**Methodology**

**Sources of Data**

The dataset analysed in this paper has been assembled from a two public records. First, the Department for Transport’s Vehicle Licensing Database represents the source of the vehicle stock data, noting the number of cars registered to private households across Northern Ireland by fuel type. Second, the Northern Ireland Population Census represents the source of the socioeconomic, household, and transport system data. The variables incorporated into the dataset are described in Table Y.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table Y:** Descriptive statistics of the dataset used in the analysis | | | | |
|  | **Min.** | **Max.** | **Mean** | **Std. Dev.** |
| *Variable of Interest* |  |  |  |  |
| Diesel Cars (%) | 22.003 | 88.089 | 49.927 | 15.619 |
| *Socioeconomics* |  |  |  |  |
| Mean Age (years) | 22.700 | 50.700 | 37.836 | 3.901 |
| Self Employed (%) | 0.905 | 20.055 | 8.587 | 4.319 |
| Level 4 Qualification – University Degree (%) | 4.307 | 63.898 | 23.336 | 9.643 |
| *Travel* |  |  |  |  |
| One Car Household (%) | 20.593 | 64.194 | 41.344 | 6.380 |
| Car Driver to Work (%) | 18.080 | 74.300 | 56.131 | 10.766 |
| Over 30 km to Work (%) | 0.445 | 28.760 | 8.284 | 5.948 |
| *Household* |  |  |  |  |
| Population Density (per hectare) | 0.100 | 143.900 | 21.397 | 24.572 |
| Mean Household Residents | 1.787 | 3.466 | 2.559 | 0.325 |
| Rent Household Socially (%) | 0.000 | 67.705 | 14.652 | 14.324 |
| Flats (%) | 0.000 | 78.025 | 8.335 | 9.806 |

**Spatial Resolution**

The variables incorporated in the dataset have been aggregated at the Super Output Area (SOA) of administrative geography in Northern Ireland. This covers 890 contiguous spatial units which contain a mean of 2,000 individuals and 700 households.

**Measurement of Nearness**

Estimating how close a spatial unit is to the Republic of Ireland represents an issue of central importance in this research project. A set of different methods have been followed in this estimation in order to approach the issue from multiple directions. Each of the methods is employed in the analysis to consider if the association between nearness to the Republic of Ireland and the ownership of diesel cars persists across different estimation procedures.

Contiguity Method

A set of distance based buffers from the border with the Republic of Ireland are set with spatial units assigned to the buffer if it intersects. These buffers incorporate spatial units that intersect a 5 kilometre (n = 104), 10 kilometre (n = 45), 15 kilometre (n = 31), and 20 kilometre (n = 36) buffer to the border as well as those which comprise the remainder of Northern Ireland (n = 674). This arrangement is illustrated in Figure Y. The hypothesis here is that spatial units that have closer contiguity to the border with the Republic of Ireland will tend to have higher rates of diesel car ownership.

Proximity Method

The centroid of each spatial unit is extracted and the position of each road crossing between Northern Ireland and the Republic of Ireland is mapped. The Euclidean distance in kilometres between each centroid and the closest road crossing to the border is calculated. The hypothesis here is that as the Euclidean distance to the closest road border crossing increases, the rate of diesel car ownership will tend to decrease.

Network Distance Method

The centroid of each spatial unit is extracted and the location of the closest fuel station on the Republic of Ireland side of the border is mapped. The network distance in kilometres between each centroid and the closest fuel station in the Republic of Ireland is calculated. This arrangement is illustrated in Figure Y. The hypothesis here is that as network distance to the closest fuel station increases, the rate of diesel car ownership will tend to decrease.

Network Time Method

The centroid of each spatial unit is extracted and the location of the closest fuel station on the Republic of Ireland side of the border is mapped. The network time to travel by car in minutes between each centroid and the closest fuel station in the Republic of Ireland is estimated. The hypothesis here is that as network time to the closest fuel station increases, the rate of diesel car ownership will tend to decrease.

|  |  |
| --- | --- |
|  |  |

Figure Y: Maps illustrating the (a) Super Output Areas that intersect set buffers to the border with the Republic of Irleand and (b)

**Limitations**

While nearness to the border has been measured in this paper using multiple methods, this may still represent a crude proxy of the level of interaction between the spatial units of Northern Ireland and the Republic of Ireland. A superior method would be the level of trips per car that originate in each spatial unit and cross into the Republic of Ireland. Commuting patterns from the 2011 Northern Ireland census could offer such a proxy, but due to the small sample sizes observed for this characteristic and the level of spatial resolution employed by the analysis, the ……. declined to provide these statistics.

**Statistical Analysis**

The assessment of the dataset progresses through a series of stages.

