Proceedings of Statistics Canada Symposium 2018
Combine to Conquer: Innovations in the Use of Multiple Sources of Data

Combining Census and Manitoba Hydro data to understand residential electricity usage

Christopher Duddek[1]

Abstract

Statistics Canada has received monthly Manitoba Hydro files from 2015 onwards. Since Manitoba Hydro is the electricity provider for most Manitobans, it would seem easy to obtain total annual residential electricity usage. When compared to published estimates, however, there are discrepancies. To better understand why, in this paper we modify geographic variables on the file in order to compare the Manitoba Hydro data to 2016 Census counts. The comparison allows us to pinpoint issues arising from the way the hydro files are created. We explore an approach to assessing coverage and evaluating the impact of coverage errors on estimated consumption estimates.

Key Words: Data Validation; Administrative Data; Census; Coverage.

1. Motivation for the study

The idea of replacing survey data with administrative data is a part of the strategic vision at Statistics Canada (Cloutier, 2010). For this reason, we are trying to obtain hydro data from all of the major electricity providers across Canada. But what is the quality of this incoming data?

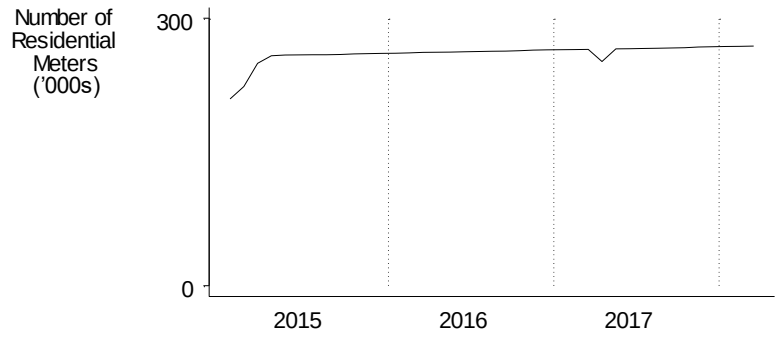
To begin the analysis of the Manitoba Hydro data, we have the luxury of being able to compare our estimates to already published data. One of the outputs of the Electrical Supply and Disposition Survey is the total residential consumption of electricity (Table 25-10-0029-01, formerly CANSIM 128-0016). If we compare those results to a summation of the kilowatt hours (kWh) for residences on the hydro file, we see that the hydro file estimates are lower for 2015 and 2016 (by 16% and 8%, respectively).

We know that neither source is perfect. On one hand, survey estimates come with sampling error and non-sampling error such as that associated with non-response. On the other hand, administrative data can have misclassification and coverage error. As an example, Figure 1-1 shows the monthly count of the number of meters in Winnipeg on the hydro file. At the beginning of the series and in one month in 2017, there are fewer meters than we would have expected, assuming that the number of meters ought to be practically constant although increasing slightly over the years.

Given that there are hundreds of thousands of electricity meters in Manitoba, how can we evaluate coverage? What follows is one attempt to break down the problem into smaller, more manageable components.

Figure 1-1
Monthly number of residential meters in Winnipeg

[1]Christopher Duddek, Statistics Canada, R.H. Coats Building, 16th Floor, Section B, Ottawa, Canada, K1A 0T6 (Christopher.Duddek@canada.ca)



2. Benefits of GPS Coordinates

If we are to look at the number of households in Manitoba at the time of the 2016 Census and compare it to the number of meters on the Hydro file in May 2016, we observe 10% fewer meters than households. How do we resolve these differences? Part of the answer lies in focusing on lower levels of geography.

Besides having address information, the Manitoba Hydro file provides Global Positioning System (GPS) coordinates for almost all meters. This incredible level of precision is an opportunity to assign Census defined geographic classifications to each meter. In this case, we wanted to go to one of the lowest levels of geography for the 2016 Census of Population, namely the Dissemination Area (DA). In Manitoba, there are just over 2,000 DAs. Using ArcGIS, we were able to assign most electrical meters to that level. For the rest, we used different approaches, such as using an associated postal code, to assign the meter to a DA.

