Estimating the Direct Rebound Effects for Passenger Vehicles in the UK

Craig Pearce

MSc Economics

University of Surrey

Guildford, UK

9th September 2024

Contents

Table of Figures 3
i. *Abstract* 5
1. Introduction 5
2. Literature Review 6
2.1. Background 6
2.2. Measuring the Rebound Effect 8
2.3. Empirical Studies on Direct Rebound Effects in Vehicles 9
3. Data: 11
3.1. Sampling 11
4. Method 15
4.1. Model Specification 15
4.2. Robustness Tests 16
5. Results 20
5.1. Fuel Price Elasticities on Vehicle Miles Travelled 20
5.2. Model Statistics 20
5.3. Other Estimates 21
5.4. Subgroup Analysis by Fuel Type, Vehicle Classification, and Geographic Classification 23
6. Caveats and Future Research 26
7. Discussion 28
7.1. Policy Implications 29
8. Conclusions 30
9. References 31
10. Appendix 34

Table of Figures

Figure 1: Illustration of Rebound Effects, sourced from [31]Sorrell (2009) ... 7
Table 1: Description of Variables ... 11
Table 2: Summary Statistics of Variables ... 14
Table 3: Summary Statistics of the Share of Vehicles by Vehicle and Postcode Classifications ... 15
Table 4: Shapiro-Wilk Test Results ... 17
Table 5: Robustness Test Results ... 19
Table 6: First-Stage Regression Statistics ... 19
Table 7: Regression Results ... 22
Figure 3: Yearly Average Mileage by Vehicle Classification 2018-2023 ... 23
Figure 4: Yearly Average Mileage by Postcode Classification 2018-2023 ... 24
Table 8: Subgroup Regressions Results ... 25
Table 9: Direct Rebound Effects from Other Studies ... 28

i. Abstract

This study investigates the short-run direct rebound effects of fuel efficiency improvements across petrol and diesel vehicles in the UK using odometer readings from MOT results from over 2.3 million vehicles tested between 2013 and 2023. Employing fixed effects and two-stage least squares (2SLS) models, the analysis estimates the elasticity of vehicle miles travelled (VMT) with respect to fuel price changes, while controlling for economic, environmental, and demographic factors. The findings suggest a direct rebound effect of approximately 8.9% across all vehicles, with petrol vehicles exhibiting a higher rebound effect than diesel vehicles, which demonstrated a negative rebound effect. These results indicate that fuel efficiency improvements leads to a small increase in driving for petrol vehicles but a reduction in mileage for diesel vehicles in the short run, particularly in urban areas. The study also highlights the importance of considering geographic and vehicular differences when designing transport and environmental policies. Our findings suggest that policies designed to discourage less-efficient vehicles should target petrol vehicles more stringently than diesel vehicles to mitigate the rise in demand for driving.

1. Introduction

The transport sector is a major contributor to energy usage and greenhouse gas emissions in the UK, accounting for 33% of the UK's oil consumption and 29.1% of all emissions and increasing each year, according to the [16,17]Department for Energy Security and Net Zero. In order to decrease this and mitigate the transport sector's impact on climate change, the UK and the rest of the world have placed significant emphasis on improving the fuel efficiency of personal cars to reduce energy demand through policies such as congestion charges ([15]Gov.uk (2020), [36]Transport for London (2019)) across urban areas of the UK and Ultra Low Emission Zones (ULEZ) ([37]Transport for London (2024)) in London. With congestion charges increasing and ULEZ being implemented in more locations in the UK, there arises a need to analyse how effective these policies are in reducing energy demand.

By increasing the fuel efficiency of vehicles, the fuel cost per mile and therefore fuel demand and emissions per mile should decrease. However, the full potential of fuel efficiency improvements may not always be realised due to the rebound effects. The direct rebound effect occurs when the cost of driving falls and leads to an increase in driving, and also depends on how drivers behave and react to these policies as well as the price of fuel. If this rebound effect is significant then it could offset the expected outcomes from these policies. The extent of this rebound effect could vary across different vehicles, fuel types and areas, so understanding these variations is crucial for designing effective transport and environmental policies.

This study is intended to build on an add to existing literature such as [6]Craglia and Cullen (2020) who conducted one of the most recent studies on the direct rebound effects in the UK using a comprehensive dataset across various subgroups of vehicles and areas. One major gap in the literature on this subject within the UK is distinguishing differences between petrol and diesel fuelled vehicles. One of the only studies that accomplishes this is from [14]Gillingham and Munk-Nielsen (2019) in Denmark, this is one of the primary goals that this study will seek to explore, whilst taking into account variation across specific characteristics such as vehicle classifications and the urban and rural areas of the UK.

The primary goal of this research is to provide a more recent estimate of the direct rebound effect in the UK in petrol and diesel vehicles following the expansion of policies aimed at improving fuel efficiency using non-aggregated data to avoid masking any geographical or social differences between vehicles and their drivers. This research also intends to distinguish these effects between rural and urban areas in preparation for the expansion of ULEZ and more stringent congestion charges across the UK and ask the question: To what extent are these policies actually reducing energy demand, and who are they impacting the most? By answering this question, this study aims to provide policymakers with useful insights that can better inform the design of future transport and environmental policies in the UK, to ensure that they achieve their intended goals in energy and environmental improvements.

2. Literature Review

This review of literature will be divided into multiple strands of study to discuss the background and theories underpinning the topic, the different methods used to measure the rebound effect and review the various empirical studies that attempted these methods that this study will seek to build upon.

2.1. Background

Improvements in energy efficiency (ε) generally leads to savings in the use of an energy service (S), and therefore monetary and emission savings, ceteris paribus. However, ceteris paribus is not true in reality and other factors such as the price of energy (P_E) also play a crucial role in determining the extent of the savings as it directly influences the energy cost of useful work (P_S), which is calculated using:

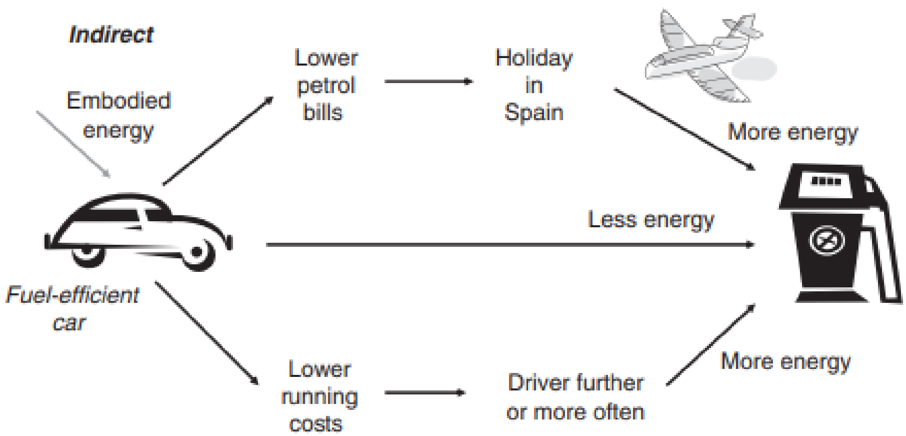
$$P_S = \frac{P_E}{\varepsilon}$$

Rebound effects are a range of mechanisms that can offset energy savings following energy efficiency improvements, such as a change in P_E, changes in the prices of substitutes (P_x) or the characteristics of individuals using the energy service.

For example, an improvement in a vehicle's fuel efficiency would mean that it requires less fuel to travel a mile, omits lower emissions per mile and can travel further on a full tank. However, the price of the fuel at the pump may change or the price of a train or bus may change, or the driver may choose to travel longer distances and/or more often. These are just a few examples of rebound effects that can all directly offset energy savings.

[31]Sorrell (2009) provides a comprehensive introduction to the rebound effect and decomposes it into two parts, the direct and indirect rebound effects. The indirect rebound effect occurs from how the individual maximises utility with their monetary savings following energy efficiency improvements, one example of this is depicted in Figure 1, where the cost of driving for the individual has fallen and they are able to accumulate savings. If they were to then spend these savings on other goods and services such as a holiday to Spain via air travel, this would result in more energy being used and higher emissions emitted from jet fuel than from the savings. Figure 1 also provides an example of the direct rebound effect, if an individual were to gain the same fuel efficiency improvements as before then instead of accumulating the same amount of savings, they could instead opt to adjust their driving patterns and increase their mileage which would lead to less or no energy savings. This effect would also be amplified by a decrease in the price of fuel or an increase in the price of transport substitutes such as buses or trains. [29]Saunders (1992), [2]Brookes (2000) and [10]Druckman et al (2011) also explored the 'backfire' effect, which occurs when an increase in energy efficiency causes an increase in energy consumption - where the total rebound effect is greater than 100%. This would lead to significant policy implications.

Figure 1: Illustration of Rebound Effects, sourced from [31]Sorrell (2009)



[31]Sorrell (2009) also decomposes the direct rebound effect into the substitution and income effects where similarly to common economic theory, the income effect is where a fall in P_S achieves the consumer a higher disposable income which allows them to attain a higher level of utility by consuming more of S. The substitution effect is where the reaction of the consumer following a fall in P_s is that to keep the same level of utility by substituting consumption of other goods and services (with a price of P_x) with the now relatively cheaper energy service.

[21]Khazzoom (1980) conducted one of the earliest and most seminal studies on the rebound effect in their investigation of increased energy efficiency in household appliances, finding that if $\eta_{P_E}(S)$ is not zero then there will be an upward pressure on energy demand following an improvement in energy efficiency. He also explains that following an improvement in a car's fuel efficiency, reducing the energy cost of driving could mean that the driver may choose to be less conservative of their fuel when driving. He argues that policymakers must account for these changes in consumer behaviour when setting minimum energy standards, highlighting the complexity of energy policies.

2.2. Measuring the Rebound Effect

[32]Sorrell, Dimitropoulos and Sommerville (2009) provide a summary for measuring the direct rebound effect and state that there are two elasticities that can be estimated, depending on the data available:

- The elasticity of demand for useful work with respect to energy efficiency, $\eta_{\varepsilon}(S)$
- The elasticity of demand for energy (E) with respect to energy efficiency, $\eta_{\varepsilon}(E)$

Where:

$$S = E \cdot \varepsilon$$

And:

$$\eta_{\varepsilon}(\mathrm{E}) = \eta_{\varepsilon}(S) - 1$$

However, the majority of studies that estimate direct rebound effect use price elasticities due to data availability, especially in vehicles where fuel price elasticities are used. [30]Small and Van Dender (2005), [6]Craglia and Cullen (2020), [12]Ficano and Thompson (2014) and [34]Stapleton et al (2016) measure $\eta_{P_E}(S)$ in their studies, because under certain assumptions $\eta_{\varepsilon}(S) \approx -\eta_{P_E}(S)$. However, these are very strong assumptions, which according to [3]Chan and Gillingham (2015) and [33]Sorrell and Dimitropoulos (2008) are:

- Fuel prices are exogenous and do not depend on fuel efficiency
- Individuals are limited in shifting their travel to alternative transport
- Individuals react the same way to changes in P_S following changes in P_E or ε

Also, under similar assumptions:

- $\eta_{\varepsilon}(S) \approx -\eta_{P_S}(S)$
- $\eta_{\varepsilon}(S) \approx -\eta_{P_E}(E)$

The second assumption may be the most limiting in the context of the UK, where there is a significant rail and bus infrastructure, especially in the urban areas. The third assumption should also hold because if drivers are rational then they should react the same way to a fall in the price of fuel to an increase in the fuel efficiency of their vehicle because using the equation for P_S, they should have an inversely proportional effect on P_S. Therefore, the best data for this assumption to hold in the UK would be from rural towns and villages. In this study, the direct rebound effect will be estimated as $-\eta_{P_E}(S)$ and we will explore the difference in $\eta_{P_E}(S)$ between the urban and rural regions, as well as the exogeneity of fuel prices because they might be a function of other variables. This study will also explore the rebound effects in different classes of vehicles because differences in energy efficiency and driving behaviour exist between vehicles and drivers, as stated by [21]Khazzoom (1980).

2.3. Empirical Studies on Direct Rebound Effects in Vehicles

The first study which measures the direct rebound effect in vehicles to be reviewed is from [6]Craglia and Cullen (2020), who take a sample of 10 million individual's cars in the UK and tracked their Ministry of Transport (MOT) test results over a 10 year period and estimate the rebound effect at a national and postcode level using fixed effects and instrumental variable regressions, finding an average $\eta_{P_E}(S)$ of -0.046. This research seems to have one of the most robust datasets used to estimate the direct rebound effect by using VMT for individuals in a panel of data rather than an aggregated time series and will be used as a methodological base for this study. However, by using a panel with a relatively short time period both studies will only be capable of estimating the short-run direct rebound effect which is generally calculated to be lower than long-run effects. [6]Craglia and Cullen (2020) also include investigative differences in populations of the postcodes, seemingly representing congestion and driving patterns, whereas including the driving population of the postcode would provide a more specific insight into this by eliminating the noise of non-drivers, such as children who are captured by total population. Also, whilst they investigate differences within vehicle class and fuel consumption, they do not consider any differences in the fuel type of each vehicle. This study will look to improve upon this, although it agrees that any consistent postcode granularity data on these variables was not available.

[30]Small and Van Dender (2005) estimate an average $\eta_{P_S}(S)$ of -0.11 in the short run, and -0.5 in the long run using their time series data between 1996 and 2001 in the US using two-stage least squares and three-stage least squares regressions, instrumenting the endogenous variables with lagged versions of themselves and CAFE regulations. There is a risk of over-identification in these models with simultaneous equations which may make it harder to ensure the robustness of the results. [12]Ficano and Thompson (2014)

however, estimate a larger elasticity in their study using cross-sectional data from the US National Household Travel Survey in 2009. They employed OLS and IV regressions to estimate that $\eta_{P_E}(S)$ varies between -0.755 and -0.778 by instrumenting hybrid ownership with the proportion of hybrid vehicles in the state of residence to factor in household driving decisions. Whilst their tests on the instrumental variable showed positive results, this may not be the most relevant instrument to explain hybrid ownership, where regional factors such as public transport availability, the urban density and fuel prices may be more relevant and stronger than their chosen instrument.

[34]Stapleton et al (2016) estimate and average η_{P_E}(S) = -0.172 and η_{P_S}(S) = -0.152 in their analysis using time series data from 1970-2011 in the UK over 108 different models. The use of different standardisations of different independent variables in both static and dynamic variables provide one of the most robust, relatively simple methodologies. However, the long time period may bring some challenges when not controlling for factors such as urbanisation or structural changes in the UK over this time, these factors would affect driving behaviour which could lead to their estimations of rebound effects to be under or over estimated. [20]Huntington (2024) provides one of the newest contributions to the rebound effect in vehicles with their study using time series data from the US between 1949-2019, excluding any pandemic-related years in an autoregressive distributed lag model, estimating the elasticity of demand for fuel consumption with respect to fuel efficiency, η_{ε}(S) = -0.2, which rises to -0.5 as they added more variables representing vehicle attributes such as weight or horsepower. They are also using a long time series, running similar risks to [34]Stapleton et al (2016), but they are also using an aggregated data on fuel efficiency (miles per gallon) which averages out any heterogeneity between vehicle types such as freights and smaller vehicles, obscuring any differences between the groups and potentially biasing estimates.

