38
2016	25.79	28.17	26.16	17.09	10.77	16.66
2017	26.17	28.33	26.51	17.35	10.77	16.91

4.3 Comparison between current and counterfactual VMTs

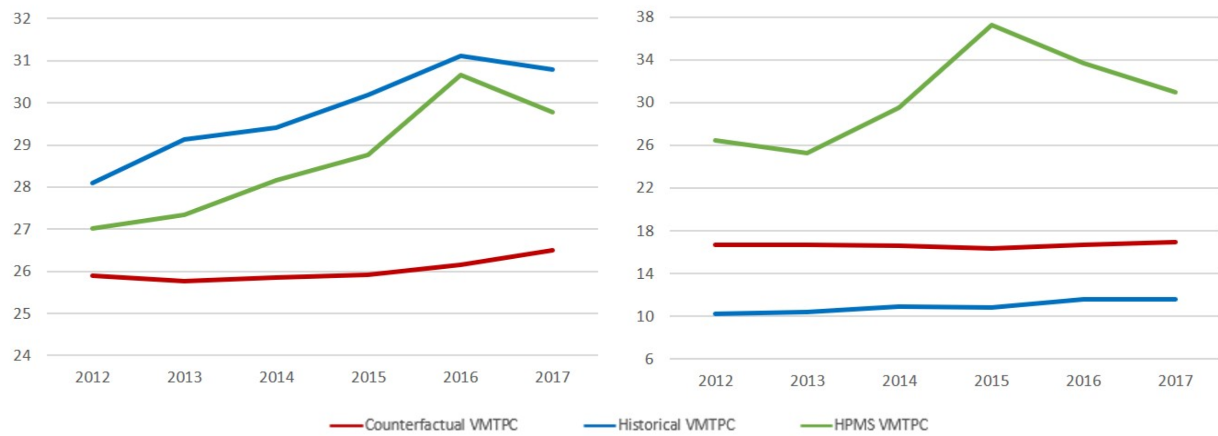
To quantify the impact of TNCs on VMT, the weighted estimate of the counterfactual VMT was compared to current VMT estimates. We evaluated the two estimates based on the following assumptions: 1) year 2011 observed conditions reflect no presence of TNCs, 2) years 2012 – 2017 counterfactual estimates reflect what the conditions would have been had there been no TNCs, and 3) years 2012 – 2017 current VMT estimates represent actual conditions in the presence of TNCs.

Table 10 shows the result of the comparison between counterfactual and current VMT per capita over the period between 2012 and 2017. For the current VMT estimates, we used two sources: 1) historical VMT per capita from GDOT and San Luis Obispo Council of Government and 2) HPMS geospatial VMT data provided by FHWA. Figure 6 depicts the comparison between counterfactual VMT per capita, historical VMT per capita, and HPMS VMT per capita. When we compared the counterfactual estimates to the two current VMT estimates for the Atlanta MSA, we found that the counterfactual VMT estimates were lower than the actual VMT estimates. The differences between the counterfactual VMT and historical VMT become larger year over year. To be specific, the difference between the counterfactual VMT per capita and the historical VMT per capita ((b)-(a)) increases from 2.20 to 4.28. Similarly, the difference between the counterfactual VMT per capita and HPMS VMT per capita ((c)-(a)) increases from 1.12 to 3.26. When we converted the absolute value of the differences to the percentage change, the diverging trend is more distinct. In 2012, the differences were 8.5 percent ((b)-(a)) and 4.1 percent ((c)-(a)), and they peaked at 18.9 percent ((b)-(a)) and 14.7 percent ((c)-(a)) in 2016. Compared to the year 2016, the differences in 2017 slightly decreased to 16.1 percent ((b)-(a)) and 11.0 percent ((c)-(a)) but these are still large. Also, the average annual growth rates between 2012 and 2017 show that the counterfactual VMT estimates have the smallest growth rate of 0.5 percent among the three VMT estimates. The results suggest that the introduction of TNCs generated more daily VMT per capita in the Atlanta MSA.

For the San Luis Obispo MSA, the results show two conflicting patterns depending on the data source used for the current VMT estimates. Due to the data discrepancy between the two currents VMT estimates, we found that the counterfactual VMT estimates were higher than the historical VMT estimates but lower than the HPMS VMT estimates over the period between 2012 and 2017. Although the differences between the two VMT estimates move in different directions, the counterfactual VMT estimates have the smallest average annual growth rate of 0.3 percent among the three VMT estimates between 2012 and 2017. This result supports the hypothesis that TNCs increased VMT per capita in the San Luis Obispo MSA.

