Calibration for Domain Totals for Business Surveys

Geoffrey Logan[1], Leon Jang[1], Michael A. Hidiroglou[1]
[1]Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, ON, K1A 0T6, Canada

Abstract
Business surveys are usually stratified by industry, geography and size. Estimation is done for domains of interest that may or may not correspond to the stratification. Auxiliary data can improve the reliability of estimated totals via calibration as follows: calibrating across a number of strata or calibrating to domains of interest. These methods have a number of advantages and disadvantages. This paper will compare these methods via an extensive numerical study illustrated using Statistics Canada's Survey of Employment, Payrolls and Hours, which provides a monthly portrait of earnings, the number of jobs and hours worked at different industrial and geographical levels.

Key Words: Auxiliary information, calibration estimation, business surveys

1. Introduction

In surveys, the population of interest is generally stratified to improve the precision of estimating certain characteristics. For business surveys, stratification of the target population usually accounts for key components such as industrial activity, geography and business size (e.g., revenue, assets, number of employees). When detailed estimates are required for several domains of interest, it is often difficult to stratify for all domains due to sampling and budgetary constraints. As such, estimation for all domains may not align with stratification levels.

The use of auxiliary information can assist with ensuring that such estimates are more reliable. Generally, this is done using calibration methods. An important factor in using auxiliary data is the determination of appropriate levels of calibration. This determination can involve several factors, including the accuracy and reliability of auxiliary information, the identification of key domains and limitations in the amount of sample in more refined domains.

In order to improve the estimates produced by Statistics Canada's Survey of Employment, Payrolls and Hours (SEPH), the current estimator is being rethought. This paper gives an overview of calibration methods being considered to produce the required domain totals as well as the results of a numerical study that was conducted to compare these methods.

Section 2 gives a brief overview of SEPH and introduces the current calibration estimator, including the drawbacks of this estimator and the goals that are desired for a new estimator. Section 3 introduces the different estimators that were part of the study, and Section 4 presents the results of the analysis that was carried out. Finally, some concluding remarks are provided in Section 5.

2. The Survey of Employment, Payrolls and Hours

Statistics Canada's SEPH (Statistics Canada, 2017) was created in 1983, providing monthly estimates on earnings, payroll jobs and hours worked at detailed industry and geography levels. Data from SEPH are used as the principal input to labour income estimates and also serve as a proxy output measure for around 15% of real gross domestic product. Among other things, SEPH data are used in the calculation of equalization payments by the federal government to provinces, in revisions to pensionable earnings and retirement savings plan contribution limits, and in contract escalation and wage rate determination purposes.

SEPH's target population consists of businesses in Canada with at least one paid employee (i.e., those remitting at least one payroll deduction form in a month). Excluded from the target population are businesses that are primarily involved in agriculture, fishing, hunting and trapping, private household services, religious organizations, international and other extraterritorial public administration and military personnel from defence services. The industry classification is based on the North American Industry Classification System (NAICS). See Statistics Canada (2012) for more details.

Data for SEPH comes from two main sources.

1) A census of administrative data is obtained from the Canada Revenue Agency (CRA), the federal agency responsible for administering tax laws for the Government of Canada. CRA uses payroll deduction forms remitted by employers to gather data on two variables on behalf of SEPH: total gross payrolls and the total number of employees for the reporting period.

2) The monthly Business Payrolls Survey (BPS) is the survey component of SEPH. Its purpose is to collect detailed monthly data on the weekly components of gross monthly payrolls, total hours and the allocation of specific information among employment categories (i.e., employees paid by the hour, salaried employees, working owners and other employees). These data are used to derive variables not available using administrative data.

   A stratified simple random sample of approximately 15,000 businesses is selected from a population of 1,000,000 businesses listed in the Business Register, a centralized frame for all businesses in Canada (Statistics Canada, 2010) that is updated on a monthly basis using survey feedback, business profiling and administrative data. The sample is stratified by province, industry and number of employees. Each month, one-twelfth of the sample is rotated out and replaced. However, some 800 businesses are permanently in sample because they represent a significant portion of employment or payroll or because they are sufficiently different from other businesses in their strata.

