

## Article

Symposium 2008 :  
Collecte des données : défis, réalisations et nouvelles orientations

### **Comparaisons des fonctions de score pour le suivi de la collecte**

par Hansheng Xie

2009



## Comparaisons des fonctions de score pour le suivi de la collecte

Hansheng Xie<sup>1</sup>

### Résumé

L'Enquête unifiée auprès des entreprises (EUE) réalisée par Statistique Canada est une enquête-entreprise annuelle dont le but est d'uniformiser plus de 60 enquêtes couvrant diverses industries. À l'heure actuelle, deux types de fonctions de score sont utilisés durant la collecte des données de l'EUE pour en faire le suivi. L'objectif est d'employer une fonction de score qui maximise les taux de réponse à l'enquête pondérés par le poids économique en ce qui a trait aux principales variables d'intérêt, sous la contrainte d'un budget de suivi limité. Les deux types de fonctions de score étant fondés sur des méthodologies différentes, leur incidence sur les estimations finales pourrait ne pas être la même.

La présente étude consiste à comparer, d'une manière générale, les deux types de fonctions de score en s'appuyant sur des données concernant la collecte recueillies au cours des deux dernières années. Aux fins des comparaisons, chaque type de fonction de score est appliqué aux mêmes données et diverses estimations de variables financières et de variables liées aux marchandises (biens et services) pour lesquelles des données sont publiées sont calculées, ainsi que leur écart par rapport à la pseudo valeur réelle et leur écart quadratique moyen, en se fondant sur chaque méthode. Ces estimations de l'écart et de l'écart quadratique moyen calculées selon chaque méthode sont ensuite utilisées pour mesurer l'effet de chaque fonction de score sur les estimations finales des variables financières et des variables liées aux biens et services.

Mots clés : Collecte des données, opération de suivi, fonction de score.

### 1. Introduction

L'Enquête unifiée auprès des entreprises (EUE) réalisée par Statistique Canada est un programme annuel dont le but est d'uniformiser les enquêtes-entreprises portant sur les industries des services, du commerce de gros et de détail, du transport, de l'aquaculture, des services bancaires et de la fabrication. Anciennement, ces enquêtes étaient conçues et exécutées indépendamment les unes des autres. L'objectif de l'EUE est de rendre le processus d'enquête plus efficace et de produire des estimations fiables et cohérentes pour les variables clés à divers niveaux. À l'heure actuelle, l'EUE comprend plus de 60 enquêtes qui ont en commun la base de sondage, le plan d'échantillonnage, la méthode de collecte, le traitement des données et les méthodes d'estimation. Pour plus de renseignements concernant l'EUE, le lecteur est invité à consulter Benhin, E., Bérard, H., DeBlois, S., Lebrasseur, D., Provençal, J.S. et Turmelle, C. (2007) (voir la bibliographie).

Le coût de la collecte des données des enquêtes-entreprises menées par Statistique Canada est important, surtout celui de l'opération de suivi. Afin de maximiser les taux de réponse pondérés par le poids économique sous la contrainte d'un budget de suivi limité, nous employons des fonctions de score de suivi dans la collecte des données pour bon nombre de ces enquêtes.

En 2001, une fonction de score appelée fonction de score 1 (SF1) a été établie pour l'Enquête annuelle des manufactures (EAM), avant qu'elle ne soit intégrée dans le programme de l'EUE. En 2002, une autre fonction de score, la fonction de score 2 (SF2), a été établie pour les enquêtes de l'EUE. En 2004, l'EAM a été entièrement intégrée dans l'EUE et, depuis, les deux systèmes de fonctions de score ont été appliqués séparément durant la collecte des données de l'EUE. La fonction SF1 est utilisée pour les enquêtes auprès des entreprises du secteur de la fabrication, tandis que la fonction SF2 l'est pour les autres enquêtes. Comme les deux fonctions de score existantes sont fondées sur des méthodologies différentes, leur effet sur les estimations finales pourrait ne pas être le même. Afin de rendre la collecte des données de l'EUE encore plus rentable et de concevoir un système de fonction de score efficace et unique pour les opérations de collecte des données, il est nécessaire d'étudier les deux fonctions de score. Une étude a été entreprise principalement pour comparer d'une manière générale les deux fonctions de score.