[25]Moshiri and Aliyev (2017) used panel data consisting of 9 of the 10 provinces of Canada over the years 1997-2009 in an Almost Ideal Demand System (AIDS), and a Quadratic AIDS (QAIDS), which are commonly used when working with household expenditure data as it can estimate both income and price elasticities and decompose the rebound effect into income and substitution effects, as shown in [5]Chitnis and Sorrell's (2015) study where they estimate direct and indirect rebound effects following energy efficiency improvements over various household appliances . [25]Moshiri and Aliyev (2017) estimate an average η_{P_E}(S) = -0.88 with large variations across provinces, this is significantly higher than estimates from other researchers in other countries and don't include factors representing driving behaviours or congestion, although they do refer to how urban each province in the study is. [7]De Borger et al (2016) also use similar data but to a finer granularity, they use household level panel data across Denmark from 2001-2011 tracking variables such as kilometres driven over time, fuel prices and driving behaviour using first difference and instrumental variable models. They estimate a η_{P_E}(S) that varies between -0.075 and -0.1, with a rebound effect lower than [25]Moshiri and Aliyev's, this could be due to Denmark's large public transport network which is generally perceived as very efficient. [9]Gillingham and Munk-Nielsen (2019) also estimate η_{P_E}(S) in Denmark using vehicular panel data similarly to [6]Craglia and Cullen (2020), over

the period 1990-2015. Using linear log-log and IV models they estimate a mean $\eta_{P_E}(S)$ = -0.3, finding a rebound effect higher than that of [7]De Borger et al (2016). These contradictory results are likely due to differences in data, [9]Gillingham and Munk-Nielsen (2019) use vehicular level panel data whilst instrumenting public transport access with municipality level population data, whereas [7]De Borger et al (2016) lack in this respect but have a better focus on fuel efficiency and car characteristics. This is something that this study will attempt to improve upon in the context of the UK.

3. Data:

In the UK, a vehicle which is over 3 years old is required to undertake a Ministry of Transport (MOT) test annually to ensure each vehicle meets minimum road safety and worthiness requirements. The results of every test are anonymised and published annually by the [8]Department for Transport (2024), providing information on descriptive statistics of each vehicle as well as odometer readings to capture vehicle mileage, these were used to calculate yearly mileage for each vehicle. For example, a vehicle ID of 1253 would be listed as a petrol fuelled Hyundai I20, tested on 24/05/2021 with a mileage of 20379 and a cylinder capacity of 998 cm^3, registered in the Wolverhampton area.

The data used in this empirical analysis is a panel and will consist of characteristics of each vehicle and their postcode, their test dates and their yearly mileage following their MOT test results across the UK over the period 2013-2023.

3.1. Sampling

Because running a regression on the full MOT dataset was computationally impractical, a random sample of 10 million unique vehicle IDs were taken from the MOT datasets and tracked over time. This sample was then cleaned by removing any vehicles which failed a test, were not tested in at least two consecutive years, any which have changed their postcode between tests, any vehicles which are not run on petrol or diesel, or any errors in the vehicle's ID, make, or model. After these procedures, the dataset was left with 10,428,847 observations tracking 2,340,314 unique vehicles. The remaining dataset was then matched with datasets containing information on the average price of the vehicles fuel source, the GDP, and the average income of the UK of the month that the vehicle was tested. A consistent monthly postcode level dataset of these variables would have been preferred but was unfortunately not available.

These results were then matched with characteristics of the postcode that the car was tested in, because it is relatively safe to assume that a driver will test their car close to where they live. They were matched with the average monthly rainfall, the yearly driving population of the postcode and the heating degree month of the postcode, which was calculated using daily temperature data from the [24]Met Office (2023) to find the mean temperature of each month then subtract it from 15.5 ^{o}C, which is the [23]Met Office's

(2006) classification of a heating degree day to measure the deviations of the postcode's temperature from the baseline each month. The use of HDM is more advantageous than average temperature because it avoids the combination of especially high and low temperatures which may average out into a moderate value and may not effectively represent thermal demand, which is expected to have a direct relationship with driving patterns and energy demand.

The vehicles were then categorised into 5 categories: City, Medium, Small Sedan, Small SUV, and SUV/MPV using characteristics of the make and model of each car, as described by the [37]Vehicle Certification Agency (VCA) database. Because the makes and models varied slightly in each dataset, for example capital letters, inverted names, or typographical errors, a fuzzy matching algorithm was employed using the make, model, year of first use, fuel type and cylinder capacity of each vehicle, this algorithm scores each match from 0-1 and uses a user-set tolerance level to remove any incorrect matches.

After these procedures, the age of the car was generated using the test date and the date of first use, and the yearly mileage was calculated by subtracting the previous year's odometer reading from the current year's reading. The yearly mileage for the first observation of each vehicle was calculated by dividing the odometer reading by the age of the car. Two dummy variables were also generated for whether the test date fell during the COVID-19 lockdowns or during the Russo-Ukrainian conflict. Finally, the dataset was then matched with the Brent Spot Price using the month of each test date.

The resulting dataset is described in Table 1, which contains an overview of the variables used in this analysis.

Table 1: Description of Variables

Variable ID:	Description:	Measurement:	Source:
VMT	Yearly vehicle miles travelled, generated using vehicle odometer readings at MOT tests	Miles	[8]Department for Transport (2024)
PP	Monthly average price of petrol at UK pumps	pence/litre	[18]Department for Energy Security and Net Zero
DP	Monthly average price of diesel at UK pumps	pence/litre	[18]Department for Energy Security and Net Zero
GDP	Monthly time series of nominal UK GDP	Index score in relation to June, 2019	[27]Office for National Statistics (2024)
Y	Monthly observations of average weekly earnings in the UK	£	[26]Office for National Statistics (2024)
AGE	Age of the vehicle at test date, generated using test date and first use	Years (1 decimal place)	[8]Department for Transport (2024)
HDM	Heating degree months, generated using daily postcode temperature data and heating degree classifications	Days per month	[24]Met Office (2023) classification: [22]Met Office Climate Data Portal (2024)
PRCP	Average monthly rainfall in each postcode, generated using daily postcode rainfall data	Millimetres	[23]Met Office (2006)
DPOP	Number of licensed vehicles in each postcode within the UK	Thousands	[9]Department for Transport (2022)
LOCK	A dummy variable indicating whether the test date was during a COVID-19 lockdown or not	= 1 if so = 0 if not	[1]Baker et al (2021)
WAR	A dummy variable indicating whether the test date was during the current conflicts in the Russo-Ukrainian war or not	= 1 if so = 0 if not	[39]Walker (2023)
BSP	Price of crude oil, converted into pound sterling	£/barrel	[11]EIA (2024) exchange rate: [13]FRED St. Louis (2024)

Table 2: Summary Statistics of Variables

Variable	Observations	Mean	Min	Max
VMT	10,428,847	6350.68	0.0223714	100000
PP	6,343,607	135.8827	104.87	191.55
DP	4,091,396	143.0878	111.7	199.22
GDP	10,428,847	98.78462	74.5691	102.8381
Y	10,428,847	182.6313	161.9497	211.1207
AGE	10,428,847	9.652286	0	123.5
HDM	10,428,847	5.080542	0.0004168	16.72138
PRCP	10,428,847	2.341273	0.0121739	20.00313
DPOP	10,428,847	365745.4	11113	966256
BSP	10,428,847	73.11	18.38	122.71
LOCK	10,428,847	0.0275732	0	1
WAR	10,428,847	0.315355	0	1

Table 2 shows the summary statistics of the model variables. One observation from this is that there are 6,343,607 observations for petrol vehicles and 4,091,396 observations for diesel vehicles, with the price of diesel being slightly higher than that of petrol. We also observe the average yearly mileage of our sample as 6350.68, this is lower than in [6]Craglia and Cullen's (2020)but can be explained by a slightly smaller sample and the COVID-19 lockdowns falling within our timeframe. This is partially shown with 2.75% of observations being tested during a lockdown, however this dummy variable does not fully capture the complete timeframe where personal travel was restricted, and therefore yearly mileage would have fallen.

Table 3 also shows the percentage makeup of the sample by the classification of each vehicle type, and by rural/urban postcode classification which was set in accordance with the [28]Office for National Statistics (2024) (ONS). We observe that 98% of vehicles in this dataset are tested in an urban postcode, this is a large majority and could have occurred from our sampling and cleaning process, such as rural based drivers may have changed their postcodes more than urban drivers during our testing period or may have been less likely to have been tested in successive years. We also observe a majority of 88.24% of vehicles in our dataset are either City or Medium class vehicles, this could be due to the user-created classification parameters or also due to the same reasons, for example the majority of Small Sedan drivers in this sample may have changed postcodes during this testing period.

Table 3: Summary Statistics of the Share of Vehicles by Vehicle and Postcode Classifications

Parameter	As a % of sample
City	23.86
Medium	64.38
SUV/MPV	4.74
Small SUV	7.01
Small Sedan	0.01
Rural	2
Urban	98

4. Method

4.1. Model Specification

This research will follow a similar analysis than that of [6]Craglia and Cullen (2020), using more recent data and will employ a fixed effects model to analyse the effects of the previously described economic and environmental factors on vehicle miles travelled (S), whilst controlling for any unobserved heterogeneity and time invariant characteristics of the vehicles and drivers. We model the VMT of each vehicle as a function of the average price of its fuel input (FP), the vehicle's age (AGE) and the UK GDP and average income at the time of its test (GDP and Y). The core model is used in a log-log form to capture the coefficients as elasticities and is described as follows:

$$lnVMT_{it} = \theta_i + \beta(X) + \epsilon_{it}$$

Where X is a vector of the following variables:

$$X = [lnFP_t + lnGDP_t + lnY_t + lnAGE_{it} + lnHDM_{it} + lnPRCP_{it} + lnDPOP_{it} + LOCK_t + WAR_t]$$

Where θ_i represents the specific fixed effect for vehicle i, ϵ_{it} is the idiosyncratic error component, and β is a vector of coefficients on each variable included in X. LnFP represents the fuel price of each vehicle, i, at the time they were tested, t, depending on the vehicles fuel input which would be lnPP for a petrol vehicle and lnDP for a diesel vehicle. This model will be used for both petrol and diesel vehicles to capture any differences in elasticities between the two, for example diesel vehicles are generally seen to be more fuel efficient than petrol and therefore could be less responsive to changes in fuel prices.

$lnHDM_{it}$ and $lnPRCP_{it}$ are included to control for the effects of weather on driving in each postcode over time, and $lnDPOP_{it}$ is included to control for potential congestion in the postcode of each vehicle. $LOCK_t$ and WAR_t are dummy variables included to control for any potential impacts on driving from the COVID-19 lockdowns, or the Russo-Ukrainian conflicts.

The primary focus from this regression will be on the coefficient on $lnFP_t$ which will represent the elasticity of demand for vehicle miles travelled with respect to the price for its fuel, $\eta_{PE}(S)$.

When analysing $\eta_{PE}(S)$, we are assuming that there is an exogenous change in fuel price causing a change in VMT, however, there may exist a short-run relationship where an increase in VMT in a particular period, for example the end of a COVID-19 lockdown, may cause an increase in fuel price. In this instance, there would be potential endogeneity in the model leading to biased and inconsistent results, which would need to be accounted for to ensure the results are robust. It is safe to assume that the international market for the inputs in production of petrol and diesel is unlikely to be unaffected by a short-run spike in VMT within the UK, therefore the proposed instrument will be Brent Spot Price, which is the main price benchmark for the international crude oil market. Crude oil is refined into diesel and gasoline and therefore its price directly affects the price of petrol and diesel at pumps in most countries. Therefore, an additional model will be estimated for both petrol and diesel vehicles in the form of fixed effects two-stage least squares regressions to test whether the results hold across all models.

4.2. Robustness Tests

Before the model can be estimated and interpreted accurately, it must undergo several robustness tests to ensure that none of the theoretical assumptions behind the model are violated and therefore the estimates are efficient and unbiased.

4.2.1. Normality

The first assumption to be tested is that the residuals of the regression are normally distributed with a zero mean and constant variance, if this is not true then the inferential statistics of our model will not be valid. To test this assumption, the Shapiro-Wilk test was implemented which carries a null hypothesis that the data is normally distributed. The results are shown in Table 4 and show that the null hypothesis was decisively rejected at a 1% significance level for every variable, this suggests that clustered standard errors may have to be used, these adjust the calculation of the standard errors to account for any intra-cluster correlation. However, the sample size is 10,428,847 which is likely to be sufficiently large enough to drop the assumption of normality.

Table 4: Shapiro-Wilk Test Results

Variable	W-Statistic
lnVMT	0.913 ***
lnPP	0.96 ***
lnDP	0.926 ***
lnGDP	0.72 ***
lnY	0.94 ***
lnAGE	0.97 ***
lnHDM	0.794 ***
lnPRCP	0.972 ***
lnDPOP	0.957 ***

Note: *, ** and *** represent as 10%, 5% and 1% significance level respectively

4.2.2. Cross-Sectional Dependence

Cross-sectional dependence is where the residuals in the model are correlated across vehicles and may influence each other, this could lead to misleading coefficients and test results. Pesaran's CD test for cross-sectional dependence in panels with a large N and a small T, is used which has a null hypothesis of weak cross-sectional dependence. Table 5 shows that the null is rejected at a 1% significance level, therefore the residuals are likely to show strong cross-sectional dependence and must be addressed using clustered standard errors.

4.2.3. Homoskedasticity

Homoskedasticity is where the variance of the errors is constant, regardless of the values of the explanatory variables, i.e. $\sigma^2_i = \sigma^2$. If this assumption is violated, then heteroskedasticity is present and the model's estimates will be inefficient. This analysis used a Modified Wald test for this assumption which carries a null hypothesis of constant variance in the error terms exists. The results are shown in Table 5, where the null hypothesis is rejected at a 1% significance level. This result also suggests the need for clustered standard errors.

4.2.4. Autocorrelation

If the error terms in two different time periods are correlated, then one of the assumptions behind the model is violated and the estimators will be inefficient as the estimated variances of the regression coefficients will be inconsistent and biased, meaning any hypothesis testing is no longer valid. This was tested for using the Wooldridge test for serial correlation in panel data models, the null hypothesis of this test being that there is no such correlation in the residuals. We can see from Table 5 that the null hypothesis is rejected at a 1% significance level, we can also see that the intra-class correlation coefficient is above 0.5 in both models which is sizeable. Due to these results and the presence of cross-sectional dependence and heteroskedasticity, clustered standard errors around vehicles will be used in the estimation of this model.

