Analysing MOT vehicle licensing data and transport model data to generate insights about car use and emissions in Strathclyde

Paul Emmerson, TRL, Jillian Anable (University of Leeds), Tim Chatterton (University of the West of England), Sally Cairns (TRL), Eddie Wilson (University of Bristol), Simon Ball (TRL)

1 Introduction

In 2010, the UK Department for Transport began publishing the records from the annual vehicle roadworthiness inspections (known in the UK as ‘MOT’ tests). These tests are required for every vehicle over three years old, with computerisation beginning in 2005. These data provide details of the make and model of each vehicle, engine size, fuel type, date of first registration and colour, along with the recorded mileage at each test. Using the latter, it is possible to estimate the annual mileage of each vehicle (see Wilson et al., 2013, Chatterton et al., 2015).

In brief, the MOT data allows fuel consumption, energy use and tax bands to be calculated for individual vehicles based on vehicle age, engine size, fuel type and CO₂ emissions (where available from the manufacturer’s rated value, or elsewhere from our own calculations, see Chatterton et al., 2015). Then, using the odometer readings from the vehicle tests, an estimate is calculated for annual mileage (see Wilson et al., 2013 for more details). The linkage to the DVLA licencing data allows the registered keeper of the vehicle to be allocate to a LSOA or datazone (in Scotland) and so allows the vehicle data to be analysed in conjunction with the 2011 Census data and other data records based on this spatial scale.

Specifically, in 2012, the EPSRC awarded a grant, via the Research Councils UK Energy Programme, to develop new methods by which car use, related emissions and energy use could be quantified and understood, building on a previous scoping study (Cairns et al, 2012). The work is currently led by the University of Leeds, working in conjunction with the TRL, Bristol University, the University of the West of England, University of Aberdeen and UCL, (www.MOTproject.net). Project activities have involved developing datasets from MOT and vehicle licensing data in order to explore vehicle ownership and use and resulting emissions and energy use (Chatterton et al, 2015). Much of this work has concentrated on England & Wales because of the availability of additional data relating to LSOAs in England and Wales such as energy statistics and journey time statistics.

The paper aims to give a flavour of the richness of this dataset with examples of the investigations of aspects of vehicle ownership and use in Scotland. It gives a brief description of the construction of the MOT dataset and its characteristics, followed by a description of the spatial pattern of the characteristics of vehicles and their use in Scotland, using the MOT data set and the 2011 Scottish Census data. This is then followed by a more detailed discussion of two particular issues in relation to the Strathclyde region:

- How does the spatial pattern of mileage per car compare with that from a traditional transport model?
- Where do the emitters of emissions, particularly CO₂ and NOₓ live?
2 The dataset

2.1 Construction of the dataset

The ‘MOT’ dataset referred to in this work is in fact the product of combining MOT testing data from the DVSA with Vehicle Stock data held by the DVLA. The relationship between the various input and output datasets is shown in Figure 1. The figure assumes that the ultimate output would be local area tables but it is possible to focus on vehicle types as well, provided that confidentiality restrictions are not compromised.

![Diagram of dataset relationship](image)

**Figure 1 Relationship between constituent data sets in the MOT project**

The MOT records consist of the data collected during the annual vehicle inspection tests for all vehicles (up to 3.5 tonnes) of more than three years old in Great Britain. A computerised system for this information was introduced in 2005, with full implementation by April 2006, and the public release of an anonymised version of this information began in 2010. About 35 million tests are recorded every year, meaning that over 300 million records are being analysed for this project. Data are organised by test, and include details of the nature of the vehicle (date of first use, make, model, colour, and engine size), an odometer mileage reading, together with details of the test (date, the area where the testing station is located and test results). There are related tables providing more details of test failures, although those have not been utilised in this project.

