THE REDESIGN OF A SURVEY TO MEASURE COMMODITY ORIGIN AND DESTINATION MOVEMENTS BY THE FOR-HIRE TRUCKING INDUSTRY IN CANADA[1]

Robert Lussier and Steven Mozes[2]

This paper firstly provides an overview of the For-hire Trucking Survey background and of the steps that were involved in the revision that led to its re-design. It secondly describes the general direction of the methodology of the re-designed survey which is being implemented for reference year 1981.

1. INTRODUCTION

The For-hire Trucking Survey was initiated by Statistics Canada in 1971 to measure commodity origin and destination movements by the for-hire trucking industry in Canada. For the purpose of the survey, this industry was defined as the sum of trucking establishments engaged in transportation of freight for compensation. The survey was a probability sample survey of shipments recorded on the shipping documents retained by Canada's for-hire trucking firms. Since 1971, the demand for more reliable and more detailed information has been increasing steadily. This increased demand can be attributed to a number of factors such as the dramatic growth of trucking activity since the early fifties; the increased sophistication of users of transportation statistics; the growing interest in the subject of economic regulation versus deregulation, and finally the increased market share of trucking, at the expense of other modes of transport, within the overall freight transportation market. In the late 1970's, Statistics Canada in cooperation with major users embarked on a complete revision of the survey.

It is the intention of this paper to serve two purposes. Firstly, it provides an overview of the background of the survey and a description of the steps in the revision process, and secondly, it describes the methodology of the redesigned survey which is being implemented for the reference year 1981. It should be noted that the details of the methodology of some phases have not yet been finalized; this paper will, however, include descriptions of the general direction of the incomplete phases.

2. BACKGROUND

2.1 Brief overview of the Canadian for-hire trucking industry

The Canadian for-hire trucking industry is characterized by a very large number of small operators and by a high degree of heterogeneity as manifested by the variety of commodities carried, size of operators, and area of operations.

Small carriers defined as carriers earning less than $100,000 represent 88% of the industry, measured in terms of numbers; however, they represent only 20% of the industry when measured in terms of operating revenues. The existence of this large number of operators, their volatility and their relative insignificance in terms of revenues lead to the decision to exclude them from the survey population.

Trucking firms are involved in the transportation of widely differing commodities, requiring different kind of equipment and operating practices. The various carrier types (e.g. general freight, bulk petroleum, household goods movers, etc.) differ from each other not only in terms of the commodities they carry but also in terms of shipment size.

Heterogeneity can also be illustrated by describing the area of operation. Some trucking firms operate locally only, others intraprovincially, and some of the larger carriers, in each province as well as internationally.

The combination of these factors have implications on the survey design especially on stratification.

2.2 History of Canadian truck origin and destination surveys

(a) The Motor Transport Traffic Survey (1957-1963)

The first attempt to measure truck traffic in Canada was made in 1957 with the introduction of the Motor Transport Traffic Survey (MTTS). This survey was a sample survey of motor vehicles engaged in freight transportation. The survey frame was a list of registered motor vehicles. It originated from the motor vehicle registration files maintained by the provincial or territorial governments. This frame was stratified by type of operation and gross vehicle weight.

The sample size was approximately 10% of all registered vehicles. The sample was selected in four quarterly segments with approximately one fourth of the total sample seleted each quarter. Each quarterly sample was spread over three survey weeks with one third of the sample being used for a seven day period per month.

As the survey was conducted on a vehicle basis, no information was requested regarding the detailed origin and destination of the commodities carried. It was a truck origin and destination survey; commodities had secondary importance. Data relating to the vehicle such as the description of the vehicle, miles travelled, fuel consumed and the operating cost associated with the vehicle were also collected.

The survey was in operation from 1957 to 1963 inclusive. It was discontinued in 1964 because of changes in vehicle licensing systems, structural changes in the industry, and most importantly because of a very se-rious deterioration in response rates.

(b) <u>The For-hire Trucking Survey (1969-1979)</u>

Initial work on a survey to measure domestic intercity origin and destination traffic movements of goods by the total Canadian for-hire trucking industry 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 a sample survey of the carriers' administrative records, namely their shipping documents, was a viable approach to the collection of the required data.

In 1970, a pilot survey was conducted to assess the effectiveness of the survey approach. The pilot survey involved the examination of the shipping documents of 187 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 on the shipping documents indicated that the survey approach was feasible.

