Project 2315 | August 2024

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Forecasting Commercial Vehicle Miles Traveled (VMT) in Urban California Areas

Steve Chung, PhD Jaymin Kwon, PhD Yushin Ahn, PhD

CSUTC California State University Transportation Consortium

FRESNO STATE Discovery. Diversity. Distinction.

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) and the Mineta Consortium for Equitable, Efficient, and Sustainable Transportation (MCEEST) funded by the U.S. Department of Transportation, the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1 and the Climate Change and Extreme Events Training and Research (CCEETR) Program funded by the Federal Railroad Administration. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 24-25

Forecasting Commercial Vehicle Miles Traveled (VMT) in Urban California Areas

Steve Chung, PhD

Jaymin Kwon, PhD

Yushin Ahn, PhD

August 2024

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
24-25		

4. Title and Subtitle	5. Report Date
Forecasting Commercial Vehicle Miles Traveled (VMT) in Urban California Areas	August 2024
	6. Performing Organization Code
7. Authors Steve Chung, PhD ORCID: 0000-0001-7255-7244 Jaymin Kwon, PhD ORCID: 0000-0002-4365-6902 Yushin Ahn, PhD ORCID: 0000-0002-5305-6396	**8. Performing Organization Report** CA-MTI-2315
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	**10. Work Unit No.**
	11. Contract or Grant No. ZSB12017-SJAUX
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Floor Long Beach, CA 90802	**13. Type of Report and Period Covered**
	14. Sponsoring Agency Code

15. Supplemental Notes
10.31979/mti.2024.2315

16. Abstract
This study investigates commercial truck vehicle miles traveled (VMT) across six diverse California counties from 2000 to 2020. The counties—Imperial, Los Angeles, Riverside, San Bernardino, San Diego, and San Francisco—represent a broad spectrum of California's demographics, economies, and landscapes. Using a rich dataset spanning demographics, economics, and pollution variables, we aim to understand the factors influencing commercial VMT. We first visually represent the geographic distribution of the counties, highlighting their unique characteristics. Linear regression models, particularly the least absolute shrinkage and selection operator (LASSO) and elastic net regressions are employed to identify key predictors of total commercial VMT. LASSO regression emphasizes feature selection, revealing vehicle population and fuel consumption as significant predictors in most counties. Elastic net regression, which balances feature selection and multicollinearity, expands the list of predictors to include variables like the number of trips, CO_2 emissions, and $PM_{2.5}$ pollution. Overall, the findings suggest that economic factors, such as fuel consumption and vehicle population, significantly impact the total commercial VMT across the counties. Pollution variables, specifically CO_2 and $PM_{2.5}$, also play a role. These insights underscore the need for nuanced transportation and environmental policies, especially in the face of economic fluctuations, to manage commercial truck VMT effectively and sustainably. Methodology using both LASSO and elastic net regression provides a robust framework for understanding these complex relationships in commercial transportation behavior.

17. Key Words	18. Distribution Statement
Vehicle miles of travel, Regression analysis, Forecasting, Predictive models, Urban areas	No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	50	

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2024.2315

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2315

Acknowledgments

The authors express their gratitude to the Mineta Transportation Institute for their support and assistance during this project. Special appreciation is extended to Dr. Hilary Nixon for her invaluable guidance and expertise, which proved instrumental in resolving queries and overcoming challenges encountered along the way. Additionally, heartfelt thanks are extended to Christine Casey and Julia Kingsley from the State Assembly Transportation Committee for their insightful feedback and recommendations, which greatly enriched the quality of our work. We also acknowledge Editing Press for their exceptional editorial services, ensuring the clarity and coherence of our manuscript. Furthermore, our sincere appreciation goes to the MTI staff, including Executive Director Karen Philbrick, PhD, and Deputy Executive Director Hilary Nixon, PhD, for their continued support and encouragement throughout this endeavor.

Contents

List of Figures

List of Tables

Executive Summary

This study focuses on six diverse California counties: Imperial, Los Angeles, Riverside, San Bernardino, San Diego, and San Francisco, each representing a unique blend of demographics, landscapes, and socioeconomic characteristics. The dataset spans from 2000 to 2020, capturing a comprehensive array of variables falling into three main categories: demographics, economics, and pollution. The outcome variable of interest is total vehicle miles traveled (VMT) by commercial vehicles primarily used for transportation purposes.

Key Findings

- Total VMT Trends: Total commercial VMT tends to be at its lowest around 2000, with a notable decline in the early 2010s likely due to the lasting effects of the 2008 global financial crisis and fluctuating fuel prices.

- Correlation Analysis: Total VMT demonstrates positive correlations with trips, CO_2 emissions, NO_2 emissions, and fuel consumption. It shows negative correlations with unemployment and poverty across all counties, indicating higher total VMT is associated with increased trips and emissions but lower unemployment and poverty rates.

- LASSO Regression:

  - A smaller subset of variables was selected, with vehicle population and fuel consumption showing significant positive impacts on total VMT in multiple counties.

  - Pollution variables CO_2 and $PM_{2.5}$ were selected, while economic factors such as median income and number of employed were chosen in specific counties.

- Elastic Net Regression:

  - Elastic net regression included a greater number of variables for each county, highlighting the balance between feature selection and managing multicollinearity.

  - Vehicle population, number of trips, and fuel consumption were consistently selected across all counties.

  - CO_2 and $PM_{2.5}$ pollution variables were chosen for all counties, while other pollution variables were not selected uniformly.

  - Economic factors like the number of employed and the house price index were included in multiple counties.

Implications

- Economic Influence: Total commercial VMT appears heavily influenced by economic factors, especially during the aftermath of the 2008 financial crisis.

- Policy Insights: Understanding the impact of economic conditions, pollution, and other factors on total VMT is crucial for informing transportation and environmental policies.

- Variable Importance: Vehicle population, fuel consumption, and pollution variables (CO_2 and $PM_{2.5}$) emerge as key predictors of total commercial VMT.

Recommendations

- Policy Adaptations: Policymakers should consider economic fluctuations when designing transportation policies, especially for commercial vehicle operations.

- Environmental Considerations: Strategies to reduce CO_2 and $PM_{2.5}$ emissions from commercial vehicles could have significant impacts on total VMT.

- Further Research: Future studies could explore additional factors influencing total commercial VMT and consider longitudinal effects for a more comprehensive understanding.

Conclusion

In conclusion, this study provides valuable insights into the dynamics of total commercial VMT across diverse California counties. By analyzing correlations, LASSO and elastic net regression results, and historical trends, we uncover key predictors and trends influencing commercial vehicle travel. These findings can inform policymakers, urban planners, and environmental agencies in developing strategies for sustainable transportation and economic resilience. For detailed analyses and specific findings, refer to the complete study.

1. Introduction

Vehicle miles traveled (VMT) serves as a fundamental cornerstone in the analysis of transportation dynamics, offering invaluable insights into the utilization and effectiveness of transportation infrastructures. With its comprehensive scope, VMT encapsulates the cumulative distance covered by all vehicles across specific regions, enabling policymakers and transportation stakeholders to discern patterns, identify challenges, and formulate strategic interventions (Federal Highway Administration, n.d.). Given its pivotal role, VMT has been the focal point of sustained scrutiny and examination over time, reflecting its criticality in gauging the evolving landscape of transportation systems. In line with projections by the Federal Highway Administration (FHWA), the trajectory of national VMT is poised to witness a 22% surge from 2019 to 2049 (Federal Highway Administration, 2020). This anticipated growth trajectory underscores the multifaceted drivers propelling the expansion of VMT, with diverse vehicle categories making substantive contributions to this upward trend.

Among the components of VMT, light-duty vehicles represent the largest portion of vehicular travel, encompassing various personal vehicles used for daily commuting, errands, and recreational purposes. Following closely are single-unit trucks, which include delivery vans, service vehicles, and small freight carriers, contributing substantially to urban logistics and last-mile transportation. Additionally, combination trucks, comprising tractor-trailers and heavy-duty freight vehicles, play a pivotal role in long-haul freight transportation, facilitating the movement of goods across vast distances. The Federal Highway Administration's forecasts project a 17% increase in light-duty vehicle VMT by 2049, reflecting ongoing societal trends such as population growth, urbanization, and economic expansion. In contrast, single-unit truck VMT is anticipated to surge by 101%, underscoring the vital role of local and regional freight networks in supporting supply chains and commerce. Similarly, combination truck VMT is projected to rise by 57% over the same period, driven by increased demand for goods movement and intermodal transportation solutions. This anticipated growth in commercial VMT highlights the importance of investigating the primary factors influencing transportation patterns and infrastructure utilization, as it has significant implications for urban mobility, environmental sustainability, and overall economic development. Effective transportation planning and policy development are essential to address the challenges and opportunities associated with this anticipated surge in vehicular travel, ensuring that infrastructure investments align with evolving mobility needs, environmental objectives, and societal priorities.

