

## Article

Symposium 2008:  
Data Collection: Challenges, Achievements and New Directions

### **Comparisons of Collection Follow-Up Score Functions**

by Hansheng Xie

2009



## Comparisons of Collection Follow-Up Score Functions

Hansheng Xie<sup>1</sup>

### Abstract

The Unified Enterprise Survey (UES) at Statistics Canada is an annual business survey that unifies more than 60 surveys from different industries. Two types of collection follow-up score functions are currently used in the UES data collection. The objective of using a score function is to maximize the economically weighted response rates of the survey in terms of the primary variables of interest, under the constraint of a limited follow-up budget. Since the two types of score functions are based on different methodologies, they could have different impacts on the final estimates.

This study generally compares the two types of score functions based on the collection data obtained from the two recent years. For comparison purposes, this study applies each score function method to the same data respectively and computes various estimates of the published financial and commodity variables, their deviation from the true pseudo value and their mean square deviation, based on each method. These estimates of deviation and mean square deviation based on each method are then used to measure the impact of each score function on the final estimates of the financial and commodity variables.

Key Words: Data collection, Follow up operation, Score function.

### 1. Introduction

The Unified Enterprise Survey (UES) at Statistics Canada is an annual program that unifies business surveys from the service, wholesale and retail, transportation, aquaculture, banking and manufacturing industries. These surveys were formerly designed and processed independently. The objective of the UES is to increase the efficiency in the survey process and to produce reliable and coherent estimates for key variables at various levels. The UES currently includes more than 60 surveys that share a common frame, sample design, collection method, data processing, and estimation methods. For more information regarding the UES, please refer to the references: Benhin, E., Bérard, H., DeBlois, S., Lebrasseur, D., Provençal, J.S. and Turmelle, C. (2007).

For business surveys at Statistics Canada, there are significant costs to data collection, especially the follow-up operation. In order to maximize the economically weighted response rates under the constraint of a limited follow-up budget, follow-up score functions have been employed in data collection for many business surveys at Statistics Canada.

In 2001, a score function referred to as Score Function 1 (SF1) was developed for the Annual surveys of Manufacture (ASM) when the ASM was not yet integrated into the UES program. In 2002, another score function, Score Function 2 (SF2), was developed for the UES Surveys. In 2004, the ASM was fully integrated into the UES, and since then these two score function systems have been used separately during the UES data collection. SF1 is used for the manufacturing surveys, while SF2 is used for the other surveys. Since the two existing score functions are based on the different methodologies, they could have different impacts on the final estimates. In order to further improve the cost-efficiency of the UES data collection and redesign an effective and unique score function system for data collection, it is necessary to conduct some studies on the two score functions. A study has been conducted firstly for generally comparing the two score functions. This study generally compares the two types of score functions using the ASM collection data collected for Reference Years (RY) of 2004 and 2005. For more information regarding the SF1 and SF2, please refer to the references: Philips, R. (2003) and Pursey, S. (2003).

The purpose of this paper is to briefly describe the characteristics of both score functions, discuss the methodology of this comparative study and summarize the main results.

---

<sup>1</sup> Hansheng Xie (hansheng.xie@statcan.gc.ca), Statistics Canada, Business Survey Methods Division, Statistics Canada, Ottawa, Canada K1A 0T6

## 2. Score function

For each UES survey, a sample is selected at the establishment level using stratified simple random sampling without replacement and data are then collected either at the establishment level or at the company (a group of establishments) level, depending on the requirement for that survey. The process of data collection consists of designing questionnaires for each survey, mailing out questionnaires to selected units, and following up the collection units that are either non-responses or that did not pass specific collection edits. The score function is applied at the collection stage to prioritize the collection effort on specific follow-up units and to provide a useful day to day operational plan for data collection staff to follow.

### 2.1 Method and application

A score function is essentially a measure of importance of collection units based on their contribution to the economically weighted response rate. This weighted response rate at the data collection phase takes account both the Sample Weight  $w_k$  and the variable of Revenue  $x_k$  at the level of domain of interest. It is defined as follows:

$$\text{Economically Weighted Response Rate} = \frac{\sum_{k \in R \cap I} w_k x_k}{\sum_{k \in I \cup U} w_k x_k}, \quad (1)$$

where  $R \cap I = \{1, \dots, r\}$  corresponds to  $r$  *In-Scope Responding Units* in the sample;  $I = \{1, \dots, i\}$  corresponds to  $i$  *In-Scope Units* in the sample and  $U = \{1, \dots, u\}$  corresponds to  $u$  *Unresolved Units* in the sample with  $r \leq i+u$ . *Responding Units* include all units that are deemed to have responded by virtue of having provided usable information, while *In-Scope Units* have been determined to belong to the target universe for the survey. *Unresolved Units* are units that are either impossible or impractical to be resolved. For example, in a business survey using a mail-back collection mode, the status of non-respondent units is questionable: non-responding units may be in-scope or out-of-scope (e.g., the business no longer exists, or it exists but it is in another industry).

