RESEARCH ISSUES IN SWEDISH ROAD TRAFFIC SURVEYS

Gösta Forsman, Swedish National Road Administration and Linköping University, Sweden
Stig Danielsson and Annica Isaksson, Linköping University, Sweden

KEY WORDS: Road traffic surveys, Multi-stage sampling, Nonsampling error

1. Introduction

The Swedish National Road Administration (SNRA) administers the construction and maintenance of Swedish national roads. In addition, the SNRA has an overall responsibility for road traffic safety in Sweden. These tasks are supported by a system of road traffic surveys, here defined as surveys based on data collection from pre-selected sites on the roads.

One part of the system is the so-called Traffic Measurement System, which includes the surveys of annual average daily traffic (AADT) and change of total vehicle mileage. Another part of the system is directly connected to road traffic safety and includes a survey of vehicle speed. This survey is based on a master sample of roads, designed for a broad use of traffic surveys. The system of road traffic surveys also encompasses surveys of vehicle acceleration and truck weight-in-motion. These surveys are still in their early stages of development.

The equipment used to collect data comprises pneumatic tubes stretched across the road and connected to a traffic analyzer. When a wheel of a passing vehicle crosses a tube, this action gives rise to a pulse in the equipment. From these pulses, vehicles are identified and information such as their speeds can be measured. For some surveys, additional equipment is needed for the data collection.

This paper presents information on the SNRA traffic surveys and discusses research activities including sampling and estimation issues, nonsampling errors, and the use of model-based inference from the surveys.

2. The Traffic Measurement System

2.1 The Survey of Annual Average Daily Traffic

The SNRA estimates yearly traffic volume at each of about 22,000 count sites on the national road network. Each count site represents a road link with supposedly homogeneous traffic volume. For each count site, the AADT is estimated, based on short-period traffic counts of lengths varying from 24 hours to three days. The high-volume part of the road network is covered over a measurement cycle of four years and the low-volume part over a cycle of eight years.

Traffic data typically have strong systematic variations within a year, with similar patterns on similar types of roads. Therefore, data from permanent count sites are used as auxiliary information in the AADT estimates. This method, known as the traditional factor approach (Aldrin, 1997; Sharma et al., 1996) can be described as a two-step procedure. Step 1 involves calculations on data from permanent count sites; step 2 involves calculations on short-period data for the actual count site for which the AADT estimate is needed.

Step 1. The permanent count sites with complete data are divided into some 40 groups with similar traffic patterns. For each group, a factor curve is calculated that describes typical variations within a year. Different factor curves are used for weekdays and weekend-days.

Step 2. When estimating AADT for a site with sample counts, the site is first assigned to one of these groups. Then the appropriate factor curve is applied to the sample counts to produce an estimate of AADT.

Step 1 is performed about every 5 to10 years (thus not each time the AADT is estimated).

2.1.1 Estimation

The observed flow f_t on a count site during selected period t is divided by its corresponding value on the factor curve, I_t, to adjust the flow for the expected multiplicative deviation. The estimator used for the true AADT, here denoted F, is

$$\hat{F} = \frac{1}{n}\sum_{t=1}^{n} c_t \frac{f_t}{I_t} \quad (1)$$

where n is the number of count periods—typically four—and c_t is a constant depending on the number of weekend-days and weekdays of the particular year and the length of count period t.

The number of time periods for which traffic flows actually are measured in a point on the road is very small, which makes estimation of the variance of $\hat{F}$ a difficult task. A cumbersome and theoretically weak method, based on repeated samples from the permanent site data, is used. Moreover, this method considerably overestimates the variance of $\hat{F}$.

2.1.2 Research issues in the AADT survey

To achieve improved variance estimates, a model-based approach has been suggested (Danielsson and Larsson, 2003). According to the model, the observed traffic flow f_t is a Poisson-distributed random variable with expected value $\lambda_t F$. The parameter λ_t follows a distribution with expected value equal to the corresponding value at the factor curve, I_t.

Using different assumptions on the λ_t-distribution, Danielsson and Larsson develop estimates of F and $Var(\hat{F})$. An important finding is the structure of the variance under the gamma distribution, when

$$Var(\hat{F}) = aF + bF^2 \qquad (2)$$

where a and b are constants. In (2), the first term reflects the variability of f_t given I_t and the second term reflects the variability of I_t. In estimating the variance, it is only the coefficient b that has to be estimated (besides F), since a is determined from the factor curve. Different estimating methods for b, all based on observational data from the permanent sites, have been examined. The research is still in progress, and definitive conclusions remain to be drawn.