---

<sup>1</sup> Hansheng Xie (hansheng.xie@statcan.gc.ca), Division des méthodes d'enquête auprès des entreprises, Statistique Canada, Ottawa, Canada K1A 0T6

Elle s'appuie pour cela sur les données concernant la collecte de l'EAM recueillies pour les années de référence 2004 et 2005. Afin d'obtenir plus de renseignements concernant les fonctions SF1 et SF2, consulter Philips, R. (2003) et Pursey, S. (2003).

Le but du présent article est de décrire brièvement les caractéristiques des deux fonctions de score, de discuter de la méthodologie de l'étude comparative et de résumer les principaux résultats.

## 2. Fonction de score

Pour chaque enquête de l'EUE, un échantillon est sélectionné au niveau de l'établissement selon un plan d'échantillonnage aléatoire simple stratifié sans remise et les données sont ensuite recueillies au niveau de l'établissement ou à celui de l'entreprise (un groupe d'établissements), selon les exigences de l'enquête. Le processus de collecte des données consiste à concevoir les questionnaires pour chaque enquête, à les envoyer par la poste aux unités sélectionnées et à faire un suivi auprès des unités de collecte qui représentent des cas de non-réponse ou des cas rejetés durant l'exécution de contrôles particuliers. La fonction de score est appliquée à l'étape de la collecte des données afin de donner la priorité à des unités de suivi particulières en ce qui concerne l'effort de collecte et offrir quotidiennement un plan opérationnel utile aux employés chargés de la collecte des données.

### 2.1 Méthode et application

Une fonction de score est essentiellement une mesure de l'importance des unités de collecte qui est fondée sur leur contribution au taux de réponse pondéré par le poids économique. Ce taux de réponse pondéré, calculé à la phase de collecte des données, tient compte à la fois du poids de sondage  $w_k$  et de la variable de revenu  $x_k$  au niveau du domaine d'intérêt. Il est défini comme suit :

$$\text{Taux de réponse pondéré par le poids économique} = \frac{\sum_{k \in R \cap I} w_k x_k}{\sum_{k \in I \cup U} w_k x_k}, \quad (1)$$

où  $R \cap I = \{1, \dots, r\}$  correspond à  $r$  unités répondantes dans le champ de l'enquête présentes dans l'échantillon,  $I = \{1, \dots, i\}$  correspond à  $i$  unités dans le champ de l'enquête présentes dans l'échantillon et  $U = \{1, \dots, u\}$  correspond à  $u$  unités non résolues présentes dans l'échantillon avec  $r \leq i+u$ . Les unités répondantes comprennent toutes les unités qui sont considérées comme ayant répondu en vertu du fait qu'elles ont fourni des renseignements utilisables, tandis que les unités dans le champ de l'enquête sont celles pour lesquelles il a été déterminé qu'elles appartiennent à l'univers cible de l'enquête. Les unités non résolues sont celles qui correspondent aux cas dont la résolution est impossible ou difficilement réalisable. Par exemple, dans une enquête-entreprise s'appuyant sur un mode de collecte avec retour des questionnaires par la poste, la classification en tant qu'unité non répondante peut être mise en doute : l'unité peut être dans le champ ou hors du champ de l'enquête (p. ex., l'entreprise n'existe plus, ou bien elle existe, mais fait partie d'une autre industrie).

La priorité accordée au suivi d'une unité particulière est d'autant plus grande que le score de cette unité est élevé. Dans chaque enquête, la fonction de score est appliquée au groupe d'unités de collecte appartenant à un même domaine d'intérêt, comme l'industrie et la géographie (habituellement province/territoire). Avant la collecte, la fonction de score est utilisée pour attribuer un score de contribution initial à chaque unité de collecte dans un groupe. Ce score dépend de la contribution de l'unité de collecte au taux de réponse pondéré et le calcul du score est fondé sur la méthodologie propre à la fonction utilisée. Cet aspect sera décrit plus loin.