For-hire trucking surveys were therefore conducted for reference years 1970 and 1971 with the above-mentioned objective. For reference year 1972, the objective was modified to restrict the survey to Canadian domiciled for-hire carriers earning $100,000 or more annually from inter-city trucking. For reference year 1973, an updated and better-defined frame of regulated motor carriers was used and a more effective sampling procedure was developed. Since reference year 1973, the survey has been conducted and the results published on an annual basis by the Transportation and Communications Division of Statistics Canada [1] [5].

3. **REVISION OF THE FOR-HIRE TRUCKING SURVEY**

The revision process consisted of two main phases. Firstly, a critical review of the existing survey was initiated. Secondly, on the basis of the recommendations made during the review process, a complete survey redesign was

undertaken.

3.1 Survey review

(a) Reasons for the review

In early 1978, Statistics Canada initiated a review of the For-hire Trucking
Survey for the following reasons. First, it has been the policy of the
Transportation and Communications Division of Statistics Canada to conduct a
periodic review of each of the ongoing surveys. The For-hire Trucking
Survey had not been reviewed since 1973. Secondly, the current and anticipa-
ted future needs of users for increased details for commodity origin and
destination statistics could not be satisfied within the constraints of the
survey. Thirdly, experience gained during the undertaking of the For-hire
Trucking Surveys and other related surveys provided additional information
upon which the frame, the stratification variables and the imputation techni-
ques could be improved. Fourthly, some developments in the trucking industry,
such as the availability of origin and destination information in machine
readable format lead to the belief that computer tapes could be utilized to
increase the data base and reduce reporting burden at the same time.

In addition, increased sophistication of users required the development of im-
proved data dissemination procedures while changes in data processing techno-
logy had made the present data processing system not only obsolete from a
technical point of view but cost inefficient as well.

(b) Phases of the review

The survey review was originally organized into two parts, namely Phase I and
Phase II.

The objective of Phase I was to outline recommendations which concentrated on
improvements to the survey within its existing framework using only limited
additional resources. The recommendations had to focus and indeed did focus

on a redefinition of the survey population, improvements in stratification variables and an increase in the sample size of shipping documents. The recommendations were presented in a report [2].

The objectives of Phase II were to assess the Phase I recommendations from a user point of view, to present various cost and implementation alternatives for the accepted recommendations and to complete further survey analyses. Phase II reformulated some of the Phase I recommendations and added additional recommendations which aimed at a smaller population of firms better stratified into more homogeneous groups. In addition to recommending the implemention of these recommendations, four alternatives for increasing the sample size were considered namely, the status quo; an increase of 50% to the sample of shipping documents; an increase of 100% to the sample of shipping documents; and finally an increase of 25% to the sample of shipping documents together with the processing of available carrier data tapes for presumably 40 or so firms. Based on an assessment of the advantages and costs of each of these alternatives, the latter one was approved in principle because it offered the potential for substantial sample size increases with a minimum of cost and data collection burden. The recommendations and the supporting details were tabled in a report [3].

A preliminary assessment of the impact of the recommendations revealed that further work was necessary especially to determine the full costs for the use of carrier waybill tapes. Therefore a Phase III was added tothe survey review process. In general, its terms of reference were to conduct the investigations required to formulate and recommend general specifications for a revised survey. The investigations had to follow the recommendations of the Phase II review.

In June 1980, Phase III proposed that the survey be redesigned to accept four types of input namely, tapes from selected respondents; transcriptions from samples of shipping documents drawn from each Document Storage Location Point (D.S.L.P.) having over 1.5 million intercity domestic revenue annually; transcriptions from samples of shipping documents drawn from a sample of D.S.L.P.'s

having between $350,000 and $1.5 million intercity domestic revenue annually; and macro information from D.S.L.P.'s with annual intercity domestic revenue between $100,000 and $350,000. The decision to collect macro information from the smaller carriers was based on that fact that these firms do not keep the documentation needed for sampling purposes.

3.2 Survey redesign

(a) Objective of the redesign

After the completion of Phase III of the survey review process, it was decided to carry out a complete redesign of all aspects of the survey. The objective of the redesign was to provide more reliable and more detailed commodity origin and destination statistics relating to the Canadian for-hire trucking industry. It was expected that both the reliability and the amount of regional and commodity detail available could be increased when compared with the "old survey".

