In automotive design - the effect of MPG on vehicle fuel consumption

Hongze Shi

Abstract:

Fuel efficiency, measured in miles per gallon (MPG), and engine power, represented by horsepower, are two critical metrics that gauge a vehicle's performance. Historically, a negative correlation has been perceived between these two variables: cars with higher horsepower were deemed to have a lower MPG. This study comprehensively explores this correlation using a dataset obtained from the UCI repository. Three regression models were applied: linear, quadratic, and cubic. Each was meticulously analyzed and evaluated using mean squared error (MSE) and residual plot to understand the data's fit and the models' predictive capabilities. Our research found that while the linear model provides initial insights, polynomial regression models, especially the quadratic one, can capture the relationship more succinctly, which revealed a relationship expressed as $f\left(x\right)=1.66x^2-7.87x+21.94$. The findings indicate that as horsepower increases, MPG decreases, but with diminishing intensity, suggesting an intricate balance between power and efficiency in automotive design.

Keywords: MPG; horsepower; regression analysis; automotive;

1. Introduction

The Miles Per Gallon (MPG) metric is a pivotal determinant of a vehicle's fuel efficiency. Elevated MPG values signify a vehicle's enhanced capability to traverse more miles for each gallon of fuel, leading to diminished fuel consumption and a consequent potential reduction in environmental footprint. Both regulatory standards concerning fuel economy and consumer preferences frequently emphasize MPG, given its profound influence on operational expenditures and ecological sustainability (Greene et al., 2023; Linn, 2016; Shalini et al., 2021)

MPG normally can be influenced by multiple factors, one of the predominant ones being the vehicle's horsepower — a metric of the engine's power (Greene et al., 2023).

Horsepower delineates the engine's power output, acting as a barometer for a vehicle's performance prowess and operational capacity. An augmented horsepower typically translates to swifter acceleration and elevated maximum velocities. While there isn't a direct correlation between horsepower and fuel efficiency, it remains an essential factor for a demographic of consumers accentuating vehicular performance and competency(Ahmed and Stater, 2023; Greene et al., 2023).

Separately, horsepower delineates the engine's power output, acting as a barometer for a vehicle's performance prowess and operational capacity. An augmented horsepower typically translates to swifter acceleration and elevated maximum velocities. While there isn't a direct correlation between horsepower and fuel efficiency, it remains an essential factor for a demographic of consumers accentuating vehicular performance and competency.

As per Galang (2017), an analysis of empirical observations suggests a general inverse relationship between MPG and horsepower; vehicles with augmented horsepower frequently manifest diminished fuel efficiency. This is attributable to the enhanced fuel requisition by potent engines to deliver necessary acceleration and elevated velocities. Conversely, engines with reduced horsepower are often engineered to optimize fuel economy, culminating in superior MPG metrics. The interplay between MPG and horsepower is multifaceted, contingent upon supplementary variables, including vehicular weight, aerodynamic properties, and advanced engine technologies. Meanwhile, according to Shalini et al. (2021) and Linn (2016), an analysis of the relationship between vehicle weight, horsepower, and fuel efficiency showed that vehicles with higher horsepower generally had lower fuel efficiency.

However, it's important to note that the magnitude and significance of the relationship between MPG and horsepower may vary based on the specific dataset and methodology used in the analysis. Therefore, it would be helpful to consult the specific references mentioned in the context information to better understand the relationship between MPG and horsepower (Ahmed and Stater, 2023).

Furthermore, Greene et al. (2013) and a study by Knittel (2011) suggested that technological advancements have led to the development of high-horsepower engines that can maintain good fuel efficiency. While horsepower and weight increase typically decrease fuel economy, technological innovations have offset these tendencies, allowing vehicles to maintain or even increase fuel efficiency.

This study aims to understand and model the relationship between horsepower and MPG, delving deeper into whether a simple linear relationship exists or if more complex polynomial relationships might provide a better fit.

2. Methodology & Result

2.1 Data Collection

2.1.1 Data Import

To examine the correlation between horsepower and fuel efficiency (MPG), we obtained data from the reputable UCI Machine Learning Repository, specifically the Auto MPG dataset. This repository, overseen by the University of California, Irvine, is a recognized source of datasets frequently used for empirical studies in the machine learning community.