2.1 Current Estimation Method

Currently SEPH uses a generalized regression estimator, introduced in the most recent redesign in January 2009. This estimator, referred to as the modified GREG estimator by Rao and Molina (2015) is approximately unbiased under the sample design as the sample size increases, even if the domain of interest is small. It is considered a modified direct estimator since it uses y-values from outside of the domain.

Model groups are defined at 125 national industry levels, by three-digit (NAICS3) or four-digit (NAICS4) industry code or grouping of several four-digit NAICS codes. Regression coefficients, calculated at the model group level, are applied to the estimates of total employment and payrolls from the administrative source in order to estimate additional variables such as the average weekly hours worked or average hourly earnings for both salaried and hourly employees.

The model used is $E(y_k) = \beta_{0,y,g} + \beta_{1,y,g} AME_k$, where y_k is the response for unit *k*, and AME_k is the average monthly earnings (monthly payroll divided by monthly employment) from the administrative data source for unit *k* and model group *g*.

A feature of SEPH's modified GREG estimator is that although the modeling is done at the model group level, calibration is done for the lowest level of domains of interest. That is, the calibrated weights are calculated at the province by NAICS4 to ensure that all domain estimates for employment and pay exactly match totals coming from the CRA.

To create estimated totals, the modified GREG estimator has the form $\hat{t}_{y,d,MGREG} = \sum_{s_g} w_k \, a_{dk} \, y_k$, where $a_{dk} = 1_{dk} + (\boldsymbol{X}_d - \hat{\boldsymbol{X}}_d)^T \left(\sum_{k \in s_g} \frac{w_k \boldsymbol{x}_k \boldsymbol{x}_k^T}{c_k} \right) \frac{\boldsymbol{x}_k}{c_k}$ is the calibration factor for unit *k* and domain *d*, 1_{dk} is the domain indicator variable, $c_k = 1/\sqrt{EMP_k}$ for administrative employment EMP_k, s_g are the sampled units in model group *g* and w_k is the non-response adjusted weight of unit *k*.

2.2 Drawbacks of the Current Estimation Method

While the SEPH estimation design is efficient in providing domain-level estimates, several concerns have been expressed by data users and survey analysts when dealing with estimated variables.

Domain calibration ensures that employment and payroll totals from SEPH are coherent with those from the administrative data source for all domains. In order to achieve this, the generalized regression estimator may sometimes produce small negative calibration factors. Given that all units in a model group (at the national level) contribute to a domain estimate (at the sub-national level), it is possible that a negative estimate results. This can especially occur among estimates for specific employment categories or rare variables (i.e., overtime or special payments). In order to ensure proper interpretability of estimates, domain estimates, aside from the CRA-calibrated employment and pay totals, could be suppressed from publication when a negative estimate occurs among some BPS variables.

Calibrating for each domain requires each sampled unit to have a calibration factor for each domain within a model group. In SEPH, a single unit could have as many as 230 calibration factors. For each month, the SEPH production file of calibration factors has around 1,900,000 records.

2.3 Properties of a New Estimator

Given the drawbacks discussed in section 2.2, there is a desire to change the current estimator.

An estimator yielding a single calibration factor for each unit would produce more coherent estimates that only involve sampled units within a domain. Having a single weight would permit the creation of custom domains crossing different model groups as requested by data users, which cannot be easily handled by the existing method. Also, the estimates could be easily reproduced using the single weight available on the microdata file.

In addition, there is a desire to reduce the number of negative weights. This would result in fewer estimates needing to be suppressed due to data inconsistency.

3. Study of the Estimators

When considering estimators producing a single weight, the calibration level and the modelling level should be consistent. Compared to the current estimator, the calibration level could be brought up from the domain level to the model group level. This would result in a single-weight estimator but would be inefficient at the expense of provincial distributions that are currently produced. Second, the modelling level could be brought down from the model group level to the domain level. This would also result in a single weight estimator that maintains provincial distributions. However, this estimator would be unstable when domain sample sizes are small.