(b) Constraints on the redesign

The main constraints imposed on the redesign were: that the survey population exclude some types of for-hire trucking firms namely, own account household goods carriers and oil field carriers; that the stratification be improved to be more in line with the economic structure of the industry; that three types of input be accepted, namely, tapes from selected respondents, transcriptions from samples of shipping documents drawn from D.S.L.P.'s of a sample of firms having more than $350,000 intercity domestic revenue annually and macro-information from a sample of firms having annual intercity domestic revenue between $100,000 and $350,000; and finally that the redesigned survey be implemented for the reference year 1981, data collection starting in the spring of 1982.

4. POPULATION AND FRAME

The population of the survey covers all shipments made during the reference year by those trucking firms which are defined as in scope for the survey. A shipment is defined as a quantity of merchandise transported by the carrier from one person or organization to another person or organization. The in-scope firms include those which earn more than $100,000 annually from intercity freight transportation, whose main activity is trucking and who are Canadian domiciled. Excluded from this population are the shipments of certain types of specialized carriers such as the oilfield carriers and own account household good movers.

However, this ideal population is not accessible. As a substitute, firms are used as natural clusters of shipments for the first-stage sampling units of the design.

For this reason, the frame consists of a list of all firms which have domestic intercity revenue over $100,000. Firms may further be segregated into D.S.L.P's. This is the case for those firms whose shipping documents are not stored at a central location. The frame is derived from an annual census survey of for-hire trucking conducted by Statistics Canada, the Motor Carriers-Freight and Household Goods Movers Survey[3].

5. ULTIMATE SAMPLING UNIT

The survey accepts three different inputs, namely tapes from selected carriers, transcribed information from sampled shipping documents and finally macro information from carriers earning between $100,000 and $350,000 annually.

---

[3] The Motor Carriers-Freight and Household Goods Movers Survey of Statistics Canada is an annual census survey of trucking establishments. Its objective is to obtain establishment-oriented input-output data such as revenues, expenses, balance sheet information and equipment operated.

The tapes contain information relating to individual shipments, the characteristics of which are the same as those which are recorded on shipping documents. Therefore, the ultimate sampling unit for those firms which either provide tapes or whose shipping documents are sampled is the shipment. For the carriers in the $100,000 to $350,000 range, macro information is obtained as these firms do not usually keep the necessary documentation relating to shipments. For these carriers, the ultimate sampling unit is the firm.

6. INFORMATION COLLECTED

The principal characteristics needed from each shipment sampled from carriers earning more than 350,000 intercity domestic revenue annually are the true origin and the final destination; the description of the commodity(ies) carried; the weight and the unit of weight; the transportation revenue earned and the interlined shipment information. Interlining 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 eliminate duplications.

The secondary characteristics needed are the date of shipment; the quantity of commodity and the unit of measurement (e.g. 5 board feet, 20 gallons, 15 sacks); some information regarding the shipment weight transcribed (e.g. minimum weight, convenient weight used for calculating revenue); the rate charged and the rate condition codes (e.g. a code indicating where rate is minimum, per 100 lb., per hour) and the revenue condition codes. (e.g. a code indicating where exact transportation revenue is not available, where the shipment is out-of-scope).

The macro information collected from the smaller carriers describe the average or typical shipments in terms of originating province, destination province, commodity, average revenue, average weight and number of shipments.

7. ADMINISTRATIVE RESTRICTIONS

The amount of resources available for data collection and processing and the goal to reduce the burden imposed on the respondent put a limit on the number of firms selected and on the number of shipments selected and transcribed.

7.1 Maximum number of firms in the sample

The population of the 1981 For-hire Trucking Survey[4] consists of 2,711 firms of which 1,288 earn more than $350,000 annually while 1,423 earn between $100,000 and $350,000 annually.

As data collection is very expensive due to the very high cost associated with travelling to remote areas, efforts are being made to limit the number of D.S.L.P.'s selected in the sample from those carriers earning over $350,000 annually. The limit is set at 875 D.S.L.P.'s per year, which has been the historical number during the last ten years of the old survey.

7.2 Maximum total number of transcriptions

The second administrative restriction relates to the total number of transcriptions. The present budget allocation allows a maximum of 418,000 transcriptions. This number may vary from year to year depending on negotiations between Statistics Canada and users who are also cofinancers of the survey.

