Alternative Methodologies for Collecting Australian Motor Vehicle Use Statistics

Sheridan Roberts
Australian Bureau of Statistics

Peter Haines

Australian Bureau of Statistics

Abstract:

The Australian Bureau of Statistics plans to conduct its next Survey of Motor Vehicle Use (SMVU) in 1997. The main objective of the SMVU is to satisfy the information needs of Commonwealth and State government agencies responsible for the allocation of funds for road development, the design and construction of highways, the regulation of road transport operators and accident exposure and energy use analysis. It collects vehicle usage data across all major vehicle types (cars, trucks, buses, etc.) and States, thus providing the direct comparisons of patterns between vehicle types and States that are vital for the above purposes. It is the only source of such information and a time series has been established. The data are often used as proxies for road use and provide the broad context for measurement of growth rates in road use, mobility and traffic congestion. A methodological review was conducted during 1995 and 1996 to identify options for improving the SMVU's design before the next survey. Three methodologies were pilot tested and assessed for quality and cost effectiveness. These were Logbook, Pre-advice and Straight Recall methodologies.

Contact Author:

Sheridan Roberts Australian Bureau of Statistics PO Box 10 Belconnen ACT 2616

telephone:

(06) 252 6068

fax: (06) 251 5324

email:

sheridan.roberts@ABS.gov.au

Introduction

The Australian Bureau of Statistics (ABS) plans to conduct its next Survey of Motor Vehicle Use (SMVU) in 1997. The main objective of the SMVU is to satisfy the information needs of Commonwealth and State government agencies responsible for the allocation of funds for road development, the design and construction of highways, the regulation of road transport operators and accident exposure and energy use analysis. It collects vehicle usage data across all major vehicle types (cars, trucks, buses, etc.) and States, thus providing the direct comparisons of patterns between vehicle types and States that are vital for the above purposes. It is the only source of such information and a time series has been established. The data are often used as proxies for road use and provide the broad context for measurement of growth rates in road use, mobility and traffic congestion.

The SMVU has traditionally been conducted about once every three years as a sample survey, commencing in 1963. Respondents have been asked to provide information on the use of selected motor vehicles based on recall of their previous 12 months usage. The populations from which the samples have been drawn included all motorised vehicles (except tractors, plant and equipment, vehicles belonging to the defence services and vehicles with diplomatic or consular plates) registered for road use with a motor vehicle registration authority at a specified date.

However, there is concern that the traditional SMVU methodology results in high levels of non-sampling error. In particular, the 12 month recall methodology has been thought to result in overestimation of distance travelled. A 1988 test surveyed about 27,000 NSW registered cars and trucks. This survey suggested that respondents to the 1988 SMVU overestimated their reported total distance travelled by 11.9%, although this varied by vehicle type. The distance travelled by trucks was overestimated by less in percentage terms than it was for cars. This may be attributable to better record keeping practices for trucks. There are no measures of bias for more detailed data (e.g. distance travelled by area of operation) but, based on the 1988 test, it is likely that all distance travelled data were biased upward in 1988. However, there are reservations about the validity of the 1988 test, and some users have expressed a contrary view that the SMVU seems to understate usage

A methodological review was conducted during 1995 and 1996 to identify options for improving the SMVU's design before the next survey
tested and assessed for quality and cost effectiveness. These were Logbook, Pre-advice and Straight Recall methodologies.

Methods of investigation

User review

The first step in doing the methodological review was to identify clients' needs and their priorities. The requirements established for previous SMVUs were used as the starting point and supplemented with knowledge from ongoing liaison with clients and providers of statistics. A SMVU user review was also conducted in 1995 to discover previously unidentified needs, and to help set priorities.

Data items: The main data item needs can be classified into one of the following groups:

- Total distance travelled:
- Vehicle characteristics:
- Fuel use:
- Area of travel:
- · Loads carried;
- Purpose of travel;
- · Passengers carried;
- Driver characteristics; and
- Trailer characteristics

Survey frequency: many users maintain that a minimum three year SMVU frequency is required for the measurement of trends, because vehicle use is such a dynamic activity. For instance, recent years have seen larger and more efficient freight vehicles such as B-Doubles and B-Triples. It is also an activity subject to government policy review including the 1996 introduction of the National Heavy Vehicle Registration Scheme which is intended to provide consistent regulations and registration fees across Australia. Finally, vehicle use is considered to be strongly correlated to economic conditions.