2.2 Change of Total Vehicle Mileage

Since 1976, the SNRA has estimated the monthly and yearly changes of total vehicle mileage on the national road network. The estimation procedures have been modified a few times. Here we restrict our attention, however, to the procedure presently in use.

As in the AADT survey, the road network is divided into road links, in such a way that within each link the flow is constant. The links are stratified according to seven regions and four road types. The total sample size is 83, and the links are Neyman-allocated (following the principle of Neyman, 1934) over the 28 strata. Within each stratum, the links are further stratified according to traffic type. The number of sub-strata is equal to the sample size, and the total vehicle mileage in each sub-stratum is approximately alike. In this way, the sample size equals one in each sub-stratum, and links are drawn with probability proportional to size (pps) with probabilities proportional to links' total vehicle mileage.

The data collection is conducted at permanent count sites.

2.2.1 Estimation

Vehicle mileage change can be estimated in several ways, but here we confine ourselves to the simplest case. Let Y and X be the unknown total mileage in period 1 and 2, respectively. The parameter to estimate is $R = X / Y$.

In sub-stratum *h,* a natural estimator is $\hat{R}_h = X_{hi} / Y_{hi}$, where Y_{hi} and X_{hi} are the flows in period 1 and 2, respectively, for the selected link. These estimates have to be weighted to a total estimator

$$\hat{R} = \sum w_h \hat{R}_h \qquad (3)$$

and the natural, but perhaps not the best, weights are simply the pps-probabilities.

Since the sample size is one in each sub-stratum, there is no possibility to construct an unbiased estimator of the variance of $\hat{R}$. Instead, we utilize the idea of collapsed strata (Cochran, 1977) to overestimate the variance.

2.2.2 Sample revision

Since the road network, as well as the link total vehicle mileage, is continuously changing, the sample of links should be renewed regularly. From a practical and economical point of view, however, it is advantageous to retain the sample for a longer time. A compromise between these two perspectives has recently been created by an elegant application of permanent random numbers (PRN) (see Ohlsson and Jacobsen, 1995).

According to the PRN method, each link is assigned a random number, which is transformed to a random observation from an exponential distribution with mean value equal to the reciprocal of the mileage. All calculated observations within a sub-stratum are ordered, and the link with the smallest observation is chosen. It is straightforward to prove that the link is chosen with probability proportional to its total vehicle mileage.

Besides giving the correct inclusion probability, the PRN method also has the ability to preserve many sampled links in a later revision. New links are assigned new PRNs, whereas old links retain their PRNs. As previously described, exponential distributed observations are calculated with actual mileage, and then ordered. Of course, there is a good chance that the link with the smallest observation in the revision is the same as the earlier one.

2.2.3 Research issues

The survey measuring changes of total vehicle mileage is one of the most important for SNRA and is therefore currently under evaluation and improvement. In the near future the most challenging issues will be:

- Improvements of the collapsed strata technique for variance estimation

- Studies of the effects of replacing the sub-stratification method with random pps-sampling within strata
- Evaluation of the recently implemented PRN method, perhaps with help of simulations
- Studies of alternative PRN techniques, allowing sample sizes greater than one

3. The SNRA Master Frame of Urban Areas

3.1 Construction of the Frame

During 1995–96 a master frame for urban areas was constructed by the SNRA. The frame was mainly intended for surveys of various aspects of traffic safety (vehicle speed, seat belt use, bicycle helmet use, etc.), but has later proved to be usable also for surveying the conditions of various types of road equipment (road signs, road fences, and lamp posts). Other fields of application may be brought to the fore in the future.

The method of using a master frame is described, for instance, by Kish (1965, pp. 478-480). In brief, a “master sample” of sampling units is selected. For each sampled unit, a frame is prepared. The sample for a particular survey is then selected from these frames, which serve for a longer time period.

The SNRA master sample was selected in two stages (with stratification in each stage.) The sampling units in the first stage are population centers. These, in turn, are divided into small geographical areas, corresponding roughly to the so-called “small area market statistics” regions administered by Statistics Sweden. The small areas serve as sampling units in the second sampling stage. Within strata, the sampling in each stage was conducted as follows:

- In **stage 1**, population centers were drawn by probability-proportional-to-size sampling with replacement, with probability proportional to the number of inhabitants.
- In **stage 2**, for every population center drawing, small areas were selected by simple random sampling.