Understanding the factors driving the growth in commercial VMT is crucial for developing sustainable and efficient transportation strategies. This paper aims to delve into the key determinants of commercial VMT growth, analyzing factors such as economic trends, demographic shifts, and variables related to pollution. Several attempts have been made to identify factors that affect VMT. Brownstone and Golob (2009) examined the relationship between residential density, vehicle usage, and energy consumption. Their study explores how variations in residential density influence travel behavior and energy usage patterns. Findings suggest that higher residential density is associated with reduced vehicle usage and lower energy consumption, highlighting the potential role of urban planning policies in promoting sustainable transportation practices. McMullen and Eckstein (2012) examined the relationship between vehicle miles traveled (VMT) and economic activity in the United States, questioning the conventional wisdom that economic growth leads to increased VMT. Using time series techniques, they empirically test for Granger causality between VMT and various measures of national economic activity. The findings suggest that, in most cases, the causal relationship is from economic activity to VMT, indicating that exogenous shocks to VMT are unlikely to affect national GDP negatively, though the relationship varies across different stages of the business cycle and is not significant for urban areas. Woldeamanuel and Kent (2014) examined the determinants of per capita vehicle miles traveled (VMT) in California. The findings suggest that factors such as income, fuel prices, vehicle ownership, and land use significantly influence VMT trends in the state. Newmark et al. (2015) investigated the relationship between income, location efficiency, and vehicle miles traveled (VMT), focusing on affordable housing as a climate strategy. The study examines how housing location influences transportation behavior and its implications for VMT patterns. They suggest that affordable housing located in transit-rich areas can contribute to reduced VMT, highlighting the potential role of housing policies in addressing climate change through transportation-related emissions reductions. Loder, Tanner, and Axhausen (2017) investigated the impact of local work and residential balance on Vehicle Miles Traveled (VMT) using a new direct approach. Their methodology involves analyzing detailed GPS data to measure VMT for individuals in Switzerland directly. The results show that a balanced distribution of work and residential locations can significantly reduce VMT, indicating the importance of spatial planning policies in mitigating transportation-related environmental impacts.

In addition to exploring the relationship between commercial VMT and other factors, our focus extends to forecasting commercial VMT. Kumapley and Fricker (1996) examined the importance of VMT estimates in transportation planning, emphasizing the necessity for objective comparisons between different estimation methods. Their review of methods used by the Indiana Department of Transportation (INDOT) compares statewide VMT estimates derived from INDOT's traffic count-based method with those from a non-traffic-data cross-classification VMT estimation model, revealing a tendency for INDOT's traffic count-based estimates to be 10 to 20 percent higher than estimates from the cross-classification model. Polzin et al. (2004) aimed to forecast future vehicle miles of travel in the United States, employing a methodology that involves analyzing transportation data and trends to develop predictive models. Their findings indicate an upward trend in vehicle miles of travel over the forecasted period, underscoring the importance of infrastructure planning and transportation policy development to accommodate increasing travel demand. Additionally, Williams et al. (2016) explored diverse methodologies for estimating and forecasting VMT, including statistical modeling, travel demand models, and data-driven approaches. Their findings provide the necessity of tailored methodologies to enhance VMT prediction accuracy for effective transportation planning and policy decisions.

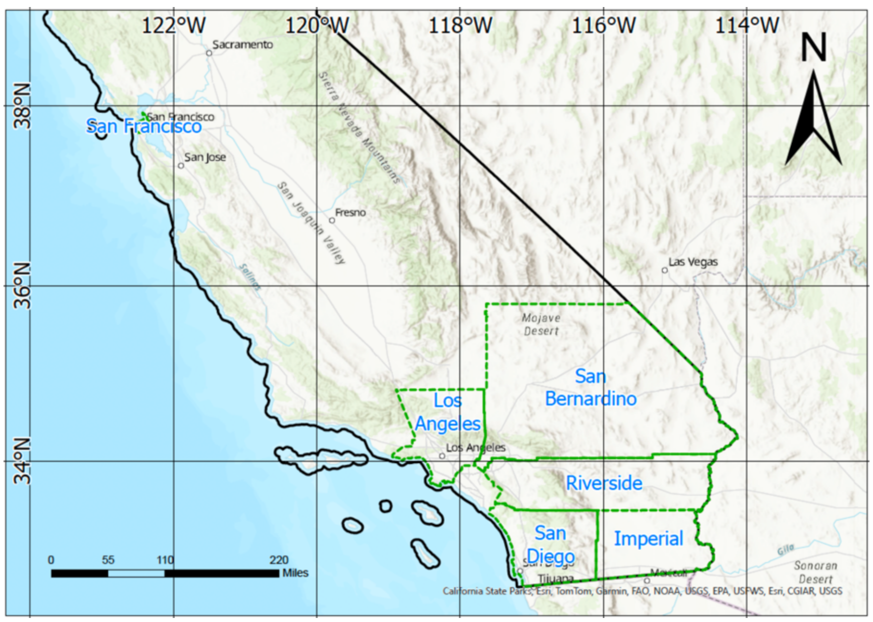
The structure of this paper is outlined as follows. Section 2 provides a detailed exposition of the data and methodology, elucidating the data sources, input-output variables, and summarizing their respective statistics. Subsequently, the paper delves into a discussion on the modeling approach, commencing with multiple linear regression and addressing any encountered limitations. Section 3 elucidates the findings derived from LASSO regression and elastic net regression, accompanied by an evaluation of forecasting capabilities. And Section 4 presents concluding remarks.

2. Data and Methodology

2.1 Data and Data Collection

In this study, we focus on six diverse California counties: Imperial, Los Angeles, Riverside, San Bernardino, San Diego, and San Francisco, each with its unique blend of demographics, landscapes, and socioeconomic characteristics. These counties represent a broad spectrum of California's geographical and cultural diversity, from the bustling metropolis of Los Angeles to San Diego's coastal regions, and from the agricultural heartland of Imperial County to the urban center of San Francisco. By examining these counties together, we aim to capture the multifaceted nature of California's communities and better understand the relationships among VMT, demographics, economics, and pollution. Figure 1 provides a visual representation of the geographic distribution of these counties, offering a glimpse into the vast and varied terrain they cover across the state.

Figure 1. A Map Showing all Imperial, Los Angeles, Riverside, San Bernardino, San Diego, and San Francisco Counties



The following provides some brief characteristics of each county in Figure 1.

1. Imperial County:

- Located in the southeastern part of California, bordering Mexico
- Known for its agriculture, particularly in the production of lettuce, carrots, and other vegetables

- Desert climate with hot summers and mild winters
- Lower population density compared to other counties

2. Los Angeles County:

- The most populous county in the United States, with a diverse population
- Home to the city of Los Angeles, known for its entertainment industry (Hollywood), beaches, and cultural diversity
- Diverse economy with sectors including entertainment, technology, finance, aerospace, and tourism
- Varied geography ranges from beaches to mountains to urban areas

3. Riverside County:

- Located east of Los Angeles County, it is one of the fastest-growing counties in California
- Known for its desert landscape, but also includes urban areas like Riverside and Palm Springs
- Economy includes agriculture, tourism (especially around Palm Springs), and healthcare
- Warm climate with hot summers and mild winters

4. San Bernardino County:

- The largest county in the United States by area
- Diverse geography with deserts, mountains (including part of the San Bernardino Mountains), and valleys
- Economically diverse with industries such as logistics, manufacturing, healthcare, and tourism
- Includes cities like San Bernardino and Ontario, as well as parts of the Mojave Desert

5. San Diego County:

- Located in the southwestern corner of California, bordering Mexico

- Known for its mild climate, beautiful beaches, and outdoor recreational opportunities
- Home to the city of San Diego, a major center for biotechnology and defense industries
- Strong military presence with military bases like Naval Base San Diego and Marine Corps Base Camp Pendleton

6. San Francisco County:

- Located in northern California, it is a compact county that includes the city of San Francisco
- Known for its iconic landmarks such as the Golden Gate Bridge, Alcatraz Island, and Fisherman's Wharf
- Diverse economy with a focus on technology (Silicon Valley is nearby), tourism, finance, and healthcare
- High cost of living, particularly in the city of San Francisco, but also offers a rich cultural scene and diverse population

These counties collectively represent a wide range of characteristics, from bustling urban centers to rural agricultural areas, and from desert landscapes to coastal regions. Each county contributes uniquely to California's diverse economy, culture, and landscape. The dataset used in this study has been sourced from two primary sources: https://fred.stlouisfed.org and https://arb.ca.gov/emfac/.

Spanning the extensive timeframe from 2000 to 2020, the dataset comprises a comprehensive array of variables falling into three main categories: demographics, economics, and pollution variables. The demographics category encompasses a range of population-related metrics, offering insights into the size, composition, and distribution of communities within the studied counties. Within the economics category, various economic indicators have been gathered, shedding light on factors such as employment rates, income levels, and economic activities prevalent in these regions over the two-decade period. Lastly, the pollution variables capture data on environmental factors, including air quality, emissions, and other pollution-related metrics sourced from the California Air Resources Board. This rich dataset provides a multifaceted view of the dynamic interplay between demographics, economic dynamics, and environmental factors within the selected counties, offering a valuable resource for our analysis and exploration of the trends and patterns across the years. Table 1 provides the variable names and their descriptions.