4.2.5. Misspecification

Misspecification refers to any errors in the form or specification of the model we are estimating, this can be tested using Ramsey's RESET test where the null hypothesis is that the model is correctly specified. The results in Table 5 show that the null hypothesis is rejected at a 1% significance level in both models, this could be causing the rejection of the null in the previously mentions tests and could be due to omitted variables or could be suggesting the use of a random effects approach instead, this will need to be tested for.

4.2.6. Hausman Test

The Hausman Test assists with the decision between using a random or fixed effects approach which each handles the constants for individual-specific effects differently. This test investigates whether the regressors are correlated with the unobserved, vehicle specific effects. The null hypothesis for this test is that there is no correlation between the two, which is the assumption of a random effects approach, so the random effects model is preferred. Table 5 shows that the null hypothesis is rejected at a 1% significance level, therefore a fixed effects approach is strongly preferred for this analysis. Following this, the models were implemented with various variables removed and added. The coefficients were varied across each regression, suggesting that the cause of the misspecification could be due omitted variables.

4.2.7. Multicollinearity

Multicollinearity is where a linear relationship between two or more explanatory variable exists, which could lead to inflated standard errors and imprecise coefficients. These problems can be exacerbated in a fixed effects model, which relies on within-group variation, where multicollinearity can be more prevalent. Multicollinearity can be detected using the Variance Inflation Factors (VIF), where a VIF value exceeding 10 is usually indicative of problematic multicollinearity, we can see from Table 5 that the mean VIF value for each model is below 2.2, in addition to this no variable specific VIF value exceeds 5.34, therefore the coefficient estimates are relatively precise and can be interpretable with confidence.

Table 5: Robustness Test Results

	Petrol		Diesel	
Test	Statistic	p-value	Statistic	p-value
Hausman	79242.38	***	74777.43	***
Wooldridge	44454.56	***	15348.96	***
Modified-Wald	1.60E+13	***	6.70E+12	***
Pesaran's CD	35.964	***	27.511	***
Ramsay RESET	466.03	***	271.09	***
VIF	2.11		2.14	
Intra-Class Correlation	0.55		0.51	

Note: *, ** and *** represent as 10%, 5% and 1% significance level respectively

4.2.8. Instrument Robustness Tests

Before we estimate the Two-Stage Least Squares regressions, we must check the strength of the chosen instrument to make an informed decision on whether to include it, this can be achieved through the first-stage regression results which are shown in Table 6. The first statistic to be analysed is the F-statistic, this statistic tests for the joint significance of the excluded instrument, we can see that the values for this are significantly higher than 10, suggesting that the instrument is very strong. The Anderson Canonical Correlation LM Statistic tests whether the model is under identified where the null hypothesis is that the model is under identified and that the Brent spot price cannot sufficiently explain the price of fuel. We can see from Table 6 that the null hypothesis is rejected at a 1% significance level in both models, suggesting that the models are identified, and Brent spot price could be capable in explaining the price of fuel. The next test is for weak instruments and is conducted by comparing the Cragg-Donald Wald F-Statistic to the Stock-Yogo critical values, if the F-Statistic is greater than the critical values then the instrument is not considered weak, and the risk of weak-instrument bias is lower. From Table 6, the Cragg-Donald Wald F-Statistics are large and exceed all Stock-Yogo critical values, suggesting that Brent spot price is a strong instrument. Finally, the coefficients on lnBSP are statistically significant, also indicating that it could be relevant as an instrument.

Table 6: First-Stage Regression Statistics

Tests	Petrol IV 2SLS	Diesel IV 2SLS
First-Stage F-Statistic	5.80E+06 ***	1.70E+06 ***
Cragg-Donald Wald F-Statistic	5.80E+06 ***	1.70E+06 ***
Anderson Canonical Correlation LM statistic	3.00E+06 ***	1.20E+06 ***
Stock-Yogo	F-Statistic > All Critical Values	F-Statistic > All Critical Values
lnBSP Coefficient	0.288 ***	0.243 ***

Note: *, ** and *** represent as 10%, 5% and 1% significance level respectively

5. Results

5.1. Fuel Price Elasticities on Vehicle Miles Travelled

The regression results from the fixed effects and two-stage least squares regressions are presented in Table 7 and the first thing to note is that we see a significant, positive sign on the coefficient for petrol prices in the two-stage least squares (2SLS) model and a negative sign in the fixed effects model, representing η_{PP}(VMT). This is unexpected and suggests that a 1% increase in pence per litre of petrol is associated with a 0.11% decrease and 0.55% increase in miles travelled in the fixed effects and 2SLS models respectively, on average, ceteris paribus. This difference could be due the misspecification and a possible omitted variable bias, or the positive value in the 2SLS due to the price of substitutes also increasing, for example the price of trains and buses rising by a larger percentage than the rise in the price of petrol and Brent spot price. This difference could also mean that there could be a potential problem with endogeneity or the validity of Brent spot price as an instrument. We know from Table 3 that 98% of our sample are from urban postcodes where there is greater alternative transport available, therefore including the price of alternatives may warrant further investigation in future research. This is also shown by a negative elasticity between income and VMT in all models, suggesting that in the petrol fixed effects model, a 1% increase in income would be associated with a 1.88% fall in mileage from petrol vehicles, on average, ceteris paribus. This suggests that the substitution effect may be dominant here, as we still observe a positive elasticity between mileage and fuel prices in all models except the petrol fixed effects – therefore including the price of alternatives in the regression would be appropriate.

We also observe from Table 7 that the statistically significant elasticity of demand for VMT from diesel vehicles with respect to the price of diesel (η_{DP}(VMT)) is 0.07% in the fixed effects model and 1.33% in the 2SLS, on average, ceteris paribus. Suggesting a strong, positive relationship both before and after instrumenting where diesel-fuelled vehicles actually increase their mileage following an increase in the price of diesel. This result is counterintuitive but could be because diesel vehicles are considered more fuel efficient, especially for long journeys so it would make sense for an individual to drive a diesel car instead of a petrol to lower their P_S and maintain a constant utility from driving. Therefore, investigating diesel vehicles in both rural and urban environments to analyse driving patterns would be useful.

5.2. Model Statistics

Table 7 also shows the model statistics, we can see that the reported R^2 statistic for all four models is low. From the fixed effects models the R^2 values suggest that roughly 11% of the within variation in the vehicle log VMT is explained by the within variation in the set of included covariates for petrol vehicles, and 13% for diesel vehicles, whilst the instrumented fixed effects models show similar estimates of 11% for both petrol and diesel. These estimates are higher than those of [6]Craglia and Cullen's (2020)who estimated similar regressions, but are still very low, capturing the difficulty in fully explaining changes in a vehicle's mileage using data as there are many additional variables that could improve the R^2 values, some of which may be immeasurable. We also note that large F and Wald Chi^2 statistics are reported in all models, suggesting that the fixed effects are found to be jointly statistically significant.

Rho represents the intra-cluster correlation coefficient and informs us on the proportion of total variance in the model that is due to vehicle specific effects. We observe from Table 7 that the estimates for rho are very high with the lowest being from the petrol fixed effects model, suggesting that 92% of the variance in the model is due to differences between vehicles. This is expected following the high F and Wald Chi^2 statistics and demonstrates why using fixed effects models is appropriate.

$Sigma_u$ and $sigma_e$ represent the standard deviations of the unobserved, vehicle specific effects and error terms respectively, with their relationship being captured in rho. From Table 7, the estimates for these are all reported as very high, especially $sigma_u$ which supports the conclusions drawn from the other statistics that significant differences between vehicles affecting VMT exist and persist over time. However, this is expected from such a large sample size where every vehicle and driver combination are so unique.

5.3. Other Estimates

We also observe statistically significant estimates for every coefficient in Table 7, with each model producing similar results. We find that a 1% increase in the age of a petrol vehicle, corresponds to a 0.05% decrease in VMT in the fixed effects model, and a 0.07% decrease in the 2SLS model. And a 1% increase in the age of a diesel vehicle leads to a 0.03% increase in VMT in the fixed effects model, and a 0.004% increase in the 2SLS model, on average ceteris paribus. These results show that older petrol vehicles are more likely to be driven by individuals with higher driving needs such as a longer commute to work or living in a rural area, but these are still very inelastic.

Vehicle Mileage is also shown to have a significant, negative elasticity with respect to the driving population of the vehicle's postcode. As observed from the diesel 2SLS results Table 7, a 1% increase in the driving population of the diesel-fuelled vehicle's postcode at test leads to a 6.5% decrease in vehicle mileage, on average, ceteris paribus. This is as expected because the variable is representing traffic in the individual's area which could motivate the driver to seek alternative methods of transport, such as in London where there is perceived to be greater traffic, the individual may prefer to travel via the London Underground. This further reinforces the evidence suggesting that the substitution effect dominates in this sample.

The weather effects captured in the model show that there is a significant, negative relationship between heating degree months of the vehicle's postcode and vehicle mileage, and a positive relationship between average monthly precipitation in the vehicle's postcode and vehicle mileage. These effects are expected because as heating degree months increase, it indicates colder weather which would be likely to discourage travel, especially for non-essential trips. And as the monthly average precipitation in the vehicle's postcode increases, this could discourage some alternative transport methods such as walking or cycling, and drivers may travel via longer routes to avoid any congestion or hazards.

Our included dummy variables, representing the time specific events for the COVID-19 lockdowns and the Russo-Ukrainian conflicts show a negative and positive relationships respectively. For example, the mileage of a diesel vehicle from the fixed effects model being tested during a lockdown month is expected to be 14% higher than that of a diesel car being tested outside of a lockdown month, on average, ceteris paribus. This result is also counterintuitive but could be due to the sample, we may have observed a number of vehicles belonging to essential workers who could have increased their travel during the lockdowns, or this could be due to MOT tests being delayed during the lockdowns where individuals would have been allowed to accumulate a higher mileage over a longer driving period. Table 7 also shows from the petrol fixed effects model that the mileage of a petrol vehicle tested during the Russo-Ukrainian conflicts increase by 28% more than that of a petrol vehicle tested before the conflicts started, on average, ceteris paribus. This could be due to fuel price uncertainty where individuals and business may have increased their travel to secure goods or services after anticipating future increases in fuel price. This could also be due to a potentially significant portion of observations in the sample being affected by the lockdowns, as we've seen a large percentage change in VMT from the lockdown variable, vehicles tested during the conflict's months could have relatively higher mileage due to a resurgence in economic activity following these months.

Table 7: Regression Results

Parameter	Petrol Fixed Effects	Diesel Fixed Effects	Petrol IV 2SLS	Diesel IV 2SLS
lnPP	-0.11 ***	x	0.55 ***	x
lnDP	x	0.07 ***	x	1.33 ***
lnGDP	2.26 ***	2.14 ***	1.6 ***	0.99 ***
lnY	-1.88 ***	-1.91 ***	-1.59 ***	-1.16 ***
lnAGE	-0.05 ***	0.03 ***	-0.07 ***	-0.004
lnHDM	-0.01 ***	-0.01 ***	-0.01 ***	-0.15 ***
lnPRCP	0.01 ***	0.01 ***	0.03 ***	0.04 ***
lnDPOP	-3.97 ***	-4.49 ***	-4.96 ***	-6.5 ***
LOCK	0.2 ***	0.14 ***	0.13 ***	0.032 ***
WAR	0.28 ***	0.24 ***	0.19 ***	0.001
R^2	0.11	0.13	0.11	0.11
ρ	0.92	0.95	0.95	0.97
σ_u	2.06	2.39	2.51	3.39
σ_e	0.6	0.57	0.6	0.58
F-Statistic	68486	36787	x	x

Note: *, ** and *** represent as 10%, 5% and 1% significance level respectively

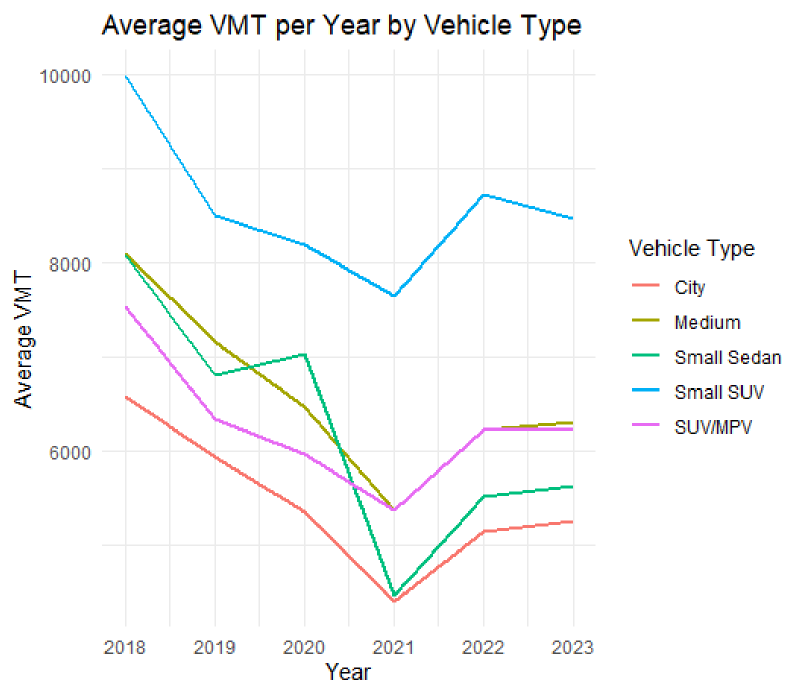
5.4. Subgroup Analysis by Fuel Type, Vehicle Classification, and Geographic Classification

This analysis will also include multiple regressions of the model described above, to test whether the results are consistent when the vehicle belongs in different subgroups.

Different car classifications are likely to have different usage patterns, for example an SUV might be more sensitive to changes in fuel price as they are generally seen as less fuel-efficient vehicles, whilst small sedans could contain more luxury vehicles, which may be less inelastic to economic changes due to the wealth required to purchase one. Different vehicle classes are also heterogenous in driving patterns, as we can see in Figure 3, where small SUVs on average record a higher mileage than any other vehicles where they typically have larger engines and are intended for longer journeys and more frequent usage, and city vehicles record a lower mileage, this could be because they typically have smaller engines and are used for short-distance travel or the owners prefer to use other forms of transport due to individual needs. Figure 3 also illustrates the fall in mileage following the COVID-19 lockdowns, although the troughs are in 2021; this is because the lockdowns occurred in between 2020 and 2021, so the effects would be highlighted at the vehicle's next test which would be in 2021.

