

COMPARAISON ENTRE DEUX SOURCES DE DONNÉES : LES DONNÉES FISCALES ET LES DONNÉES DES RÉPONDANTS DE L'ENQUÊTE UNIFIÉE AUPRÈS DES ENTREPRISES DE STATISTIQUE CANADA

Eric Pelletier, Wisner Jocelyn et Lenka Mach ¹

RÉSUMÉ

Depuis 2002, l'utilisation des données fiscales en remplacement aux données d'enquête s'est accrue, et ce pour plusieurs enquêtes auprès des entreprises de Statistique Canada. Le tout, notamment dans le but de réduire le fardeau de réponse et le coût des enquêtes. Ceci a été rendu possible grâce à une entente entre Statistique Canada et l'Agence du revenu du Canada. En 2005, le Projet d'intégration des données a été développé à Statistique Canada dans le but d'étudier comment les données fiscales peuvent être utilisées de façon optimale dans les enquêtes. Une des premières étapes de ce projet a été de comparer les concepts et les définitions utilisées pour les données fiscales et les données recueillies auprès des entreprises. Les données provenant de l'Enquête unifiée auprès des entreprises (EUE) seront utilisées dans cet article pour comparer les deux sources de données.

MOTS CLÉS : Concepts et définitions; erreur de mesure; intégration des données.

ABSTRACT

Since 2002, several Statistics Canada enterprise surveys are making more use of tax data in lieu of survey data to reduce response burden and costs. This is feasible because of an agreement with the Canada Revenue Agency. In 2005, the Data Integration Project was launched at Statistics Canada in order to determine the optimal methods for surveys' use of tax data. The first step of the project was to compare the concepts and definitions used for tax data and the data collected via a survey. In this paper, the Unified Enterprise Survey (UES) data will be used to compare the two data sources.

KEY WORDS: Concepts and definitions, Data integration, Measurement error.

1. INTRODUCTION

1.1 Projet d'intégration des données

Le Projet d'intégration des données (PIDD), comme son nom l'indique, vise à combiner des données provenant de diverses sources afin de produire des estimations respectant les standards de qualité de Statistique Canada. Reconnaisant le fait que l'implantation d'une méthodologie appropriée exige une bonne compréhension des données et donc des concepts sous-jacents, nous avons entrepris le projet en analysant les différents concepts et définitions de variables utilisés au sein des deux sources de données.

Notre objectif ici n'est pas de définir ou de re-définir les concepts utilisés mais plutôt d'arriver à une compréhension suffisante de ceux-ci, afin que nous puissions proposer les méthodes les mieux adaptées pour l'obtention et l'analyse des données correspondant aux concepts à mesurer.

¹ Eric Pelletier, Wisner Jocelyn, Division des méthodes d'enquêtes auprès des entreprises, 11^e étage, édifice R.-H.-Coats, Statistique Canada, Ottawa, Ontario, K1A 0T6, eric.pelletier@statcan.ca, wisner.jocelyn@statcan.ca. Lenka Mach, Division des méthodes d'enquêtes sociales, 15^e étage, édifice R.-H.-Coats, Statistique Canada, Ottawa, Ontario, K1A 0T6, lenka.mach@statcan.ca.

Les données proviennent de sources différentes (données fiscales versus données d'enquête), ce qui implique des modes de collecte différents. Nous devons vérifier que les divergences entre les données fiscales et celles de l'enquête ne cachent pas un problème d'erreur de mesure ou une disparité dans le processus de collecte de chacune des deux sources de données.

Concrètement, nous aimerions donc nous assurer que les variables clés découlant des concepts comptables provenant de l'état des résultats des entreprises aient la même signification, d'un point de vue statistique, qu'on les obtienne via un questionnaire d'enquête ou d'un fichier administratif.

1.2 Programme de remplacement des données d'enquête par des données fiscales

À l'origine du Projet d'intégration des données, le Programme de remplacement des données d'enquête par des données fiscales a été mis en place pour l'année de référence 2003 pour l'Enquête unifiée auprès des entreprises (EUE) (voir Laniel et Royce (1998) et Pelletier (2004)). Ce projet a pour objectif une utilisation accrue des données fiscales dans les programmes d'enquête dans le but de réduire de fardeau de réponses des entreprises, de réduire le coût des programmes et de potentiellement améliorer la qualité des données.