7.3 Maximum number of transcriptions per firm

There is also an administrative restriction which relates to the maximum number of transcriptions per firm. There is an implicit limit imposed on the number of days the data collection team can spend at any particular location, so that the respondents are not burdened by the presence of the Statistics Canada regional operations personnel.

---

[4] 1981 For-hire Trucking Survey means the survey conducted in 1982 for reference year 1981.

8. STRATIFICATION AND SAMPLE ALLOCATION

Using the results of the previous year's Motor Carriers-Freight and Household Goods Movers Survey, the firms are stratified according to their in-scope transportation revenue, type of operation and area of operation. These variables were chosen because they characterize the heterogeneity of the industry. The in-scope transportation revenue indicates if the firm is a Class 1, a Class 2 or a Class 3 firm i.e. if the firm earned $2.7 million or more, between $350,000 and $2.7 million or between $100,000 and $349,999 dollars of revenue respectively from Canadian intercity non-armoured and non-household goods freight transport. The type of operation characterizes the firms as specializing in general freight small shipments, general freight large shipments, automobiles, liquid petroleum, dump trucking, forest products, building materials, dry bulk and/or refrigerated liquids, heavy machinery, refrigerated solids, explosive and/or other dangerous goods, agricultural products, animals and van lines. The general freight small shipment carriers are general freight carriers for which the average revenue per shipment is less than $85.00; the general freight large shipment carriers are the rest of the general freight carriers. The area of operation indicates the specific Canadian province, Yukon or Northwest Territories, or that combination in which the which the firm operates. For example, an area of operation could be New Brunswick, meaning that the firm operates in New Brunswick only. Another example would be Atlantic which means that the firm operates in 2 or more of Newfoundland, Prince Edward Island, Nova Scotia and New Brunswick but nowhere else in Canada. There are 20 of these areas of operation.

The dollar cut-offs used in the stratification by revenue and by type (i.e. $85, $350,000 and $2.7 million) are flexible and may vary in the years to come depending on the changes occurring in the population.

The above stratification creates 840 strata of which 355 were non-empty in the 1981 For-hire Trucking Survey.

Once the frame is stratified, subject matter officers may identify take-all

firms i.e. firms that they want to be included in the sample with probability one. Next, a methodologist determines the number of firms to be selected among the non take-all firms in the stratum. To do so, he goes through several steps from which the take-all firms are excluded.

First, a computer program calculates the initial number of firms to be selected in each stratum to meet a target coefficient of variation of the estimate of in-scope revenue in the stratum. This target coefficient of variation is the coefficient that one would like to obtain if the estimate were calculated using the reported total revenue from a sample of firms selected using simple random sampling from a population of firms for which the distribution of the in-scope revenue is the same as the distribution of the in-scope revenue of the previous year's Motor Carriers-Freight and Household Goods Movers Survey. The formula is:

$$_1 n_h = \frac{N_h^2 S_h^2}{N_h S_h^2 + Y_h^2 (C.V._h)^2}$$

where $_1 n_h$: initial number of firms to be selected among the non take-all firms in stratum h;

N_h : number of non take-all firms in stratum h;

Y_h : total in-scope revenue of the non take-all firms in stratum h;

S_h^2 : variance of the in-scope revenue of the non take-all firms in stratum h; and

$C.V._h$: target coefficient of variation in stratum h (the value used is the same for all strata of a given class but may vary from class to class).

Secondly, the initial sample sizes are revised to ensure that a minimum number of firms is selected from each stratum i.e.

$$_2 n_h = \min \left\{ \max \left(m, \ _1 n_h \right), \ N_h \right\}$$

where $_2 n_h$: revised initial number of firms to be selected among the non take-all firms in stratum h; and

m : minimum number of firms to be selected in stratum h if possible[5].

Then the revised initial sample sizes are summed over the strata to get a total revised initial sample size.

Next, the sample sizes are again reviewed to ensure that the sample size in a given stratum is greater or equal to the sample size that one would have obtained if he had distributed the total revised initial sample size of a class across the strata of the class proportionally to the square root of the number of firms in each stratum i.e.

$$_3 n_h = \max \left\{ \frac{\sqrt{N_h}}{\sum\limits_{h} \sqrt{N_h}} \sum\limits_{h} {_2 n_h}, \ _2 n_h \right\}$$

where the summation is done over all strata of the same class than stratum h.

Finally, the survey manager may subjectively adjust the sample sizes to $_4 n_h$.