Table 1. Variable Descriptions

Variables	**Descriptions**
Median Income	Median household income (in Dollars)
Unemployment	Unemployment rate (in %)
Employed	Employed Person (10,000's)
GDP	County Gross Domestic Product (in 1,000,000 Dollars)
House Price Index	All-Transaction House Price Index
SNAP	Supplemental Nutrition Assistance Program Benefit Recipients
Poverty	Estimate of People of All Ages in Poverty
Population	Resident Population (in 1,000)
Premature	Age-Adjusted Premature Death Rate (Rate per 100,00)
Vehicle Population	County Vehicle Populations
Trips	Number of Trips
NOx	Nitrogen Oxides
$PM_{2.5}$	Particulate matter 2.5 with diameter 2.5 microns or less
PM_{10}	Particulate matter 10 with diameter 10 microns or less
CO_2	Carbon Dioxide
NO_2	Nitrogen Dioxide
CO	Carbon Monoxide
SO_x	Sulfur Dioxide
Total VMT	Total Vehicle Miles Traveled

Table 2. Summary Statistics of the Variables

	Imperial Mean (Std Dev.)	LA Mean (Std Dev.)	Riverside Mean (Std Dev.)	SB Mean (Std Dev.)	SD Mean (Std Dev.)	SF Mean (Std Dev.)
County Pop	167.35(13.71)	9890.43(143.7)	2142.38(264.04)	2024.71(124.81)	3109.86(172)	824.07(39.22)
MedIncome	38517.3(5171.5)	55019.7(9786.1)	56011.5(8958.3)	52885.5(7473.6)	63331.1(11220.9)	78654.2(23118.0)
Unemploymt	21.51(4.97)	7.63(2.9)	7.89(3.21)	7.59(3.21)	6.10(2.62)	5.47(2.12)
Employed	5.48(0.38)	454(18.84)	87.2(11.40)	81.5(6.05)	143(5.56)	46.6(5.90)
GDP	6.46(1.60)	572(121)	65.0(168)	68.8(16.2)	179(38.2)	114(32.4)
Poverty	34891.2(4800.42)	1599259(169982)	297845.9(66878.5)	328084.55(54825.43	372518(63651)	92432(12287.66)
SNAP	29981(11015.2)	913117(2335907)	186389(94208)	268405(109656)	186438(89060)	41974(12008)
Premature	314.93(61.06)	290.88(28.4)	332.38(32.65)	349.39(43.3)	279.9(16.58)	313.34(32.44)
House Price	149.03(36.05)	200.28(49.46)	173.5(45.16)	184.81(48.61)	179.27(37.57)	185.059(49.8)
Vehicle	16762.36(1715.94)	316782.38(24255.41)	85431.73(11929.06)	103323.56(12433.73	118388.04(11996.87)	11808.76(748.67)
Trips	248126.5(24868.8)	4488793.77(331079.9)	1224863.1(167823.9)	1485832.8(171787.3	1636303.9(154277.4)	158379.2(9712.8)
NOx	12.62(5.44)	132.46(50.56)	62.15(26.21)	69.71(29.49)	42.46(14.91)	3.59(1.15)
$PM_{2.5}$	0.34(0.17)	3.37(1.75)	1.69(0.83)	1.81(0.89)	1.06(0.48)	0.084(0.038)
PM_{10}	0.35(0.17)	3.52(1.83)	1.77(0.87)	1.89(0.93)	1.11(0.5)	0.087(0.0399)
CO_2	1758.76(112.14)	22782.63(1481.59)	8728.57(619.77)	10194.44(679.15)	7809.24(544.73)	781.09(49.74)
NO_2	0.25(0.017)	2.95(0.18)	1.26(0.092)	1.46(0.096)	0.96(0.057)	0.091(0.0058)
CO	5.57(2.69)	79.88(38.97)	20.17(9.07)	27.19(12.16)	25.38(11.51)	2.63(1.28)
SO_x	0.051(0.055)	0.57(0.56)	0.25(0.27)	0.29(0.31)	0.195(0.19)	0.019(0.018)
Total VMT	1184033(81406)	16199054(118237)	5906822.2 (513502.7)	6889779(54967)	5552359.5(488700.4)	498712.7(32774)

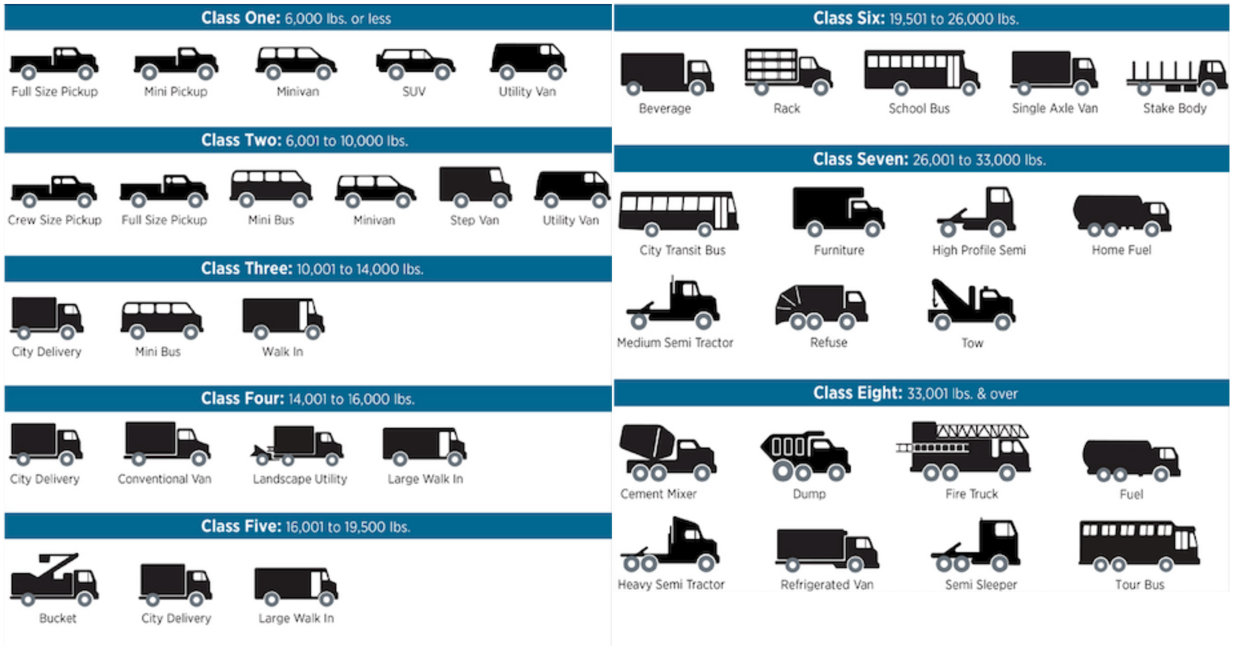
The outcome variable in our study is total VMT, which represents the total vehicle miles traveled by commercial truck vehicles primarily used for transportation purposes. In accordance with the California Vehicle Code CVC §260, a commercial vehicle is one that is required to be registered and is used for transporting persons for hire, compensation, or profit, as well as for the transportation of property. Specifically, a motor truck is defined as a motor vehicle designed, used, or maintained primarily for transporting property.

While our dataset does not explicitly designate vehicles as commercial or private, we have classified them based on their Vehicle Class. For this study, we have focused on vehicles with a Vehicle Class of 4 or higher, excluding passenger cars, light-duty trucks with a gross vehicle weight rating (GVWR) less than 6000 lbs, buses, motor homes, and motorcycles. The commercial truck vehicles considered in our analysis fall into the following categories:

- Medium-duty vehicles (GVWR 6000–8500 lbs)
- Light-heavy duty trucks (GVWR 8501–10,000 lbs)
- Light-heavy duty trucks (GVWR 10,001–14,000 lbs)
- Medium-heavy duty trucks (GVWR 14,001–33,000 lbs)
- Heavy-heavy duty trucks (GVWR greater than 33,000 lbs)

By focusing on these categories, we aim to capture the VMT of trucks primarily used for commercial purposes, which play a significant role in transporting goods and services across the studied California counties. This classification allows us to analyze and understand the trends and patterns in the travel behavior of these commercial truck vehicles over the period from 2000 to 2020, providing valuable insights into their impact on transportation systems and the environment. Figure 2 provides examples of trucks in different truck classes.

Figure 2. Examples of Trucks in Truck Class



(Source: https://afdc.energy.gov/data)

Figure 3 presents correlation plots illustrating the relationships between the input variables and the outcome variable, total VMT, across the six selected California counties. These plots are color-coded, with blue indicating a positive correlation and red indicating a negative correlation, providing a visual representation of the linear relationships between variable pairs. Notably, the pollution variables (excluding CO_2 and NO_2) exhibit negative correlations with the economics and demographics variables across all counties. This may suggest that as pollution levels increase, there tends to be a decrease in economic and demographic indicators such as income levels, population density, and employment rates. However, the implication can also be the opposite.

Furthermore, the analysis reveals that total VMT demonstrates positive correlations with the number of trips, CO_2 emissions, NO_2 emissions, and fuel consumption. This may imply that as total VMT increases, there is a corresponding rise in the number of trips taken, as well as in CO_2 and NO_2 emissions and fuel consumption. In our regression models, we consider VMT as the outcome variable and emissions as independent variables. It helps to understand how changes in the independent variables are associated with the changes in the outcome variable; but it only shows correlation, not causation. Conversely, total VMT shows negative correlations with unemployment and poverty, indicating that higher levels of total VMT are associated with lower unemployment rates and poverty levels across the counties.

For a broader perspective, correlation plots (Figure 6) for the entire dataset have also been included for comparison. These plots provide insights into the overall trends and relationships among the variables across the entire dataset spanning from 2000 to 2020. By examining these correlations, we can gain a deeper understanding of how various factors, such as pollution, economic indicators, and demographics, are interrelated and how they influence total VMT in the context of the selected California counties. These visualizations serve as valuable tools for exploring the complex dynamics at play within these regions and their implications for transportation and environmental policies.

