ASSESSING THE REPRESENTATIVENESS OF SURVEY RESPONDENTS FOR MICRO-LEVEL NETWORK RESILIENCE MODELING[1]

Georgiana Madar, University of Windsor[1]
Hanna Maoh, University of Windsor

Introduction

Freight transportation is an integral component of economic activity in Canada, with a significant portion of Canadian freight being transported by trucks across the road network. It is expected that any disruptions or impediments to these flows can result in significant economic costs and loss of productivity, in addition to having significant negative impacts on local traffic congestion and the excessive wear of infrastructure elements.As such, it is essential that the planning of the transportation network is done in a thoughtful and informed manner to ensure that freight flows can continue moving efficiently throughout the network and the negative effects of disruptions are mitigated and minimized. Freight fluidity, a concept that has been of increasing interest to Transport Canada (Gregory and Kwiatkowski,2011) and their counter partsin the United States (US), refers to the ease with which freight flows are able to move across a transportation network. This fluidity can be characterized through indicators of network attributes including robustness (the network's ability to remain intact in the presence of a disruptive event) and resilience (the network's ability to return to a state of equilibrium following a disruption). Considering the complexities involved in quantifying such characteristics, a multidisciplinary exercise is encouraged, collaborating to produce informed and holistic analyses of network conditions leading to policies and decisions about infrastructure investments that maximize benefits to economic productivity.

Traditional modeling exercises in network robustness and related areas have primarily been done using aggregate data. Recent thinking suggests thata micro-level approach, using individual establishments as the unit of analysis,can improve the accuracy of models for characterizing existing conditions on the network. As the freight flows originate and terminate with individual firms, such an approach would allow for a more accurate depiction of freight flows in a regional context.However, the data requirements are more intensive than aggregate methods due to the need to collect detailed information about shippers. Even when such data is collected, the representativeness of the collected records must be examined to avoid any potential bias when performing statistical analysis.

To this end, this study examines a dataset consisting of survey responses from Canadian business establishments that is considered for application in a micro-level modeling exercise. The survey collected data from over one thousand Canadian firms, including information about the respondent establishments' characteristics, such as industry classification and employment size, and their shipping activities, such as goods classifications and origin and destination locations. For this data to be applied to a modeling exercise with any accuracy, it must be ensured to be representative of the entire population of Canadian shippers. The aim of the current paper is to determine the reliability of this dataset through exploring its representativeness. This will be done by analyzing the distribution of responses within the survey records, as well as comparisons with Canadian national averages for industry classifications, establishment locations, and other characteristics. Comparisons will also be drawn with the United States Commodity Flow Survey (CFS) results, as some comparisons can be made between the two countries' economic structures. The remainder of this paper will be organized as follows: the next section will provide a background literature review concerning survey data representativeness and collection of data through business establishment surveys. The following section will outline the data and study area, and

[1]Presented at the 54th Annual Meetings of the Canadian Transportation Research Forum, May 26-29, 2019 at Vancouver, BC

summarize the analysis undertaken for the dataset. Next, the results of the analysis will be discussed, followed by a final section summarizing conclusions that can be drawn and recommendations for next steps.

Background

Freight fluidity is a concept that refers to the ease with which freight flows can traverse the network between origin and destination regions. The concept of freight fluidity became of increasing interest within Transport Canada, who developed a set of indicators to quantify the reliability of the nation's supply chain networks (Gregory and Kwiatkowski,2011).Truck movement data was obtained and processed to create a database, including trip origin and destination (OD) locations, travel time between OD pairs, travel distance, and stops. The data could then be filtered toquantify various characteristics of truck trips. The concept of freight fluidity has subsequently been adopted in the United States, building on the Canadian implementation, as a measure of the overall performance of the supply chain system with respect to freight (Turnbull, 2014). Measures of fluidity include the reliability and variability of travel times for various commodities, identifying critical portions of the network (such as the locations of bottlenecks), and estimates of border crossing times.Similar principles may be applied to the dataset under study in this work to identify the relevant measures of network resilience, robustness, and criticality that impact freight fluidity.

