Using MOT records to estimate local level vehicle annual mileages

Rachel Moyce[1] and Daryl Lloyd[1, 2]

[1]Statistics Travel and Safety Division, Transport Statistics, Department for Transport, 2/13 Great Minster House, 33 Horseferry Road, London SW1P 4DR

Tel. 020 7944 3077

rachel.moyce@dft.gsi.gov.uk, daryl.lloyd@dft.gsi.gov.uk

[2]UCL Centre for Advanced Spatial Analysis, Gower Street, London WC1E 6BT

KEYWORDS: transport, car ownership and use, roadworthiness, odometer readings, MOT tests, vehicle mileage rates, open data

1. Introduction

Although a number of ways of estimating average vehicle mileages have been demonstrated in the past, the majority of these have been based around the collection and processing of project-specific survey data, see, for instance, Yang et al (2004). An alternative data source that has become more available over recent years has been in-vehicle GPS systems such as those used by TrafficMaster in the UK. On the whole, though, data from these systems are commercial datasets with restrictions on their use and have not, generally, been used for wider research topics (Tao and Tsai 2007). Furthermore, such devices tend to be restricted to larger fleets of vehicles or high-value vehicles where the owner is concerned about tracking the vehicle in case of theft. In these cases, it can be argued that such vehicles are not ‘typical’ and therefore any attempt to derive annual mileages would be subject to significant biases.

In the UK and elsewhere, some of the most reliable mileage estimates have been produced from household surveys (Sana et al 2010). In Great Britain specifically, the National Travel Survey (DfT 2011) has traditionally been used to create estimates of how far, on average, both cars and drivers travel each year. Mokhtarian and Cao (2008) provide a critic of methodologies used for the collection of travel information through surveys.

During 2011 the Vehicle and Operator Services Agency (VOSA) made the complete MOT Test dataset available to the Department for Transport and an anonymised version available to the public (http://data.gov.uk/dataset/anonymised_mot_test). As this release includes odometer readings recorded at the time of the test, a new source of information has been made available which can help understand how far vehicles are driven each year. By combining the MOT data with DVLA keepership information (i.e. the postcode of the keeper) it has become possible to create mileage estimates for local areas of where vehicles are kept.

2. The MOT dataset

The MOT test is a test of vehicle safety, roadworthiness and emissions which is required for most vehicles used on the public highways in the UK. Most vehicles need to undertake a test annually from the third anniversary of the vehicle first being registered.

The MOT test in GB is overseen by VOSA; they specify what components and tests are carried out on each vehicle and authorise individual testers at testing stations (usually local garages) to carry out such tests and issue certificates. Although MOT tests have been carried out in the UK since 1960, many of the requirements of the test are specified as part of EU Directive 2009/40/EC.

Since 2005, all MOT tests have been logged on VOSA's central database, MOTComp. Each vehicle testing station (VTS) has secure access to MOTComp and the outcome and records for each test are recorded as part of the MOT process.

One variable that is captured on MOTComp at the time of the test is the odometer reading for the vehicle. Although this is an optional variable around 95 per cent of the MOT records include it.

The MOT dataset that has been released by VOSA and used for this analysis does not include heavy goods vehicles (i.e. with a gross maximum permitted weight over 3.5 tonnes) or large buses and coaches. In total, roughly 38 million MOT tests are carried out on cars, motorcycles, minibuses vans in Great Britain each year. As shown in Table 1, this has resulted in around 192 million MOT records (from MOTs undertaken between April 2005 and December 2010) which were available for this analysis.

Table 1. Number of MOT tests in MOTCOMP by year and test class

Thousands of tests

Vehicle type	2005	2006	2007	2008	2009	2010	Total
Motorcycles (Class 1 and 2)	198.1	1,038.9	1,067.7	1,111.1	1,153.3	1,161.4	5,730.5
Cars, vans (<=3 t), taxis, ambulances and private passenger vehicles (up to 12 seats) (Class 4 and 4a)	7,541.3	32,280.9	34,028.1	35,113.7	36,095.9	36,736.5	181,796.5
Private passenger vehicles and ambulances (over 12 seats) (Class 5 and 5a)	17.3	65.2	65.4	67.2	68.2	67.0	350.4
Goods vehicles (3-3.5 t) (Class 7)	161.9	702.0	759.3	795.7	829.8	877.8	4,126.5
Three-wheel vehicles (Class 3)	5.5	19.5	17.5	16.7	16.2	15.4	90.9
Unknown class	162.1	5.1	0.0	0.0	0.0	0.0	167.1

This paper will not seek to explain or discuss the MOT dataset, details of MOTComp or how the data are structured. The best information about this can be found in DfT (2012) and Wilson et al (2013).

