

Estimation de la répartition annuelle du patrimoine dans le Système de comptabilité nationale du Canada

Margaret Wu et Cilanne Boulet¹

Résumé

Au cours des dernières années, il y a eu un intérêt grandissant envers les mesures de répartition du bien-être économique. Pour répondre à ce besoin, Statistique Canada est en train d'élaborer une série de tableaux annuels qui intègrent des données macroéconomiques sur les comptes nationaux avec des microdonnées d'enquête sur le patrimoine, le revenu et la consommation. Ce produit, Répartition des comptes économiques des ménages, ajoute une composante de répartition aux comptes macroéconomiques du Canada, donnant ainsi un portrait plus complet du bien-être économique des ménages canadiens dans le cadre des comptes nationaux.

Le présent article décrit la méthodologie utilisée pour établir les tableaux de patrimoine. Par le passé, Statistique Canada ne menait des enquêtes sur le patrimoine qu'occasionnellement. L'un des principaux défis de ce projet est de trouver un moyen de combler l'absence relativement longue de données entre les années d'enquête. La modélisation, l'étalonnage, l'analyse comparative et le ratissage sont combinés afin de relever les défis pour combler ces absences de données et assurer la cohérence avec les totaux des comptes macroéconomiques et les données d'enquête.

Mots-clés : Comptes nationaux; intégration des données; calage; modèle au niveau du domaine; données administratives; méthode itérative.

1. Introduction

Au cours des dernières années, l'attention portée aux inégalités financières et à la stabilité financière s'est accrue, et, par conséquent, il y a eu un intérêt grandissant envers les mesures de répartition du bien-être économique. Pour les ménages canadiens en tant que groupe, les données sur le patrimoine des ménages sont accessibles depuis de nombreuses années par l'intermédiaire des comptes du bilan national (CBN) dans les comptes nationaux. Cependant, il leur manque la composante de répartition qui est importante pour comprendre les disparités et les vulnérabilités dans la population canadienne. Les Comptes économiques répartis pour le secteur des ménages (CERSM) sont un produit relativement nouveau dans le Système de comptabilité nationale (SCN) et visent à combler cette lacune statistique. Les CERSM intègrent les CBN aux renseignements de répartition provenant de l'Enquête sur la sécurité financière (ESF), laquelle est une enquête menée auprès des ménages au sujet du patrimoine. L'objectif est de combiner les forces des deux sources : le niveau de détail de l'enquête, ainsi que la couverture, les concepts et la comparabilité internationale des CBN (Statistique Canada, 2018).

La composante du patrimoine des CERSM consiste en une série annuelle de tableaux sur la richesse distribuée en fonction des caractéristiques des ménages. Une version préliminaire a d'abord été diffusée en décembre 2016, et depuis, des diffusions régulières ont suivi. En ce moment, les tableaux couvrent la période s'échelonnant de 2010 à 2017 et sont fondés sur les CBN et les deux versions de l'ESF effectuées pendant cette période. Les tableaux sur le patrimoine sont accessibles selon les ventilations des ménages suivantes : quintile de revenu disponible rendu équivalent, groupe d'âge du principal soutien économique, type de ménage et province. Le tableau 1-1 ci-dessous est un exemple de tableau couvrant une seule année en fonction du groupe d'âge. La colonne des totaux provient des CBN, alors que les autres colonnes sont fondées sur la source de microdonnées.

¹Margaret Wu, Statistique Canada, 100 promenade Tunney's Pasture, Ottawa, Canada K1A 0T6, (margaret.wu@canada.ca); Cilanne Boulet, 100 promenade Tunney's Pasture, Ottawa, Canada K1A 0T6, (cilanne.boulet@canada.ca)

Le présent article présentera la méthodologie pour l'intégration des sources de données dans le but de bâtir les tableaux sur le patrimoine. La méthodologie mise en œuvre suit l'approche graduelle recommandée par l'Expert Group on Disparities within the National Accounts (groupe d'experts sur les disparités à l'intérieur des comptes nationaux) de l'Organisation de coopération et de développement économiques (OCDE) (OCDE, 2013). Dans la section 2, on décrit les sources de microdonnées et de macrodonnées, ainsi que les deux principaux défis que présentent leurs combinaisons; dans les sections 3 et 4, on décrit les méthodologies passées et actuelles; et dans la section 5, on compare les méthodologies en question.