The Auto MPG dataset contains several attributes, among which we have focused primarily on 'MPG' (Miles per Gallon) and 'Horsepower.' The dataset provides realistic data for several types of vehicles and encompasses various characteristics that may affect the fuel efficiency of a vehicle.

2.1.2 Data Cleaning

Our method of obtaining data consisted of retrieving the dataset from the UCI repository via a web link. Real-world datasets commonly involve discrepancies and absent values, confirmed after initial assessment of the 'horsepower' attribute. We ensured the reliability of our analysis by substituting these absent values with the dataset's mean horsepower.

The dataset presented a systematic arrangement of vehicle names and their corresponding attributes, providing a thorough outlook on several vehicle specifications and fuel efficiencies. Possessing this dataset allowed for the following stages: preprocessing, exploratory data analysis, and the core modeling processes.

2.2 Descriptive Analysis

2.2.1 Draw a scatter plot

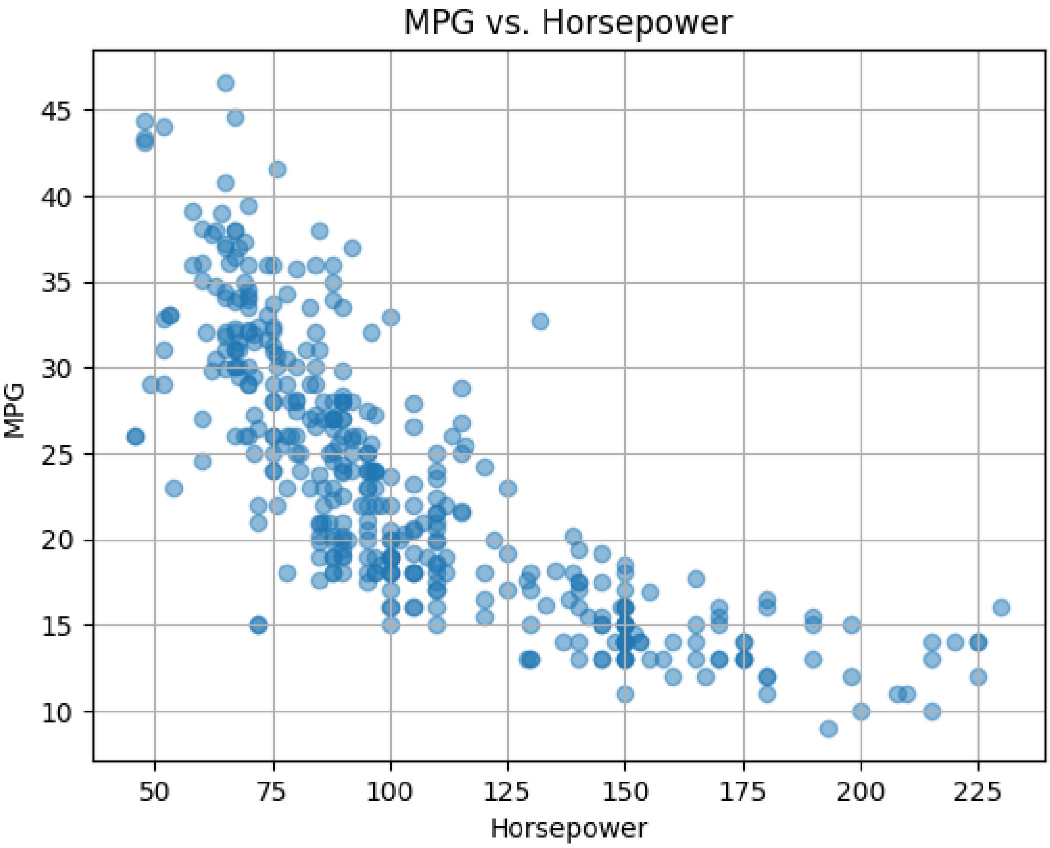

Figure 1 scatterplot

This scatter diagram displays the correlation between horsepower and miles per gallon (MPG). Most data points cluster in the low horsepower and high MPG categories. As the horsepower increases, the MPG exhibits a downward trend, indicating a negative association between the two. The graph has almost no outliers, just a few examples in the high horsepower and high MPG areas. In conclusion, this graph clearly shows the trend where horsepower increases can lead to MPG decreases.