The above sample allocation method has been retained because it is an algorithm which has given satisfactory results during the testing phase as well as has made use of the only variable that was available for all firms namely the in-scope revenue of the firms. Nevertheless, it should be realised that the in-scope revenues of the firms are not collected directly in the For-hire Trucking Survey but revenues from a sample of shipments are collected. Therefore the above method of sample allocation ignores completely the second stage of sampling.

---

[5] For reference year 1981, this minimum was set to 3 for all strata.

9. FIRST STAGE SAMPLE DESIGN

The first stage consists of selecting in each stratum a number of firms corresponding to the number of firms $_4n_h$, determined at the sample allocation stage.

All firms earning \$2.7 million or more of in-scope transportation revenue were made take-all i.e. were selected with probability one in the 1981 For-hire Trucking Survey. The reason for this approach is that these firms are known to be heterogeneous with respect to the principal statistics to be estimated and are known to be contributing a large proportion of the revenue figures to be estimated.

The sample of firms is finally converted to a sample of D.S.L.P.'s by including in the latter sample all D.S.L.P.'s of the selected firms.

10. SECOND STAGE SAMPLE DESIGN

The second stage of the sample design for D.S.L.P's of Class 1 and Class 2 firms consists of selecting a systematic sample of shipments from the files of each selected D.S.L.P. This selection is done by Statistics Canada Regional Operations Division interviewers at the D.S.L.P. The sampling intervals used are different depending on the number of shipments carried by the firms. They are generally obtained from a table provided to the interviewers. This table gives various file size ranges with their corresponding sampling interval. However, the sampling intervals may be pre-determined for any given firm by Statistics Canada Head Office staff. This is especially the case of multi-D.S.L.P. firms because the interviewer at a given D.S.L.P. may not know how many shipments were carried by the firm as a whole. This is also the case for firms having special characteristics, such as firms carrying dangerous goods, and others for which the survey manager may want a larger data base. In subsequent years, this may also be the case for firms contributing to domains where the reliability of the estimates in the previous year was less or more

than what was desired.

For D.S.L.P.'s of Class 3 firms, there is no second stage sampling design. Individual shipments are not selected from the files of the D.S.L.P.'s. Instead, aggregated data are collected at the D.S.L.P. level.

11. FIELD OPERATIONS

The field operations are different for class 1 and class 2 firms than for class 3 firms. For class 1 and class 2 firms, the operations consist of selecting shipments from the files of their D.S.L.P.'s and of transcribing the characteristics of the selected shipments on coding sheets. For class 3 firms, they consist of obtaining aggregated data over the telephone about their trucking 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 at the D.S.L.P.'s of class 1 and class 2 firms, the collection from the class 3 firms and finally the profiling of class 1 and 2 D.S.L.P.'s.

11.1 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 session is four days long and is conducted during the month of March. It is broken down into two components: an in-class-training and an on-the-job training. The in-class training consists of a series of talks and exercises given by the survey project manager and the methodologist(s). The on-the-job training consists of having groups of three to four people visiting a D.S.L.P. and applying and discussing the knowledge acquired during the in-class training.

11.2 Planning of the collection

Having been trained, 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 and plan their itineraries for their visits to D.S.L.P.'s of class 1 and class 2 firms. The itineraries are drawn to avoid unnecessary travel and to achieve maximum productivity. The interviewers mail to the D.S.L.P. officials introductory letters which provide a brief explanation of the survey. Subsequently, the interviewers telephone D.S.L.P. officials for appointments. The collection of the data takes place between May and September for the survey covering the previous calendar year.

11.3 Overall description of the collection in the D.S.L.P.'s of Class 1 and Class 2 firms

At the time of the appointment, the interviewer conducts an interview with the D.S.L.P. officials. During the interview, he/she explains the survey, describes the uses of the data, estimates the time required to do the work and asks information about the firm. This information concerns mainly revisions to the names and addresses, changes of ownership, type(s) of document and filing system used and aggregated data about the operations of the D.S.L.P. during the reference year.

The most common types of shipping documents are the probills, bills of lading, load manifests, trip reports, and invoices. A firm may use any combination of these.

The types of filing system include: in complete numeric sequence; in broken numeric sequence; in chronological order; in alphabetical order (e.g. by customer name); by terminals; by commodity type 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 sets of binders or shannon files, on shelves, in drawers, or even in books.