3.1 Estimators Studied

In addition to the existing SEPH calibration estimator, two alternative estimators were considered for the production of estimates. For comparison purposes, the Horvitz-Thompson estimator was also used in the study. Table 1 gives a comparison of the four estimators in the study.

Table 1: Formulae for estimators used in the study

Estimator	*Formulae*
Horvitz-Thompson (HT)	$\hat{Y}_{HT,d} = \frac{N}{n}\sum_{s_d} y_k$
$\text{GREG}_{\text{Canada}}$	$\hat{Y}_{GREG1,d} = \hat{Y}_{HT,d} + (\mathbf{X} - \hat{\mathbf{X}}_{HT})^T \hat{\mathbf{B}}_1$ $\hat{\mathbf{B}}_1 = \left(\sum_{s_g} w_k c_k \mathbf{x}_k \mathbf{x}_k^T\right)^{-1} \sum_{s_g} w_k c_k \mathbf{x}_k y_{dk}$
$\text{GREG}_{\text{Domain}}$	$\hat{Y}_{GREG2,d} = \hat{Y}_{HT,d} + (\mathbf{X}_d - \hat{\mathbf{X}}_{HT,d})^T \hat{\mathbf{B}}_2$ $\hat{\mathbf{B}}_2 = \left(\sum_{s_d} w_k c_k \mathbf{x}_k \mathbf{x}_k^T\right)^{-1} \sum_{s_d} w_k c_k \mathbf{x}_k y_k$
Current	$\hat{Y}_{GREG3,d} = \hat{Y}_{HT,d} + (\mathbf{X}_d - \hat{\mathbf{X}}_{HT,d})^T \hat{\mathbf{B}}_3$ $\hat{\mathbf{B}}_3 = \left(\sum_{s_g} w_k c_k \mathbf{x}_k \mathbf{x}_k^T\right)^{-1} \sum_{s_g} w_k c_k \mathbf{x}_k y_k$

3.2 Construction of Estimators

An empirical analysis of the four estimators given in Table 1 was performed using monthly SEPH data from January 2015 to December 2015. For the purposes of this analysis, all units identified by the existing system as being influential values were not treated prior to

producing estimates. This was done in order to study the true impact of each estimator, as the outliers are detected using the regression models associated with the current estimator. As such, estimates produced for this analysis using the current SEPH estimator differ from the final published results.

Calibration and modelling were done at the same level for each individual estimator in order to produce a single weight. That means the $GREG_{Canada}$ estimator was modelled and calibrated at the Canada by model group level, while the $GREG_{Domain}$ estimator was modelled and calibrated at the province by NAICS3 level.

4. Analysis and Results

This section compares and contrasts the two alternate calibration estimators with SEPH's current calibration estimator, as well as the HT, in terms of the number of observed negative weights, the number of observed negative estimates, the precision of the estimates produced and the observed differences between the different sets of estimates.

4.1 Negative Weights

The strongest motivation for this study, other than having only a single weight per unit, is to reduce the number of negative weights and negative estimates created by SEPH's current estimator. This subsection will compare the frequency of negative calibration factors between the current estimator and the two alternate estimators.

4.1.1 Current estimator

Currently every month, approximately 30% of the calibration factors are negative. This relatively high proportion of negative factors is a side effect of the form of the modified GREG estimator SEPH uses, which borrows strength from observations outside of the domain of interest.

The formula for the adjustment factors, a_{dk}, for the current estimator is given in section 2.1. Notice that units within the same model group, g, but outside of the domain of interest, d, will have a calibration factor of 0 plus an adjustment which could be negative. As these factors centre at 0 and there are no bounding constraints, many of them are negative. In fact, Table 2 illustrates that almost all of the negative calibration factors observed in a particular month, which is representative of other months, are from units that fall outside of the domain of interest.