Une fois que le taux de réponse pondéré cible est établi pour chaque groupe, les unités de collecte comprises dans chaque groupe sont classées par ordre décroissant en fonction de ce score de contribution. Ensuite, les unités classées sont réparties en deux sous-ensembles en utilisant comme seuil le taux de réponse pondéré cible. Un score de un est attribué aux unités appartenant au sous-ensemble supérieur, et ces unités feront l'objet d'un suivi en cas de non-réponse (« unités suivies »). Les autres unités, comprises dans le sous-ensemble inférieur, reçoivent un score de

priorité de zéro, ce qui indique que les unités non répondantes de cette catégorie ne feront pas l'objet d'un suivi (« unités non suivies »).

Le *score de contribution* est recalculé périodiquement en se basant sur l'information mise à jour. Dans le nouveau calcul, toutes les unités qui ont répondu contribuent au taux de réponse pondéré et un score de priorité nul leur est attribué; un score de priorité nul est également attribué aux cas de non-réponse, puisqu'ils ne fourniront aucune donnée; les unités hors du champ de l'enquête sont retirées entièrement du processus et certaines unités sont promues d'un score de priorité de zéro à un score de un pour compenser les non-réponses.

## 2.2 Principales différences méthodologiques entre les deux fonctions de score

Soit  $x_{ij}$  la valeur du revenu produit par la  $j^{\text{e}}$  marchandise (bien ou service) de la  $i^{\text{e}}$  unité et  $w_{ij}$  le poids de sondage pour  $x_{ij}$ . Soit  $N$  le nombre total d'unités dans un domaine d'intérêt et  $M$ , le nombre total de marchandises de sortie produites dans ce domaine d'intérêt. Supposons que l'unité en question produit  $m$  marchandises de sortie; alors, durant la collecte, le score fondé sur la fonction de score 1 est calculé au moyen de la formule suivante :

$$FS1_i^t = \sum_{j=1}^m \left\{ w_{ij}^t x_{ij}^t * \frac{\sum_{j=1}^m w_{ij}^t x_{ij}^t}{\sum_{i=1}^N \sum_{j=1}^M x_{ij}^{t-1}} * \frac{\sum_{i=1}^N x_{ij}^{t-1}}{\sum_{i=1}^N \sum_{j=1}^M x_{ij}^{t-1}} \right\}, \quad (2)$$

où  $t$  désigne l'année courante et  $t-1$ , l'année précédente.

En plus des hypothèses susmentionnées, supposons que l'échantillon annuel contient  $n$  unités dans le domaine d'intérêt en question. Alors, la formule de calcul du score fondée sur la fonction de score 2 peut s'écrire :

$$FS2_i^t = \frac{\sum_{j=1}^m w_{ij}^{t-1} x_{ij}^{t-1}}{\sum_{i=1}^n \sum_{j=1}^m w_{ij}^{t-1} x_{ij}^{t-1}}. \quad (3)$$

Le tableau 2.2-1 décrit les principales différences de conception entre les deux fonctions de score.

**Tableau 2.2-1**  
**Principales différences méthodologiques entre les deux fonctions de score**

|                                                 | <b>Fonction de score 1</b>                                                                                                                               | <b>Fonction de score 2</b>                                                                              |
|-------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| <b>Calcul du score</b>                          | Basé sur la contribution de l'unité ainsi que celle des marchandises qu'elle produit à la valeur totale d'une variable financière clé à un niveau donné. | Basé sur la contribution de l'unité à la valeur totale d'une variable financière clé à un niveau donné. |
| <b>Estimation de la variable financière clé</b> | Basée sur l'information mise à jour durant la collecte courante et celle recueillie l'année précédente.                                                  | Basée sur l'information obtenue l'année précédente.                                                     |

Durant l'application d'une fonction de score, la variable financière clé est habituellement le revenu, tel que le revenu d'exploitation ou le revenu d'expédition. En général, la structure de la fonction FS1 est plus compliquée que celle de la fonction FS2.