The vehicle licensing data (in the vehicle stock dataset) are, instead, organised by vehicle, and includes details of about 56 million vehicles from 2005 onwards. Unlike the MOT records (where test dates are vehicle dependent), data are reported at regular, quarterly intervals. For this project, key details provided (which are not available from the MOT records) include the location of the registered vehicle ‘keeper’, an indication of whether the vehicle is in personal or commercial ownership, and the vehicle’s CO\(_2\) rating (according to the manufacturer).

The availability of MOT data up to November 2014 means that all vehicles in existence at the time of the 2011 Census should appear in the MOT records by that date. Because the DVLA vehicle stock data can ascribe each registered keeper through their postcode to a Lower-layer Super Output Area (LSOA) in England and Wales, or datazone in Scotland. It is possible to analyse the data set at a finer
degree of spatial disaggregation than that available through current published sources. Through this output area link the characteristics of the vehicle population and use can be related to the socio-demographic characteristics of the output area using 2011 Census data sources and other data sources with links to such output areas, such as Experian income data and accessibility data.

2.2 Characteristics of the dataset

Before describing the spatial characteristic of vehicle ownership and use in Scotland (and Strathclyde in particular) it is useful to provide some more information on the definition of the variables in the data that can be provided by the new MOT dataset and its output files.

2.2.1 Type of vehicle

The MOT dataset uses a definition of vehicle type that is not exactly the same as that used either by either the DfT traffic estimates or used in the Census definition of ‘cars available to households’. For much of the work that the research group has undertaken we have aggregated the classes in the DVLA classification (shown in Figure 3) into 3 broad bands, with nearly all the work focused on the middle band which corresponds to the DVLA Class 4. It has not been possible yet to distinguish ‘cars up to 8 passenger seats’ from other vehicles in class 4 but they will dominate the vehicle numbers in that category. The inclusion of a ‘type of keeper’ variable in the vehicle stock dataset does help distinguish cars likely to be available to households. However, as the next sub-section explains, this is not fool proof.

2.2.2 Form of ‘keepership’

Perhaps the most important issue to note is the definition of ‘ownership used in the MOT dataset. The DVLA vehicle stock dataset defines the postcode of the ‘Registered Keeper’ of the vehicle. It further distinguishes the type of Keeper into ‘Private’ and ‘Commercial’ based on the title on the vehicle record. Vehicles are coded as private or commercial based on the registered keeper’s title. This is either personal (e.g. Mr, Ms, Dr, Rev etc) or business related (e.g. Company, Messrs etc.). With a company car, the day-to-day user is likely to be the registered keeper. This, in reality, is not a perfect correlation to what we would define as Private and Commercial. For instance many commercial vehicles are registered to Private keepers. The MOT-dataset also suggests that there are cases where ‘private’ vehicles are actually commercial vehicles. In the England & Wales, data records there are a limited number of LSOAs where there are many more vehicles registered to private keepers than there are people! In Scotland, such obvious anomalies are not present but Error! Reference source not found., which shows a comparison of cars registered to private keepers (‘P4 vehicles’) with the 2011 Census using the ‘car availability’, does highlight one data zone (circled in red) in the Charing Cross area of Glasgow where the number of cars registered with a private keeper is twice that of the number of cars available to the population in the 2011 Census! The figure also shows that, in general, the number of Class 4/4A vehicles registered with private keepers tends to be less than the total number of cars ‘available to the population’ at the time of the 2011 Census. This probably reflects the fact that a proportion of vehicles registered to a commercial keeper are available for use by households. In general, however, the correlation is very high. The analysis of the dataset present in this paper has been solely concerned with vehicles in Class 4/4A with a ‘Private’ keeper variable attached which are referred to here as ‘Cars’.