Tableau 1-1
Échantillon de tableau sur le patrimoine pour une seule année

	Groupe d'âge du principal soutien économique					Total des CBN
	<35	35-44	45-54	55-64	65+	
Total des actifs						
Total des actifs financiers						
Assurance-vie et rentes						
Autres actifs financiers						
Total des actifs non financiers						
Immobilier						
Autres actifs non financiers						
Passif total						
Passif d'hypothèque						
Autres éléments de passifs						
Patrimoine/valeur nette						

2. Sources des données

2.1 Comptes du bilan national (CBN) : source de macrodonnées

Les CBN s'inscrivent dans le Système de comptabilité nationale du Canada, qui est le produit d'un cadre comptable dans lequel les données provenant de sources diverses sont combinées et comparées. Les CBN sont un bilan des actifs, des passifs et de la valeur nette à un moment donné, pour chaque secteur de l'économie canadienne (secteur des ménages, secteur des sociétés, secteur gouvernemental, etc.). Ce bilan est produit chaque trimestre. Dans la mesure du possible, les CBN suivent les lignes directrices internationales en ce qui a trait aux concepts et à la couverture du SCN pour des raisons de comparabilité internationale. L'objectif des CERSM est de distribuer les totaux relatifs à la richesse du secteur des ménages, lesquels couvrent l'ensemble des ménages canadiens, parmi les diverses classifications des ménages présentant un intérêt, et ce, tout en préservant l'harmonisation avec le cadre du SCN.

Les CBN de la fin du quatrième trimestre de chaque année de 2010 à 2017 sont utilisés pour créer les tableaux sur le patrimoine.

2.2 Enquête sur la sécurité financière (ESF) : source de microdonnées

L'ESF sert de source de microdonnées qui fournit la composante de répartition. L'ESF est une enquête transversale par échantillon effectuée par Statistique Canada, qui brosse un portrait détaillé de la valeur nette du patrimoine des Canadiens. Elle combine les données recueillies au moyen d'interviews sur place au sujet de la valeur de l'ensemble des principaux avoirs financiers et non financiers et sur les sommes dues à l'égard des prêts hypothécaires, des véhicules, des cartes de crédit, des prêts étudiants et d'autres dettes, avec d'autres données provenant de fichiers administratifs et d'autres sources sur le revenu et les pensions au moyen du couplage d'enregistrements.

Dans le passé, l'ESF était effectuée selon une fréquence occasionnelle, en 1999, en 2005 et en 2012. À compter de 2016, elle sera réalisée tous les trois ans. L'ESF couvre la population habitant dans les 10 provinces du Canada et

exclut certains groupes, lesquels représentent environ 2 % de la population (par exemple, les personnes vivant dans des réserves ou d'autres établissements autochtones, ainsi que les malades chroniques vivant dans les hôpitaux ou les établissements de soins infirmiers).

Les ESF de 2012 et de 2016 ont été utilisées pour créer les tableaux sur le patrimoine de 2010 à 2017, alors que les ESF de 1999 et 2005 ont été utilisées seulement pour valider les méthodes.

2.3 Défis relatifs à l'intégration des sources

L'un des principaux défis de l'intégration des sources est l'harmonisation des microdonnées d'enquête et des macrodonnées des CBN de manière qu'elles soient suffisamment cohérentes pour être combinées. L'ESF et les enquêtes auprès des ménages ont différentes raisons d'être et ne comportent pas nécessairement la même couverture ou les mêmes concepts. Par exemple, l'ESF exclut les territoires et les populations institutionnalisées, lesquels sont toutefois inclus dans les CBN. Un exemple de différence dans les définitions est celui de la dette découlant d'une carte de crédit. Dans l'ESF, le concept mesuré est le solde reporté d'une période à la suivante, alors que dans les CBN, le concept est celui du solde en souffrance à un moment donné. L'intégration des deux sources nécessite l'examen attentif des différences et des corrections pour les faire cadrer dans la mesure du possible. Pour faciliter cette pratique, les catégories de patrimoine ont été agrégées à partir des CBN à un niveau où il y avait davantage de cohérence avec les données de l'ESF en matière de concept et de totaux à l'échelle du Canada.