As mentioned, the kind of micro-level analysis proposed for this work is quite data intensive, both in terms of cost and time, and one of the potential methodologies for data collection, which was employed in this case, is through a web-based establishment survey. The implementation of such surveys has been investigated since the advent of the internet and researchers have noted possible sources of errors, such as low response rates and nonresponse bias. Cook et al (2000) examines the characteristics of early web surveys compared to other survey methodologies, including telephone and paper-and-pen surveys. It was noted that a survey with a low response rate still has the potential to produce accurate results if the sampled population is representative. By examining the characteristics of a set of electronic surveys the authors concluded that higher response rates are related to the number of contacts, personalized contacts, and pre-contacts. Researchers were cautioned to ensure that the survey responses are free from nonresponse error. This is precaution is of general concern for all surveys, regardless of the mechanism used for data collection (web or otherwise). Wilcox et al (1994) state that the representativeness of a survey sample must be assessed in light of the objectives of the survey. A more recent assessment, explicitly of web-based freight surveys, notes that biasdue to low response rates is avoided when the characteristics of respondents and non-respondents are not significantly different (Samimi et al, 2013). The aim of this paper, therefore, is to analyze the responses from the web-survey to ensure that the characteristics of the respondents are representative of the national population of firms so that such biases do not arise.

Although not as common as more aggregate data collection methods, the use of establishment surveys to collect micro-level freight data has begun to increase in recent years, as researchers become more interested in micro-level analysis. One pioneering study in the province of Alberta (Hunt et al, 2006) analyzed commercial vehicle movements at the establishment level in the cities of Calgary and Edmonton through an extensive business establishment survey.The collected data was used to develop a tour-based microsimulation model of freight movements. The data collected included commodity information and details of freight vehicle activities. The results obtained from the survey were scaledbased on industry classification, employment size, and geographical location,in order to ensure that the combination of characteristics for each establishment would be representative of the study area. Further validation of the dataset was done through a simulated network analysis of the flows, to ensure that the freight flow information provided through the survey responses corresponded to the actual flow conditions present on the network.

In a similar vein, a tour-based survey has been implemented to collect urban freight data from establishments for informing policy (Ambrosini et al, 2010). The authors note that conventional urban data collection methods, such as loop detectors or roadside surveys, none can capture an exhaustive description of the characteristics of urban freight movements. To this end, establishment-based surveys are considered to bemore effective. It is noted that appropriate sampling must be undertaken so that the survey results can be accurately weighted for expansion to represent the full population under examination. The survey data should collect data of sufficient detail to enable good quality simulations and forecasts and inform policy. Establishment surveys have been noted to have low response rates (Samimi et al, 2013) due to the sensitive nature of the information being collected and the high value of time of the company officials who would be required to answer the survey.Data on establishment characteristics as well as detailed information on recent shipments is generally included in such surveys.

The establishment-based shippers' survey under analysis in this study collected information from a variety of establishments throughout Canada. The purpose of this paper is to examine the characteristics of the respondents and the responses to ensure that the data can accurately represent the population of Canadian shippers and their freight activities. Observations will be made on the distribution of industry classifications, geographical locations, and employment size. The surveyed population will be compared to Canadian national distributions of the same characteristics to ensure representativeness. An additional source for comparison will be with the Commodity Flow Survey (CFS) in the US. This large-scale shippers' survey is well established and can serve as a benchmark for the analysis of national freight activity. Although there are differences between the economic structures of the two nations, there are also strong similarities that can allow for this kind of comparison. Also following the findings from literature, weighting factors will be proposed to ensure the representativeness of this data.

Methods of Analysis

The data under consideration in the present work was collected through a shipper's survey of Canadian firms conducted in 2016 by the Cross Border Institute (CBI) at the University of Windsor.The survey collected micro-level data on goods movements within Canada and cross-border between Canada and the US. A market research firm contacted business establishments throughout the nation and a link to the web-based survey was provided to firms engaging in shipping activities and agreeing to respond. The survey contained forms for the collection of information about the establishment (e.g. industry classification, employment size), as well as details about inbound and outbound shipments (e.g. goods classification, origin and destination locations) and their cross-border shipping activities (e.g. top trading markets, participation in border security programs).Maoh et al (2017) provides a more detailed summary description of the survey structure and an outline of general observations drawn from the dataset with respect to the collected variables.

