



**Retraits des REER avant l'âge de la retraite : comment les néo-données administratives nous  
aident-elles à comprendre le phénomène?**

**par**

**Simon Dagenais**

**Professeur Daniel Parent et Professeur Pierre-Carl Michaud**

**HEC Montréal**

**Directeurs de recherche**

**Sciences de la gestion**

**(Spécialisation Économie appliquée)**

*Mémoire présenté en vue de l'obtention du grade de maîtrise ès sciences en gestion*

*(M. Sc.)*

Septembre 2020

©Simon DAGENAI, 2020

*« Old men sit in the shade because they planted a tree many years before. »*

Proverbe ougandais

# *Remerciements*

Je tiens d'abord à remercier mes directeurs de recherche Professeur Daniel Parent et Professeur Pierre-Carl Michaud. Ils ont témoigné d'une patience exemplaire à mon égard. Leurs encouragements m'ont certainement permis de traverser ce périple qui fut mon chemin de croix. Je sors néanmoins grandi de cette épreuve grâce à vous. Merci mille fois.

Je voudrais également remercier Professeur Daniel Parent, le Centre interuniversitaire du Québec en statistiques sociales et spécialement la Chaire de recherche Industrielle Alliance pour le financement du projet. Cette aide m'a été indispensable à l'époque.

Je remercie mes collègues de travail chez MoovAI qui m'ont permis de m'absenter du bureau pour me rendre au laboratoire. Vous avez fait preuve d'une empathie à la hauteur des valeurs humaines que ne nous sommes proposées.

Je voudrais remercier mes parents et mes amis qui n'ont cessé de me poser la question exponentiellement embarrassante dans le temps : « Et puis ton mémoire ? ». Le stress et les malaises provoqués à l'écoute de cette interlocution m'ont fait comprendre qu'il n'était pas question que j'abandonne la rédaction de ce texte. Je vous remercie de vous soucier de l'avancement de mes projets, fous et moins fous.



## *Résumé exécutif*

Le régime enregistré d'épargne retraite (REER) est un outil financier hautement prisé par la population canadienne. Au cours des années 2000, c'est plus de 25% des Canadiens qui contribuent au REER chaque année. Que ce soit pour les avantages fiscaux qu'il procure ou par le peu de contraintes de liquidités qu'il impose à l'épargne, le REER offre maints incitatifs aux Canadiens et Canadiennes d'épargner. Or, nous tenons à approfondir la recherche sur l'utilisation que font les Canadiens de ce régime d'épargne. Plus particulièrement, nous nous intéressons aux retraits des REER des Canadiens et Canadiennes de 30 à 50 ans. Nous tentons de cerner les profils sociodémographiques des contribuables qui retirent de leur régime. Nous analysons par le fait même les chocs de revenus ou d'événements de la vie qui poussent les contribuables à retirer. Nous portons également une attention particulière à l'incidence des taux effectifs marginaux d'imposition sur l'événement et montant du retrait. Pour ce faire, nous employons des méthodes économétriques qui exploitent la forme longitudinale des données utilisées. Nous évaluons des modèles à variable dépendante dichotomique et continue. Nous développons un modèle d'apprentissage statistique qui nous permet de différencier entre deux types de retrait des REER et ajustons notre analyse économétrique en conséquence. Finalement, nous obtenons des résultats empiriques qui, pour une certaine part, ne concordent pas avec les hypothèses du modèle théorique employé.



# Table des matières

Remerciements . . . . .	3
Résumé exécutif . . . . .	5
<b>1 Introduction</b>	<b>15</b>
1.1 Introduction . . . . .	15
1.1.1 Règles entourant les REER . . . . .	17
Cotisation . . . . .	17
Retrait . . . . .	17
1.1.2 Règles entourant le Régime d'accession à la propriété . . . . .	18
1.1.3 Synthèse . . . . .	18
<b>2 Revue de littérature</b>	<b>21</b>
2.1 L'utilisation Régimes enregistrés d'épargnes-retraites . . . . .	21
2.1.1 Retraits des REER . . . . .	22
2.1.2 Aucune entre distinction retraits des REER et défauts au RAP . . . . .	24
2.1.3 Rôle du régime d'accession à la propriété . . . . .	25
2.1.4 Contributions aux REER . . . . .	26
2.2 L'utilisation des techniques d'apprentissage statistique en économétrie . . . . .	29
2.3 Synthèse . . . . .	30
<b>3 Modèles</b>	<b>33</b>
3.1 Modèle théorique et hypothèses . . . . .	33
3.1.1 Modèle de cycle de vie . . . . .	33
3.2 Modèles économétriques . . . . .	36
3.2.1 Wooldridge 1995 . . . . .	36
3.2.2 Probit à effets aléatoires . . . . .	38
3.3 Modèle d'apprentissage statistique . . . . .	39
<b>4 Apprentissage statistique : distinction entre retraits et défauts</b>	<b>43</b>
4.1 Données : La Banque de données administratives longitudinales . . . . .	44
4.2 Apprentissage statistique . . . . .	46