For Scotland the impact of commercially-kept cars is quite small with such cars being slightly less than 10% of all cars registered in Scotland and undertaking slightly more than 10% of distance travelled (Table 1).
Figure 2 Comparison of Census and MOT definitions of ‘cars’ (Scotland, 2011 data)

<table>
<thead>
<tr>
<th>Class</th>
<th>Vehicle Description</th>
<th>1st Test</th>
<th>MOT Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Motorbike (engine size up to 200 cm$^3$)</td>
<td>3 Years</td>
<td>£29.65</td>
</tr>
<tr>
<td>Class 1</td>
<td>Motorbike with sidecar (engine size up to 200 cm$^3$)</td>
<td>3 Years</td>
<td>£37.80</td>
</tr>
<tr>
<td>Class 2</td>
<td>Motorbike (engine size over 200 cm$^3$)</td>
<td>3 Years</td>
<td>£29.65</td>
</tr>
<tr>
<td>Class 2</td>
<td>Motorbike with sidecar (engine size over 200 cm$^3$)</td>
<td>3 Years</td>
<td>£37.80</td>
</tr>
<tr>
<td>Class 3</td>
<td>3-wheeled vehicles (up to 450kg unladen weight)</td>
<td>3 Years</td>
<td>£37.80</td>
</tr>
<tr>
<td>Class 3</td>
<td>3-wheeled vehicles (over 450kg unladen weight)</td>
<td>4 Years</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4</td>
<td>Cars (up to 8 passenger seats)</td>
<td>3 Years</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4</td>
<td>Motor caravans</td>
<td>3 Years</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4</td>
<td>Quads (max unladen weight 400kg - for goods vehicles 550kg and max net power of 15kw)</td>
<td>3 Years</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4</td>
<td>Dual purpose vehicles</td>
<td>3 Years</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4</td>
<td>Private hire and public service vehicles (up to 8 seats)</td>
<td>3 Years</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4</td>
<td>Ambulances and taxis</td>
<td>1 Year</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4</td>
<td>Private passenger vehicles and ambulances (9 to 12 passenger seats)</td>
<td>1 Year</td>
<td>£57.30</td>
</tr>
<tr>
<td>Class 4</td>
<td>Goods vehicles (up to 3,000kg design gross weight)</td>
<td>3 Years</td>
<td>£54.85</td>
</tr>
<tr>
<td>Class 4A</td>
<td>Class 4 vehicles (9 to 12 passenger seats) with a seat belt installation check</td>
<td>n/a</td>
<td>£64.00</td>
</tr>
<tr>
<td>Class 5</td>
<td>Private passenger vehicles and ambulances (13 to 16 passenger seats)</td>
<td>1 Year</td>
<td>£59.55</td>
</tr>
<tr>
<td>Class 5</td>
<td>Private passenger vehicles and ambulances (more than 16 passenger seats)</td>
<td>1 Year</td>
<td>£80.65</td>
</tr>
<tr>
<td>Class 5A</td>
<td>Class 5 vehicles (13 to 16 passenger seats) with a seatbelt installation check</td>
<td>n/a</td>
<td>£80.50</td>
</tr>
<tr>
<td>Class 5A</td>
<td>Class 5 vehicles (more than 16 passenger seats) with a seatbelt installation check</td>
<td>n/a</td>
<td>£124.50</td>
</tr>
<tr>
<td>Class 7</td>
<td>Goods vehicles (over 3,000kg up to 3,500kg design gross weight)</td>
<td>3 Years</td>
<td>£58.60</td>
</tr>
</tbody>
</table>

Figure 3 Classification of vehicle types distinguished in the MOT-dataset.
Table 1 Impact of type of keeper on numbers and distance travelled. (MOT dataset -Scotland only 2011)

<table>
<thead>
<tr>
<th>Keeper type</th>
<th>Number of Class 4 vehicles (‘cars’)</th>
<th>Average distance per Class 4 vehicle (.000km)</th>
<th>Total distance travelled by Class 4 Vehicles (.000km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private</td>
<td>2,178,788</td>
<td>12.472</td>
<td>27,172,839</td>
</tr>
<tr>
<td>Commercial</td>
<td>194,716</td>
<td>17.926</td>
<td>3,490,464</td>
</tr>
<tr>
<td>Private as % of total</td>
<td>91.8%</td>
<td>96.5%</td>
<td>88.6%</td>
</tr>
</tbody>
</table>