Figure 3. Correlation Plots Correlation Plots of the Variables Among Trucks

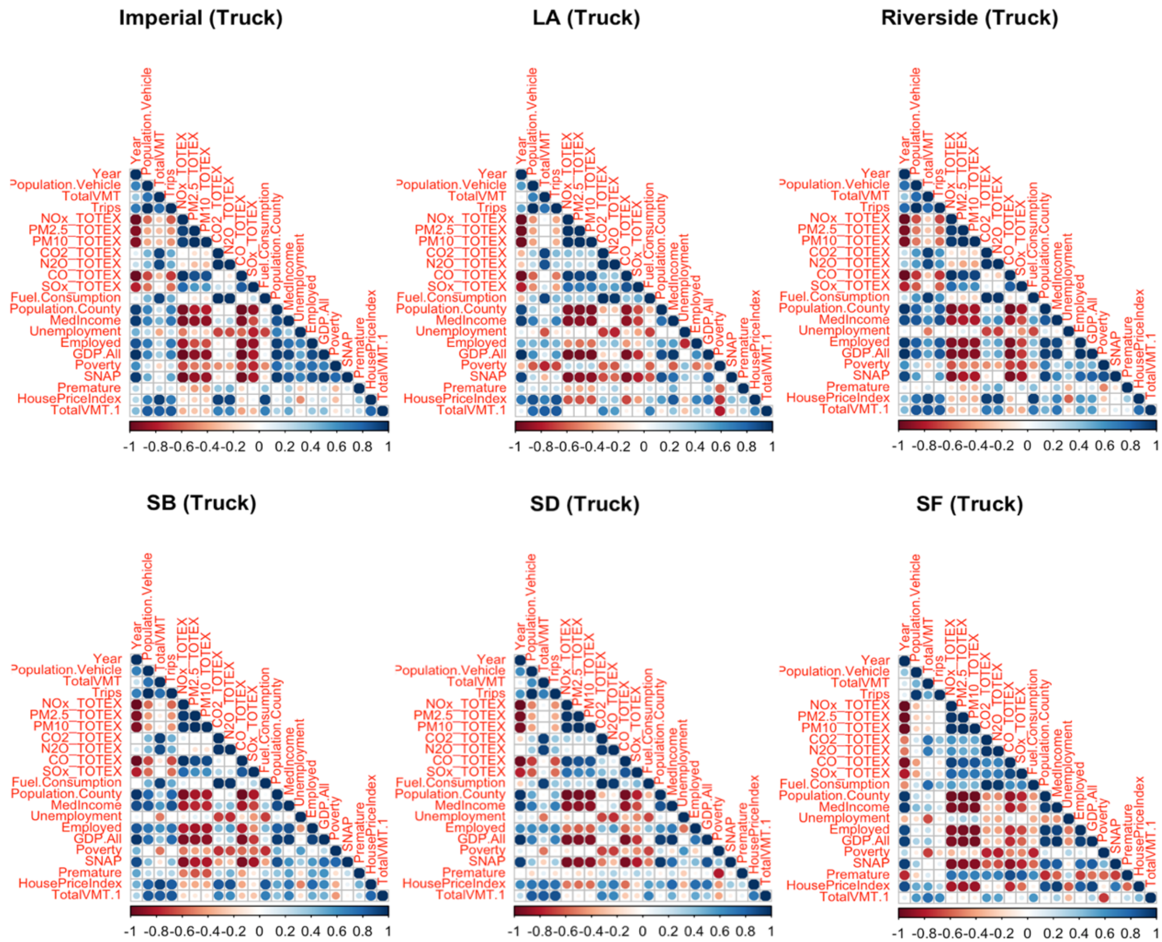


Figure 4 presents line plots depicting the total commercial VMT for each of the six California counties over the period from 2000 to 2020. A notable trend in the plots is that VMT tends to be at its lowest around the year 2000, with the exception of San Francisco County. However, across all counties, there is a sharp decrease in VMT around the early 2010s. This decline can likely be attributed to the lasting effects of the 2008 global financial crisis, which significantly impacted the economy.

During the aftermath of the financial crisis, high unemployment rates and slow economic recovery led to reduced consumer spending, including on transportation. This cautious approach to spending could have translated into fewer trips being taken and subsequently reduced VMT among commercial vehicles. Additionally, the economy's gradual recovery during this period may have influenced people to be more mindful of their spending, especially on transportation-related expenses. Another factor to consider is the volatility of fuel prices during the early 2010s. Fluctuating fuel costs can have a substantial impact on driving habits, with higher fuel prices often prompting individuals and businesses to cut down on unnecessary trips or find ways to save on fuel expenses.

Comparing the plots in Figure 4 with the correlation plots for the entire dataset in Figure 5, it becomes evident that total VMT for commercial vehicles differs significantly from the overall VMT trends (whole data). The commercial VMT appears to be more heavily influenced by economic factors, particularly the financial crisis, highlighting the vulnerability of commercial transportation to economic downturns. This insight underscores the importance of understanding how economic conditions can shape transportation behaviors and emphasizes the need for adaptive policies and strategies to mitigate the impacts of such fluctuations on commercial VMT and the broader economy.

