

Exploring the Complex Influences on Sustainable Driving: A Structural Equation Modeling Approach with Multisource Data

Virginia Petraki^{1*} and George Yannis¹

¹ National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heron Polytechniou str., 15773, Athens, Greece
*vpetraki@mail.ntua.gr

Abstract. This study aims to investigate the factors and their complex interrelationships that influence key indicators of driving sustainability, with a focus on both safety and efficiency. A Structural Equation Model is developed using multi-source data to capture how unsustainable driving practices are shaped by the following three latent variables: driving volatility, route characteristics, and weather conditions. These latent variables are modeled using multiple observed indicators derived from smartphone sensors, OpenStreetMap, NASA satellite data, and weather sensors. The dataset consists of approximately 33,000 trips recorded via smartphone sensors in Athens, Greece, between March and May 2024. The model examines fuel consumption as a measure of efficiency, and four Surrogate Safety Measures (SSMs), speeding events, mobile phone use, and harsh acceleration and braking as safety-related variables. SEM enables estimation of both direct and indirect pathways, including how latent factors interact and how safer and more efficient driving behaviors may reinforce each other. This approach supports the generation of targeted insights to inform behavioral interventions for promoting safer and more efficient urban mobility.

Keywords: Sustainable driving behavior, SEM, road safety, fuel consumption.

1 Introduction

The increase in vehicle ownership, road fatalities, excessive energy consumption, and environmental pollution are pressing issues. Driving behavior is a major determinant of road safety and is responsible for approximately 95% of road crashes [1]. Furthermore, driving behavior greatly affects fuel consumption and vehicle exhaust emissions [2]. Therefore, enhancing driving behavior is crucial to promote sustainable road transportation addressing its externalities.

Current literature often investigates risky and inefficient driving behavior as two distinct problems. However, numerous studies have investigated the relationship between eco and safe driving practices, unveiling important similarity between them or compromises. Pinals et al. [3] proved that improvements in safe driving behavior, i.e., reduced harsh braking and high-speed driving, were considerably associated with reduced fuel consumption. Şimşek et al. [4] have shown with GPS data that by addressing

speeding events and idle durations, driver performance can balance safety and fuel efficiency. Yao et al. [5] studied the impact of the external condition on driver safety and fuel consumption and concluded that both fuel consumption and crash risk significantly increase when driving in junctions at night, in a traffic jam, or in poor weather such as light rain, snow, or dust. Yet other studies have revealed that although intersections between safe and fuel-efficient driving practices are significant, their interaction is typically complex and sometimes contradictory [6, 7].

Therefore, the need to investigate both eco-friendly and safe driving simultaneously is evident since they are two interdependent phenomena. Structural Equation Model (SEM) is a multivariate analysis tool that supports multiple-input and multiple-output modeling. In driver behavior studies, SEM is employed to represent latent constructs like risky driving with a range of Surrogate Safety Measures (SSMs) like speeding, harsh events, sudden lane change, tailgating, time to collision etc. [8-10], distraction [11-13], driving skills and volatility (speed variation, jerk) [10, 14], eco-efficiency (fuel use, strong accelerations, stop density) [15]. Exogenous conditions such as weather, road conditions, and traffic are also modeled as latent variables [16, 17], so SEM can capture the interaction between driving behavior and exogenous conditions. This study is guided by the following research questions:

- How can unobserved latent factors influencing both risky and inefficient driving behavior be measured within a unified framework?
- What is the relationship between risky and inefficient driving behavior at the trip level?
- How can multiple datasets encompassing driving behavior, road environment, temporal, and weather data be integrated to investigate sustainable driving behavior?

2 Methodology

2.1 Data Collection

Driving Behavior Data

The data was collected through an innovative data collection scheme, developed by the OSeven Telematics (www.oseven.io), which records personalized driving behavior analytics in real time, using smartphone sensors. For more information about the collection, storage, management, and processing of data, the reader can refer to previous research [18]. The driving behavior dataset captures various driving parameters, including GPS speed, acceleration and several SSMs like harsh acceleration and deceleration events, speeding events and mobile phone use. The data was collected in an anonymized format from 32,815 trips over a 3-month period, March to May, during 2024.

Fuel Consumption

The fuel consumption was calculated through Virginia Tech Comprehensive Power-Based Fuel Consumption Model (VT-CPFM) model based on telematics driving data. We used VT-CPFM as a microscopic fuel consumption model that estimates vehicle instantaneous fuel consumption rates by calculating instantaneous power as an input variable. This model can avoid bang-bang control as it has a second-degree polynomial relationship with vehicle-specific power (VSP). More details on the model and its implementation can be found in [19].