The pilot tests

The second stage in the methodological review was to conduct a series of pilot tests to:

- identify differences in total distance travelled estimated by different methods;
- allow an assessment of relative data quality; and
- provide information on operational aspects of different methods.

Two alternative methodologies, Logbooks and Pre-advice, were pilot tested during 1995 and 1996. The traditional Straight Recall methodology was also trialed, for control purposes and to help bridge between the old and any new methodology. These methodologies and their component methods are described broadly, overpage.

The methodological options

Logbook methodology

The Logbook methodology employed three different methods to obtain all the necessary data: Trip Logbooks, Fuel Logbooks and Odometer Readings. Fleet Data Capture, discussed later in this paper, would also be incorporated into a Logbook methodology

Trip Logbooks: an initial Contact Information Form was sent to get basic information to help tailor the logbook to the usage characteristics of the vehicle Six different types of "trip" logbook were used. They were tailored by vehicle type - truck, bus and other (other included cars and motorcycles) - and by whether the vehicle was normal or high use. For example, taxis do lots of trips so they were sent a high use logbook that requested aggregate data for each shift, rather than details of each trip.

Trip Logbooks were designed to obtain detailed data for a short period, which would be expanded with annual or quarterly estimates of total distance travelled. They were initially tested over one and two week reporting periods. Two week reporting periods were found to be better; having lower error rates, a higher average number of drivers for trucks, and an equivalent daily (and weekly) average number of reported trips. It was thought that a decrease in the number of reported trips, especially toward the end of the logbook period, would result if respondent fatigue occurred. There was no evidence of such a decrease for either the one or two week reporting periods

All drivers of selected vehicles were asked to report details of every trip (or shift) they made in the nominated vehicle. The Logbooks obtained odometer readings at the start of the reference period and at the end of each trip (or shift) so we could derive accurate distance travelled for each trip (or shift). We also included a question asking for the suburb of destination for each trip. It was intended that a Geographic Information System (GIS) would be used to apportion travel within different geographic areas, such as urban and rural. Several other data items were obtained for each trip, varying depending on the vehicle type.

Because Logbooks only cover a small time period (e g two weeks), individual responses would not provide accurate estimates of total distance travelled over a year for that vehicle Two possibilities were considered to obtain estimates of total distance travelled using a Logbook methodology.

1) A different vehicle sample could be surveyed each fortnight throughout the year. This would allow estimates of total distance travelled to be produced for broad vehicle type categories by amalgamating the responses over the year. However, this approach would be expensive, logistically difficult and result in small sample sizes.

2) Alternatively, the year could be split into a few time strata. Only one fortnight would be selected for survey from each time strata, with different vehicle samples selected for each fortnight. Total distance travelled by vehicle type would be derived from the Odometer Readings method (see below), for a separate sample. These total distance travelled data would be applied to the detailed Trip Logbook data, by broad vehicle type, to expand the detailed information to annual estimates. However, it would require a good knowledge of seasonality to establish representative time strata.

Fuel Logbooks: these were designed to provide information on average fuel consumption by vehicle and fuel type, across all vehicle categories. They were based on Statistics Canada's Fuel Logbook, and involved the provider filling in details of their fuel purchases over a one month period. They also provided a measure of the total distance travelled during the month. This method was tested because Trip Logbooks can not readily provide fuel use information. Fuel Logbooks were also considered to be a potentially more accurate source of fuel use information than Straight Recall or Pre-advice. As for Trip Logbooks, a technique would need to be developed to expand short duration information to represent a year.

Odometer Readings: The Odometer Readings method involved sending a form out at the beginning of a period to obtain an odometer reading and vehicle characteristics information, and then sending another form out at the end of the period to obtain a second odometer reading. This was designed to provide accurate total distance travelled by vehicle characteristics (e.g. vehicle type and year of manufacture) data for all vehicle categories. These data are valuable in their own right, and would also be used to expand the detailed data obtained using Trip and Fuel Logbooks if time sampling of the logbooks were undertaken.