4.2.1	Méthodologie . . . . .	46
4.2.2	Résultats de l'entraînement . . . . .	47
<b>5</b>	<b>Analyse statistique</b>	<b>55</b>
5.1	Données : l'Étude longitudinale et internationale des adultes . . . . .	55
5.2	Statistiques descriptives . . . . .	56
5.2.1	Taux effectif d'imposition marginal . . . . .	56
5.2.2	Tendances entre les provinces . . . . .	61
5.2.3	Utilisation des REER . . . . .	63
5.2.4	Analyse univariée . . . . .	63
<b>6</b>	<b>Analyse économétrique</b>	<b>73</b>
6.1	Régression linéaire et régression logistique . . . . .	73
6.2	Modèle de probabilité linéaire à effets fixes . . . . .	74
6.3	Probit ordonné à effets aléatoires . . . . .	78
6.4	Wooldrige 1995 : modèle à variable dépendante continue . . . . .	79
6.5	Synthèse . . . . .	83
6.5.1	Modèles à variable dépendante dichotomique . . . . .	83
6.5.2	Modèles à variable dépendante continue . . . . .	84
<b>7</b>	<b>Conclusion</b>	<b>85</b>
<b>A</b>	<b>Taux effectif d'imposition marginal</b>	<b>93</b>
A.1	Calcul du TEMI . . . . .	93
A.2	Régressions linéaires et logistiques . . . . .	95
A.3	Résultats des modèles multivariés avec forme fonctionnelle sans revenu . . . . .	100
A.4	Résultats des régressions probits à effets aléatoires . . . . .	104

# Table des figures

4.1	Distribution des retraits et défauts confondus dans la base de donnée du DAL	45
4.2	Distributions des retraits des REER et défauts du RAP dans la base de donnée du DAL . . . . .	45
4.3	Résultats de l'entraînement du modèle . . . . .	48
4.4	Performance du modèle d'apprentissage statistique . . . . .	50
4.5	Interprétation du modèle d'apprentissage statistique . . . . .	51
4.6	Distribution des montants retirés dans la base de donnée du DAL et dans l'ELIA	52
4.7	Distribution des montants retirés dans la base de donnée du DAL et prédicts dans l'ELIA . . . . .	52
5.1	Distribution des TEMI . . . . .	60
5.2	Comparaison des TEMI en fonction de différentes variables sociodémographiques . . . . .	60
5.3	Retraits et défauts agrégés par provinces et années . . . . .	62
5.4	Fréquences des retraits par contribuable . . . . .	63
5.5	Tendances de revenus et retraits entre les contribuables . . . . .	65
5.6	Tendances des événements de la vie des contribuables . . . . .	68



# Liste des tableaux

4.1	Résultats de la recherche quadrillée . . . . .	49
5.1	Régression linéaire : TEMI en fonctions des revenus, provinces, années et autres variables socio-démographiques . . . . .	57
5.2	Analyse statistique : variables continues . . . . .	67
5.3	Analyse statistique : sans distinction entre retraits et défauts . . . . .	70
5.4	Analyse statistique : distinction entre retraits et défauts . . . . .	71
6.1	Modèle de probabilité linéaire à effet fixe . . . . .	75
6.2	Effets marginaux partiels évalués à la moyenne évalués aux groupes de montants : Échantillon Complet . . . . .	78
6.3	Effets marginaux évalués à la moyenne évalués aux groupes de montants : Échantillon <i>Retrait</i> . . . . .	79
6.4	Variable dépendante continue : Wooldrige 1995 . . . . .	81
7.1	Sommaire des résultats . . . . .	88
A.1	Présentation des variables utilisées par CTACS pour le calcul du TEMI . . . . .	93
A.2	Régression linéaire : Retrait en fonctions des TEMI, revenus, provinces, années et autres variables sociodémographiques . . . . .	95
A.3	Régression logistique : Événement du retrait en fonctions des TEMI, revenus, provinces, années et autres variables sociodémographiques . . . . .	98
A.4	Modèles de probabilité linéaire : spécification sans revenu . . . . .	100
A.5	Modèle à variable dépendante continue (Wooldrige 1995) : spécification sans revenu . . . . .	102
A.6	Résultats des régressions probits à effets aléatoires . . . . .	104







# Chapitre 1

## Introduction

### 1.1 Introduction

Administré par l'Agence du revenu du Canada, le Régime enregistré d'épargne retraite (REER) permet aux Canadiens d'épargner tout en bénéficiant d'avantages fiscaux attrayants. En effet, les contributions faites au sein du régime sont déductibles d'impôts et le rendement composé du capital en est complètement exonéré. L'épargne accumulée est seulement imposée lorsqu'un montant est retiré hors du régime d'épargne. Le REER permet ainsi aux Canadiens de minimiser leur charge fiscale lors de périodes de revenus élevés. Il fournit par la même occasion aux Canadiens les incitatifs nécessaires pour décaisser seulement lors de périodes de faibles revenus, telles que la retraite.

