

ENQUÊTE SUR LE MILIEU DE TRAVAIL ET LES EMPLOYÉS : APERÇU DE LA MÉTHODOLOGIE

B. Lapierre, Z. Patak et J. Morel¹

RÉSUMÉ

L'Enquête sur le milieu de travail et les employés (EMTE) est une nouvelle enquête longitudinale de Statistique Canada. Elle constitue la première tentative d'envergure, au niveau national, pour recueillir de l'information à la fois sur les employeurs et sur leurs employés. Cette initiative permettra de faire des inférences sur les employeurs à partir des données des employés, et vice-versa. Pour ce faire, on liera les employeurs et les employés au niveau des micro-données, par l'entremise d'un plan d'échantillonnage à deux degrés. Dans la présente communication, on décrira la méthodologie développée pour l'EMTE.

MOTS-CLÉS : Enquête longitudinale; échantillons liés; non-réponse.

ABSTRACT

The Workplace and Employee Survey (WES) is a new Statistics Canada longitudinal survey. It constitutes the first attempt at the national level to collect information on employers and their employees. This initiative will allow us to make inferences about employers using employee data and vice-versa. To do so, we will link employers to their employees at the microdata level. In this paper, we will describe the methodology developed for WES.

KEY WORDS : Longitudinal Survey; Linked Samples; Nonresponse.

1. INTRODUCTION

Au cours des dernières années, Statistique Canada a constamment développé son expertise en matière d'analyse longitudinale. Relier longitudinalement les données d'enquête et les données administratives permet d'étudier l'impact qu'ont les changements observés dans le marché du travail sur les entreprises et leurs employés. Les clients intéressés à la compétitivité de l'industrie canadienne ont financé des enquêtes sur les innovations, l'utilisation de la technologie et le succès des petites et moyennes entreprises. Ces enquêtes commencent à mettre en lumière les facteurs expliquant la croissance et le déclin des entreprises, particulièrement l'influence des capacités d'adaptation et d'innovation des entreprises sur leur réussite.

Les facteurs déterminant la capacité d'une entreprise à répondre au changement ne peuvent être étudiés que dans un cadre longitudinal, couvrant tous

les comportements et caractéristiques des entreprises sur le plan de la performance. On s'intéresse également aux politiques et aux pratiques face aux employés puisqu'ils se doivent d'être les agents de changement dans l'entreprise. Ainsi, l'enquête idéale suivrait des échantillons liés d'entreprises et d'employés sur une période indéfinie.

L'Enquête sur le milieu de travail et les employés (EMTE) est ainsi conçue. Elle a pour objectif d'explorer les relations entre la compétitivité, les innovations, l'utilisation de la technologie et la gestion des ressources humaines du côté de l'employeur. Du côté de l'employé, elle s'attarde aux relations entre l'utilisation de la technologie, la formation, la stabilité d'emploi et les revenus. L'EMTE permettra de mettre en lumière les causes des embauches et des départs, la stabilité d'emploi et la perception qu'en ont les employés. Le type d'employés utilisant certaines technologies et la mesure dans laquelle l'utilisation de ces technologies

¹ Bruno Lapierre, Zdenek Patak et Josée Morel (lapibru@statcan.ca; patazde@statcan.ca; morejos@statcan.ca), Division des méthodes d'enquêtes auprès des entreprises, Statistique Canada, 11^{ième} étage, Édifice R.H. Coats, Ottawa, Ontario, Canada, K1A 0T6

affecte leur rémunération est un autre volet qui pourra être examiné. L'EMTE permettra également d'étudier le lien entre la gestion des ressources humaines et la performance de l'entreprise.

À la section 2, on décrira la stratégie longitudinale. La section 3 est consacrée au plan d'échantillonnage, tandis que les méthodes de collecte, de détection des valeurs aberrantes et d'imputation sont exposées à la section 4. On poursuivra avec la pondération et l'estimation à la section 5 pour conclure, à la section 6, avec les développements futurs.