3. Methodology

The main methodological challenge in analysing the MOT records is to correctly match consecutive tests for individual vehicles. Matching tests in this way allows researchers to calculate both the time and the difference in odometer readings between tests. As MOT tests are not undertaken precisely on yearly intervals, these calculations are the key to estimating annual average mileages.

Two main challenges make such a matching process hard. First is the opportunity for testers to make recording mistakes whilst entering data. Although MOTComp contains data from the DVLA vehicle register, testers are able to manually enter vehicle details. Any mistakes in entering either the vehicle registration mark (VRM) or vehicle identification number (VIN) can result in either the MOT record not matching any other record, or, even more serious, erroneously being matched to a different vehicle's sequence of MOT tests.

The second difficulty is caused by cherished transfers (CTs). A CT takes place when a vehicle owner decides to move a VRM from one vehicle to another. This could either be between two vehicles owned simultaneously by the keeper or from an old vehicle to a new one. The result of a CT is that MOTComp can end up with two (or more) different vehicles with exactly the same VRM at different points in time. If the different vehicles are not separated the estimated annual average mileage will be incorrect and possibly even negative.

MOTComp does attempt to ensure that CTs are correctly dealt with by cross-referencing the VRM with VIN and other vehicle details. For instance, if a VRM is added for a second time and what seems to be a new vehicle, a new entry is created in the vehicles table and specific MOT tests cross-reference the unique entries in the vehicles table.

To help deal with these problems, each MOT test record was matched to the correct VIN for the individual vehicle that was being tested as closely as possible. This was carried out using data in both the MOT dataset and the DVLA vehicle register. An example of the resulting test records is shown in Figure 1 (note that the VRM and VIN entries have been partially redacted to protect the vehicle keeper's privacy). The records in rows 1 to 10 (with a VRM of *B16G##*) all have exactly the same VRM but three different VINs. Therefore we can conclude that this specific VRM was transferred between three vehicles during 2005 to 2010.

	TestVehVRM	VIN	VehTestCntr	TestVehDetHistCntr	VehicleTestDate	MOTTestResult	VehTstRecMileage	VehTstExpiryDt
1	B16G###	SHHCG###	-1	0	2006-04-21 00:00:00.000	P	120061	2007-04-20 00:00:00.000
2	B16G###	SHHCG###	0	0	2006-04-19 00:00:00.000	P	119991	2007-04-18 00:00:00.000
3	B16G###	SHHEU###	-7	-1	2008-04-02 00:00:00.000	P	131298	2009-04-04 00:00:00.000
4	B16G###	SHHEU###	-6	-6	2008-04-01 00:00:00.000	F	131292	NULL
5	B16G###	SJNFD###	-9	-8	2010-02-15 00:00:00.000	P	109389	2011-02-26 00:00:00.000
6	B16G###	SJNFD###	-8	-8	2010-02-11 00:00:00.000	F	109389	NULL
7	B16G###	WBABG###	-5	-1	2007-12-12 00:00:00.000	P	136438	2008-12-11 00:00:00.000
8	B16G###	WBABG###	-4	-4	2007-12-12 00:00:00.000	F	136438	NULL
9	B16G###	WBABG###	-3	-3	2006-12-08 00:00:00.000	P	128756	2007-12-07 00:00:00.000
10	B16G###	WBABG###	-2	-3	2006-12-07 00:00:00.000	F	128737	NULL
11	B16H###	WDB20###	-6	0	2010-03-07 00:00:00.000	F	83596	NULL
12	B16H###	WDB20###	-5	0	2010-03-07 00:00:00.000	P	83596	2011-03-06 00:00:00.000
13	B16H###	WDB20###	-4	0	2009-02-04 00:00:00.000	P	67370	2010-02-12 00:00:00.000
14	B16H###	WDB20###	-3	0	2008-02-13 00:00:00.000	P	52272	2009-02-12 00:00:00.000
15	B16H###	WDB20###	-2	0	2008-01-31 00:00:00.000	F	51880	NULL
16	B16H###	WDB20###	-1	0	2007-02-05 00:00:00.000	P	44286	2008-02-04 00:00:00.000
17	B16H###	WDB20###	0	0	2006-02-01 00:00:00.000	P	37963	2007-01-31 00:00:00.000

Figure 1. Example of MOT records (limited fields)

Once individual vehicles have been identified it becomes a relatively simple task to match consecutive tests together and study the sequence of test dates, results and odometer readings for each vehicle. An example of this can be seen in Figure 2.