L'utilisation des REER n'est, par contre, pas restreinte à cet unique cas de figure. En fait, les règles entourant les retraits des REER contraignent seulement les Canadiens de cesser de contribuer à partir de l'âge de 71 ans. Contrairement au système d'épargne retraite américain *401 (k)*, le REER n'impose aucune pénalité sur le montant retiré et ne stipule aucune contrainte « d'événement distribuable ». Cette contrainte restreint les participants du *401 (k)* à retirer des fonds seulement qu'à la suite d'une perte d'emploi ou d'un départ à la retraite. Le REER, par sa nature permissive et flexible, agit en fait comme un régime d'épargne à impôt différé.

Véhicule d'épargne prisé auprès de la population canadienne, près de 25% des déclarants d'impôt ont utilisé les REER des années 1999 à 2013. Les montants médians de cotisations ont avoisiné 2700 \$ au cours de la même période, pour un total moyen de 25 milliards de dollars par année. C'est donc la nature particulière du régime d'épargne et son utilisation répandue qui ont poussé certains académiciens et décideurs publics à étudier et remettre en question son fonctionnement. Par exemple, Kevin Milligan (2002) s'est penché sur l'incidence des taux marginaux d'imposition dans la décision de contribuer aux REER. Mawani et Paquette (2011) ont à leur tour étudié les retraits des REER avant l'âge de la retraite. Dereck Messacar (2017)

donnait un aperçu descriptif de l'étendue des retraits des REER de la période de 2000 à 2013. Steele (2007) a étudié l'utilisation du Régime d'accèsion à la propriété (RAP). Ce régime permet aux Canadiens de déboursier de leur REER pour l'achat d'une première propriété.

Ce présent mémoire vise à faire progresser les connaissances sur la problématique des retraits des REER avant l'âge de la retraite grâce à l'exploitation de nouvelles données administratives ainsi que d'une approche méthodologique novatrice. Nous tenons pour objectif de discerner les facteurs qui poussent les individus à retirer des REER ou faire défaut au Régime d'accèsion à la propriété. Comme Mawani et Paquette, nous tenons également à comprendre l'incidence des taux effectifs marginaux d'impositions (TEMI) sur la décision de retirer des REER ou de faire défaut au RAP. Nous testons par le fait même certaines hypothèses induites d'un modèle théorique développé par Daly (1981). Nous présentons à la fin de ce chapitre certaines caractéristiques propres aux des retraits des REER et défauts du RAP.

L'aspect novateur de cette recherche repose sur l'emploi de techniques d'apprentissages statistiques pour ajouter précision et nuances à nos analyses. Nous utilisons ces techniques et algorithmes pour distinguer les retraits des REER des défauts du RAP dans la base de données de l'étude longitudinale et internationale des adultes (ELIA) puisque celle-ci ne contient pas le niveau de détails suffisants pour distinguer entre ces deux types de retraits des REER. Nous utilisons donc une seconde base de données, la Banque de données administrative longitudinale (DAL), précise au niveau des données administratives, pour construire un algorithme d'apprentissage statistique qui permet ensuite de distinguer défauts et retraits dans la base de données de l'ELIA. Cette démarche est possible grâce aux nombreuses variables qui s'entrecoupent entre l'ELIA et le DAL et puisque les deux échantillons couvrent les mêmes populations. Cette démarche permet donc de bénéficier de l'ensemble des données d'enquête de l'ELIA tout en ajoutant une estimation précise (plus de 90% de succès) sur la classification des retraits.

De plus, contrairement à ce qui a été effectué à l'heure actuelle dans la littérature à ce sujet, nous utilisons un ensemble de modèles économétriques conçus pour exploiter les données longitudinales. Finalement, grâce à l'enquête de l'ELIA et des données administratives qui lui sont couplées, nous sommes en mesure d'analyser ce phénomène à l'aide de variables qui ne se retrouvent pas dans les fichiers administratifs, telle que l'éducation.

