



**Enquête sur le financement et la croissance des petites et
moyennes entreprises de 2014**

Rapport sur la méthodologie

**Préparé par la
Division des méthodes d'enquêtes auprès des entreprises
Novembre 2014**

Enquête sur le financement et la croissance des petites et moyennes entreprises de 2014

Rapport sur la méthodologie

1 INFORMATION CONTEXTUELLE

L'Enquête sur le financement des petites et moyennes entreprises est souvent considérée comme l'enquête sur la demande de financement par les petites et moyennes entreprises (PME). Statistique Canada fait enquête auprès de ces entreprises partout au pays, afin de déterminer le type d'endettement, de location-acquisition et de financement par capitaux propres dont les petites et moyennes entreprises dépendent, ainsi que pour recueillir des données sur les tentatives en vue d'obtenir du nouveau financement. Une enquête distincte sur le financement des entreprises, du point de vue de l'offre, est aussi menée par Statistique Canada sur une base annuelle. Le présent document traite uniquement de l'enquête portant sur la demande.

En 2000, une enquête pilote auprès d'environ 2 000 entreprises a été menée pour l'année de référence 1999, afin de déterminer la pertinence de l'approche servant à recueillir ce type de données, ainsi que pour produire quelques estimations nationales pour certaines des principales questions d'intérêt. À la suite de cette enquête pilote, de nombreuses recommandations ont été faites, en vue de la mise en œuvre d'une enquête définitive à grande échelle. Tout d'abord, des changements importants ont été apportés au questionnaire, afin de le simplifier et de le raccourcir. En deuxième lieu, certaines modifications ont été apportées à la façon de définir la population des PME. Finalement, les résultats de l'enquête pilote ont servi à concevoir l'échantillon de la première enquête de l'année de référence 2000 et l'enquête supplémentaire en 2001.

Dans le cas des enquêtes de 2004, 2007 et 2011, les questions étaient similaires à celles des enquêtes de 2000 et de 2001, mais celles qui n'étaient pas pertinentes compte tenu des objectifs particuliers ont été supprimées, tandis que d'autres étaient ajoutées. Ces changements correspondent à l'objectif secondaire, mais néanmoins important de l'enquête, visant à apporter certaines améliorations et quelques ajouts au contenu et au libellé de l'enquête.

L'enquête de 2014 a été remaniée par Statistique Canada, Industrie Canada et leurs partenaires, en conservant l'ensemble des questions de base des éditions antérieures, en ajoutant certaines questions sur des nouveaux sujets d'intérêt et en laissant de côté d'autres. Avant de lancer l'enquête, le questionnaire révisé a fait l'objet d'un essai sur le terrain auprès de répondants potentiels afin de déterminer les problèmes. Ce qui a mené à d'autres révisions pour améliorer le questionnaire.

Le présent document fait état de la méthodologie utilisée pour la tenue de l'enquête sur le financement des PME en 2014.

Note aux utilisateurs

L'Enquête sur le financement des petites et moyennes entreprises de 2014 a été menée auprès des entreprises qui étaient actives au cours de la période visée par l'enquête. Les résultats de

l'enquête n'ont pas tenu compte des entreprises qui avaient cessé leurs activités en raison d'un manque de financement juste avant la tenue de l'enquête. Les entrepreneurs qui avaient tenté de lancer une entreprise, mais qui avaient échoué pendant la période visée par cette enquête, étaient aussi hors du champ de l'enquête.

2 Populations cibles

La population cible est formée de l'ensemble des unités d'une population pour laquelle on désire produire des estimations. Pour l'enquête sur le financement et la croissance des petites et moyennes entreprises, la population cible comprend toutes les entreprises qui ont entre 1 et 499 employés et un revenu brut d'au moins 30 000\$. Un certain nombre d'entreprises sont exclues de la population cible. Il s'agit des entreprises suivantes :

- 1- Organismes sans but lucratif (écoles, hôpitaux, organismes de charité, etc.)
- 2- Coentreprises
- 3- Organismes publics
- 4- Entreprises qui, d'après le Système de classification des industries de l'Amérique du Nord (SCIAN), ne sont pas d'intérêt. Elles comprennent les services publics (22), la finance et les assurances (52), la gestion de sociétés et d'entreprises (55), les services d'enseignement (61), les administrations publiques (91), la location et la location à bail de matériel automobile (5321), la location et la location à bail de machines et matériel d'usage commercial et industriel (5324), les centres de soins ambulatoires (6214), les laboratoires médicaux et d'analyses diagnostiques (6215), les autres services de soins ambulatoires (6219), les hôpitaux généraux et hôpitaux de soins chirurgicaux (6221), les hôpitaux psychiatriques et hôpitaux pour alcooliques et toxicomanes (6222), les hôpitaux spécialisés (sauf les hôpitaux psychiatriques et les hôpitaux pour alcooliques et toxicomanes) (6223) et les services communautaires d'alimentation et d'hébergement, services d'urgence et autres secours (6242).

En plus de cette grande population, un intérêt particulier est consacré à des sous-populations particulières. Celles-ci sont listées ici :

- Les coopératives (COOPS)
- Les unités du programme de financement des petites entreprises du Canada (FPEC)
- Les unités ayant conclu des contrats avec Travaux publics et services gouvernementaux (TPSGC)
- Les technologies de l'information et des communications (TIC).

Pour cette dernière population, les unités sont définies comme étant les entreprises dont le code SCIAN se trouve dans la liste des codes SCIAN à 4 chiffres présentés dans le tableau 8 de l'annexe.