Figure 4. Line Plot of VMT for Commercial Trucks

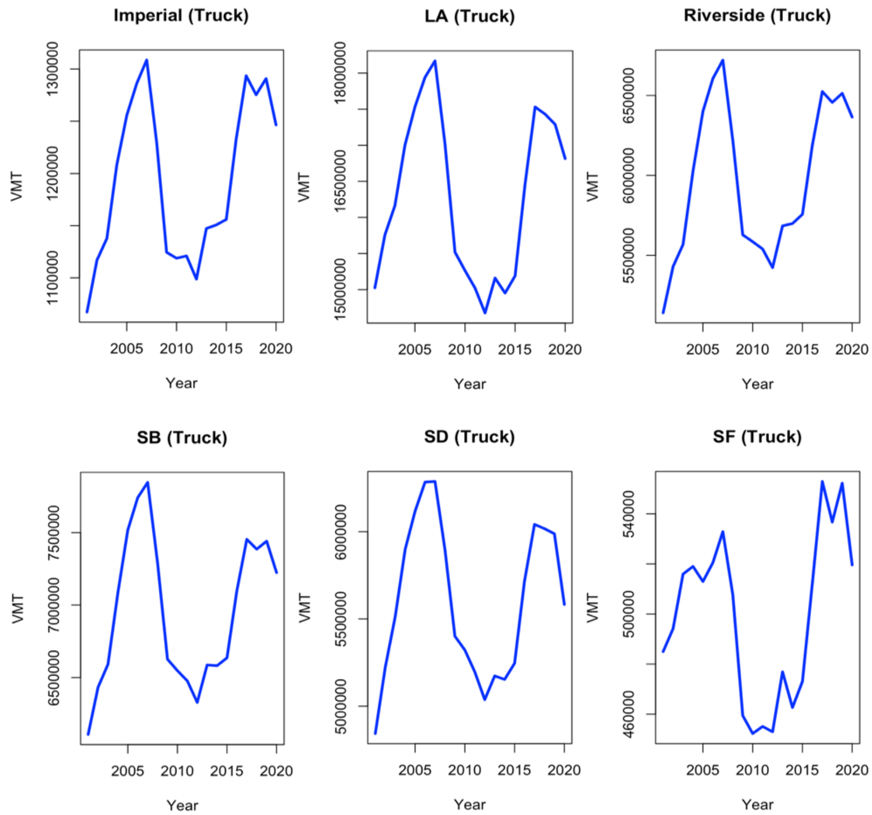


Figure 5. Line Plot of VMT for the Whole Data

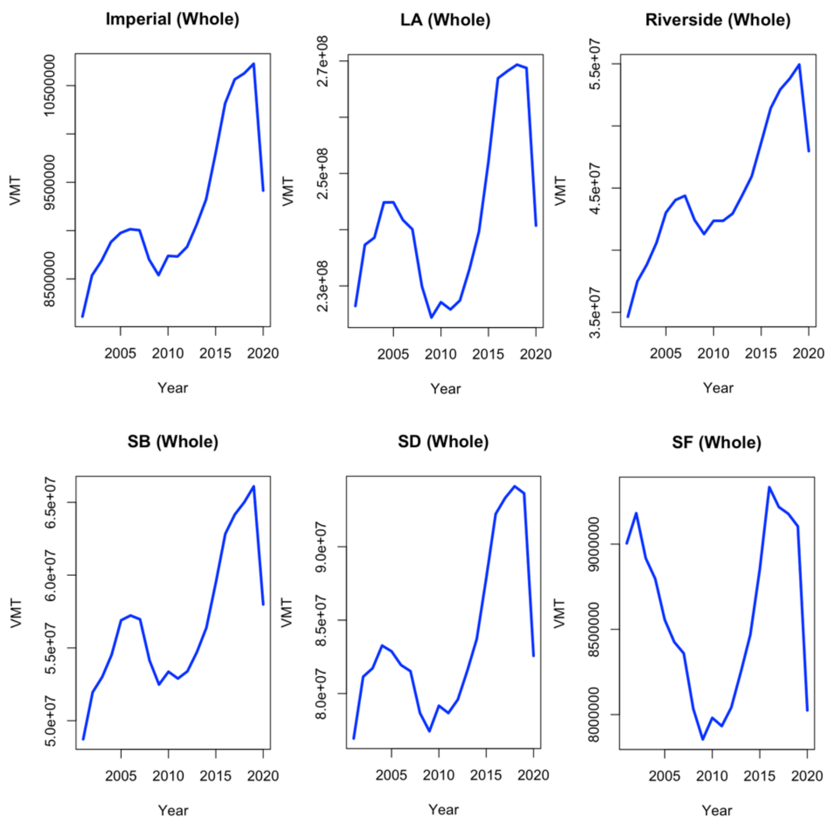
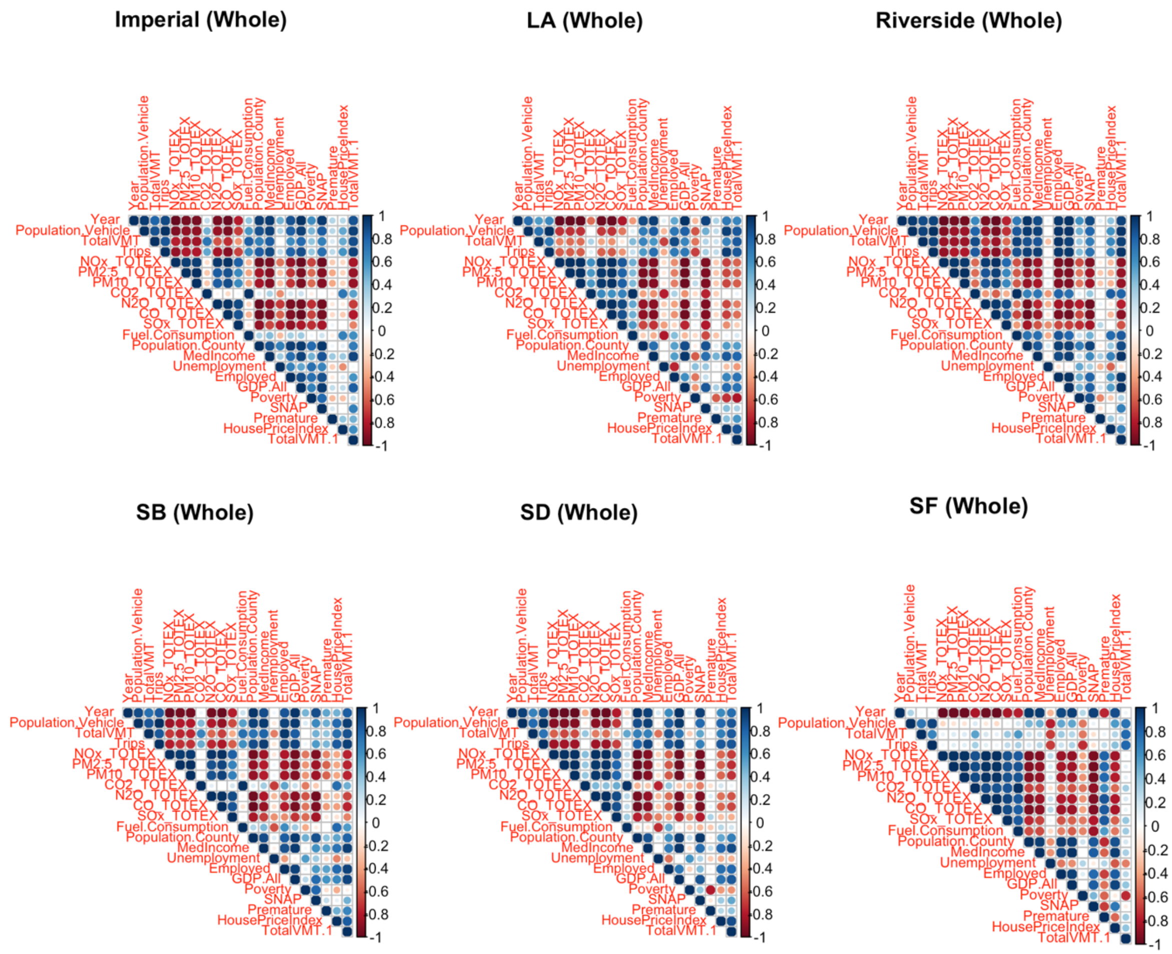


Figure 6. Correlation Plots for the Whole Dataset



2.2 Methodology

Linear regression is a versatile and widely applied statistical approach crucial for understanding relationships within data. It stands as a foundational method in data analysis due to its simplicity and interpretability. The core purpose of linear regression is to model and quantify the linear association between one or more independent variables (x) and a dependent variable (y). This relationship is expressed through a linear equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i, \text{ for i = 1,…, n.}$$

Here, y is the outcome variable, β_0 is the intercept, $\beta_1, \beta_2, \dots, \beta_k$ are the coefficients representing the effect of each input variables $x_1, x_2, \dots, x_k$, and ϵ is the error term. The list of input variables was presented in Table 1. In the context of predicting the commercial VMT, linear regression can help quantify the impact of socioeconomic factors like poverty and median income on VMT. While median income may be influenced by other factors, such as geographical location, our primary interest lies in its effect on commercial VMT. Moreover, median income can be considered an independent variable because it reflects the overall economic status of an area, which impacts the demand for goods and services, and subsequently, commercial VMT.

To estimate the parameters, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$, the ordinary least square (OLS) estimation method is often used. This method provides what is called the Best Linear Unbiased Estimator (BLUE). Many of the software packages provide parameter estimates based on the OLS method. Under the OLS method, we attempt to minimize the objective function.

$$\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})]^2.$$

However, this approach encounters difficulties when the number of observations (n) is smaller than the number of input variables (k). This is because the inverse matrix under the OLS does not exist when n $\leq$ k, presenting a challenge in estimating the parameters effectively. In our dataset, this becomes particularly problematic as we have n = 21 years and k = 21 variables, resulting in an equal number of variables and observations. As a result, the parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ cannot be reliably estimated using the ordinary least squares (OLS) method.

To address this challenge, a penalty term can be introduced into the objective function as follows:

$$\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})]^2 + \left[\alpha \sum_{j=1}^{k} |\beta_j| + (1-\alpha) \sum_{j=1}^{k} \beta_j^2\right].$$

Here, α is the regularization parameter. When α = 1, this method is known as the least absolute shrinkage and selection operator (LASSO) regression, and when α = 0, it is referred to as the "ridge regression." Any value of α between 0 and 1 is often considered the elastic net regression. This penalty term offers a solution for our dataset, where the number of input variables matches the number of observations. In our analysis, we explore two cases: α = 1 (LASSO regression) and α = 0.5 (elastic net regression). LASSO regression helps to prevent overfitting by encouraging the model to select only the most relevant features for prediction. Elastic net regression, on the other hand, combines the strengths of LASSO and ridge regression. It is particularly valuable when we want to perform feature selection and regularization simultaneously, enhancing the model's interpretability and robustness.

In our study, we utilize the total commercial VMT (y) as the dependent variable, while the other input variables (x's) are considered predictors. Additionally, we incorporate the lagged total VMT variable, which examines whether the previous year's total VMT has any influence on the current total VMT. This allows us to explore potential temporal dependencies in the data, providing insights into how past VMT levels may affect current VMT patterns.

3. Analyses and Results

In this section, we discuss the analyses and findings derived from the LASSO and elastic net regression methods. Prior to conducting these analyses, all variables underwent standardization using the z-score method, expressed as:

$$z = \frac{x - mean}{standard\ deviation}.$$

This standardization process ensures that all variables are on a comparable scale, facilitating the comparison of the impact of parameter estimates on the outcome variable across the different input variables. By standardizing the data, we are able to assess the relative importance of each input variable in influencing the total commercial VMT (outcome variable). The z-score transformation provides a standardized representation of each variable's deviation from its mean in terms of standard deviations. This normalization is particularly beneficial when working with regression models, as it places all variables on a common scale. Consequently, the coefficients obtained from the LASSO and elastic net regression analyses can be directly compared to gauge the strength and direction of their impact on the total commercial VMT.

In our analysis, the parameter estimates obtained from these regression techniques reveal which variables have the most significant effect on total commercial VMT, allowing us to identify key predictors that contribute to variations in commercial vehicle miles traveled. This standardized approach ensures a fair and meaningful comparison of the coefficients, aiding in the interpretation of the results and the identification of influential factors in the transportation patterns across the selected California counties.

3.1 LASSO Regression

As discussed in the preceding section, LASSO regression introduces a penalty term to the conventional linear regression model's sum of squared residuals. This penalty term is the absolute value of the coefficients, which results in some coefficients being shrunk to precisely zero. This aspect of LASSO regression facilitates variable selection by effectively eliminating less influential predictors from the model. By shrinking coefficients, LASSO regression imposes a constraint on the estimates, aiding in the prevention of overfitting and promoting the development of simpler and more interpretable models.

Tables 2 and 3 display the variables that have been chosen by the LASSO regression and elastic net regression, respectively. Since LASSO regression shrinks the parameter estimates to zero, eliminating the variables, it provides a smaller number of input variables than the elastic net regression. The p-values presented in the table are derived from fitting a multiple linear regression model based on this subset of selected variables. These p-values indicate the significance of each term within the linear model, providing insights into which variables are statistically significant predictors of the total commercial VMT.