Nos résultats empiriques concernant les retraits des REER et défauts du RAP concordent avec les résultats empiriques observés dans la littérature, mais ne correspondent toutefois pas entièrement aux résultats théoriques attendus. D'une part, tout comme observé par Mawani et Paquette et comme attendu par Daly, nous observons que les diminutions de revenus

détiennent une incidence indubitable sur les retraits et défauts. Nous observons également que certains événements de la vie tels que le divorce et la perte involontaire d'un emploi sont également associés à des hausses de retraits et défauts. Par contre, d'une autre part, tout comme Mawani et Paquette, nous ne sommes pas en mesure de confirmer que les TEMI ont une incidence négative sur retraits et défauts du RAP. Tout comme les auteurs, nous observons que les contribuables retirent malgré des TEMI élevés.

Ce mémoire se divise comme suit. Dans un premier temps, nous poursuivrons l'introduction en fournissant au lecteur un sommaire complet des règles entourant les REER afin de faciliter sa compréhension du sujet. Dans un second chapitre, nous présentons une revue de littérature détaillée sur les contributions et retraits des REER. Nous offrons également au lecteur un aperçu du présent état de l'utilisation des techniques d'apprentissages statistiques en économétrie. Au chapitre trois, nous détaillons les modèles théoriques, statistiques et économétriques qui ont été utilisés pour cette recherche. Le quatrième chapitre divulgue les résultats d'apprentissage statistique. Nous exposons dans le même chapitre la base de données du DAL qui nous a permis de former l'algorithme. Le cinquième chapitre vulgarise les données de l'ELIA grâce à une suite d'analyses statistiques. Nous y présentons par la même occasion le simulateur CTACS qui a permis le calcul des taux effectifs marginaux d'impositions. Un sixième chapitre divulgue les résultats des modèles d'analyse économétriques à variable dépendante dichotomique et continue. Nous concluons dans un septième et final chapitre.

### **1.1.1 Règles entourant les REER**

#### **Cotisation**

Le REER est un mécanisme de report d'impôt : les cotisations à un REER sont déductibles d'impôts et les revenus de placement accumulés sont exempts d'impôts jusqu'à leur encaissement. La cotisation maximale est fixée par l'Agence Revenu canadienne. Il est possible de cotiser une somme égale à 18 % du revenu admissible, jusqu'au maximum permis. Les déductions inutilisées sont reportées jusqu'à l'âge limite de 71 ans.

#### **Retrait**

Les retraits des REER doivent être déclarés par les individus et sont imposés à la source. Les seuls retraits non imposés sont faits dans les cadres du régime d'accession à la propriété (RAP) ou du régime d'encouragement à l'éducation permanente (REEP). Dans ces cas, il est

possible de retirer des fonds de ses REER sans que l'émetteur retienne d'impôt à la source. Lorsqu'un contribuable retire des fonds de ses REER dans le cadre d'un de ces programmes, il ne lui est pas requis d'inclure ce montant comme revenu dans sa déclaration de revenus et de prestations.

L'institution financière qui administre le REER retient l'impôt à la source lorsqu'un individu retire des fonds. Le taux d'imposition dépend de son lieu résidence et du montant retiré. Pour les fonds gérés dans la province de Québec, un impôt provincial est également prélevé. Il est important de mentionner que ce montant n'est pas une pénalité. L'excédent retenu est remboursé au moment du rapport d'impôt. Les taux de retenue sont affichés dans le tableau suivant :

Taux de retenu		Montant
Canada	Québec	
10%	5%	$x \leq 5\,000\$$
20%	10%	$5\,000 < x \leq 15\,000$
30%	15%	$x > 15\,000 \$$

### 1.1.2 Règles entourant le Régime d'accession à la propriété

En place depuis 1992, le régime d'accession à la propriété permettait aux premiers acheteurs d'emprunter un montant maximal de 20 000\$ par contribuable d'un ménage à partir des fonds accumulés dans les REER<sup>1</sup>. L'habitation doit servir de résidence principale et être occupée par les acheteurs. Le participant détient une période de 15 ans pour rembourser son emprunt, sous forme de cotisation annuelle à son REER. Un défaut complet ou partiel au RAP se traduit comme un montant retiré des REER. Le montant est alors assujetti aux mêmes lois fiscales qu'un retrait.

### 1.1.3 Synthèse

Une première différence entre les retraits des REER et défauts du RAP provient de l'aspect « involontaire » ou « inconscient » des défauts du RAP. Effectivement, si un individu ne cotise pas la somme due au RAP, le retrait des REER est automatiquement enregistré. La seconde différence provient des sommes retirés. Puisqu'une limite de 20, 000 dollar subsiste sur le montant décaissé dans pour l'achat d'une première maison, le montant en souffrance du RAP

1. Ce montant est passé à 35 0000\$ en mars 2019

ne peut excéder un montant de 1, 330 dollars (20, 000 \$ / 15 an). Distinguer entre retraits et défauts pour notre analyse prend donc toute son importance. Les individus qui déboursent des REER ou font défaut sont soumis à des contraintes et incitatifs distincts.