Quant aux trois autres sous-populations, Statistique Canada ne dispose pas de classifications permettant l'identification de leurs unités. Comme nous allons le voir dans les sections subséquentes, l'identification des unités n'a été rendue possible que suite à l'appariement des listes d'unités reçues au Registre des entreprises.

Dans la suite de ce document, nous allons distinguer la population principale des quatre autres sous-populations.

3 Plan d'échantillonnage

3.1 Bases de sondage

La base de sondage dans une enquête est la liste des unités correspondant à la population cible de l'enquête. On y retrouve l'information de contact pour tous les éléments de base ainsi que les variables de stratification (l'emploi et la variable de revenu). L'unité statistique de l'enquête est l'entreprise telle que défini selon le Registre des Entreprises (RE).

Dans cette section, nous décrivons la création de la base de sondage associée à chacune des populations cibles. Celle de la population principale (ou de base) est construite en sélectionnant du RE toutes les entreprises qui ont entre 1 et 499 employés et un revenu brut d'au moins 30 000\$. De plus, un certain nombre d'entreprises sont exclues de la population comme décrit dans la section portant sur la définition des populations d'intérêt. La base de sondage de la population principale comprend 812 095 entreprises.

D'autre part, la base de sondage de la population TIC telle que décrite à la section précédente compte 32 071 entreprises.

Quant aux trois autres populations, les bases de sondage ont été créées essentiellement en deux étapes. À la première, des listes d'entreprises fournies à Statistique Canada ont été appariées au RE. L'exercice d'appariement a été réalisé par le Centre des Projets Spéciaux Entreprises (CPSE). Le centre a effectué l'appariement sur la base du nom légal des entreprises, du code postal et de l'adresse. Il s'agit d'un appariement exact. Une fois que les fichiers ont été appariés, la seconde étape consistait à créer une base de sondage pour chacune des populations. Comme pour la population principale, les entreprises de la base devaient avoir entre 1 et 499 employés et un revenu brut d'au moins 30 000 \$. Les exclusions appliquées à la population principale ont également été appliquées aux populations spéciales. Toutefois, dans le cas de la population FPEC, une exception a été faite afin d'inclure dans la base de sondage des unités qui pouvaient être « inactives » sur le registre des entreprises. Le concept d'inactivité est relié au fait que certaines entreprises ont que très récemment reçu un numéro d'entreprise mais n'ont pas encore d'information sur le registre des entreprises. Faute d'inclure ces unités, il est possible de laisser place à une sous-couverture de la population. Toutefois, seul le résultat de la collecte permettra d'identifier le nombre d'unités dans le champ de l'enquête et de pouvoir évaluer l'ampleur de cette population particulière.

3.2 Estimation de taille d'échantillons et allocation

Le processus d'estimation des tailles d'échantillon de la population principale s'est fait en plusieurs phases. Une première estimation des tailles d'échantillon s'est d'abord faite en fonction

des besoins d'Industrie Canada. Ensuite, pour satisfaire aux besoins statistiques des partenaires de l'enquête, les tailles d'échantillons ont été ajustées à la hausse.

La base de sondage de la population principale a été stratifiée. Les variables de stratification choisies étaient l'emploi, l'industrie, la géographie et le type d'entreprise. Le type d'entreprise réfère à « l'âge » de l'entreprise. En fait, les entreprises ayant 2 ans et moins (les entreprises en démarrage) sont séparées des entreprises ayant plus de deux ans (la population générale). De plus, les cibles de précision étaient différentes en fonction du type d'entreprise.

Les strates ont été créées en croisant les catégories de chacune des variables. Pour la population générale, la variable emploi a servi à créer 4 catégories. Les entreprises de 1 à 4 employés, les entreprises de 5 à 19 employés, les entreprises de 20 à 99 employés et finalement les entreprises d'au moins 100 employés. Quant aux entreprises en démarrage, l'emploi était divisé en 2 groupes : les entreprises ayant entre 1 et 4 employés et la catégorie des entreprises ayant au moins 5 employés. L'industrie (le code SCIAN) était utilisée pour créer 10 secteurs de l'activité économique. La liste des codes SCIAN utilisée pour définir les industries d'intérêt de l'enquête est présentée en annexe, dans le tableau 6. Finalement, la géographie a été définie selon la province d'opération de l'entreprise. À partir de celle-ci, l'univers des petites et moyennes entreprises couvraient les provinces de l'Atlantique, le Québec, l'Ontario, le Manitoba & Saskatchewan, l'Alberta et la Colombie-Britannique & Territoires.

L'estimation de la taille d'échantillon a débuté par la population des entreprises en démarrage. Nous avons supposé que le taux de réponse attendu était de 35% dans chacune des strates. L'erreur-type maximale visée pour toutes les proportions était fixée à 2,5% au niveau national. Nous avons aussi supposé que la proportion à estimer était fixée à 50%. Autrement dit, si nous supposons que la quantité que l'on veut estimer c'est la proportion des femmes propriétaires d'entreprise, cela veut dire que pour des fins d'estimation de la taille d'échantillon, la proportion des femmes propriétaires d'entreprise est fixée à 50%. La valeur de 50% permet d'avoir la taille d'échantillon la plus conservatrice car c'est la taille d'échantillon la plus élevée qu'il est possible d'obtenir parmi toutes les valeurs possibles de la proportion. En plus de ces hypothèses, l'estimation des tailles d'échantillon se faisait sous la contrainte d'un nombre minimal d'entreprises par strate. Toutes les strates devaient compter au moins 10 entreprises dans l'échantillon sinon toutes les entreprises de la strate étaient recensées. La répartition des tailles d'échantillon dans chacune des strates se faisait en utilisant la méthode de répartition proportionnelle à la racine carrée de la taille de la strate. La taille finale de l'échantillon obtenue était de 1 603 entreprises.

