

Proceedings of Statistics Canada Symposium 2024: The Future of Official Statistics

A New Origin-to-Destination Table of Canadian Manufacturing Sales: Challenges with Imputing a Distribution from Annual Survey Data

by Nicholas Huliganga and Sébastien Landry

Release date: September 8, 2025



A New Origin-to-Destination Table of Canadian Manufacturing Sales: Challenges with Imputing a Distribution from Annual Survey Data

Nicholas Huliganga and Sébastien Landry¹

Abstract

Until now, detailed data on the destination of manufacturing sales have not historically been available to Canadians. Through integration of annual survey data, a destination of sales table by industry and province of origin was developed for the annual and monthly manufacturing surveys at Statistics Canada.

Respondents for the annual survey are asked for their distribution of sales as a percentage across 15 destinations. To tackle the difficulty of generating an establishment-level distribution for multi-province respondents, three approaches were compared: using the respondents' total distribution for all their establishments, using optimization, and using the distributions of the single-province respondents. The imputed distribution of destination sales from the annual data was then applied to the monthly survey's sales value.

This paper delves into challenges faced with imputing the destination sales (especially for respondents with establishments in multiple provinces), ensuring sales match marginal origin province totals, and allocating a distribution of destinations based on data from the annual program to the monthly estimates.

Key Words: Data integration; Raking; Imputation; Manufacturing.

1. Introduction

An origin-destination of sales table gives a picture of the distribution of sales going from each origin to a set of disjoint destinations. This type of information was not previously available for manufacturing sales in Canada until July 2024 when the new tables were published online (Statistics Canada, 2024a). Producing these tables using data from Statistics Canada's Annual Manufacturing Survey posed several challenges, particularly in imputing distributions for certain respondents. Three different imputation methods were explored to address these challenges. Future plans to publish a monthly version also involve certain considerations due to important differences between the annual and monthly manufacturing surveys.

2. Manufacturing Surveys

The two surveys necessary to gather the data for an origin-destination of sales table were the Annual Survey of Manufacturing and Logging (ASML) (Statistics Canada, 2023) and the Monthly Survey of Manufacturing (MSM) (Statistics Canada, 2024b). Point estimates and quality indicators were published for both surveys for the North American Industry Classification System (NAICS) codes 31-33, which is the manufacturing sector, including food manufacturing (311), wood products (321), machinery (333), and so on. These published estimates only provide information about the geographical origin of sales; no information was published about the destination of sales. For example, the two surveys published how much was sold by manufacturing establishments in Ontario, but not how much of that was sold to destinations like Quebec or Manitoba.

In the annual survey (ASML), there was a question in which respondents report their sales across 15 destinations as a percentage. The destinations included the 13 provinces and territories of Canada, as well as the USA and "Other",

¹Nicholas Huliganga, Statistics Canada, 300 W Georgia St, Vancouver, BC Canada, V6B 6B1 (nicholas.huliganga@statcan.gc.ca); Sébastien Landry, Statistics Canada, 150 Tunney's Pasture Ottawa, Canada, K1A 0T6 (sebastien.landry@statcan.gc.ca)

representing the rest of the world. This type of question did not exist at all on the MSM, so the only way to produce a monthly origin-destination table was to integrate annual data in some manner.

The responses received from the ASML could be used to estimate information about the proportion of sales going to each destination at the Canada level. But, to create a final 13x15 origin-destination annual table of the 13 provinces and territories of Canada as the origin and the 15 destinations, information needs to be at the level of each individual manufacturing establishment. This allocation from the respondent-level to the establishment-level was difficult for respondents of the annual survey who had establishments in multiple provinces.