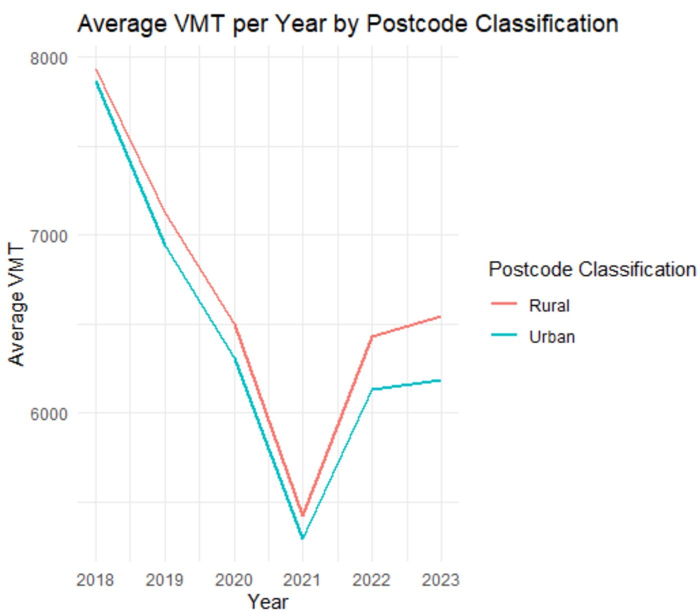
Understanding the difference in rebound effects between these classes may help any targeted policy implementations for sufficient fuel efficiency improvements in all vehicles in the UK. Separate regressions will be implemented for each vehicle class, which will also be split by fuel type to investigate this.

Figure 3: Yearly Average Mileage by Vehicle Classification 2018-2023



Rural and urban areas are likely to have very different driving behaviours due to the availability of transport alternatives such as public transport. We know that for the negative of our coefficients, $-\eta_{PE}(S)$ to approximate $\eta_{\varepsilon}(S)$, one of the assumptions is that the drivers are limited in shifting travel to other types of transport ([3]Chan and Gillingham (2015), [33]Sorrell and Dimitropoulos (2008)), therefore the rural based vehicles would have stronger theoretical standing when estimating the rebound effect. Therefore 4 additional models will be implemented, using petrol and diesel vehicles from rural and urban areas in the UK. Figure 4 illustrates the average VMT of vehicles in both urban and rural postcodes, we can see that both lines follow the same pattern over time with a significant drop following the lockdowns, but rural vehicles have a steeper slope than urban vehicles in the rise between 2022 and 2023, indicating that urban vehicles may have been more sensitive to hikes in fuel prices during this period following the outbreak of the Russo-Ukrainian conflict.

Figure 4: Yearly Average Mileage by Postcode Classification 2018-2023



These additional models must be estimated separately because vehicle classification is a time invariant characteristic, and so is the rural/urban indicator because vehicles that changed postcodes were removed from the sample, therefore these would be perfectly collinear with the fixed effects term if included as dummy variables. Fixed effects models will be implemented with the same model specifications as in Section 4.1, using subsets of the data for each subgroup.

5.5. Fuel Price Elasticities by Vehicle Classification

Table 8 presents the results of these regressions and we observe statistically significant elasticities of demand for VMT with respect to fuel price in every subgroup of both petrol and diesel vehicles except for small sedans, this could be due to not enough observations from this group, as we know from Table 3 that only 0.01% of the sample has been observed as small sedans so it is possible for the roughly 100,000 observations to not yield significant results, especially as they are split between fuel types. We also observe positive elasticities from diesel vehicles and negative elasticities from petrol vehicles, largely similar to the results from Table 7.

We also observe that the diesel-fuelled vehicles with larger engines are more elastic in their response to yearly mileage after changes in their fuel price, whereas petrol-fuelled vehicles follow similar trends except for city class vehicles which are associated with an $\eta_{PP}(VMT)$ = -0.96%, and diesel-fuelled SUV/MPV's with an $\eta_{DP}(VMT)$ = 0.134%. This could be because the larger vehicles, particularly diesel-fuelled, may be more prevalent in rural areas and tend to drive longer distances and more frequently.

Our findings also show that medium vehicles have more inelastic responses in regard to changes in their fuel price compared to the larger vehicle classes, with petrol-fuelled medium vehicles returning an $\eta_{PP}(VMT)$ = -0.114%, and their diesel counterparts returning $\eta_{DP}(VMT)$ = 0.068%. One possible reason behind this would be that medium cars are the most general-purpose class of vehicles, used for both short and long journeys, therefore they may be used more as a necessity form of transport and are therefore less responsive to changes in price. We can see in Figure 3 that medium vehicles follow the same trends in mileage as the other classes but tends to stay in the middle, indicating that it may be representing the average, everyday use vehicle.

Table 8: Subgroup Regressions Results

Petrol			Diesel		
Parameter	Coefficient	$\eta_{PE}(S)$	Parameter	Coefficient	$\eta_{PE}(S)$
lnVMT (City)	-0.96 ***	-0.96%	lnVMT (City)	0.042 ***	0.04%
lnVMT(Medium)	-0.114 ***	-0.11%	lnVMT(Medium)	0.068 ***	0.07%
lnVMT(SUV/MPV)	0.089 ***	0.09%	lnVMT(SUV/MPV)	0.134 ***	0.13%
lnVMT(Small SUV)	-0.55 *	-0.55%	lnVMT(Small SUV)	0.115 ***	0.12%
lnVMT(Small Sedan)	0.444	0.44%	lnVMT(Small Sedan)	0.791	0.79%
lnVMT (Rural)	-0.214 ***	-0.21%	lnVMT (Rural)	-0.041 ***	-0.04%
lnVMT(Urban)	-0.104 ***	-0.10%	lnVMT(Urban)	0.0786 ***	0.08%

Note: *, ** and *** represent as 10%, 5% and 1% significance level respectively

5.6. Fuel Price Elasticities by Postcode Classification

The section investigates how the elasticity of demand for VMT with respect to fuel price varies by whether the postcode is classified as urban or rural in both petrol and diesel vehicles.

The first result to note from Table 8 is that the coefficient for the logged price of diesel on logged VMT is positive in urban areas and negative and more inelastic in rural areas on average, ceteris paribus. The substitution effect could be affecting the sample during this time period, where the prices of alternative such as trains, undergrounds, or trams may be rising by a higher rate than diesel prices and could be causing a shift towards driving. We must also consider commuting patterns, if there is an economic boom then they may be an increase in employment opportunities where individuals may need to commute further or more often; the positive elasticities shown for GDP in Table 7 suggest that this may be the case. Also, in rural areas there is more of a necessity for driving over potentially longer distances and few alternatives, this could be why we observe rural-based vehicles recording a higher yearly mileage than urban-based vehicles in Figure 4. It is also possible for rural areas of the UK to experience less or less noticeable diesel price fluctuations, or even delayed effects – given the data used in this analysis is on UK average diesel prices at pumps then it is possible these postcode-specific effects are not being captured in the rural areas of the UK which is why this coefficient is less elastic.

We also note that VMT is more elastic in petrol-fuelled vehicles that are tested in rural areas than in urban areas on average, ceteris paribus. This could be because there are typically shorter trip distances in urban areas, or we may be observing an income effect here where the individuals in our sample that reside in urban areas may be achieving relatively higher incomes, and the increase in fuel prices may not significantly impact their driving behaviour.

The negative elasticities of demand for mileage with respect to petrol prices in both urban and rural postcodes shown in Table 8 align more with expectations and suggest that in rural areas they are more sensitive to changes in the price of petrol, possibly due to fewer transport substitutes available or lower average incomes in their area, and will change their driving patterns to respond to this. Rural-Based vehicles will typically have a higher mileage due to the necessity of use and potentially longer driving distances which we observe in Figure 4, therefore it may be possible for the individuals to reduce non-essential or consolidate trips.

6. Caveats and Future Research

One limitation of the method used in this analysis is that the models failed the RESET test and therefore is misspecified. Also, because the coefficients changed upon the addition and removal of different variables, this means that there may be an omitted variable bias leading to incorrect estimations of the resulting elasticities, this problem is likely to never be fully solved due to the sheer number of factors that influence vehicle mileage – but can be improved by including variables such as the prices of alternatives and government policies such as ULEZ or congestion charges since we observe a strong substitution effect in our data, or employment rates to better capture commuting trends in the UK. However, even if the models are misspecified and have an omitted variable bias, they can still provide practical and useful information for policymakers.

Of course, local data rather than nationwide average or postcode on all variables including GDP, income and fuel prices would be ideal and better improve this research to provide more accurate estimates which would be more robust to the regional heterogeneity within the UK, although this may not be obtainable at the granularity required or consistent throughout all local authorities.

A major limitation to this research is a potential sample selection bias, after the cleaning process 98% of remaining observations were tested in an urban location, therefore there is likely an urban bias skewing the results which may not be generalising rural areas well. This is likely to have occurred because most rural-tested vehicles in the original sample changed their postcode during the sample period. Also, there is potential for the misclassification of some vehicles in the sample arising from the fuzzy-matching procedure which may be slightly skewing the results.

Another limitation would be that whilst Brent spot price was found to not be weak or overidentified, the relationship between this and fuel prices may vary by region which would mean that we haven't fully captured the exogenous variation if fuel prices, leading to biased results. Including other instruments such as exchange rates or fuel tax policies which are likely to be correlated with fuel prices but not with vehicle mileage may improve the robustness of the regressions.

Using this data in a fixed effects approach where there is a relatively small time period and a large number of fixed effects, could be reducing the degrees of in the model which would compromise the efficiency of the estimators. Also, due to this data being an unbalanced panel where some vehicles drop out of the sample at different points, there was no suitable test found for stationarity which could lead to spurious regressions. There is likely little risk of stationarity due to the small time period, but this could still present a problem.

Another suggestion for future research would be that the [37]VCA (2024) database also provides information on each specific vehicles Worldwide Harmonised Light Vehicles (WLTP), which, according to the [36]AA (2018), scores vehicles based on measurements of its fuel economy (miles per gallon), carbon dioxide and other pollutants when travelling at different speeds and temperatures with different loads and tyre pressures. This could be interesting if included as a proxy for fuel efficiency, however this testing system has only been introduced in 2018 and differs from the previous New European Driving Cycle (NEDC) test. Therefore, it was not appropriate to include WLTP scores in this data but for future research when the testing procedures are consistent throughout the sample, this variable may yield some interesting results.

7. Discussion

Table 9: Direct Rebound Effects from Other Studies

Authors	Style of Data	Time-Period	Country	Metric	Estimated Direct Rebound Effect
[6]Craglia and Cullen (2020)	Panel, Odometer readings	2006-2016	United Kingdom	Miles	4.6%
[7]De Borger et al (2016)	Panel, Odometer readings	2001-2011	Denmark	Kilometres	7.5-10%
[30]Small and Van Dender (2005)	Panel, Aggregate	1966-2001	United States	Miles	11% in SR, 50% in LR
[34]Stapleton et al (2016)	Time Series, Aggregate	1970-2011	United Kingdom	Kilometres	19%
[19]Greene et al (1999)	Panel, Household Survey	1979-1994	United States	Miles	20%
[14]Gillingham and Munk-Nielsen (2019)	Panel, Odometer Readings	1998-2011	Denmark	Kilometres	30%
[20]Huntington (2024)	Time Series, Aggregate	1949-2019	United States	Miles	20-50%
[4]Chitnis et al (2020)	Panel, Aggregate	1964-2015	United Kingdom	Kilometres	54%
[12]Ficano and Thompson (2014)	Cross-Sectional, Household Survey	2009	United States	Miles	75.5-77.8%
[25]Moshiri and Aliyev (2017)	Panel, Household Survey	1997-2009	Canada	Kilometres	88%

Since the 2SLS regressions show highly counterintuitive results, indicating that the instrument used may be inappropriate or more instruments should be included, the fixed effects estimates are preferred between the two, suggesting a direct rebound effect of approximately 11% for petrol vehicles and that diesel vehicles have no direct rebound effect and reduce their mileage by 7% following an improvement in fuel efficiency. Across the vehicle classification subsamples we observe a weighted average rebound effect of approximately 33.67% in petrol vehicles and once again no rebound effect for diesel vehicles, who decrease their mileage by 6.82%. Across rural and urban subsamples, we find a weighted average rebound effect of approximately 10.62% for petrol vehicles and no rebound effect for diesel vehicles who reduce their mileage by 7.62%. Finding the vehicle classification subgroup results more robust, we estimate a weighted average rebound effect of approximately 8.9% across all vehicles.

Table 9 presents estimates of direct rebound effects in personal vehicles from various studies and highlights how much the estimates differ between time periods, countries, and methods of estimation. We can see that our estimate lies in the lower end, closer to the estimate provided by [7]De Borger et al (2016), who used the same style of data in Denmark. The result is lower than most estimates, suggesting that vehicle efficiency improvements are leading to lower increases in demand for mileage in the UK than countries such as the US or Canada, this is likely because the UK has a very different transport infrastructure than these countries. In the UK there are fuel taxes, urban congestion charges and extra parking costs if a vehicle that does not meet minimum emission standards and drives or parks in a 'clean air zone', and ULEZ in certain cities which would discourage driving, whereas The US and Canada are much more reliant on personal transport as they are much larger countries and have potentially longer journeys between destinations.

We also estimate a larger rebound effect than [6]Craglia and Cullen (2020) (4.6%) whilst sampling from the same MOT dataset including 4 more recent years of results, suggesting that the short-run direct rebound effect has increased in the UK over the last 3 years. However, this could be due to the heterogeneity between the two samples and sample sizes and differences in specifications.

Interestingly, we find counterintuitive results in diesel vehicles which differ from those of [14]Gillingham and Munk-Nielsen (2019), they find that diesel vehicles are more price elastic and tend to have larger rebound effects than petrol whereas we find that diesel vehicles in our sample actually tend to drive less following increases in fuel efficiency. Again, this could be from differences in countries and policies between the UK and Denmark, or the rising popularity in electric and hybrid vehicles between 2011 and 2023.

Another important contribution of this study is the examination of the rebound effect across urban and rural areas and different vehicle classifications, the finding that petrol city cars exhibit a higher rebound effect compared to larger vehicles, such as SUVs, is consistent with the expectation that smaller, more fuel-efficient cars may be used more intensively when cost of fuel, and therefore the energy cost driving falls. In contrast, the lower rebound effect in rural diesel vehicles highlights that where there exists greater dependency on diesel vehicles for essential travel, this could limit their response to fuel efficiency improvements.

7.1. Policy Implications

These results show that while improvements in fuel efficiency do result in a small increase in the demand for mileage in petrol vehicles, this effect is relatively small compared to findings from other studies. These findings suggest that policies which are aimed at improving fuel efficiency will still lead to significant reductions in both emissions and fuel consumption, especially from diesel vehicles. Additionally, the large variation in results across different subgroups suggest the need for targeted policies across different types of vehicles and regions of the UK. For example, promoting public transport in urban areas by improving efficiency or reducing prices would be effective in mitigating these rebound effects, especially in petrol vehicles. Additionally, expanding public transport in rural areas would likely have an impact by reducing the necessity of driving.