L'estimation de la taille d'échantillon pour la population principale s'est faite sur des hypothèses similaires à celles de la population des entreprises en démarrage. La taille finale pour cette population est obtenue après plusieurs vagues d'estimation afin de tenir compte des besoins particuliers de chacun des partenaires de l'enquête. On commençait par estimer les tailles d'échantillon en combinant la population des entreprises en démarrage et la population générale. On partait de l'hypothèse que le taux de réponse attendu dans chacune des strates était de 40% pour chacune des strates. On visait une erreur-type maximale de 2,7% pour chacune des catégories de l'emploi, de l'industrie et de la géographie. La taille minimale de chacune des strates était fixée à 10 entreprises. On utilisait la technique d'allocation proportionnelle à la

racine carrée de la taille de la strate pour la répartition des tailles dans chacune des strates. Les strates initiales étaient obtenues en croisant les catégories de l'emploi, de l'industrie et de la géographie. Après cette étape, la taille d'échantillon totale était de 17 577 entreprises.

Les vagues d'estimation subséquentes ont consisté à ajouter des nouvelles dimensions de strates dans la province de l'Ontario et de la région de l'Atlantique. Au niveau de l'Atlantique, la stratification additionnelle a consisté à ajouter la catégorie « Atlantique rurale ». Du point de vue de l'Ontario, 14 Régions Métropolitaines de Recensement (RMR) ont été définies. Pour ces RMR, on visait une erreur-type maximale de 6,5%. De plus, l'Ontario a aussi été stratifiée afin de séparer le Nord et le Sud. La cible visée était de 3,5% en termes d'erreur-type pour le Nord de l'Ontario. De façon globale, le taux de réponse attendu était de 40%. Par contre, à la différence de la première vague d'estimation, nous avons utilisé la technique d'allocation proportionnelle à la taille pour répartir les tailles d'échantillon dans les nouvelles strates. Ceci garantissait que la précision au niveau des strates initiales n'était pas amoindrie par l'ajout de nouvelles strates. Au final, la taille d'échantillon principale estimée était de 19 998 (l'échantillon de la population générale et de la population des entreprises en démarrage).

En ce qui concerne les populations spéciales, les tailles d'échantillon ont été estimées en supposant un taux de réponse de 40%. Le paramètre d'intérêt était une fois de plus la proportion. Elle a été fixée à 50% comme dans le cas de la population principale. Il n'y pas eu de stratification pour les populations TIC, COOPS et TPSGC. Les cibles de précision visées étaient 4%, 2,7% et 3% respectivement. Finalement, pour respecter les cibles de précision, les tailles estimées étaient de 388 pour TIC, 610 pour COOPS et 673 entreprises pour TPSGC.

Quant à la population FPEC, la base de sondage a été stratifiée en trois groupes d'unités. Le premier correspondait aux entreprises « actives » pour lesquelles toute l'information sur l'emploi et le revenu était disponible sur le RE. Tandis que le second groupe incluait les entreprises (pas encore actives) qui venaient de recevoir un Numéro d'Entreprise (NE) ou celles qui étaient actives mais dont l'information sur l'industrie et sur l'emploi n'étaient pas disponibles. Finalement, le troisième groupe comprenait des unités actives sur le RE qui n'avaient pas d'information sur l'industrie mais dont l'information sur l'emploi était disponible. La taille d'échantillon sélectionnée dans cette population était de 1 500 entreprises.

Le tableau suivant présente un sommaire des tailles d'échantillon et des cibles de précision visée pour la population principale ainsi que les populations spéciales.

Tableau 1 : Estimation de la qualité

Domaine	Population	Échantillon	Répondants prévus	E.-T.
Canada	812 095	19 998	7 919	0,7%
Âge de l'entreprise				
Population générale	779 386	18 395	7 358	0,7%
Entreprises en démarrage	32 709	1 603	561	2,5%
Catégorie d'emploi				
1 à 4	470 072	8 625	3 450	1,0%
5 à 19*	243 723	5 729	2 292	1,1%

20 à 99*	76 072	3 379	1,352	1,5%
100 à 499*	10 763	1 539	616	2,5%
Région				
Atlantique	50 830	2 223	889	1,9%
Atlantique rurale*	12 715	725	295	3,0%
Québec	175 950	3 634	1 454	1,4%
Ontario	290 676	6 451	2 580	1,3%
Nord de l'Ontario*	16 440	757	306	3,5%
Sudbury (580)*	2 823	209	85	6,5%
Thunder Bay (595)*	2 355	203	82	6,5%
Ottawa/Gatineau (505)*	22 069	477	193	4,8%
Kingston (521)*	2 791	209	84	6,5%
Peterborough (529)*	2 410	214	86	6,5%
Oshawa (532)*	4 715	219	88	6,5%
Hamilton (537)*	13 660	252	102	6,2%
St. Catharines - Niagara (539)*	7 242	222	88	6,5%
Kitchener (541)*	9 445	224	90	6,5%
Brantford (543)*	2 518	202	82	6,5%
Guelph (550)*	3 060	213	84	6,5%
London (555)*	8 768	224	90	6,5%
Windsor (559)*	5 656	217	88	6,5%
Barrie (568)*	3 721	216	86	6,5%
Manitoba et Saskatchewan	50 661	1 925	770	2,0%
Alberta	118 383	2 832	1 133	1,7%
C.-B. et Territoires	125 595	2 933	1 173	1,6%
Industrie				
Hébergement	64 501	1 918	767	2,0%
Construction	126 941	2 567	1 027	1,8%
Fabrication	44 427	1 619	648	2,3%
Autres services	57 640	1 624	650	2,3%
Industries primaires	51 931	1 725	690	2,2%
Services professionnels	113 450	2 168	867	2,1%
Commerce de détail	94 561	2 278	911	1,9%
Transport	52 010	1 630	652	2,4%
Commerce de gros	44 024	1 584	634	2,3%
Tous les autres	162 610	2 885	1 154	1,7%
Autres populations				
TIC	32,371	388	155	4%
TPSGC	8,235	673	269	3%
COOPS	843	610	244	2,7%
FPEC	2240	1 500	600	3%