3. Multi-Province (M-P) Respondents

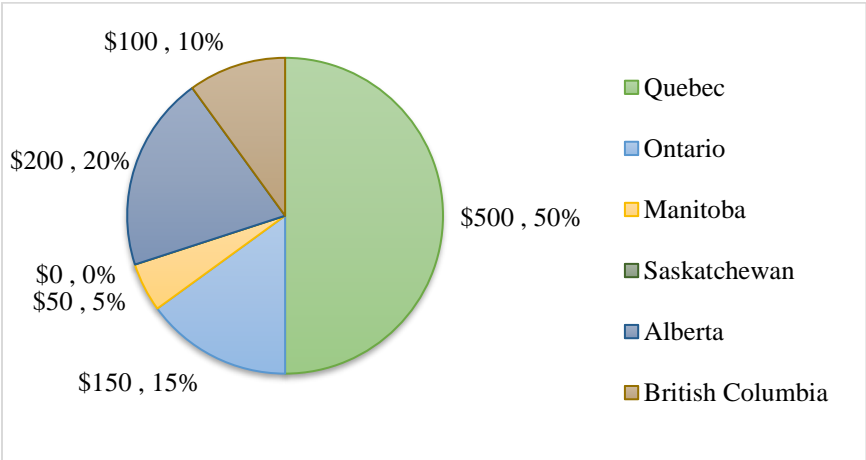
3.1 Description

Manufacturing respondents of the ASML are sampled from a frame consisting of a list of enterprises on Statistics Canada’s Business Register (BR). They can comprise multiple establishments across multiple provinces/territories. Therefore, they can be categorized as either single-province (all establishments reside in the same province/territory; S-P) or multi-province (establishments reside in multiple provinces/territories; M-P). To estimate the aggregate totals for the annual tables, establishment-level distributions were required instead of the overall respondent-level distributions reported on the ASML.

For establishments under S-P respondents, the overall distribution could be simply copied to the individual establishments. On the other hand, for M-P respondents it may not have been valid to obtain the establishment-level distributions this way. For example, if an M-P respondent had establishments in Ontario and British Columbia (BC) and reported 60% of their sales were destined for Ontario, it may not have been correct to assume their establishments in BC were sending 60% of their own sales to Ontario.

Three different approaches were explored to impute a distribution for these M-P respondents. A fictitious example of an annual respondent can help illustrate the differences between the approaches. This example respondent had a total sales value of \$1,000 and was multi-province with establishments in Quebec (totaling \$780) and Alberta (totaling \$220). Additionally, they reported an overall distribution for destination of sales as percentages (Figure 3.1-1).

Figure 3.1-1
Distribution for destination of sales for a fictitious respondent.



3.2 Approach 1

The most basic and least desirable approach to imputing distributions for M-P respondents was to treat them as if they were S-P respondents and to copy the overall distribution to each establishment. Results from this approach were often incongruent with a general understanding of how manufacturing establishments operate. Specifically, establishments are more likely to sell to their own province or the ones closest to them. With approach 1 though, the fictitious respondent was imputed to have the establishments in Alberta send more than double their sales to Quebec than stay in Alberta (Table 3.2-1).

Table 3.2-1
Approach 1 destination of sales for example respondent.

	Destination of Sales (\$)						
	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia	Total
Respondent	500 (50%)	150 (15%)	50 (5%)	0	200 (20%)	100 (10%)	1000
Origin:							
Quebec	390 (50%)	117 (15%)	39 (5%)	0	156 (20%)	78 (10%)	780
Alberta	110 (50%)	33 (15%)	11 (5%)	0	44 (20%)	22 (10%)	220

3.3 Approach 2

To better account for the distances between origins and destinations when imputing a distribution for M-P respondents, the second approach utilized optimization. Specifically, the cost of selling from an origin to a destination was minimized, with cost defined as the average distance between each origin and each destination. The distance for each origin-destination pair was calculated using the average distance between the postal codes of the origin and the destination, with the postal code being geocoded. Each respondent used the same overall cost-distance matrix, which included the 13 origins and 15 destinations, for the optimization process.

The optimization also incorporated constraints to ensure the marginal totals were preserved for each respondent. This was especially important for the column (destination) totals such that the final establishment-level distributions would sum back to the reported overall distribution for each respondent.