8. Conclusions

This study sought to estimate the short-run direct rebound effects of fuel efficiency improvements across petrol and diesel vehicles in the UK, using a combination of fixed effects and instrumental variable approaches. Our results suggest that the average rebound effect across all vehicles is approximately 8.9%, with petrol vehicles exhibiting a much higher rebound effect compared to diesel vehicles, which showed a negative rebound effect. These results suggest that fuel efficiency improvements in diesel cars have led to a fall in mileage, especially in urban areas.

These findings carry important policy implications, while improvements in fuel efficiency can significantly decrease fuel consumption, the complementary policies within the UK such as congestion charges may need to be more stringent on petrol vehicles rather than diesel vehicles to mitigate the increases in driving from this group. The variation across different subgroups of vehicles suggest that policymakers must tailor any interventions for specific areas and vehicle types to achieve optimal outcomes.

This study does have its limitations, the counterintuitive results from the instrumental variables approaches suggest that more robust instruments are required to properly account for potential endogeneity in fuel prices. The estimations would also benefit from a finer granularity of economic data rather than aggregations to estimate regional differences more effectively. This study also does not consider the indirect rebound effect which may be more prevalent in diesel vehicle owners, who are likely to be accumulating savings by reducing their driving after increases in fuel efficiency.

In conclusion, this study provides valuable insights into the rebound effects following fuel efficiency improvements, offering evidence that diesel and petrol vehicles respond differently to these changes. As the UK transport sector continues to evolve, understanding these dynamics will be essential for crafting policies that balance efficiency gains with sustainable travel behaviour.

9. References

1- Baker, C., Brown, J., Barber, S. and Kirk-Wade, E. (2021). Coronavirus: A history of English lockdown laws. *commonslibrary.parliament.uk*, [online] 1(1). Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-9068/.

2- Brookes, L., 2000. Energy efficiency fallacies revisited. *Energy policy*, *28*(6-7), pp.355-366.

3- Chan, N. W., & Gillingham, K. (2015). The Microeconomic Theory of the Rebound Effect and Its Welfare Implications. *Journal of the Association of Environmental and Resource Economists*, *2*(1), 133–159. https://doi.org/10.1086/680256

4- Chitnis, M. et al. (2020) Rebound Effects for Household Energy Services in the UK. *The Energy journal (Cambridge, Mass.)*. [Online] 41 (4), 31–60.

5- Chitnis, M., & Sorrell, S. (2015). Living up to expectations: Estimating direct and indirect rebound effects for UK households. *Energy Economics*, *52*, S100–S116. https://doi.org/10.1016/j.eneco.2015.08.026

6- Craglia, M., & Cullen, J. (2020). Do vehicle efficiency improvements lead to energy savings? The rebound effect in Great Britain. *Energy Economics*, *88*, 104775-. https://doi.org/10.1016/j.eneco.2020.104775

7- De Borger, B., Mulalic, I., & Rouwendal, J. (2016). Measuring the rebound effect with micro data: A first difference approach. *Journal of Environmental Economics and Management*, *79*, 1–17. https://doi.org/10.1016/j.jeem.2016.04.002

8- Department for Transport, (2024). *Anonymised MOT tests and results*. [online] Available at: https://www.data.gov.uk/dataset/e3939ef8-30c7-4ca8-9c7c-ad9475cc9b2f/anonymised-mot-tests-and-results [Accessed 15 May 2024].

9- Department for Transport (2022). *Vehicle licensing statistics data tables*. [online] GOV.UK. Available at: https://www.gov.uk/government/statistical-data-sets/vehicle-licensing-statistics-data-tables. [Accessed 15 May 2024]

10- Druckman, A., Chitnis, M., Sorrell, S., & Jackson, T. (2011). Missing carbon reductions? Exploring rebound and backfire effects in UK households. *Energy Policy*, *39*(6), 3572–3581. https://doi.org/10.1016/j.enpol.2011.03.058

11- EIA. (2024). *Petroleum & Other Liquids*. [online] Available at: https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rbrte&f=m [Accessed 4 May 2024].

12- Ficano, C. C., & Thompson, P. (2014). ESTIMATING REBOUND EFFECTS IN PERSONAL AUTOMOTIVE TRANSPORT: GAS PRICE AND THE PRESENCE OF HYBRIDS. *The American Economist (New York, N.Y. 1960)*, *59*(2), 167–175. https://doi.org/10.1177/056943451405900207

13- FRED.ST.LOUIS. (2024). *U.S. Dollars to U.K. Pound Sterling Spot Exchange Rate*. [online] Available at: https://fred.stlouisfed.org/series/DEXUSUK/ [Accessed 4 May 2024].

14- Gillingham, K., & Munk-Nielsen, A. (2019). A tale of two tails: Commuting and the fuel price response in driving. *Journal of Urban Economics*, *109*, 27–40. https://doi.org/10.1016/j.jue.2018.09.007

15- GOV.UK (2020). *Driving in a Clean Air Zone*. [online] GOV.UK. Available at: https://www.gov.uk/guidance/driving-in-a-clean-air-zone.
16- Gov.uk. (n.d.). *Energy Consumption in the UK (ECUK) 1970 to 2022*. [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1187617/Energy_Consumption_in_the_UK_2023.pdf [Accessed 23 Aug. 2024].
17- Gov.uk. (2024). *2023 UK greenhouse gas emissions, provisional figures*. [online] Available at: https://assets.publishing.service.gov.uk/media/6604460f91a320001a82b0fd/uk-greenhouse-gas-emissions-provisional-figures-statistical-release-2023.pdf [Accessed 23 Aug. 2024].
18- GOV.UK (2022). Weekly road fuel prices. [online] GOV.UK. Available at: https://www.gov.uk/government/statistics/weekly-road-fuel-prices. [Accessed 23 May. 2024].
19- Greene, D. L., Kahn, J. R., & Gibson, R. C. (1999). Fuel Economy Rebound Effect for U.S. Household Vehicles. *The Energy Journal (Cambridge, Mass.)*, *20*(3), 1–31. https://doi.org/10.5547/issn0195-6574-ej-vol20-no3-1
20- Huntington, H.G., 2024. US gasoline response to vehicle fuel efficiency: A contribution to the direct rebound effect. Energy Economics, p.107655.
21- Khazzoom, J. D. (1980). Economic Implications of Mandated Efficiency in Standards for Household Appliances. The Energy Journal, 1(4), 21–40. https://doi.org/10.5547/issn0195-6574-ej-vol1-no4-2
22- Met Office Climate Data Portal. (2024). *Annual Heating Degree Days - Projections (12km)*. [online] Available at: https://climatedataportal.metoffice.gov.uk/datasets/TheMetOffice::annual-heating-degree-days-projections-12km/about [Accessed 20 Jul. 2024].
23- Met Office (2006): MIDAS: UK Daily Rainfall Data. NCAS British Atmospheric Data Centre, *[online]*. https://catalogue.ceda.ac.uk/uuid/c732716511d3442f05cdeccbe99b8f90/ [Accessed 20 Jul. 2024].
24- Met Office (2023): MIDAS Open: UK daily temperature data, v202308. NERC EDS Centre for Environmental Data Analysis, [online] doi:10.5285/220b9b8ffbed43fcbbd323e739118f6c. https://dx.doi.org/10.5285/220b9b8ffbed43fcbbd323e739118f6c [Accessed 20 Jul. 2024].
25- Moshiri, S., & Aliyev, K. (2017). Rebound effect of efficiency improvement in passenger cars on gasoline consumption in Canada. *Ecological Economics*, *131*, 330–341. https://doi.org/10.1016/j.ecolecon.2016.09.018
26- Office for National Statistics. (2024). EARN01: Average weekly earnings. [online] Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/averageweeklyearningsearn01 [Accessed 14 Jul. 2024].

27- Office for National Statistics. (2024). *Monthly GDP and main sectors to four decimal places*. [online] Available at: https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/monthlygdpandmainsectorstofourdecimalplaces [Accessed 16 Jul. 2024]

28- Office for National Statistics. (2024). *Online ONS Postcode Directory (Live)*. [online] Available at: https://geoportal.statistics.gov.uk/datasets/ons::online-ons-postcode-directory-live/about [Accessed 20 Jul. 2024].

29- Saunders, H. D. (1992). The Khazzoom-Brookes Postulate and Neoclassical Growth. The Energy Journal (Cambridge, Mass.), 13(4), 131–148. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol13-No4-7

30- Small, K. A., & Van Dender, K. (2005). The Effect of Improved Fuel Economy on Vehicle Miles Traveled: Estimating the Rebound Effect Using U.S. State Data, 1966-2001.

31- Sorrell, S., 2009. The Rebound Effect: definition and estimation. *International handbook on the economics of energy*, pp.199-233.

32- Sorrell, S., Dimitropoulos, J., & Sommerville, M. (2009). Empirical estimates of the direct rebound effect: A review. *Energy Policy*, *37*(4), 1356–1371. https://doi.org/10.1016/j.enpol.2008.11.026

33- Sorrell, S., & Dimitropoulos, J. (2008). The rebound effect: Microeconomic definitions, limitations and extensions. *Ecological Economics*, *65*(3), 636–649. https://doi.org/10.1016/j.ecolecon.2007.08.013

34- Stapleton, L., Sorrell, S., & Schwanen, T. (2016). Estimating direct rebound effects for personal automotive travel in Great Britain. *Energy Economics*, *54*, 313–325. https://doi.org/10.1016/j.eneco.2015.12.012

35- Theaa.com. (2018). *Official fuel consumption figures and WLTP | The AA*. [online] Available at: https://www.theaa.com/driving-advice/fuels-environment/official-fuel-consumption-figures#:~:text=The%20new%20WLTP%20test%20lasts%20longer [Accessed 8 Jul. 2024].

36- Transport for London (2019). *Congestion Charge*. [online] Transport for London. Available at: https://tfl.gov.uk/modes/driving/congestion-charge [Accessed 8 Jul. 2024].

37- Transport for London (2024). *Ultra Low Emission Zone*. [online] Transport for London. Available at: https://tfl.gov.uk/modes/driving/ultra-low-emission-zone [Accessed 8 Jul. 2024].

38- Vehicle Certification Agency. (2024). *Car fuel data, CO2 and vehicle tax tools*. [online] Available at: https://carfueldata.vehicle-certification-agency.gov.uk/ [Accessed 1 Jul. 2024].

39- Walker, N. (2023). *Conflict in Ukraine: A timeline (2014 – eve of 2022 invasion)*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-9476/?trk=public_post_comment-text [Accessed 25 Aug. 2024].

10. Appendix

10.1. *Summary*

Variable	Obs	Mean	Std. dev.	Min	Max
VMT	10,428,847	6350.68	5117.789	.0223714	100000
PP	6,343,607	135.8827	18.69171	104.87	191.55
DP	4,091,396	143.0878	21.92049	111.7	199.22
GDP	10,428,847	98.78462	4.468308	74.5691	102.8381
Y	10,428,847	182.6313	14.46524	161.9497	211.1207
AGE	10,428,847	9.652286	5.196006	0	1003.4
HDM	10,428,847	5.080542	4.11768	.0004168	16.72138
PRCP	10,428,847	3.325158	2.341273	.0121739	20.00313
DPOP	10,428,847	365743.4	166251.1	11113	966256
bsp	10,428,847	73.11196	19.57388	18.38	122.71
LOCK	10,428,847	.0275732	.1637466	0	1
WAR	10,428,847	.315355	.4646571	0	1

10.2. *Shapiro-Wilk Test Results*

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lnVMT	10,431,773	0.91276	1.5e+04	26.808	0.00000
lnPP	6,344,744	0.96047	6747.278	24.743	0.00000
lnDP	4,093,188	0.92590	1.2e+04	26.494	0.00000
lnGDP	10,431,773	0.72344	4.8e+04	30.024	0.00000
lnY	10,431,773	0.94429	9604.111	25.558	0.00000
lnAGE	10,431,771	0.97132	4944.589	23.708	0.00000
lnHDM	10,431,773	0.79378	3.6e+04	29.206	0.00000
lnPRCP	10,431,773	0.97169	4880.206	23.671	0.00000
lnDPOP	10,431,773	0.95723	7374.149	24.822	0.00000

Note: The normal approximation to the sampling distribution of W'
is valid for 4<=n<=2000.

10.3. *Petrol Fixed Effects*

10.3.1. Pesaran's CD Test Result

```
Pesaran's test of cross sectional independence =    35.964, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.423
```

10.3.2. Modified Wald Test Result

```
Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (1432400)  = 1.6e+13
Prob>chi2 =      0.0000
```

10.3.3. Wooldridge Test Result

```
Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1, 1151475) =  44454.560
           Prob > F =      0.0000
```

10.3.4. Ramsey's RESET Test Result

```
( 1)  fitted_values_sq = 0
( 2)  fitted_values_cu = 0

      F(  2,4906174) =  466.03
           Prob > F =    0.0000
```

10.3.5. Hausman Test Result

	Coefficients (b) Pfixed	(B) Prandom	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
lnPP	-.1105445	-.603814	.4932694	.0025661
lnGDP	2.255287	2.224217	.0310702	.0020096
lnY	-1.883961	-2.543299	.6593379	.0085004
lnAGE	-.0528345	-.3448857	.2920511	.0024285
lnHDM	-.0050144	-.0023911	-.0026234	.00018
lnPRCP	.0146032	.0224487	-.0078455	.0001693
lnDPOP	-3.970646	-.0079646	-3.962682	.0162099
LOCK	.1987007	.2733767	-.074676	.0004992
WAR	.280408	.3118739	-.0314659	.000358

```
                b = Consistent under H0 and Ha; obtained from xtreg.
 B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

    chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            = 112326.77
Prob > chi2 =     0.0000
```

10.3.6. VIF Test Results

Variable	VIF	1/VIF
lnPP	5.14	0.194539
WAR	3.77	0.265501
lnY	2.77	0.360604
lnGDP	2.72	0.367562
LOCK	1.37	0.729259
lnPRCP	1.15	0.872918
lnHDM	1.10	0.906542
lnDPOP	1.03	0.970419
rural	1.01	0.988913
lnAGE	1.00	0.995650
Mean VIF	2.11	

10.3.7. One-Way Analysis of Variance Output

One-way analysis of variance for lnVMT:

Number of obs = 6,338,585
R-squared = 0.6484

Source	SS	df	MS	F	Prob > F
Between vehicle_id	3620636.8	1432399	2.5276733	6.32	0.0000
Within vehicle_id	1963281.4	4906185	.40016457		
Total	5583918.2	6338584	.88094094		

Intraclass correlation	Asy. S.E.	[95% conf.	interval]
0.54575	0.00041	0.54495	0.54656

Estimated SD of vehicle_id effect .6933806
Estimated SD within vehicle_id .6325856
Est. reliability of a vehicle_id mean 0.84169
(evaluated at n=4.43)

10.3.8. Regression Output

Fixed-effects (within) regression Number of obs = 6,338,585
Group variable: **vehicle_id** Number of groups = 1,432,400