Stage One

First, the spatial variation in diesel car ownership is considered. A boxplot of the proportion of the local authority car fleets that are fuelled by diesel are grouped across the Government Office Regions of the UK is formatted. A choropleth map using equal bin counts and depicting the rate of diesel car ownership across the SOAs of Northern Ireland is produced. The degree to which the rate of diesel car ownership in Northern Ireland displays spatial dependence (i.e. non-random spatial patterning) is evaluated through a spatial autocorrelation analysis. A spatial weights matrix, which allows for the calculation of spatial lags of variables, is specified following a binary queen contiguity approach whereby spatial units are classified as neighbours if they share a line or a point border. This matrix is summarised in Equation 1 where *Wij* is the contiguity between spatial units *i* and *j*. The global spatial autocorrelation Moran’s-I statistics is calculated to consider the degree to which the rate of diesel ownership is correlated across neighbours. The Local Indicator of Spatial Association (LISA, reference) is also calculated to assess if particular regions are exhibiting similar rates of diesel car ownership, indicating the presence of hot-spots and low-spots.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Stage Two

Second, the association between nearness to the Republic of Ireland and the proportion of the local car stock that is diesel fuelled is evaluated. A boxplot of the rate of diesel car ownership across the SOAs grouped by buffer category (i.e. the contiguity method) is produced. The Kruskal-Wallis test is applied to determine if these groups of SOAs significantly differ in terms of their rate of diesel car ownership. A set of scatterplots are produced which have the rate of diesel car ownership across the SOAs on the y-axis and Euclidean distance to the closest road crossing (i.e. the proximity method), network distance to the closest fuel station in the Republic of Ireland (i.e. the network distance method), and network time to the closest fuel station (i.e. the network time method) on the x-axis. A Spearman’s correlation analysis is utilised to determine if these variables are significantly related to one another.

Stage Three

Third, a series of log-log regression models are specified in order to explain variation in diesel car ownership across the SOAs. A set of benchmark OLS models are initially produced which have the following arrangements of independent variables:

OLS Model 1

This base model is summarised in Equation 2 were *y* is a vector of observations of diesel car ownership, *α* is a constant parameter, *βa* is a vector of coefficients associated with the area characteristic (i.e. socioeconomic, transport, and household) independent variables, *xa* is a vector set of observations of the area characteristic independent variables, and *ɛ* is the model residual.

|  |  |
| --- | --- |
|  | (2) |

OLS Model 2

This extends OLS Model 1 through the inclusion of dummy variables which cover the buffer categories outlined in the contiguity method. Equation 3 summarises this model, where *βc* is a vector of coefficients associated with the dummy independent variables and *xa* is a vector set of observations of the dummy variables.

|  |  |
| --- | --- |
|  | (3) |

OLS Model 3

This extends OLS Model 1 through the inclusion of a variable measuring the Euclidean distance from the SOA centroid to the closest road crossing which is outlined in the proximity method. Equation 4 summarises this model, where *βp* is a coefficient associated with the proximity variable and *xp* is a vector of observations of the proximity variable.

|  |  |
| --- | --- |
|  | (4) |

OLS Model 4

This extends OLS Model 1 through the inclusion of a variable measuring the network distance from the SOA centroid to the closest fuel station in the Republic of Ireland which is outlined in the network distance method. Equation 5 summarises this model, where *βd* is a coefficient associated with the network distance variable and *xd* is a vector of observations of the network distance variable.

|  |  |
| --- | --- |
|  | (5) |

OLS Model 5

This extends OLS Model 1 through the inclusion of a variable measuring the network time from the SOA centroid to the closest fuel station in the Republic of Ireland which is outlined in the network time method. Equation 6 summarises this model, where *βt* is a coefficient associated with the network time variable and *xt* is a vector of observations of the network time variable.

|  |  |
| --- | --- |
|  | (6) |

In order to determine if the benchmark OLS models need to be extended through the integration of spatial interaction effects to account for persisting spatial autocorrelation, the robust Lagrange Multiplier diagnostics are calculated. Following this, the extension of the benchmark OLS to cover the occurrence of local spatial spillovers is conducted through the specification of the Spatial Durbin Error Model (SDEM). The SDEM incorporates spatial lags of the model independent variables to allow for direct, indirect, and total effects to be estimated. The model also contains a spatial lag of the benchmark OLS model’s residual to account for spatial dependence in the variables omitted from the analysis. Equation 7 and 8 summarise the structure of the SDEM, where *Ɵ* is a vector of coefficients associated with the independent variables, *X* is a vector set of observations of spatially lagged independent variables, *λ* is a coefficient associated with the spatially lagged OLS benchmark model’s residual, and *u* is a vector of observations of the spatially lagged benchmark OLS model’s residual.

|  |  |
| --- | --- |
|  | (7) |
|  | (8) |