SSC Annual Meeting, May 2008
Proceedings of the Survey Methods Section

SMALL AREA ESTIMATION TO STUDY THE IMPACT OF GLOBALIZATION

Catalin Dochitoiu, Susana Rubin-Bleuer[1]

ABSTRACT

We study the feasibility of producing relevant statistics to measure the impact of globalization in the Canadian economy. In this project we require means of key economic variables for the wholesale industry at the level of "Trade Group" by "Province" by "Globalization Indicators". We consider small area estimation using area level models with random effects and use penalized splines in order to accommodate departures from linearity. We propose a new bootstrap method to estimate the mean squared errors of the small area estimators. We illustrate the methodology with data from a particular trade group.

KEY WORDS: Bootstrap, Business data, Mean squared error, Penalized splines, Small area.

RÉSUMÉ

Nous étudions la faisabilité de produire des statistiques pertinentes pour mesurer l'effet de la mondialisation sur l'économie canadienne. Dans ce projet, les moyennes de variables économiques pour l'industrie du commerce de gros au niveau du « groupe de commerce » par « province » et par « indicateur de mondialisation » sont requises. Nous examinons l'estimation sur petits domaines en utilisant des modèles au niveau du domaine avec effets aléatoires. Afin de tenir compte des écarts par rapport à la linéarité, nous utilisons des splines pénalisées. Nous dérivons une méthode d'estimation d'erreur quadratique moyenne en utilisant une nouvelle méthode de bootstrap et nous donnons un exemple avec des données d'un groupe de commerce particulier.

MOTS CLÉS: Bootstrap; donnés des entreprises; erreur quadratique moyenne; petit domaines; spline.

1. BACKGROUND

1.1 Purpose of the study

Globalization is a major force shaping the world economies and Canadian businesses. Many policy research programs are working towards understanding how these forces influence Canadian businesses' behaviours and outcomes. However, there is a lack of basic information on the impact of globalization on the Canadian economy. There are few reliable statistics on key economic variables measuring the economic activity of Canadian or foreign controlled multinational firms. Some of the relevant statistics include intra-firm trade, trade of intermediate products and activities of Canadian multinationals abroad per industry (Gervais, 2006). A desirable goal would be to obtain these statistics for the major industries. For this purpose an integrated data warehouse was developed, combining the Unified Enterprise Survey (UES, 2008) database with other data sources like the exporter and importer registries produced by the Canadian International Trade Division (International Trade Division, 2007).

One of the main goals of the integrated database was to facilitate the production of new statistics such as totals and means at the level of Trade Group by province by international trade characteristics. Since these domain identifiers were not even in the survey frame at the time of the design, the sample is not large enough to yield direct survey estimates with an acceptable precision for many domains. In this particular project we study the feasibility of producing small area estimates in the Wholesale Industry using administrative tax data as auxiliary information.

In Section 2 we describe the survey and the auxiliary data at our disposal. We also describe the main issues with the data and the techniques we used to deal with them. In Section 3 we describe the models and approach we propose and in Section 4 we illustrate the methodology with data from a specific Trade Group within the Wholesale Industry. Finally, in Section 5 we present our conclusions and planned future work.

[1] Catalin Dochitoiu, Catalin.Dochitoiu@statcan.gc.ca 100 Tunney's Pasture Driveway, Ottawa, ON, K1A 0T6 RHC-11 D
Susana Rubin-Bleuer, Susana.Rubin-Bleuer@statcan.gc.ca 100 Tunney's Pasture Driveway, Ottawa, ON, K1A 0T6 RHC-16 K

2. CHARACTERISTICS OF THE DATA

2.1 The Canadian Annual Wholesale Trade Survey and the Administrative Tax Data

The Canadian Annual Wholesale Trade Survey (WTS) is currently administered as part of the UES program, which itself has been designed to integrate about 200 business surveys into a single master survey program. The objective of the WTS is to present timely information on the operating revenues, expenditures and inventory of wholesalers in Canada by trade group and at national and provincial or territorial levels for the previous calendar year. The WTS target population consists of all wholesale establishments operating in Canada during the reference year, while the survey frame is defined as all statistical establishments coded to NAICS 41 (North American Industry Classification System 41) that are present on the Business Register (about 110 000 establishments). Survey data are collected directly from respondents. The sample consists of approximately 14 000 establishments. For further details, see the Statistics Canada web-site.