The VT-CPFM model is a black-box model that requires vehicle-specific parameters, instantaneous speed, and acceleration, as well as external factors such as road grade and altitude. Since detailed vehicle data are not available for our telematics drivers, we assumed an “applied vehicle”. Based on the available information, that the telematics fleet consists of passenger cars, and the majority is gasoline-powered, we used the profile of an average gasoline passenger car in Greece, with a mean age of 17.3 years [20]. However, this approach allows for an efficiency analysis methodology that can be refined with future driver-specific vehicle information from drivers or insurers. Finally, information on road grade and altitude at one-second resolution was obtained through the map-matching of the driving data, as described in the following section.

Road Infrastructure Data

To map match the naturalistic trips within the Attica region, a bounding box was dynamically generated based on the spatial extent of the GPS coordinates of the driving data. This area was then used to extract information about the junctions (nodes) and the road edges (ways) from the OSM database using the Python library OSMnx [21]. Each trip trajectory was processed at a one-second resolution, with GPS points converted to geometric representations and map-matched to the corresponding road network edges. Finally, the map-matched database was integrated with the extended OSM database, extracting parameters like road type, road segment geometry characteristics etc.

An additional step was to add the information road grade in each road segment through altitude of the start and end point of each road segment. To integrate elevation data, we utilized the Mapzen Global Terrain dataset, which provides open-source elevation database derived from NASA SRTM and other global sources. These elevation values were assigned to the node of each road edge. Then, the road slope was calculated based on the elevation difference between its start and end nodes.

Road Infrastructure Data

Hourly weather conditions were collected using the Open-Meteo, a free weather data provider offering global forecasts and historical climate records. Specifically, data were retrieved for a weather station near Athens, Greece, including parameters such as temperature, wind speed, relative humidity, snowfall and daylight. This hourly resolution weather dataset was used to capture the environmental conditions during the time periods relevant to each trip.

2.2 Structural Equation Modeling Background

SEM is a technique within the family of latent variable analysis. It is a multivariate method that supports both multiple-input and multiple-output modeling addressing various endogenous and exogenous variables to define any direct or indirect relationship between latent and manifest variables [22]. In this study, a SEM was designed to capture the influence of external factors as long as human factors on both fuel consumption and risky driving behavior. One of the most common goodness-of-fit measures is standardized root average square residual (SRMR), which is an index of the average of standardized residuals between the observed and the hypothesized co-variance matrices. Values of the SRMR range between zero and one, with well-fitting models having values <0.08 [23].

3 Results and Discussion

This section presents the results of the SEM analysis, focusing solely on the final models, presented in Table 2. All statistical analyses were conducted in R-studio and SEM analysis utilized the lavaan R package [24]. Ultimately, the proposed SEM structure retained four latent unobserved variables. The “DrivingVolatility” which reflected the instability in driving performance and was measured through stop frequency, acceleration variability, and high positive jerk during the trip. The “RouteCharacteristics” which summarized the roadway environment, combining the proportion of motorway and residential roads, the share of flat road grades, the road grade variability, and the number of intersections per kilometer during the trip. The “Weather” latent variable captured the weather conditions represented by daylight status, temperature, and humidity. The “AggressiveDriving” was modeled as latent using the harsh braking and harsh acceleration events per 10 kilometers.

Table 1 presents the set of variables used in the SEM model, grouped under the corresponding latent constructs and endogenous variables. To enhance model fit, several variables were scaled linearly by factors of 10, which reduced variance.

Table 1. Description of variables.

| Latent | Variable | Description |
|------------------------------|-------------------------|---|
| Driving Volatility | stop_count_per_km | Stop frequency per trip kilometer |
| | pos_jerk_Q90 | 90 th percentile of positive jerk |
| | acc_SD | Standard deviation of longitudinal acceleration during the trip |
| | perc_motorway | Proportion of trip distance driven on motorways (speed limit > 80 km/h) |
| Route Characteristics | perc_residential | Proportion of trip distance driven on residential roads (speed limit < 50 km/h) |
| | inzone | 1 if trip occurs within one municipality, 0 if it crosses more |
| | unique_u_count_perkm_sc | Number of intersections per kilometer (scaled by /10) |
| | std_grade | Standard deviation of road grade during a trip |
| | perc_grade_flat_sc | Proportion of flat road sections (scaled by /100) |
| Weather | is_day | Daytime indicator (1=day, 0=night) |
| | temperature_sc | Temperature scaled by /10 (°C) |
| | humidity_pct_sc | Relative humidity scaled by /100 (%) |
| Aggressive Driving | harsh_brk_per_10km | Harsh braking events per 10 km |
| | harsh_acc_per_10km | Harsh acceleration events per 10 km |
| | speeding_per_100m | Number of speeding events per 100 m |
| | mobile_use_per_100m | Mobile phone interactions per 100 m |
| - | FC_100 | Fuel consumption in liters per 100 km of trip |

Following SEM calibration, the produced model results are presented on Table 2. Statistically significant p-values (<0.05) are shown in bold. Collectively, all the four examined goodness-of-fit measure values suggest good model fit.