## Chapitre 2

# Revue de littérature

Cette revue de littérature vise à familiariser le lecteur sur deux courants de recherches distincts. Dans un premier temps, nous couvrons la recherche accomplie sur l'utilisation des régimes enregistrés d'épargnes-retraites. Nous amorçons cette première partie en survolant la littérature théorique en lien aux REER. Nous poursuivons en couvrant la recherche qui traite des retraits des REER avant l'âge de la retraite en distinguant la recherche en lien aux défauts du RAP. Nous discutons finalement de la recherche qui traite des cotisations aux REER. Dans un second temps, nous abordons la recherche relative à l'utilisation des techniques d'apprentissage statistique en économétrie.

### 2.1 L'utilisation Régimes enregistrés d'épargnes-retraites

Notre cadre théorique repose sur le modèle cycle de vie proposé par Daly (1981). L'auteur formule un modèle de cycle de vie avec altruisme afin d'illustrer le rôle d'épargne des REER (DALY, 1981). Dans son modèle, un agent maximise son utilité au cours de sa vie et lègue une somme à son décès. En plus de choisir sa consommation, l'agent détermine ses investissements dans les REER (contributions) et ses investissements hors REER. Les contributions dans le régime d'épargne sont déductibles du revenu imposable de l'agent. Les contributions font également face à une limite exprimée par un minimum entre un pourcentage du revenu ou un montant fixe. Les investissements hors REER sont assujettis au taux d'imposition effectif en vigueur. Daly énonce que les cotisations et retraits des REER permettent à l'agent de déterminer son profil de « taux marginal d'imposition optimal ». Cette règle lui permet de minimiser la valeur actualisée de ses paiements en impôts. Il démontre que les investissements hors REER permettent à l'agent d'atteindre son profil de consommation optimal. Il subsiste ainsi un effet prix où les agents épargnent lors des périodes où leur TEMI est plus élevé. Il démontre également que les REER permettent à l'agent de lisser sa

consommation et de soutenir des chocs de revenus négatifs. Cet effet revenu pousse l'agent à épargner lors des périodes de revenu élevé et à retirer lors des périodes de revenu faible.

En opposition à Daly (1981), Ragan argumente que les REER ont l'effet de réduire l'épargne privée des ménages (RAGAN, 1994). Grâce à un modèle à deux périodes avec un régime d'imposition uniforme, Ragan montre que les REER n'ont aucun effet de substitution. Il montre ensuite, à l'aide d'un modèle à deux périodes jumelé à un régime d'imposition progressif, que les REER ont pour effet de réduire l'épargne des ménages. Ragan supporte cette position puisqu'il émet l'hypothèse que les ménages sont assujettis à des TEMI supérieurs lors de la seconde période.

### 2.1.1 Retraits des REER

Mawani et Paquette analysent les facteurs expliquant les retraits des régimes enregistrés d'épargne-retraite des Québécois de 24 à 59 ans (MAWANI et PAQUETTE, 2011). Les auteurs tentent de discerner l'effet prix des TEMI, les effets fiscaux et effets de variables sociodémographique sur les montants retirés ou la probabilité d'observer un retrait. Les auteurs formulent leurs hypothèses grâce au modèle de Daly (1981). Les auteurs différencient les comportements d'épargne des individus qui retirent du REER et des individus qui font défaut au RAP. Ils utilisent les données de la Banque de données administratives longitudinales (DAL) des années 1991 à 2003 et restreignent leur analyse à la province de Québec. L'accès aux annexes détaillées des fichiers T1 et T4 des contribuables leur permet de calculer le TEMI et les montants maximaux de contribution aux REER avec un niveau de précision élevé. Les auteurs utilisent un modèle logarithmique et un modèle de moindres carrés généralisés pour analyser les effets des différentes variables sur la décision du retrait et sur le montant retiré lors de l'année 2000.

Grâce à leurs résultats économétriques, Mawani et Paquette rejettent que la relation entre retraits des REER et TEMI soit strictement négative. Ils montrent plutôt qu'il subsiste une relation non linéaire en forme de « U », avec un point d'inflexion au seuil de TEMI de 35% à 40%. De plus, contrairement aux hypothèses préétablies, ils démontrent que les individus qui contribuent fréquemment aux REER sont également plus susceptibles de retirer régulièrement. Finalement, les auteurs valident l'hypothèse qu'il subsiste une relation négative importante entre les chocs négatifs de revenu et le montant retiré.