Comme le nom du programme l'indique, des données fiscales sont utilisées en remplacement de données d'enquête pour une portion de l'échantillon principal. Ainsi, aucun questionnaire n'est envoyé à ces entreprises et les données sont prises directement dans les fichiers administratifs. À noter que seules les unités incorporées (unités T2) ayant une structure simple, i.e. des entreprises incorporées opérant dans une seule province et dans une seule industrie, sont éligibles pour la substitution de leurs données d'enquête par des données fiscales.

L'article est organisé ainsi : La section 2 fournit une brève discussion sur la définition des principaux concepts clés et les variables dérivées de ces concepts. À la section 3, nous décrivons une étude menée et à la section 4, nous discutons brièvement des résultats obtenus tout en présentant quelques pistes expliquant les différences entre les résultats provenant de sources de données différentes. Nous concluons finalement à la section 5, non sans avoir proposé quelques recommandations.

2. CONCEPTS ET DÉFINITIONS

Un des objectifs de l'Enquête unifiée auprès des entreprises (EUE) implique la production d'estimations provinciales de qualité afin de déterminer la répartition des recettes provenant de la taxe de vente harmonisée. Pour y arriver, les enquêtes économiques couvertes par l'EUE ont adopté un certain nombre de concepts en commun pour entre autres, mesurer les revenus, les dépenses et ultimement les profits. Cela correspond aux concepts trouvés généralement dans la portion de l'état des résultats de l'entreprise. Par exemple, ce sommaire comptable périodiquement effectué par chaque entreprise, fournit des données sur les revenus générés, le coût des marchandises vendues ainsi que les dépenses contractées dans l'exploitation de l'entreprise.

2.1 Revenus

La fédération internationale des comptables définit le revenu en ces termes : « valeur brute, pour une période de déclaration, des avantages économiques ou du potentiel de service qui donnent lieu à une augmentation des capitaux propres, exception faite des augmentations liées aux apports des propriétaires ».

La grande majorité des questionnaires d'enquêtes auprès des entreprises à Statistique Canada demandent aux répondants de rapporter le revenu dans les termes suivants : « déclarez le total de vos ventes, de vos commissions et de vos autres recettes ». Cette définition se rapproche beaucoup de celle de l'Agence du revenu du Canada (ARC) qui spécifie sur son formulaire que le revenu est constitué de : « ventes, commissions et autres rétributions ». Ces définitions semblent donc être similaires et semblent être conformes aux principes comptables généralement acceptés. Donc, s'il y a une différence entre le revenu collecté par une enquête et le revenu fourni par l'ARC, elle doit provenir soit de la structure des entités comparées, soit de la période de référence différente des données déclarées ou de différentes erreurs de

mesure. Afin de minimiser l'impact de ces facteurs, nous avons effectué une étude dont nous donnerons plus de détails à la prochaine section.

2.2 Dépenses

Du point de vue des dépenses, la fédération internationale des comptables définit les Dépenses comme: « une diminution des avantages économiques ou du potentiel de service, sous forme de sorties ou de diminution d'actifs ou de constitution de passifs, au cours d'une période de déclaration, qui donnent lieu à une diminution des capitaux propres, exception faite de celles découlant des distributions d'actifs aux propriétaires ». Cette définition est d'un point de vue pratique l'inverse de celle portant sur les revenus. Cependant, les sources de dépenses semblent beaucoup plus nombreuses que celles de revenus, qu'on songe aux dépenses de publicité, aux salaires, aux impôts fonciers, aux taxes d'affaires, etc. Comparer les dépenses provenant de sources diverses est un peu plus compliqué en effet.

Considérons la façon dont les salaires sont rapportés au niveau des questionnaires d'enquêtes auprès des entreprises à Statistique Canada. On demande généralement de déclarer les Salaires et traitements des employés dans un premier temps et dans un deuxième temps on demande de rapporter la part des avantages sociaux versée par l'employeur. Du côté de l'ARC on demande de rapporter les salaires et traitements (y compris les cotisations de l'employeur). Pour qu'il y ait correspondance entre ces deux définitions il faudrait que les répondants comprennent que les cotisations de l'employeur correspondent à la part des avantages sociaux versée par l'employeur. Est-ce le cas?