The purpose of this work is to assess the representativeness of this sample and propose methods for ensuring that use of this sample can be applied to accurately model freight-related activities in Canada. Several comparisons were employed to this end. First, the sample of 1,168 responses was divided into ten random 10% samples. The distributions of key variables, including industry classification, employment size, and square footage, were examined and compared among the ten random samples. It is expected that a sample population that is representative of the whole will exhibit a higher level of internal consistency. Next, the geographic distribution of the entire sample population was compared to the national distribution of firms within the country. It was expected that a representative sample would follow a similar distribution. As noted in the Background section, discrepancies that may arise in this comparison can be corrected through the application of corrective weight factors to the sample population. Lastly, the responses were also compared to findings from other sources of data, including the CFS in the US and a dataset of GPS tracked freight truck flows collected through the CBI.These external sources also serve to validate the distributions that are noted in the sampled responses.

Results

As noted in the previous section, the respondent population was divided into ten random 10% samples. For brevity, comment will be made on the analysis of two of the variables under examination, namely, the employment size (Figure 1) and establishment age (Figure 2). As expected, some variation is notable among each of the categories of comparison, between the ten random samples. However, a trend does appear in each case. In the case of the full surveyed population, the most prevalent employment size reported was between 1 and 5 employees, followed by the categories for 51 to 200 employees and over 200 employees, respectively. The same general trend is observed across the comparison in Figure 1 below. The variation observed does point to the fact that the sample is not perfectly representative. A further analysis and comparison will be done with respect to employment size at the national level, as this work continues.

Figure 1: Distribution of Employment Size for 10 Random Respondent Samples

With respect to establishment age, Figure 2 also shows a clear trend among the random sample distributions. Firms that reported being at their current location for over 20 years comprise the largest proportion, and this proportion decreases with establishment age. This pattern follows what is observed for the full surveyed population. Again, some variation is noted among the categories; minor discrepancies, however, can be corrected through the application of weights to the respondent records.