2.2.3 Estimating mileage in the MOT dataset

The generation of mileage estimates has been a major analytical task of the project, and is reported in more detail elsewhere (see, for example, early work on this topic by Wilson et al 2013a and 2013b). Particular challenges include the fact that although roughly annual, test intervals vary; there are erroneous readings recorded (due to both accidental and deliberate error); there will be variation in vehicle use between test intervals (not least due to changes in the weather and associated activities – e.g. the traditional British summer holiday); there are seasonal patterns of car purchasing (which generate an alternative set of seasonal impacts); and, finally, most vehicles do not provide a mileage until they are at least three years old. Processes for dealing with these range from complex heuristics to identify erroneous values, through to simple assumptions (e.g. that mileage increases linearly from zero to the three year value). Given seasonal issues and the processes used, the mileage data is only considered to be robust for annual estimates. It should be noted that all distances quoted in this paper are in kilometres.

2.2.4 Emission characteristics of vehicles

As mentioned earlier the MOT-dataset contains the vehicle manufacturer’s estimate of the CO₂ emission level for that vehicle (in grams per km). These values are based on the NEDC test cycles. However, this is only available for vehicles produced since about 2000 and so is not available for some 30% of the vehicle stock in 2011 (but will of course be a smaller proportion now). The research group has undertaken a detailed piece of work to ascribe emission factors for CO₂, NO₂, and PM₁₀ to virtually all ‘cars’ in the dataset, making use of data on vehicle manufacturer, brand and engine size and type that are available in the dataset. These emission factors are based on the COPERT vehicles’ tests and probably give a more realistic estimate of emissions from the vehicles although they do assume constant driving conditions across all vehicles, and by expansion, across all areas. This is a point to bear in mind when comparing the results with more results from transport model-based estimates of emissions discussed later on.

2.2.5 Other characteristics of the MOT dataset – time-series analysis

All the work so far has been on creating a comprehensive and checked dataset for the situation around the time of the 2011 Census, but the existence of data for a number of years means there will be opportunities to consider changes over time as well as space in vehicle use. The current time series is restricted by the fact that the MOT data does not normally include cars until they have their first MOT after 3 years and the data for the early years of computerisation were problematic. Enabling the dataset to investigate time-series issues is being investigated.
3 Car ownership and use in Scotland

3.1 Car ownership

Before turning to the Strathclyde region in particular, it is useful to look at the spatial characteristics of car ownership and use in Scotland as a whole, as picked out by the MOT dataset. As might be expected, the spatial pattern of car ownership per person is very similar to that obtained using 2011 Census data. One of the issues that has concerned the research group is whether there are spatial variations in car ownership that are not being picked up in regression models based on traditional aspatial assumptions, or in car ownership models developed on household data such as the UK car ownership model. Early analysis of the England & Wales data has found that there are spatial patterns that are not explained by a traditional regression model, based on LSOA or MSOA level input variables including Census, income and accessibility input variables (Yeboah, et al, 2016).

Specifically, the values of the regression residuals from these models are more clustered than one might expect if the data was spatially independent. The residuals show spatial autocorrelation. Work on this topic in relation to other characteristics of vehicles and their use continues but an initial study of Scottish car ownership also shows a clustering of positive (higher car ownership than a normal aspatial regression model would predict) and negative (lower than predicted car ownership) residuals. There is higher than predicted car ownership in the Shetlands, Speyside and large parts of Dumfries and Galloway and, conversely, lower than expected car ownership in Caithness, North-East Scotland and Argyll and Bute.

Whilst such spatial variations can be described it is more of a challenge to explain why they occur, although one obvious avenue that can be explored is that the spatial patterns are caused by the effects of variables not already considered. Another is that the impact of variables in the regression model is not stationary – that is their effect may vary spatially.