### 3. Étude comparative

#### 3.1 Données utilisées dans l'étude

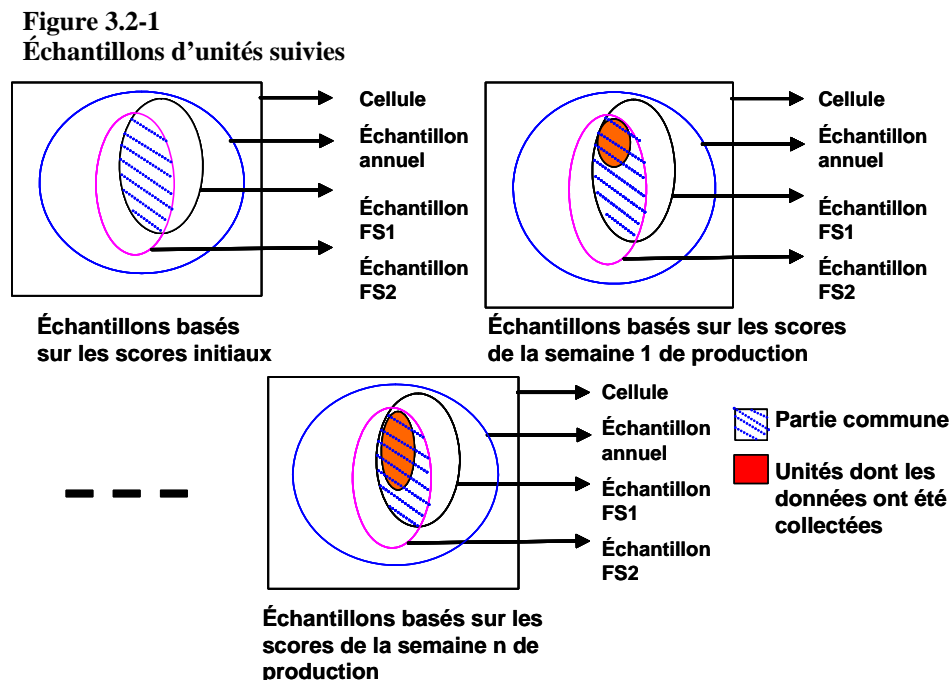
L'étude porte sur les données des années de référence 2004 et 2005 pour l'EAM. Ces données sont celles concernant la collecte de l'EAM, les données financières finales de l'EAM, les données finales sur les entrées de marchandises de l'EAM et les données finales sur les sorties de marchandises de l'EAM. Les données sur la collecte contiennent tous les renseignements recueillis pendant une période particulière durant la collecte, comme le code de réponse, l'étape courante de collecte, etc. L'information sur la collecte est utilisée pour mettre à jour le statut courant des unités de collecte. Durant la collecte, chaque fonction de score est appliquée deux fois par semaine pendant plus de 25 semaines. Donc, il existe plus de 50 ensembles de données sur la collecte pour chaque année de référence.

L'étude consiste à comparer les résultats obtenus pour 21 industries, 12 variables financières (VF) et deux variables relatives aux marchandises (M1 : entrée d'une marchandise et M2 : sortie d'une marchandise).

#### 3.2 Méthodes employées dans l'étude

Une population est établie pour chacun des trois types de données finales, c'est-à-dire les données financières finales, les données finales sur les entrées de marchandises, et les données finales sur les sorties de marchandises, pour chacune des deux années de référence. La population est constituée de toutes les unités échantillonnées et non échantillonnées annuellement. Après l'application de la méthode liée à chaque fonction de score aux données sur la collecte au niveau de la cellule définie par le recoupement de l'industrie et de la province, un sous-ensemble d'unités est désigné pour le suivi durant chaque cycle de production parmi les unités échantillonnées annuellement en se fondant sur chaque méthode.

Le graphique suivant illustre les deux échantillons d'unités suivies, au niveau de la cellule.