Une chose est sûre cependant, faire correspondre les définitions de dépenses du côté fiscal à celles des enquêtes semble être potentiellement plus complexe. Ces différences au niveau de la façon de rapporter les dépenses nous font penser qu'au-delà des correspondances définies (« mapping »), il pourrait subsister des différences dues non pas à des anomalies dans les données mais à des façons différentes de les rapporter ou encore à une interprétation différente du répondant selon qu'il remplit un questionnaire d'enquête ou un formulaire d'impôt.

3. ÉTUDE SUR LES VARIABLES CLÉS DE L'EUE

3.1 Introduction

Plusieurs études ont été réalisées dans le passé comparant les données d'enquête aux données fiscales (Pelletier et autres, 2003, Tebrake 2003a et 2003b). Ces études ont surtout considérées la corrélation entre les données d'enquête et les données fiscales pour une utilisation pour le Programme de remplacement des données d'enquête par des données fiscales. Toutefois, considérant que les données fiscales sont maintenant de meilleures qualités, il est justifié de refaire une étude à l'aide de données plus récentes et d'explorer une autre méthodologie pour la comparaison des données.

Ainsi, l'objectif de cette nouvelle étude a été d'expliquer les différences entre les données d'enquête et les données fiscales et d'explorer plus en profondeur les différences pour des unités qui ont un impact important dans les comparaisons.

3.2 Enquêtes utilisées et variables clés

Nous rapportons dans cette section la description ainsi que les résultats de cette étude menée à partir de quelques enquêtes de l'EUE, spécifiquement celles qui ont adopté la stratégie de remplacement des données d'enquête par des données administratives. À cette fin, nous avons considéré des enquêtes sur les industries de services et sur l'enquête annuelle sur le commerce de gros. L'étude considère tant les données provenant des corporations (T2) que celles provenant des entreprises individuelles (T1).

L'étude a porté principalement sur les variables clés recueillies par l'EUE et par les fichiers administratifs. Ces variables sont, par exemple, le revenu total d'exploitation, les dépenses totales, le coûts des marchandise vendues, etc. Nous avons vu à la section précédente que les définitions semblent les mêmes pour les variables de revenus et de dépenses mais que

du côté des dépenses, pour plusieurs composantes de cette catégorie, certaines sont rapportées différemment. Ainsi, l'étude a permis de vérifier cette hypothèse avec des données observées.

3.3 Méthodologie

L'étude a porté sur les années de référence 2003 (AR2003) et 2004 (AR2004). Les données d'enquête obtenues lors de la collecte avant l'étape de la vérification et de l'imputation ont été utilisées. À noter que seules les entreprises ayant une structure simple, i.e. les entreprises opérant dans une seule province et dans une seule industrie, ont été utilisées pour l'étude. Les unités inactives et hors du champ de l'enquête ont été enlevées. Pour les fichiers contenant les unités T1, comme il n'était pas possible de distinguer les unités imputées des unités non imputées, toutes les unités ont été utilisées pour l'étude. Cependant, pour les fichiers contenant les unités T2, seulement les unités non imputées ont été utilisées. Pour effectuer les comparaisons des données, un autre critère a été ajouté; un enregistrement devait avoir une valeur positive pour la variable du revenu total et celle du total des dépenses, autant du côté des données d'enquête que du côté des données fiscales, et avoir une valeur non manquante pour les autres variables clés. Ce critère a été mis en place dans le but d'uniformiser les comparaisons et obtenir le même nombre d'unités à analyser pour chacune des variables.

Les comparaisons ont été réalisées à deux niveaux : au niveau macro (i.e. au niveau global) ainsi qu'au niveau micro (i.e. au niveau des enregistrements). Au niveau macro, pour une variable donnée, une différence relative a été calculée en sommant les enregistrements provenant de l'enquête et des données fiscales. Le calcul a été fait de la façon suivante :

$$\% \text{ différence relative} = \frac{(\sum_i \text{donnée d'enquête}_i - \sum_i \text{donnée fiscale}_i)}{\sum_i \text{donnée d'enquête}_i} \times 100$$

Lors de la production des premières comparaisons au niveau macro, certains résultats étaient grandement influencés par une ou deux unités seulement. Nous observons pour ces unités des différences démesurées entre la donnée d'enquête et la donnée fiscale. Ainsi, dans le but d'obtenir des résultats plus précis, quelques enregistrements ont été enlevés des comparaisons au niveau macro. Ces unités, qu'on appelle unités aberrantes, ont été identifiées graphiquement et manuellement.