Contrairement à la dernière itération de l'enquête, aucun traitement spécial n'a été fait sur la sous-population des Industries Fondées sur le Savoir (IFS). Autrement dit, il n'y avait pas de cible de précision à atteindre pour cette sous-population et nous n'avons donc pas estimé la taille d'échantillon requise pour cette population particulière. Parmi les 812 095 unités de la base de sondage, 22 599 entreprises étaient des IFS et 409 a été sélectionné dans l'échantillon de base. Pour la liste des codes SCIAN permettant d'identifier les IFS, se référer au tableau 7 présenté à l'annexe de ce document.

Les entreprises du secteur du tourisme ont aussi été d'intérêt dans cette itération de l'Enquête. Elles ne l'étaient pas dans l'enquête de 2010. Toutes les entreprises dans la base de sondage pour lesquelles, le code SCIAN était identifié dans le tableau 9, étaient d'intérêt. Parmi les 812 095 entreprises de la base de sondage, 22 599 entreprises étaient liées au tourisme. L'échantillon de base sélectionné comprenait 835 entreprises.

4 Collecte et traitement des données

4.1 Collecte

La collecte de cette enquête a été menée au moyen d'interviews téléphoniques assistées par ordinateur (ITAO). La période de référence pour l'enquête correspond à l'année civile 2014. Les interviews ont commencé en février 2015 et ont pris fin en mai 2015

Un sommaire des résultats de la collecte de l'échantillon de base (après les contrôles) est présenté dans le tableau ci-après. Le nombre de répondants à l'enquête était de 10 397 tandis que le nombre de non-répondants et d'unités hors du champ de l'enquête était respectivement égal à 7 308 et 2 293 entreprises.

Des données similaires sont disponibles en annexe pour les populations supplémentaires (tableaux 10 à 13).

Tableau 2 : Résultats de la collecte de l'échantillon principal

Catégories	Fréquence
Répondants	10 397
Répondants	10 397
Hors du champ de l'enquête	2 293
Éliminés	2 004
Fermeture	236
Temporairement inactifs	40
Autres cas hors du champ de l'enquête	13
Nonréponse	7 308
Non-réponse à l'échéance de l'enquête	4 976
Refus	1 623
Impossible à localiser	198
Changement de propriétaire	55
Activités saisonnières ou à temps partiel	0

4.2 Vérification des données

Les données de l'enquête ont été vérifiées par les analystes de CPSE. Pour ce faire, une application informatique permettant d'identifier les unités qui ne respectaient pas les règles de cohérence interne entre les variables a été utilisée. Les unités qui ne satisfaisaient pas aux règles de cohérence ou qui étaient manquantes étaient imputées. L'étape de l'imputation a été effectuée par la division des méthodes d'enquête auprès des entreprises.

Le questionnaire compte un certain nombre de variables quantitatives. Ces variables sont analysées pour le traitement des valeurs aberrantes. Elles sont identifiées en utilisant la technique proposée dans Hidiroglou-Berthelot (1986). Cette méthode est implémentée dans le système généralisé de vérification et d'imputation développée par Statistique Canada (BANFF). Toutes les valeurs aberrantes sont vérifiées par les analystes de CPSE et par Industrie Canada. Le traitement des cas problématiques se fait au moyen de l'imputation.

4.3 Taux d'achèvement et de réponse

Le taux d'achèvement est défini comme le nombre de répondants divisé par l'échantillon total. Le taux d'achèvement de l'enquête principale est de 52 %. Il s'agit d'un taux plus élevé que le taux d'achèvement de l'enquête de 2011 qui était de 40 %.

Pour déterminer le taux de réponse, le nombre total d'unités dans le champ de l'enquête doit être déterminé. Ce chiffre comprend tous les répondants, les activités saisonnières ou à temps partiel comprises dans le champ de l'enquête et une estimation du nombre d'unités dans le champ de l'enquête parmi les non-répondants qui restent. Cette estimation est calculée à partir de la proportion d'unités connues dans le champ de l'enquête et d'unités connues hors du champ de l'enquête, ce qui donne le « taux d'unités dans le champ de l'enquête ».

