

## FOR-HIRE TRUCKING SURVEY:

## SURVEY DESIGN

R. Lussier<sup>1</sup>

The methodology of the For-hire Trucking Survey is discussed in this paper. This survey provides good examples of administrative and operational constraints faced by survey statisticians and field data collection teams.

## 1. INTRODUCTION

This paper presents the methodology of the For-hire Trucking Survey, a multi-stage probability survey of shipping documents retained by for-hire trucking carriers in Canada. The paper is structured as follows. In the second section, the survey context is described. In the third section, the ultimate sampling unit, namely the shipment is defined. Then the universe and the frame of the survey are depicted in the fourth section. After that, some major administrative considerations are listed in the fifth section and the stratification and sample allocation are covered in the sixth section. Then the first stage sample design and the subsequent stage(s) sample design are presented respectively in the seventh and eighth sections. In the ninth section, the field operations are discussed. The data processing and the estimation methods respectively are explained in the tenth and eleventh sections. Finally, a comment about the future of the survey concludes the paper in the twelfth section.

---

<sup>1</sup> R. Lussier, Business Survey Methods Division, Statistics Canada.

## 2. SURVEY CONTEXT

### 2.1 Primary Objective and Uses of the Survey

The primary objective of this survey is to provide information about the domestic intercity movements of goods by the Canadian for-hire trucking industry. The for-hire trucking industry covers any carrier which for compensation undertakes the transport of goods by truck. This survey measures the output of this industry in terms of revenues earned, tons carried and ton-miles performed by commodity group.

Requests for estimates from this survey come from a wide variety of sources such as government departments concerned with trade; transport regulatory officials at both federal and provincial levels; carriers; university consultants; industry associations; and many other organizations and individuals who share a common interest in transportation.

The estimates are used extensively to serve four basic requirements. First, they measure the volume of domestic trade transported by intercity for-hire carriers provincially and interprovincially. Secondly, they provide a cross-check on the rate of industrial growth reflected by intercity commodity movements and they provide information on regional development. Thirdly, they assist in transportation studies (e.g. [1], [2] and [3]). Finally, they support the presentation of briefs, submissions and other inquiries to regulatory authorities and commissions.

### 2.2 Background

Initial work on the For-hire Trucking Survey began in 1969. At that time, a study of various methods of collecting commodity origin and destination statistics was carried out. The study results showed that, from a cost-benefit point of view, a sample survey of shipping documents was the only viable approach to collect the data.

In 1970, a pilot survey was conducted to assess the survey approach effectiveness. The pilot survey involved the examination of the shipping documents of about 200 for-hire trucking firms throughout the country. The favourable response to the pilot survey and the availability of origin, destination, commodity, weight and revenue information of the shipping records indicated that the survey approach was feasible.

The For-hire Trucking Survey has been conducted on an annual basis since 1970 by the Transportation and Communications Division of Statistics Canada. However, the survey design has changed over time. Examples of changes mentioned in this paper are changes to the frame and changes to the sample allocation technique.

### 3. THE ULTIMATE SAMPLING UNIT

The 1969 study and the 1970 pilot survey mentioned in Section 2 indicated that the ultimate sampling unit be the shipment.

The principal characteristics needed from each sampled shipment are the true origin and the final destination; the description of the commodity(ies) carried; the weight; the transportation revenue earned and the interlined shipment information. An interlined shipment occurs when a consignment is moved by a carrier to an intermediate point and then moved by another carrier to another point. The interlined shipment information is used to unduplicate interlined shipments.

The secondary characteristics needed are the month and year of shipment; the quantity of commodity and the unit of measurement (e.g. 5 board feet, 20 gallons, 15 sacks, etc.); the method of movement (e.g. heated van, refrigerated van, piggyback, fishyback, container, etc.); some remarks on shipment weight transcribed (e.g. minimum weight, convenient weight used for calculating revenue, etc.); the rate charged and the rate

condition codes (e.g. code indicating where rate is minimum, per 100 lb., per hour, etc.); and the revenue condition codes (e.g. code indicating where exact transportation revenue is not available, where shipment is out of scope, etc.).