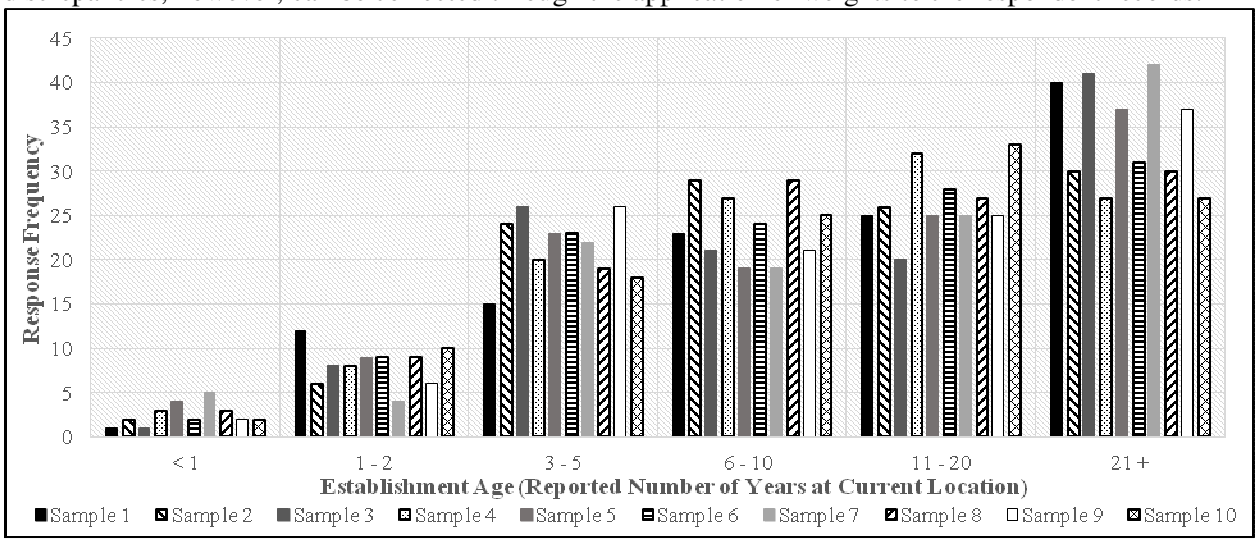

Figure 2: Distribution of Establishment Age for 10 Random Respondent Samples

Table 1 shown below summarizes the geographic distribution of the survey respondents throughout Canada in 2016 (for a close comparison to the survey year) side-by-side with the same distribution among the entire population of firms registered nation-wide. As can be noted, Ontario comprises the majority share of firms, with British Columbia, Alberta, and Quebec also making up a significant portion. One disparity that can be noted is the under-representation of the province of Quebec, which is expected to be a major contributor to economic activity and freight transportation. This underrepresentation can be attributed to the survey tool, which was administered exclusively in English; it is plausible that some Francophone establishment representatives neglected to fill the survey for this reason. To overcome this limitation, the data can be adjusted using weight factors, in order to ensure an accurate representation of the geographic distribution of establishments and account for any under- or over-representation. The data will also be compared in detail with other data sources, such as the Canadian Freight Analysis Framework database, to ensure that the freight flow conditions modeled using this sample data will correctly depict the current conditions within Canada and will be able to produce accurate predictions for future conditions.

Table 1: Distribution of Respondent Establishment Locations by Province

Province	National Distribution	Survey Distribution
Alberta	14.2%	12.7%
British Columbia	15.5%	15.8%
Manitoba	3.2%	3.7%
New Brunswick	2.1%	2.9%
Newfoundland and Labrador	1.7%	1.3%
Northwest Territories	0.1%	0.2%
Nova Scotia	2.5%	2.8%
Nunavut	0.1%	0.0%
Ontario	36.3%	49.6%
Prince Edward Island	0.5%	0.4%
Quebec	20.3%	8.4%
Saskatchewan	3.4%	2.2%
Yukon	0.1%	0.1%

One important characteristic obtained through the survey was the self-reported industry classification of each establishment, given by its NAICS code at the 2-digit level. Table 2 presents a comparison between the national distribution of industry classificationsobtained from CBP and that reported through the survey.Some deviations from the national distribution can be noted in a few industry classes, such as NAICS groups 11, 23, and 53, which appear under-represented in the sample, and NAICS groups 44-45 and 48-49, which appear over-represented. Before discarding the survey sample as poorly representative, it is important to consider the types of firms that participated in the survey, namely, those who engage in freight shipping. When examining the national population of firms, it is to be expected that not all industry groups will necessitate or participate in freight transportation. These would have been screened out of the survey where the firm representative would have responded negatively when asked whether the firm engaged in any such activity.

As such, the survey respondents are more closely representative of the subset of all firms which encompasses shippers. This becomes apparent when examining a GPS dataset of truck trips obtained by the CBI, which contains trip tracking data obtained from on-board GPS units on approximately 30,000 Canadian-owned trucks, corresponding to about 580 freight carriers (Gingerich et al, 2016). Table 3 summarizes the industry categories of the firms participating in the latter mentioned shipping activities. Evidently, the Retail Trade and Transportation sectors exhibit the largest share of truck trips. Conversely, the industries for Agriculture and related activities, Construction, and Real Estatecomprise relatively

smaller shares. This serves to show that, when considering the population of firms engaged in shipping activities, the survey sample does follow a similar distribution and can be considered representative.

The two rightmost columns of Table 1 also provide a breakdown of the proportions of freight activities undertaken in each industry class for establishments who report engaging in cross-border movements to the US compared to those who donot (whose freight activities are only within Canada). An important element of the survey was to collect data relating to cross-border movements. This information will be useful in enhancing freight resilience and fluidity modeling efforts by informing the international trade component.Similar trends can be seen among these two distributions as to that noted for the entire surveyed population, with a few exceptions. The proportion of cross-border movements in the industries of Retail Trade and Transportation and Warehousing are higher than the domestic movements undertaken in these industries. This observation is intuitive, considering the nature of these establishments. Retail Trade firms are likely to have commerce on both sides of the border and require the transportation of their goods. The Transportation and Warehousing industry class includes freight carriers, which are often hired by firms in other sectors to transport goods in 3PL arrangements. Analysis of this dataset with respect to cross-border and domestic comparisons revealed that a larger proportion of firms who ship across the border will outsource their transportation activities to a 3PL carrier than those who ship within Canada (Maoh et al, 2016). Other variations of less significant magnitude are observed, showing slightly lower shipments across the border than domestically. This can be attributed to the nature of the respective industries, which are less likely to have reason to ship products across the border (e.g. NAICS 61, Educational Services).