Table 10. Counterfactual VMT estimates and current VMT estimates between 2012 and 2017

Year	Counterfactual VMT estimates (a)	Current VMT estimates		Differences	
		Historical VMT (b)	HPMS VMT (c)	(b)-(a)	(c)-(a)
Atlanta MSA					
2011	-	28.70	27.78	-	-
2012	25.89	28.09	27.01	2.20 (8.5%)	1.12 (4.1%)
2013	25.76	29.14	27.34	3.38 (13.1%)	1.58 (5.8%)
2014	25.86	29.42	28.17	3.56 (13.8%)	2.31 (8.2%)
2015	25.91	30.19	28.76	4.28 (16.5%)	2.85 (9.9%)
2016	26.16	31.11	30.66	4.95 (18.9%)	4.50 (14.7%)
2017	26.51	30.79	29.77	4.28 (16.1%)	3.26 (11.0%)
Average annual growth rate	0.5%	1.9%	2.0%	-	-
San Luis Obispo MSA					
2011	-	10.40	26.02	-	-
2012	16.64	10.22	26.45	-6.42 (-38.6%)	9.81 (37.1%)
2013	16.67	10.41	25.29	-6.26 (-37.6%)	8.62 (34.1%)
2014	16.59	10.90	29.52	-5.69 (-34.3%)	12.93 (43.8%)
2015	16.38	10.83	37.22	-5.55 (-33.9%)	20.84 (56.0%)
2016	16.66	11.61	33.67	-5.05 (-30.3%)	17.01 (50.5%)
2017	16.91	11.58	30.92	-5.33 (-31.5%)	14.01 (45.3%)
Average annual growth rate	0.3%	2.7%	3.4%	-	-



(a) Atlanta MSA (b) San Luis Obispo MSA

Figure 6. Daily VMT per capita comparison

Chapter V: Conclusions

The primary purpose of this study is to examine the impact of TNCs on VMT in Atlanta and San Luis Obispo regions and investigate whether TNCs increase or decrease daily VMT per capita. Since the TNC dataset is not publicly available, this study employed the EB approach which is a well-established method for evaluating traffic safety at a location where treatment is applied. The framework developed to quantify the impact of the introduction of TNCs on VMT changes was based on the counterfactual theory, which compares VMT estimates after the introduction of the TNCs in a region to what the VMT would have been without the TNCs. The latter is the counterfactual VMT, and the EB method was applied to estimate VMT in a counterfactual scenario. The EB method requires two models: a cross-sectional analysis using SEM from reference regions to the study area and a time-series analysis using longitudinal data from the study area to estimate counterfactual VMTs. These models were developed based on 2011 conditions when TNC activities were negligible and these conditions were used as the basis for estimating counterfactual VMT after 2011. The outputs from the two analyses were combined using a weight factor based on the variance of each model. The weighted combination of counterfactual VMT estimates was compared to the current VMT estimates to quantify the impact of TNCs.

The results show that the difference between counterfactual and current VMT per capita increased over the period between 2012 and 2017. When we compared the counterfactual and the current VMT estimates for the Atlanta MSA and the San Luis Obispo MSA, we found that the counterfactual VMT estimates were lower than the actual VMT estimates, and the differences between two VMT estimates gradually increased year over year. Also, the average annual growth rates of the counterfactual VMT estimates were lower than those of the current VMT estimates between 2012 and 2017. The average annual growth rates of the counterfactual VMT estimates in the Atlanta MSA and the San Luis Obispo MSA were 0.5 percent and 0.3 percent, respectively, while those of the current VMT estimates were ranging between 1.9 percent and 3.4 percent. The results support the hypothesis that the introduction of TNCs generated more daily VMT per capita in the Atlanta MSA and the San Luis Obispo MSA.

The research design of this study has several limitations that future research could address. First, the longitudinal analysis using the time-series model did not involve any other factors that affect VMT. The results of the univariate time-series model that uses only a single variable, which is the historical VMT trend, may not be a powerful predictor of VMT estimates. This affects the data and results, especially from the San Luis Obispo region. Therefore, the application of a multivariate time-series model that accounts for multiple variables affecting VMT may help to increase the reliability of VMT estimates. Other estimation techniques such as machine learning can also be applied to estimate and predict VMT estimates for the counterfactual scenario.