For each selected small area, frames of its roads, cycle tracks, signal-controlled intersections, etc. were prepared. From these frames, survey-specific samples of observation sites were drawn according to a survey-specific sampling design.

3.2 The Vehicle Speed Survey

An important cause of traffic fatalities in Sweden is speeding. Large resources are therefore spent to improve observance of the speed limits. To learn whether these efforts pay off, every summer since 1996, the SNRA has conducted the Vehicle Speed Survey. The survey is conducted on both national and urban roads; here we restrict our attention to the urban part of the survey. From the roads in the small areas included in the SNRA master sample, road sites (that is, one-meter sections of the roads) are selected for observation. This selection represents **stage 3** in the full sampling design. Each road site is observed over a random 24-hour period. From the collected data, various parameters are estimated. Parameters of interest include total vehicle mileage t_y, total travel time t_z, and, especially, average speed $R = t_y/t_z$. The estimator used for average speed is

$$\hat{R} = \frac{\hat{t}_y}{\hat{t}_z} \tag{4}$$

where $\hat{t}_a$ is an unbiased estimator of the total t_a for variable a (which may be y or z)—that is, unbiased with respect to the sampling design. Basically, the estimator $\hat{t}_a$ is a weighted sum of estimated small area totals:

$$\hat{t}_a = \sum \sum w_{iq} \hat{t}_{aiq} \tag{5}$$

where summation is made over all population center drawings i and selected small areas q. The weights w_{iq} are determined by the drawing probabilities of selected population centers, and by the inclusion probabilities of selected small areas. The estimator of the a-total for small area (i,q), t_{aiq}, is given by

$$\hat{t}_{aiq} = N_{iq} a_k \tag{6}$$

where N_{iq} is the total length of all roads in area (i,q), and a_k is the study variable value for the observed road site.

3.2.1 Research Issues in the Vehicle Speed Survey

The design and performance of the urban part of the speed survey has been evaluated in a recent doctoral thesis by Isaksson (2003), who investigated the uncertainties due to imperfect sampling frames and missing data on the survey estimators. Consider the estimator of t_{aiq} presented in equation (6). First, in practice, the true road length N_{iq} is not known, but is substituted by the road length according to the sampling frame N_{Fiq}. The sampling frame was constructed in 1996 by manual measurements of road lengths on city maps and is known to contain errors. Thus, N_{Fiq} may differ from N_{iq}. Second, the measurement equipment typically fails to observe all vehicles passing a selected road site. Then, the true study variable value a_k must be replaced by some estimator $\hat{a}_k$ (the present estimator arising from simply ignoring any missing data.)

In the thesis, error models for N_{Fiq} and $\hat{a}_k$ are formulated (and evaluated from experimental data). The expectation and approximate variance of $\hat{R}$, taken with respect jointly to the sampling design and the error models, is derived. In this way, the isolated impact of each type of error is analyzed, as well as the total error arising from all error sources. It is concluded that the frame error under consideration probably does not bias the estimator of R, and that it only implies a minor increase of its variance. It remains unclear whether the estimator needs to be adjusted for missing data. The possibility of reducing the sampling error by reallocating the total sample over sampling stages is also investigated in the thesis. More precisely, the components of the total variance of $\hat{R}$ arising from each sampling stage are estimated and their relative sizes analyzed.

At present, within strata, the number of population center drawings in stage 1 is 10, whereas one small area and one road site is selected in stages 2 and 3, respectively. Estimation of the total variance of $\hat{R}$ is made possible since units are drawn with replacement in stage 1. Due to the small sample sizes in subsequent stages, the variance components arising from each sampling stage can however not be estimated. In the thesis, this problem is circumvented by use of conditional estimates (conditional on the stage 1 sample) and some experimental data. It is concluded that for unchanged total sample size, the precision of $\hat{R}$ is likely to improve if more road sites are sampled in stage 3, and the number of population center drawings in stage 1 decreased correspondingly (the sample size in stage 2 is already at minimum). This encouraging result, however, remains to be confirmed by a full-scale experiment.