Selon la méthode de chaque fonction de score, un échantillon d'unités suivies est sélectionné dans l'échantillon annuel. Les échantillons obtenus au moyen des deux fonctions de score ont une partie de leurs unités suivies en commun. Cette partie commune contient aussi les unités auprès desquelles la collecte a effectivement été réalisée. À mesure que le cycle de collecte progresse, le nombre d'unités effectivement soumises à la collecte augmente.

Aux fins de la présente étude, l'estimation est faite au niveau de l'emplacement comme dans le cas de la production des estimations pour le secteur de la fabrication. Le niveau de l'emplacement est inférieur au niveau de l'établissement, celui-ci pouvant comporter plusieurs emplacements. En se fondant sur chaque échantillon d'unités suivies, des estimations calées des totaux sont produites à divers niveaux pour les variables choisies. Trois niveaux sont utilisés pour les variables financières : l'*EAM totale*, l'*industrie* et la *cellule*. Trois niveaux sont également considérés pour les variables liées aux marchandises : l'*EAM totale*, la *marchandise* et la *marchandise dans la cellule*. Pour les comparaisons, les totaux de population sont également calculés à divers niveaux pour les variables choisies, puis ils sont utilisés comme valeur de « référence » afin de mesurer l'effet de chaque fonction de score sur les estimations finales des variables financières et des variables liées aux marchandises.

La présente étude s'appuie sur l'information concernant la collecte recueillie durant des périodes différentes. En fait, il n'existe qu'un seul échantillon annuel à partir duquel est sélectionné un échantillon d'unités suivies. Donc, deux mesures sont définies comme suit pour les comparaisons :

Pour mesurer l'écart d'une estimation (basée sur un échantillon) par rapport à la valeur de référence, nous définissons l'écart relatif (ER) comme étant

$$\text{Écart relatif} = \frac{\hat{X} - X}{X}, \quad (4)$$

où  $\hat{X}$  est une *estimation* basée sur une méthode de fonction de score et  $X$  est la *pseudo valeur réelle* d'une variable.

Pour mesurer la variation totale par rapport à la valeur de référence, nous définissons la racine carrée relative de l'écart quadratique moyen (RCREQM) comme étant

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

où  $S$  est l'*erreur type*.

L'ER et la RCREQM sont calculés à divers niveaux d'agrégation pour chaque échantillon d'unités suivies.

Puisque la RCREQM tient compte de l'information provenant de l'écart ainsi que de l'erreur type, nous l'utilisons comme norme pour les comparaisons et nous définissons le critère suivant : si  $|\text{RCREQM1} - \text{RCREQM2}| < 1\%$ , les deux méthodes sont considérées comme produisant à peu près les mêmes résultats; sinon, une méthode est meilleure que l'autre.

Sur la base des données recueillies pour les deux années, nous concluons que :

- 1) les deux méthodes donnent à peu près les mêmes résultats si, pour les deux années, elles donnent à peu près les mêmes résultats ou si, pour une année, une méthode est meilleure que l'autre, mais que, pour l'autre année, elle est pire que l'autre;
- 2) une méthode est meilleure que l'autre si, les deux années, elle est meilleure ou si, une année, elle est meilleure et que, l'autre année, les deux méthodes donnent à peu près les mêmes résultats.

### 3.3 Analyse des principaux résultats

Pour simplifier l'étude, nous choisissons les échantillons obtenus à trois points particuliers de la collecte pour les comparaisons qui suivent. Le tableau 3.3-1 donne le nombre de variables financières par catégorie de comparaison aux niveaux de l'*EAM totale* et de l'*industrie*. Les échantillons sont ceux fondés sur les scores initiaux, les scores de la huitième semaine et ceux de la quinzième semaine, respectivement. Les chiffres en caractères gras indiquent les cas où la fonction FS2 donne de meilleurs résultats que la fonction FS1, en ce qui concerne le nombre de variables financières.

