Small Area Estimation for Visitor Travel Survey

The Visitor Travel Survey (VTS) provides a full range of statistics on the volume of international visitors to Canada and detailed characteristics of their trips. In recent years, there has been an increased interest in estimating sub-provincial inbound travel spending. Direct estimates of foreign travel spending can be obtained from the VTS, but they would be reliable only if the sample sizes are large enough. Therefore, a Small Area Estimation (SAE) methodology is now used to improve the quality of sub-provincial estimates, using Payment processors' (acquirer) data provided by Destination Canada. This document briefly describes this methodology.

1. Introduction

The VTS was introduced in January 2018 to replace the U.S. and overseas visitors to Canada component of the International Travel Survey (ITS). The objective of the VTS is to provide a full range of statistics on the volume of international visitors to Canada and detailed characteristics of their trips such as expenditures, activities, places visited and length of stay. The target population of the VTS is all U.S. and overseas residents entering Canada. Excluded from the survey's coverage are diplomats and their dependents, refugees, landed immigrants, military, crew and former Canadian residents.

The demand for inbound travel spending estimates at smaller geographical levels has greatly increased in recent years. Standard weighted estimates (or direct estimates) at sub-provincial levels can be obtained from the VTS. However, these direct estimates can be considered reliable as long as the sample size in the area of interest is large enough. To address this issue, a SAE methodology is used to improve the quality of sub-provincial estimates, using Payment processors' data provided by Destination Canada.

SAE methods attempt to produce reliable estimates when the sample size in the area is small. In this application of the methodology, the small area estimate is a function of two quantities: the direct estimate from the survey data, and a prediction based on a model – sometimes referred to as the indirect, or synthetic estimate. The model involves survey data from the geographical area of interest, but also incorporates data from other areas (as input to the model parameters) and auxiliary data. The auxiliary data must come from a source that is independent of the VTS, and it must be available at the appropriate levels of geography. The SAE model uses the Payment processors' data which includes a portion of credit and debit card payments made by international visitors to Canada, as the auxiliary data. More precisely, the Payment data along with the direct survey estimates, are used to derive the small area estimates. For the smallest areas, the direct estimates are not reliable and the small area estimates are driven mostly by the predictions from the model. However, for the largest areas, this is the opposite and the small area estimates tend to be close to the direct estimates.

There are two types of SAE models: area-level (or aggregate) models that relate small area means to area-specific auxiliary variables, and unit-level models that relate the unit values of the study variable to unit-specific auxiliary variables. The VTS uses an area-level model as the auxiliary information (i.e., Payment data) which is aggregated.

Section 2 describes the requirements to produce sub-provincial inbound travel spending estimates. In section 3, diagnostics used for model validation and evaluation of small area estimates are briefly discussed.

2. Area-level model

The small area estimates were obtained through the use of the small area estimation module of the generalized software G-EST[1] version 2.02 (Estevao et al., 2017a, 2017b). Three inputs need to be provided to the G-EST for each area in order to obtain small area estimates:

   i) Direct estimates $\hat{\theta}_i$, which are calculated using survey weights

$$\hat{\theta}_i = \sum_{k \in s_i} w_k y_k,$$

   where y_k represents spending by unit k in domain i, and w_k is the sampling weights assigned to unit k on the VTS sample
   i) Smoothed variance estimates , which are obtained by applying a piecewise smoothing approach on the variance estimates that are calculated using mean bootstrap weights
   ii) Vector of auxiliary variables $\mathbf{z}_i$

For the estimation of inbound travel spending, the domain of interest are defined as: 11 country / country groups $\times$ 22 tourism regions / grouped tourism regions (M=242).

The 11 country / country groups are as follows:

Table 1: Country / country groups

Group	Country
1	Australia
2	China
3	Japan
4	South Korea
5	India
6	United Kingdom
7	France
8	Germany
9	Mexico
10	United States
11	Other countries

---

[1] The G-EST is an estimation software developed at Statistics Canada. It includes a set of programs for basic area-level models.

The 84 tourism regions are grouped into 22 domains, as shown in the following table.