Le taux d'unité dans le champ de l'enquête peut être calculé comme suit :

$$\frac{\text{unités connues dans le champ de l'enquête}}{\text{Unités connues dans le champ de l'enquête} + \text{Unités connues hors du champ de l'enquête}}$$

En utilisant les chiffres présentés au tableau 2, ce taux est estimé à 81,9%. À partir de ce taux, nous pouvons donc estimer le taux des unités dans le champ de l'enquête comme suit :

$$\text{Total des unités dans le champ de l'enquête} = 10\,397 + 0 + 81,9\% * (7\,308) \approx 16\,384$$

Le taux de réponse est par la suite calculé et correspond au nombre de répondants divisé par le nombre d'unités estimées dans le champ de l'enquête :

$$\text{Taux de réponse} = \frac{\text{Répondants}}{\text{Total des unités dans le champ de l'enquête}} = 63,5 \%$$

On a aussi demandé aux répondants s'ils souhaitaient partager les résultats du questionnaire avec l'Institut de la Statistique de Québec. Une ventilation des répondants, du taux d'achèvement, du taux de réponse et du taux de partage par domaine figure dans le tableau 3 pour l'échantillon de base. Dans le tableau 4, des chiffres similaires sont présentés pour les populations additionnelles.

Tableau 3 : Taux d'achèvement, de réponse et de partage pour la population de base

Domaine	Échantillon	Répondants	Taux d'achèvement	Taux de réponse	Taux de partage
NATIONAL	19 998	10 397	51,99%	63,46%	95,14%
1 à 4 employés	8 625	3 604	41,79%	62,04%	94,18%
5 à 19 employés	6 345	3 714	58,53%	62,93%	94,84%
20 à 99 employés	3 476	2 165	62,28%	66,08%	96,95%
100 à 499 employés	1 552	914	58,89%	67,59%	96,18%
Atlantique	2 222	1 215	54,68%	61,93%	.
Colombie-Brit. et Territoires	2 936	1 411	48,06%	57,94%	.
Ontario	6 452	3 436	53,26%	63,56%	.
Prairies	4 756	2 565	53,93%	69,60%	.
Québec	3 632	1 770	48,73%	60,63%	95,14%
Services d'hébergement	1 885	974	51,67%	57,24%	96,49%
Construction	2 561	1 373	53,61%	64,12%	96,17%
Fabrication	1 591	962	60,47%	66,81%	94,12%
Autres	3 165	1 319	41,68%	55,96%	95,00%
Autres secteurs	1 613	853	52,88%	62,86%	97,35%
Industrie primaire	1 650	883	53,52%	64,67%	94,48%
Service professionnel	2 128	1 079	50,71%	71,24%	97,28%
Commerce de détail	2 251	1 310	58,20%	65,26%	93,97%
Transport	1 610	765	47,52%	64,47%	92,54%
Commerce de gros	1 544	879	56,93%	66,78%	93,24%

Tableau 4 : Taux d'achèvement, de réponse et de partage pour les populations additionnelles

Populations	Domaine	Échantillon	Répondants	Taux d'achèvement	Taux de réponse
TIC	NATIONAL	388	171	44,07%	76,03%
TPSGC	NATIONAL	673	419	62,26%	68,50%
COOPS	NATIONAL	610	420	68,85%	78,52%
FPEC	NATIONAL	1500	743	49,53%	55,86%

4.4 Imputation

Après la vérification des microdonnées, nous avons créé une variable pour chacune des variables de l'enquête afin de déterminer celles qui avaient échoué aux règles de vérification ou celles dont les valeurs étaient manquantes. Deux catégories d'unités ont été créées : les cas de non-réponse totale et les cas de non-réponse partielle. Les unités de non-réponse totale ont été traitées au moyen d'une pondération, les poids des unités déclarantes de la même strate ayant été corrigés de sorte que les unités non-répondantes soient aussi représentées. Le traitement des cas de non-réponse partielle a été effectué par imputation.

L'imputation due à la non-réponse partielle a été effectuée de façon indépendante pour chacune des diverses populations cibles. Par exemple, les données complètes de l'échantillon des coopératives ne pouvaient pas servir à imputer les données manquantes des unités de l'échantillon des technologies de l'information et des communications. Les variables manquantes ont fait l'objet d'une imputation par la méthode du plus proche voisin. On utilise la fonction de distance minimax pour trouver le donneur le plus proche. La fonction de distance minimax détermine le donneur le plus proche comme étant celui dont la valeur des variables d'appariement présente la plus petite différence maximale absolue par rapport à celle des variables d'appariement du receveur. Pour la plupart des variables, la variable d'appariement utilisée était les données de l'emploi. L'imputation a été effectuée au sein des groupes d'unités appelés « classes d'imputation ». Ces classes d'imputation ont été formées d'unités de taille semblable (emploi), du même niveau géographique et du même secteur d'activité.

Un nombre minimal d'unités a été requis dans chacune des classes d'imputation. Quand les classes d'imputation étaient trop petites, de plus grandes étaient créées en les combinant.

Aux fins de la cohérence interne (cohérence entre les variables du même enregistrement), la valeur des variables manquantes ou incohérentes a été imputée selon l'ordre dans lequel elles

figuraient dans le questionnaire. De cette façon, si une question posée à une certaine étape du questionnaire a donné lieu à une imputation, on a pu s'en servir comme variable d'appariement pour une question située plus loin dans le questionnaire.

La plupart des imputations de données d'enquête ont été faites de façon automatisée à l'aide de BANFF, un système généralisé conçu par Statistique Canada.

Le taux d'imputation d'une variable donnée est défini de la façon suivante :

$$\text{Taux d'imputation} = \frac{\text{Nombre d'unités imputées}}{\text{Nombre total d'unités qui devraient répondre à cette question}}$$

Dans le tableau 5, nous présentons les taux d'imputation minimal et maximal dans chacune des sections pour la population de base. La distribution des variables est également présentée en fonction des taux d'imputation observés.