4. Other Road Traffic Surveys

4.1 A Pilot Survey on Vehicle Acceleration

In 1999, the SNRA conducted a pilot survey for estimating vehicle accelerations in signal-controlled intersections (SNRA, 2000). The measurements were accomplished at sites selected from the SNRA master frame of urban areas; the resulting estimates of mean accelerations were reasonable and had acceptable precision.

The survey was possible due to the research about measuring accelerations with a traffic analyzer, reported in a licentiate thesis (Walter, 2001). In the thesis, a thorough validation of different acceleration measurements is carried out. The parameter "mean acceleration," A, is defined as the quotient between total change of speed and total travel time. It is possible to unbiasedly estimate the numerator and the denominator, and in the simple case with data from only one site, the estimator is

$$\hat{A} = \frac{\sum (a_i / v_i)}{\sum (1 / v_i)} \tag{7}$$

where v_i and a_i are the speed and acceleration, respectively, for vehicle i.

In the thesis, acceleration profiles of all traffic close to intersections were studied. The profiles seem to be rather stable, and could be used to improve the estimator of A. However, much more research is needed in order to identify suitable estimation methods.

4.2 Truck Weight-in-Motion

A difficult but highly needed task is the development of an information system for truck weights on road links. Such information would have a strong economic impact on road construction and maintenance.

A major problem in the development of such a system is the lack of reliable equipment for weighing trucks in motion. During 2002 and 2003, the SNRA evaluated a weighing system, developed in Slovenia, which uses bridges as weight-in-motion locations. The results from this evaluation are promising and steps are now taken towards a system that provides estimates of road link parameters such as total yearly tonnage and total overload.

By necessity, estimates from such systems will be model-based since bridges are not randomly located on a road network. In addition, for cost reasons, the number of measurement spots will be very small.

The research is in progress and we will here only mention the basic steps of the model. Critical for the model is a stratification of the road network of study. The auxiliary information used for stratification is taken from (1) the Commodity Flow Survey conducted by Statistics Sweden, (2) AADT data, and (3) weight data from a number of bridges. Within strata, truck flows and weight distributions are assumed approximately constant for all road links. A small number of bridges are selected as weight-in-motion locations in each stratum and annual average truck weights, annual average overload weights, etc., are estimated from measurements on these bridges.

Under the model, the total annual tonnage, T, of the road network of study is estimated by

$$\hat{T} = \sum_{h=1}^{H} \hat{\mu}_h (VM)_h \tag{8}$$

where $\hat{\mu}_h$ is the estimated annual average truck weight for stratum h and $(VM)_h$ is the total truck vehicle mileage in stratum h. Other parameters of interest, such as total overload, are estimated analogously.

5. Acknowledgements

Support from the Swedish National Road Administration and the Bank of Sweden Tercentenary Foundation (Grant no. 2000-5063) is gratefully acknowledged.

6. References

Aldrin (1997), “A New Method for Traffic Volume Estimation Based on Counts – Calibrated to Short-period Counts from Oslo,” Technical report STAT/07/97, Norwegian Computing Center, Oslo, Norway. (In Norwegian)

Cochran, W.G. (1977), *Sampling Techniques* (3rd ed.), New York: Wiley.

Danielsson, S., and Larsson, A. (2003), “Estimation of Traffic Flows,” unpublished manuscript, Linköping University, Sweden.

Isaksson, A. (2003), “Survey Models for a Vehicle Speed Survey,” Linköping Studies in Statistics No. 2 (doctoral thesis), Department of Mathematics, Linköping University, Sweden.

Kish, L. (1965), *Survey Sampling,* New York: Wiley.

Neyman, J. (1934), “On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection,” *Journal of the Royal Statistical Society*, 97, 558-625.

Ohlsson, E., and Jacobsen, J. (1995), “Revidering av Vägverkets basurval för statligt vägnät inför undersökningsåret 1996,” unpublished manuscript, Stockholm University, Sweden. (In Swedish)

Sharma, S.C., Gulati, B.M., and Rizak, S.N. (1996), “Statewide Traffic Volume Studies and Precision of AADT Estimates,” *Journal of Transportation Engineering,* 122, 430-439.

Swedish National Road Administration (2000), “Undersökning av accelerationer och retardationer 1999 – resultat - och metodrapport,” Publication 2000:19. (In Swedish)

Walter, L. (2001), “Estimating the Speed and Acceleration in Actual Road Traffic by Spot Measurements,” unpublished thesis no. FiF-a 48, Linköping University, Sweden.