Table 2: Tourism region / Grouped tourism regions

Tourism region / Grouped Tourism Regions	Tourism regions	Province/Territory
1000 (Newfoundland & Labrador)	001, 005, 010, 015, 020, **099²**	Newfoundland and Labrador
1100 (Prince Edward Island)	101	Prince Edward Island
1200 (Nova Scotia)	202, 206, 211, 215, 220, 225, 232, **299**	Nova Scotia
1300 (New Brunswick)	300, 302, 304, 308, 318, **399**	New Brunswick
2400 (Rest of Quebec)	401, 405, 410, 420, 425, 430, 435, 440, 445, 450, 455, 465, 470, 475, 480, 485, 491, 492, 493, 495, **499**	Quebec
0415 (Quebec)	415	
0460 (Montreal	460	
3500 (Rest of Ontario)	502, 511, 516, 526, 531, 536, 541, 551, 556, 560, 565, 570, **599**	Ontario
0506 (Niagara Falls and Wine Country)	506	
0521 (Greater Toronto Area)	521	
0546 (Ottawa and Countryside)	546	
4600 (Manitoba)	601, 605, 610, 615, 620, 625, 630, 635, **699**	Manitoba
4700 (Saskatchewan)	701, 705, 710, 715, 720, 725, 730, **799**	Saskatchewan
4800 (Rest of Alberta)	801, 805, 810, 825, **899**	Alberta
0815 (Canadian Rockies)	815	
0820 (Calgary and Area)	820	
5900 (Rest of British Columbia)	901, 910, 920, 925, **999**	British Columbia
0905 (Vancouver, Coast & Mountains)	905	
0915 (Kootenay Rockies)	915	
6000 (Yukon)	981	Yukon
6100 (Northwest Territories)	991	Northwest Territories
6200 (Nunavut)	992	Nunavut

It should be mentioned that for the VTS, a modification of the basic area-level model, piecewise area-level model, was used. The piecewise area-level is useful when a single linear model does not provide an adequate explanation on the relationship between the variable of interest and the covariates. The area specific auxiliary variable i.e., spending from the Payment data, is partitioned into intervals and a separate line segment is fit to each interval.

3. Evaluation of small area estimates

The accuracy of small area estimates depends on the reliability of the model. It is therefore essential to make a careful assessment of the validity of the model before releasing estimates. For

---

² Not stated tourism regions are presented in bold font in Table 2. Spending in not stated tourism regions are assigned to domains which encompass multiple tourism regions.

instance, it is important to verify that a linear relationship actually holds between direct estimates from VTS ($\hat{\theta}_i$) and payment data (z_i), at least approximately.
For the VTS, diagnostic plots and tests in the G-EST are used to assess the model, and outliers are identified iteratively by examining the standardized residuals from that model.

A concept that is useful to evaluate the gains of efficiency resulting from the use of the small area estimate $\hat{\theta}_i^{SAE}$ over the direct estimate $\hat{\theta}_i$ is the Mean Square Error (MSE. The MSE is unknown but can be estimated (see Rao and Molina, 2015). Gains of efficiency over the direct estimate are expected when the MSE estimate is smaller than the smoothed variance estimate or the direct variance estimate. In general, the small area estimates in the VTS were significantly more efficient than the direct estimates, especially for the areas with the smallest sample size.

References

Estevao, V., You, Y., Hidiroglou, M., Beaumont, J.-F. (2017a). Small Area Estimation-Area Level Model with EBLUP Estimation- Description of Function Parameters and User Guide. Statistics Canada document.

Estevao, V., You, Y., Hidiroglou, M., Beaumont, J.-F. and Rubin-Bleuer, S. (2017b). Small Area Estimation-Area Level Model with EBLUP Estimation- Methodology Specifications. Statistics Canada document.

Rao, J.N.K., and Molina, I. (2015). Small Area Estimation. John Wiley & Sons, Inc., Hoboken, New Jersey.

Statistics Canada. (2017). Monthly Labour Force Survey Small Area Estimation- Documentation to accompany small area estimates. Statistics Canada document.