Another limitation is that our unit of analysis, which is MSA, may be too large to generalize the results in explaining VMT trends in smaller geographical boundaries such as cities and neighborhoods. However, this study formulated a framework for estimating VMTs by any geographic boundaries (e.g. urbanized area, county, and city) depending on research interests. Our proposed methodology is very robust in that it can be applied to any geographical level, which may serve the purposes of the various stakeholders who need to quantify the impact of TNCs at different levels of aggregation.

Despite these limitations, the proposed quasi-experimental EB approach to assess VMT changes after TNC introduction can also be useful for other disruptive technologies including the advent of connected and automated vehicles (CAVs). Also, a similar approach would be useful in more precisely quantifying the impact of TNCs on DUI-related crashes, which remains a relatively less researched area in the existing literature.

References

Arrington, G., & Cervero, R. (2008). *Effects of TOD on housing, parking, and travel*. (TCRP Report 128). Transportation Research Board. Retrieved from https://www.transitwiki.org/TransitWiki/images/5/51/TCRP_TOD_Report.pdf

Brazil, N., & Kirk, D. S. (2016). Uber and metropolitan traffic fatalities in the United States. *American journal of epidemiology*, *184*(3), 192-198.

Boarnet, M., & Crane, R. (2001). The influence of land use on travel behavior: Specification and estimation strategies. *Transportation Research Part A: Policy and Practice, 35*(9), 823-845. doi:10.1177/0739456X9801700302

Brons, M., Nijkamp, P., Pels, E., & Rietveld, P. (2008). A meta-analysis of the price elasticity of gasoline demand. A SUR approach. *Energy Economics*, *30*(5), 2105-2122.

Handy, S., & Boarnet, M. G. (2014). Impact of Highway Capacity and Induced Travel on Passenger Vehicle Use and Greenhouse Gas Emissions. *California Environmental Protection Agency, Air Resources Board, Retrieved August*, *28*, 2015.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, *21*(2), 230-258.

Caulfield, B. (2009). Estimating the environmental benefits of ride-sharing: A case study of Dublin. *Transportation Research Part D: Transport and Environment*, *14*(7), 527-531.

Cervero. R. (2002). Induced travel demand: Research design, empirical evidence, and normative policies. *Journal of Planning Literature*, 17(1): 3–20.

Cervero, R., & Murakami, J. (2010). Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and planning A*, *42*(2), 400-418.

Choi, K., & Zhang, M. (2017). The net effects of the built environment on household vehicle emissions: A case study of Austin, TX. *Transportation Research Part D: Transport and Environment*, *50*, 254-268.

Choo, S., Mokhtarian, P. L., & Salomon, I. (2005). Does telecommuting reduce vehicle-miles traveled? An aggregate time series analysis for the US. *Transportation*, 32(1), 37-64.

Dills, A. K., & Mulholland, S. E. (2018). Ride-Sharing, Fatal Crashes, and Crime. *Southern Economic Journal*, *84*(4), 965-991.

Ding, C., Wang, Y., Xie, B., & Liu, C. (2014). Understanding the role of built environment in reducing vehicle miles traveled accounting for spatial heterogeneity. *Sustainability*, *6*(2), 589-601.

DuPuis, N., Martin, C., & Rainwater, B. (2015). City of the Future: Technology and Mobility. Center for City Solutions and Applied Research, National League of Cities, 31.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion?. *Science advances*, *5*(5), eaau2670.

Ewing, R. (2008). Highway-induced development: research results for metropolitan areas. *Transportation Research Record*, *2067*(1), 101-109.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, *76*(3), 265-294.

Ewing, R., Hamidi, S., Gallivan, F., Nelson, A. C., & Grace, J. B. (2014). Structural equation models of VMT growth in US urbanised areas. *Urban Studies*, 51(14), 3079-3096.

Federal Highway Administration (2016). Highway Performance Monitoring System Field Manual, Federal Highway Administration: Washington D. C.

Espey. M. (1998). Gasoline demand revisited: An international meta-analysis of elasticities. *Energy Economics* 20: 273–295.

Fricker, J. D., & Kumapley, R. K. (2002). *Updating procedures to estimate and forecast vehicle-miles traveled* (Joint Transportation Research Program Publication No. FHWA /IN/JTRP-2002/10). West Lafayette, IN: Purdue University. http://dx.doi.org/10.5703/1288284313337

Garceau, T. J. (2015). A Multi-Scalar Model to Identify the Causes of Decreased Vehicle Miles Traveled (VMT) in the United States.