The shipment characteristics can be found on documents known in the trucking industry as Probills, Bills of Lading, Load Manifests, Trip Reports, Invoices, or a combination of the above in either a computer listing format or other media of storage.

These documents can be filed in complete numeric sequence; in broken numeric sequence; in chronological order; in alphabetical order (e.g. by customer name); by terminal; by commodity type, i.e. usually contracts; or in no order at all. The documents may even be cross-filed; for example, by serial number and by customer's name. Within a filing system, documents may be kept in a set of file drawers, in a set of binders or shannon files, on shelves, in drawers, or even in a book.

#### 4. UNIVERSE AND FRAME

The first choice for the frame is ideally a list of all shipments. However, such a list is not available. Instead, D.S.L.P.'s (Document Storage Location Points) are used as natural clusters of shipments for the first stage sampling units of the design. A D.S.L.P. is a site at which shipping documents suitable for sampling are kept. A trucking company may have more than one D.S.L.P. in the case where shipping documents are stored at several terminals but not at the company head office. In some cases, a D.S.L.P. contains shipping documents for more than one company.

The universe and the frame of D.S.L.P.'s have changed since the survey started in 1970.

In 1970, the universe was defined to include the D.S.L.P.'s of all provincially regulated or licensed carriers, regardless of size, type or major activity. The frame for this universe was derived from provincial license lists and included about 15,000 D.S.L.P.'s.

Since 1975, the universe has been limited to the D.S.L.P.'s of the carriers earning \$100,000 or more annually from intercity trucking. Other exclusions are mentioned in the survey publication [4]. Also, the frame for a given year has been the list of the D.S.L.P.'s of the carriers whose reports to the Motor Carriers Freight and Household Goods Movers Survey for the previous year show earnings of \$100,000 or more from the domestic movement of goods over more than 15 miles of public roads.

The description of the frame of the 1978 survey and of the principal statistics estimated from the survey data are given in Table 1 at the end of the paper.

## 5. ADMINISTRATIVE RESTRICTIONS

The survey design consists of selecting shipments from the files of selected D.S.L.P.'s and of transcribing the characteristics of the selected shipments on coding sheets. However, administrative restrictions limit the number of transcriptions. The administrative restrictions are presented in this section and the survey design is detailed in the next sections.

### 5.1 Maximum Total Number of Transcriptions

The cost of the survey is relatively high. As an example, the 1977 For-hire Trucking Survey cost \$494,000. The distribution of this expenditure by function is given in Table 2 at the end of the paper.

The cost has set the maximum total number of transcriptions to 225,000 shipments. This size has remained the same since 1972 although the

number of shipments carried by the for-hire trucking industry has increased.

#### 5.2 Minimum Number of Transcriptions per Selected D.S.L.P.

A minimum number of shipments are to be transcribed from the filing systems of the D.S.L.P.'s in the sample to justify travel and salary expenditures. Under the present design, a minimum of 200 shipments are selected from each D.S.L.P. in the sample.

#### 5.3 Maximum Number of Transcriptions per Selected D.S.L.P.

Identification of shipment records in the sample and the transcription of information from these records are done at the D.S.L.P.'s. There is a constraint on the number of days the data collection team spends at a particular location, so that the respondents are not burdened by the presence of the team. This constraint translates to a maximum of 3000 shipment records to be transcribed from any one D.S.L.P.