The WTS was designed to produce reliable statistics at the level of trade group by province. The main stratification variables were province, trade group and size. There are 17 Trade Groups (TG) are assembled according to their NAICS code. Some TGs consist of a single four-digit NAICS code, for instance, the TG of petroleum product wholesalers, whereas other TGs contain many four-digit codes. For each province and trade group, the population was divided into one take-all stratum or census, that included the establishments with the largest sizes and were selected to the sample with certainty, two take some strata, comprising establishments classified according to size and selected to the sample with a sampling fraction smaller than 1, and a take-none stratum, which consisted of the smallest establishments according to a pre-set threshold, and none of them were selected to the sample. An expansion estimator (see Särndal et al, 1992) was used to derive totals and means at the aggregation levels needed.

The integrated database contains globalization indicators like MNE (whether the establishment belongs or not to a Multi-National Enterprise), TRADER (whether the establishment deals with international trade in goods, in services or in both), TIG (whether the establishment exports or imports Goods) and TIS (whether the establishments exports or imports Services) (Economic Globalization Indicators Project, 2006). These indicators were added to the database after the data was collected and processed. Therefore they could not be taken into account at the design stage.

Through an agreement with the Canadian Revenue Agency (CRA), Statistics Canada has access to data collected from the tax reports of incorporated and unincorporated businesses. For this project, we linked the integrated data warehouse with administrative data obtained from corporate and personal tax return files to be used as auxiliary data in the models proposed.

The economic variables studied were Total Operating Revenue, Total Revenue, Total Labour Expenses, Total Operating Expenses, Total Expenses, and Cost of Goods Sold. They were collected by the survey, and also have equivalents in the tax data.

For each Trade Group, the small domains (areas) were defined by a set of three characteristics. For example, the set of domains defined by PROVINCE by MNE by TRADER had a potential maximum of m=104 small domains (13x2x4). However, for some TGs many of these domains were empty. Direct estimates of means in each small area were obtained by domain estimation (see Särndal et al, 1992). Domain population sizes $N_i, i = 1,...m$ were not large and sample sizes $n_i, i = 1,..., m$ ranged from 1 to 20, depending on the Trade Group under consideration.

2.2 Issues with the data

Small area estimation is model-based, and the quality of the output relies mainly on the quality of the auxiliary data. There are often conceptual incompatibilities between the available administrative data and the variable of interest (see Nadeau, 2004).

Furthermore, administrative (tax) data are usually obtained at the enterprise level. When the enterprise consists of only one establishment (called here a simple single business), tax data are directly allocated to the establishment. A business is called complex if the enterprise includes more than one establishment. In that case, allocation of the tax information to the level of establishment is done under certain assumptions, and it is not as reliable as for simple single businesses.

As noted in Hidiroglou and Smith (2005), some problems of business surveys are exacerbated when dealing with small areas. Due to the skewness of the population distributions, estimates can be dominated by one or two units, especially when the sample size is small. Model outliers can have even larger effects when they are the dominant units. Moreover, we observed that the correlation between direct estimates of the domain (area) level and corresponding tax data aggregates was considerably higher than the correlation between unit level quantities. Hence, though auxiliary data is available at the unit level, here we used area level models with the expectation that by aggregation some of the errors in the data cancel out. Nevertheless, the nature of business data - in particular the presence of dominant units or the wrong tax data allocation – yielded some aggregate values that were outliers and/or influential observations.

During the exploratory phase we observed that for each TG, most domains that were census by design (whose units belonged to take-all strata) were outliers and/or influential observations. These observations do not need small area techniques to improve their quality, so we eliminated them outright from the modelling: they represented approximately 14% of the data points. We used two main detection techniques to identify influential domains and outliers in the domains that were not completely enumerated by design (which included some containing complex units), namely Studentized residuals and Cook's distance (see Besley et al, 1980, Chapter 2). While the first technique allowed the detection of outliers in the category of "gross errors", Cook's distance helped find data points that have an unreasonably large influence on the parameters. Setting thresholds for separating the outliers was done by trial and error, judging from the scatter-plots whether the results were consistent.