2. STRATÉGIE LONGITUDINALE

L'EMTE est une enquête longitudinale. Elle utilise deux unités d'échantillonnage, l'emplacement et l'employé. L'emplacement désigne un lieu physique où se déroule une activité économique et où les données des employeurs et des employés peuvent être liées directement. La présente section s'attarde à la stratégie longitudinale.

2.1 Emplacements

L'enquête est menée sur une base annuelle, la première collecte de données étant prévue pour le printemps 1999. La première vague est constituée d'un échantillon transversal d'emplacements. À partir de la deuxième vague, l'accent sera mis sur l'analyse longitudinale. À chaque vague, les unités toujours vivantes sont gardées dans l'échantillon et réenquêtées. À la troisième vague, un échantillon de naissances sera sélectionné. Les naissances seront suivies de façon longitudinale. Cette addition à l'échantillon aura lieu à tous les deux ans et assurera une représentativité transversale lors de ces vagues. Les emplacements n'ayant aucun employé sont hors du champ de l'enquête. Si de telles unités acquièrent des employés par la suite, elles seront considérées comme des naissances et échantillonnées en conséquence.

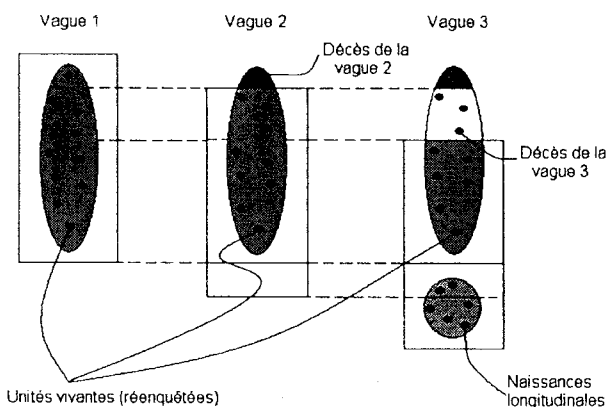
L'échantillon d'emplacements a été conçu pour être robuste longitudinalement. Pour parer à l'érosion de l'échantillon, on s'assure que chaque strate à tirage partiel contient au moins dix unités. Si, malgré cette précaution, une strate se retrouve nettement victime de l'érosion, il est possible qu'un second panel soit sélectionné pour y ajouter des unités.

Les emplacements sélectionnés demeureront dans l'échantillon pendant au moins quatre vagues. Durant cette période, la lassitude des répondants

pourrait causer une baisse des taux de réponse et la stratification risque d'être moins efficace suite aux changements survenus dans l'univers des emplacements.

Le fardeau de réponse pour les emplacements de petite taille et leurs employés sera évalué à chaque nouvelle vague. Pour le moment, aucune rotation n'est prévue. Cette stratégie pourrait être revue si la lassitude des répondants causait une érosion significative dans certaines strates. Pour remédier à la désuétude de la classification des unités, on analysera l'impact de restratifier la population et de resélectionner l'échantillon à la vague 5. Maximiser le chevauchement entre les deux échantillons permettra de conserver un maximum d'unités longitudinales dans le nouvel échantillon. Si de nombreux changements de strate surviennent, il est possible que même un chevauchement maximal ne soit pas suffisant pour conserver toutes les unités longitudinales sans augmenter les fractions d'échantillonnage. S'il s'avérait qu'un nombre significatif d'unités longitudinales soient perdues dans le processus, on explorerait alors d'autres méthodes que la restratification.