Table 2. SEM model of fuel inefficient and risky driving behavior.

| SEM Components | | Parameters | Estimate | Std.Err | P(> z) | |
|---------------------------------|--------------------------|-----------------------------|--------------------|---------|--------------|--------------|
| Latents: | DrivingVolatility | stop_count_per_km | 1 | - | - | |
| | | pos_jerk_Q90 | 0.634 | 0.002 | 0.000 | |
| | | acc_SD | 0.081 | 0 | 0.000 | |
| | RouteCharacteristics | perc_motorway | 1 | - | - | |
| | | perc_residential | -0.788 | 0.01 | 0.000 | |
| | | inzone | -1.301 | 0.02 | 0.000 | |
| | | unique_u_count_perkm_sc | -3.199 | 0.029 | 0.000 | |
| | | std_grade | -3.553 | 0.053 | 0.000 | |
| | Weather | perc_grade_flat_sc | 0.639 | 0.008 | 0.000 | |
| | | is_day | 1 | - | - | |
| | | temperature_sc | 2.19 | 0.028 | 0.000 | |
| | AggressiveDriving | humidity_pct_sc | -0.867 | 0.011 | 0.000 | |
| | | harsh_brk_per_10km | 1 | - | - | |
| | | harsh_acc_per_10km | 0.931 | 0.028 | 0.000 | |
| Regressions: | FC_100 | DrivingVolatility | 1.663 | 0.01 | 0.000 | |
| | | RouteCharacteristics | -3.801 | 0.103 | 0.000 | |
| | | Weather | 0.015 | 0.048 | 0.747 | |
| | AggressiveDriving | DrivingVolatility | 0.001 | 0.013 | 0.964 | |
| | | RouteCharacteristics | -3.182 | 0.147 | 0.000 | |
| | | Weather | -0.12 | 0.068 | 0.076 | |
| | speeding_per_100m | DrivingVolatility | -0.12 | 0.003 | 0.000 | |
| | | RouteCharacteristics | 0.856 | 0.04 | 0.000 | |
| | | Weather | -0.122 | 0.018 | 0.000 | |
| | mobile_use_per_100m | DrivingVolatility | 0.032 | 0.011 | 0.004 | |
| | | RouteCharacteristics | -2.266 | 0.122 | 0.000 | |
| | | Weather | 0.3 | 0.06 | 0.000 | |
| | Covariances: | perc_motorway | speeding_per_100m | -0.023 | 0.001 | 0.000 |
| | | std_grade | perc_grade_flat_sc | -0.117 | 0.001 | 0.000 |
| acc_SD | | harsh_acc_per_10km | 0.126 | 0.002 | 0.000 | |
| | | harsh_brk_per_10km | 0.07 | 0.002 | 0.000 | |
| DrivingVolatility | | RouteCharacteristics | -0.112 | 0.002 | 0.000 | |
| | | Weather | 0.019 | 0.002 | 0.000 | |
| RouteCharacteristics | | Weather | 0 | 0 | 0.016 | |
| AggressiveDriving | | FC_100 | 0.53 | 0.02 | 0.000 | |
| | | speeding_per_100m | 0.139 | 0.007 | 0.000 | |
| | | mobile_use_per_100m | -0.034 | 0.023 | 0.139 | |
| FC_100 | | speeding_per_100m | 0.042 | 0.005 | 0.000 | |
| | | mobile_use_per_100m | 0.06 | 0.016 | 0.000 | |
| speeding_per_100m | | mobile use per_100m | -0.036 | 0.006 | 0.000 | |
| Goodness-of-fit measures | | Comparative Fit Index (CFI) | | 0.962 | | |
| | Tucker-Lewis Index (TLI) | | 0.947 | | | |
| | RMSEA | | 0.055 | | | |
| | SRMR | | 0.035 | | | |

The SEM path diagram developed through lavaangui R package and is presented in Figure 1. The green arrows denote positive correlations, whereas red arrows denote negative correlations. It should be noted that covariances between latent variables are also depicted in the path diagram.