3. SMALL AREA MODELING

3.1 Framework

We model the direct estimate of the small area means (means over the all the establishments in the small area) of an economic variable as a function of the corresponding tax variable mean. We propose to use a generalization of the Fay-Herriot (1979) area level model (denoted by FH from now on) where the functional form of the response variable is more flexible. In the traditional FH model, the direct survey means $\bar{y}_i, i = 1,...,m$ are modelled as a linear function of the auxiliary variables (tax means $\bar{X}_i, i = 1,..., m$) plus a random coefficient to account for the small area effect:

$$\bar{y}_i = \theta_i + e_i, \quad \theta_i = \underset{\sim}{\bar{X}}_i'\beta + v_i, \qquad v_i \overset{\text{i.i.d.}}{\sim} (0,\sigma_v^2), \quad e_i \overset{\text{i.d.}}{\sim} (0,\psi_i), \qquad (1)$$

where $\underset{\sim}{\bar{X}}_i = (1, \bar{X}_i)'$, $\beta = (\beta_0, \beta_1)'$ is the 2×1 vector of fixed regression coefficients, v_i is the area-specific random effect assumed to be independent and identically distributed with $E_\xi(v_i) = 0, V_\xi(v_i) = \sigma_v^2$. The subscripts ξ and *p* applied to the expectation and variance symbols denote the expectation and variance with respect to the model and design, respectively. The $e_i, i = 1,..., m$, represent the stochastically independent sampling errors and are independent of the area effects. The direct survey estimator is assumed approximately design-unbiased, that is, $E_p(e_i) \approx 0$, and the design variances of the sampling errors $\psi_i = V_p(e_i) = V_p(\bar{y}_i), i = 1,..., m,$ are assumed known constants whereas the variance component σ_v^2 is generally unknown and estimated from the data via the restricted maximum likelihood method (REML).

Let $X = (\underset{\sim}{\bar{X}}_1',..., \underset{\sim}{\bar{X}}_m')'$ be the $m \times 2$ matrix of covariate means. Let $y = (\bar{y}_1,..., \bar{y}_m)'$ be the $m \times 1$ vector of direct estimates of the small area means. Let $v = (v_1,..., v_m)'$ and $e = (e_1,..., e_m)'$. Let D be the matrix of design variances $D = \text{Diag}(\psi_1,..., \psi_m)$ and let $V = V_\xi(y) = D + \sigma_v^2 I_m$. Note that since D is composed by constant elements, $E_\xi(D) = D$. If the variance matrix is known, the estimators of the regression vector β and of the vector of random effects are obtained as

$$\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \text{ and } \tilde{v} = \sigma_v^2 V^{-1}(y - X\tilde{\beta}). \qquad (2)$$

Note that for the FH model, $\tilde{v}_i = \kappa_i(\bar{y}_i - \underset{\sim}{\bar{X}}_i'\beta)$ where $\kappa_i = \sigma_v^2 / (\sigma_v^2 + \psi_i), i = 1,..., m.$

The Best Linear Unbiased Predictor (BLUP) of the small area mean θ_i, is given by a convex combination of its direct and predicted estimators: $\hat{\theta}_i = \underset{\sim}{\bar{X}}_i'\tilde{\beta} + \tilde{v}_i = \kappa_i \hat{\bar{y}}_i + (1 - \kappa_i)\underset{\sim}{\bar{X}}_i'\tilde{\beta}, i = 1,..., m.$ When σ_v^2 is estimated from the data, the estimators of the of the small area means are then called Empirical Best Linear Unbiased Predictors (EBLUPs) and are obtained by replacing σ_{v}^{2} in (2) by $\hat{\sigma}_{\mathrm{v}}^{2}$.