Figure 2.1 Stratégie longitudinale : emplacements



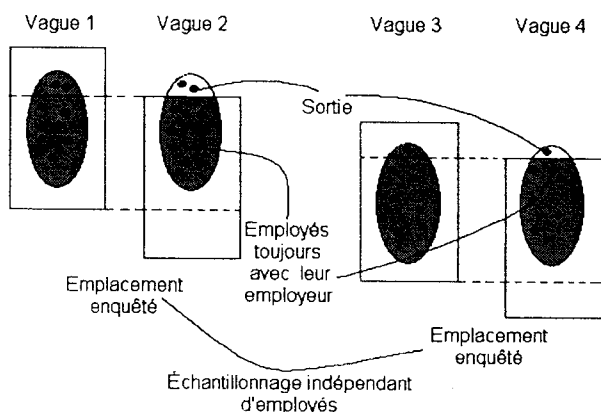
2.2 Employés

Les interviewers sélectionnent un échantillon d'employés lors de leur visite chez l'employeur. Ceci étant un processus coûteux, on limite l'échantillonnage des employés à une fois tous les deux ans. Lors de la deuxième vague, les employés qui sont toujours avec leur employeur sont gardés dans l'échantillon et répondent au même questionnaire que lors de la première vague. Un questionnaire de sortie est administré à tous ceux qui ne sont plus avec leur employeur, que ce soit parce que l'entreprise a fermé ses portes ou parce que l'employé a quitté ou perdu son emploi. On s'enquiert alors des raisons de

son départ et de son nouveau statut, c'est-à-dire : s'il s'est trouvé un nouvel emploi, s'il est au chômage ou s'il a quitté le marché du travail. Une série de questions lui est alors adressée, en fonction de son nouveau statut.

Un faible mais significatif pourcentage des employeurs ne sont pas en mesure de fournir une liste des nouvelles personnes embauchées. Plutôt que d'implanter deux stratégies d'échantillonnage différentes, selon l'information rendue disponible par l'employeur, un nouvel échantillon indépendant est tiré lors de la troisième vague.

Figure 2.2 Stratégie longitudinale : employés



3. PLAN D'ÉCHANTILLONNAGE

Pour satisfaire les objectifs de l'enquête, l'EMTE a été conçue selon un plan d'échantillonnage à deux degrés. Un échantillon d'emplacements est sélectionné au premier degré et un échantillon d'employés au deuxième degré. On exposera dans cette section les détails du plan d'échantillonnage.

3.1 Emplacements

On a créé la base de sondage à partir des informations fournies par le Registre des entreprises de Statistique Canada, une base de données où sont listées toutes les entreprises opérant au Canada. Sur le Registre des entreprises, chaque entité commerciale est structurée selon une hiérarchie à quatre niveaux. Le dernier niveau de cette hiérarchie, désigné par le terme "emplacement", est celui se rapprochant le plus de la notion de lieu de travail, et représente l'unité d'intérêt pour l'enquête.

La population cible est définie comme étant tous les emplacements situés au Canada qui ont des employés rémunérés. Quelques régions et secteurs d'activité en sont exclus.

On divise la population en quatre-vingt-quatre groupes, résultant du croisement de six régions et de quatorze secteurs d'activité industrielle. Les définitions des industries sont basées sur le Système de classification des industries d'Amérique du Nord (SCIAN), qui a remplacé la Classification type des industries (CTI). Trois strates de taille sont créées à l'intérieur de chaque groupe, la taille d'un emplacement étant basée sur son nombre d'employés. La méthode utilisée pour la stratification par taille est similaire à l'approche basée sur un modèle, expliquée plus en détails dans Särndal, Swensson et Wretman (1992, p.456)

Avec cette approche, on suppose que l'on peut utiliser le modèle $y_i = \beta x_i + \varepsilon_i$, où $\varepsilon_i \approx (0, \sigma^2 x_i^\gamma)$, et que la variable d'intérêt y (par exemple, les salaires) a une variance proportionnelle à une variable auxiliaire connue x (par exemple, le nombre d'employés). Plusieurs valeurs sont possibles pour γ . Pour l'estimateur par quotient, on utilise $\gamma = 1$. Selon Mahalanobis (1952), $\gamma = 2$ est aussi approprié pour plusieurs populations asymétriques. Une étude empirique a montré que, étant donné un ensemble de contraintes sur les CV, $\gamma = 2$ résultait en une plus petite taille d'échantillon pour l'EMTE.