Pour une section donnée, le taux d'imputation est calculé sur l'ensemble des variables de la section. Le taux d'imputation minimal (maximal) de la section correspond au taux le plus bas (le plus élevé) observé pour l'ensemble des variables de la dite section. Par exemple, à partir du tableau 5, nous pouvons voir que le taux d'imputation minimal sur l'ensemble des 21 variables de la section « Financement Général » est de 3,62% tandis que le taux d'imputation maximal dans la même section est de 20,98%.

La seconde partie du tableau 5 montre la distribution des variables présentes dans une section donnée en fonction de différents intervalles des taux d'imputation. Par exemple, dans la section sur le Financement Général, nous pouvons voir que 12 des 21 variables ont un taux d'imputation qui est inférieur à 15% et le reste des variables a un taux d'imputation compris entre 15 et 30%.

Les tableaux 14 à 17 présentent les taux d'imputation similaires pour les populations additionnelles. Ils sont présentés en annexe de ce document.

Tableau 5 : Fourchettes de taux d'imputation selon la section pour l'échantillon de base

Sections	Taux minimum (%)	Taux maximum (%)	< 15%	15-30%	30-40%	40-50%	50-100%
Questions de sélection	0,00	0,22	6	0	0	0	0
Financement général	3,62	20,98	12	9	0	0	0
Prêts hypothécaires	0,61	78,89	2	1	2	0	8
Marge de crédit	0,88	59,80	2	1	8	0	1
Prêts à termes	0,74	77,59	2	2	1	0	8
Carte de crédit commercial	0,94	80,00	3	0	0	1	8
Financement par emprunt général	5,96	16,95	12	3	0	0	0
Financement par location-acquisition	0,88	84,25	3	1	0	0	1
Financement par crédit commercial	1,33	13,46	3	0	0	0	0
Financement par capitaux propres	1,29	63,74	1	0	0	0	1

Subventions ou contributions non remboursables de l'état	1,19	27,67	1	2	0	0	0
Raisons pour lesquelles il n'y a pas eu de demande de financement	3,32	3,32	1	0	0	0	0
Renseignements généraux sur l'entreprise	4,79	62,66	46	0	23	5	1
Renseignements sur la personne principalement responsable de la prise de décisions concernant l'entreprise	4,02	21,83	8	1	0	0	0

4.5 Pondération (traitement de la non-réponse par unité)

La pondération est l'étape du processus où les poids finaux sont calculés. Ces derniers servent à la production des estimations ponctuelles et des estimations de variance.

La pondération de la population générale est effectuée en deux étapes. Lors de la première étape, les poids de sondage sont calculés dans chacune des strates comme le quotient entre la taille de la strate et le nombre d'unités sélectionnées dans l'échantillon. Ensuite, à la deuxième étape, les poids finaux sont calculés. Les unités non-répondantes sont identifiées et exclues de toutes les strates. Des strates avec peu de répondants sont regroupées. À l'intérieur des nouvelles strates, les poids finaux sont calculés par le quotient du nombre total d'entreprises de chaque strate sur le nombre d'entreprises répondantes de l'échantillon.

La pondération des populations spéciales est effectuée en deux étapes comme dans le cas de la population générale. À la première étape, les poids de sondage sont calculés sur l'ensemble de l'échantillon et à la seconde étape, les poids de sondage sont ajustés pour tenir compte de la non-réponse par unité. Contrairement à la population générale, les poids de sondage et les poids ajustés pour la non-réponse sont calculés au niveau de l'échantillon c'est-à-dire qu'aucune strate ni classe de repondération a été créée.

5 Confidentialité

De nombreux tableaux sont produits à partir des données de l'enquête. La protection de la confidentialité des variables quantitatives se fait en utilisant la règle du seuil et la règle C2 du système généralisé de confidentialité (G-CONFID). Tandis que la protection de la confidentialité pour les variables qualitatives se fait à partir de la règle du seuil.

La règle du seuil consiste à déterminer un nombre minimum d'unités répondantes acceptables par cellule. Toutes les cellules dont le nombre de répondants est inférieur au seuil fixé sont identifiées comme étant des cellules sensibles. Quant à la règle C2, elle consiste à établir la contribution de toutes les unités présentes dans chacune des cellules. Une cellule est dite sensible selon cette règle, lorsque la contribution d'une ou de plusieurs unités de la dite cellule domine largement sur celle des autres unités répondantes.

La méthodologie de protection de la confidentialité a été développée suite à des consultations avec le centre de ressources en contrôle de divulgation et le CPSE.

Bibliographie

Hidiroglou, M.A., and Berthelot, J.-M. (1986). "Statistical Editing and Imputation for Periodic Business Surveys" *Survey Methodology*, 12, 73-83.