When the mean of θ_{i}, say $\mathrm{m}(\bar{\mathrm{X}}_{\mathrm{i}})$, is not a linear function of the tax mean $\bar{\mathrm{X}}_{\mathrm{i}}$, we use the penalized-spline method (denoted by PS from now on), which can accommodate the non-linearity and still fit within the general mixed linear model framework. The model we propose is similar to that of Opsomer et al (2008) but at the area level rather than the unit level studied by them and where the design variances $\psi_{\mathrm{i}}, \mathrm{i}=1,\ldots,\mathrm{m}$, are assumed known constants and allowed to vary with the survey estimators. The mean function $\mathrm{m}(\bar{\mathrm{X}}_{\mathrm{i}})$ is approximated by a piecewise linear function with K knots $\mathrm{k}_1 < \ldots < \mathrm{k}_{\mathrm{K}}$ plus an area effect. Let $\gamma=(\gamma_1,\ldots,\gamma_{\mathrm{K}})'$, and $\mathrm{Z}_{\mathrm{i}}=\left[(\bar{\mathrm{X}}_{\mathrm{i}}-\mathrm{k}_1)_{+},\ldots,(\bar{\mathrm{X}}_{\mathrm{K}}-\mathrm{k}_{\mathrm{K}})_{+}\right]$, where $(\bar{\mathrm{X}}_{\mathrm{i}}-\mathrm{k}_{\mathrm{j}})_{+}=(\bar{\mathrm{X}}_{\mathrm{i}}'-\mathrm{k}_{\mathrm{j}})\mathrm{I}_{\left\{(\bar{\mathrm{X}}_{\mathrm{i}}-\mathrm{k}_{\mathrm{j}})>0\right\}}$. Our PS model is given by

$$\bar{\mathrm{y}}_{\mathrm{i}}=\theta_{\mathrm{i}}+\mathrm{e}_{\mathrm{i}},\quad \theta_{\mathrm{i}}=\underset{\sim}{\bar{\mathrm{X}}}_{\mathrm{i}}'\beta+\mathrm{Z}_{\mathrm{i}}\gamma+\mathrm{v}_{\mathrm{i}}+\mathrm{e}_{\mathrm{i}},\mathrm{i}=1,\ldots,\mathrm{m},\quad \gamma_{\mathrm{j}}\overset{\text{i.i.d.}}{\square}(0,\sigma_{\gamma}^{2}),\quad \mathrm{v}_{\mathrm{i}}\overset{\text{i.i.d.}}{\square}(0,\sigma_{\mathrm{v}}^{2}),\quad \mathrm{e}_{\mathrm{i}}\overset{\text{i.d.}}{\square}(0,\psi_{\mathrm{i}}),\qquad (3)$$

where the random effects are assumed independent. Let $\mathrm{Z}=(\mathrm{Z}_1',\ldots,\mathrm{Z}_{\mathrm{m}}')'$. The $\mathrm{m}\times\mathrm{m}$ covariance of $\bar{\mathrm{y}}$ is given by $\mathrm{V}=\sigma_{\gamma}^{2}\mathrm{ZZ}'+\sigma_{\mathrm{v}}^{2}\mathrm{I}_{\mathrm{m}}+\mathrm{D}$. As in the FH model (1) the $\psi_{\mathrm{i}}, \mathrm{i}=1,\ldots,\mathrm{m}$, are known constants and the variance components σ_{v}^{2} and σ_{γ}^{2} are generally unknown and estimated from the data by the REML method. The coefficients $\tilde{\beta}$ and $\tilde{\mathrm{v}}$ are estimated by (2) replacing σ_{γ}^{2} and σ_{v}^{2} in (2) by $\hat{\sigma}_{\gamma}^{2}$ and $\hat{\sigma}_{\mathrm{v}}^{2}$, and $\tilde{\gamma}=\hat{\sigma}_{\gamma}^{2}\mathrm{Z}'\hat{\mathrm{V}}^{-1}(\mathrm{y}-\mathrm{X}\tilde{\beta})$ (see Rao, 2003, section 6.2). The EBLUP of the small area mean θ_{i} is given by a convex combination of its direct and predicted estimates $\mathrm{K}_{\mathrm{i}}\bar{\mathrm{y}}_{\mathrm{i}}+(1-\mathrm{K}_{\mathrm{i}})\underset{\sim}{\bar{\mathrm{X}}}_{\mathrm{i}}'\tilde{\beta}$ plus other less influential terms (here $\mathrm{K}_{\mathrm{i}}=1-\psi_{\mathrm{i}}\hat{\mathrm{V}}_{\mathrm{ii}}^{-1}$, where $\hat{\mathrm{V}}_{\mathrm{ii}}^{-1}$ is the i-th diagonal element of the inverse of the covariance matrix $\hat{\mathrm{V}}$). Furthermore, the dominant term in the mean squared error (MSE) of the EBLUP for the i-th area is $\mathrm{g}_{1\mathrm{i}}=\psi_{\mathrm{i}}\mathrm{K}_{\mathrm{i}}$ and if we define Γ_{i} as the ratio of the model variance over the overall variance, then $\mathrm{K}_{\mathrm{i}}\approx\Gamma_{\mathrm{i}}\,\square\,(\sigma_{\gamma}^{2}\mathrm{Z}_{\mathrm{i}}\mathrm{Z}_{\mathrm{i}}'+\sigma_{\mathrm{v}}^{2})/(\sigma_{\gamma}^{2}\mathrm{Z}_{\mathrm{i}}\mathrm{Z}_{\mathrm{i}}'+\sigma_{\mathrm{v}}^{2}+\psi_{\mathrm{i}})$ (see Rubin-Bleuer and Dochitoiu, 2008).