2.2.2 Correlation Analysis

Our study commenced with evaluating the straight-line connection between a car's fuel economy, recognized as miles per gallon (MPG), and its horsepower, which gauges this relationship by the Pearson correlation coefficient. The obtained coefficient value of r2 = −0.78 indicates a significant inverse correlation between MPG and horsepower, indicating that automobiles with higher horsepower tend to have lower fuel efficiency. Nevertheless, correlation mustn't imply causation; other factors may affect these variables.

2.3 Model Fitting

2.3.1 Data Preprocessing

Initially, horsepower was designated as our independent variable (denoted as 'X'), while the fuel efficiency, measured in MPG, was taken as the dependent variable (represented as 'y').

The dataset was further partitioned into training and testing subsets to ensure an unbiased evaluation of our regression model. Specifically, 80% of the data was reserved for training, while the remaining 20% was set aside for validation purposes.

Furthermore, we subjected the data to standardization to maintain uniformity and ensure a valid comparison across the horsepower values. This process entails adjusting the values to have a mean of zero and a standard deviation of one. Standardizing the data is vital, especially in regression analysis, to ensure that the scale of the predictors does not influence the coefficients.

To account for potential non-linearities in the relationship between MPG and horsepower, polynomial features of the standardized horsepower were generated, specifically squared and cubed terms extended to our dataset. Specifically, squared and cubed terms of the standardized horsepower values were computed, thus enabling the exploration of quadratic and cubic regression models.

2.3.2 Regression Analysis

Taking into account the previously established nomenclature where horsepower was designated as our independent variable (X) and fuel efficiency (in MPG) as the dependent variable (y), our analysis further explored the nature of their relationship using varying degrees of polynomial regression (Given the precision required for our analysis, all coefficients mentioned in this section are approximated to three decimal places):

1. Linear Relationship: Our primary step was to gauge the linear association between (X) and (y). Mathematically, this relationship is characterized by

$$y = a_1 + b_1 X$$

The OLS regression results rendered the following coefficients:

- Intercept (a_1): 23.608
- Slope (b_1): -6.047

2. Quadratic Relationship: We considered a quadratic model involving the squared term of (X) to capture potential nonlinear patterns. This is mathematically described as

The derived coefficients from this regression were:

$$y = a_2 + b_2 X + c_2 X^2$$

- Intercept (\[{a}_{1}\]): 21.943
- Linear term coefficient (b_2): -7.870
- Quadratic term coefficient (c_2): 1.665

3. Cubic Relationship: We incorporated a cubic term to delve deeper, positing a more intricate polynomial relationship. This takes the form

$$y = a_3 + b_3 X + c_3 X^2 + d_3 X^3$$

The coefficients for this model were:

- Intercept (a_3): 21.827
- Linear term coefficient (b_3): -7.730
- Quadratic term coefficient (c_3): 1.893
- Cubic term coefficient (d_3): -0.102

Through this structured approach, we endeavored to holistically apprehend the intricacies between horsepower (X) and MPG (y), facilitating a thorough examination to identify the model that best describes the empirical observations.

2.4 Model Evaluation

We conducted a comprehensive evaluation using the test dataset to ascertain the predictive power of the models we've generated. Specifically, we assessed the fit of each model (linear, quadratic, and cubic) by computing both the Mean Squared Error (MSE) and the R-squared (R^2) statistics:

Table 1 MSE and R-squared of three models

	Linear Regression	Quadratic Regression	Cubic Regression
MSE	19.1512	13.7609	13.8134
R^2	0.6438	0.7441	0.7431

To further dissect the performance, we examined the residuals — the differences between the observed and predicted fuel efficiencies. The following observations can be drawn from the residual plots:

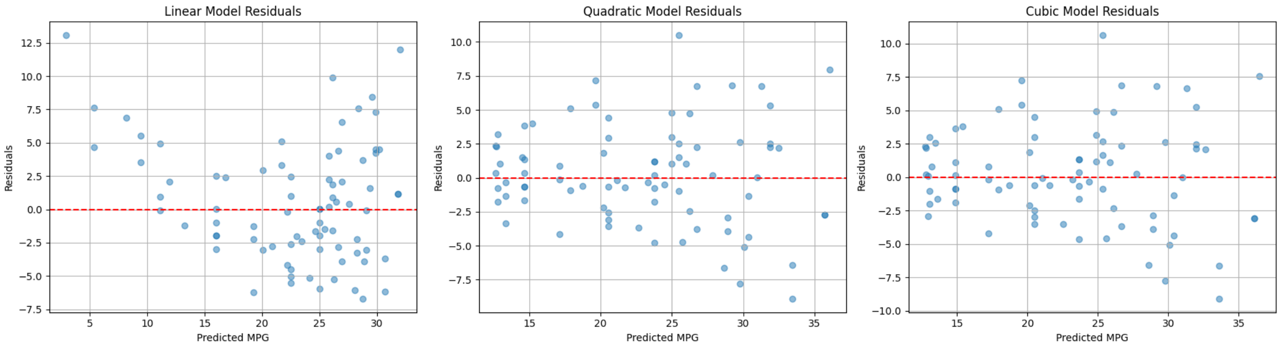

Figure 2 Residual plots

Linear Model: The distribution of residuals against predicted MPG values shows a certain pattern, suggesting potential non-linearity in the actual relationship.

Quadratic Model: The residuals are more evenly distributed around the zero line than the linear model, indicating a better fit.

Cubic Model: While the cubic model also depicts residuals distributed around the zero line, the scattering hints at potential overfitting, where the model might capture noise rather than the underlying trend.

In summary, upon rigorous evaluation of the model's predictive capabilities, as demonstrated by the MSE, R^2 values, and the dispersion in the residual plots, the quadratic model consistently outperforms the linear and cubic alternatives. In terms of both error magnitude and explained variance, the quadratic model exhibits superior performance. Moreover, the residual plots corroborate this, manifesting a more homogenous distribution for the quadratic regression than its counterparts. Consequently, from an academic standpoint, we can deduce that the quadratic model furnishes the most optimal representation of the relationship between horsepower and MPG.

2.5 Graphical Representation of Regression Models

Visual representation plays a crucial role in the academic analysis of our data. The following describes the methodology and results of our graphical analysis:

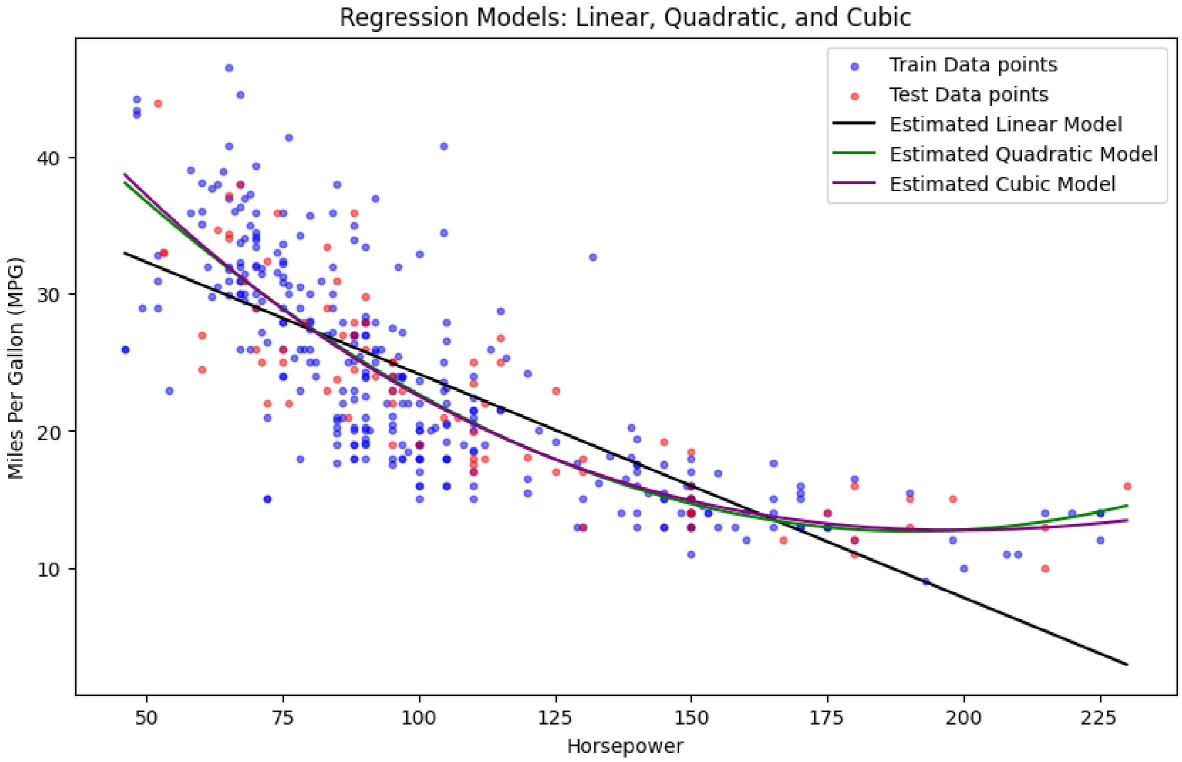

Figure 3 Visualisation of regression models

For visualization purposes, the data was plotted with the original training data points in blue and the test data points in red. Superimposed on this scatter plot were the three regression models. The linear regression was rendered in black, the quadratic in green, and the cubic in purple. This juxtaposition provides a clear comparative view of how each model fits the data.

From the graph, the curves of the cubic and square models almost overlap, which makes it evident that the two models offer a more nuanced representation of the data, curving to capture the underlying trend in the relationship between horsepower and MPG. On the other hand, the linear model doesn‘t appear to encapsulate the data trend as proficiently.

In conclusion, the graphical representation fortifies our previous findings, asserting the quadratic model as the superior fit for this dataset.

3. Conclusion

The relationship between an automobile’s horsepower and fuel efficiency, as measured in miles per gallon (MPG), has long been a topic of interest in the automobile industry. In our study, we comprehensively explored this relationship by employing linear, quadratic, and cubic regression models.

Our regression results showed a negative correlation between horsepower and MPG, indicating that fuel efficiency tends to decrease as horsepower increases. This is consistent with conventional wisdom, as higher horsepower vehicles consume more fuel to produce that power, resulting in lower MPG.

Additionally, our analysis’s crux was determining which of the three regression models - linear, quadratic, or cubic - provided the best representation of our data set. Our criteria for model evaluation encompassed various statistical measures, including the Mean Squared Error (MSE) and the R-squared value, as well as a thorough inspection of residual plots.

Conclusively, while each model has its merits, the quadratic model emerged as the most fitting for this specific dataset and purpose. It strikes a balance, accurately representing the data without veering into overfitting territory. This has implications for the automobile industry, suggesting that there might be a non-linear, possibly quadratic, relationship between horsepower and fuel efficiency. Future research could delve deeper into the factors influencing this relationship, providing even more clarity and aiding in developing more fuel-efficient vehicles.

4. Discussion

4.1 Strength

It’s crucial to address some key considerations that arose during our investigation:

1. Feature Standardization: Our data underwent standardization, ensuring all features contribute equitably to the model by bringing them to a common scale. This step was pivotal, especially for polynomial regression, ensuring that higher-degree terms did not unduly influence the model due to their larger scales.
2. Model Complexity: While the quadratic model was optimal in this scenario, it’s imperative to remember the risks associated with increasing model complexity. The model becomes more intricate with each added polynomial term, and the potential for overfitting grows. The cubic model showcased hints of this, capturing noise rather than the underlying trend. This serves as a reminder that a more complex model isn’t always the better choice.
3. Quantitative Metrics: Using MSE and R-squared provided quantifiable measures to compare model performances. This makes the evaluation objective and based on statistically sound criteria.
4. Comprehensive Visualization: Through various plots like the residuals plot and regression models plot, the analysis provided a visual understanding of the model’s behavior and its fit. This aids in non-technical interpretation and conveys insights effectively.