The aggregated data about the operations of the D.S.L.P. cover several variables among which are the total transportation revenue earned; the total tonnage carried; the total number of shipments transported; the percentages of each of these three items represented by intercity shipments and the percentages represented by international shipments; the types of commodities carried and the percentage each type represents in the total transportation revenue.

Often the interviewer has a choice of filing systems which provide information on the items needed in the survey. The interviewer assesses the completeness of the various filing systems with regard to the information on the five principal characteristics and on the reference year, and then chooses the system having the smallest under-coverage. However, if two or more systems have the same under-coverage (if any), the interviewer selects the one that includes the smallest number of out-of-scope records or the one that allows out-of-scope records to be removed from the file or not to be counted.

Next, the interviewer selects the sample of shipments as follows. Using the number of shipments reported by the official of the D.S.L.P., he/she gets from a table the corresponding sampling interval and random start. In some instances, the interval and the random start may have already been pre-determined by Statistics Canada Head Office. Next, he/she adds the random start and/or the interval to the document numbers to get the selected shipments in numeric filing system. Otherwise, he/she has to count a number of documents equal to the random start and/or to the interval to get the selected shipments.

Once a shipment is selected, the interviewer transcribes its characteristics. The transcribing operation is often difficult because it can be hard to understand the various documents and the coding used on some documents. This is especially true for the commodity names. The interviewer must avoid the use of brand names, proper names and names which have more than one meaning. 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.

11.4 Overall description of the collection in the Class 3 firms

The interviewer mails an introductory letter two to three weeks prior to any attempt to contact the firm by phone. Subsequently, he/she contacts the official in the firm that is best suited to provide the required information. This may, however, take several phone calls. The interviewer then conducts an interview over the phone.

During the interview, he/she will ask questions similar to the questions asked for class 1 and 2 D.S.L.P.'s. There is a major difference however; no questions are asked about the types of documents utilized and the filing systems used by the firm. Once this first part of the interview is completed, the interviewer proceeds to have the respondent describe his types of shipments. For each type of shipment, the description is to be made in terms of province of origin; province of destination and name of commodity carried. Then the official is asked to report an estimate of the number of shipments, the average weight and the average revenue of each type of shipment.

It is a general subject matter belief that the operations of any given class 3 firms are fairly homogeneous. Therefore each has only a few types of shipments to report. The coverage obtained throught this approach is also believed to be acceptable from a user point of view. No testing was done of this hypothesis.

11.5 "Profiling" of Class 1 and 2 D.S.L.P.'s

It sometimes happens that a class 1 or 2 D.S.L.P. cannot provide any documents, does not keep documents suitable for sampling or cannot provide a portion of the shipping documents and that this portion cannot be represented by the available documents. The latter may happen for example when the missing documents represent specific contracts that have been removed for audit purposes. In these cases, the interviewer has to "profile" the missing documents. The profiling consists in having a D.S.L.P. official describe the types of shipments on the missing documents. The profile is similar to the description

of the types of shipments of the class 3 firms with the exception that the precise origin and the precise destination of the shipments (i.e. the village, town, city, etc.) is wanted in this case.

The profiling activity can be long in some D.S.L.P.'s because their operations can be quite extensive. It requires good cooperation from the D.S.L.P.'s official.

12. DATA PROCESSING

12.1 Manual processing

The completed documents are sent to Statistics Canada Head Office in Ottawa. Upon receipt, the documents are logged in and the identification numbers verified. Two short tasks are also undertaken at this point.

First, a brief scan is conducted to identify and code closings of D.S.L.P.'s, death of firms, out-of-scope firms and abortions. Out-of-scope firms are active firms for which the in-scope revenue is nil for the reference year. Abortions are D.S.L.P.'s for which no information was collected although it was known that the D.S.L.P. had in-scope revenue for the reference year. As examples, a firm found in the field to have earned its revenue 100% from local shipments would be an out-of-scope firm while a single D.S.L.P. firm that refuses to cooperate or is on strike would be an abortion.

Secondly, the profile data of the class 1 and 2 D.S.L.P.'s are examined to determine the number of shipments that should have been transcribed for each reported type of shipments if the documents had been available. These numbers are determined by performing calculations using the total number of shipments covered by the profile, the random start and the sampling interval that should have been used if the documents had been available. These numbers are then coded so that the computer could generate the required number of transcription records for each type of shipments as if transcriptions were obtained.