3.2 Variance Smoothing

We assumed that ψ_i is known, but actually we work with estimates from the sample, which are known to be unstable. It is customary to smooth the estimated variances for the purpose of small area modelling as this yields results that are more stable. We tried several auxiliary variables for smoothing: sample size, tax data, and the variable of interest, *y*. On a regular scatter-plot the relationships appeared rather noisy. On a log-log scatter-plot relationships were more apparent. The auxiliary variable with the most noticeable trend seemed to be the *y* variable. After smoothing in the log-log domain, the smoothed values were transformed back by exponentiation, and used as inputs to the two models described above.

3.3 Estimation of Mean Squared Error (MSE) of the EBLUP under the PS model

The estimators of the fixed and random regression coefficients β,γ and v, are not linear in the response variables $\bar{\mathrm{y}}_{\mathrm{i}}$ because the variance components have to be estimated. Here we extend to the PS model a "parametric bootstrap" technique proposed for the classical FH model by Pfeffermann and Glickman (2004). This technique assumes that the spline slopes, the area effects and the sampling errors are distributed approximately as normal random variables. We follow the steps of Pfeffermann and Glickman (2004) in order to reduce the bias due to the estimation of the variance components. For a detailed description of the method see Rubin-Bleuer and Dochitoiu (2008).

4. EXAMPLE

Most TGs comprised a small number of areas ($\mathrm{m}=20$ in average). Hence our PS models contain two or three knotts at the most (K≤3) and we define fixed knots by looking at the scatter-plots. In total there were 102 model groups, defined by the six economic variables and the 17 TGs. The scatter-plots showed that 16 model groups had clear non-linear trends.

For these we determined the knots and fitted both the PS and FH models.. The other model groups were fitted with the FH model.

Table 1 lists the direct estimates and the small area estimates from fitting the PS and FH models for Labour Expenses as the characteristic of interest and say Trade Group A, for the sake of confidentiality. The first two columns show the direct estimates and estimated design standard errors $\sqrt{\psi_i}, i = 1, ..., m$ (obtained after smoothing the estimated sampling variances). We observe that 6 of the 18 estimates have a standard error of zero. These correspond to small area means from domains that were completely enumerated by design. However, they contribute to the modeling of the random effects. The next four columns show the estimated EBLUPs $\hat{\theta}_i^{PS}$ under the penalized spline model, the bootstrap root mean squared errors (RMSE), the Γ_i factors (ratios of the model variance to the overall variance) and the associated synthetic estimates $\bar{\underset{\sim}{X}}'_i \tilde{\beta}^{PS}, i = 1, ..., m,$ resulting from fitting the PS model. Similarly, columns seven to ten show the EBLUPs $\hat{\theta}_i^{FH}$ under the FH model, the bootstrap RMSEs, the κ_i factor and the associated synthetic estimates $\bar{\underset{\sim}{X}}'_i \tilde{\beta}^{FH}, i = 1, ..., m,$ resulting from fitting the FH model.