R-squared:
Within = 0.1116
Between = 0.0000
Overall = 0.0004

Obs per group:
min = 1
avg = 4.4
max = 6

F(9,4906176) = 68486.31
corr(u_i, Xb) = -0.9248 Prob > F = 0.0000

lnVMT	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnPP	-.1105445	.0050102	-22.06	0.000	-.1203643	-.1007247
lnGDP	2.255287	.0087525	257.67	0.000	2.238132	2.272441
lnY	-1.883961	.0101421	-185.76	0.000	-1.903839	-1.864083
lnAGE	-.0528345	.0025494	-20.72	0.000	-.0578312	-.0478379
lnHDM	-.0050144	.000289	-17.35	0.000	-.0055808	-.004448
lnPRCP	.0146032	.0004567	31.97	0.000	.013708	.0154983
lnDPOP	-3.970646	.0160954	-246.69	0.000	-4.002193	-3.9391
LOCK	.1987007	.0018775	105.83	0.000	.1950208	.2023806
WAR	.280408	.0011503	243.76	0.000	.2781534	.2826626
_cons	58.70324	.1740262	337.32	0.000	58.36216	59.04433
sigma_u	2.0631823					
sigma_e	.59624031					
rho	.92292167	(fraction of variance due to u_i)				

F test that all u_i=0: F(1432399, 4906176) = 6.58 Prob > F = 0.0000

10.3.9. Regression Output Following the Removal of GDP

Fixed-effects (within) regression Number of obs = 6,337,451
Group variable: **vehicle_id** Number of groups = 1,432,400

R-squared:
Within = 0.1005
Between = 0.0001
Overall = 0.0005

Obs per group:
min = 1
avg = 4.4
max = 6

F(8,1432399) = 51567.70
corr(u_i, Xb) = -0.9306 Prob > F = 0.0000

(Std. err. adjusted for **1,432,400** clusters in **vehicle_id**)

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnPP	.6280379	.0042334	148.35	0.000	.6197406	.6363352
lnY	-1.592458	.0125068	-127.33	0.000	-1.616971	-1.567946
lnAGE	-.1560389	.0032246	-48.39	0.000	-.1623591	-.1497187
lnHDM	-.0035672	.0003576	-9.98	0.000	-.004268	-.0028663
lnPRCP	.0311364	.0004718	66.00	0.000	.0302118	.032061
lnDPOP	-4.141222	.0194881	-212.50	0.000	-4.179418	-4.103026
LOCK	-.0214703	.0014657	-14.65	0.000	-.0243431	-.0185975
WAR	.2280039	.0011215	203.31	0.000	.2258059	.230202
_cons	66.31567	.2103445	315.27	0.000	65.9034	66.72794
sigma_u	2.1369483					
sigma_e	.59769582					
rho	.92744605	(fraction of variance due to u_i)				

10.3.10. Regression Output Following the Removal of Average Income

```
Fixed-effects (within) regression               Number of obs     =  6,337,451
Group variable: vehicle_id                      Number of groups  =  1,432,400

R-squared:                                      Obs per group:
     Within  = 0.1063                                         min =          1
     Between = 0.0003                                         avg =        4.4
     Overall = 0.0007                                         max =          6

                                                F(8,1432399)      =   51304.16
corr(u_i, Xb) = -0.9590                         Prob > F          =     0.0000

                       (Std. err. adjusted for 1,432,400 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnPP	.0986836	.0045703	21.59	0.000	.0897261	.1076411
lnGDP	2.071383	.007893	262.43	0.000	2.055912	2.086853
lnAGE	-.3450414	.0024059	-143.41	0.000	-.3497569	-.3403258
lnHDM	-.0028544	.0003587	-7.96	0.000	-.0035573	-.0021514
lnPRCP	.0047907	.0004796	9.99	0.000	.0038507	.0057306
lnDPOP	-5.51468	.0178193	-309.48	0.000	-5.549605	-5.479754
LOCK	.1578326	.0016327	96.67	0.000	.1546325	.1610326
WAR	.198913	.0010414	191.01	0.000	.196872	.2009541
_cons	69.00125	.2083471	331.18	0.000	68.59289	69.4096
sigma_u	2.7660165					
sigma_e	.59576546					
rho	.95566499	(fraction of variance due to u_i)				

10.3.11. Regression Output Following the Removal of Vehicle's Age

```
Fixed-effects (within) regression               Number of obs     =  6,337,451
Group variable: vehicle_id                      Number of groups  =  1,432,400

R-squared:                                      Obs per group:
     Within  = 0.1125                                         min =          1
     Between = 0.0000                                         avg =        4.4
     Overall = 0.0002                                         max =          6

                                                F(8,1432399)      =   53452.05
corr(u_i, Xb) = -0.9263                         Prob > F          =     0.0000

                       (Std. err. adjusted for 1,432,400 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnPP	-.114442	.004919	-23.27	0.000	-.1240831	-.1048008
lnGDP	2.286256	.0078024	293.02	0.000	2.270964	2.301549
lnY	-2.015193	.0089372	-225.48	0.000	-2.03271	-1.997676
lnHDM	-.0048732	.0003552	-13.72	0.000	-.0055693	-.0041771
lnPRCP	.0145232	.000475	30.57	0.000	.0135922	.0154542
lnDPOP	-4.006366	.019153	-209.18	0.000	-4.043905	-3.968826
LOCK	.198367	.001641	120.88	0.000	.1951507	.2015834
WAR	.2815665	.0011658	241.51	0.000	.2792815	.2838515
_cons	59.60302	.1986195	300.09	0.000	59.21373	59.99231
sigma_u	2.0824923					
sigma_e	.59366842					
rho	.92483986	(fraction of variance due to u_i)				

10.4. *Diesel Fixed Effects*

10.4.1. Pesaran's CD Test Result

```
Pesaran's test of cross sectional independence =    27.511, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.414
```

10.4.2. Modified Wald Test Result

```
Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (907121)  = 6.7e+12
Prob>chi2 =      0.0000
```

10.4.3. Wooldridge Test Result

```
Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,   742989) =  15348.955
           Prob > F =       0.0000
```

10.4.4. Ramsey's RESET Test Result

```
 ( 1)  fitted_values_2 = 0
 ( 2)  fitted_values_3 = 0

       F(  2,3179895) =  271.09
            Prob > F =    0.0000
```

10.4.5. Hausman Test Result

	(b) Dfixed	(B) Drandom	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
lnDP	-.1178501	-.2377436	.1198935	.003795
lnRGDP	1.579581	.9737143	.6058667	.0051824
lnRY	-7.119682	-6.599851	-.5198316	.0150241
lnAGE	.177191	-.3992466	.5764376	.0026296
lnHDM	-.0115198	-.0072826	-.0042373	.0002091
lnPRCP	.0019399	.008223	-.0062831	.0002195
lnDPOP	-2.529526	.0035556	-2.533082	.0189237
LOCK	-.040756	-.0279279	-.0128281	.0005906
WAR	.1571334	.1636612	-.0065279	.0004518

```
                   b = Consistent under H0 and Ha; obtained from xtreg.
    B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic

    chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            = 74777.43
Prob > chi2 =   0.0000
```

10.4.6. VIF Test Results

Variable	VIF	1/VIF
lnDP	5.34	0.187363
WAR	4.23	0.236210
lnY	2.64	0.379334
lnGDP	2.51	0.398904
LOCK	1.37	0.728167
lnPRCP	1.16	0.865064
lnHDM	1.10	0.912004
lnDPOP	1.05	0.953090
rural	1.02	0.983939
lnAGE	1.01	0.992222
Mean VIF	2.14	

10.4.7. One-Way Analysis of Variance

One-way analysis of variance for lnVMT:

Number of obs = 4,087,029
R-squared = 0.6166

Source	SS	df	MS	F	Prob > F
Between vehicle_id	1910068.4	907120	2.1056402	5.64	0.0000
Within vehicle_id	1187716.1	3179908	.37350644		
Total	3097784.5	4087028	.75795528		

Intraclass correlation	Asy. S.E.	[95% conf.	interval]
0.50722	0.00053	0.50618	0.50825

Estimated SD of vehicle_id effect .6200397
Estimated SD within vehicle_id .6111517
Est. reliability of a vehicle_id mean 0.82262
(evaluated at n=4.51)

10.4.8. Regression Output

Fixed-effects (within) regression Number of obs = 4,087,027
Group variable: **vehicle_id** Number of groups = 907,121

R-squared: Obs per group:
Within = 0.1267 min = 1
Between = 0.0001 avg = 4.5
Overall = 0.0002 max = 6

F(9,907120) = 36787.19
corr(u_i, Xb) = -0.9540 Prob > F = 0.0000

(Std. err. adjusted for **907,121** clusters in **vehicle_id**)

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnDP	.0739492	.005449	13.57	0.000	.0632694	.084629
lnGDP	2.142432	.0093711	228.62	0.000	2.124065	2.160799
lnY	-1.907316	.0153431	-124.31	0.000	-1.937388	-1.877244
lnAGE	.0319941	.0039527	8.09	0.000	.024247	.0397412
lnHDM	-.0070893	.0004148	-17.09	0.000	-.0079023	-.0062762
lnPRCP	.0136381	.0005785	23.57	0.000	.0125042	.0147721
lnDPOP	-4.488908	.0236186	-190.06	0.000	-4.535199	-4.442616
LOCK	.1435349	.0019628	73.13	0.000	.1396879	.147382
WAR	.2359007	.0015642	150.81	0.000	.2328349	.2389665
_cons	65.26116	.2500386	261.00	0.000	64.77109	65.75123
sigma_u	2.3942108					
sigma_e	.57113498					
rho	.94615856	(fraction of variance due to u_i)				

10.5. *Petrol 2SLS*

10.5.1. First-Stage Regression Output Part 1

First-stage regressions

First-stage regression of lnPP:

Statistics consistent for homoskedasticity only
Number of obs = 6338585

lnPP	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnBSP	.2880053	.00012	2400.81	0.000	.2877702	.2882404
lnGDP	.2610587	.0006119	426.65	0.000	.2598594	.262258
lnY	.376676	.0003468	1086.19	0.000	.3759963	.3773556
lnAGE	-.0002651	.0000291	-9.10	0.000	-.0003222	-.000208
lnHDM	-.0044928	9.49e-06	-473.41	0.000	-.0045114	-.0044742
lnPRCP	-.0008299	.0000238	-34.83	0.000	-.0008766	-.0007832
lnDPOP	-.0008626	.0000355	-24.30	0.000	-.0009322	-.0007931
LOCK	.1863624	.0001281	1455.25	0.000	.1861114	.1866133
WAR	.057981	.0000664	873.41	0.000	.0578509	.0581111
_cons	.5102246	.0029032	175.75	0.000	.5045344	.5159148

F test of excluded instruments:
F(1,6338575) = 5.8e+06
Prob > F = 0.0000
Sanderson-Windmeijer multivariate F test of excluded instruments:
F(1,6338575) = 5.8e+06
Prob > F = 0.0000

Summary results for first-stage regressions

Variable	F(1,6338575)	P-val	(Underid) SW Chi-sq(1)	P-val	(Weak id) SW F(1,6338575)
lnPP	5.8e+06	0.0000	5.8e+06	0.0000	5.8e+06

Stock-Yogo weak ID F test critical values for single endogenous regressor:

10% maximal IV size	16.38
15% maximal IV size	8.96
20% maximal IV size	6.66
25% maximal IV size	5.53

Source: Stock-Yogo (2005). Reproduced by permission.
NB: Critical values are for Sanderson-Windmeijer F statistic.

10.5.2. First-Stage Regression Output Part 2

```
Underidentification test
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Anderson canon. corr. LM statistic       Chi-sq(1)=3.0e+06  P-val=0.0000

Weak identification test
Ho: equation is weakly identified
Cragg-Donald Wald F statistic                                      5.8e+06

Stock-Yogo weak ID test critical values for K1=1 and L1=1:
                                   10% maximal IV size             16.38
                                   15% maximal IV size              8.96
                                   20% maximal IV size              6.66
                                   25% maximal IV size              5.53
Source: Stock-Yogo (2005).  Reproduced by permission.

Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and orthogonality conditions are valid
Anderson-Rubin Wald test           F(1,6338575)=3328.76      P-val=0.0000
Anderson-Rubin Wald test           Chi-sq(1)=    3328.76      P-val=0.0000
Stock-Wright LM S statistic        Chi-sq(1)=    3327.01      P-val=0.0000

Number of observations               N  =     6338585
Number of regressors                 K  =          10
Number of endogenous regressors      K1 =           1
Number of instruments                L  =          10
Number of excluded instruments       L1 =           1
```

10.5.3. First-Stage Regression Output Part 3

```
IV (2SLS) estimation

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =   6338585
                                                      F(  9,6338575) = 67479.85
                                                      Prob > F      =    0.0000
Total (centered) SS     =  5583918.173                Centered R2   =    0.0876
Total (uncentered) SS   =  435452175.4                Uncentered R2 =    0.9883
Residual SS             =  5094662.616                Root MSE      =     .8965
```

lnVMT	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
lnPP	-.5081735	.0088069	-57.70	0.000	-.5254348	-.4909123
lnGDP	2.274778	.0145541	156.30	0.000	2.246252	2.303303
lnY	-2.445501	.0078538	-311.38	0.000	-2.460894	-2.430108
lnAGE	-.4084937	.0006158	-663.34	0.000	-.4097007	-.4072867
lnHDM	.0017359	.0002037	8.52	0.000	.0013365	.0021352
lnPRCP	.0274741	.0005055	54.35	0.000	.0264834	.0284648
lnDPOP	.0148049	.0007509	19.72	0.000	.0133332	.0162765
LOCK	.2014868	.00259	77.79	0.000	.1964105	.2065632
WAR	.2365702	.0016873	140.21	0.000	.2332632	.2398772
_cons	13.5869	.0596341	227.84	0.000	13.47002	13.70378

```
Underidentification test (Anderson canon. corr. LM statistic):          3.0e+06
                                                     Chi-sq(1) P-val =   0.0000

Weak identification test (Cragg-Donald Wald F statistic):               5.8e+06
Stock-Yogo weak ID test critical values: 10% maximal IV size              16.38
                                         15% maximal IV size               8.96
                                         20% maximal IV size               6.66
                                         25% maximal IV size               5.53
Source: Stock-Yogo (2005).  Reproduced by permission.

Sargan statistic (overidentification test of all instruments):            0.000
                                                (equation exactly identified)

Instrumented:         lnPP
Included instruments: lnGDP lnY lnAGE lnHDM lnPRCP lnDPOP LOCK WAR
Excluded instruments: lnBSP
```