Table 2: Distribution of the calibration factors produced by the current estimator, January 2015

Negative calibration factor?	*Unit inside domain of interest?*	*Frequency*	*Percent(%)*
No	No	1,230,345	65.52
No	Yes	85,066	4.53
Yes	No	562,243	29.94
Yes	Yes	206	0.01

These negative calibration factors result in negative weights and could also lead to negative estimates, which is undesirable.

4.1.2 Alternate estimators

Unlike SEPH's current modified GREG estimator, both alternate estimators are considered domain direct estimators, meaning that only observations within a domain contribute to its estimate. Thus, each unit only has a single calibration factor.

Table 3 compares the number of negative calibration factors observed for the current estimator and the two alternate estimators using the same data from January 2015.

Table 3: Percent of negative calibration factors by estimator, January 2015

Estimator	*Number of calibration factors*	*Number of negative calibration factors*	*Percent of negative calibration factors*
Current	1,877,860	562,449	29.95
$GREG_{Domain}$	12,691	489	3.85
$GREG_{Canada}$	12,691	129	1.02

From Table 3, it is clear that both alternate estimators offer a great reduction in the percent of negative calibration factors. The $GREG_{Domain}$ reduces the frequency of negative calibration factors from 30% to 4%, while the $GREG_{Canada}$ further reduces the frequency of negative calibration factors to 1%. These frequencies could be reduced even further by exploring the implementation of a lower bound on the weights in a future study.

Table 3 also serves to demonstrate the sheer magnitude of the number of weights that are created by the current estimator each month. For this particular month, each of the 12,691 responding units had an average of 148 different calibration factors: one factor for each domain within a unit's model group.

4.2 Negative Estimates

While the previous subsection detailed the frequency of negative weights, this subsection will focus on how often those negative weights can lead to negative estimates for SEPH's current estimator as well as the two alternate estimators.

One of the criteria for publishing an estimate for SEPH is that there needs to be at least five responding units inside of the domain. Table 4 looks at the frequency of there being at least one negative estimate amongst domains that have five or more respondents.

Table 4: Percent of publishable domains with at least one negative estimate, by estimator, averaged across 2015

Domain	*$GREG_{Canada}$*	*$GREG_{Domain}$*	*Current*
Province NAICS2	1.20%	7.36%	41.4%
Province NAICS3	2.29%	9.50%	38.5%
Province NAICS4	1.69%	6.75%	43.0%

With the current estimator, roughly 40% of domains at the province by NAICS2, province by NAICS3 and province by NAICS4 levels have at least one negative estimate. The frequency of negative estimates is reduced with the $GREG_{Domain}$ estimator to between 6.8 and 9.5% and is reduced even further to between 1.2 and 2.3% with the $GREG_{Canada}$ estimator.

While the percentages for the current estimator look high, it is important to make clear that the majority of these negative estimates are for rarer variables related to infrequent special payments or bonuses. These are often very small negative estimates which occur when no respondents in the domain of interest report a type of special payment, or report very little, and the estimate is dominated by units outside of the domain with negative calibration factors. However, these negative estimates do not prevent the publication of SEPH's key variable of interest, namely, the average weekly earnings (AWE).

Table 5 narrows the focus to negative estimates which prevent the publication of the AWE at the province by NAICS3 or province by NAICS4 domains. The majority of the time this is caused by negative estimates related to overtime pay or overtime hours but can also be related to hourly or salaried variables as well.

Table 5: Percent of publishable domains where the AWE is suppressed due to a negative estimate, by estimator, averaged across 2015

Domain	*$GREG_{Canada}$*	*$GREG_{Domain}$*	*Current*
Province NAICS3	1.0%	4.5%	14.0%
Province NAICS4	0.5%	2.0%	11.8%

With the current estimator, the AWE is suppressed by negative estimates in 14.0% of province by NAICS3 domains and 11.8% of province by NAICS4 domains. This is reduced to 4.5% and 2.0% with the $GREG_{Domain}$ estimator and to 1.0% and 0.5% with the $GREG_{Canada}$ estimator. As mentioned in the previous subsection, these frequencies could be reduced even further by exploring bounds on the weights in a future study.