The higher the unit's score, the higher the priority for the unit to be followed up. In each survey, the score function is applied to the group of collection units belonging to the same domain of interest such as industry and geography (usually province/territory). Before collection, the score function is used to assign an initial *contribution score* to each collection unit within each group. This score depends on the collection unit's contribution to the weighted response rate and the calculation of the score is based on its specific methodology. This will be described further later.

With a target of weighted response rate set for each group, collection units within each group are sorted by this *contribution score* in descending order. These ordered units are divided into two subsets using the target of weighted response as the cut-off. The units belonging to the top subset are assigned a priority score of one; hence these units will be followed up in case of non-response ("follow-up units"). The remaining units in the bottom subset receive a priority score of zero, indicating that non-responding units will not be followed-up ("non-follow-up units").

The *contribution score* is recalculated periodically based on the updated information. In the recalculation, all the units that have responded contribute to the weighted response rate and are reassigned a priority score of zero; confirmed non-responses are also reassigned a priority score of zero since they will not provide any data; out-of-scope units are removed entirely from the process, and the priority scores of some units are promoted from zero to one to make up for the non-responses.

### 2.2 Main differences of the methodology between the two score functions

Let  $x_{ij}$  represent the revenue value produced by the  $j^{\text{th}}$  commodity of the  $i^{\text{th}}$  unit and  $w_{ij}$  be the sample weight for  $x_{ij}$ . Let the total number of units in a domain of interest be  $N$  and the total number of output commodities produced in this domain of interest be  $M$ . Suppose that this unit produces  $m$  output commodities, then during collection, the score based on Score Function 1 is calculated by the following formula:

$$SF1_i^t = \sum_{j=1}^m \left\{ w_{ij}^t x_{ij}^t * \frac{\sum_{j=1}^m w_{ij}^t x_{ij}^t}{\sum_{i=1}^N \sum_{j=1}^M x_{ij}^{t-1}} * \frac{\sum_{i=1}^N x_{ij}^{t-1}}{\sum_{i=1}^N \sum_{j=1}^M x_{ij}^{t-1}} \right\}, \quad (2)$$

where t denotes the current year and t-1 denotes the previous year.

In addition to the above assumptions, suppose that the annual sample contains  $n$  units in this domain of interest. Then the formula of calculating the score based on Score Function 2 can be written as follows:

$$SF2_i^t = \frac{\sum_{j=1}^m w_{ij}^{t-1} x_{ij}^{t-1}}{\sum_{i=1}^n \sum_{j=1}^m w_{ij}^{t-1} x_{ij}^{t-1}}. \quad (3)$$

Table 2.2-1 describes the main differences of the designs between the two score functions.

**Table 2.2-1**  
**Main differences of the methodology between the two score functions**

	<b>Score Function 1</b>	<b>Score Function 2</b>
<b>Score Calculation</b>	Based on both the contribution of a unit and the contribution of its output commodities to the total value of a key financial variable at a certain level.	Based on the contribution of a unit to the total value of a key financial variable at a certain level.
<b>Estimate of the key financial variable</b>	Based on the information updated during the current collection and the information obtained last year.	Based on the Information obtained last year.

During the application of a score functions, the key financial variable is usually Revenue, such as Operating Revenue or Shipment Revenue. In general, the structure of SF1 is more complicated than that of SF2.

### 3. Comparative study

#### 3.1 Data used in the study

This study is conducted using data RY2004 and RY2005 for ASM. These data are the ASM Collection Data, ASM Final Financial Data, ASM Final Input Commodity Data and ASM Final Output Commodity Data. The Collection Data contains all the collection information in a specific period during collection, such as Response Code, Current Collection Stage, etc. The collection information is used by updating the current status of collection units. During collection, each score function is run twice per week for over 25 weeks. Thus there are more than 50 collection datasets for each reference year.