Le deuxième principal défi a trait aux données manquantes, surtout dans l'estimation des distributions du patrimoine pendant les années où l'ESF n'a pas eu lieu. On aborde cette situation en modélisant le patrimoine au moyen de l'information auxiliaire, comme il est décrit dans la section suivante. De plus, la publication était limitée à compter de l'année 2010, ce qui a donné lieu à des écarts plus courts entre les années où il y a eu des enquêtes.

3. Méthodologie pour la modélisation du patrimoine

Deux approches ont été utilisées pour modéliser la valeur nette du patrimoine et ses composantes : l'une est fondée sur le calage et l'autre sur les modèles au niveau du domaine, et les deux approches recourent à différents types de données auxiliaires. L'approche par calage sera utilisée pour la prochaine diffusion en mars 2019. Les diffusions antérieures de 2016 à 2018 étaient fondées sur l'approche au niveau du domaine ou utilisaient une approche mixte. D'autres précisions sur cette méthodologie sont accessibles à Statistique Canada (2018).

3.1 Méthodologie actuelle : approche par calage

La méthodologie actuelle pour combiner les sources et estimer les distributions de patrimoine pour les années avec et sans enquête est fondée sur le calage (Deville et Sarndal 1992), et l'idée est similaire au vieillissement statique. Le calage est une étape du processus habituel de pondération dans les enquêtes et consiste à corriger les poids des unités échantillonnées afin que les estimations provenant de l'enquête coïncident avec les totaux connus à l'échelle de la population.

La première étape du processus est un recalage de l'ESF de 2012 et de 2016 aux chiffres de population pour les années avec et sans enquête. Les totaux de calage sont des estimations de la population échantillonnée en fonction des projections des chiffres de population de Statistique Canada. Les totaux utilisés comprennent les comptes d'individus par catégorie de sexe et de groupe d'âge, les comptes de ménages selon la taille du ménage et les comptes de familles économiques selon la taille de la famille économique pour des tailles de famille économique sélectionnées à l'intérieur des provinces. Ces chiffres sont très similaires à ceux utilisés dans la production de l'ESF, même si l'on procède à certaines corrections pour mieux s'harmoniser avec les catégories de ménages diffusées dans le cadre des CERSM. Le recalage permet de rajuster les poids des unités échantillonnées afin que les estimations de l'enquête coïncident avec ces totaux de population pour les années sans enquête, essentiellement en rajustant les poids de sondage pour refléter les changements démographiques.

Autrement dit, les ensembles suivants de poids, $w_{i,y}$, sont déterminés pour l'ensemble des ménages i dans l'ESF de

2012, pour chaque année $y = 2010, \dots, 2017$ où \mathbf{x}_i représente un vecteur de caractéristiques démographiques pour un ménage i et \mathbf{X}_y représente les chiffres de population pour l'année y qui servent de totaux de contrôle du calage. Le même processus est répété pour tous les ménages de l'ESF de 2016 au moyen des mêmes chiffres de population \mathbf{X}_y .

$$w_{i,2010} \text{ de sorte que } \sum w_{i,2010} \mathbf{x}_i = \mathbf{X}_{2010}$$

$$w_{i,2011} \text{ de sorte que } \sum w_{i,2011} \mathbf{x}_i = \mathbf{X}_{2011}$$

...

$$w_{i,2017} \text{ de sorte que } \sum w_{i,2017} \mathbf{x}_i = \mathbf{X}_{2017}$$

Dès que les poids des échantillons de l'ESF ont fait l'objet d'un recalage pour chaque année, les diverses composantes du patrimoine au niveau des microdonnées sont mises en correspondance avec les totaux des CBN de chaque année. Dans ce contexte, il est possible d'obtenir les estimations de la valeur nette du patrimoine et de ses composantes pour les années sans enquête au moyen des données d'enquêtes mises en correspondance et les poids corrigés.