On minimise la variance de l'estimateur pour obtenir une stratification par taille quasi-optimale. Pour ce faire, on procède aux étapes suivantes. À l'intérieur de chaque groupe industrie/région, on trie les unités de la population en ordre croissant de la variable auxiliaire. On divise ensuite chaque groupe en H groupes de telle sorte que les sommes des variables auxiliaires élevées à la puissance γ soient égales entre elles. On a ainsi, pour une industrie/région g ,

$$\sum_{U_{g1}} x_{(i)}^\gamma \cong \dots \cong \sum_{U_{gH}} x_{(i)}^\gamma \cong \sum_{U_g} x_i^\gamma / H.$$

Une fois la population stratifiée par taille, les tailles d'échantillon à l'intérieur de chaque strate sont déterminées à l'aide de la répartition de Neyman, sous la contrainte que les CV selon le plan au niveau du croisement industrie/région soient inférieurs à 0,09. On a ainsi obtenu un échantillon de 7400 emplacements. Les CV marginaux variaient de 0,043 à 0,058 au niveau de la l'industrie et de 0,032 à 0,035 au niveau de la région. Au niveau de la strate, les CV atteignaient un maximum de 0,276. On a ensuite tiré un échantillon aléatoire simple sans remise de chaque strate, à l'aide du Système généralisé d'échantillonnage de Statistique Canada. On a échantillonné de façon exhaustive vingt-trois strates

formées d'unités de grande taille, créant quelque 292 unités à tirage complet.

3.2 Employés

La population cible est définie comme étant tous les employés recevant une rémunération pour services rendus ou pour un congé avec solde, et pour qui l'employeur, faisant partie du champ de l'enquête pour la composante "emplacements", doit compléter un formulaire supplémentaire T-4 de Revenu Canada. Les employés sont sélectionnés en utilisant l'échantillonnage systématique à partir d'une liste fournie par l'employeur. On sélectionne trois, six, neuf ou douze employés par emplacement, selon la taille de l'emplacement. Ceci permet à la fois d'obtenir une meilleure couverture des groupes d'occupation (gestionnaires, professionnels, techniciens, etc.) et d'éviter d'obtenir des unités ayant des poids de sondage trop élevés. L'échantillon comprendra environ 40000 unités. Tel que décrit à la section 2.2, ce processus n'est effectué que lors des vagues impaires.

4. COLLECTE DES DONNÉES, DÉTECTION DES VALEURS ABERRANTES ET IMPUTATION

Les données des deux portions de l'enquête seront recueillies par les sept bureaux régionaux de Statistique Canada. On a développé des applications IPAO (interview personnelle assistée par ordinateur) et ITAO (interview téléphonique assistée par ordinateur) pour les emplacements et les employés.

4.1 Collecte des données

Lors des vagues impaires, où de nouvelles unités sont sélectionnées, la première étape de la collecte des données consiste à identifier et contacter le répondant primaire, habituellement le responsable des ressources humaines, pour recueillir quelques informations de base sur l'emplacement. Les emplacements qui déclarent n'avoir aucun employé sont hors du champ de l'enquête et ne sont pas enquêtés. Une IPAO est ensuite menée auprès du répondant primaire. Étant donné la grande variété de données recueillies pour l'EMTE, il est possible que plus d'un répondant soit nécessaire. Si tel est le cas, le répondant primaire fournira, lors du premier contact, les coordonnées des personnes en mesure de fournir l'information demandée. Lors des vagues paires, on procède par entrevues téléphoniques.

Lors de l'entrevue en personne, l'employeur fournit une liste d'employés à partir de laquelle on

sélectionne un échantillon. Par l'entremise de l'employeur, on remet aux employés échantillonnés un formulaire de participation, qu'ils retournent en indiquant s'ils acceptent de participer à l'enquête. Seul le formulaire de participation, qui contient six questions, est obligatoire, l'enquête se déroulant sur une base volontaire. L'employeur assurera un suivi auprès des employés ne retournant pas le formulaire de participation. Ceci place un fardeau indu sur l'employeur. C'est la raison pour laquelle on se limite à douze employés par emplacement. On procède ensuite à une ITAO auprès de ceux ayant accepté de participer. Les entrevues par téléphone seront menées à partir des bureaux régionaux de Statistique Canada. Lors des vagues paires, on procède également par entrevues téléphoniques.