3.2 Characteristics of cars

3.2.1 Vehicle age

The MOT dataset enables us to explore the spatial aspects of some of the characteristics of the vehicles in the dataset. Two of these have been analysed for this paper; namely vehicle age and vehicle size, both of which might be expected to influence the emission from cars.

Figure 4 shows the average age of car by datazone in Scotland in 2001. The highest average ages of cars (> 2 standard deviations = >9.6 years old) are to be found in datazones in the far North (Sutherland and the Orkneys), in parts of North-East Scotland and also in parts of Edinburgh. In contrast, the data zones with the youngest average age (<2 standard deviations or < 7 years old) are to be found largely around the Clydeside conurbation but are also relatively widely scattered including semi-rural areas such as Mearnskirk, but also urban areas such as Darmoyne and Easterhouse. The association of average age of car with median income and ‘cars per person’ is complex. Whilst the average age of a car in a datazone declines as the median income of a data zone increases there is evidence that for those data zones with the highest car ownership levels (>0.55 cars per person) that the average age of the car starts to increase again.

3.2.2 Vehicle size

The spatial pattern of mean engine size is slightly different. On average, the largest cars are found in the rural areas especially in North-East Scotland, but there are pockets of larger than average (>2 standard deviations - >1900 cc) cars in Central Edinburgh (especially the New Town) and in the centre of Glasgow. In contrast, the datazones with the smallest average car size (<2 standard deviations, <1550 ccs) are found scattered throughout the Clydeside conurbation both in the inner and outer suburbs. There are also fewer of these because the distribution of average size is skewed to the right.
Whilst median household income per datazone explained more of the variation in average car size (32%) than it did for average age (19%) and whilst there was a significant negative correlation between average age and average car size, the actual correlation between the two characteristics was very small (R^2=0.2\%). Consequently, to all intents and purposes the two characteristics are unrelated at a datazone level.

Figure 4 Average age of vehicle (based on the 2011 MOT dataset)

Figure 5 Average size of vehicle (based on the 2011 MOT dataset)
4 Car use and emissions in Strathclyde

4.1 Comparisons with transport model outputs

At this stage it was hoped to provide a comparison of the MOT dataset outputs for distance travelled by cars against similar outputs from the new Strathclyde Regional Transport model (SRTM) and the associated new version of the Strathclyde Integrated Transport and Land-Use Model (SITLUM). In the event, the timetable for the development of these models has precluded their use in this paper but a comparison is made with the current version of SITLUM which has a base year of 2011 (but based on aspects of both the 2001 and 2011 Censuses). With this model it is also possible to contrast the spatial pattern of car emissions by where they are emitted (from SITLUM) against the residence of those that use the cars (from the MOT dataset). The results from the MOT dataset have been aggregated up to the SITLUM zoning system and only those zones within the Strathclyde Region have been considered.

In making these comparisons the differences between the two sources of data will also be discussed.

4.2 Distance that cars are driven

In comparing the estimates of how much cars are driven by residents in different areas one needs to be aware of how the estimates are obtained. The estimates from the MOT dataset, shown in Figure 6 Distance per car from 2011 MOT dataset (,000km per annum) are based on interpolated average annual distances by car from the MOT dataset for those cars assumed to be kept by private individuals. Some additional mileage will be undertaken by commercial cars registered in Scotland, and some by those registered as Commercial cars in England & Wales but based in Scotland.

In the case of the estimates from the transport model, in this case the SITLUM model for Strathclyde, (Figure 7) the estimates are outputs from the modelling process - total kilometres over a weekday. They have been estimated by adding together all home-based trips and their trip distance and ascribing it to the home zone. Non-home based travel is ignored since it cannot be ascribed to a given residential zone. The pattern of trips, and their distances, is a complex function of the distribution of
trip origins and destinations, and assumptions about the deterrent effect of time and cost of travel by car.