Greenwood, B. N., & Wattal, S. (2015). Show me the way to go home: An empirical investigation of ride sharing and alcohol related motor vehicle homicide. Fox School of Business Research Paper No. 15-054. Retrieved from https://ssrn.com/abstract=2557612 or http://dx.doi.org/10.2139/ssrn.2557612

Goodwin, P., Dargay, J., & Hanly, M. (2004). Elasticities of road traffic and fuel consumption with respect to price and income: a review. *Transport reviews*, *24*(3), 275-292.

Grace, J. B. (2006). *Structural equation modeling and natural systems*. Cambridge: Cambridge University Press.

Graham, D. J., & Glaister, S. (2004). Road traffic demand elasticity estimates: a review. *Transport reviews*, *24*(3), 261-274.

Greenblatt, J. B., & Shaheen, S. (2015). Automated vehicles, on-demand mobility, and environmental impacts. *Current sustainable/renewable energy reports*, *2*(3), 74-81.

Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, *37*(1), 1-25.

Hale, C. (2014). TOD versus TAD: The great debate resolved…(?). *Planning Practice & Research, 29*(5), 492-507. doi:10.1080/02697459.2012.749056 Hall, Palsson, and Price 2018

Hall, J. D., Palsson, C., & Price, J. (2018). Is Uber a substitute or complement for public transit?. *Journal of Urban Economics*, *108*, 36-50.

Hauer, E. (1997). *Observational before-after studies in road safety.* Pergamon/Elsevier Science, Inc., Tarrytown, NY.

Hauer, E., Harwood, D., Council, F., & Griffith, M. (2002). Estimating safety by the empirical B ayes method: a tutorial. Transportation Research Record: *Journal of the Transportation Research Board*, (1784), 126-131.

Henao, A., & Marshall, W. E. (2019). The impact of ride-hailing on vehicle miles traveled. *Transportation*, *46*(6), 2173-2194.

Hoffmann, K., Ipeirotis, P., & Sundararajan, A. (2016). Ridesharing and the Use of Public Transportation. Proceedings of the 37th International Conference on Information Systems (ICIS), Dublin, Ireland.

Hoyle, R. H. (2011). *Structural equation modeling for social and personality psychology*. SAGE Publications Ltd.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*(4), 424.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Kim, S., Park, D., Heo, T. Y., Kim, H., & Hong, D. (2016). Estimating vehicle miles traveled (VMT) in urban areas using regression kriging. *Journal of Advanced Transportation*, *50*(5), 769-785.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kumapley, R. K., & Fricker, J. D. (1996). Review of methods for estimating vehicle-miles traveled. *Transportation Research Record*, 1551, 59–66.

Laham, M. L., & Noland, R. B. (2017). Nonwork trips associated with transit-oriented development. *Transportation Research Record: Journal of the Transportation Research Board*(2606), 46-53. doi:10.3141/2606-06

Langlois, M., van Lierop, D., Wasfi, R. A., & El-Geneidy, A. M. (2015). Chasing sustainability: Do new transit-oriented development residents adopt more sustainable modes of transportation? *Transportation Research Record: Journal of the Transportation Research Board*(2531), 83-92. doi:10.3141/2531-10

Leard, B., Linn, J., & Munnings, C. (2016). Explaining the evolution of passenger vehicle miles traveled in the United States. Washington, DC: Resources for the Future. Retrieved from http://www.rff.org/files/document/file/DP-16-38.pdf

Li, Z., Hong, Y., & Zhang, Z. (2016, September 25). An Empirical Analysis of On-Demand Ride Sharing and Traffic Congestion. Proc. International Conference on Information Systems 2016 (ICIS'16). Retrieved from https://ssrn.com/abstract=2843301

Liu, F., & Kaiser, R. (2006). *Forecasting truck VMT growth at the county and statewide levels*. Linthicum, MD: Michael Baker Jr., Inc.

Lu, Z., Zhang, H., Southworth, F., & Crittenden, J. (2016). Fractal dimensions of metropolitan area road networks and the impacts on the urban built environment. *Ecological indicators*, *70*, 285-296.