D'une approche plus descriptive, Messacar, de la Division de l'analyse sociale et de la modélisation chez Statistiques Canada, utilise les données du DAL pour décrire les tendances de contributions et de retraits des REER avant l'âge de la retraite pour l'ensemble du Canada

(MESSACAR, 2017). L'auteur montre que les contributions aux REER chez les 25 à 54 ans ont largement diminué de 2000 à 2013, passant de 30,6 milliards de dollars à 22,5 milliards de dollars, alors que les contributions au compte d'épargne libre d'impôt (CELI) ont augmenté de 9 milliards de dollars en 2009 à 15,9 milliards de dollars en 2013. De manière cohérente avec Mawani et Paquette, il montre les individus qui retirent de leurs REER sont plus susceptibles d'être des hommes célibataires qui ont un faible revenu d'emploi, qui ont peu de revenu de capital et qui reçoivent de l'assurance emploi. En observant la distribution des montants des retraits, Messacar observe un effet de regroupement autour des valeurs nominales de 5 000 \$ , 10 000 \$ et 15 000 \$. Messacar attribue ce phénomène au barème d'imposition en vigueur pour les retraits des REER qui crée des effets inframarginaux élevés aux valeurs de 5 000 \$ et 15 000 \$.

Dans sa thèse de doctorat, Messacar analyse l'élasticité-prix des retraits de régimes d'épargne préretraite grâce à une modification apportée à la structure du barème des taux de retenue à la source au Québec en 2005 (MESSACAR, 2016). La réforme mise en place par le gouvernement du Québec harmonise les taux de retenues à 16 % pour l'ensemble des retraits alors que les taux moyens étaient de 20 % pour des retraits d'au-delà de 5000 dollars \$. L'auteur utilise deux méthodes pour évaluer l'effet de ce changement de loi. D'abord, il utilise une méthode de différence en différence pour analyser le choc exogène de la baisse du taux de retenue entrée en vigueur. Il compare les changements de tendances au Québec avec le reste du Canada. Cette expérience naturelle permet à Messacar d'évaluer le comportement d'épargne des consommateurs en fonction des théories de lissage de consommation et d'impatience. Il utilise finalement une méthode non paramétrique développée par Saez (2010) et Kleven et Wassem (2013) pour estimer l'élasticité-prix aux points de discontinuité de barème des taux de retenue à la source (SAEZ, 2010, KLEVEN et WASEEM, 2013). L'auteur utilise les microdonnées de l'Enquête sur la dynamique du travail et du revenu (EDTR) de Statistique Canada. Pour des fins de robustesses, il répète son analyse en utilisant les données publiques de l'Enquête sur les dépenses des ménages. L'EDTR permet d'analyser le comportement de 17,000 ménages de 2002 à 2009 sur des panels de 3 ans. Il restreint son analyse aux contribuables âgés de 35 à 64 ans. Puisqu'il ne peut distinguer retrait et défauts du RAP, Messacar impose un seuil minimal sur la taille des retraits de 2,500 dollars pour éliminer les défauts du RAP. Grâce à sa régression de différence en différence, l'auteur trouve une élasticité-prix de près de 1,5 et une augmentation des retraits de 2,430 dollars. Ses résultats confirment l'hypothèse que les individus sont sensibles aux seuils de taux de retenues à la source. Il obtient élasticité-prix de 1.2 grâce à la méthode non paramétrique de Saez (2010), ce résultat confirme la validité des résultats de la méthode différence en différence.

## 2.1.2 Aucune entre distinction retraits des REER et défauts au RAP

Pour sa thèse de doctorat, LeBlanc s'intéresse également au rôle des REER ainsi qu'aux facteurs déterminants les retraits préretraites (LEBLANC, 2002). Il utilise une base de données longitudinales qui regroupe 5 % des payeurs d'impôts canadiens des années 1982 à 1999. L'auteur fait face à une contrainte similaire à la nôtre : il ne peut distinguer la somme accumulée dans les REER pour chaque individu. Il ne distingue pas entre défauts du RAP et retraits du REER dans son analyse. Il restreint son étude aux individus âgés de 25 à 54 ans. LeBlanc cherche à déterminer si les individus utilisent l'épargne accumulée dans les REER pour subvenir à des pertes de revenu. LeBlanc pose l'hypothèse que la relation entre une baisse de revenu et un retrait du REER est non linéaire : les larges pertes de revenus devraient proportionnellement entraîner des retraits plus importants. Leblanc évalue cette hypothèse grâce à un modèle de correction d'échantillon. Comme Mawani et Paquette, il spécifie les niveaux changements de revenus et TEMI avec une série de variables dichotomiques pour prendre en compte de l'effet non linéaire des changements de revenus et TEMI sur le retrait des REER. Les résultats de LeBlanc suggèrent que les individus qui subissent des pertes de revenus majeures sont plus susceptibles de retirer de leur REER. Il trouve, par exemple, que pour les individus de revenus moyens qui subissent une diminution de 50 % des revenus sont 10 à 15 fois plus susceptibles de retirer de leur REER. LeBlanc observe que les montants retirés sont plus importants pour les individus qui reçoivent des prestations d'assurances sociales, qui ont des TEMI faibles, qui n'ont pas de revenus d'investissements et qui subissent des événements de vie majeurs (divorce, séparation, décès du conjoint).