With all meters assigned a Census geography, we were able to compare the number of meters from the hydro files to the number of households enumerated by the Census. What we observed was a tendency for there to be one meter for the dwellings that were classified as single detached dwellings. As long as there was only one household in that dwelling, we then had a 1:1 ratio of meters to households.

Apartments were another story. Sometimes, each unit had its own meter. Other times, however, there were only one or two meters servicing a multi-unit apartment. Moreover, these meters servicing multiple units were sometimes classified as commercial rather than residential meters. What this implied was that we needed a way of evaluating the meters versus dwellings at low levels of geography.

3. Validating Sampled Dissemination Areas in Winnipeg

In order to assess the coverage of the hydro file, we decided to concentrate our efforts on the city of Winnipeg. We did this because Winnipeg, in addition to having more than half of the Manitoba population, is an urban area where it is easy to verify addresses and the type of buildings by using Google Maps.

Figure 3-1
Selected sample (large centroids) of dissemination areas making up Winnipeg

Out of 1115 DAs that made up Winnipeg, we selected 27 at random (Figure 3-1). We then proceeded to resolve the breakdown of dwelling types, DA by DA.

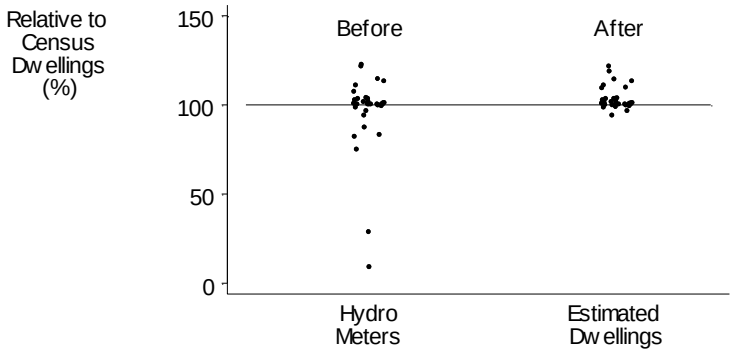
For each DA, we obtained counts of Census dwellings. The key dwelling types for this study were (1) single detached dwellings; and (2) apartments. The 2016 Census was our source for knowing how many dwellings there were for each dwelling type. If the total number of dwellings reported by the Census was close to the number of residential hydro meters, then we assumed that we had good coverage. Where it was not, we proceeded to go through a verification process.

The first step in verification was to look at the hydro meter data and see if any of the residential meters were businesses by looking at the name associated with the meter. If the name was that of a business, we changed the meter's classification to a commercial meter.

The second step was to look at the Census counts by dwelling type and then assign a dwelling type to each meter. The strategy was a manual exercise which consumed a lot of time but could be accomplished by strategizing depending on the distribution of dwellings within a DA.

If, for example, the preponderance of dwellings were single detached dwellings with a few apartments, we would search for the apartments and then find the meters that matched those addresses. We did not try a direct record linkage between the hydro files and the Census. Instead, we looked at meters with high consumption patterns and looked at the name associated with the meters to see if it was an obvious apartment or we would see if the Google Maps street-view image at that address was an apartment.

Figure 3-2
Resolution of number of meters to estimated dwellings



Although this was sometimes difficult to do, we found we could achieve our objectives by putting in enough effort. Occasionally, we would find that the meter that serviced an apartment was classified as a commercial meter. For those cases, we changed the classification to residential. In the end, we obtained a table with the hydro meters broken down by the same dwelling type classification as found on the Census.

The third step was to map the meters to an estimated count of dwellings. For single detached dwellings, we assumed that there would be one main meter per dwelling. More interesting was the mapping for apartments. If there were multiple meters for an apartment, we assumed that each of the units would have its own meter. On the other hand, if there was one meter for an apartment, it was back to Google Maps to estimate the number of units the meter served. In the end, we either assigned a meter a 1:1 or a 1:N ratio of meters to dwellings.