Table 2: Proportions of Establishments by Industry Classification in Canada (NAICS), Including Cross Border and Non-Cross Border Movements

2-Digit NAICS	NAICS Description	National Distribution	Survey Distribution	Cross Border	Within Canada
11	***Agriculture, forestry, fishing and hunting***	8.1%	0.6%	0.8%	0.5%
21	Mining, quarrying, and oil and gas extraction	0.6%	0.0%	0.0%	0.0%
22	Utilities	0.1%	0.0%	0.0%	0.0%
23	***Construction***	9.7%	0.2%	0.0%	0.3%
31-33	Manufacturing	1.7%	2.2%	4.2%	1.3%
41	Wholesale trade	2.1%	0.3%	0.8%	0.0%
44-45	***Retail trade***	4.5%	21.7%	26.1%	20.3%
48-49	***Transportation and warehousing***	5.3%	11.8%	17.8%	9.3%
51	Information and cultural industries	1.4%	5.1%	6.5%	4.7%
52	Finance and insurance	6.0%	4.1%	3.7%	4.5%
53	***Real estate and rental and leasing***	24.4%	2.2%	1.3%	2.8%
54	Professional, scientific and technical services	12.9%	15.5%	13.8%	17.0%
55	Management of companies and enterprises	3.3%	3.6%	3.1%	4.0%
56	Administrative and support, waste management and remediation services	3.9%	1.7%	2.1%	1.6%
61	Educational services	0.9%	3.9%	2.9%	4.7%
62	Health care and social assistance	6.1%	5.2%	2.6%	6.8%
71	Arts, entertainment and recreation	1.8%	4.6%	3.4%	5.5%
72	Accommodation and food services	1.6%	2.1%	2.1%	2.3%
81	Other services (except public administration)	5.4%	9.5%	7.6%	10.9%
91	Public administration	0.0%	2.7%	1.3%	3.6%

Table 3:Distribution of Truck Trips by Origin Establishment Industry Classification (SIC), GPS Truck Data

SIC Group	SIC Codes	SIC Description	Truck Trips
A	***01 – 09***	***Agriculture, Forestry & Fishing***	1.1%
E	48 – 49	Communications, Electric, Gas & Sanitary Services	0.5%
C	***15 – 17***	***Construction***	6.1%
H	***60 – 67***	Finance, Insurance &***Real Estate***	1.5%
D	20 – 39	Manufacturing	17.6%
B	10 – 14	Mining	0.4%
J	99	Nonclassifiable Establishments	2.4%
J	90 – 98	Public Administration	1.4%
G	***52 – 59***	***Retail Trade***	18.4%
I	70 – 89	Services	16.9%
E	***40 – 47***	***Transportation***	24.6%
F	50 – 51	Wholesale Trade	9.0%

One last comparison was made between the survey responses and the US CFS data. The data table that was applicable to this analysis, Shipment Characteristics by Industry, was extracted from the preliminary data of the CFS for the year 2017. Since the survey collected for this project aggregated industry classes to the 2-digit NAICS level.Table 4 presents a summary of the CFS data for comparison.It should be noted that the CFS collects shipment information from a specific subset of firms. That is to say, not all establishments covered under NAICS 44 – 45, for example, are sampled. These results are reflective of the economic activity in the US and include those shippers who contribute most significantly within that context. For this reason, the category for NAICS 44 – 45 (Retail Trade) appears overrepresented in the Canadian data, where a wider range of establishments in this industry sector were sampled.An additional inference that can be made is that a significant proportion of manufactured goods (NAICS 31-33, which has a large share in the CFS reporting) are transported to retail establishments for distribution to clients and would be accounted for through the Retail Trade sector's activities in the context of the Canadian survey.