12.2 Data capture

The forms are next sent to data capture. The capture is done on a mini-computer which allows edits and other processing to be performed on-line.

There are many edits performed on the mini. Some edits generate error messages and require corrective action; others generate warning messages that require verification of the entered data and corrective action only if necessary Some edits consider the validity of each response individually while others consider the relationships between valid characteristics of the same shipment. The operators of the mini-computer are expected to possess subject matter knowledge to perform corrections on-line. Manual imputations are performed when necessary because there is no automated imputation performed for class 3 D.S.L.P.'s.

As part of the other processing, the weight is converted to metric and the rate to $/100 kilograms. Also, the origin and destination names (i.e. village, town, etc.) if present, are matched against a municipality library to obtain a Standard Geographical Classification (S.G.C.) code, a latitude and a longitude. Whenever there is a nonmatch, the operator is instructed to enter a synonym. Similarly, the commodity name is matched against a commodity library to get a 3-digit Standard Commodity Classification (S.C.C.) code. Whenever there is a nonmatch, the operator uses a synonym or enters "unknown". There is therefore always an S.C.C. code for each shipment. Also, the mini generates the required number of transcription records for each type of shipments from the profile data of the class 1 and 2 D.S.L.P.'s.

Finally, the data are unloaded from the minis and two data sets are created; a data set of shipments of class 1 and 2 firms and a data set of type of ship-data of class 3 firms. The principal difference between the two data sets are that the first one is at the shipment level while the second one is at an aggregated level. Note also that the first one has more variables (e.g. rate, place of origin rather than province of origin, etc.) than the second one.

12.3 Main frame edits and imputations

A road distance between the origin and the destination of each in-scope class 1 or 2 shipment has to be obtained in order to be able to provide tonne-kilometres estimates for class 1 and 2 firms. Therefore, the origin S.G.C. - destination S.G.C. pair of each in-scope class 1 or 2 record is matched against a distance library to get a road distance in kilometres between the two locations. Whenever there is a nonmatch, an aerial distance (X) is calculated using the latitudes and longitutes of the origin and of the destination. using the latitudes and longitudes of the origin and of the destination. Then X is converted to a road distance Y using the simple linear regression model

$$Y = a X + b$$

where a and b vary according to 12 regions of origin and 12 regions of destination. The road distance is assigned to the record.

Missing data of partially transcribed shipments of class 1 and 2 firms are also imputed. The imputation technique used depends on the missing variable or the pair of missing variables. Major imputations are performed using fixed relationships between reported figures, unit weight conversion factors and pro-rate tables. An example of a fixed relationship between reported figures is

$$\text{weight} = \frac{\text{revenue} \times 100}{\text{rate}}$$

This relationship can be used to impute weight when revenue and rate are present or revenue when weight and rate are present. Unit weight conversion factors are coefficients determined by unit type (e.g. case, bag, litre, etc.) by S.C.C. code. Knowing the unit and the S.C.C. code of the commodity, the proper conversion factor can be applied to the quantity of units to derive the weight. Finally, pro-rate tables show rates by commodity section, by distance block and by revenue or weight group. These tables are based on the previous years' data modified by incoming valid current-year data. The pro-rate tables are used to calculate the weight when the revenue is present or the revenue when the weight is present.

In cases where too many characteristics of a shipment have to be imputed, the shipment is flagged as not usable.

Expansion edits are subsequently performed. For class 1 and 2 firms, these edits consist of weighting up crudely the number of shipments transcribed, the transcribed revenue and the transcribed tonnage and comparing the results to the total number of shipments, revenue and tonnage provided during the interview by the D.S.L.P. official. Similar edits are performed for class 3 firms. Discrepancies in both cases are followed up.

13. ESTIMATION PROCEDURES

For the estimation procedures, it was decided to consider the second stage systematic sampling in the class 1 and 2 firms as simple random sampling without replacement (S.R.S.W.O.R.). This decision was made because first the documents were considered to be in random order and secondly the use of S.R.S.W.O.R. allows the computation of an estimate of the sampling variance.

As the first step of the estimation procedures, weights are calculated. There are first stage and second stage weights for class 1 and 2 records but only first stage weights for class 3 records. In general, first stage weights correspond to the inverse of the probability of selecting of a D.S.L.P. in its stratum and second stage weights correspond to the inverse of the probability of selecting of a shipment in its D.S.L.P. supposing S.R.S.W.O.R. was used. First stage weights are adjusted by computer to reflect the contribution of abortions. No adjustments are made for the closing of D.S.L.P.'s, deaths of firms and out-of-scope firms because they are considered as having generated no shipments. Final weights are attached to each record on the data set of class 1 and 2 firms and on the data set of class 3 firms.