The study is done by comparing the results obtained for 21 industries, 12 financial variables (FV) and 2 commodity variables (C1: Input Commodity and C2: Output Commodity).

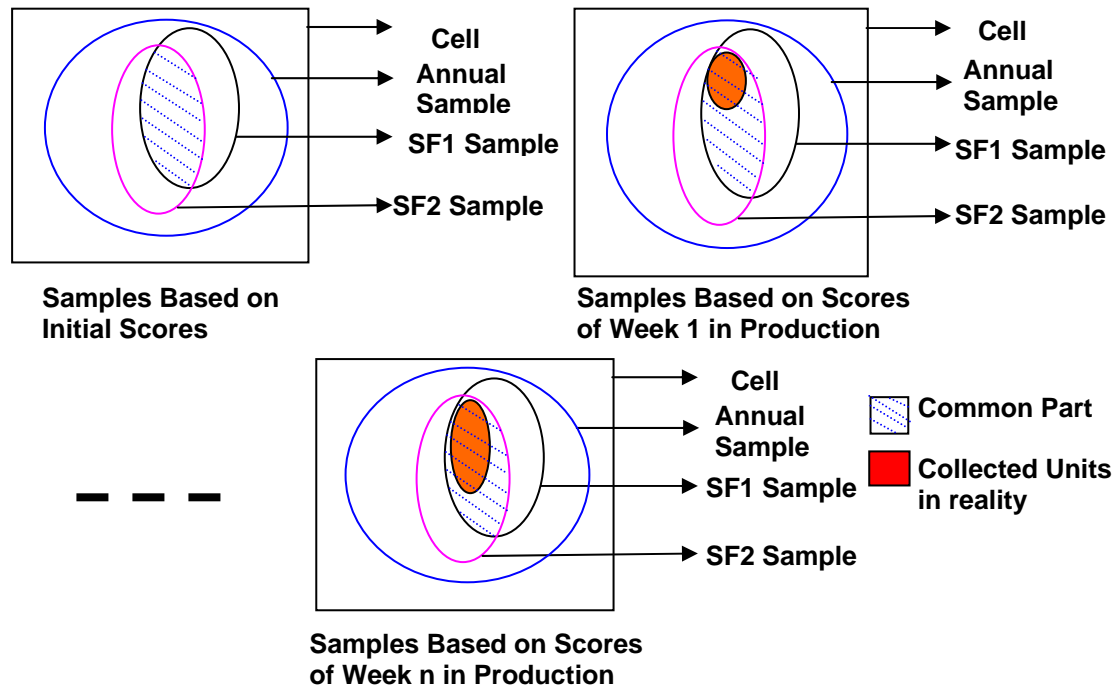
#### 3.2 Methods employed in the study

In our study, a population is established for each of the three types of the final data, i.e., the Final Financial Data, Final Input Commodity Data and Final Output Commodity Data, for each of the two reference years. The population consists of all annually sampled units and non-sampled units. After applying each score function method to the

collection data at the cell level that is defined as the combination of industry and province, a subset of units is identified for follow-up for each production run, from the annually sampled units based on each score function method.

The following graph shows the two samples of follow-up units at the Cell level.

**Figure 3.2-1**  
**Samples of follow-up units**



Based on each score function method, a sample of follow-up units is selected within the annual sample. There is a common part of follow-up units for the two score function samples. This common part also contains the actually collected units during collection. As collection goes along, there are more and more collected units.

For this study, the element of estimation is at the location level, similar to what is done in the production of the estimates for the manufacturing sector. The location level is lower than the establishment level and an establishment could have several locations. Based on each sample of follow-up units, calibrated estimates of totals are produced at various levels for the chosen variables. There are 3 levels for the financial variables: *ASM Total*, *Industry* and *Cell*. There are also 3 levels for the commodity variables: *ASM Total*, *Commodity* and *Commodity within Cell*. For comparison purposes, population totals are also calculated at various levels for the chosen variables and they are used as “reference” values in order to measure the impact of each score function on the final estimates of the financial and commodity variables.

This study uses the collection information contained in the collection data of different specific periods. As a fact, there is only one annual sample from which a sample of follow-up units is selected. Thus, two measurements for the comparisons are defined as follows:

For measuring the deviation of a sample estimate from the reference value, the Relative Deviation (RD) is defined as

$$\text{Relative Deviation} = \frac{\hat{X} - X}{X}, \quad (4)$$

where  $\hat{X}$  is a *Sample Estimate* based on a score function method and  $X$  is the *Pseudo True Value* of a variable.