10.5.4. 2SLS Regression Output

```
Fixed-effects (within) IV regression            Number of obs     =  6,338,585
Group variable: vehicle_id                      Number of groups  =  1,432,400

R-squared:                                      Obs per group:
     Within  = 0.1085                                         min =          1
     Between = 0.0000                                         avg =        4.4
     Overall = 0.0003                                         max =          6


                                                Wald chi2(9)      =  515944.08
corr(u_i, Xb) = -0.9498                         Prob > chi2       =     0.0000

                      (Std. err. adjusted for 1,432,400 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
lnPP	.5468998	.0078577	69.60	0.000	.531499	.5623007
lnGDP	1.599845	.0100506	159.18	0.000	1.580147	1.619544
lnY	-1.587136	.0130655	-121.48	0.000	-1.612744	-1.561528
lnAGE	-.0688586	.0033697	-20.43	0.000	-.0754632	-.062254
lnHDM	-.0045202	.0003577	-12.64	0.000	-.0052213	-.0038192
lnPRCP	.0279437	.0004898	57.05	0.000	.0269837	.0289037
lnDPOP	-4.955824	.0222039	-223.20	0.000	-4.999343	-4.912305
LOCK	.1343456	.0017474	76.89	0.000	.1309208	.1377703
WAR	.1913919	.0013445	142.35	0.000	.1887568	.1940271
_cons	69.51864	.2410797	288.36	0.000	69.04613	69.99115
sigma_u	2.5123863					
sigma_e	.5972857					
rho	.94650487	(fraction of variance due to u_i)				

```
Instrumented: lnPP
 Instruments: lnGDP lnY lnAGE lnHDM lnPRCP lnDPOP LOCK WAR lnBSP
```

10.6. Diesel 2SLS

10.6.1. First-Stage Regression Output Part 1

```
First-stage regressions


First-stage regression of lnDP:

Statistics consistent for homoskedasticity only
Number of obs =                 4087027
```

lnDP	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnBSP	.2429211	.0001882	1290.65	0.000	.2425522	.24329
lnGDP	.3415714	.0009649	353.99	0.000	.3396802	.3434626
lnY	.3180631	.0005375	591.73	0.000	.3170096	.3191167
lnAGE	.0013309	.0000488	27.30	0.000	.0012354	.0014265
lnHDM	.003033	.0000152	199.89	0.000	.0030033	.0030628
lnPRCP	-.0034475	.000038	-90.83	0.000	-.0035219	-.0033731
lnDPOP	.0001789	.0000531	3.37	0.001	.0000748	.000283
LOCK	.155744	.0002005	776.70	0.000	.155351	.1561371
WAR	.1160123	.0001027	1130.01	0.000	.1158111	.1162135
_cons	.6524884	.0045495	143.42	0.000	.6435717	.6614052

```
F test of excluded instruments:
  F(  1,4087017) =  1.7e+06
  Prob > F      =   0.0000
Sanderson-Windmeijer multivariate F test of excluded instruments:
  F(  1,4087017) =  1.7e+06
  Prob > F      =   0.0000


Summary results for first-stage regressions


                                           (Underid)            (Weak id)
Variable     | F(  1,4087017)  P-val | SW Chi-sq(  1) P-val | SW F(  1,4087017)
lnDP         |      1.7e+06   0.0000 |       1.7e+06   0.0000 |        1.7e+06

Stock-Yogo weak ID F test critical values for single endogenous regressor:
                                   10% maximal IV size             16.38
                                   15% maximal IV size              8.96
                                   20% maximal IV size              6.66
                                   25% maximal IV size              5.53
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Sanderson-Windmeijer F statistic.
```

10.6.2. First-Stage Regression Output Part 2

```
Underidentification test
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Anderson canon. corr. LM statistic       Chi-sq(1)=1.2e+06  P-val=0.0000

Weak identification test
Ho: equation is weakly identified
Cragg-Donald Wald F statistic                                      1.7e+06

Stock-Yogo weak ID test critical values for K1=1 and L1=1:
                                   10% maximal IV size             16.38
                                   15% maximal IV size              8.96
                                   20% maximal IV size              6.66
                                   25% maximal IV size              5.53
Source: Stock-Yogo (2005).  Reproduced by permission.

Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and orthogonality conditions are valid
Anderson-Rubin Wald test           F(1,4087017)= 628.85     P-val=0.0000
Anderson-Rubin Wald test           Chi-sq(1)=    628.85     P-val=0.0000
Stock-Wright LM S statistic        Chi-sq(1)=    628.75     P-val=0.0000

Number of observations               N  =    4087027
Number of regressors                 K  =         10
Number of endogenous regressors      K1 =          1
Number of instruments                L  =         10
Number of excluded instruments       L1 =          1
```

10.6.3. First-Stage Regression Output Part 3

```
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                  Number of obs =  4087027
                                                  F(  9,4087017) = 61380.71
                                                  Prob > F      =   0.0000
Total (centered) SS     =  3097778.116            Centered R2   =   0.1191
Total (uncentered) SS   =  314598261.1            Uncentered R2 =   0.9913
Residual SS             =  2728839.286            Root MSE      =    .8171
```

lnVMT	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
lnDP	-.2975949	.0118671	-25.08	0.000	-.320854	-.2743358
lnGDP	1.862395	.017738	104.99	0.000	1.827629	1.897161
lnY	-2.249766	.0088896	-253.08	0.000	-2.267189	-2.232342
lnAGE	-.4853635	.0007465	-650.20	0.000	-.4868266	-.4839004
lnHDM	.0013453	.0002358	5.71	0.000	.0008831	.0018074
lnPRCP	.0167422	.0005921	28.27	0.000	.0155816	.0179028
lnDPOP	.0116759	.0008137	14.35	0.000	.0100809	.0132708
LOCK	.1551883	.0029519	52.57	0.000	.1494026	.1609739
WAR	.2128529	.0024893	85.51	0.000	.2079739	.2177319
_cons	14.13462	.0666351	212.12	0.000	14.00402	14.26522

```
Underidentification test (Anderson canon. corr. LM statistic):          1.2e+06
                                                Chi-sq(1) P-val =        0.0000

Weak identification test (Cragg-Donald Wald F statistic):              1.7e+06
Stock-Yogo weak ID test critical values: 10% maximal IV size              16.38
                                         15% maximal IV size               8.96
                                         20% maximal IV size               6.66
                                         25% maximal IV size               5.53
Source: Stock-Yogo (2005).  Reproduced by permission.

Sargan statistic (overidentification test of all instruments):           0.000
                                                  (equation exactly identified)

Instrumented:         lnDP
Included instruments: lnGDP lnY lnAGE lnHDM lnPRCP lnDPOP LOCK WAR
Excluded instruments: lnBSP
```

10.6.4. 2SLS Regression Output

```
Fixed-effects (within) IV regression            Number of obs     =  4,087,027
Group variable: vehicle_id                      Number of groups  =    907,121

R-squared:                                      Obs per group:
     Within  = 0.1125                                         min =          1
     Between = 0.0000                                         avg =        4.5
     Overall = 0.0001                                         max =          6

                                                Wald chi2(9)      =  400232.30
corr(u_i, Xb) = -0.9771                         Prob > chi2       =     0.0000

                       (Std. err. adjusted for 907,121 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
lnDP	1.332974	.0125352	106.34	0.000	1.308406	1.357543
lnGDP	.9881048	.0139384	70.89	0.000	.960786	1.015423
lnY	-1.16087	.0167814	-69.18	0.000	-1.193761	-1.127979
lnAGE	-.0040652	.0039819	-1.02	0.307	-.0118695	.0037392
lnHDM	-.0152329	.000421	-36.18	0.000	-.0160581	-.0144077
lnPRCP	.0396509	.0006246	63.48	0.000	.0384267	.0408751
lnDPOP	-6.50483	.0311002	-209.16	0.000	-6.565785	-6.443875
LOCK	.032413	.002212	14.65	0.000	.0280775	.0367485
WAR	.0011952	.0025093	0.48	0.634	-.003723	.0061134
_cons	86.15328	.3286615	262.13	0.000	85.50912	86.79745
sigma_u	3.3867649					
sigma_e	.5757491					
rho	.97191184	(fraction of variance due to u_i)				

```
Instrumented: lnDP
 Instruments: lnGDP lnY lnAGE lnHDM lnPRCP lnDPOP LOCK WAR lnBSP
```

10.7. Petrol Subgroup Analysis

10.7.1. City Vehicle Output

```
Fixed-effects (within) regression               Number of obs     =  2,137,397
Group variable: vehicle_id                      Number of groups  =    469,208

R-squared:                                      Obs per group:
     Within  = 0.1060                                         min =          1
     Between = 0.0002                                         avg =        4.6
     Overall = 0.0008                                         max =          6

                                                F(9,469207)       =   15689.12
corr(u_i, Xb) = -0.9233                         Prob > F          =     0.0000

                       (Std. err. adjusted for 469,208 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnPP	-.0957342	.0082546	-11.60	0.000	-.111913	-.0795553
lnGDP	2.095807	.0136311	153.75	0.000	2.069091	2.122524
lnY	-1.104523	.0219294	-50.37	0.000	-1.147504	-1.061542
lnAGE	-.2357476	.0055053	-42.82	0.000	-.2465378	-.2249575
lnHDM	-.0041398	.0005922	-6.99	0.000	-.0053004	-.0029791
lnPRCP	.0171316	.0008031	21.33	0.000	.0155575	.0187056
lnDPOP	-3.795159	.0325856	-116.47	0.000	-3.859025	-3.731292
LOCK	.2055117	.0027551	74.59	0.000	.2001118	.2109116
WAR	.2618089	.0019326	135.47	0.000	.258021	.2655968
_cons	53.4076	.3506804	152.30	0.000	52.72028	54.09493
sigma_u	1.9741282					
sigma_e	.57822294					
rho	.92098786	(fraction of variance due to u_i)				

10.7.2. Medium Vehicle Output

```
Fixed-effects (within) regression               Number of obs     =  3,814,165
Group variable: vehicle_id                      Number of groups  =    875,081

R-squared:                                      Obs per group:
     Within  = 0.1142                                         min =          1
     Between = 0.0000                                         avg =        4.4
     Overall = 0.0003                                         max =          6

                                                F(9,875080)       =   29983.90
corr(u_i, Xb) = -0.9276                         Prob > F          =     0.0000

                        (Std. err. adjusted for 875,081 clusters in vehicle_id)
------------------------------------------------------------------------------
             |               Robust
       lnVMT | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnPP |  -.1137382    .006393   -17.79   0.000    -.1262683   -.1012081
       lnGDP |   2.287313   .0104365   219.16   0.000     2.266857    2.307768
         lnY |  -2.034045   .0162716  -125.01   0.000    -2.065937   -2.002153
       lnAGE |  -.0098242   .0042795    -2.30   0.022    -.0182119   -.0014365
       lnHDM |  -.0053602   .0004602   -11.65   0.000    -.0062623   -.0044582
      lnPRCP |    .014187   .0006083    23.32   0.000     .0129948    .0153792
      lnDPOP |  -4.021968   .0248339  -161.95   0.000    -4.070642   -3.973294
        LOCK |   .1990478   .0021281    93.53   0.000     .1948767    .2032189
         WAR |   .2834341   .0015134   187.28   0.000     .2804678    .2864003
       _cons |   59.95515   .2663041   225.14   0.000      59.4332    60.47709
-------------+----------------------------------------------------------------
     sigma_u |  2.0813249
     sigma_e |  .59402095
         rho |  .92467921   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

10.7.3. SUV/MPV Output

```
Fixed-effects (within) regression               Number of obs     =    217,762
Group variable: vehicle_id                      Number of groups  =     52,547

R-squared:                                      Obs per group:
     Within  = 0.1330                                         min =          1
     Between = 0.0001                                         avg =        4.1
     Overall = 0.0003                                         max =          6

                                                F(9,52546)        =    2003.75
corr(u_i, Xb) = -0.9114                         Prob > F          =     0.0000

                         (Std. err. adjusted for 52,547 clusters in vehicle_id)
------------------------------------------------------------------------------
             |               Robust
       lnVMT | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnPP |   .0899827    .035103     2.56   0.010     .0211804    .1587849
       lnGDP |   2.275303   .0573678    39.66   0.000     2.162861    2.387744
         lnY |  -3.758323   .0834794   -45.02   0.000    -3.921944   -3.594703
       lnAGE |    .152011   .0269003     5.65   0.000     .0992862    .2047358
       lnHDM |  -.0085306   .0022885    -3.73   0.000    -.0130161   -.0040451
      lnPRCP |   .0054581   .0033174     1.65   0.100    -.0010441    .0119602
      lnDPOP |  -4.687842   .1278517   -36.67   0.000    -4.938432   -4.437251
        LOCK |   .1161483   .0111262    10.44   0.000     .0943408    .1379558
         WAR |   .3342789   .0083135    40.21   0.000     .3179844    .3505735
       _cons |   75.61534   1.363829    55.44   0.000     72.94223    78.28846
-------------+----------------------------------------------------------------
     sigma_u |  2.5147764
     sigma_e |  .76139983
         rho |  .91602782   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

10.7.4. Small SUV Output

```
Fixed-effects (within) regression               Number of obs     =    168,697
Group variable: vehicle_id                      Number of groups  =     39,117

R-squared:                                      Obs per group:
     Within  = 0.1300                                         min =          1
     Between = 0.0003                                         avg =        4.3
     Overall = 0.0001                                         max =          6

                                                F(9,39116)        =    1524.99
corr(u_i, Xb) = -0.9200                         Prob > F          =     0.0000

                         (Std. err. adjusted for 39,117 clusters in vehicle_id)
------------------------------------------------------------------------------
             |               Robust
       lnVMT | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnPP |  -.0551917   .0311314    -1.77   0.076      -.11621    .0058266
       lnGDP |   2.333435   .0510644    45.70   0.000     2.233348    2.433522
         lnY |  -2.841217   .0777096   -36.56   0.000     -2.99353   -2.688905
       lnAGE |   .0863193   .0278658     3.10   0.002     .0317016     .140937
       lnHDM |  -.0032845   .0023206    -1.42   0.157     -.007833     .001264
      lnPRCP |   .0092387    .003014     3.07   0.002     .0033313    .0151462
      lnDPOP |  -3.892011   .1134472   -34.31   0.000     -4.11437   -3.669651
        LOCK |   .1796814   .0101978    17.62   0.000     .1596935    .1996694
         WAR |   .3121178   .0075191    41.51   0.000     .2973802    .3268555
       _cons |   61.69556    1.22777    50.25   0.000      59.2891    64.10202
-------------+----------------------------------------------------------------
     sigma_u |   2.077757
     sigma_e |  .61079217
         rho |  .92045718   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