Because odometer readings are a key component of both the Logbook and Pre-advice methodologies, several trials were conducted. These trials covered three, six and 12 month periods; and included sub-trials involving interim telephone calls (e.g. half of the sample allocated for quarterly Odometer Readings tests received telephone calls after one and two months from the start of the period, to obtain interim odometer readings). These trials were designed to identify the best period and procedures for obtaining odometer readings, and to help bridge between an odometer based estimate of Total Distance Travelled and the traditional Straight Recall estimates. Each of the different time periods proved feasible, although the interim telephone calls proved onerous on ABS resources.

Fleet Data Capture

This is a potential adjunct both to the Logbook and Pre-advice methodologies. The main purpose of this method is to overcome difficulties in obtaining some data items for some vehicle categories using standard methods. Examples are taxi and urban bus data. Large bus, taxi and freight companies were approached to ascertain what electronic administrative records they possessed, and any other data that might be of use. The intention would be to get diskettes with information about all vehicles owned by such companies. That would reduce the reporting burden for those companies, while raising the effective sample size.

Pre-advice methodology

Pre-advice was designed to get data for all SMVU data categories, including fuel use and total distance travelled. It involved sending a letter at the beginning of a period, advising the vehicle owner of the vehicle's selection, the purpose of the survey, and the items that would be requested at the end of the period. The letter included suggestions on how to maintain useful records during the period. Provision of an optional worksheet to help compile the data during the period was considered but rejected. Draft worksheets tended to be very long trip logbooks, which providers were unlikely to use. It is intended to investigate further the possibility of a simplified worksheet, through a series of discussions and tests involving respondents

Because a letter was sent at the start of the period, the provider was also asked to return a form reporting their contact details and the vehicle's odometer reading. The odometer component provided a records based estimate of total distance travelled when combined with a second odometer reading obtained at the end of the period, thus reducing recall bias. The provider-estimated distance travelled provides an alternative measure of distance travelled which is especially useful for those vehicles with faulty odometers.

One and three month reporting periods were trialed for Pre-advice. The one month tests required dual despatches (there are despatches for future and past months at the start and end of each period) of forms to untrained respondents over a one month time period. They proved logistically too complex and were discontinued by the end of the first one month test. The three month tests proved manageable. For periods less frequent than quarterly it is unlikely that pre-advice would encourage record keeping or substantially reduce recall bias

Fleet Data Capture: a Fleet Data Capture element would be included as part of a Pre-advice methodology, just as it would for the Logbook methodology

Straight Recall

Straight Recall is the way SMVU information was collected from 1963 to 1995, inclusive. It involved sending a form out at the end of the reference period without prior notice, asking the person to record details from memory or any existing records. There were three form types, tailored by vehicle type (truck, bus and other). This was tested for control purposes, against which the other two methodologies were assessed, and to help bridge between the old and new methodologies

Bridging

Results obtained under a new methodology will not be directly comparable with the information from 1963 to 1995 obtained using Straight Recall, due to the reduction in reporting biases. The methodological review has included work to ensure that clients can bridge between the old and any new methodology. A second form was sent in September 1996 to a sub-sample of the selections for the September 1995 SMVU. This second form asked for an odometer reading and, based on recall, the distance travelled during the twelve months ended 30 September 1996. A separate sample was also sent a form in September 1996 and asked to recall distance travelled over the previous twelve months. This has allowed us to compare distance reported by recall and odometer readings over the same period. A set of bridging factors tailored to each major vehicle type will be published separately

A 1988 test of bias suggested that respondents to the 1988 SMVU overestimated their reported total distance travelled by 11.9%, although this varied by vehicle type. However, there have been reservations about the validity of the 1988 test, and some users have expressed a contrary view that the SMVU seems to understate usage. Preliminary results from the 1996 tests confirm the 1988 findings that the Straight Recall methodology results in overestimation of distance travelled for most vehicle types. However, distance travelled was underestimated for buses (not assessed in 1988)

Findings of the review

The comments in this section derive from three sources:-

- (a) Forms analysis, post enumeration studies and statistical analysis of the first two of the four quarters of trials. The third and fourth quarter trials focused on the bridging exercise, and helped establish the extent of bias in estimates of distance travelled derived from recall
- (b) A two day workshop comprising staff involved with the trials
- (c) Views of major users of the SMVU outputs, expressed at various Transport Statistics User Group meetings and during general liaison.