We observe that the gains in efficiency over the direct estimator, from fitting either model, are considerable. Note that unlike the FH estimates, the PS small area estimates $\hat{\theta}_i^{PS}$ do not always lie in the convex segment determined by the direct and synthetic estimates, as is evident in in areas 4 and 12. But the convex combination of the direct and synthetic estimates is in general influential in the composition of the PS estimate (see Section 3.1). Due to the random nature of the spline slopes, it is expected that the PS estimators would be slightly less efficient than the FH estimators, even if the model fits well. Table 1 shows that in 5 of the 12 areas with non-zero sampling variances, the factor Γ_i is very small and the bootstrap RMSE of the PS estimator is smaller than the corresponding bootstrap RMSE of the FH estimate. In these cases, the relative difference in estimated RMSEs (relative to the design variances) is small. However, in the seven areas where the bootstrap RMSEs are smaller, the relative differences are larger. This is not enough to decide which model fits better. A test of $\sigma_\gamma^2 = 0$ vs σ_γ^2 positive, using an approach similar to that of Opsomer et al (2008), might be useful. Figure 1 shows the data and the synthetic estimates produced by the two models, FH and PS.

5. FINAL CONSIDERATIONS AND FUTURE WORK

Outlier detection and deletion is always needed for business data. We have used here two generic methods. Further, we fit only linear models when determining outlier related statistics, but we fit spline models on the cleaned data set. Even though we believe the approximation thus introduced may be acceptable, it would be interesting to have outlier detection customized to the penalized spline case. In addition, the production of final estimates requires calibration to higher order totals. Further work is planned to develop appropriate goodness-of-fit tests to decide between the PS and FH models. Finally, the approach taken here involves modeling means after removing the take-none strata. Administrative tax data is scarce for the units in the Take-none strata: for most units, only Total Revenue is available. For characteristics of interest other than Total Revenue, we plan in the future, to fit the direct estimates of small area domains as functions of the total revenue rather than the corresponding tax variable, in order to obtain estimates that include the take-none strata.

6. REFERENCES

Besley, D., Kuh, E. and Welsch, R. (1980). Regression Diagnostics. Identifying Influential data and Sources of Collinearity, Wiley series in probability and Statistics.

Economic Globalization Indicators Project (2006), http://icn-rci.statcan.ca/fpt/fpt02/btsreports/fpt02_egip_2006_e.htm - internal

Fay, R.E., and Herriot, R.A. (1979), *Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data* Journal of the American Statistical Association, 74, 269-277.

Gervais, Y. (2006). *Proposal Summary: Globalization Project (A-530-01*) (Internal Document, BTS, Statistics Canada.

Hidiroglou, M.A. and Smith, P.A. *Developing Small Area Estimates for Business Surveys at the ONS.* Statistics in Transition, December 2005, vol. 7, No. 3, pp. 527-539.

International Trade Division (2007), http://icn-rci.statcan.ca/fpt/fpt02/fpt02_cbs_200710min-eng.htm - internal

Nadeau, C., *Challenges associated with the increased use of fiscal data for the Unified Enterprise Survey*, Proceedings of the Survey Research methods section of the American statistical association, 2004.

Opsomer, D., Claeskens, G., Ranalli, M.G., Kauermann, F. and Breidt, F.J. (2008). *Nonparametric small area estimation using penalized spline regression*. Journal of the Royal Statistical Society, Series B, 70, pp. 265-286

Pfeffermann D., Glickman H. *Mean Squarred Eerror Approximation in Small Area Estimation by Use of Parametric and Non-Parametric Bootstrap* Proceedings of the Section on Survey Research Methods (2004) Alexandria, VA: American Statistical Association. 4167-78.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: Wiley.

Rubin-Bleuer, S. and Dochitoiu, C. (2008). Bootstrap Mean Squared Errors for Penalized Splines Models in Small Area Estimation. Statistics Canada Series, SRID-2008.