Giles et Maser se penchent sur l'utilisation des fonds accumulés au sein du REER avant la retraite (GILES et MASER, 2004). Ils examinent les retraits préretraites entre 1993 et 2001 et tentent de déterminer si les événements majeurs de la vie ou certaines caractéristiques sociodémographiques influencent la probabilité d'observer un retrait. Ils ne distinguent pas entre défauts au RAP et retraits. Ils observent que les contribuables des déciles de revenus inférieurs sont moins susceptibles de cotiser à un REER, mais également plus susceptibles de décaisser de leur REER. Les auteurs observent les effets de certains événements de la vie. Ils constatent que le décès d'un conjoint est l'événement le plus susceptible de mener à un retrait alors que la perte d'emploi involontaire a peu d'incidence. Ces résultats sont cohérents avec ceux de Mawani et Paquette. Giles et Maser omettent les effets prix du TEMI et du lissage d'impôt pour expliquer les retraits.

### 2.1.3 Rôle du régime d'accession à la propriété

Steele, dans son article « The Canadian Home Buyers' Plan : Tax Benefit, Tax Expenditure, and Policy Assessment », analyse l'utilisation du RAP chez les contribuables canadiens. Elle décèle d'abord que les Canadiens ont peine à rembourser les sommes retirées de leur REER par l'entremise du RAP. En 2003, les contributions totales aux REER désignées au remboursement du RAP totalisaient 525 millions \$, ce qui représentait 5,1 pour cent des balances du RAP. Cette proportion est bien en deçà du 6,7 pour cent (1/15) qui représenteraient un remboursement complet des sommes dues. Elle note également une forte croissance de la popularité du RAP, qui passe de 370,000 utilisateurs à 1,3 million entre 1995 et 2003.

Steele analyse la participation au RAP en fonction du revenu des contribuables et ménages. Elle distingue une forte participation au RAP chez les contribuables qui détiennent un revenu en deçà de 20,000 \$. Ce résultat est contre-intuitif étant donné que ces contribuables sont assujettis à des TEMI faibles et accumulent probablement peu d'épargne. Steele émet l'hypothèse que ces contribuables participent au RAP à l'aide d'un époux et feraient en fait partie de ménages à TEMI élevés. L'auteure note d'ailleurs qu'au cours des années 2001 à 2003, la participation au RAP à titre d'outil pour diviser les revenus au sein d'un ménage a augmenté. En ce qui a trait aux remboursements, elle estime que plus du tiers des contribuables qui devaient rembourser un montant ont fait défaut au cours de leur période de remboursement. L'auteure distingue que les contribuables de moins de 45 ans à faible revenu forment le groupe le plus susceptible de détenir des montants en souffrance. Ce résultat suggère que le rôle du RAP comme mesure incitative à épargner est limité pour ce sous-échantillon. Steele décerne tout de même un impact positif du RAP sur l'épargne à long terme pour la majorité des Canadiens.

Mawani et Paquette analysent également l'utilisation du RAP chez les Canadiens (MAWANI et PAQUETTE, 2011). Grâce à la base de données du DAL, ils constatent que les participants et non-participants du RAP diffèrent sur plusieurs aspects sociodémographiques. En effet, les participants du RAP sont plus jeunes, détiennent un revenu ajusté plus faible, acquièrent des revenus d'investissements hors REER plus faibles et font face à des TEMI plus élevés. Ils décèlent également que les participants au RAP retirent des montants plus faibles de manière persistante. Ils notent un lien significatif entre les défauts et les chocs de perte d'emploi, séparation ou divorce. Ils concluent finalement que les participants au RAP sont moins réceptifs aux TEMI que les contribuables qui ne participent pas au RAP. Mawani et Paquette observent que les participants au RAP qui voient leur TEMI diminuer de 5% de par rapport à leur TEMI moyen des quatre années précédentes sont plus susceptibles de faire défaut. Finalement, les auteurs émettent l'hypothèse que les participants au RAP sont moins enclins

à rembourser le montant dû lors de baisses de revenus puisqu'ils sont possiblement plus contraints en termes de liquidités comparativement aux non-participants du RAP.