In Table 2, it is evident that vehicle population has a positive influence on total commercial VMT in four counties. Similarly, fuel consumption demonstrates a positive impact on this outcome across all six counties. Specifically, vehicle population was a selected variable in four counties, with statistical significance found in three of them. On the other hand, fuel consumption was selected in all six counties, with statistical significance present in four of them. Regarding the pollution variables (NOx, $PM_{2.5}$, PM_{10}, CO_2, NO_2, CO, and SOx), the model only selected CO_2 and $PM_{2.5}$, and the remaining variables were not included. Among the economic factors, median income was chosen in three counties, the number of employed was selected in four counties, and poverty was included in two counties.

Table 3. Output from the LASSO Regression

Imperial	**Variable**	**Estimate**	**Std Error**	**t value**	**Pr(>|t|)**	
	Population	2.33E-01	5.30E-02	4.387	0.000735	***
	CO_2	-2.38E-01	1.60E+00	-0.149	0.883907	
	Fuel	1.09E+00	1.61E+00	0.679	0.508936	
	MedIncome	8.10E-02	4.98E-02	1.626	0.127939	
	Employed	9.91E-02	7.34E-02	1.35	0.20017	
	Premature	-1.19E-01	3.82E-02	-3.121	0.008118	**
Los Angeles	Variable	Estimate	Std Error	t value	Pr(>|t|)	
	Trips	2.73E-01	7.44E-02	3.661	0.00326	**
	Fuel	9.20E-01	8.66E-02	10.631	1.84E-07	***
	MedIncome	3.56E-01	1.33E-01	2.676	0.02019	*
	Employed	6.78E-02	5.80E-02	1.168	0.26553	
	Poverty	8.18E-02	4.90E-02	1.669	0.12091	
	Premature	2.96E-02	5.57E-02	0.532	0.60436	
	House Price	-1.99E-01	8.19E-02	-2.432	0.03161	*
Riverside	Variable	Estimate	Std Error	t value	Pr(>|t|)	
	Population	3.32E-01	2.54E-02	13.046	3.18E-09	***
	$PM_{2.5}$	2.53E-04	2.44E-02	0.01	0.9919	
	Fuel	7.87E-01	2.85E-02	27.62	1.30E-13	***
	Poverty	-6.59E-02	3.09E-02	-2.131	0.0513	.
	Premature	-1.39E-01	2.12E-02	-6.557	1.28E-05	***
San Bernardino	Variable	Estimate	Std Error	t value	Pr(>|t|)	
	Population	3.67E-01	2.62E-02	14.036	1.22E-09	***
	CO_2	-3.42E-01	9.54E-01	-0.358	0.725622	
	Fuel.Consumpti	1.12E+00	9.62E-01	1.162	0.264849	
	Employed	3.52E-02	2.83E-02	1.243	0.234336	
	Premature	-9.62E-02	1.97E-02	-4.885	0.000241	***

San Diego	Variable	Estimate	Std Error	t value	Pr(>|t|)	
	Population	1.07E+00	8.13E-01	1.317	0.2104	
	Trips	-8.24E-01	8.16E-01	-1.011	0.3305	
	Fuel	8.33E-01	6.65E-02	12.53	1.24E-08	***
	Employed	1.42E-02	3.39E-02	0.419	0.682	
	Premature	-3.57E-02	3.07E-02	-1.163	0.2658	
	House Price	1.35E-01	4.76E-02	2.827	0.0143	*
San Francisco	Variable	Estimate	Std Error	t value	Pr(>|t|)	
	Trips	6.03E-02	3.69E-02	1.634	0.13676	
	$PM_{2.5}$	-4.28E-01	1.77E-01	-2.425	0.03827	*
	Fuel	1.07E+00	5.53E-02	19.431	1.17E-08	***
	Population	4.88E-02	1.97E-01	0.247	0.81023	
	MedIncome	-7.40E-02	1.96E-01	-0.378	0.71412	
	Unemployment	1.65E-01	5.28E-02	3.12	0.01232	*
	GDP All	3.22E-01	1.56E-01	2.064	0.06906	.
	SNAP	-2.93E-01	7.11E-02	-4.117	0.00261	**
	Premature	-2.20E-01	5.24E-02	-4.19	0.00234	**
	House Price	1.57E-01	9.69E-02	1.62	0.13963	

3.2 Elastic Net Regression

Elastic net regression represents a variant of linear regression that merges the regularization penalties of both LASSO and ridge techniques. This method is specifically crafted to address scenarios characterized by a multitude of features, some of which may exhibit correlations with each other. Elastic net regression stands out as a potent approach for linear regression when dealing with datasets encompassing numerous features that are potentially correlated among themselves. By blending the characteristics of LASSO and ridge regularization, the elastic net provides a balance between feature selection and managing multicollinearity. Elastic net regression offers a comprehensive solution by simultaneously performing feature selection and mitigating the effects of multicollinearity. By finding the optimal balance between these two aspects, the elastic net enables the creation of robust and interpretable models, making it a valuable tool for uncovering meaningful insights from complex datasets.

Table 3 displays the outcomes of the elastic net regression, which has resulted in the inclusion of a greater number of variables for each county compared to the results in Table 2. This increase in selected variables can be attributed to the characteristic of elastic net regression, which does not force parameter estimates to shrink toward zero, as LASSO regression does. Specifically, in this table, the variable representing vehicle population was included for five counties, the number of trips for all six counties, and fuel consumption for all six counties. Moreover, both CO_2 and/or $PM_{2.5}$ pollution variables were chosen for all six counties, while other pollution variables were not selected, aligning with the findings from LASSO regression. Among the economic factors, the number of employed individuals was included for four counties, and the house price index was selected for three counties.

Table 4. Output from the Elastic Net Regression

Imperial	**Variable**	**Estimate**	**Std Error**	**t vale**	**Pr(>|t|)**	
	Population Vehicle	3.03E+00	5.08E-01	5.974	9.27E-05	***
	Trips	-3.09E+00	5.61E-01	-5.512	0.000183	***
	CO_2	-9.03E-03	9.28E-01	-0.01	0.992417	
	Fuel Consumption	1.05E+00	9.32E-01	1.122	0.285626	
	MedIncome	2.02E-01	5.01E-02	4.028	0.001988	**
	Employed	5.25E-02	6.10E-02	0.861	0.407647	
	GDP.All	2.21E-01	7.75E-02	2.85	0.015802	*
	Premature	-9.71E-02	2.33E-02	-4.165	0.001576	**
Los Angeles	Variable	Estimate	Std Error	t vale	Pr(>|t|)	
	Population Vehicle	-3.57E-02	1.02E+00	-0.035	0.9725	
	Trips	2.31E-01	1.06E+00	0.218	0.8313	
	CO_2	-1.61E+00	1.07E+00	-1.512	0.1564	
	Fuel Consumption	2.31E+00	1.10E+00	2.101	0.0574	.
	Employed	6.58E-02	4.08E-02	1.611	0.1331	
	Poverty	-9.91E-02	7.69E-02	-1.289	0.2217	
	House Price Index	1.19E-01	7.70E-02	1.547	0.1478	
Riverside	Variable	Estimate	Std Error	t vale	Pr(>|t|)	
	Population Vehicle	2.55E+00	1.03E+00	2.463	0.0315	*
	Trips	-2.33E+00	1.08E+00	-2.148	0.05486	.
	$PM_{2.5}$	-1.10E-01	6.73E-02	-1.638	0.1296	
	CO_2	-5.22E-01	9.42E-01	-0.554	0.59069	
	Fuel Consumption	1.40E+00	9.59E-01	1.455	0.17357	
	Poverty	-4.06E-03	6.43E-02	-0.063	0.95079	
	SNAP	-3.77E-02	9.93E-02	-0.38	0.71128	
	Premature	-1.01E-01	2.68E-02	-3.75	0.00321	**

San Bernardino	Variable	Estimate	Std Error	t vale	Pr(>|t|)	
	Population Vehicle	1.97E+00	1.10E+00	1.783	0.098	.
	Trips	-1.65E+00	1.14E+00	-1.45	0.1708	
	CO_2	-3.56E-01	9.18E-01	-0.388	0.7043	
	Fuel Consumption	1.17E+00	9.27E-01	1.262	0.229	
	Employed	7.27E-02	3.76E-02	1.934	0.0751	.
	Premature	-7.37E-02	2.45E-02	-3.011	0.01	*
San Diego	Variable	Estimate	Std Error	t vale	Pr(>|t|)	
	Population Vehicle	1.79E+00	6.61E-01	2.707	0.019058	*
	Trips	-1.55E+00	6.63E-01	-2.332	0.037945	*
	CO_2	-1.51E+00	4.72E-01	-3.2	0.007635	**
	Fuel Consumption	2.39E+00	4.89E-01	4.886	0.000374	***
	Employed	2.42E-02	2.61E-02	0.926	0.372466	
	Premature	-8.28E-03	2.50E-02	-0.332	0.745792	
	House Price Index	1.22E-01	3.66E-02	3.341	0.005881	**
San Francisco	Variable	Estimate	Std Error	t vale	Pr(>|t|)	
	Trips	5.54E-02	2.74E-02	2.024	0.077586	.
	$PM_{2.5}$	-2.17E-01	1.49E-01	-1.45	0.185233	
	CO_2	-1.49E+00	5.11E-01	-2.906	0.019718	*
	Fuel Consumption	2.47E+00	4.81E-01	5.127	0.000899	***
	Population County	2.78E-01	1.66E-01	1.674	0.132572	
	MedIncome	1.71E-02	1.48E-01	0.116	0.910797	
	Unemployment	1.63E-01	3.90E-02	4.178	0.003087	**
	GDP All	1.30E-01	1.33E-01	0.983	0.354392	
	SNAP	-3.15E-01	5.32E-02	-5.927	0.000351	***
	Premature	-1.21E-01	5.16E-02	-2.341	0.047354	*
	House Price Index	2.26E-01	7.55E-02	2.991	0.017297	

3.3 Forecasting Capabilities

In order to assess the forecasting capability of each model, we have used adjusted R^2. R-squared (R^2) is a statistical measure that represents the proportion of the variance in the dependent variable (the variable one is trying to predict) that is explained by the independent variables (the variables used for prediction) in a regression model. In simple terms, it tells us how well the independent variables explain the variability of the dependent variable. The formula for R^2 is as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}},$$

Here, SS_{res} is the sum of squared residuals, which measures the differences between the actual values and the predicted values by the model, and SS_{total} is the total sum of squares, which measures the total variability in the dependent variable.

