Transportation Research Part E 60 (2013) 126–139

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

On the estimation of temporal mileage rates

R.E. Wilson [a,*], J. Anable [b], S. Cairns [c,d], T. Chatterton [e], S. Notley [c], J.D. Lees-Miller [a]

[a] *Transport and Mobility Modelling Group, University of Bristol, UK*
[b] *Centre for Transport Research, University of Aberdeen, UK*
[c] *TRL, Wokingham, Berkshire, UK*
[d] *Centre for Transport Studies, University College London, UK*
[e] *Air Quality Management Centre, University of the West of England, UK*

ARTICLE INFO

Article history:
Received 3 June 2013
Received in revised form 20 June 2013
Accepted 21 June 2013

Keywords:
Car use
Mileage rates
Mathematical methods

ABSTRACT

Mathematical and computational techniques are developed for the analysis of annual roadworthiness (MOT) test data that the UK Department for Transport has placed in the public domain. This paper develops a new theory to estimate fine-scale temporal (e.g., monthly) variations in vehicle mileage at a population level – derived from coarse-scale (e.g., annual) mileage data at an individual vehicle level. Numerical time-stepping schemes are derived from the theory and are tested on synthetic data to permit comparison with a known ground-truth mileage rate. Finally, we consider first steps in applying the methods directly to the MOT data set.

1. Introduction

Driven by a need to reduce emissions of greenhouse gases in order to mitigate climate change, there is an increasing policy interest in initiatives that reduce car use and encourage people to own cleaner, more efficient vehicles. These initiatives range from large-scale national projects, such as the promotion of electric vehicles through to small-scale projects, such as schemes that make it more attractive to walk or cycle in a given local area. But how can we tell whether these initiatives are effective, and if they are, how can we determine the scale of their effects? And how do the effects vary over time and with geography? (For example, between regions; between urban and rural areas; between different towns?)

To answer these questions, we require robust data concerning vehicle ownership and usage, and to date there have been two main sources. These are (i) surveys of individuals or households – which are potentially subject to bias due to misreporting (either deliberate or accidental) and who chooses to respond and (ii) on-street traffic counts, which are subject to bias in the manner in which counting sites are chosen – since an estimation procedure is required to convert spot counts into some measure of total vehicle kilometres travelled (Department for Transport, 2010a). Furthermore, automatic traffic counts cannot robustly disaggregate traffic over vehicle classes, or (for example) identify the fuel type or engine capacity of the counted vehicles.

Thus there is a demand for new estimation techniques which are either cheap, or perhaps even free – because they are based on secondary analysis of data that was originally collected for other purposes. In the United Kingdom, one such source of data originates from the annual roadworthiness test (known as the MOT test, since it used to stand for 'Ministry of Transport') which since 1967 has been compulsory for vehicles over three years old. In 2005, a computerised system was introduced for reporting the MOT test results and storing them in a Department for Transport (DfT) database. In November 2010,

* Corresponding author. Address: Department of Engineering Mathematics, Faculty of Engineering, University of Bristol, BRISTOL BS8 1TR, UK. Tel.: +44 1173315690.

E-mail address: RE.Wilson@bristol.ac.uk (R.E. Wilson).

1366-5545/$ - see front matter
http://dx.doi.org/10.1016/j.tre.2013.06.004

the DfT published this data (Department for Transport, 2010b) – consisting of the results of approximately 150 million MOT tests from 2005 to the Spring of 2010. Although some fields, such as vehicle registration plates and unique Vehicle Test Station (VTS) identities, have been withheld from the published data, what remains still contains a wealth of information that is not available elsewhere. In addition to the results of the test itself, the data include: the date of the test; the vehicle odometer (mileage) reading; the vehicle manufacturer, model and engine capacity; the vehicle's year of first use; and the top-level postal area (first letters only from the postcode) of the VTS.

An unadvertised feature of the published data is that an internal database index may be used to track many individual vehicles from year to year throughout their test history.[1] Consequently one may infer the mileage of a vehicle between a pair of tests on known dates. Our purpose is to analyse this data to provide an important missing link in the analysis of vehicle usage. Note that although the focus here is on the UK situation because of data availability, the methods that we develop are generic and could be applied internationally to any data set where the odometer readings of individual vehicles are monitored from time to time. (A situation where vehicles are monitored at regular intervals is directly addressable by our technique; irregular monitoring intervals would require further extensions to the theory.)