General

Response rates for all methods were generally poorer than would be expected in a "live" survey due to resource limitations and the large number of trials being conducted. Relatively little intensive follow-up was conducted. A comparison between the final 1995 SMVU response rates (78.4%) and those for the comparable Straight Recall tests (about 60%) may be indicative of the size of the difference between a live survey and the pilot tests.

Statistical analyses of the data from the first two quarters of pilot tests, to establish whether there were statistical differences in key variables between methods, generally identified no statistical differences. The sample sizes for the tests were small, and the findings were inconclusive rather than indicative of there being no difference. The results from subsequent larger scale "bridging tests" are still being analysed.

Logbook methodology

The Logbook methodology proved too complex with low response rates and a number of reporting problems, giving few useable responses. Considerable clerical effort would be required to correct and/or query the forms with respondents. The errors often related to misreported odometer readings on Irip and Fuel logbooks with the benefits of query action, weeks after the event, being doubtful. The findings in respect of the different logbook methods are detailed, below.

Trip Logbooks the main finding was that there was an unacceptably low number of useable responses due to a combination of low response/reporting rates and poor data quality. This was reinforced by detailed analyses of completed survey forms which identify the types and number of reporting errors. The number of useable responses probably could not be raised sufficiently even with extensive editing. The main problems with Trip Logbooks are listed, below

- They proved very expensive and would have been even more expensive if thorough editing had been performed. The vast amount of data (i.e. separate details for every trip) resulted in many errors and would require far more extensive editing than other methods. It is likely that even extensive clerical editing would be ineffective in raising the data quality to sufficient standards. Many providers would probably not be able to recall accurately the correct odometer reading several weeks after the event. Where the correct data were reasonably obvious from other information on the form, they were corrected clerically without contact with the provider.
- The high processing cost per form means that sample sizes would be extremely small
- Response rates were disappointing Several techniques were trialed to raise response rates, and this was the one method for which detailed follow-up action was pursued Resource intensive techniques such as telephoning all respondents prior to the start of the reporting period raised response rates to around 70% but, unlike other methods, we probably could not achieve much of an improvement beyond this level. Response rates of 50% to 60% were common when only mail based prompts and reminders were used.
- They required significant detail to be recorded of individual trips over a two week period, which was very burdensome on providers.
- For high use shift logbooks, the geographic component was excluded and the travel was assumed to be in the local area. While this is likely mainly to hold true, it will have limitations for State/Territory of operation information in border areas such as Albury/Wodonga and the ACT "region".

Fuel Logbooks: this proved to be an expensive method for collecting only one data item, and produced poor quality data. The main problems are outlined, below.

• The response rate was disappointing at 47%, but this could probably be improved using some of the intensive follow-up techniques trialed for Trip Logbooks

- The quality of average fuel consumption data was poor for "low use" vehicles At least two fuel fills were required during the month to generate reliable information, which means that this method cannot produce accurate estimates of fuel use for low use vehicles. The low use vehicles were often older vehicles that travelled less than new vehicles so this flaw could result in some bias.
- Fuel Logbooks suffered from the same problem as Trip Logbooks with errors in an early odometer reading reducing the usefulness of subsequent information.

Odometer Readings: This method was relatively cheap and could support a large sample size, providing accurate total distance travelled data with low sampling errors. However, it does have some limitations

- While providing a key statistic, it does not provide the detailed information required of a SMVU. The Odometer Reading method is not sufficient in its own right and needs to be combined with other methods to form a comprehensive, integrated methodology
- About 10% of odometers were reported as not working. Faulty odometers appeared to be more prevalent among older vehicles, which also tended to travel shorter distances than newer vehicles. This potential bias can be overcome by requesting an estimated total distance travelled on the second odometer form.
- There was some evidence, both from forms analyses of errors and statistical analyses of data, that respondents did not all understand the term "odometer". Instructions were provided and tested to clarify the meaning, but were not entirely successful. A pictorial representation of an odometer on the form may help clarify the meaning. Further forms testing is proposed to refine the form design.

Fleet Data Capture

Fleet Data Capture was found not to be feasible for taxi companies because they do not maintain suitable electronic records for sufficient periods of time. There were also not considered to be significant gains, either for the respondents or the ABS, in obtaining electronic records from companies with large trucking fleets. Road freight is predominated by small operators and most companies do not maintain electronic records of the information that we require.