### 6. STRATIFICATION AND SAMPLE ALLOCATION

Using the results of the previous year's Motor Carriers Freight and Household Goods Movers Survey, the D.S.L.P.'s are stratified according to their intercity transportation revenue class, their type of operation and their area of operation. The intercity transportation revenue class indicates if the D.S.L.P. earned \$2 million or more, between \$500,000 and \$1,999,999, or between \$100,000 and \$499,999 dollars of revenue from intercity freight transport. The type of operation says if the D.S.L.P. is a general freight carrier, an automobile carrier, a household goods mover, a van line, a bulk (e.g. petroleum, milk, etc.) carrier or an other specialized (e.g. heavy machinery, livestock, etc.) carrier. The area of operation is different depending on the D.S.L.P. total transportation revenue. If the revenue is greater than or equal to 2 million, the area of operation indicates if the revenue is between 2 and 20 million or

greater than 20 million dollars and if the predominant source of revenue is earned east of the Manitoba/Ontario border, west or from international trucking. If the revenue is less than 2 million dollars, the area of operation indicates which of the 10 provinces, the Yukon, the Northwest Territories or the international brought the predominant source of revenue to the D.S.L.P. There were 102 non-empty strata in the 1978 For-hire Trucking Survey.

Once the frame of D.S.L.P.'s is stratified, the number of shipments to be selected and transcribed ultimately from each stratum is determined by allocating the maximum total number of transcriptions to strata. The allocation technique has changed since the survey started in 1970.

Originally, the allocation was calculated in three steps. First, the total number of transcriptions was allocated to 5 domains so that the coefficients of variation were equal for ton-miles over the 5 domains. These domains were the geographic regions of origin of the shipments. The coefficients of variation were estimated using historical data. This first step gave a number of transcriptions to each domain. Secondly, the number of transcriptions to each domain was allocated to strata using essentially a Neyman allocation. This second step gave a number of transcriptions to each domain within each stratum. Finally, the numbers of transcriptions to each domain within each stratum were summed over the domains to get the total stratum allocation of transcriptions.

In 1975, the allocation scheme was revised because it was felt that the estimates of the true variance of each domain within each stratum needed for the Neyman allocation were not reliable and because the resultant strata allocations were highly variable across years and adversely affected longitudinal analyses.

The revised allocation procedure is radically different than the original one. It has been developed using years of experience and is partially a

judgement allocation based on results of the previous year's survey. It consists of allocating workloads defined as 100 transcriptions. The allocation is performed in three steps.

First, the total number of transcriptions, i.e. 2250 workloads, is allocated to the groups of strata having the same intercity transportation revenue class as follows:

Intercity Transportation Revenue Class	Workload Allocation to Groups of Strata
\$2 million or more	908
\$500,000 to \$1,999,999	832
\$100,000 to \$499,999	510
TOTAL	2,250

Secondly, the allocation of workloads to strata within a group of strata having the same intercity transportation revenue class is performed using a stratum size measure. This measure for a stratum is the total intercity transportation revenue in units of 10,000 dollars of the D.S.L.P.'s in the stratum. For the group of strata having intercity transportation revenue of \$2 million or more, the allocation is proportional to stratum size measure. For the others, the allocation is proportional to the square root of the stratum size measure. The square root is used in the latter case because otherwise some strata having little contribution to the revenue would almost be ignored in the sample.

Finally, the allocations obtained from the previous steps are adjusted as follows. The allocation is reduced to 2 workloads in the strata of the international carriers and household goods movers with intercity transportation revenue of \$2 million or more. It is increased in strata



where detailed data are needed. It is also adjusted to meet the administrative restrictions mentioned in Section 5 and to preserve consistency with previous years' allocations.

## 7. FIRST STAGE SAMPLE DESIGN

The current first stage consists of selecting a twice replicated stratified sample of D.S.L.P.'s. The sample selection is different in class 1 strata than in the other strata. Class 1 strata cover class 1 D.S.L.P.'s which are D.S.L.P.'s earning 2 million or more dollars of intercity transportation revenue. The two sample selection procedures are described below. The selection is done by methodologists of Business Survey Methods Division.

### 7.1 Selection of D.S.L.P.'s in Class 1 Strata and Allocation of Workloads to Selected D.S.L.P.'s.

All class 1 D.S.L.P.'s are selected with probability one. The reason for this approach is that these D.S.L.P.'s are known to be heterogeneous with respect to the principal statistics estimated.