However, more interesting questions are dynamic (i.e., time-dependent) and concern how usage is changing with time. If these changes occur over rapid time-scales, the *straddling rate* is unable to capture them and the goal of this paper is to develop and demonstrate more sophisticated measures for calculating the temporal dependence of vehicle usage from the MOT data. As an over-arching challenge, we would like methods that might identify sharp changes in vehicle usage due to (for example) fluctuations in the price of fuel, or perhaps more severely, economic crises such as that resulting from the collapse of the banking system in Autumn 2008. Since data from automatic counters will give some measure of these effects, the potential added value in the MOT data is examining how these changes distribute themselves over vehicles of different ages and classes and the locations in which they reside.

At first sight, the challenge seems impossible, and the fact that it might potentially be viable is the major surprise and fresh contribution of this paper – to the best of our knowledge, this idea is in entirely virgin territory. The seemingly intractable problem is that we cannot know from MOT data how an individual vehicle's mileage is distributed between two consecutive tests (typically one year apart): most likely that distribution is highly non-uniform. For example, a typical vehicle may have a rather constant level of underpinning mileage due to (e.g.) regular commuting trips, that is interspersed with a heavy tail of occasional long distance trips, whose occurrence is essentially unmodellable. So how can we use this kind of annual mileage data to infer anything about mileages over time-scales of months?

The answer is that whilst a fine-scale temporal analysis is impossible at the level of individual vehicles, it *might be* possible at the population level. To this end, we shall introduce the concept of the population *spot rate*, which is a mileage rate common to all vehicles in some segment that we wish to analyse, and which captures in average terms how their usage is modulated in time. Clearly, individual vehicles have different levels of total usage, which we may model by assigning to them different (time-independent) proportional factors of the population spot rate. Further, individual vehicles have apparently random temporal fluctuations in their usage as we have described above. However, our assertion is that in some kind of population-averaged way, the spot rate modulates the usage of all vehicles in the given segment, and describes how that usage depends on time. The *spot rate* is formulated in mathematical terms in Section 3, which also develops a calculus that computes the *straddling rate* in terms of the *spot rate*, in the form of an integral equation.

However, since the straddling rate can be computed directly from the MOT data, the desired result is rather a formula that gives the spot rate in terms of the straddling rate, and this theory is the subject of Section 4. The output is a possible time-stepping scheme for evolving the spot rate forward in time. However, a drawback is that two years of initial data for the spot rate are required to start the scheme up. Hence Section 5 considers generalisations to the definition of the straddling rate, which are also directly computable from MOT data, but which have a crisper relationship with the spot rate, and which consequently might be more amenable to time-stepping schemes: a detailed presentation of the resulting schemes is then given in Section 6.

Section 7 then tests the proposed time-stepping schemes on synthetic data. The value in synthetic data is that it may be generated from a known (i.e., ground-truth) spot rate that we may then attempt to reconstruct. Because the synthetic data is 'idealised' (this means that certain approximations to the real-world data that are used in the development of the theory hold exactly), we may focus attention on a particular difficulty of the schemes which involves a trade-off between truncation error and statistical sampling error. In particular, although our method appears feasible for large enough data sets (which tend to reduce the problems due to sampling error), it requires the use of smoothing filters – a full investigation of which are beyond the scope of this paper.

Section 8 is concerned with the potential application of our method to the real-world MOT data (Department for Transport, 2010b). Because of the difficulties with sample sizes and smoothing techniques, this work is at a very early stage. In addition, due to the breakdown of simplifying assumptions used in the theory, various extensions are desirable that are discussed in Section 9, that also presents the conclusions.

[1] Since the initial submission of this paper, there has been a second release of the MOT data (in March 2012) that provides additional fields, including a unique vehicle identifier that links tests to individual vehicles. A further release is planned in Autumn 2012, which is expected to provide further enhancements to the data (DfT, personal correspondence).