However, Fleet Data Capture was found to have the potential to provide larger effective sample sizes for urban buses, and reduce respondent load significantly for them. These companies typically have a large number of buses selected in each survey. Several urban bus companies reported good electronic records that we expect to be able to access. This typically included passenger boardings but not passenger kilometres. However, some

companies have results from their own surveys that estimate average distance travelled for passengers. This could be applied to passenger boardings to generate passenger-kilometres for those companies. Discussions have indicated that these companies can provide data readily in a range of formats.

We propose to start with those urban bus companies for which data provision is relatively straightforward and where there are clear benefits both to the company and the ABS In the longer term, we intend to extend Fleet Data Capture to all urban bus companies, and possibly to taxi and truck fleets as electronic record keeping improves in the future This will require ongoing liaison and research.

Pre-advice methodology

Pre-advice, combined with odometer readings, is the intended new methodology. Odometer reading data will be used to adjust the reported total and component distances travelled. The main advantages of Pre-advice are listed, below

- Pre-advice is simpler and cheaper than Trip Logbooks.
- It can provide all data items where Odometer Readings alone cannot
- Useful estimates of total distance travelled can be obtained if only the end of period form is returned, which is not the case for a pure Odometer Readings approach. However, if the first Pre-advice form is not returned it might imply that the provider never read the instructions, and so did not maintain records. Our test sample sizes were too small to indicate whether or not there were statistically significant differences between those who returned both forms and those who only returned the second form. We would adjust the reported distance travelled with an odometer based bias factor.
- It has the further advantage over pure Odometer Readings that estimates of total distance travelled can be compiled for older vehicles with faulty odometers, by imputing based on reported distance. This provides the best of both worlds: the more accurate odometer readings to derive distance travelled for most of the sample, and a separate estimate to overcome the problem of odometers that are faulty or misreported.
- There was evidence that data quality is significantly better than for Straight Recall Forms analyses of errors indicated that there were far fewer errors of omission (where respondents failed to report an item that should have been reported) for passenger vehicles with Pre-advice than Straight Recall. Those items that are most difficult to report are either left blank or guessed, so the fewer omissions might also imply that the quality of the reported data was higher for Pre-advice Straight Recall was only tested for passenger vehicles so comparisons for other vehicle types were not possible.

• The trials indicated that provider load was comparable for Pre-advice and Straight Recall. The two Pre-advice forms, combined, took respondents an average of 45 minutes to complete. The one Straight Recall form took 43 minutes. This does not take into account the records that respondents maintained for the Pre-advice during the period.

Both one and three month periods were trialed for Pre-advice. Periods longer than three months were considered likely to reduce the effectiveness of the pre-advice

- The logistics of the one month approach proved insurmountable. The dual despatches (there are despatches for future and past months at the start and end of each period) also increased the costs of Pre-advice compared with Straight Recall.
- The costs and logistics of a three month Pre-advice period were manageable

The estimate of the total distance travelled based on the reported distance was not significantly different from the Straight Recall estimate. Also, the estimate of the total distance travelled based on the two odometer readings was not significantly different from the Straight Recall estimate. The small sample sizes render this finding inconclusive.

Survey frame

The poor quality of motor registry data poses problems, as does the volatility of vehicle ownership details. About 40% of name and address details change each year. This adversely affects response rates and follow up action as the sampling frame becomes older. It is essential that a frame be established each year (assuming an ongoing quarterly design), with sub-annual (probably once more per year) updates of names and addresses for selected units. The trials attempted to obtain quarterly updates to name and address details and this proved too frequent, both for ABS and some motor registries, to be able to manage.

A major potential improvement in our frame is a national scheme being implemented by the NSW Roads and Traffic Authority. This will be a centralised national repository of motor registration records called NEVDIS (National Exchange of Vehicle and Driver Information System). This has the potential to provide all of our frame requirements from a single source (instead of the current 12 different sources), and in a single consistent format. It is also likely to result in general improvements in the underlying quality of the frame data as the various motor registries (presumably) systematically compare notes. This scheme is expected to be implemented during the second half of 1998. Combined with a direct "read only" EDI link from NEVDIS to the ABS, it has the potential to reduce costs of frame creation and maintenance substantially. We will continue to monitor developments in this area, and maintain liaison.