One final adjustment before the optimization procedure was to maximize intra-provincial sales within the marginal constraints. For the example respondent, the maximum amounts of \$500 and \$200 were allocated to the Quebec-to-Quebec and Alberta-to-Alberta pairs, respectively (Table 3.3-1). This step was necessary because, without it, the optimization could sometimes allocate all the sales to a destination as originating from a different province, even when establishments in the destination province had nonzero sales. The general and imprecise nature of the distance measure was the cause of this. For instance, the optimization might avoid sending sales from a province like Quebec a long and costly distance to BC, choosing instead to allocate those as originating from Alberta. This could result in none of Alberta's sales staying in Alberta, which is highly unlikely according to our prior understanding of how manufacturing establishments tend to operate.

Table 3.3-1
Approach 2 destination of sales for example respondent.

	Destination of Sales (\$)						
	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia	Total
Respondent	500 (50%)	150 (15%)	50 (5%)	0	200 (20%)	100 (10%)	1000
Origin:							
Quebec	500 (64.1%)	150 (19.2%)	50 (6.4%)	0	0	80 (10.3%)	780
Alberta	0	0	0	0	200 (90.9%)	20 (9.1%)	220

3.4 Approach 3

In approach 3, establishment-level imputation was based on the aggregate distributions of the S-P respondents. For each establishment under a M-P respondent, their initial distribution across 15 destinations (expressed as percentages) was calculated from the aggregate totals of the S-P respondents. Specifically, the establishments under S-P respondents in the most similar NAICS-Origin Province group with at least five units.

An initial set of 15 factors corresponding to each destination was calculated by summing the allocated sales going to each destination of the S-P respondents in the group and dividing by the sum of the total sales of the S-P respondents in the group. For example, if the 5 S-P respondents in a group sent \$50, \$150, \$0, \$0, and \$200 to Quebec and their total sales values were \$1000, \$500, \$50, \$50, & \$2000, then the initial factor for the Quebec destination factor would be 400/3600 or 0.11. Then, for each M-P respondent, the destinations with zero sales had their factors set to zero. For instance, even if S-P respondents in their group sold to Saskatchewan the example respondent would have their Saskatchewan destination factor set to zero.

A raking process was then performed to ensure the marginal totals were preserved similar to approach 2. Final results tended to be similar to approach 2 with the intra-provincial sales higher than any other destination. For the example respondent, approach 2 and 3 were quite similar with the Quebec-to-Quebec and Alberta-to-Alberta values being the highest for each origin province (Table 3.4-1).

Table 3.4-1
Approach 3 destination of sales for example respondent.

	Destination of Sales (\$)						Total
	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia	
Respondent	500 (50%)	150 (15%)	50 (5%)	0	200 (20%)	100 (10%)	1000
Origin:							
Quebec	500 (64.1%)	150 (19.2%)	10 (1.3%)	0	20 (2.6%)	100 (12.8%)	780
Alberta	0	0	40 (18.2%)	0	180 (81.8%)	0	220

3.5 Approach Comparisons

With an establishment-level distribution for the annual respondents using all three approaches, the approaches were compared by running estimation to calculate the final tables at all NAICS levels for each one. Overall, the results were similar between all three at the national level, as 87% of total respondents on the annual survey are single-province which have the same distributions for each approach. Differences became more apparent at more specific NAICS-Province levels, but in general approach 2 (distance-cost optimization) and 3 (S-P respondent aggregates) were most similar with the highest number of sales tending to stay within the same origin province, while approach 1 (copy overall respondent distribution) had sales more spread out. For example, this was observed in NAICS 311 (food manufacturing) for 2021 ASML data (Figure 3.5-1).

After consulting with subject-matter experts of the ASML, approach 3 was chosen as the preferred imputation method. The choice was made due to the use of actual respondent data for approach 3, whereas approach 2 relied on a general distance measure involving postal codes and assumptions that are harder to validate. While we cannot definitively prove the results from approach 3, it was considered a more reliable option.