Deux séries d'estimations sont obtenues pour la période allant de 2010 à 2017 : l'une utilise l'ESF de 2012 et l'autre, l'ESF de 2016. La série fondée sur l'ESF de 2012 est utilisée pour produire les estimations pour la période allant de 2010 à 2012. Pour les années 2013 à 2015, les séries sont combinées par interpolation linéaire entre les séries. Par exemple, pour 2013, l'estimation combinée est calculée comme suit : $\frac{3}{4} \times$ l'estimation à partir de l'ESF de 2012 recalée à 2013 + $\frac{1}{4} \times$ l'estimation à partir de l'ESF de 2016 recalée à 2013. La série fondée sur l'ESF de 2016 est utilisée pour produire les estimations de 2016 à 2017.

Pour limiter les points de retournement créés lors de la combinaison de deux séries d'estimation, une moyenne mobile pondérée centrée de trois points est appliquée à la série combinée de 2012 et de 2016. Cela modifie légèrement les estimations dérivées pour les années d'enquête. Comme dernière étape, la méthode itérative est appliquée dans chaque année afin de rétablir ces relations parmi les colonnes et les rangées qui ont été perturbées à l'étape de lissage, et ce, tout en réduisant au minimum les modifications apportées aux cellules individuelles du tableau (Dagum et Cholette, 2006).

Il convient de noter que l'approche axée sur le calage au moyen d'une seule année de l'ESF ne saisit que les tendances relatives au patrimoine lié aux variables démographiques utilisées, et repose sur l'hypothèse que ces relations sont stables au fil du temps. Or, en utilisant différentes années avec enquête, en les combinant en plus d'intégrer les totaux des CBN, cette méthode devrait pouvoir permettre de saisir d'autres tendances que celles qui sont liées aux données démographiques.

3.2 Méthodologie passée : approche fondée sur le modèle au niveau du domaine

Lors des premières diffusions des CERSM, une seule version de l'ESF était accessible et l'approche fondée sur le modèle au niveau du domaine était plutôt utilisée pour obtenir les distributions du patrimoine dans les années sans enquête. Cette approche est motivée par les modèles au niveau du domaine qui sont une technique classique d'estimation sur petits domaines (Rao et Molina, 2015). Elle met à contribution le fait qu'il existe une relation entre le revenu et le patrimoine et emploie les données de l'impôt sur le revenu accessibles de l'Agence du revenu du Canada.

En particulier, les modèles de Fay-Herriot ont été utilisés pour estimer le patrimoine au moyen de covariables provenant des déclarations d'impôt sur le revenu, surtout le fichier sur la famille T1 (FFT1). Voici, en guise d'exemple, un tel modèle pour l'estimation de la valeur nette du patrimoine chez le plus jeune groupe d'âge :

$$\hat{\theta}_i^{DIR} = \mathbf{z}_i^T \boldsymbol{\beta} + b_i v_i + e_i \text{ pour les domaines } i = 1, \dots, 49 \text{ où}$$

- $\hat{\theta}_i^{DIR}$ est l'estimation directe du logarithme du total de la valeur nette du patrimoine pour le domaine du plus jeune groupe d'âge de l'ESF de 2012
- \mathbf{z}_i est le vecteur des covariables du revenu provenant des données fiscales
- b_i est une constante égale à 1
- $v_i \sim iid(0, \sigma_v^2)$ sont des effets aléatoires propres aux domaines
- $e_i \sim iid(0, \psi_i)$ sont des erreurs d'échantillonnage, et ψ_i est estimé à partir de l'ESF de 2012

- les domaines i sont définis en croisant les régions économiques combinées dans les 10 provinces avec la catégorie d'intérêt en matière de ménages (comme le groupe d'âge)

Les données de l'ESF de 2012 fournissant les estimations directes ont été les premières mises à l'échelle des microdonnées pour correspondre aux CBN de 2012. Les modèles ont alors été ajustés aux données mises à l'échelle de l'ESF de 2012, puis appliquées toutes les années où les données fiscales auxiliaires étaient accessibles. Cela a donné lieu aux estimations synthétiques $\mathbf{z}_i^T \hat{\boldsymbol{\beta}}$ du logarithme de la valeur nette du patrimoine pour chaque domaine dans les années sans enquête en plus de l'enquête de 2012. Afin d'obtenir les distributions requises pour les tableaux sur le patrimoine, qui se trouvent à des niveaux géographiques plus élevés, les estimations de la valeur nette du patrimoine ont été additionnées dans les domaines pertinents, au sens de $\sum e^{\mathbf{z}_i^T \hat{\boldsymbol{\beta}}}$. Par exemple, le total estimé de la valeur nette du patrimoine pour le plus jeune groupe d'âge était la somme de l'exponentielle des estimations synthétiques du logarithme de la valeur nette du patrimoine, $e^{\mathbf{z}_i^T \hat{\boldsymbol{\beta}}}$, sur toutes les régions géographiques dans le plus jeune groupe d'âge.