4.2 Limitations

This research has provided valuable insights into the relationship between a vehicle’s horsepower and fuel efficiency regarding miles per gallon (MPG). However, as with any empirical study, several limitations need to be acknowledged:

1. Scope of Model Complexity: While the quadratic regression emerged as the optimal model for this dataset, it’s essential to recognize the inherent limitations of polynomial models. Though capturing the underlying trend, the quadratic representation might omit nuanced interactions in the actual relationship.
2. Generalizability: Our study’s findings are based on a specific dataset, which, while informative, might not universally represent all cars or conditions. For broader applications, one must consider external variables such as vehicle weight, engine type, or driving conditions, which could influence the MPG.
3. Explanatory Variables: The research predominantly fixated on horsepower as the sole predictor. However, MPG is a multifaceted metric influenced by numerous variables—engine type, vehicle weight, and aerodynamics, to name a few. Future studies incorporating these factors

might yield more holistic models.

4. Causation and Correlation: It is vital to distinguish between correlation and causation. While a strong correlation was observed, it does not inherently imply a causal relationship between horsepower and MPG.

In light of these limitations, the findings of this study should be approached with an informed perspective. Future research endeavors can build upon this foundation, incorporating a wider array of variables and utilizing more advanced modeling techniques to elucidate the intricate dynamics between horsepower and MPG.

4.3 Further Work

The findings from our study open several avenues for future research that can deepen our understanding of the relationship between vehicle horsepower and miles per gallon (MPG). Some potential directions include:

1. Expanding Dataset Scope: To enhance the generalizability of our findings, future studies could consider a more diverse dataset, possibly encompassing different types of vehicles (e.g., hybrid, electric), various manufacturers, and different geographical locations.
2. Advanced Modelling Techniques: Given the intricacies of vehicle performance metrics, advanced machine learning algorithms or non-linear regression models might offer improved predictive performance and insights.
3. Causal Inference Techniques: Techniques like instrumental variable regression or propensity score matching can be explored to move beyond mere correlations and delve into causative relationships.
4. Technological Advancements: The relationship between horsepower and MPG could shift with emerging technologies and evolving design philosophies in the automobile sector. Electric vehicles, hybrid technologies, and advanced fuel management systems might alter this dynamic, necessitating periodic reassessments.

4.4 Implications

4.4.1 Automotive Industry

1. Marketing and Positioning: As Anair and Mahmassani (2012) note, consumers are more attracted to fuel-efficient vehicles. Cars that balance horsepower and impressive MPG could be uniquely positioned in the market.
2. Regulatory Compliance: Bandivadekar et al. (2008) discuss the stringent fuel efficiency standards imposed worldwide. Automakers should leverage insights from horsepower and MPG relationships to ensure compliance and minimize penalties.

4.4.2 Environmental Implications

Jacobson (2009) observes that vehicles with higher MPG generally contribute less to greenhouse gas emissions. The trade-off between horsepower and MPG is thus essential in environmental sustainability discussions.

4.4.3 Consumer Implications

Purchase Decisions: Axsen and Kurani (2012) explore how MPG affects consumer choices. By understanding the relationship between horsepower and MPG, consumers can make more informed decisions, considering performance and efficiency.

4.5 Discussion

Our exploration of the relationship between horsepower and MPG (miles per gallon) underscores a nuanced interplay that is corroborated by academic research and industry articles. The quintessential study by Christopher R. Knittel in “Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector” provides a robust groundwork for understanding the technological advancements in the automobile sector since the 1980s. Specifically, Knittel postulates that as vehicle manufacturers grapple with the intricate balance between power and efficiency, they are confronted by certain trade-offs among fuel economy, weight, and engine power characteristics.