The last step was to calculate the sum of estimated dwellings that the hydro meters served and to compare it to the Census counts. Figure 3-2 shows the "before" and "after" distributions of hydro meters and estimated dwellings relative to the number of Census dwellings for the sampled DAs. It is thanks to this resolution process that that we are confident that we have good coverage of the dwellings that were counted by the 2016 Census. If it were not such a labourious process, we would have applied the same verification procedure to the rest of the DAs in Winnipeg. The results from the sample that we drew were comforting.

4. Estimation of Totals

Lavigne and Nadeau (2014) use a framework for determining the fitness-for-use of administrative data. One of the dimensions of quality they mention accuracy, recognising that errors lead to measurement error. Having drawn a sample and assessed it, we can proceed to evaluating accuracy.

We know that our estimation of electricity consumption is going to be affected by misclassification of meters. First, there are the meters which are classified as residential but that actually service a business. Then there are meters which provide electricity to apartment blocks but are classified as commercial. How does this misclassification impact on consumption estimates?

Figure 4-1
Impact of coverage errors on quarterly estimates of electricity consumption

The hydro data is mostly monthly data with the start and end dates uniformly distributed over the month. Calendarizing the data to monthly intervals means we have to attribute part of the monthly reported consumption to different months, which introduces measurement error (see Fortier et al, 2010). To mitigate the effect, we've calendarized the hydro data to quarterly time periods.

Going again with our sample, we calculated consumption using the original and corrected classification. This gave an indication of the bias due to misclassification. Figure 4-1 shows that the impact is modest. Part of the reason is the misclassifications cancel each other out to a certain degree. From what we can see, the quarterly differences change from 1% to 5%, depending on the quarter.

5. Conclusion

We used the GPS coordinates on the hydro files to map meters to Census geography. That allowed us to compare Census dwelling counts with estimated dwelling counts from the hydro files. We were able to say from the sample that we selected that the coverage of the hydro files is good. There is some misclassification of commercial and residential meters but it goes both ways and has a modest impact on consumption estimates.

The work presented was done to understand the quality of the coverage of the hydro files. A benefit of the work was that we have now added a covariate to a small part of the file which can be used to get an idea of how consumption changes between different types of dwellings. We observed that electricity consumption was much lower for apartments than for single detached dwellings. This will be useful for assessing outliers for a survey such as the Households and the Environment Survey.

Future work should focus on parts of Manitoba outside of Winnipeg, to assure that the rest of the file is also of high quality in terms of coverage. We might find that misclassification error is smaller if there are fewer apartments as a proportion of dwelling outside of the city of Winnipeg.

The notion that administrative records would become an integral part of National Statistical Organizations was identified at least thirty years ago (Brackstone, 1987). The reality has been that the integration alternate data sources has been slower than desired due to issues of access, privacy and fitness-for-use. The current climate of reduced response rates and higher costs has revived interest in obtaining new streams of existing data and will result in more

demands for assessing the quality of the data that we receive. Validating data by comparing to data sources like the Census will help in reaching that goal.

References

Brackstone, G. J. (1987), "Issues in the Use of Administrative Records for Statistical Purposes", *Survey Methodology*, 13, pp. 29-43.

Cloutier, M. (2010), "A Strategic vision for the use of administrative data at Statistics Canada", *Proceedings of Statistics Canada Symposium 2010*, pp. 432-436.

Fortier, S., B. Quenneville, and F. Picard (2010), "Calendarization methods and applications", unpublished documentation prepared for Statistics Canada's Advisory Committee on Statistical Methods, Ottawa, Canada: Statistics Canada.

Lavigne, M., and C. Nadeau (2014), "A Framework for the Evaluation of Administrative Data", *Proceedings of Statistics Canada Symposium 2014*.