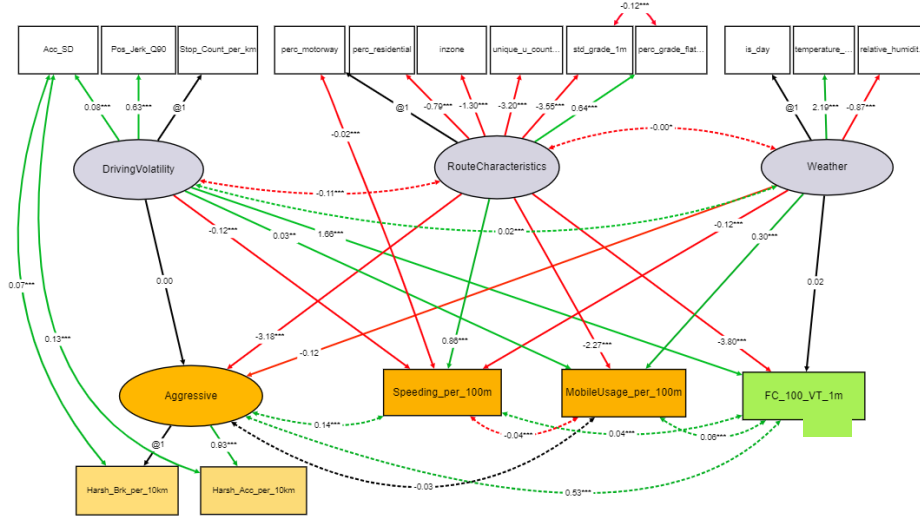


Fig. 1. Path diagram of SEM model.

Leading contributor in causing fuel consumption was driving volatility. Traveling trips with high stop rates, increased acceleration variability, and positive jerks require more energy [25]. Safety-wise, unsafe trips with higher mobile phone use, were also linked positively with driving volatility, possibly indicating that, for example stop-and-go conditions foster secondary task involvement by drivers. Interestingly, volatility had a negative correlation with speeding, indicating that volatility emerges in congested or urban environments where opportunities for speeding are scarce, whereas risks of distraction are greater. Thus, it can be concluded that volatile driving situations penalize efficiency but minimize speeding risk, while increasing distraction-related risk.

Route increased complexity, capturing higher residential share, dense urban intersections, and steeper road grades, significantly increases fuel consumption and aggressive driving, thereby validating the benefits of motorway prevalence and low-grade conditions for eco performance and safe performance [26]. But those same conditions were considerably linked with higher speeding events. This outcome validates an old compromise, planning for smoother, high-capacity flow reduces inefficiency and harsh events but increases risk from high-speed crashes, which are typically more harmful. Therefore, although route planning has a role in eco-efficiency, it must be balanced with caution using speed management measures to prevent neutralizing safety benefits.

Weather showed no significant effect on fuel consumption but was weakly and positively associated with distraction. A likely explanation is that conditions during March–May were mostly moderate, and the fuel model did not account for air-conditioning use – both of which dampen weather – fuel sensitivity. The residual covariance structure also provides evidence for eco-safety interdependencies. Fuel consumption was positively correlated with aggressive driving ($r=0.53$, $p < 0.001$) and with speeding and mobile phone use, although less so. This is in line with the view that inefficient trips tend to be unsafe trips [3, 6].

4 Conclusion

This study aims to investigate the interrelationships between driving behavior, roadway and environmental conditions, and their combined effects on fuel consumption and safety-related outcomes on a trip level using a large-scale naturalistic driving dataset. By applying a SEM framework, we unveiled direct and indirect pathways linking driving volatility, route characteristics, and weather factors to fuel consumption, aggressive driving, speeding, and distraction. The analysis shows that fuel inefficiency and unsafe driving are not independent, and they share behavioral and contextual roots. This finding underscores the need for integrated approaches that simultaneously target eco-driving and road safety objectives, instead of addressing them in isolation.

Acknowledgement

This study is based on the research project “OptiMo - Optimising driver behaviour for safe, green and energy efficient mobility” (project code: 016529), carried out within the framework of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU (Implementation body: HFRI).