4.3 Precision of the Estimates

This subsection focuses on the precision of the estimates produced by each of the two alternate estimators, the current estimator, as well as the HT. Often at Statistics Canada, an estimate's precision is represented by the coefficient of variation (CV). For SEPH, the CV is used to define a quality rating.

Table 6 defines the quality ratings used for SEPH's publications.

Table 6: Quality ratings used for SEPH

Quality rating	*CV range (%)*	*Description*
A	0-5	Excellent
B	5-10	Very good
C	10-15	Good
D	15-25	Acceptable
E	25-35	Use with caution
F	35+	Too unreliable to publish

Figure 1 presents the distribution of quality ratings of the AWE for province by NAICS3 and province by NAICS4 domains, across the four estimators, averaged across the 12 months of 2015.

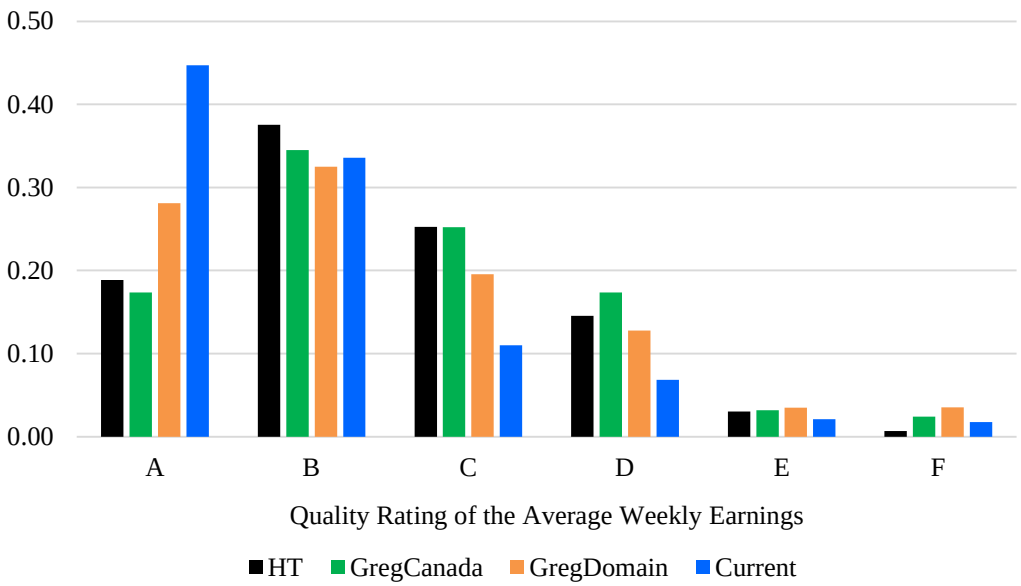

Figure 1: The distribution of quality ratings across the four estimators, averaged across 2015

From Figure 1, it is clear that the current estimator outperforms the two alternate estimators as well as the HT. This is the known strength of the current estimator. As a modified GREG estimator, it borrows strength from units across the model group, helping to build a more robust model, calibrating amongst more units and resulting in a reduced variance.

Comparing the two alternate estimators, the $GREG_{Canada}$ does not look much different from the HT. Meanwhile the $GREG_{Domain}$ shows some efficiency gains, with about 10% more of the AWE estimates having a quality rating of A than $GREG_{Canada}$ or HT.

The following two tables summarize the frequency at which provincial estimates at the NAICS3 or NAICS4 level meet their target CV.