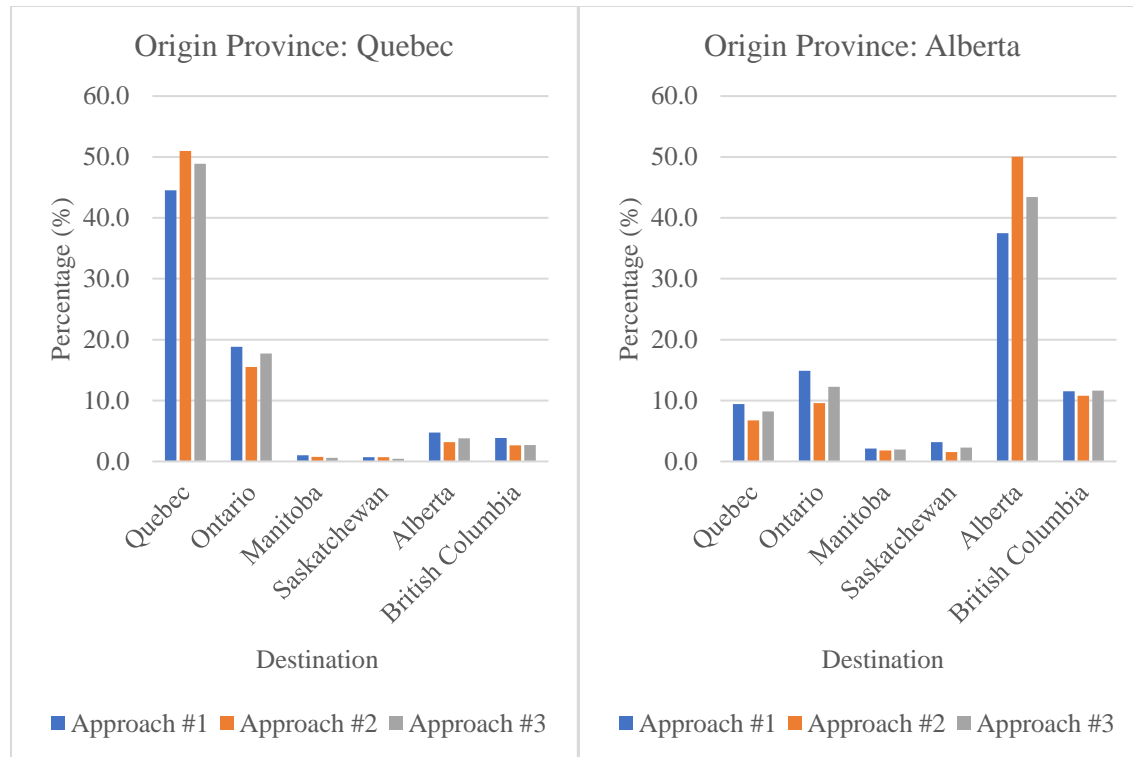
4. Annual Survey Estimation

ASML estimation was simple once each establishment in the sampled portion had its sales distributed to the 15 destinations (from the original question on the survey or by imputation). The values for the 15 destinations could be easily treated as new variables with estimates generated for each NAICS-Origin Province level that existed for the ASML. The weights for the new variables were chosen to be the same as the original sales value for each establishment. Finally, the new variables would be labelled as imputed if: the overall sales value was imputed, the

response for the percentage distribution of sales across the 15 destinations was imputed, or if the establishment was under a multi-province respondent. The final tables were successfully published online on Statistics Canada’s website for annual data from 2020 to 2022.

Figure 3.5-1

Comparison of sales proportions for 2021 ASML data (NAICS 311).



4.1 Take-None Portion

To generate the final estimates, the take-none (non-sampled) portion of the population also needed values for the new 15 destination variables. Since this portion represented the smallest establishments of the population, it was decided to allocate all the sales of these establishments to the same origin province. For example, a take-none establishment in Quebec would be assumed to have 100% of its sales destined for Quebec.