	TestVehVRM	VIN	VehicleTestDate1	VehTstRecMileage1	MOTTestResult1	VehicleTestDate2	VehTstRecMileage2	MileageDone	DaysBwTst	AnnualAvgMileage
1	B16G###	SHHCG###	2006-04-19 00:00:00.000	119991	P	2006-04-21 00:00:00.000	120061	P	2	12783
2	B16G###	SHHEU###	2008-04-01 00:00:00.000	131292	F	2008-04-02 00:00:00.000	131298	P	1	2191
3	B16G###	SJNFD###	2010-02-11 00:00:00.000	109389	F	2010-02-15 00:00:00.000	109389	P	4	NULL
4	B16G###	WBABG###	2007-12-12 00:00:00.000	136438	F	2007-12-12 00:00:00.000	136438	P	0	NULL
5	B16G###	WBABG###	2006-12-08 00:00:00.000	128756	P	2007-12-12 00:00:00.000	136438	F	369	7305
6	B16G###	WBABG###	2006-12-07 00:00:00.000	128737	F	2006-12-08 00:00:00.000	128756	P	1	6939
7	B16H###	WDB20###	2010-03-07 00:00:00.000	83596	P	2010-03-07 00:00:00.000	83596	F	0	NULL
8	B16H###	WDB20###	2009-02-04 00:00:00.000	67370	P	2010-03-07 00:00:00.000	83596	P	396	14610
9	B16H###	WDB20###	2008-02-13 00:00:00.000	52272	P	2009-02-04 00:00:00.000	67370	P	357	15340
10	B16H###	WDB20###	2008-01-31 00:00:00.000	51880	F	2008-02-13 00:00:00.000	52272	P	13	10957
11	B16H###	WDB20###	2007-02-05 00:00:00.000	44286	P	2008-01-31 00:00:00.000	51880	F	360	7670
12	B16H###	WDB20###	2006-02-01 00:00:00.000	37963	P	2007-02-05 00:00:00.000	44286	P	369	6209

Figure 2. Example of matched pairs of consecutive tests (limited fields)

Unfortunately, though, our analysis has shown that any attempted reference to a table of unique vehicles (as opposed to just unique VRMs) is not perfect and it still leaves a significant number of tests which cannot be accurately matched together. In order to deal with this problem, and the issue of logging errors as described earlier, we have used a set of thresholds to remove erroneously matched vehicles. Examples of these thresholds are the removable of matched pairs where a single vehicle seems to have two passes within relatively short time (row 1 in Figure 2 shows a specific example of this) and annual average mileages that are under 250 miles or over 50,000 miles per year). Fewer than 5 per cent of the MOT records were removed as they either could not be matched to specific vehicles, matched to other tests or they failed one of these thresholds.

Although we were working with the full set of MOT records, most of the work as described to this point can be completed independently using the publicly-available anonymised MOT records. The main extra value we were able to bring was in using DVLA vehicle register records in order to more accurately match pairs of tests. This results in a greater number of matched pairs than would be achievable with the anonymised data.

4. National mileage estimates

Given a set of matched pairs of consecutive test the production of national average annual mileages is relatively straightforward. In theory, every (non-HGV or larger coach / bus) vehicle over three years of age is represented within the database meaning that, even taking into account the mismatches or rejected pairs, the underlying data is closer to a census of vehicle mileages than a sample survey.

The outputs are of greater value if the vehicle / car fleet is segmented by, for instance, age, keeper type (private / business) body type or fuel type. Average mileages for a segmented fleet can be used to feed the assumptions of a number of models, such as the transport section of the National Atmospheric Emissions Inventory (NAEI) and DfT's own National Transport Model (NTM).

Table 2. Annual average mileage by body type

Vehicle Body Type	Annual Average Mileage miles
AGRICULTURAL	12,400
BUSES & COACHES	18,025
CARS	10,764
GOODS - HEAVY	18,483
GOODS - LIGHT	12,972
MOTORCYCLES, MOPEDS & SCOOTERS	6,477
NOT RECORDED	10,272
OTHERS	15,071
SPECIAL PURPOSE	15,275
TAXIS	26,768
TRICYCLES	12,655

Segmenting the records by the DVLA registered vehicle body type, as can be seen in Table 2, gives an annual average for car mileage of 10,764. Compared with this, light goods vehicles, buses & coaches and taxis (Hackney Carriages) all travel further on average, as would be expected.

Table 3. Annual average mileage by age of vehicle for cars

Vehicle Age years	Annual Average Mileage miles
3	9,587
4	8,934
5	8,478
6	8,048
7	7,652
8	7,294
9	6,923
10	6,573
11	6,248
12	5,911
13	5,592
14	5,285
15	4,959

Table 3 shows how the average mileage for cars varies with vehicle age. The table starts at 3 years of age as this is when most cars undertake the first MOT. This analysis demonstrates that newer vehicles tend to drive greater distances than older vehicles.