Martin-Buck, F. (2017). Driving Safety: An Empirical Analysis of Ridesharing's Impact on Drunk Driving and Alcohol-Related Crime. *University of Texas at Austin Working Paper*.

Retrieved from http://www.frankmartinbuck.com/Ridesharing%20and%20Alcohol-Related%20Crime%20by%20Frank%20Martin-Buck.pdf

McMullen, B. S., & Eckstein, N. (2013, September). Determinants of VMT in urban areas: A panel study of 87 US Urban Areas 1982-2009. In *Journal of the Transportation Research Forum*, 52(3).

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological methods*, *7*(1), 64.

Nasri, A., & Zhang, L. (2014). The analysis of transit-oriented development (TOD) in Washington, DC and Baltimore metropolitan areas. *Transport policy, 32*, 172-179. doi:10.1016/j.tranpol.2013.12.009

Newmark, G. L., Haas, P. M., Pappas, J., Schwartz, M., Kenyon, A., & Unit, M. O. (2015). Income, location efficiency, and VMT: Affordable housing as a climate strategy. *Center for neighborhood technology working paper, produced for the california housing partnership*.

Park, K., Ewing, R., Scheer, B. C., & Tian, G. (2018). The impacts of built environment characteristics of rail station areas on household travel behavior. *Cities, 74*, 277-283. doi:10.1016/j.cities.2017.12.015

Rodier, C., Alemi, F., & Smith, D. (2016). Dynamic ridesharing: Exploration of potential for reduction in vehicle miles traveled. *Transportation Research Record*, *2542*(1), 120-126.

Salon, D., Boarnet, M. G., Handy, S., Spears, S., & Tal, G. (2012). How do local actions affect VMT? A critical review of the empirical evidence. *Transportation research part D: transport and environment*, *17*(7), 495-508.

Soltani-Sobh, A., Heaslip, K., Bosworth, R., & Song, Z. (2016). An Aggregate Time Series Analysis to Investigate Effect of Compressed Natural Gas (CNG) Vehicles on Vehicle Miles Traveled (VMT). *Int. J. Mech. Eng. Autom*, *3*(10), 399-413.

Small, K. A., & Van Dender, K. (2007). Fuel efficiency and motor vehicle travel: the declining rebound effect. *The Energy Journal*, 25-51.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. In *the annual meeting of the Psychometric Society.* Iowa City, IA.

Heres-Del-Valle, D., & Niemeier, D. (2011). CO2 emissions: Are land-use changes enough for California to reduce VMT? Specification of a two-part model with instrumental variables. *Transportation Research Part B: Methodological*, *45*(1), 150-161.

Wallsten, S. (2015). The competitive effects of the sharing economy: how is Uber changing taxis. *Technology Policy Institute*, *22*.

Wang, S. G., Bai, L., & Bao, Y. (2011, January). Modeling and algorithms of VMT and AADT estimation for community area traffic networks. In *IFAC World Congress 2011*.

Yu, R., & Abdel-Aty, M. (2014). Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accident Analysis & Prevention*, *62*, 161-167.

Zhang, L., Hong, J., Nasri, A., & Shen, Q. (2012). How built environment affects travel behavior: A comparative analysis of the connections between land use and vehicle miles traveled in US cities. *Journal of transport and land use*, *5*(3), 40-52.

Zhou, H., Zhao, J., Hsu, P., & Huang, J. (2013). Safety effects of median treatments using longitudinal channelizers: Empirical bayesian before-and-after study. *Journal of transportation engineering*, *139*(12), 1149-1155.

CENTER FOR TRANSPORTATION
EQUITY, DECISIONS & DOLLARS

The **Center for Transportation, Equity, Decisions and Dollars (CTEDD)** is a USDOT University Transportation Center, leading **transportation policy research** that aids in **decision making** and **improves economic development** through **more efficient, and cost-effective use of existing transportation systems,** and offers **better access to jobs and opportunities.** We are leading a larger consortium of universities focused on **providing outreach** and **research to policy makers**, through **innovative methods** and **educating future leaders of the transportation field.**

UNIVERSITY OF TEXAS ARLINGTON

CAL POLY SAN LUIS OBISPO

Georgia Tech

USF UNIVERSITY OF SOUTH FLORIDA

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

C-TEDD@UTA.EDU | CTEDD.UTA.EDU | @UTACTEDD | @C__TEDD | @ CENTER FOR TRANSPORTATION, EQUITY, DECISIONS AND DOLLARS | 817.272.5138