Annexe A – Tableaux additionnels

Tableau 6 : Codes du SCIAN inclus dans les domaines d'industrie

Domaine	Codes du SCIAN inclus
Industries primaires	11,21
Construction	23
Fabrication	31-33
Commerce de gros	41
Commerce de détail	44-45
Transport	48-49
Services professionnels	54
Hébergement	72
Autre services	81
Autre	Tous les autres codes du SCIAN faisant partie du champ de l'enquête

Tableau 7 : IFS définies par Industrie Canada

SCIAN	Description
325410	Fabrication de produits pharmaceutiques et de médicaments
333310	Fabrication de machines pour le commerce et les industries de services
334110	Fabrication de matériel informatique et périphérique
334210	Fabrication de matériel téléphonique
334220	Fabrication de matériel de radiodiffusion, de télédiffusion et de communication sans fil
334290	Fabrication d'autres types de matériel de communication
334310	Fabrication de matériel audio et vidéo
334410	Fabrication de semi-conducteurs et d'autres composants électroniques
334511	Fabrication d'instruments de navigation et de guidage
334512	Fabrication d'appareils de mesure et de commande et d'appareils médicaux
335920	Fabrication de fils et de câbles électriques et de communication
336410	Fabrication de produits aérospatiaux et de leurs pièces
511210	Éditeurs de logiciels
512110	Production de films et de vidéos
512190	Postproduction et autres industries du film et de vidéo
515210	Télévision payante et spécialisée
517111	Fournisseurs de services de télécommunications par fil (sauf câblodistribution)
517112	Câblodistribution et autres activités de distribution d'émissions de télévision
517210	Télécommunications sans fil (sauf par satellite)
517410	Télécommunications par satellite
517910	Autres services de télécommunications
518210	Traitement de données, hébergement de données et services connexes
541360	Services de prospection et de levé géophysiques
541370	Services d'arpentage et de cartographie (sauf les levés géophysiques)
541510	Conception de systèmes informatiques et services connexes
541620	Services de conseils en environnement
541690	Autres services de conseils scientifiques et techniques
541710	Recherche et développement en sciences physiques, en génie et en sciences de la vie
541990	Tous les autres services professionnels, scientifiques et techniques

Tableau 8 : TIC définies par Industrie Canada

SCIAN	Description
3341	Fabrication de matériel informatique et périphérique
3342	Fabrication de matériel de communication
3343	Fabrication de matériel audio et vidéo
3344	Fabrication de semi-conducteurs et d'autres composants électroniques
3346	Fabrication et reproduction de supports magnétiques et optiques
4173	Grossistes-marchands d'ordinateurs et de matériel de communication
5112	Éditeurs de logiciels
5182	Traitement de données, hébergement de données et services connexes
5415	Services professionnels, scientifiques et techniques
8112	Réparation et entretien de matériel électronique et de matériel de précision

Tableau 9 : Tourisme définies par Industrie Canada

SCIAN	Description
4811	Transport aérien régulier
4812	Transport aérien non régulier
4821	Transport ferroviaire
4831	Transport hauturier, côtier et sur les Grands Lacs
4832	Transport sur les eaux intérieures
4851	Services urbains de transport en commun
4852	Transport interurbain et rural par autocar
4853	Services de taxi et de limousine
4854	Transport scolaire et transport d'employés par autobus
4855	Services d'autobus nolisés
4859	Autres services de transport en commun et de transport terrestre de voyageurs
4871	Transport terrestre de tourisme et d'agrément
4872	Transport par eau de tourisme et d'agrément
4879	Autres services de transport de tourisme et d'agrément
51213	Présentation de films et de vidéos
5615	Services de préparation de voyages et de réservation
7111	Compagnies d'arts d'interprétation
7112	Sports-spectacles
7115	Artistes, auteurs et interprètes indépendants
7121	Établissements du patrimoine
7131	Parcs d'attractions et salles de jeux électroniques
7132	Jeux de hasard et loteries
7139	Autres services de divertissement et de loisirs
7211	Hébergement des voyageurs
721114	Hôtels
721198	Tous les autres services d'hébergement des voyageurs
721211	Parcs pour véhicules récréatifs (VR) et campings
721212	Camps de chasse et de pêche
7224	Débits de boissons alcoolisées

Tableau 10 : Résultats de la collecte de l'échantillon TIC

Catégories	Fréquence
Répondants	171
Répondants	171
Hors du champ de l'enquête	124
Éliminés	118
Fermeture	6
Temporairement inactifs	0
Autres cas hors du champ de l'enquête	0
Nonréponse	93
Non-réponse à l'échéance de l'enquête	61
Refus	21
Impossible à localiser	9
Changement de propriétaire	0
Activités saisonnières ou à temps partiel	0
Autres cas de non-réponse	2
Total	388

Tableau 11 : Résultats de la collecte de l'échantillon TPSGC

Catégories	Fréquence
Répondants	419
Répondants	419
Hors du champ de l'enquête	42
Éliminés	37
Fermeture	3
Temporairement inactifs	1
Autres cas hors du champ de l'enquête	1
Nonréponse	212
Non-réponse à l'échéance de l'enquête	160
Refus	44
Impossible à localiser	1
Changement de propriétaire	1
Activités saisonnières ou à temps partiel	0
Autres cas de non-réponse	6
Total	673

Tableau 12 : Résultats de la collecte de l'échantillon COOPS

Catégories	Fréquence
Répondants	420
Répondants	420
Hors du champ de l'enquête	59
Éliminés	51
Fermeture	7
Temporairement inactifs	0
Autres cas hors du champ de l'enquête	1
Nonréponse	131
Non-réponse à l'échéance de l'enquête	97
Refus	30
Impossible à localiser	0
Changement de propriétaire	1
Activités saisonnières ou à temps partiel	0
Autres cas de non-réponse	3
Total	610

Tableau 13 : Résultats de la collecte de l'échantillon FPEC

Catégories	Fréquence
Répondants	743
Répondants	420
Hors du champ de l'enquête	95
Éliminés	89
Fermeture	6
Temporairement inactifs	0
Autres cas hors du champ de l'enquête	0
Nonréponse	662
Non-réponse à l'échéance de l'enquête	369
Refus	87
Impossible à localiser	21
Changement de propriétaire	0
Activités saisonnières ou à temps partiel	0
Autres cas de non-réponse	185
Total	1 500