Table 4. Car annual average mileage by fuel type

Fuel Type	AvMileage miles
DIESEL	10,019
ELECTRIC DIESEL	13,132
ELECTRICITY	2,327
FUEL CELLS	10,749
GAS	7,974
GAS BI-FUEL	11,641
GAS DIESEL	9,394
HYBRID ELECTRIC	10,600
NEW FUEL TECHNOLOGY	9,039
PETROL	6,627
PETROL/GAS	6,731

Table 4 shows the final segmentation in this paper: splitting cars by fuel types. Of the two most common fuels, diesel-fuelled cars travel considerably further on average than petrol-fuelled.

4. Spatial distribution of vehicles

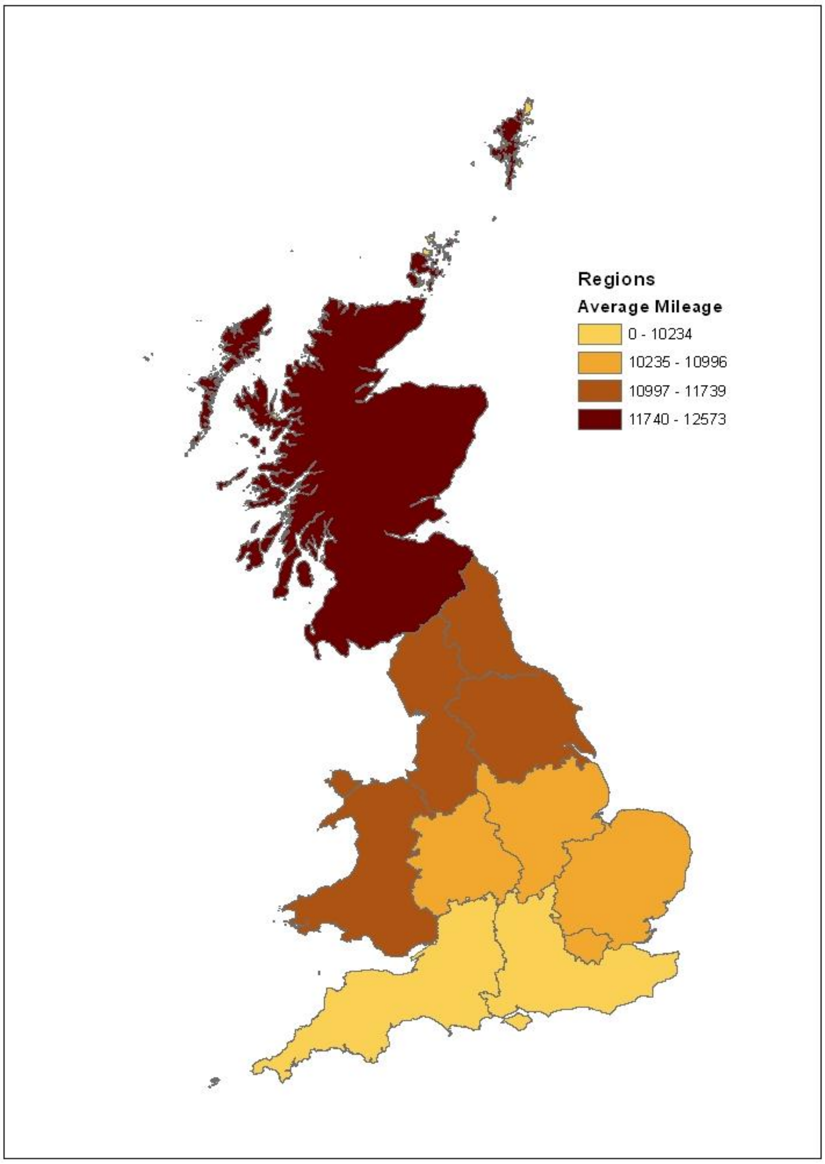

Figure 3. Map of GB showing annual average car mileage, by region of registered keeper

As well as segmenting the fleet by vehicle characteristics, it is possible to analyse annual average mileage by the location of the registered keeper. Figure 3 shows some basic analysis of mileages by region, highlighting slightly increased mileages in Scotland, the north of England and Wales.

The real value of this analysis will come from a more detailed understanding of fine-scale spatial distributions of vehicle mileage along with the relevant vehicle and geodemographic segmentations. Further work on this will be completed during winter 2012/13 with the aim of extracting and exploring detailed spatial patterns.