Table 7: Percent of province by NAICS3 domains where the AWE CV meets the target, January 2015

Estimator	*Frequency of meeting the target CV*
HT	80.22%
$GREG_{Canada}$	77.86%
$GREG_{Domain}$	84.75%
Current	88.05%

Table 8: Percent of province by NAICS4 domains where the AWE CV meets the target, January 2015

Estimator	*Frequency of meeting the target CV*
HT	75.45%
$GREG_{Canada}$	72.56%
$GREG_{Domain}$	68.59%
Current	88.20%

Similar to Figure 1, the current estimator gives the most precise AWE estimates, especially for the finest domains at the province by NAICS4 level, meeting the target in almost 90% of all domains. Comparing the two alternate estimators, the $GREG_{Domain}$ outperforms the $GREG_{Canada}$ for province by NAICS3 domains, performing similarly to the current estimator. While both estimators struggle to meet the target for province by NAICS4 domains more than 70% of the time, this is not a surprise as neither of these estimators calibrate to the NAICS4 level.

Table 9 shows the average provincial CV achieved by each estimator across 2015.

Table 9: Average CVs for provincial AWE estimates, 2015

Province	*HT*	*$GREG_{Canada}$*	*$GREG_{Domain}$*	*Current*
N.L.	2.90%	4.15%	2.17%	1.53%
P.E.I.	4.05%	4.48%	1.54%	0.91%
N.S.	2.78%	2.95%	1.57%	1.61%
N.B.	2.43%	3.55%	1.61%	0.93%
Que.	2.03%	1.70%	1.12%	0.78%
Ont.	2.25%	1.86%	1.65%	0.90%
Man.	2.82%	2.68%	1.36%	1.13%
Sask.	2.87%	3.09%	1.57%	1.09%
Alta.	2.66%	2.16%	1.67%	1.21%
B.C.	2.69%	2.30%	1.56%	0.88%
Y.T.	4.07%	8.08%	1.17%	1.34%
N.W.T.	3.13%	4.36%	1.13%	1.27%
Nvt.	4.22%	4.57%	1.79%	1.92%

Again, it is clear that the current estimator outperforms the other three in the majority of provinces, with an average CV between 0.78% and 1.92%. The $GREG_{Domain}$ gives similar results, with an average CV between 1.12% and 2.17%. The HT and $GREG_{Canada}$ both have average CVs in the 2% to 4% range, aside from one spike to 8% in the Yukon.

This subsection has demonstrated the strength of the current estimator, producing the most precise provincial estimates for the AWE. The $GREG_{Domain}$ has proven to be the next most precise estimator, while the $GREG_{Canada}$ gives similar results to the HT.

4.4 Differences between the Estimates

This subsection turns the focus toward the estimates themselves. While each of the four estimators give unbiased estimates, it might be considered an advantage if the new estimator produced estimates which were more similar to our current estimates, ensuring a certain level of coherence between the two sets of estimates. Table 10 attempts to summarize this coherence.

Table 10: Percent difference in province by NAICS3 and province by NAICS4 estimates relative to the current estimator, June 2015

		Distribution of the % difference in AWE estimates compared to the current estimator		
Estimator	*Domain size*	*0-10%*	*10-20%*	*20+%*
HT	5-19	53.06	26.00	20.94
HT	20+	73.17	19.94	6.90
$GREG_{Canada}$	5-19	53.05	25.70	21.25
$GREG_{Canada}$	20+	74.68	18.29	7.03
$GREG_{Domain}$	5-19	65.66	19.77	14.57
$GREG_{Domain}$	20+	89.19	8.16	2.65

The last row of the table says that 89.19% of the $GREG_{Domain}$'s AWE estimates are within 10% of the current estimator's set of AWE estimates, when there are at least 20 respondents in the domain. Another 8.16% of AWE estimates have between a 10% and a 20% relative difference compared to the current estimates, and finally 2.65% of AWE estimates differ by more than 20% compared to the current set of estimates.

Comparing the three estimators in Table 10, it is clear that the HT and the $GREG_{Canada}$ have a very similar distribution of differences compared to the current estimator. This is because the $GREG_{Canada}$ is only calibrated to national totals.

To visualize the estimates over time, time series plots of the estimates were created for many different domains. Figure 2 gives an example of a particular province by NAICS3 domain.