Afin d'assurer la cohérence avec les ESF, les distributions estimées ont été étalonnées d'après les distributions de l'ESF de 2012. Cette étape a imposé les distributions de l'ESF en 2012 tout en préservant dans la mesure du possible le mouvement d'une année à l'autre dans les séries estimées (Dagum et Cholette, 2006). L'étalonnage était également une manière de composer avec le biais résultant de l'utilisation d'un modèle logarithmique. En guise d'étape finale, on a appliqué la méthode itérative pour chaque année afin de rétablir ces relations parmi les colonnes et les rangées puisque chaque composante du patrimoine était modélisée de façon individuelle.

4. Comparaison des approches

Lorsque l'on compare l'approche fondée sur le modèle au niveau du domaine à l'approche par calage, on constate un certain nombre d'avantages et de désavantages. Le principal avantage de l'approche fondée sur le modèle au niveau du domaine est qu'elle permet théoriquement de cerner les tendances en matière de richesse qui sont liées au revenu. En fait, toutes autres données auxiliaires pourraient être utilisées pour la modélisation, pourvu qu'elles soient accessibles au niveau du modèle. En revanche, pour intégrer de nouvelles variables auxiliaires à l'approche par calage, il faudrait qu'elles soient accessibles au niveau des ménages, ce qui est une exigence plus restrictive.

Or, il y a certains désavantages pratiques qui sont uniques à l'approche fondée sur le modèle au niveau du domaine, soit la rapidité et l'utilisation des ressources. L'approche fondée sur le modèle au niveau du domaine nécessite beaucoup plus de temps et de ressources, car chaque composante doit être modélisée séparément pour chaque catégorie de ménage. Cela rend ce processus plus complexe pour l'obtention des tableaux sur le patrimoine en fonction des catégories supplémentaires de ménages, alors que l'approche par calage permet d'obtenir de nouvelles distributions facilement, dès que les poids réétalonnés ont été produits. En outre, la modélisation au niveau du domaine nécessite des données auxiliaires opportunes; dans le cas des données de l'impôt sur le revenu, il y a un retard d'une année ou deux entre le moment de l'accessibilité et le calendrier de diffusion des CERSM; l'utilisation de cette approche signifie que la dernière ou les dernières années des CERSM doivent être extrapolées autrement.

Un autre désavantage de l'approche fondée sur le modèle au niveau du domaine est qu'il n'est pas certain qu'un modèle approprié puisse être trouvé au moyen des covariables accessibles. Cette situation s'est produite dans le cas de certaines des plus petites catégories d'avoirs et de dettes, et pour ces cas, l'approche par calage a été utilisée.

En ce qui concerne les estimations du patrimoine, les deux approches ont donné lieu à des estimations similaires pour la période de 2010 à 2017 dès que les ESF de 2012 et de 2016 ont été accessibles. Cela était vrai pour la plupart des catégories de patrimoine, et les plus grandes différences ont été réduites lorsque les distributions de l'ESF ont été imposées par l'étalonnage. On trouvera un exemple des différences de ces deux approches avant l'étalonnage dans le tableau 4-1 ci-dessous pour la valeur nette du patrimoine par groupe d'âge. Les différences comparées à l'ESF sont similaires du point de vue de la taille et de la direction; la somme des différences absolues par rapport aux groupes d'âge est indiquée en orangé. En fait, une fois que les étapes finales décrites dans la prochaine section ont été accomplies, la différence entre les parts détenues par chaque groupe est de moins de 1 % dans presque toutes les cellules du tableau.