Each class 1 D.S.L.P. next must be assigned a number of workloads for each replicate of the sample. This assignment has to be derived from the stratum allocation which was obtained through the procedure described in Section 6. The distribution of the stratum allocation to individual D.S.L.P.'s is done as follows. Let the stratum allocation be  $x$  workloads and let the number of D.S.L.P.'s in the stratum be  $d$ . One workload is assigned per replicate for each D.S.L.P. in the stratum so a total of  $2d$  workloads are assigned. The remaining  $w = (x - 2d)$  workloads are equally distributed to the two replicates. Then a probability proportional to size systematic sample of  $(\frac{w}{2})$  D.S.L.P.'s is

drawn for each replicate. The measure of size used in the intercity transportation revenue. One workload is assigned per selection to the selected D.S.L.P.

#### 7.2 Selection of D.S.L.P.'s in the Other Strata and the Allocation of Workloads to Selected D.S.L.P.'s

The selection of D.S.L.P.'s in other strata and the allocation of workloads to selected D.S.L.P.'s is done simultaneously as follows. Let the stratum allocation be  $k$  workloads. Each replicate in the stratum is assigned  $\left(\frac{k}{4}\right)$  workloads. A probability proportional to size systematic sample of  $\left(\frac{k}{4}\right)$  D.S.L.P.'s is drawn for each replicate. The measure of size used is the intercity transportation revenue. One workload is assigned per selection to the selected D.S.L.P. Finally, the workload assignments to selected D.S.L.P.'s are doubled.

#### 7.3 Review of the Sample

The selected D.S.L.P.'s and the distribution of workloads are reviewed using the information from the previous survey, and adjustments to assignments are made, if warranted for practical reasons.

### 8. SUBSEQUENT STAGE(S) SAMPLE DESIGN

The subsequent stage(s) of the sample design consist(s) of selecting shipments from the files of each selected D.S.L.P. This selection is done by Statistics Canada Regional Operations Division staff at the D.S.L.P. The sample design is different depending on whether the filing system of the D.S.L.P. is small or large.

If the filing system is small (i.e. less than or equal to 2,000 shipments), then two independent systematic samples of 50 shipments are selected for each workload assigned to the D.S.L.P..

If the filing system is large, (i.e. greater than 2,000 shipments), then a sample of about 100 shipments is selected independently for each workload in two stages. For the first stage, an estimate of the number of bundles (defined as 100 shipments) is obtained and divided into 8 equal sections. Then a bundle is selected at random from each section and the selected bundle is located in the filing system. For the second stage, a systematic sample of shipments is selected from the selected bundle. The sample interval used for the systematic sampling is 8 so that 12 or 13 shipments are usually transcribed. Thus, the 8 sections provide approximately 100 transcriptions for each workload.

## 9. FIELD OPERATIONS

This section discusses the activities that involve the Statistics Canada Regional Operations staff: namely the training of the Regional Operations project managers, the planning of the collection, the collection itself and some special cases.

### 9.1 Training of the Regional Operation Project Managers

Every year, the Statistics Canada Regional Operations project managers are trained on all aspects of the survey. The training is five days long and is conducted during the month of March. The survey project manager as well as methodologists are involved in the training. A collection procedures manual is used for the in-class training. The Regional Operations project managers also receive on the job training by visiting a number of D.S.L.P.'s with different filing systems.

### 9.2 Planning of the Collection

Every spring, the Regional Operations project managers recruit the interviewers and administer a thorough training program. Then the interviewers with the advice of their Regional Operations project manager schedule their work, plan their itineraries and telephone D.S.L.P

officials for appointments. The collection takes place between May and September for the survey covering the shipments of the previous calendar year.