R^2 ranges from 0 to 1. A higher R^2 indicates that the independent variables explain a larger proportion of the variance in the dependent variable. For example, an R^2 of 0.80 means that the independent variables in the model explain 80% of the variability in the dependent variable. However, R^2 has a limitation. It tends to increase as more independent variables are added to the model, even if they are not improving the model's performance. This can lead to overfitting, where the model fits too closely to the training data but performs poorly on new, unseen data.

Adjusted R-squared (adjusted R^2) addresses this issue by penalizing the addition of unnecessary variables to the model. It considers the number of independent variables in the model and adjusts R^2 accordingly. The formula for adjusted R2 is as follows:

$$adjusted\ R^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1},$$

where n is the number of observations and k is the number of independent variables in the model. Adjusted R^2 ranges from $-\infty$ to 1. It will be lower than R^2 when unnecessary variables are added to the model, and it will be equal to R^2 when all variables are necessary and contribute to the model's performance.

Table 5. R^2 and Adjusted R^2 Values from LASSO Regression

LASSO		Imperial	Los Angeles	Riverside	San Bernardino	San Diego	San Francisco
Truck	R^2	0.9941	0.9964	0.9979	0.9966	0.9959	0.9975
	Adjusted R^2	0.9914	0.9942	0.9971	0.9953	0.994	0.9948
Whole	R^2	0.9986	0.999	0.9939	0.9934	0.9967	0.9865
	Adjusted R^2	0.9978	0.9976	0.9927	0.991	0.9947	0.9828

Table 6. R^2 and Adjusted R^2 Values from Elastic Net Regression

Elastic Net		Imperial	Los Angeles	Riverside	San Bernardino	San Diego	San Francisco
Truck	R^2	0.9984	0.9925	0.9985	0.9971	0.9978	0.9988
	Adjusted R^2	0.9973	0.9881	0.9975	0.9957	0.9965	0.9972
Whole	R^2	0.9986	0.9985	0.9942	0.9977	0.9979	0.9924
	Adjusted R^2	0.9978	0.9968	0.9916	0.996	0.9955	0.9868

Tables 5 and 6 display the outcomes of the LASSO and elastic net regression models, respectively. These models incorporate multiple input variables, and to gauge their performance, we focus on the adjusted R^2 metric. Adjusted R^2 penalizes the model as the number of input variables increases, providing a more robust measure of goodness of fit. Upon examination of both models, it is evident that all counties' models exhibit adjusted R^2 values exceeding 0.98. This indicates that the selected input variables, as determined by the models, collectively explain at least 98% of the variability in commercial vehicle miles traveled (VMT). In other words, with their chosen inputs, the linear models considered in this analysis effectively account for the vast majority of the variation observed in commercial VMT across the selected California counties. This high adjusted R^2 value suggests that the variables selected by the models are strong predictors of commercial VMT, capturing the underlying trends and patterns with remarkable accuracy. The models' performance, as indicated by the adjusted R^2 values, underscores the effectiveness of the chosen variables in explaining the fluctuations in commercial VMT over the study period.

4. Summary & Conclusions

This study focused on six California counties: Imperial, Los Angeles, Riverside, San Bernardino, San Diego, and San Francisco, each offering a unique blend of demographics, landscapes, and socioeconomic characteristics. By analyzing these counties together, the aim is to capture the multifaceted nature of California's communities. The dataset spans from 2000 to 2020 and includes variables falling into three main categories: demographics, economics, and pollution. The primary outcome variable is total commercial vehicle miles traveled (VMT), representing the total miles traveled by commercial trucks used for transportation purposes.

The analysis employs linear regression, LASSO regression, and elastic net regression to understand the relationships between various factors and total commercial VMT across the six counties. LASSO and elastic net regressions allow for feature selection and regularization, which is particularly valuable when dealing with datasets with many variables. The results reveal key predictors influencing total commercial VMT, including variables such as vehicle population, fuel consumption, CO_2 and $PM_{2.5}$ pollution, median income, number of employed individuals, and poverty.

The correlation plots illustrate relationships between input variables and total VMT, highlighting positive correlations with the number of trips, CO_2 emissions, NO_2 emissions, and fuel consumption, as well as negative correlations with unemployment and poverty. The positive correlation suggests that policies aimed at reducing total VMT could have multiple benefits, such as lower emissions and reduced fuel consumption. This can guide policymakers in prioritizing interventions that target VMT reduction to achieve environmental and energy conservation goals. The negative correlations with unemployment and poverty indicate that economic and social factors are important considerations in transportation policy. Policies that stimulate economic growth and reduce poverty could indirectly influence VMT and its associated impacts.

The key findings include the positive impact of vehicle population and fuel consumption on total VMT, the selection of CO_2 and $PM_{2.5}$ pollution variables, and the influence of economic factors such as median income and employment on transportation patterns. The analysis also highlights the vulnerability of commercial transportation to economic fluctuations, as seen during the aftermath of the 2008 financial crisis. These findings can inform the development of targeted policies aimed at managing and optimizing commercial transportation, improving air quality, and addressing socioeconomic disparities. By understanding the interplay of demographics, economics, and pollution on the total commercial VMT, stakeholders can work towards sustainable and efficient transportation systems that benefit the economy and the environment across California's various counties.

In conclusion, this study sheds light on the factors influencing total commercial VMT in six diverse California counties over a two-decade period. Through the application of various regression techniques, significant predictors have been identified, providing valuable insights for policymakers, urban planners, and transportation authorities.

Appendix A

Table 7. LASSO Regression for the Whole Data

Imperial (Whole)	**Estimate**	**Std Error**	**t value**	**Pr(>|t|)**	
$PM_{2.5}$	-4.20E-01	3.68E-02	-11.417	8.41E-08	***
CO_2	3.96E-01	1.38E-01	2.866	0.014185	*
Fuel Consumption	1.86E-01	1.33E-01	1.404	0.18571	
MedIncome	2.78E-01	3.98E-02	6.992	1.45E-05	***
Unemployment	-1.22E-01	2.96E-02	-4.139	0.001374	**
Poverty	1.08E-01	2.48E-02	4.332	0.000976	***
House Price Index	-1.32E-01	2.07E-02	-6.381	3.50E-05	***
Los Angeles	Estimate	Std Error	t value	Pr(>|t|)	
Trips	-3.03E-01	7.64E-02	-3.959	0.004185	**
X.$PM_{2.5}$_TOTEX	-2.41E-01	2.07E-01	-1.161	0.27914	
X.CO_TOTEX	5.36E-01	2.42E-01	2.211	0.057966	.
Fuel Consumption	7.86E-01	8.75E-02	8.985	1.88E-05	***
Population County	2.11E-01	1.63E-01	1.299	0.23015	
Employed	1.03E-01	6.40E-02	1.612	0.145582	
GDP.All	1.13E+00	1.82E-01	6.177	0.000266	***
Poverty	-9.29E-02	7.20E-02	-1.29	0.23305	
SNAP	3.53E-02	9.89E-02	0.357	0.730288	
Premature	1.50E-01	8.31E-02	1.804	0.108942	
House Price Index	-1.11E-01	1.09E-01	-1.022	0.336681	
Riverside	Estimate	Std Error	t value	Pr(>|t|)	
$PM_{2.5}$	-4.84E-01	7.86E-02	-6.165	1.36E-05	***
Fuel Consumption	6.64E-01	6.22E-02	10.664	1.12E-08	***
Employed	-3.81E-02	1.20E-01	-0.317	0.755	