4.2 Quality Indicators

Obtaining a quality indicator was also simple since the 15 destination variables were treated the same as the overall sales value. Thus, the quality indicator for each cell in the 13 by 15 table was derived the exact same way as the overall sales, being based on two dimensions: weighted response rate and coefficient of variation (CV).

5. Monthly Survey Testing

Current work is underway to produce a similar version of the origin-destination tables for the monthly survey of manufacturing. Unlike the ASML, the MSM does not include a question where respondents give information about

the destinations of their sales. Therefore, annual data must be integrated to impute destination information for each establishment on the monthly survey.

5.1 Challenges

Two issues when considering the use of annual data for the monthly survey are that establishments may not necessarily exist in the sampled portion of both surveys, and there are publication timeline differences between the two surveys.

Another important consideration is that without any calendarization, annual data would be the same for each of the 12 months in a single year.

5.2 Integrating Annual Data

Of course, some monthly establishments exist in the annual survey, and therefore can copy the same establishment-level distribution for the 15 destinations that they received for the ASML. The rest of the monthly establishments received a distribution using a method similar to approach 3. For each of these establishments, the annual aggregate distribution of the most specific NAICS-Province or NAICS group with at least 5 units was used.

With a distribution for each establishment, estimation can be run using a process similar to the one that the MSM uses for the overall sales value. In testing, a year's worth of monthly tables was successfully produced. But since no calendarization process was done, the tables remained very similar from month to month, as the inputs from the annual survey were the same.

5.3 Production Timeline Differences

In general, for a specific reference period the MSM will publish data much sooner than the ASML, up to two years. For example, estimates for July 2022 would first be published in September 2022 for the MSM. On the other hand, estimates for the year 2022 would not be published for the ASML until the end of 2023 or beginning of 2024. Therefore, if the origin-destination tables were to publish on the same schedule as the MSM, it would need to integrate annual data from two years prior.

To investigate the possible effect of this, monthly tables were produced in testing for 12 months of 2021 using both 2019 and 2021 annual survey data as input. Early testing revealed differences were not large, especially at more general NAICS levels, but more investigation is necessary.

6. Conclusions

The new origin-destination of sales table including quality indicators was published in July 2024 for annual manufacturing sales from 2020 to 2022 using the annual survey of manufacturing at Statistics Canada. To do so, multi-province respondents of the ASML had their distribution of sales across 15 destinations imputed at the establishment level. The imputation approach chosen for this involved generating initial factors based on the aggregate distributions for the single-province respondents of a similar NAICS-Origin Province. It was preferable over other approaches due to its use of actual respondent data. In addition, raking was done to ensure the marginal totals (total sales to each destination and total sales of each origin) were preserved for each respondent.

6.1 Future Work

It is possible to extend the tables to the Monthly Survey of Manufacturing at Statistics Canada by integrating annual data. For monthly establishments that cannot be matched to an annual counterpart, aggregate annual distributions from a similar NAICS-Province or NAICS group could be used to impute a distribution.

The effects of publication timeline differences also may not be large, but further investigation is necessary. The use of calendarization would also be valuable to mitigate the issue of tables remaining very similar month-to-month if the same year of annual data is used for each month.

Potentially, a similar origin-destination table could also be produced for the annual and monthly wholesale trade surveys at Statistics Canada.

7. Acknowledgements

The authors would like to thank the numerous people who worked on this project for their contribution. The authors would also like to thank the reviewers for their useful comments.

References

Statistics Canada (2023, March 6), Annual Survey of Manufacturing and Logging Industries (ASML).
<https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2103>.

Statistics Canada (2024a, July 9), Table 36-10-0698-01 Origin and destination of manufacturing sales by industry, annual. doi: <https://doi.org/10.25318/3610069801-eng>.

Statistics Canada (2024b, November 22), Monthly Survey of Manufacturing (MSM).
<https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2101>.