Les applications IPAO et ITAO contiennent des règles de vérification pour réduire l'incidence d'incohérences dans les données. Il existe trois types de règles. Une règle de validité permet par exemple de s'assurer que les dépenses annuelles totales rapportées soient supérieures ou égales à zéro. Une règle de fourchette exige que ce total soit compris entre deux bornes. Une règle inter-champs assure que la somme de certaines composantes des dépenses totales, telles que les salaires, les dépenses de formation et les coûts d'implantation de technologies, ne dépasse pas ce total. À partir de la deuxième vague, on mettra en place des règles de vérification historiques.

4.2 Détection des valeurs aberrantes (Vague 1)

Une fois les données recueillies, on procède à la détection des valeurs aberrantes pour les variables continues. Les règles de vérification insérées dans les outils de collecte devraient réduire le nombre d'erreurs de réponse et de saisie. Si certains types d'erreurs ne sont toujours pas détectés, une méthode multivariée de détection des valeurs aberrantes est appliquée avant l'imputation. La méthode utilise la distance robuste de Mahalanobis, $(\mathbf{x}_i - \bar{\mathbf{x}}_{rob})^T \Sigma_{rob}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{rob})$, pour identifier les unités pour lesquelles cette statistique excède un percentile pré-spécifié de la loi χ^2 . $\bar{\mathbf{x}}_{rob}$ est l'estimateur- L_1 de localisation multivariée (Rousseeuw et Leroy, 1987), Σ_{rob} est l'estimateur robuste de dispersion, calculé en utilisant une suite de projections sur une base orthogonale et \mathbf{x}_i est le vecteur de réponses pour l'unité i . Ce type de détection des valeurs aberrantes est appliqué aux deux portions de l'enquête, au niveau des micro-données. La flexibilité inhérente à la méthode permet d'adapter le processus aux besoins de l'enquête.

La validation des données se fait aussi au niveau des macro-données. Pour certaines variables-clés, on identifie les dix observations contribuant le plus aux estimations à des niveaux agrégés. Les niveaux d'agrégation sont industrie/région, industrie/taille et industrie. Les spécialistes du sujet identifient les anomalies et corrigent les erreurs. Le cycle de validation est ensuite répété. Les valeurs qui demeurent aberrantes sont conservées mais ne sont pas utilisées comme donneurs lors de l'imputation.

4.3 Imputation (Vague 1)

La non-réponse totale sera traitée par la repondération des répondants. Cette approche suppose une non-réponse aléatoire (Little et Rubin, 1987). Différentes méthodes d'imputation seront utilisées pour la non-réponse partielle.

Trois méthodes sont utilisées pour l'imputation de la composante "emplacements": l'imputation déterministe, l'imputation à l'aide de distributions et le hotdeck pondéré. L'imputation déterministe est utilisée lorsque qu'une valeur peut être déduite avec certitude à l'aide des réponses fournies.

L'imputation à l'aide de distributions est utilisée lorsque les différentes composantes d'un total connu sont manquantes, en tout ou en partie. La distribution des composantes est calculée à un niveau agrégé et appliquée à l'enregistrement à imputer.

Le hotdeck pondéré consiste à former un bassin de donneurs potentiels, à partir duquel on sélectionne un donneur avec une probabilité proportionnelle à son poids final. On a choisi cette approche pour quatre raisons: la méthode est facile à implanter, elle procure des estimations sans biais selon le plan (Rao, 1996), un estimateur convergent de la variance peut être calculé en présence de données imputées (Rao, 1996), et la plupart des questions sont indépendantes l'une de l'autre, gardant ainsi le nombre d'ajustements post-imputation à un minimum.

L'imputation pour la composante "employés" se fait à l'aide du SGVI, le Système généralisé de vérification et d'imputation de Statistique Canada. On utilise l'imputation par plus proche voisin. D'autres méthodes ont été envisagées mais, plusieurs relations existant entre les différentes variables, toutes les méthodes d'imputation ayant une approche variable-par-variable ont mené à des incohérences dans les données. Un système de post-imputation aurait dû être très sophistiqué pour maintenir intactes toutes ces relations.

5. PONDÉRATION ET ESTIMATION

L'estimation pour l'EMTE est effectuée à l'aide du SGE, le Système généralisé d'estimation de Statistique Canada.