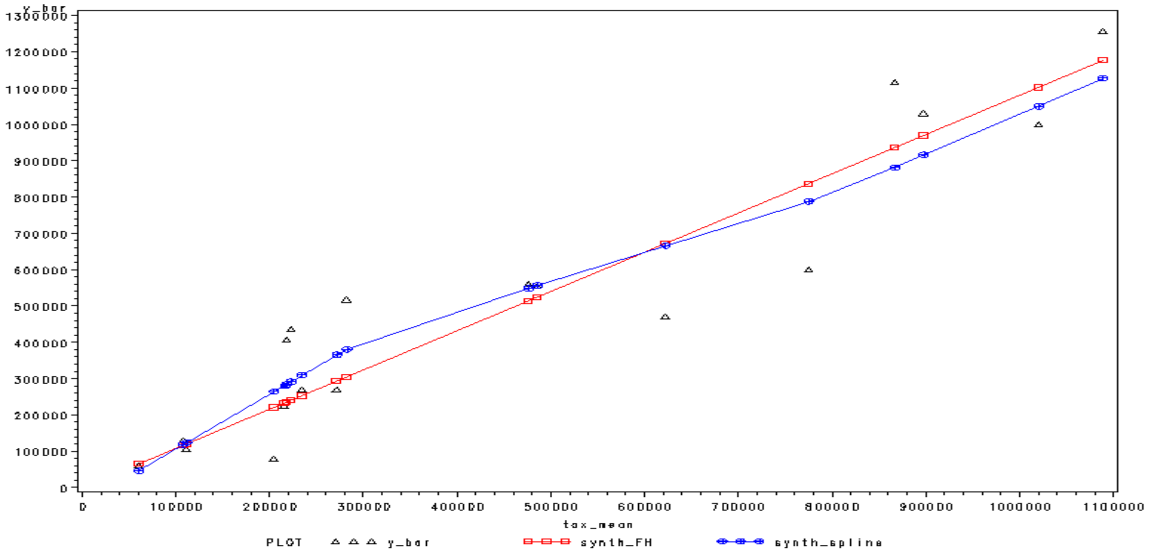
Särndal. C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

UES (2008), www.statcan.ca/bsolc/english/bsolc?catno=68F001X

Table 1. Estimates and their estimates standard deviations for Trade Group A - in thousands of dollars.

Dir	$\sqrt{\psi_i}$	PS	RMSE	Γ	Syn PS	FH	RMSE	κ	Syn FH
269	(0)	269	(0)	1.00	364	269	(0)	1.00	293
130	(0)	130	(0)	1.00	118	130	(0)	1.00	116
517	(155)	371	(58)	0.04	380	326	(81)	0.10	304
601	(181)	809	(103)	0.85	788	819	(73)	0.07	837
561	(0)	561	(0)	1.00	548	561	(0)	1.00	514
557	(0)	557	(0)	1.00	556	557	(0)	1.00	524
224	(65)	271	(32)	0.18	281	230	(36)	0.38	233
1030	(315)	932	(107)	0.75	915	971	(75)	0.02	970
80	(22)	183	(81)	0.64	264	103	(27)	0.83	221
1116	(341)	901	(103)	0.69	881	940	(76)	0.02	936
408	(121)	281	(38)	0.06	283	262	(50)	0.15	235
1000	(305)	1054	(159)	0.83	1050	1099	(77)	0.02	1102
106	(0)	106	(0)	1.00	123	106	(0)	1.00	120
61	(0)	61	(0)	1.00	46	61	(0)	1.00	65
437	(130)	288	(40)	0.05	291	267	(48)	0.13	240
1256	(386)	1123	(198)	0.79	1126	1178	(75)	0.01	1176
269	(79)	299	(39)	0.13	308	258	(62)	0.29	253
470	(140)	674	(54)	0.81	666	648	(40)	0.11	672

Figure 1. Trade Group A, Labour Expenses. Synthetic Estimates for the PS and FH models


y_bar
tax_mean
PLOT △△△ y_bar
synth_FH
synth_spline