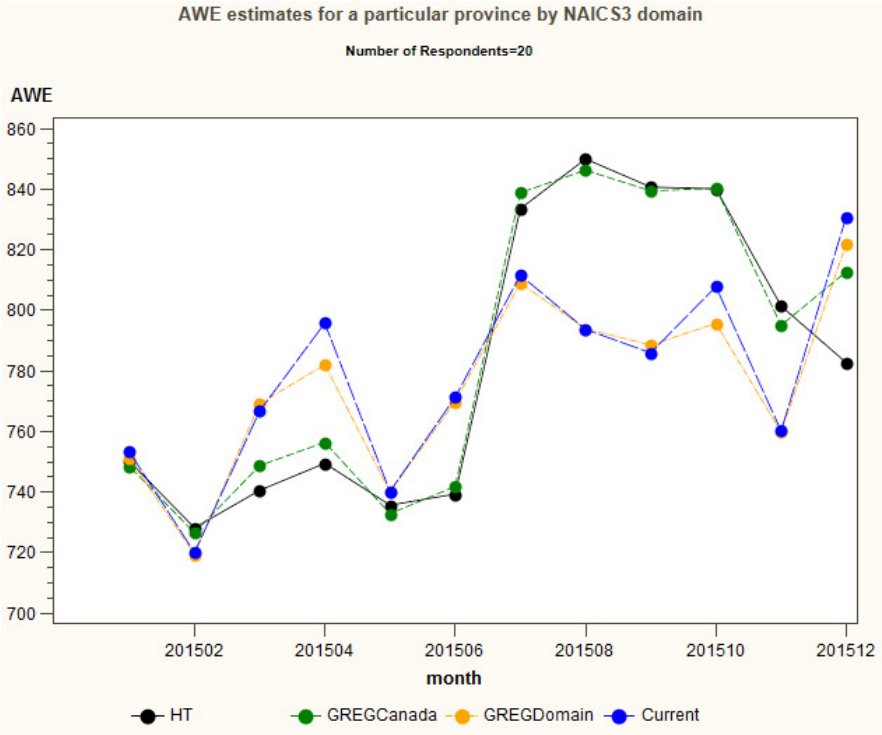

Figure 2: 2015 estimates of the average weekly earnings for a province by NAICS3 domain with 20 respondents

The time series plot in Figure 2 shows a similar story as Table 10 did earlier. The estimates produced by the $GREG_{Domain}$ tend to be more similar to the estimates produced by the current estimator, while the estimates produced by the $GREG_{Canada}$ tend to be more similar to the HT estimator.

5. Conclusion

The two alternate estimators achieve the goal of having a single weight and demonstrate a drastic reduction in the proportion of negative weights and, subsequently, the proportion of negative estimates. This can be reduced even further by exploring bounding constraints on the calibration factors.

In terms of precision, the $GREG_{Domain}$ tends to have more precise provincial estimates than the $GREG_{Canada}$, while giving similar or slightly less precise estimates at the national level.

Comparing the provincial estimates, the $GREG_{Domain}$ gives more similar results as our current estimator while the $GREG_{Canada}$ more closely resembles the HT.

Since provincial estimates are very important to users of SEPH data, the $GREG_{Domain}$ looks to be the preferred estimator. Future studies will focus on refining the $GREG_{Domain}$ estimator.

Acknowledgements

The authors would like to thank the reviewers for their time and their thoughtful feedback.

References

Rao, J.N.K., and Molina, I. (2015). *Small area estimation, Second Edition.* John Wiley & Sons.

Statistics Canada. (2010). *A Brief Guide to the BR – July 2010.* Statistics Canada, Ottawa, Canada. Retrieved from the Statistics Canada website: http://www23.statcan.gc.ca/imdb-bmdi/document/1105_D2_T1_V3-eng.pdf

Statistics Canada. (2012). *North American Industry Classification System (NAICS) Canada, 2012.* Statistics Canada catalogue number 12-501-X.

Statistics Canada. (2017). *Guide to the Survey of Employment, Payrolls and Hours, 2017.* Statistics Canada catalogue number 72-203-G.