5.1 Emplacements

Tel que mentionné à la section 3.1, un échantillon aléatoire simple stratifié à un degré est tiré sans remise. On utilise l'estimateur par quotient séparé. On améliore la précision des estimations en ajustant les données de l'échantillon à des totaux connus de la population, fournis par le Registre des entreprises. Le calage s'effectue au niveau de la strate. Pour les strates où le nombre d'unités est insuffisant (inférieur à 15) pour obtenir des estimations stables, le calage pourra s'effectuer à un niveau agrégé.

Avant de décrire le processus d'estimation, voici quelques notations. Soit U_{1h} la population d'emplacements de taille N_{1h} et s_{1h} l'échantillon de taille n_{1h} , dans la strate h , où $h = 1, \dots, L$. Le premier terme de l'indice dénote le degré d'échantillonnage.

Soit y la variable d'intérêt, y_{1hi} sa valeur pour la $i^{\text{ième}}$ unité de la strate h au premier degré d'échantillonnage et x la variable auxiliaire (nombre d'employés). À une vague donnée, l'estimation transversale du total pour le domaine d'intérêt U_d est donnée par

$$\hat{Y}_{1,RAT}(d) = \sum_{h=1}^L \hat{Y}_{1h}(d) \frac{X_{1h}}{\hat{X}_{1h}},$$

$$\text{où } \hat{Y}_{1h}(d) = \frac{N_{1h}}{n_{1h}} \sum_{i=1}^{n_{1h}} y_{1hi}(d), \quad X_{1h} = \sum_{i \in U_{1h}} x_{1hi}, \quad \text{et}$$

$$\hat{X}_{1h} = \frac{N_{1h}}{n_{1h}} \sum_{i \in s_{1h}} x_{1hi}.$$

L'estimation de la variance est donnée par

$$\hat{V}(\hat{Y}_{1,RAT}(d)) = \sum_{h=1}^L N_{1h}^2 \left(\frac{1}{n_{1h}} - \frac{1}{N_{1h}} \right) \frac{1}{n_{1h}-1} \sum_{i=1}^{n_{1h}} (g_{1hi} e_{1hi}(d) - \overline{g_{1h} e_{1h}}(d))^2,$$

$$\text{où } g_{1hi} = \frac{x_{1h}}{\hat{x}_{1h}}, \quad e_{1hi}(d) = y_{1hi}(d) - \frac{\hat{y}_{1h}}{\hat{x}_{1h}} x_{1hi}, \quad \text{et}$$

$$\overline{g_{1h} e_{1h}}(d) = \frac{1}{n_{1h}} \sum_{i=1}^{n_{1h}} g_{1hi} e_{1hi}(d).$$

5.2. Employés

On se souviendra que la composante "employés" de l'enquête est conçue selon un plan stratifié à deux degrés. Les emplacements sont

sélectionnés au premier degré et les employés au deuxième degré. À l'intérieur d'un emplacement (défini par l'indice hi), on échantillonne systématiquement n_{2hi} unités parmi N_{2hi} . On obtient un échantillon d'employés: $s_2 = \bigcup_{h=1}^L \bigcup_{i=1}^{n_h} s_{2hi}$.

L'observation pour une unité est notée y_{2hij} .

À une vague donnée, l'estimation transversale du total pour le domaine d'intérêt U_d est donnée par

$$\hat{Y}_2(d) = \sum_{h=1}^L \frac{N_{1h}}{n_{1h}} g_{1h} \sum_{i=1}^{n_{1h}} \frac{N_{2hi}}{n_{2hi}} \sum_{j=1}^{n_{2hi}} y_{2hij}(d),$$

où $g_{1h} = \frac{x_{1h}}{\bar{x}_{1h}}$.