**Tableau 3.3-1**  
**Comparaisons : nombre de variables financières par catégorie**

| Industrie    | Scores initiaux  |                  |                               | Semaine 8        |                  |                               | Semaine 15       |                  |                               |
|--------------|------------------|------------------|-------------------------------|------------------|------------------|-------------------------------|------------------|------------------|-------------------------------|
|              | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes |
| <b>Total</b> | 5                | <b>6</b>         | 1                             | 0                | <b>9</b>         | 3                             | 2                | <b>5</b>         | 5                             |
| N° 1         | 6                | 0                | 6                             | 1                | <b>6</b>         | 5                             | 4                | 1                | 7                             |
| N° 2         | 12               | 0                | 0                             | 11               | 0                | 1                             | 11               | 0                | 1                             |
| N° 3         | 0                | <b>5</b>         | 7                             | 0                | 0                | 12                            | 12               | 0                | 0                             |
| N° 4         | 11               | 0                | 1                             | 1                | 0                | 11                            | 1                | <b>6</b>         | 5                             |
| N° 5         | 12               | 0                | 0                             | 10               | 0                | 2                             | 9                | 0                | 3                             |
| N° 6         | 12               | 0                | 0                             | 10               | 0                | 2                             | 11               | 0                | 1                             |
| N° 7         | 5                | 1                | 6                             | 7                | 1                | 4                             | 10               | 0                | 2                             |
| N° 8         | 5                | 2                | 5                             | 5                | 0                | 7                             | 11               | 0                | 1                             |
| N° 9         | 5                | 1                | 6                             | 5                | 3                | 4                             | 5                | 3                | 4                             |
| N° 10        | 1                | 11               | 0                             | 12               | 0                | 0                             | 12               | 0                | 0                             |
| N° 11        | 9                | 0                | 3                             | 0                | <b>11</b>        | 1                             | 2                | <b>9</b>         | 1                             |
| N° 12        | 11               | 0                | 1                             | 10               | 0                | 2                             | 8                | 0                | 4                             |
| N° 13        | 6                | 2                | 4                             | 5                | 2                | 5                             | 10               | 1                | 1                             |
| N° 14        | 1                | <b>11</b>        | 0                             | 1                | 0                | 11                            | 12               | 0                | 0                             |
| N° 15        | 1                | <b>8</b>         | 3                             | 7                | 2                | 3                             | 2                | <b>3</b>         | 7                             |
| N° 16        | 10               | 0                | 2                             | 12               | 0                | 0                             | 5                | 0                | 7                             |
| N° 17        | 1                | 0                | 11                            | 5                | 0                | 7                             | 7                | 3                | 2                             |
| N° 18        | 12               | 0                | 0                             | 0                | <b>3</b>         | 8                             | 0                | <b>2</b>         | 10                            |
| N° 19        | 1                | <b>7</b>         | 4                             | 1                | <b>8</b>         | 3                             | 4                | 0                | 8                             |
| N° 20        | 1                | <b>10</b>        | 1                             | 12               | 0                | 0                             | 4                | <b>7</b>         | 1                             |
| N° 21        | 7                | 0                | 5                             | 12               | 0                | 0                             | 9                | 0                | 3                             |

L'examen du tableau 3.3-1 révèle qu'au niveau de l'*EAM totale*, FS2 donne de meilleurs résultats que FS1, tandis qu'au niveau de l'*industrie*, FS1 donne de meilleurs résultats que FS2 pour la plupart des industries.

Dix marchandises importantes ont été choisies pour les comparaisons. Les noms sont les mêmes pour les entrées et les sorties de ces dix marchandises. Le tableau 3.3-2 donne les entrées de marchandises selon la catégorie de comparaison pour l'*EAM totale* et pour les *dix marchandises importantes*.