Detailed diagnostic reports are produced. These reports are tables which present the data under various aggregates. They are useful tools to analyse the data and to perform final quality checks.

The data set of class 1 and 2 firms is cleared by discarding out-of-scope shipments. 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 did not generate any intercity transportation revenue; and records which relate to non-transportation services such as storage, packing, equipment rental, labour loading and unloading.

Estimates of revenue, tonnage and tonne-kilometres for the publication are finally generated by summing the weighted data over the appropriate domains. Measures of error such as the coefficients of variation are also provided with the estimates. The coefficients of variation are obtained from the formula derived from the sample design but supposing the systematic sample of shipments is a simple random sample of shipments.

14. USE OF THE DATA AND METHODS OF DATA DISSEMINATION

14.1 Use of data

Requests for estimates yielded from the old survey came from a wide variety of sources. The nature of these requests has also varied a great deal. It is expected that the nature of the demand for data from the new survey will be similar to that in the past.

The estimates have been used extensively to satisfy five main requirements, namely, to measure the volume of domestic trade transported by intercity for-hire carriers provincially and interprovincially; to measure the rate of industrial growth reflected by intercity commodity movements; to provide information on regional development; to assist in transportation studies; and to support the presentation of briefs, submissions and other inquiries to

regulatory authorities and commissions.

One specific use of the data was to define the characteristics of trucking markets using variables such as commodities carried, average lengths of haul and shipment weight. Another specific study examined and analyzed selected aspects of performance of carriers operating in regulated and unregulated environment. The cost behaviour of these carriers was examined by using traffic characteristics such as shipment size and average length of haul.

In the past, special requests for estimates from this survey have come from 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.

14.2 Methods of dissemination

The redesigned survey will provide information in three modes similar to the old survey.

First the publication will present the estimates that are generated by the regular system of the survey. Measures of error such as the coefficients of variation will be given with the estimates. Secondly, special requests will be processed subject to cost and reliability constraints. Finally, the data base of shipments generated by the survey may be made available on magnetic tape to selected users subject to constraints of confidentiality.

15. FUTURE WORK

As mentioned earlier in this paper, the survey accepts three types of input, one of which is computer tape from selected respondents. This type of input has been found difficult to handle and, although work has commenced on this

subject, progress so far is disappointing. Extensive negotiations are requi-
red with the firms to obtain the requested data on tape and then a further
analysis is needed to evaluate the data. For reference year 1981, only one
tape will be used for a firm which handled about 5 million shipments during
reference year 1981. More firms will provide data on tape in subsequent
years. Agreements are presently being reached with 5 additional companies for
reference year 1982.

Nevertheless, when a firm's computer tape is finally obtained and found to
meet our requirements, extensive systems manipulation will still be required
to handle the tape. Also, manual interventions will be necessary to handle
non matches to the various libraries. Therefore, records will most likely
have to be sampled on each tape using the same second stage sampling design as
for the sampling of documents of class 1 and 2 firms.

Another area where future work is needed is on having firms themselves sample
their documents. As an example, a company could photocopy the pro-bills
ending in a given number when the pro-bills are issued, and send the photoco-
pies to Statistics Canada on a monthly basis.

Finally, major efforts will be made to evaluate thoroughly the various phases
of the survey and to formulate recommendations for improvements. These recom-
mendations will hopefully be implemented for the 1982 reference year survey.

REFERENCES

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

[2] "Report on the Findings and Recommendations of the Working Group on the
For-hire Trucking Survey Phase I Review"; prepared by the Transportation
and Communications Division of Statistics Canada, July 7, 1978.

[3] "Report of the Interdepartmental Working Group on For-hire Trucking -
     Phase II Review"; prepared by the Transportation and Communications
     Division of Statistics Canada, April 1979.

[4] Statistics Canada Motor Carriers - Freight and Household Goods Movers,
     Catalogue No. 53-227, Annual.

[5] Lussier, R. (1981), "For-hire Trucking Survey: Survey Design", Survey
     Methodology, Statistics Canada, Vol. 7, No. 1, pp. 74-92.