5. An imperfect model?

As with all models, the analysis and methodology presented within this paper is far from perfect. Perhaps the greatest advantage analysis based on the MOT dataset has over other methods for estimating distances is sample size and lack of bias. The National Travel Survey, currently used as the main official estimate of average distance driven, samples around 10,000 vehicles from 8,000 households each year. Whilst this sample size is high enough for generalised estimates it is not enough for vehicle segmentations or spatial disaggregations.

However, the MOT database is significantly lacking in newer vehicles, especially cars under 3 years of age. It is highly likely that these vehicles will have a different usage pattern than older cars especially as over half of the cars registered for the first time in recent years in Great Britain have been licensed to businesses, not individuals (DfT 2012b). More work needs to be carried out to understand how this biases the outputs.

As demonstrated, although it is possible to extract valuable information about the spatial distribution of mileages driven, this does not necessarily tell us anything about **where** the mileage is driven. However, it can be assumed that most mileage is driven near to where the vehicle is kept as around 94 per cent of car trips are under 25 miles and over 50 per cent are under 5 miles (DfT 2011 – table NTS0308). Nevertheless, vehicles which are driven significantly larger distances each year than the average vehicle are likely to operate at greater distances from where the vehicle is registered.

6. Conclusion

The MOT odometer data is a rich dataset which has only just started to be explored, see Cairns et al (forthcoming) and Wilson et al (2013). The information about how vehicles are used can be improved even more by coupling it with DVLA vehicle data. This initial analysis of annual mileages segmented by vehicle type and geography has given a glimpse of the potential end uses.

The Department for Transport recognises that whilst some of the analysis can be carried out by anyone using the MOT data released on data.gov.uk, some of the detailed mileage and location analysis can only be produced in conjunction with the DVLA vehicles register. Therefore the aim is to produce a mileage dataset which can be published and reused by academics, researchers and consultants anywhere.

Acknowledgments

The authors would like thank staff at VOSA for supplying the MOT data and providing documentation about the data source. They are also grateful for the research team of Sally Cairns, Eddie Wilson, Jillian Anable and Tim Chatterton for discussing their research and sharing their findings with us.

References

Department for Transport (2011) National Travel Survey 2010 http://www.dft.gov.uk/statistics/releases/national-travel-survey-2010/ (retrieved 1/11/2012).

Department for Transport (2012a) MOT testing data user guide v2, London. Available at http://data.dft.gov.uk/anonymised-mot-test/12-03/MOTTestingDataUserGuide_v2.doc (retrieved 1/11/2012).

Department for Transport (2012b) Cars registered for the first time by keepership (private and company) – table VEH0252. Available at http://www.dft.gov.uk/statistics/tables/veh0252/ (retrieved

1/11/2012).

Cairns S, Wilson RE, Chatterton T, Anable J, Notley S and McLeod F (forthcoming) Using MOT data to analyse travel behaviour change – scoping report *TRL PPR578*.

Mokhtarian PL and Cao X (2008) Examining the impacts of residential self-selection on travel behavior: A focus on methodologies *Transportation Research Part B* Vol 42 pp 204-228.

Sana B, Konduri KC, Pendyala RM (2010) Quantitative analysis of impacts of moving toward a vehicle mileage-based user fee *Transportation Research Record* **2187** pp 29-35.

Tao CC and Tsai CJ (2007) A digital mileage calculating program for electronic toll collection by using vehicle position system in *Proceedings of the Second International Conference on Space Information Technology* p 7951.

Wilson RE, Cairns S, Notley S, Anable J, Chatterton T, McLeod F (2013) Techniques for the inference of mileage rates from MOT data *Transportation Planning and Technology*.

Yang F, Yu L, Song GH, Wang LS (2004) Application of small sampling approach to estimating vehicle mileage accumulations for Beijing *Energy and Environmental Concerns 2004* pp 77-82, Transportation Research Record.

Biography

Rachel Moyce is a higher statistical officer at the Department for Transport, working on vehicles, MOTs, driving licence and accessibility statistics. Prior to this she has worked on local government finance statistics, the national travel survey, buses and taxis statistics. This is the first research she has contributed to.

Daryl Lloyd is the lead statistician for road safety statistics at the Department for Transport. Prior to this he has worked as a government statistician on vehicles, MOTs & driving licences; accessibility to key services; road traffic; and housing. He acts as the Department's transparency practitioner, has sat on the Government Statistical Service transparency sub-group and spends a significant amount of his time bullying colleagues into releasing more 'open data'. Before joining the civil service he was a research fellow at the UCL working on spatial patterns of surname distributions and modelling town centre boundaries.