For measuring the total variation about the reference value, the Relative Square Root of Mean Square Deviation (RSRMSD) is defined as

$$RSRMSD = \frac{\sqrt{(\hat{X} - X)^2 + \hat{S}^2}}{X}, \quad (5)$$

where  $S$  is the *Standard Error*.

These RD and RSRMSD are computed at various levels for each sample of follow-up units.

Since RSRMSD takes into account the information of both deviation and standard error, it is used as a standard for the comparisons and the following criterion is defined: if  $|RSRMSD1 - RSRMSD2| < 1\%$ , the two methods are considered to be about the same; otherwise, one method is better than the other.

Based on the two years' data, we conclude that:

- 1) The two methods are about the same if for both years, they are about the same; or for one year one method is better, but for the other year the method is worse than the other method.
- 2) One method is better than the other if for both years, it is better; or for one year it is better, and for the other year the two methods are about the same.

### 3.3 Analysis of the main results

For simplicity, the samples of three specific periods during collection are chosen for the following comparisons. Table 3.3-1 contains Number of Financial Variables by Comparison Category at the *ASM Total* and *Industry* levels. They are the samples based on the Initial Scores, Week 8 and Week 15 respectively. The numbers highlighted in bold are the cases that SF2 performs better than SF1, in terms of the number of financial variables.

**Table 3.3-1**  
**Comparisons: number of financial variables by category**

Industry	Initial scores			Week 8			Week 15		
	SF1 better	SF2 better	About same	SF1 better	SF2 better	About same	SF1 better	SF2 better	About same
<b>Total</b>	5	<b>6</b>	1	0	<b>9</b>	3	2	<b>5</b>	5
<b>#1</b>	6	0	6	1	<b>6</b>	5	4	1	7
<b>#2</b>	12	0	0	11	0	1	11	0	1
<b>#3</b>	0	<b>5</b>	7	0	0	12	12	0	0
<b>#4</b>	11	0	1	1	0	11	1	<b>6</b>	5
<b>#5</b>	12	0	0	10	0	2	9	0	3
<b>#6</b>	12	0	0	10	0	2	11	0	1
<b>#7</b>	5	1	6	7	1	4	10	0	2
<b>#8</b>	5	2	5	5	0	7	11	0	1
<b>#9</b>	5	1	6	5	3	4	5	3	4
<b>#10</b>	1	11	0	12	0	0	12	0	0
<b>#11</b>	9	0	3	0	<b>11</b>	1	2	<b>9</b>	1
<b>#12</b>	11	0	1	10	0	2	8	0	4
<b>#13</b>	6	2	4	5	2	5	10	1	1
<b>#14</b>	1	<b>11</b>	0	1	0	11	12	0	0
<b>#15</b>	1	<b>8</b>	3	7	2	3	2	<b>3</b>	7
<b>#16</b>	10	0	2	12	0	0	5	0	7
<b>#17</b>	1	0	11	5	0	7	7	3	2
<b>#18</b>	12	0	0	0	<b>3</b>	8	0	<b>2</b>	10
<b>#19</b>	1	<b>7</b>	4	1	<b>8</b>	3	4	0	8
<b>#20</b>	1	<b>10</b>	1	12	0	0	4	<b>7</b>	1
<b>#21</b>	7	0	5	12	0	0	9	0	3

It can be seen from Table 3.3-1 that at the *ASM Total* level, SF2 performs better than SF1, while at the *Industry* level, SF1 performs better than SF2 for most industries.

Ten large commodities are chosen for comparisons. For these ten commodities, they have the same names for their input and output commodities. Table 3.3-2 shows Input Commodity by Comparison Category for the *ASM Total* and the *Ten Large Commodities*.

**Table 3.3-2**  
**Comparisons: input commodity by category**

Commodity	Initial scores			Week 8			Week 15		
	SF1 better	SF2 better	About same	SF1 better	SF2 better	About same	SF1 better	SF2 better	About same
<b>Total</b>		*			*			*	
<b>#1</b>	*			*			*		
<b>#2</b>			*		*			*	
<b>#3</b>			*			*	*		
<b>#4</b>	*			*				*	
<b>#5</b>			*	*			*		
<b>#6</b>			*			*			*
<b>#7</b>		*		*					*
<b>#8</b>		*				*		*	
<b>#9</b>			*			*			*
<b>#10</b>	*			*			*		

\* Comparison Status

At the *ASM Total* level, SF2 performs better than SF1; however, SF1 appears to perform at least similarly as SF2 for most of the *Ten Large Commodities*.