What both maps show is that the greatest intensity of distance per car is found in the rural areas surrounding the Clydeside conurbation whilst the least used cars are those in Glasgow itself, as well as the Isle of Bute. There are, however, some differences between the patterns from the two models. (Both models show the distance per car in quintile bands.) The rural-urban split is more evident from the MOT data. Meanwhile the zones in the southern part of Argyll and Bute have relatively lower mileages than forecast from the SITLUM model. The outliers are better appreciated by graphing the estimates of distance per car by the two approaches (Figure 8). From the figure is evident that the model estimates for some of the islands and peninsulas in the Strathclyde region are much higher than would be expected on the basis of their MOT estimates - Campeltown and Cumbrae especially. One of the Glasgow city centre zones also appears to be a large overestimate. Large underestimates by the transport model are rare but distances per car appear to be underestimated in East Kilbride.

To understand why a transport model may not provide an estimate of the distance per car consistent with the MOT estimate one needs to consider how a typical model estimates the number and distribution of car driver trips.

Firstly, the estimate of trips generated by a zone will take into account the socio-economic characteristics of the zone and use the relevant trip rates (by purpose) but these will only be crudely split by location.

Secondly the sensitivity to costs in the trip distribution model (and mode choice mode) will normally only be disaggregated by sector/region at most. The SITLUM model shown here calibrated zone specific values for travel to work by car based on Census data for 2001 but used TMfs model parameters for other trip purposes which were, at the time, only disaggregated by 5 area types. This
can cause problems for certain types of zone such as islands and peninsulas were trip opportunities are highly constrained even when allowance is made for the high cost of travel for ferries.

Thirdly, the distribution of trips by car drivers is further affected by the modelling of the choice of mode for a trip which will affect the proportion of trips undertaken as a car driver. Again the estimated parameters will usually only have very simple spatial structures.

Fourthly, it should be remembered that the comparison between the two estimates is not perfect in that they are estimating slightly different things. The transport model estimates here does not include travel on none-home based trips and the MOT estimate excludes the travel in cars registered to commercial keepers but used on household business including business travel. The problem is that the spatial pattern of both this quantities is not known.

4.3 Emissions

The other comparison that can be made between the outputs from the MOT data set and that from a transport model is in the spatial pattern of emissions. Comparison of outputs from the two sources is more complicated than that for distance per car. Emissions per car from the MOT dataset are based on estimating the mean CO₂ per km for each class of vehicle in the MOT dataset using the COPERT4 emission tests. The actual emissions depend on the assumption of how much travel is undertaken on urban, semi-urban and rural roads, and at what assumed speed (including motorways) but these travel assumptions are standardised across all vehicles. So the MOT dataset measures the emissions with an assumed speed profile but allows for variations in vehicle types. The transport model emission outputs works the other way around. Since they are estimated where they are emitted, they can take account of the speeds undertaken at that point, but they do assume (for cars) an ‘average car’. Because of the different spatial aggregation, it is not possible to use a consistent output measure for comparison.

Consequently, using the MOT dataset Figure 9 shows the amount of CO₂ emitted per annum per car ‘kept’ in a zone. As might be expected the pattern of CO₂ emissions largely reflects that from the distance per car. The amount of CO₂ emitted per car by rural areas is generally much greater than that.
from the cars in urban areas of Glasgow. There is a hint that CO₂ emissions per car in the central zones of Glasgow could be higher than might be expected from the distances travelled – this is likely to be because the average CO₂ per km for these cars is higher than the average, not because they are kept in a congested area.

Meanwhile in the case of the transport model, the emissions are plotted ‘per km²’ rather than per car since this gives a better picture of the intensity of emissions where they are emitted (Figure 10). The importance of travel along both side of the River Clyde and along the motorways is especially highlighted.