References

1. Singh, S.: Critical reasons for crashes investigated in the national motor vehicle crash causation survey. DOT HS 812 115 (2015).
2. Singh, H., Kathuria, A.: Profiling drivers to assess safe and eco-driving behavior—A systematic review of naturalistic driving studies. *Accident Analysis & Prevention* 161, 106349 (2021).
3. Pinals, L., Guo, F., Ahn, K., Madden, S., Rakha, H.: Safe driving is sustainable driving: An interpretable telematics methodology. *Transportation Research Record*, 03611981251346774 (2025).
4. Şimşek, B., Pakdil, F., Dengiz, B., Testik, M.C.: Driver performance appraisal using GPS terminal measurements: A conceptual framework. *Transportation Research Part C: Emerging Technologies* 26, 49–60 (2013).
5. Yao, Y., Zhao, X., Zhang, Y., Chen, C., Rong, J.: Modeling of individual vehicle safety and fuel consumption under comprehensive external conditions. *Transportation Research Part D: Transport and Environment* 79, 102224 (2020).
6. Haworth, N., Symmons, M.: The relationship between fuel economy and safety outcomes (2001).
7. Jain, N., Mittal, S.: Review of computational techniques for modelling eco-safe driving behavior. *International Journal of Automotive and Mechanical Engineering* 20(2), 10422–10440 (2023).
8. Dong, X., Xie, K., Yang, H.: How did COVID-19 impact driving behaviors and crash severity? A multigroup structural equation modeling. *Accident Analysis & Prevention* 172, 106687 (2022).
9. Yang, D., Xie, K., Ozbay, K., Yang, H.: Fusing crash data and surrogate safety measures for safety assessment: Development of a structural equation model with conditional

- autoregressive spatial effect and random parameters. *Accident Analysis & Prevention* 152, 105971 (2021).
10. Ding, N., Jiao, N., Zhu, S., Liu, B.: Structural equations modeling of real-time crash risk variation in car-following incorporating visual perceptual, vehicular, and roadway factors. *Accident Analysis & Prevention* 133, 105298 (2019).
 11. Das, A., Ahmed, M.M.: Structural equation modeling approach for investigating drivers' risky behavior in clear and adverse weather using SHRP2 naturalistic driving data. *Journal of Transportation Safety & Security* 15(11), 1116–1147 (2023).
 12. Sheykhfard, A., Haghghi, F.: Driver distraction by digital billboards? Structural equation modeling based on naturalistic driving study data: A case study of Iran. *Journal of Safety Research* 72, 1–8 (2020).
 13. Rupp, M.A., Gentzler, M.D., Smither, J.A.: Driving under the influence of distraction: Examining dissociations between risk perception and engagement in distracted driving. *Accident Analysis & Prevention* 97, 220–230 (2016).
 14. Papantoniou, P., Antoniou, C., Yannis, G., Pavlou, D.: Which factors affect accident probability at unexpected incidents? A structural equation model approach. *Journal of Transportation Safety & Security* 11(5), 544–561 (2019).
 15. Zhu, X.H., Lu, K.F., Peng, Z.R., Gao, H.O.: Characterizing carbon emissions from China V and China VI gasoline vehicles based on portable emission measurement systems. *Journal of Cleaner Production* 378, 134458 (2022).
 16. Sheykhfard, A., Haghghi, F., Saecidi, S., SafariTaherkhani, M., Das, S.: Understanding the influence of environmental factors on driver speed: A structural equation modeling analysis. *IATSS Research* 48(3), 427–439 (2024).
 17. Xu, C., Li, D., Li, Z., Wang, W., Liu, P.: Utilizing structural equation modeling and segmentation analysis in real-time crash risk assessment on freeways. *KSCE Journal of Civil Engineering* 22(7), 2569–2577 (2018).
 18. Papadimitriou, E., Argyropoulou, A., Tselentis, D.I., Yannis, G.: Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. *Safety Science* 119, 91–97 (2019).
 19. Rakha, H.A., Ahn, K., Moran, K., Sacerens, B., Van den Bulck, E.: Virginia Tech comprehensive power-based fuel consumption model: Model development and testing. *Transportation Research Part D: Transport and Environment* 16(7), 492–503 (2011).
 20. ACEA: Average age of the EU vehicle fleet, by country (2024).
 21. Boeing, G.: OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65, 126–139 (2017).
 22. Shaaban, K., Gaweesh, S., Ahmed, M.M.: Characteristics and mitigation strategies for cell phone use while driving among young drivers in Qatar. *Journal of Transport & Health* 8, 6–14 (2018).
 23. Chen, F.: Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling* 14, 464–504 (2007).
 24. Rosseel, Y.: lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(1), 1–36 (2012).
 25. Fernandes, P., Ferreira, E., Macedo, E., Coelho, M.C.: Unraveling roundabout dynamics: Analysis of driving behavior, vehicle performance, and exhaust emissions. *Transportation Research Part D: Transport and Environment* 133, 104308 (2024).
 26. Charly, A., Mathew, T.V.: Identifying risky driving behavior: A field study using instrumented vehicles. *Transportation Letters* 16(7), 688–702 (2024).