Imperial (Whole)	**Estimate**	**Std Error**	**t value**	**Pr(>|t|)**	
San Bernardino	Estimate	Std Error	t value	Pr(>|t|)	
$PM_{2.5}$	-3.78E-01	1.25E-01	-3.029	0.00902	**
Fuel Consumption	5.38E-01	6.27E-02	8.579	6.01E-07	***
Unemployment	-4.18E-02	3.10E-02	-1.349	0.19872	
Employed	4.35E-01	1.77E-01	2.465	0.02724	*
House Price Index	-1.82E-01	6.07E-02	-2.999	0.00956	**
San Diego	Estimate	Std. Error	t value	Pr(>|t|)	
Trips	-1.67E-01	6.91E-02	-2.412	0.032779	*
Fuel Consumption	5.97E-01	4.02E-02	14.862	4.32E-09	***
Population County	1.03E+00	4.48E-02	22.907	2.85E-11	***
Unemployment	-1.18E-01	3.37E-02	-3.485	0.004501	**
Poverty	-2.44E-01	5.46E-02	-4.463	0.000775	***
Premature	9.42E-02	5.43E-02	1.734	0.108539	
House Price Index	-4.04E-02	5.15E-02	-0.784	0.448346	
San Francisco	Estimate	Std. Error	t value	Pr(>|t|)	
Population Vehicle	8.77E-04	5.14E-02	0.017	0.9866	
Fuel Consumption	1.17E+00	5.55E-02	21.001	1.55E-12	***
Employed	1.77E-01	9.05E-02	1.953	0.0698	.
GDP.All	8.31E-01	1.14E-01	7.271	2.74E-06	***

Table 8. Elastic Net Regression for Whole Data

Imperial (Whole)	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	
$PM_{2.5}$	-4.20E-01	3.68E-02	-11.417	8.41E-08	***
CO_2	3.96E-01	1.38E-01	2.866	0.014185	*
Fuel Consumption	1.86E-01	1.33E-01	1.404	0.18571	
MedIncome	2.78E-01	3.98E-02	6.992	1.45E-05	***
Unemployment	-1.22E-01	2.96E-02	-4.139	0.001374	**
Poverty	1.08E-01	2.48E-02	4.332	0.000976	***
House Price Index	-1.32E-01	2.07E-02	-6.381	3.50E-05	***
Los Angeles	Estimate	Std. Error	t value	Pr(>|t|)	
Trips	-2.52E-01	7.01E-02	-3.596	0.005786	**
CO_2	-2.90E-01	2.75E-01	-1.056	0.318294	
Fuel Consumption	1.15E+00	2.39E-01	4.801	0.000973	***
Population County	1.01E-01	7.88E-02	1.278	0.233303	
Employed	2.29E-01	4.68E-02	4.899	0.000849	***
GDP.All	8.17E-01	1.38E-01	5.921	0.000223	***
Poverty	-7.01E-02	6.51E-02	-1.078	0.309208	
SNAP	1.25E-01	9.88E-02	1.268	0.236631	
Premature	3.04E-01	3.51E-02	8.65	1.18E-05	***
House Price Index	-2.63E-01	6.69E-02	-3.93	0.003459	**
Riverside	Estimate	Std. Error	t value	Pr(>|t|)	
NO_x	-4.83E-02	2.03E-01	-0.238	0.8155	
$PM_{2.5}$_TOTEX	-4.44E-01	1.07E-01	-4.169	0.0011	**
CO_2_TOTEX	2.53E-01	4.51E-01	0.562	0.5836	
Fuel Consumption	3.42E-01	4.96E-01	0.691	0.502	
Employed	2.30E-01	3.25E-01	0.707	0.4919	
GDP.All	-2.08E-01	2.48E-01	-0.841	0.4157	

Imperial (Whole)	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	
San Bernardino	Estimate	Std. Error	t value	Pr(>|t|)	
$PM_{2.5}$	-3.95E-01	8.94E-02	-4.425	0.001019	**
CO_2	1.70E-01	3.75E-01	0.453	0.659378	
Fuel Consumption	3.96E-01	3.74E-01	1.058	0.312735	
MedIncome	-3.00E-02	8.43E-02	-0.355	0.729126	
Unemployment	5.67E-02	6.73E-02	0.841	0.418072	
Employed	5.90E-01	1.92E-01	3.068	0.010703	*
Poverty	-1.18E-01	5.44E-02	-2.177	0.052124	.
House Price Index	-3.00E-01	5.50E-02	-5.449	0.000201	***
	Estimate	Std. Error	t value	Pr(>|t|)	
Trips	-2.64E-01	8.97E-02	-2.945	0.01635	*
$PM_{2.5}$	-1.20E-01	2.34E-01	-0.512	0.620865	
CO_2	-4.08E-01	2.71E-01	-1.508	0.165937	
Fuel Consumption	9.16E-01	2.38E-01	3.857	0.003862	**
Population County	1.21E+00	2.22E-01	5.449	0.000406	***
Unemployment	-1.53E-01	3.69E-02	-4.138	0.002528	**
Poverty	-2.41E-01	6.68E-02	-3.612	0.005641	**
SNAP	-3.35E-01	1.65E-01	-2.027	0.073233	.
Premature	6.40E-02	5.30E-02	1.208	0.258003	
House Price Index	-1.76E-02	5.38E-02	-0.327	0.751114	
	Estimate	Std. Error	t value	Pr(>|t|)	
Population Vehicle	2.43E-02	5.08E-02	0.478	0.641979	
CO_2	-9.87E-01	5.39E-01	-1.831	0.094273	.
Fuel Consumption	2.17E+00	4.82E-01	4.499	0.000902	***
Population County	3.39E-01	2.06E-01	1.645	0.128121	
Employed	-2.79E-01	2.12E-01	-1.319	0.213843	
GDP.All	1.09E+00	1.86E-01	5.875	0.000107	***
SNAP	-2.76E-02	1.21E-01	-0.229	0.823401	

Imperial (Whole)	Estimate	Std. Error	t value	Pr(>|t|)
House Price Index	-1.75E-01	1.03E-01	-1.701	0.117097

Bibliography

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65(1), 91-98.

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

Ewing, R., & Cervero, R. (2001). Travel and the built environment: a synthesis. *Transportation research record*, *1780*(1), 87-114.

Federal Highway Administration (FHWA). (n.d.). What is Vehicle Miles Traveled (VMT)? Retrieved from https://www.fhwa.dot.gov/tpm/guidance/glossary/vmt.cfm

Federal Highway Administration (FHWA). (2020). Freight Facts and Figures 2020. Retrieved from https://www.fhwa.dot.gov/policyinformation/statistics/2020/index.cfm

Feng, W., & Figliozzi, M. (2013). An economic and technological analysis of the key factors affecting the competitiveness of electric commercial vehicles: A case study from the USA market. *Transportation Research Part C: Emerging Technologies*, *26*, 135-145.

Fruin, S. A., Winer, A. M., & Rodes, C. E. (2004). Black carbon concentrations in California vehicles and estimation of in-vehicle diesel exhaust particulate matter exposures. *Atmospheric environment*, *38*(25), 4123-4133.

Janssen, N. A., Schwartz, J., Zanobetti, A., & Suh, H. H. (2002). Air conditioning and source-specific particles as modifiers of the effect of PM (10) on hospital admissions for heart and lung disease. *Environmental health perspectives*, *110*(1), 43-49.

Kumapley, R. K., & Fricker, J. D. (1996). Review of methods for estimating vehicle miles traveled. *Transportation Research Record*, *1551*(1), 59-66.

Loder, A., Tanner, R., & Axhausen, K. W. (2017). The impact of local work and residential balance on vehicle miles traveled: A new direct approach. *Journal of Transport Geography*, 64, 139-149.

McMullen, B. S., & Eckstein, N. (2012). Relationship between vehicle miles traveled and economic activity. *Transportation Research Record*, 2297(1), 21-28.

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.

Noland, R. B. (2001). Relationships between highway capacity and induced vehicle travel. *Transportation Research Part A: Policy and Practice*, *35*(1), 47-72.

Polzin, S. E., Chu, X., & Toole-Holt, L. (2004). Forecasts of future vehicle miles of travel in the United States. *Transportation research record*, *1895*(1), 147-155.

Williams, T. A., Chigoy, B., Borowiec, J. D., & Glover, B. (2016). Methodologies used to estimate and forecast Vehicle Miles Traveled (VMT).

Woldeamanuel, M., & Kent, A. (2014). Determinants of per capita vehicle miles traveled (VMT): The case of California. *Journal of the Transportation Research Forum*, 53(3), 35-46.

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.

About the Authors

Steve Chung, PhD

Dr. Chung has received PhD in Statistics. He has been actively involved in many research fields including engineering, transportation, medicine, nursing, and data science. And he has worked with students, faculty, practitioners, and professionals from Fresno State Transportation Institute, Community Regional Medical Center, UC San Francisco, Heart and Artery Center at Fresno, and Fresno VA Hospital.

Jaymin Kwon, PhD

Dr. Kwon has a background in environmental exposure science, with specific training and expertise in research on personal exposure assessment to air pollutants and data analysis on spatial and temporal variation of air pollutants. His current (STARTRAQ_2020) and recently completed (CHAPS) research include spatial and temporal changes of particulate matter and other air pollutants associated with fossil fuel combustion emissions, such as traffic. As a PI or co-investigator at several universities, NIOSH, and NIEHS/USEPA-funded grants, he constructed spatial GIS databases and analyzed the proximity effect of air pollutants.

Yushin Ahn, PhD

Dr. Ahn is a member of the geomatics engineering faculty at California State University in Fresno, CA. He is an active member of the American Society for Photogrammetry and Remote Sensing and has been a certified Photogrammetrist since 2014. His research interest is remote sensing, mapping, sensor modeling/calibration, and geographic information system assisted data analysis.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

MTI MINETA TRANSPORTATION INSTITUTE