Recommendations

Methodology

The main conclusion is that a quarterly Pre-advice methodology should be implemented. This should incorporate an odometer component which would be used to derive total distance travelled and to impute other data (e.g. State of operation) based on percentage splits reported by providers The reasons for this conclusion are listed, below.

- Forms analyses of errors indicate that it has the best quality and highest useable reporting rates of all methods.
- All required data items can potentially be obtained using Pre-advice, giving it the important advantage of simplicity
- Pre-advice is also more robust than a pure Odometer Reading method for measuring total distance travelled.
- It is substantially cheaper per unit than Trip Logbooks.
- Fuel Logbooks are not considered to be a viable method to obtain reliable fuel use information, and Pre-advice is recommended for this data item, although further forms design work needs to be undertaken

It is intended that the SMVU become an ongoing quarterly survey, producing broad annual estimates. This would entail a different set of selections each quarter reporting only once for the year. A regular biennial or triennial publication containing detailed estimates from pooled data is also proposed. The exact periodicity and nature of the publication is subject to further user consultation. It is intended to implement this methodology in respect of the "quarter" commencing 1 August 1997. It is desirable to lag one month behind traditional quarters to minimise the logistical complications for despatch and reminder action during the Christmas/New Year period.

Further forms design work is required to refine the Pre-advice forms. This needs to include testing of pictorial representations of odometers; the best ways to assist respondents at the start of the period (including refinement of the covering letter and investigation of a workbook); and an assessment of the viability of each data item (especially those found to be problematic across all methodologies).

Fleet Data Capture is to be pursued for those urban bus companies identified as having good electronic records. An ongoing program to monitor trends among other urban bus companies, and taxi and truck fleets, needs to be implemented with a view to extending

Fleet Data Capture in the future.

Sample size and frequency

A significant problem with the traditional SMVU approach was the infrequent nature of the survey (it has been conducted about triennially since 1976) It was considered that a more frequent (e.g. ongoing or biennial) survey would provide a better time series of key information, thus removing the need for users to extrapolate results for non-survey years. Importantly, it would also allow maintenance of ABS subject matter expertise and facilitate the evolution of systems and procedures The principles of continuous quality improvement can not as readily be applied to a triennial survey as to a more frequent SMVU. A major disadvantage of a more regular survey is that the funding available for the triennial surveys would be spread more thinly, particularly as there is no extra budget available The consequence is that sample sizes would need to be lower resulting in higher sampling errors

Three options for a more frequent survey were discussed with major users. These were:

1) a biennial survey of all vehicle types;

2) an ongoing quarterly survey with freight vehicles surveyed in the first year, non-freight carrying vehicles surveyed in the second year and broad comparative information obtained for all vehicle types in the third year; and

3) an ongoing quarterly survey of all vehicle types.

Users expressed strong support for an ongoing survey of all vehicle types. This would produce an annual publication of broad aggregates, and a more detailed publication less frequently based on pooled data over two or three years.

The tentative estimate of the ongoing quarterly sample size is 5,000 vehicles. Implementation will allow sufficient resources for a continuous quality improvement program While some improvements will have an ongoing component (e.g. refinements to systems), more comprehensive and structured reviews of key elements will help the survey remain relevant and efficient

Data items

Several data items proved troublesome to collect reliably, irrespective of the method employed to obtain them Tare weight will be discontinued for all vehicle types except freight carrying trucks, which is subject to further forms design work. Gross Vehicle Mass and Gross Combination Mass are also subject to further forms design work Engine Capacity and Number of Cylinders have been included as part of general forms design testing, but will be discontinued if significant improvements in quality are not found Driver information for taxis and buses will be dropped, as will the two questions on the age that drivers were licensed for trucks. Passenger information will be discontinued for all vehicles types Further form design work is required to improve the responses for Purpose of Iravel and Other Urban area of operation. The quality of fuel use information needs further review, in particular an assessment of how well a Pre-advice methodology estimates total distance travelled (the bridging tests combined with specific pre-advice trial data should help) and average fuel use (post enumeration studies are recommended) The various freight items require further analysis to ascertain the extent of any problems and should be included in a general forms design and post enumeration exercise.

Survey frame

It is intended that a frame be established each year, with sub-annual updates of names and addresses for selected units Developments with NEVDIS need to be monitored because it has the potential to reduce costs and time in generating the frame, and increase its quality