Figure 9 Generation of CO₂ per car by home zone from 2011 MOT dataset (kgs per annum)

The difference between these two approaches to ascribing emissions is probably more important when considering emissions which can be thought of as directly injurious to health such as NOx and PM₁₀. In this paper we focus on NO₂ since the PM₁₀ plots would be very similar because we are only in a sense dealing with the impact of diesel cars. Figure 11 shows the amount of NO₂ emitted per car based on the MOT dataset, based on quintiles. The rural/urban split is even more pronounced than for CO₂, because of the greater proportion of diesel cars in rural areas who are responsible for most of the NO₂ emitted by cars in recent years. The impact of petrol cars on NO₂ emissions is now very slight.
Figure 10 CO\textsubscript{2} per area from SITLUM transport model (kgs per km\textsuperscript{2} per weekday) -2011 forecast year

Figure 12 shows where the NO\textsubscript{2} is actually emitted, again based on quintiles, from the SITLUM model. It is obvious that the NO\textsubscript{2} is emitted in a very different pattern to that of the distribution of emittance per car, with the highest rates along the urban motorway corridors where traffic flows are greatest. These differences in the spatial patterns of those doing the emitting and those potentially suffering from the emissions is an issue related to ideas of social justice and is being pursued by the research group (see Barnes and Chatterton, 2014).
5 Conclusions

This paper has aimed to provide a flavour of the potential of the newly constructed MOT dataset in relation to Scotland and to the Strathclyde region in particular. The existence of the dataset opens up opportunities to investigate spatial patterns of vehicle ownership, characteristics and use. Whilst the dataset is not without its quirks, and its relationship with existing datasets such the census or transport
model outputs are not straightforward, it is capable of providing insights into an area of research at a highly spatially disaggregated scale that have not been possible before.

6 References


Acknowledgements

The work has been undertaken as part of the MOT project (EP/K000438/1), funded by the UK Engineering and Physical Sciences Research Council under the Research Councils UK Energy Programme. Grateful thanks to members of DfT, DVSA, DVLA and DECC, who have provided advice and support for this work, and to other members of the project team: Dr Jo Barnes at the University of the West of England, Dr Oliver Turnbull at University of Bristol, and Dr Godwin Yeboah.

Contains National Statistics data © Crown copyright and database right 2012
Use also made of Experian income data provided via the UK Data Archive.
MOT Project website http://www.MOTproject.net

1 Estimates of emissions: Manufacturers estimate of emissions are based on the New European Driving Cycle (NEDC). This is a driving cycle, last updated in 1997, designed to assess the emission levels of car engines and fuel economy in passenger cars (which excludes light trucks and commercial vehicles). It is also referred to as MVEG cycle (Motor Vehicle Emissions Group). It is supposed to represent the typical usage of a car in Europe. It consists of four repeated ECE-15 urban driving cycles (UDC) and one Extra-Urban driving cycle (EUDC) with about half the distance being on the extra-urban cycle.

Emissions of NOx, PM10 and CO2 for vehicles in the MOT dataset have been calculated from a set of generic emission factors developed by AQMRC (Barnes and Bailey, 2013). These are based on the currently best available data for the UK from a range of sources (primarily COPERT 4 (v8.1), TRL, NAEI and EMEP/EEA). The emission factors for NOx, PM10 and CO2 used were for cars of different fuel types: Petrol (<1400cc, 1400-2000cc, and >2000cc: Pre-Euro and Euro 1-6), Diesel (<2000cc and >2000cc: Pre-Euro and Euro 1-6), LPG (all engine sizes: Euro 1-6 (pre-Euro 1 treated as Euro 1)), Hybrid (single factor). These emission factors were available for urban, rural and motorway driving. A
compound emission factor was calculated based on figures from Transport Statistics Great Britain (DfT, 2012) that split total mileage for cars, motorbikes and LGVs between 19% motorways, 28% urban and 43% rural. Euro standards for the vehicles have been based on date of first registration in relation to the EU compliance date for the relevant standards. Due to the absence of data on body type, all P4 vehicles have been treated as ‘cars’.