Pour ce qui est de l'analyse au niveau micro, un calcul similaire a été réalisé mais cette fois-ci pour chacun des enregistrements. Le calcul a été fait de la façon suivante :

$$\% \text{ différence relative}_i = \frac{(\text{donnée d'enquête}_i - \text{donnée fiscale}_i)}{\text{donnée d'enquête}_i} \times 100$$

Par la suite, les différences relatives obtenues ont été classifiées par catégorie : (0%, 1%), [1%, 5%), etc. Les résultats des analyses macro et micro seront présentés à la section suivante.

4. RÉSULTATS ET ANALYSE

4.1 Analyse macro

Le tableau 4.1 présente les différences relatives entre les principales variables clés pour deux des enquêtes étudiées soit l'Enquête sur les industries de services : comptabilité et tenue de livres, et l'Enquête annuelle sur le commerce de gros pour les unités T2. À noter que des résultats similaires ont été obtenus pour l'AR2003.

Nous remarquons que pour les variables de revenus, les différences relatives sont très près de 0 donc, il y a peu de différence entre les données d'enquête et les données fiscales. Cependant, pour certaines variables, comme par exemple pour l'Enquête annuelle sur le commerce de gros pour les variables Salaire et traitements des employés et Total – Dépréciation et amortissement, les différences relatives sont assez élevés. Ceci semble confirmer les hypothèses soulevées à la section 2 au sujet des variables de dépenses.

Tableau 4.1 Différence relative en pourcentage entre les données d'enquête et les données fiscales pour les unités T2 pour l'AR2004

Variables	Enquête sur les industries de services : comptabilité et tenue de livres ⁽¹⁾	Enquête annuelle sur le commerce de gros ⁽²⁾
Revenu total d'exploitation	2,4%	-0,4%
Revenu total	3,1%	1,4%
Salaires et traitements des employés	-3,7%	19,0%
Total – Rémunération du travail	5,4%	26,4%
Total – Dépréciation et amortissement	-0,8%	5,9%
Total des dépenses	-1,2%	4,2%
Coûts des marchandises vendues	n.d.	-0,8%

(1) 101 unités ont été utilisés pour l'analyse pour l'Enquête sur les industries de services : comptabilité et tenue de livre

(2) 519 ont été utilisés pour l'analyse pour l'Enquête annuelle sur le commerce de gros

Une analyse similaire a été produite pour les unités T1 pour l'AR2003 et l'AR2004. Pour les unités T1, les différences relatives au niveau macro sont beaucoup plus importantes pour pratiquement toutes les variables. Une première explication est que certaines unités à structure simple identifiées par l'EUE ne sont pas en réalité des unités simples mais plutôt des unités complexes (voir Girard (1999)). Les unités complexes peuvent être associées à plusieurs rapports fiscaux aussi, il est possible que les données fiscales allouées ne correspondent pas directement à la portion de l'entreprise échantillonnée. Une autre explication possible est que les unités T1 en partenariat sont traitées comme des unités à structure simple dans l'EUE, ce qui ne devrait pas être le cas. Finalement, des erreurs de mesures ont été observées autant du côté des données d'enquête que du côté des données fiscales, et ce autant pour les unités T1 que T2.

4.2 Analyse micro

Le tableau 4.2 présente pour l'Enquête sur les industries de services : comptabilité et tenue de livre les différences relatives classifiées par catégorie pour les unités pour l'AR2004. À noter que des résultats similaires ont été obtenus pour l'AR2003.