**Table 3.3-2**  
**Comparaisons : entrées de marchandises par catégorie**

| Marchandise  | Scores initiaux  |                  |                               | Semaine 8        |                  |                               | Semaine 15       |                  |                               |
|--------------|------------------|------------------|-------------------------------|------------------|------------------|-------------------------------|------------------|------------------|-------------------------------|
|              | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes |
| <b>Total</b> |                  | *                |                               |                  | *                |                               |                  | *                |                               |
| <b>N° 1</b>  | *                |                  |                               | *                |                  |                               | *                |                  |                               |
| <b>N° 2</b>  |                  |                  | *                             |                  | *                |                               |                  | *                |                               |
| <b>N° 3</b>  |                  |                  | *                             |                  |                  | *                             | *                |                  |                               |
| <b>N° 4</b>  | *                |                  |                               | *                |                  |                               |                  | *                |                               |
| <b>N° 5</b>  |                  |                  | *                             | *                |                  |                               | *                |                  |                               |
| <b>N° 6</b>  |                  |                  | *                             |                  |                  | *                             |                  |                  | *                             |
| <b>N° 7</b>  |                  | *                |                               | *                |                  |                               |                  |                  | *                             |
| <b>N° 8</b>  |                  | *                |                               |                  |                  | *                             |                  | *                |                               |
| <b>N° 9</b>  |                  |                  | *                             |                  |                  | *                             |                  |                  | *                             |
| <b>N° 10</b> | *                |                  |                               | *                |                  |                               | *                |                  |                               |

\* Résultats de la comparaison

Au niveau de l'*EAM totale*, FS2 donne de meilleurs résultats que FS1; par contre, FS1 semble donner des résultats au moins semblables à FS2 pour la plupart des *dix marchandises importantes*.

Le tableau 3.3-3 donne les sorties de marchandises par catégorie de comparaison pour l'*EAM totale* et les *dix marchandises importantes*.

**Table 3.3-3**  
**Comparaisons : sorties de marchandises par catégorie**

| Marchandise  | Scores initiaux  |                  |                               | Semaine 8        |                  |                               | Semaine 15       |                  |                               |
|--------------|------------------|------------------|-------------------------------|------------------|------------------|-------------------------------|------------------|------------------|-------------------------------|
|              | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes | FS1<br>meilleure | FS2<br>meilleure | À peu<br>près<br>les<br>mêmes |
| <b>Total</b> |                  |                  | *                             |                  |                  | *                             | *                |                  |                               |
| <b>N° 1</b>  | *                |                  |                               | *                |                  |                               | *                |                  |                               |
| <b>N° 2</b>  | *                |                  |                               | *                |                  |                               |                  |                  | *                             |
| <b>N° 3</b>  |                  | *                |                               | *                |                  |                               | *                |                  |                               |
| <b>N° 4</b>  | *                |                  |                               |                  |                  | *                             |                  |                  | *                             |
| <b>N° 5</b>  | *                |                  |                               |                  | *                |                               | *                |                  |                               |
| <b>N° 6</b>  |                  |                  | *                             |                  |                  | *                             | *                |                  |                               |
| <b>N° 7</b>  | *                |                  |                               |                  |                  | *                             |                  | *                |                               |
| <b>N° 8</b>  | *                |                  |                               | *                |                  |                               | *                |                  |                               |
| <b>N° 9</b>  | *                |                  |                               |                  |                  | *                             |                  |                  | *                             |
| <b>N° 10</b> |                  | *                |                               |                  | *                |                               | *                |                  |                               |

\* Résultats de la comparaison

Le tableau 3.3-3 montre qu'en général, FS1 semble donner de meilleurs résultats et qu'en particulier, pour la semaine 15, elle possède un avantage important par rapport à FS2. Cela pourrait être dû au fait que FS1 tient compte de la contribution des sorties de marchandises d'une unité.

Le tableau 3.3-4 donne le pourcentage par catégorie de comparaison à divers niveaux d'agrégation calculé d'après les données de 2004. Les niveaux d'agrégation sont la *cellule*, la *marchandise* et la *marchandise dans la cellule*.