Tableau 14 : Fourchettes de taux d'imputation selon la section pour l'échantillon de TIC

Sections	Taux minimum (%)	Taux maximum (%)	< 15%	15-30%	30-40%	40-50%	50-100%
Questions de sélection	0,00	0,00	6	0	0	0	0
Financement général	4,68	38,46	5	7	9	0	0
Prêts hypothécaires	0,00	40,00	10	0	1	1	0
Marge de crédit	0,00	25,00	4	8	0	0	0
Prêts à termes	1,17	100,00	1	1	1	0	10
Carte de crédit commercial	0,58	100,00	3	0	1	0	8
Financement par emprunt général	5,71	20,00	12	3	0	0	0
Financement par location-acquisition	0,58	100,00	4	0	0	0	1
Financement par crédit commercial	1,17	13,33	3	0	0	0	0
Financement par capitaux propres	1,17	37,5	1	0	1	0	0
Subventions ou contributions non remboursables de l'état	0,58	10,00	3	0	0	0	0
Raisons pour lesquelles il n'y a pas eu de demande de financement	3,96	3,96	1	0	0	0	0
Renseignements généraux sur l'entreprise	6,43	33,33	46	24	5	0	0
Renseignements sur la personne principalement responsable de la prise de décisions concernant l'entreprise	4,09	15,49	8	1	0	0	0

Tableau 15 : Fourchettes de taux d'imputation selon la section pour l'échantillon de COOPS

Sections	Taux minimum (%)	Taux maximum (%)	< 15%	15-30%	30-40%	40-50%	50-100%
Questions de sélection	0,00	0,00	6	0	0	0	0
Financement général	4,76	54,93	3	7	2	0	9
Prêts hypothécaires	0,00	33,33	12	0	1	0	0
Marge de crédit	0,24	66,67	2	2	0	0	8
Prêts à termes	0,24	100	4	1	0	0	8
Carte de crédit commercial	0,48	100	3	0	0	1	8
Financement par emprunt général	3,03	15,38	12	3	0	0	0
Financement par location-acquisition	0,48	100,00	4	0	0	0	1
Financement par crédit commercial	0,00	8,67	3	0	0	0	0
Financement par capitaux propres	0,24	13,33	2	0	0	0	0
Subventions ou contributions non remboursables de l'état	0	12,5	3	0	0	0	0
Raisons pour lesquelles il n'y a pas eu de demande de financement	0,79	0,79	1	0	0	0	0
Renseignements généraux sur l'entreprise	2,38	63,64	46	0	12	1	16
Renseignements sur la personne principalement responsable de la prise de décisions concernant l'entreprise	4,76	64,86	5	3	0	0	1

Tableau 16 : Fourchettes de taux d'imputation selon la section pour l'échantillon de TPSGC

Sections	Taux minimum (%)	Taux maximum (%)	< 15%	15-30%	30-40%	40-50%	50-100%
Questions de sélection	0,00	0,24	6	0	0	0	0
Financement général	4,06	25,00	3	18	0	0	0
Prêts hypothécaires	1,19	77,78	2	1	2	0	8
Marge de crédit	1,19	63,64	2	1	1	0	8
Prêts à termes	0,72	80,00	2	3	0	0	8
Carte de crédit commercial	0,95	100	3	0	0	1	8
Financement par emprunt général	5,39	16,56	13	2	0	0	0
Financement par location-acquisition	0,95	80,00	3	1	0	0	1
Financement par crédit commercial	1,91	17,24	2	1	0	0	0
Financement par capitaux propres	1,43	66,67	1	0	0	0	1
Subventions ou contributions non remboursables de l'état	1,43	26,32	1	2	0	0	0
Raisons pour lesquelles il n'y a pas eu de demande de financement	5,00	5,00	1	0	0	0	0
Renseignements généraux sur l'entreprise	4,30	68,29	39	30	5	0	1
Renseignements sur la personne principalement responsable de la prise de décisions concernant l'entreprise	4,06	24,44	8	1	0	0	0

Tableau 17 : Fourchettes de taux d'imputation selon la section pour l'échantillon de FPEC

Sections	Taux minimum (%)	Taux maximum (%)	< 15%	15-30%	30-40%	40-50%	50-100%
Questions de sélection	0	0	6	0	0	0	0
Financement général	2,29	11,35	21	0	0	0	0
Prêts hypothécaires	0,27	55,56	4	1	1	0	7
Marge de crédit	1,08	42,39	3	7	0	2	0
Prêts à termes	0,67	80	4	1	1	0	7
Carte de crédit commercial	0,81	61,11	3	0	0	1	8
Financement par emprunt général	2,54	5,56	15	0	0	0	0
Financement par location-acquisition	1,21	73,33	4	0	0	0	1
Financement par crédit commercial	1,75	16,13	2	1	0	0	0
Financement par capitaux propres	2,02	65,22	1	0	0	0	1
Subventions ou contributions non remboursables de l'état	2,15	27,94	1	2	0	0	0
Raisons pour lesquelles il n'y a pas eu de demande de financement	6,33	6,33	1	0	0	0	0
Renseignements généraux sur l'entreprise	4,17	60,61	46	0	6	7	16
Renseignements sur la personne principalement responsable de la prise de décisions concernant l'entreprise	3,36	17,88	8	1	0	0	0