Table 4: Excerpt from US CFS Shipment Characteristics by Industry

NAICS	Value ($ Millions)	Tons (Thousand)	Ton-Miles (Million)
212	96,671	2,966,494	819,648
31-33	5,672,939	4,634,436	1,456,460
42	6,392,775	4,387,750	749,333
44-45	570,852	60,357	16,418
48-49	1,397,621	305,757	61,914
51	42,719	7,435	1,140
55	193,035	116,621	25,785

Conclusion

The work presented in this paper served as a preliminary analysis of the representativeness of a survey sample dataset of Canadian business establishments and their freight transportation activities. Comparisons were made between the population of surveyed respondents and external data sources, including data on the entire population of Canadian firms as reported by Statistics Canada, data on freight movements in the US as summarized in the CFS, and data on individual truck movements by Canadian carriers obtained through GPS tracking. An analysis of the dataset itself found the responses to exhibit internal consistency when comparing ten random 10% samples.Comparison with national establishment

characteristics revealed that the geographic distribution of firms in the sample population matches the national distribution, with the exception of an under-representation for the province of Quebec, an issue that can be corrected through the inclusion of weight factors. Some discrepancy in the distribution of industry classifications between the survey respondentsand national-level data, which can be attributed to the nature of businesses that engage in freight transportation. When considering these types of firms, it is apparent that the industry sectors are distributed similarly to existing data. Given this preliminary analysis, it can be concluded that the data collected from the surveyed establishments can be applied to an analysis of freight activitiesin Canada. As cautioned by other researchers, the data should be appropriately weighted to ensure that any discrepancies that do exist are corrected and the data is able to produce accurate outputs when used in any modeling or analysis applications.

This analysis is of value to ensure that this dataset is able to be applied toward modeling and analysis efforts that will aim to quantify and qualify freight fluidity in Canada. Future work that is proposed includes converting the survey responses into truck flows and assigning these trips to the road network. Further validation can then be done by comparing the routing patterns observed from the survey observations with current freight flow conditions. Additional external sources can also be introduced for comparison and a more in-depth analysis involving additional response variables can be done to ensure that the sample population is representative.

References

Ambrosini, C., Patier, D., & Routhier, J. L. (2010). Urban freight establishment and tour based surveys for policy oriented modelling. *Procedia-Social and Behavioral Sciences*, 2(3), 6013-6026.

Bureau of Transportation Statistics. (2017). Table 5: Shipment Characteristics by Industry. *2017 CFS Preliminary Data*. https://www.bts.gov/sites/bts.dot.gov/files/Table%205_Industry.xlsx

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, 60(6), 821-836.

Gingerich, K., Maoh, H., & Anderson, W. (2016). Big Data Analysis to Measure Delays of Long-haul Truck Trips. In *TAC 2016: Efficient Transportation-Managing the Demand-2016 Conference and Exhibition of the Transportation Association of Canada.*

Gregory, A. and Kwiatkowski, K. (2011). The fluidity of the Canadian transportation system: A commercial trucking perspective. *Canadian Transportation Research Forum Proceedings*, 37-51. Gatineau, QC.

Hunt, J., Stefan, K., & Brownlee, A. (2006). Establishment-based survey of urban commercial vehicle movements in Alberta, Canada: survey design, implementation, and results. *Transportation Research Record: Journal of the Transportation Research Board*, (1957), 75-83.

Maoh, H., Khan, S., & Anderson, W. (2017). Cross-Border Impediments Facing Canadian Shippers Trading with US Markets: Insights from a Recent Survey. *TRB 96th Annual Meeting Compendium of Papers*(No. 17-03926).

Samimi, A., Mohammadian, A., & Kawamura, K. (2013). A nationwide web-based freight data collection. *Canadian Journal of Civil Engineering*, 40(2), 114-120.

Statistic Canada. (2016) CANSIM Table 33-10-0041-01 Canadian Business Counts, with employees, December 2016. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3310004101.

Turnbull, K. F. (2014). Developing Freight Fluidity Performance Measures: Supply Chain Perspective on Freight System Performance. Summary of a Workshop, May 21-22, 2014, Washington, DC. *Transportation Research Circular*, (E-C187).

Wilcox, J. B., Bellenger, D. N., & Rigdon, E. E. (1994). Assessing sample representativeness in industrial surveys. *Journal of Business & Industrial Marketing*, 9(2), 51-61.