**Table 3.3-4**  
**Comparaisons : pourcentage par catégorie à divers niveaux d'agrégation**

| Niveau                       | Scores initiaux         |                         |                                   | Semaine 8               |                         |                                   | Semaine 15              |                         |                                   |
|------------------------------|-------------------------|-------------------------|-----------------------------------|-------------------------|-------------------------|-----------------------------------|-------------------------|-------------------------|-----------------------------------|
|                              | FS1<br>meilleure<br>(%) | FS2<br>meilleure<br>(%) | À peu<br>près les<br>mêmes<br>(%) | FS1<br>meilleure<br>(%) | FS2<br>meilleure<br>(%) | À peu<br>près les<br>mêmes<br>(%) | FS1<br>meilleure<br>(%) | FS2<br>meilleure<br>(%) | À peu<br>près les<br>mêmes<br>(%) |
| Cellule<br>(VF)              | 36,8                    | 26,7                    | 36,5                              | 41,3                    | 22,3                    | 36,4                              | 39,2                    | 16,7                    | 44,0                              |
| Marchandise<br>(M1)          | 30,8                    | 27,2                    | 42,1                              | 37,7                    | 18,8                    | 43,7                              | 29,9                    | 14,5                    | 55,6                              |
| Marchandise<br>(M2)          | 27,9                    | 25,2                    | 46,9                              | 34,6                    | 16,9                    | 48,5                              | 26,8                    | 13,0                    | 60,2                              |
| Marchandise*<br>Cellule (M1) | 16,1                    | 14,3                    | 69,6                              | 19,3                    | 8,4                     | 72,4                              | 14,1                    | 6,1                     | 79,8                              |
| Marchandise*<br>Cellule (M2) | 15,8                    | 12,3                    | 71,9                              | 17,8                    | 6,9                     | 75,3                              | 13,8                    | 5,3                     | 81,0                              |

Le tableau 3.3-4 montre qu'à tous ces niveaux, le pourcentage de cas « FS1 meilleure » est plus élevé que celui de cas « FS2 meilleure », et que le pourcentage dans la catégorie « À peu près les mêmes » est important. En outre, les résultats obtenus en se servant des données de 2005 sont semblables à ceux présentés dans le tableau.

#### 4. Conclusions

Dans notre étude, l'écart quadratique moyen est utilisé comme norme pour procéder à des comparaisons générales des deux fonctions de score, de manière à mesurer l'effet de chacune sur les estimations finales de variables financières et de variables liées aux marchandises.

L'étude montre qu'au niveau de l'*EAM totale*, la fonction de score 2 semble donner de meilleurs résultats que la fonction de score 1 pour les variables financières et pour la variable d'entrées de marchandises, tandis qu'au niveau de l'*industrie*, la fonction de score 1 semble surpasser la fonction de score 2 pour les variables financières et pour la variable de sorties de marchandises. Aux niveaux de la *cellule*, de la *marchandise* et de la *marchandise dans la cellule*, les nombres de cellules, de marchandises et de marchandises dans la cellule classés dans la catégorie « FS1 meilleure » sont plus élevés que ceux figurant dans la catégorie « FS2 meilleure ».

À cause de certaines limites de la collecte des données, de nombreux facteurs sont confondus dans cette étude générale et les résultats préliminaires présentés ici pourraient également être limités. Dans l'avenir, il serait intéressant d'utiliser simplement l'écart relatif comme norme pour les comparaisons des deux fonctions de score. Cette approche pourrait aboutir à des conclusions différentes, telles que certaines estimations préférables ayant un écart relatif plus petit, mais une variance plus grande. En outre, des études spécifiques doivent être exécutées afin de déterminer les caractéristiques optimales pour le remaniement de la fonction de score.

#### Bibliographie

- Benhin, E., Bérard, H., DeBlois, S., Lebrasseur, D., Provençal, J.S. et Turmelle, C. (2007). Unified Enterprise Survey (UES) Future Directions: Methodological Perspectives, document interne de Statistique Canada, Ottawa, Canada.
- Philips, R. (2003). La théorie et les applications de la fonction de score pour déterminer la priorité de suivi pour le Sondage annuel des manufactures, *Actes de la Section des méthodes d'enquête*, la Société statistique du Canada, 121-126.

Pursey, S. (2003), L'utilisation de la fonction de caractérisation pour optimiser les ressources de la collecte des données dans l'Enquête unifiée auprès des entreprises, *Actes de la Section des méthodes d'enquête*, la Société statistique du Canada, 117-120.