En présumant que la liste d'employés utilisée pour sélectionner l'échantillon est ordonnée aléatoirement, on peut approcher la variance du deuxième degré en utilisant la formule de variance pour un échantillon aléatoire simple sans remise. L'estimation de la variance totale de $\hat{Y}_2(d)$ correspond à la somme de $\hat{V}_I(\hat{Y}_2(d))$ et de $\hat{V}_{II}(\hat{Y}_2(d))$, où le premier indice désigne le degré d'estimation, c'est-à-dire,

$$\hat{V}(\hat{Y}_2(d)) = \hat{V}_I(\hat{Y}_2(d)) + \hat{V}_{II}(\hat{Y}_2(d)). \quad (5.1)$$

La première composante de (5.1) est définie par

$$\hat{V}_I(\hat{Y}_2(d)) = \sum_{h=1}^L N_{1h}^2 \left(\frac{1}{n_{1h}} - \frac{1}{N_{1h}} \right) \frac{1}{n_{1h}-1} \sum_{i=1}^{n_{1h}} (g_{1hi} e_{1hi}(d) - \overline{g_{1hi} e_{1hi}(d)})^2,$$

où $e_{1hi}(d) = \hat{Y}_{2hi}(d) - \bar{Y}_{2h}(d)$. Les deux termes dans

l'expression du résidu sont $\hat{Y}_{2hi}(d) = \frac{N_{2hi}}{n_{2hi}} \sum_{j=1}^{n_{2hi}} y_{2hij}(d)$ et

$$\bar{Y}_{2h}(d) = \frac{1}{n_{1h}} \sum_{i=1}^{n_{1h}} \hat{Y}_{2hi}(d).$$

La deuxième composante de (5.1) est de la forme

$$\hat{V}_{II}(\hat{Y}_2(d)) = \sum_{h=1}^L \frac{N_{1h}}{n_{1h}} \sum_{i=1}^{n_{1h}} g_{1hi}^2 N_{2hi}^2 \left(\frac{1}{n_{2hi}} - \frac{1}{N_{2hi}} \right) \frac{1}{n_{2hi}-1} \sum_{j=1}^{n_{2hi}} (y_{2hij}(d) - \bar{y}_{2hi}(d))^2,$$

$$\text{où } \bar{y}_{2hi}(d) = \frac{1}{n_{2hi}} \sum_{j=1}^{n_{2hi}} y_{2hij}(d).$$

6. DÉVELOPPEMENTS FUTURS

Les préparatifs pour la première vague sont en cours. La stratification, la répartition et la sélection de l'échantillon, les véhicules de collecte, la détection des valeurs aberrantes, l'imputation et l'estimation ont été mis en place. Un échantillon-test formé d'une

centaine d'unités sera enquêté en septembre 1998 pour tester les nouveaux systèmes mis en place depuis l'enquête pilote de 1995. Cet échantillon-test sera aussi utilisé ultérieurement pour tester différentes stratégies longitudinales.

Une étude débutera sous peu pour identifier la meilleure méthode d'imputation longitudinale. Il est concevable que différentes stratégies soient développées pour les entreprises de petite, moyenne et grande taille. Présentement, on ne prévoit pas de révision rétroactive des poids et des estimations. Selon le plan proposé, on remédierait à la non-réponse de vague par l'imputation, pour les unités ayant été des répondants lors d'une vague antérieure. La non-réponse de vague chronique (ou érosion) serait traitée par une repondération des répondants.

L'estimation du nombre d'employés fourni par le Registre des Entreprises est actuellement utilisée pour stratifier la population des emplacements selon leur taille. Ces estimations peuvent ne pas être à jour. Un accord conclu récemment avec Revenu Canada permettra d'obtenir des données d'emploi plus fiables. Ceci permettra à la fois d'améliorer la qualité de la stratification par taille et de réduire l'incidence des changements de strate. Les estimations obtenues seront ainsi plus efficaces.

REMERCIEMENTS

Les auteurs remercient Yanick Beaucage et Jean Dumais de Statistique Canada pour leurs commentaires utiles lors de la rédaction de cette communication.

RÉFÉRENCES

- Mahalanobis, P. C. (1952). Some aspects of the design of sample surveys. *Sankhya* 12, 1-7.
- Little, R. J. et Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- Särndal, C. E., Swensson, B., et Wretman, J. H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Rousseeuw, P. J. et Leroy, A. M. (1987). *Robust regression and outlier detection*. John Wiley & Sons.
- Rao, J. N. K. (1996). On Variance Estimation With Imputed Survey Data. *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 499-520