In a similar vein, Clark Williams-Derry, in his 2008 exposition “Horsepower vs. MPG - Sightline Institute,” delves into the progression of the automotive landscape since 1975. Williams-Derry underlines the initial focus on improving fuel efficiency from 1975 to 1987, driven by stringent federal standards. The subsequent years, however, witnessed a paradigm shift, with engineers channeling their innovations towards augmenting raw power, often to the detriment of fuel efficiency. This dichotomy, where fuel economy improvements stagnate while horsepower escalates, particularly from the late 1980s, can be attributed to an amalgam of factors, ranging from plummeting gas prices to the revival of the economy and the establishment of federal fuel economy benchmarks.

Synthesizing these perspectives with our empirical analyses based on regression models (linear, quadratic, and cubic), we find that the quadratic model provides the best representation of the data. This is consistent with the findings of both Knittel and Williams-Derry, which emphasize the multifaceted nature of the relationship between MPG and horsepower. While raw power and fuel economy often appear antithetical, technological advances and vehicle designs can sometimes blur these boundaries.

In addition, the residual plots from our regression models provide insight into the behavior of the data around the regression lines, revealing patterns that may not be immediately apparent from the regressions alone.

For example, the residuals from the quadratic model highlight the article's emphasis on the importance of multiple variables in deciphering vehicle efficiency and performance.

In conclusion, while our analyses confirm the obvious trade-offs between power and fuel efficiency, they also underscore the monumental influence of technological innovation and policy directives in shaping this relationship. Future efforts in this area would benefit from an interdisciplinary approach that integrates technological, environmental, and policy perspectives to provide a holistic understanding of the evolving dynamics of the automotive industry.

References

[1]. Ahmed, R., & Stater, M. (2023). Automobile technology and tradeoffs: How does automobile footprint affect its fuel economy? Transportation Research Interdisciplinary Perspectives, 21, 100897. https://doi.org/10.1016/j.trip.2023.100897

[2]. Anair, D., & Mahmassani, A. (2012). State of charge. Union of Concerned Scientists, 10. State of Charge | Union of Concerned Scientists (ucsusa.org)

[3]. Axsen, J., & Kurani, K. S. (2012). Interpersonal influence within car buyers' social networks: applying five perspectives to plug-in hybrid vehicle drivers. Environment and Planning A, 44(5), 1047-1065. https://doi.org/10.1068/a43221x

[4]. Bandivadekar, A., Bodek, K., Cheah, L., Evans, C., Groode, T., Heywood, J., ... & Weiss, M. (2008). On the road in 2035: reducing transportation's petroleum consumption and GHG emissions. Laboratory for Energy and the Environment, Massachusetts Institute of Technology. http://web.mit.edu/sloan-auto-lab/research/beforeh2/otr2035/

[5]. Galang, A. (2017). Predicting Hybrid Vehicle Fuel Economy and Emissions with Neural Network Models Trained with Real-World Data (Master of Science, Colorado State University, Fort Collins, Colorado, Spring)

[6]. Greene, D. L., Kahn, J. R., & Gibson, R. C. (2023). Fuel Economy Rebound Effect for U.S. Household Vehicles. https://www.jstor.org/stable/41322836

[7]. Jacobson, M. Z. (2009). Review of solutions to global warming, air pollution, and energy security. Energy & Environmental Science, 2(2), 148-173. https://doi.org/10.1039/B809990C

[8]. Knittel, Christopher R. 2011. "Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector." American Economic Review, 101 (7): 3368-99. http://dx.doi.org/10.1257/aer.101.7.3368

[9]. Linn, J. (2016). The Rebound Effect for Passenger Vehicles. The Energy Journal, 37(2). https://doi.org/10.5547/01956574.37.2.jlin

[10]. Shalini, L., Naveen, S., & Ashwinkumar, U. M. (2021). Prediction of Automobile MPG using Optimization Techniques. 2021 IEEE Madras Section Conference (MASCON), 1–6. https://doi.org/10.1109/MASCON51689.2021.9563597

[11]. Sightline Institute. (n.d.). Horsepower vs. MPG. Retrieved from Horsepower vs. MPG - Sightline Institute