### 9.3 Overall Description of the Collection

When the interviewer gets on the D.S.L.P. premises, he/she has first to conduct an interview with the D.S.L.P. officials. During the interview, he/she will explain the survey, will mention the uses of the data, will estimate the time required to do the work and will complete a control form. The control form records information about the operations of the firm such as the total tonnage transported; the total number of shipments carried; the types of commodities carried and the percentage each type represents in the total transportation revenue; and the filing system(s) used.

Often, the interviewer has a choice of filing systems which provide information on the items needed in this survey. The interviewer assesses the completeness of information on principal characteristics from various filing systems, and then chooses the most appropriate system.

Then the interviewer estimates the number of shipments in the selected filing system of the carriers. This estimation is needed to be able to properly select a sample of shipments and to calculate the weights of the sampled shipments. This estimation involves measurement when the filing system is neither numeric nor broken numeric.

After then, the interviewer selects the sample shipments and transcribes their characteristics. This latter operation is often difficult because it can be hard to understand the various documents and the coding used on some documents. The interviewer often has to interpret the information on the documents and to enter on the coding sheets the data in a format that would be accepted by the computer system.

Finally, the survey project manager and the methodologists are consulted whenever necessary to assist in the field operations.

#### 9.4 Special Cases

This sub-section discusses synthesising, abortions and cancellations.

Synthesising is the construction of hypothetical workloads when a D.S.L.P. does not keep documents suitable for sampling. In such a case, the interviewer collects through an interview with the D.S.L.P. officials macro-information about the D.S.L.P. Then this information is sent to Transportation and Communications Division who constructs shipment data in a format accepted by the computer system.

Abortions are in-scope D.S.L.P.'s for which we have not obtained transcriptions nor macro-information. Examples of abortions are a D.S.L.P. which refuses to cooperate or a D.S.L.P. on strike. The contributions of the abortions are reflected in the estimates via imputations using previous year's data or via adjustments to the weights.

Finally, cancellations are D.S.L.P.'s which are identified as out-of-scope for the reference period. In spite of the efforts made in verifying the D.S.L.P.'s in the universe against several sources on the activity of the carriers, interviewers find out that some D.S.L.P.'s in the sample are out-of-scope for the reference period. In such cases, Head Office makes adjustments to the weights when a large number of cancellations occur within a stratum.

### 10. DATA PROCESSING

All data processing is carried out at Head Office. Incoming data first go through a manual edit procedure which uses the data collected on the control forms. If the shipment data are correct and complete, they are sent to Key-edit. A standard data record file is created by matching the incoming transcriptions and accompanying material with a check-in list called workload master file. Out-of-scope shipments are discarded. There

were about 59,000 out-of-scope shipments in the 1978 For-hire Trucking Survey, i.e. about 26.2% of the transcribed shipments were out-of-scope. Some types of out-of-scope shipments are shipments to or from the U.S.A.; shipments transported 15 miles or less from origin to destination; shipments which were off-highway; shipments which would be double counted as a result of interlining between road carriers; shipments which would be double counted because they were recorded by household goods movers who are van line agents and by the van lines themselves; shipments which relate to a period other than the reference period; shipments which did not bring any intercity transportation revenue; and records which relate to non-transportation services such as storage, packing, equipment rental, labour loading and unloading.

The in-scope shipment records are assigned Standard Geographic Codes, Standard Commodity Codes and the distance between the origin and destination of the shipment using respective computer libraries.

Missing fields and those failing edits are imputed. There are two major imputation procedures that may take place in the system. These are prorating which is arithmetic imputation using fixed relationships and simple Hot Deck which consists of matching the record with missing data to a "similar" record with complete data. Essentially these procedures take values (or codes) to be imputed from valid or complete records and are applied to records which are incomplete. These procedures are premised on the assumption that the characteristics of records within the same workload are similar.

Weights are finally assigned and the data file is passed to the estimation module.

Detailed diagnostics produced at each stage of data processing are used as a quality check on the data passing through the system.