Table 3.3-3 shows output Commodity by Comparison Category for the *ASM Total* and the *Ten Large Commodities*.

**Table 3.3-3**  
**Comparisons: output commodity by category**

Commodity	Initial scores			Week 8			Week 15		
	SF1 better	SF2 better	About same	SF1 better	SF2 better	About same	SF1 better	SF2 better	About same
<b>Total</b>			*			*	*		
<b>#1</b>	*			*			*		
<b>#2</b>	*			*					*
<b>#3</b>		*		*			*		
<b>#4</b>	*					*			*
<b>#5</b>	*				*		*		
<b>#6</b>			*			*	*		
<b>#7</b>	*					*		*	
<b>#8</b>	*			*			*		
<b>#9</b>	*					*			*
<b>#10</b>		*			*		*		

\* Comparison Status

It can be seen from Table 3.3-3 that, in general, SF1 seems to perform better, and especially in Week 15, it has a strong advantage over SF2. This may be due to the fact that SF1 takes into account the contribution of a unit's output commodities.

Table 3.3-4 gives Percentage by Comparison Category at Various Levels based on the RY2004 Data. These levels are *Cell*, *Commodity* and *Commodity within Cell*.

**Table 3.3-4****Comparisons: percentage by category at various levels**

Level	Initial scores			Week 8			Week 15		
	SF1 better (%)	SF2 better (%)	About same (%)	SF1 better (%)	SF2 better (%)	About same (%)	SF1 better (%)	SF2 better (%)	About same (%)
Cell (FV)	36.8	26.7	36.5	41.3	22.3	36.4	39.2	16.7	44.0
Commodity (C1)	30.8	27.2	42.1	37.7	18.8	43.7	29.9	14.5	55.6
Commodity (C2)	27.9	25.2	46.9	34.6	16.9	48.5	26.8	13.0	60.2
Commodity * Cell (C1)	16.1	14.3	69.6	19.3	8.4	72.4	14.1	6.1	79.8
Commodity * Cell (C2)	15.8	12.3	71.9	17.8	6.9	75.3	13.8	5.3	81.0

Table 3.3-4 shows that across all these levels, the percentage of 'SF1 better' is higher than that of 'SF2 better', and there is a big percentage in the category of 'About same'. As well, the results obtained based on the RY2005 data are similar to those shown in this table.

#### 4. Conclusions

In our study, the Mean Square Deviation is used as a standard for the general comparisons of the two score functions, in order to measure the impact of each score function on the final estimates of the financial and commodity variables.

The study has shown that at the level of *ASM Total*, Score Function 2 seems to perform better than Score Function 1 for the financial variables and for the variable of Input Commodity; while at the level of *Industry*, Score Function 1 seems to perform better than Score Function 2 for the financial variables and for the variable of Output Commodity. At the levels of *Cell*, *Commodity* and *Commodity within Cell*, the numbers of Cells, Commodities and Commodities within Cell in the Category of 'SF1 better' are higher than those in the Category of 'SF2 better'.

Due to some limitations of the collection data, many factors are confounded in this general study and the preliminary results shown in this paper may also be limited. In the future, it would be interesting to just use the Relative Deviation as a standard for the comparisons of the two score functions. This could lead to some different conclusions, such as some preferable estimates that have a smaller relative deviation but a larger variance. As well, some specific studies need to be conducted in order to determine optimal features for the redesigned score function.

#### References

- Benhin, E., Bérard, H., DeBlois, S., Lebrasseur, D., Provençal, J.S. and Turmelle, C. (2007). Unified Enterprise Survey (UES) Future Directions: Methodological Perspectives, internal document, Statistics Canada, Ottawa, Canada.
- Philips, R. (2003). The Theory and Application of the Score Function for Determining the Priority of Follow Up in the Annual Survey of Manufactures, *Proceedings of the Survey Methods Section*, Statistical Society of Canada, 121-126.
- Pursey, S. (2003), Use of the Score Function to Optimize Data Collection Resources in the Unified Enterprise, *Proceedings of the Survey Methods Section*, Statistical Society of Canada, 117-120.