Un dernier point d'intérêt est qu'il n'était pas clair si la modélisation au niveau du domaine permettait davantage de saisir les effets des grands événements, spécifiquement la récession de 2008. Étant donné que les modèles au niveau du domaine reposent sur des données sur le revenu, on pourrait s'attendre, par exemple, à ce que les modèles au niveau du domaine montrent que la récession n'a pas eu la même incidence sur la richesse des ménages appartenant au quintile de revenu le plus bas que sur la richesse des ménages du quintile de revenu le plus élevé. Malheureusement, les estimations comportaient trop de bruit pour que l'on puisse distinguer quelques différences que ce soit.

À compter de la prochaine diffusion en mars 2019, les CERSM seront créés à l'aide de l'approche par calage seulement. Les désavantages pratiques de l'approche fondée sur le modèle au niveau du domaine comptent parmi les raisons motivant ce changement, surtout à la lumière du calendrier serré de production des CERSM. Une autre raison provient de la connaissance du fait que le patrimoine ne change pas rapidement habituellement. On l'accumule tout au long d'une vie et, maintenant que l'ESF est effectuée aux trois ans, les écarts entre les années d'enquête ne sont plus aussi longs qu'avant.

Néanmoins, il convient de noter que la modélisation au niveau du domaine demeure plus utile si des tableaux sont requis pour la période antérieure de 1999 à 2009, où les écarts entre les années d'enquête sont plus longs, ou, dans le cas où l'on désire d'autres distributions par ménage qui sont moins liées aux variables de calage, dans l'hypothèse où les covariables de modélisation appropriées au niveau du domaine sont disponibles.

Tableau 4-1

Comparaison des estimations provenant de l'ESF, de l'approche fondée sur le modèle au niveau du domaine et de l'approche par calage, avec la somme des différences absolues (orangé)

		Distribution			Différence	
		ESF	Modèle au niveau du domaine	Calage	Modèle au niveau du domaine	Calage
2016	<35	5,6 %	5,5 %	5,9 %	-0,1 %	0,4 %
	35-44	12,1 %	11,6 %	12,2 %	-0,5 %	0,1 %
	45-54	21,6 %	22,3 %	23,5 %	0,7 %	1,9 %
	55-64	30,9 %	29,0 %	27,9 %	-1,9 %	-3,0 %
	65+	29,8 %	31,6 %	30,4 %	1,7 %	0,6 %
	Total	100,0 %	100,0 %	100,0 %	5,0 %	6,0 %
2005	<35	4,4 %	6,1 %	6,4 %	1,8 %	2,0 %
	35-44	17,3 %	14,5 %	15,4 %	-2,8 %	-1,9 %
	45-54	26,6 %	27,8 %	28,3 %	1,2 %	1,7 %
	55-64	26,8 %	25,7 %	24,7 %	-1,1 %	-2,1 %
	65+	25,0 %	25,9 %	25,2 %	0,9 %	0,3 %
	Total	100,0 %	100,0 %	100,0 %	7,7 %	8,1 %
1999	<35	7,0 %	7,0 %	7,4 %	0,0 %	0,3 %
	35-44	18,8 %	17,8 %	18,8 %	-1,0 %	-0,1 %
	45-54	25,8 %	28,1 %	27,4 %	2,3 %	1,6 %
	55-64	22,7 %	23,0 %	21,5 %	0,3 %	-1,2 %
	65+	25,6 %	24,1 %	25,0 %	-1,5 %	-0,6 %
	Total	100,0 %	100,0 %	100,0 %	5,2 %	3,8 %

Bibliographie

Boulet, C., et M. Wu (2017), « Annual Household Distribution Tables: Wealth Tables Methodology », rapport non publié, Ottawa, Canada: Statistics Canada.

Dagum, E. B., et P. A. Cholette (2006), *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series Data*, New York: Springer-Verlag.

Deville, J.-C., et C.-E. Sarndal (1992), « Calibration Estimators in Survey Sampling », *Journal of the American Statistical Association*, 87(418), pp. 376-382.

OECD (2013), *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth*, Paris: OECD Publishing.

Rao, J. N. K., et A. Molina (2015), *Small Area Estimation, 2nd Edition*, Hoboken: Wiley.

Statistique Canada (2018), *Comptes économiques répartis pour le secteur des ménages, estimations des répartitions des actifs, des passifs et de la valeur nette, 2010 à 2017 : rapport technique sur la méthodologie et la qualité*.