## 11. ESTIMATION METHOD

The estimates are generated from a very small sample relative to the size of the population. As an example, the 1978 sample represented only about 0.5% of the shipments in-scope to the survey.

Each workload generates an independent estimate for its stratum as follows. First, using the estimated size of the filing system in shipments, the individual shipment data of the workload are expanded to obtain an estimate at the D.S.L.P. level. Next, the relative size of the D.S.L.P. to the stratum size in terms of intercity transportation revenue is used to expand the D.S.L.P. estimate to the stratum level.

The average of estimates from all workloads in a given replicate within a given stratum provides the replicate stratum estimate. These replicate estimates are averaged to derive an overall stratum estimate. Finally, these overall stratum estimates are aggregated to provide national estimates.

Standard errors of the estimates are calculated using the two replicated estimates for each stratum.

The tabulated estimates are reviewed to ensure a check on the accuracy of the weights as well as a general check on the quality of the estimates generated. Examples of the review are the comparisons of the estimates to the previous year estimates from the For-hire Trucking Survey and to the previous year estimates from the Motor Carrier Freight Survey.

## 12. FUTURE OF THE SURVEY

The survey is being re-designed. One of the objectives is to develop a new approach to the data collection to reduce the cost of collection per in-scope sampled shipment so that the sample size can be enlarged. The

possibility of using computer tapes as a vehicle for the reporting firms falls into this approach. The progress of the re-design project is being monitored by an interdepartmental committee involving Transport Canada, the Canadian Transport Commission and Statistics Canada.

#### ACKNOWLEDGEMENT

The content of this paper is based largely in work done by R. Sugavanam, Business Survey Methods Division, Statistics Canada.

#### RESUME

Cet article décrit la méthodologie de l'enquête sur le Transport routier de marchandises pour le compte d'autrui. Cette enquête fournit de bons exemples de contraintes administratives et opérationnelles rencontrées par les statisticiens d'enquêtes et par les équipes de collecte de données sur le terrain.

#### REFERENCES

- [1] Cairns, M. and Kirk, B.D. (1980), "Canadian For-hire Trucking and the Effects of Regulation: A cost Structure Analysis", Canadian Transport Commission, Research Report No. 10-80-03.
- [2] Chow, G. (1980), "An Analysis of Selected Aspects of Performances of For-hire Motor Carriers in Canada", report prepared for the Bureau of Competition Policy, Consumer and Corporate Affairs Canada.
- [3] McRae, J.R. and Prescott, D.M. (1980), "Definition and Characteristics of the Trucking Markets: A Statistical Analysis", report prepared for Transport Canada.



- [4] Statistics Canada, "For-hire Trucking Survey", Catalogue 53-224, Annual.

TABLE 1

Description of the Frame of the 1978 Survey and of the Principal Statistics Estimated from the Survey Data by Intercity Revenue Group.

Statistics	Intercity Revenue			
	\$2,000,000 and more	\$500,000 to \$1,999,999	\$100,000 to \$499,999	Total
Number of D.S.L.P.'s in the frame	218	540	1,339	2,097
Sample allocation in shipments	90,800	83,200	51,000	225,000
Number of shipments used in tabulation	69,620	62,729	33,646	165,995
Estimated number of shipments in-scope to this survey	20,146,157	7,016,909	5,261,362	32,424,428
Estimated revenue (\$,000)	1,720,578	449,781	303,575	2,473,934
Estimated number of tons (,000)	62,703	36,743	22,800	122,426
Estimated number of ton-miles (,000)	16,594,065	5,277,815	3,120,894	24,992,774

TABLE 2

1977 For-hire Trucking Survey Expenditure by Functions

<u>Function</u>	<u>Expenditure</u>
Field Data Collection	\$295,000
Mangement, Operations	\$ 98,000
Data Processing	\$ 65,000
Methodological Support	\$ 30,000
Travel, Printing, Misc.	6,000
<hr/>	<hr/>
Total	\$494,000