Tableau 4.2 Classification des différences relatives par catégorie pour la variable Revenu total d'exploitation pour l'Enquête sur les industries de services : comptabilité et tenue de livre pour les unités T2 pour l'AR2004

Différence relative	# d'unités	# d'unités aberrantes	% du total
Moins de -100%]	1	1	1,9%
(-100% à -50%]	0	0	0,0%
(-50% à -20%]	5	0	4,9%
(-20% à -10%]	2	0	1,9%
(-10% à -5%]	5	0	4,9%
(-5% à -1%]	19	0	18,6%
(-1% à 0%]	24	0	23,5%
Pas de différence	23	0	22,5%
(0% à 1%]	10	0	9,8%
[1% à 5%]	5	0	4,8%
[5% à 10%]	1	0	0,9%
[10% à 20%]	1	0	0,9%
[20% à 50%]	3	0	2,9%
[50% à 100%]	2	0	1,9%
Total	101	1	100,0%

Nous remarquons que pour cette variable pour cette enquête, la majorité des différences relatives se situent entre -10% et 10%, ce qui entraîne, comme on l'a vu, de faibles différences relatives au macro (tableau 4.1).

Une analyse similaire a été produite pour les unités T1 pour l'AR2003 et l'AR2004. Pour les unités T1, la majorité des différences relatives ne sont pas regroupées autour de 0 mais sont beaucoup plus dispersées de manière presque uniforme entre -100% et 100%. Ceci reflète les résultats obtenus avec l'analyse au niveau macro pour les unités T1 et discutés à la section 4.1.

5. CONCLUSION, RECOMMANDATIONS ET TRAVAUX FUTURS

En conclusion, ce projet a permis de démontrer qu'il ne semble pas y avoir un problème de concepts et de définitions des variables clés. Les différences obtenues à la section 4 proviennent principalement des erreurs de mesure (autant du côté de l'enquête que du côté des données fiscales) et d'erreurs sur la base de sondage.

Des différences plus importantes ont été observées pour les unités T1 que pour les unités T2. Pour les unités T1, il semble y avoir divergence dans la façon d'identifier les unités simples. Plusieurs unités sont identifiées comme unités simples du côté de l'EUE mais selon les dossiers administratifs, ces unités sont considérées comme des unités complexes. Ainsi, les unités T1 en partenariat sont identifiées comme des unités simples du côté de l'enquête mais elles sont en réalité des unités complexes. Finalement, certains liens entre les données d'enquête et les données fiscales se sont avérés être incorrects pour les unités T1.

En conséquent, le remplacement direct pour les unités T2 donne de très bons résultats pour les variables Revenu total d'exploitation, Dépenses totales et pour plusieurs des variables clés. Cependant, pour certaines enquêtes et composantes des dépenses, les résultats ne sont pas aussi satisfaisants mais ils sont acceptables. Pour les unités T1, l'identification des unités simples doit être modifiée pour pouvoir effectuer du remplacement direct. Il faut cependant noter que la contribution des unités T1 est petite pour l'estimation totale pour la plupart des enquêtes de l'EUE. Par conséquent, nous recommandons que pour l'identification des unités simples, autant pour les unités T1 que pour les unités T2, que les variables décrivant l'entité légale et le type d'entreprise provenant du Registre des entreprises soient utilisées.

Pour les travaux futurs, certaines analyses avec les données de l'AR2004 seront reproduites, cette fois-ci avec le plan comptable (« Charts of Accounts »). Également, des analyses supplémentaires seront effectuées sur les unités T1 dans le but d'expliquer clairement les différences observées entre les deux sources de données.

REMERCIEMENTS

Les auteurs tiennent à remercier Yves Morin, Patrice Martineau, Hélène Bérard et Danielle Lebrasseur pour leurs commentaires pertinents à la révision de ce document.

RÉFÉRENCES

Girard, C. (1999). BRD's and BSMD's complexity flags. Document interne de Statistique Canada.

Laniel, N., Royce, D. (1998). Projet d'amélioration des statistiques économiques provinciales : objectifs et enquête-pilote. Société Statistique du Canada, Recueil 1998 de la Section des méthodes d'enquête, 59-63.

Pelletier, E., Muggaberg, V., Gauthier, S., Balsdon, M. (2003). Overview of the Estimation Studies for the UES 2003 Tax Replacement Project (TRP). Document interne de Statistique Canada.

Pelletier, E. (2004). L'utilisation accrue des données fiscales dans le cadre de l'Enquête unifiée sur les entreprises. Société Statistique du Canada, Recueil 2004 de la Section des méthodes d'enquête.

Tebrake, J. (2003a). Tax Simulation Results Management Consultants. Document interne de Statistique Canada.

Tebrake, J. (2003b). Tax Simulation Results Wholesale Trade. Document interne de Statistique Canada.