10.7.5. Small Sedan Output

```
Fixed-effects (within) regression               Number of obs     =        564
Group variable: vehicle_id                      Number of groups  =        129

R-squared:                                      Obs per group:
     Within  = 0.2065                                         min =          2
     Between = 0.0003                                         avg =        4.4
     Overall = 0.0003                                         max =          6

                                                F(9,128)          =      10.57
corr(u_i, Xb) = -0.9801                         Prob > F          =     0.0000

                            (Std. err. adjusted for 129 clusters in vehicle_id)
------------------------------------------------------------------------------
             |               Robust
       lnVMT | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnPP |   .4445334   .5633511     0.79   0.432     -.670153     1.55922
       lnGDP |   2.498573   .9325138     2.68   0.008     .6534349    4.343711
         lnY |  -3.052804   1.239695    -2.46   0.015    -5.505753   -.5998559
       lnAGE |   .6126429   .2428165     2.52   0.013      .132189    1.093097
       lnHDM |  -.0469404   .0327709    -1.43   0.154    -.1117832    .0179023
      lnPRCP |    .060022   .0609343     0.99   0.326    -.0605469    .1805908
      lnDPOP |   -8.69965   2.349174    -3.70   0.000    -13.34789   -4.051409
        LOCK |   .2287018   .3421353     0.67   0.505    -.4482713    .9056749
         WAR |   .2472777   .1344656     1.84   0.068    -.0187855    .5133409
       _cons |    118.811   22.69871     5.23   0.000     73.89771    163.7243
-------------+----------------------------------------------------------------
     sigma_u |  5.5034457
     sigma_e |  .69679478
         rho |  .98422266   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

10.7.6. Rural Classification Output

```
Fixed-effects (within) regression               Number of obs     =    142,677
Group variable: vehicle_id                      Number of groups  =     32,769

R-squared:                                      Obs per group:
     Within  = 0.1225                                         min =          2
     Between = 0.0000                                         avg =        4.4
     Overall = 0.0002                                         max =          6

                                                F(9,32768)        =    1162.76
corr(u_i, Xb) = -0.9641                         Prob > F          =     0.0000

                         (Std. err. adjusted for 32,769 clusters in vehicle_id)
------------------------------------------------------------------------------
             |               Robust
       lnVMT | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnPP |   -.213594   .0305933    -6.98   0.000    -.2735579     -.15363
       lnGDP |   2.702277   .0535416    50.47   0.000     2.597333     2.80722
         lnY |  -2.424271   .0780044   -31.08   0.000    -2.577163    -2.27138
       lnAGE |   .0479257   .0208487     2.30   0.022     .0070615    .0887898
       lnHDM |  -.0002511    .002626    -0.10   0.924    -.0053982     .004896
      lnPRCP |   .0147431   .0026386     5.59   0.000     .0095715    .0199148
      lnDPOP |  -4.956207   .1364356   -36.33   0.000    -5.223626   -4.688788
        LOCK |   .2234532   .0112159    19.92   0.000     .2014696    .2454368
         WAR |    .303817   .0073535    41.32   0.000     .2894038    .3182302
       _cons |   71.02132   1.479628    48.00   0.000     68.12119    73.92144
-------------+----------------------------------------------------------------
     sigma_u |  3.1101161
     sigma_e |  .58998928
         rho |   .9652639   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

10.7.7. Urban Classification Output

```
Fixed-effects (within) regression               Number of obs     =  6,195,908
Group variable: vehicle_id                      Number of groups  =  1,399,631

R-squared:                                      Obs per group:
     Within  = 0.1115                                         min =          1
     Between = 0.0000                                         avg =        4.4
     Overall = 0.0004                                         max =          6

                                                F(9,1399630)      =   47666.52
corr(u_i, Xb) = -0.9241                         Prob > F          =     0.0000

                      (Std. err. adjusted for 1,399,631 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnPP	-.103563	.0050278	-20.60	0.000	-.1134173	-.0937088
lnGDP	2.245498	.0082308	272.82	0.000	2.229366	2.26163
lnY	-1.86173	.0129902	-143.32	0.000	-1.88719	-1.83627
lnAGE	-.0547762	.0034119	-16.05	0.000	-.0614634	-.0480891
lnHDM	-.0050296	.0003599	-13.98	0.000	-.0057349	-.0043243
lnPRCP	.0145395	.0004849	29.98	0.000	.0135891	.01549
lnDPOP	-3.984646	.0195787	-203.52	0.000	-4.023019	-3.946272
LOCK	.1973449	.0016679	118.32	0.000	.1940758	.2006141
WAR	.2794355	.001186	235.62	0.000	.277111	.28176
_cons	58.80387	.2104027	279.48	0.000	58.39148	59.21625
sigma_u	2.0511184					
sigma_e	.59634088					
rho	.92205892	(fraction of variance due to u_i)				

10.8. Diesel Subgroup Analysis

10.8.1. City Vehicle Output

```
Fixed-effects (within) regression               Number of obs     =    350,864
Group variable: vehicle_id                      Number of groups  =     75,134

R-squared:                                      Obs per group:
     Within  = 0.1308                                         min =          1
     Between = 0.0001                                         avg =        4.7
     Overall = 0.0007                                         max =          6

                                                F(9,75133)        =    3375.76
corr(u_i, Xb) = -0.9611                         Prob > F          =     0.0000

                         (Std. err. adjusted for 75,134 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnDP	.0424217	.0181032	2.34	0.019	.0069396	.0779039
lnGDP	2.072221	.0303588	68.26	0.000	2.012718	2.131724
lnY	-1.434943	.0479895	-29.90	0.000	-1.529002	-1.340883
lnAGE	-.1078453	.0124089	-8.69	0.000	-.1321668	-.0835238
lnHDM	-.0067004	.0013571	-4.94	0.000	-.0093603	-.0040404
lnPRCP	.0151002	.0019071	7.92	0.000	.0113623	.018838
lnDPOP	-4.397135	.0790042	-55.66	0.000	-4.551983	-4.242287
LOCK	.1643779	.0065005	25.29	0.000	.1516369	.1771189
WAR	.2691489	.0052523	51.24	0.000	.2588543	.2794435
_cons	62.35699	.8303514	75.10	0.000	60.7295	63.98447
sigma_u	2.3316407					
sigma_e	.54421094					
rho	.94833764	(fraction of variance due to u_i)				

10.8.2. Medium Vehicle Output

```
Fixed-effects (within) regression               Number of obs     =  2,897,212
Group variable: vehicle_id                      Number of groups  =    635,564

R-squared:                                      Obs per group:
     Within  = 0.1300                                         min =          1
     Between = 0.0000                                         avg =        4.6
     Overall = 0.0002                                         max =          6

                                                F(9,635563)       =   26649.32
corr(u_i, Xb) = -0.9570                         Prob > F          =     0.0000

                       (Std. err. adjusted for 635,564 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnDP	.0680965	.00628	10.84	0.000	.0557879	.0804051
lnGDP	2.124126	.0107994	196.69	0.000	2.10296	2.145293
lnY	-1.764897	.0182595	-96.66	0.000	-1.800685	-1.729109
lnAGE	.006043	.0048459	1.25	0.212	-.0034549	.015541
lnHDM	-.0065495	.0004746	-13.80	0.000	-.0074797	-.0056192
lnPRCP	.0142587	.0006618	21.55	0.000	.0129616	.0155559
lnDPOP	-4.475612	.0269917	-165.81	0.000	-4.528515	-4.422709
LOCK	.1524517	.0022646	67.32	0.000	.1480132	.1568903
WAR	.2393442	.0018072	132.44	0.000	.2358021	.2428863
_cons	64.52416	.2870846	224.76	0.000	63.96148	65.08683
sigma_u	2.3733719					
sigma_e	.55350499					
rho	.94841652	(fraction of variance due to u_i)				

10.8.3. SUV/MPV Output

```
Fixed-effects (within) regression               Number of obs     =    276,836
Group variable: vehicle_id                      Number of groups  =     64,925

R-squared:                                      Obs per group:
     Within  = 0.1278                                         min =          1
     Between = 0.0006                                         avg =        4.3
     Overall = 0.0000                                         max =          6

                                                F(9,64924)        =    2536.98
corr(u_i, Xb) = -0.9520                         Prob > F          =     0.0000

                        (Std. err. adjusted for 64,925 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnDP	.134443	.0232372	5.79	0.000	.0888981	.1799879
lnGDP	2.327713	.0399119	58.32	0.000	2.249486	2.40594
lnY	-2.537498	.0587453	-43.19	0.000	-2.652639	-2.422357
lnAGE	.1815294	.0141387	12.84	0.000	.1538176	.2092413
lnHDM	-.0071867	.0017578	-4.09	0.000	-.010632	-.0037415
lnPRCP	.0151785	.0025022	6.07	0.000	.0102741	.0200829
lnDPOP	-5.006711	.0991608	-50.49	0.000	-5.201067	-4.812356
LOCK	.1179373	.008313	14.19	0.000	.1016439	.1342308
WAR	.2190188	.006584	33.27	0.000	.2061141	.2319235
_cons	73.45531	1.038894	70.71	0.000	71.41908	75.49154
sigma_u	2.6959485					
sigma_e	.63145268					
rho	.9479928	(fraction of variance due to u_i)				

10.8.4. Small SUV Output

```
Fixed-effects (within) regression               Number of obs     =    561,867
Group variable: vehicle_id                      Number of groups  =    133,016

R-squared:                                      Obs per group:
     Within  = 0.1156                                         min =          1
     Between = 0.0011                                         avg =        4.2
     Overall = 0.0000                                         max =          6

                                                F(9,133015)       =    4581.95
corr(u_i, Xb) = -0.9405                         Prob > F          =     0.0000

                       (Std. err. adjusted for 133,016 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnDP	.1146469	.0161465	7.10	0.000	.0830001	.1462936
lnGDP	2.161127	.0286778	75.36	0.000	2.104919	2.217335
lnY	-2.518564	.0432406	-58.25	0.000	-2.603315	-2.433813
lnAGE	.1451502	.0099168	14.64	0.000	.1257135	.1645869
lnHDM	-.0100765	.0012734	-7.91	0.000	-.0125724	-.0075807
lnPRCP	.0092982	.0017538	5.30	0.000	.0058608	.0127357
lnDPOP	-4.473411	.0733772	-60.96	0.000	-4.617229	-4.329593
LOCK	.0929143	.0059376	15.65	0.000	.0812767	.1045518
WAR	.2010828	.0046001	43.71	0.000	.1920667	.210099
_cons	67.82849	.7719826	87.86	0.000	66.31542	69.34156
sigma_u	2.4501799					
sigma_e	.64117392					
rho	.93591006	(fraction of variance due to u_i)				

10.8.5. Small Sedan Output

```
Fixed-effects (within) regression               Number of obs     =        248
Group variable: vehicle_id                      Number of groups  =         51

R-squared:                                      Obs per group:
     Within  = 0.1840                                         min =          2
     Between = 0.0037                                         avg =        4.9
     Overall = 0.0208                                         max =          6

                                                F(9,50)           =      11.99
corr(u_i, Xb) = -0.8927                         Prob > F          =     0.0000

                               (Std. err. adjusted for 51 clusters in vehicle_id)
------------------------------------------------------------------------------
             |               Robust
       lnVMT | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnDP |   .7913358   .5435923     1.46   0.152    -.3005014    1.883173
       lnGDP |   1.833578   1.290946     1.42   0.162    -.7593631    4.426519
         lnY |   1.029488   1.745578     0.59   0.558    -2.476609    4.535584
       lnAGE |  -.9579289   .5010714    -1.91   0.062     -1.96436    .0485026
       lnHDM |  -.0200242   .0325108    -0.62   0.541     -.085324    .0452757
      lnPRCP |  -.0003891   .0575412    -0.01   0.995     -.115964    .1151859
      lnDPOP |  -1.982603   2.366314    -0.84   0.406    -6.735484    2.770277
        LOCK |   .5141936   .1485333     3.46   0.001     .2158557    .8125316
         WAR |  -.0395092   .1277586    -0.31   0.758    -.2961198    .2171015
       _cons |    17.9567   27.76939     0.65   0.521    -37.81977    73.73317
-------------+----------------------------------------------------------------
     sigma_u |  1.5278246
     sigma_e |  .48844021
         rho |  .90727153   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

10.8.6. Rural Classification Output

```
Fixed-effects (within) regression               Number of obs     =     89,226
Group variable: vehicle_id                      Number of groups  =     20,347

R-squared:                                      Obs per group:
     Within  = 0.1256                                         min =          1
     Between = 0.0020                                         avg =        4.4
     Overall = 0.0004                                         max =          6

                                                F(9,20346)        =     769.20
corr(u_i, Xb) = -0.9732                         Prob > F          =     0.0000

                           (Std. err. adjusted for 20,347 clusters in vehicle_id)
------------------------------------------------------------------------------
             |               Robust
       lnVMT | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        lnDP |  -.0411181   .0337143    -1.22   0.223    -.1072009    .0249647
       lnGDP |   2.365902   .0617369    38.32   0.000     2.244893    2.486911
         lnY |  -2.592543   .0947304   -27.37   0.000    -2.778223   -2.406864
       lnAGE |    .109038   .0237571     4.59   0.000     .0624723    .1556038
       lnHDM |    .002154   .0030633     0.70   0.482    -.0038502    .0081583
      lnPRCP |   .0217812   .0032914     6.62   0.000     .0153298    .0282327
      lnDPOP |  -4.195214   .1633578   -25.68   0.000    -4.515408   -3.875019
        LOCK |   .1499254   .0138388    10.83   0.000     .1228003    .1770505
         WAR |   .2503949   .0099515    25.16   0.000     .2308892    .2699006
       _cons |   62.97114   1.733393    36.33   0.000     59.57355    66.36873
-------------+----------------------------------------------------------------
     sigma_u |  3.1675953
     sigma_e |  .55343857
         rho |   .9703776   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

10.8.7. Urban Classification Output

```
Fixed-effects (within) regression               Number of obs     =  3,997,801
Group variable: vehicle_id                      Number of groups  =    886,774

R-squared:                                      Obs per group:
     Within  = 0.1268                                         min =          1
     Between = 0.0001                                         avg =        4.5
     Overall = 0.0002                                         max =          6

                                                F(9,886773)       =   36035.53
corr(u_i, Xb) = -0.9530                         Prob > F          =     0.0000

                      (Std. err. adjusted for 886,774 clusters in vehicle_id)
```

lnVMT	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lnDP	.0786752	.0055251	14.24	0.000	.0678462	.0895043
lnGDP	2.138374	.0094844	225.46	0.000	2.119785	2.156963
lnY	-1.885581	.015545	-121.30	0.000	-1.916048	-1.855113
lnAGE	.0310376	.004007	7.75	0.000	.0231839	.0388912
lnHDM	-.007212	.0004187	-17.22	0.000	-.0080326	-.0063913
lnPRCP	.0132931	.0005878	22.62	0.000	.012141	.0144452
lnDPOP	-4.516482	.0239502	-188.58	0.000	-4.563423	-4.46954
LOCK	.142907	.0019832	72.06	0.000	.1390201	.146794
WAR	.2353051	.0015841	148.54	0.000	.2322004	.2384098
_cons	65.53193	.2535405	258.47	0.000	65.035	66.02886
sigma_u	2.3680897					
sigma_e	.57149853					
rho	.94496372	(fraction of variance due to u_i)				