#### 2.1.4 Contributions aux REER

Il est fondamental de comprendre les motifs qui poussent les contribuables à cotiser aux REER. Les bases théoriques amenées par Daly (1981) sont entre autres formulées en fonction des cotisations. Ainsi, comme Mawani et Paquette, nous posons l'hypothèse que les choix de contributions et retraits sont dépendants des effets prix et revenus analogues.

Tout d'abord, Veall, en 2001, reporte des coefficients associés l'élasticité-prix entre contributions au REER et TEMI non significatifs (VEALL, 2001). Grâce à la base de données du DAL, Veall utilise l'expérience naturelle de la réforme des seuils de taux marginaux d'imposition de 1988 au Canada pour analyser les changements de contributions des individus. Cette réforme a eu pour effet de diminuer le nombre de paliers d'impositions, qui passe de 11 à 4, d'aplatir la structure des taux d'imposition marginaux, variants de 6 % à 34 % versus 17 % à 29 %. Veall n'observe aucun coefficient significatif dans son estimation de différence en différence. Les signes des coefficients des TEMI qu'il estime sont négatifs.

Comme Veall, Milligan analyse les incidences des TEMI et TEMI futurs sur les contributions aux REER des Canadiens afin de tester l'hypothèse du lissage des charges fiscales des agents (MILLIGAN, 2002). Milligan formule l'hypothèse que les contributions dans les REER sont positivement associées aux TEMI. L'auteur utilise les données transversales de l'Enquête sur les dépenses des ménages (EDM) pour les années 1982, 1984, 1986, 1990, 1992 et 1996. L'auteur conserve les contribuables âgés de 19 à 59 ans en son échantillon final. Il exclut les ménages ayant retiré de leur REER. Dans un premier temps, l'auteur agrège les données par année et province pour examiner la contribution aux REER moyenne en fonction du TEMI moyen. Il distingue une corrélation positive entre les variables, ainsi qu'un fort effet fixe pour la province de Québec, qui, pour des TEMI similaires, contribue moins en moyenne que le reste du Canada. Dans un second temps, l'auteur utilise une régression de type probit pour mesurer les effets des variables explicatives sur la probabilité d'observer une contribution aux REER. Il inclut un riche ensemble de variables descriptives ainsi que de variables dichotomiques pour l'année et la province. Il contrôle pour le revenu et sa transformation quadratique. L'auteur estime un effet marginal significatif, mais minime du TEMI. Ses résultats indiquent qu'une augmentation de 10 points de pourcentages de TEMI augmente la probabilité d'une contribution de 2.8 pour cent. Finalement, Milligan estime le TEMI futur des ménages grâce à une modélisation du revenu des contribuables au cours de leur vie. Il utilise une variable

instrumentale (*participation à un syndicat*) reliée aux revenus et plausiblement non reliée à la contribution pour permettre l'identification du modèle. L'auteur constate un coefficient positif du TEMI, qui signifie qu'une augmentation du TEMI est associée à une augmentation du montant cotisé. Ce résultat démontre qu'un TEMI futur élevé diminue la contribution présente au REER des ménages. Ce résultat est compatible avec la théorie de l'optimisation du profil d'impôt marginal des contribuables de Daly.

Dans sa thèse de doctorat, LeBlanc étudie également les contributions aux REER en fonction des fluctuations annuelles de revenus (LEBLANC, 2002). Après avoir contrôlé pour le TEMI, il trouve qu'une augmentation/diminution de 1000 \$ du revenu est associée à une augmentation/diminution de 79 \$ de contribution aux REER. Contrairement à Milligan (2002), il ne trouve aucun lien significatif entre les changements courants de contributions aux REER et les changements de revenus futurs ou TEMI futur. Il montre une relation positive entre la contribution aux REER et la volatilité des revenus grâce à un modèle tobit censuré à zéro et à la limite de contribution. LeBlanc précise son analyse en évaluant comment la volatilité du revenu affecte les contributions en fonction du groupe de revenu des contribuables qu'il sépare entre bas, moyen et élevé. L'auteur évalue un coefficient associé à l'effet de la volatilité positif et significatif pour les contribuables à revenus moyens. Selon l'auteur, les contribuables à revenu faible ont recouru à des programmes d'assurances sociales alors que les contribuables à revenus élevés utilisent des actifs financiers hors REER, ce qui limiterait leurs contributions.

Sujet d'analyse	Auteur (année de publication)	Années observées	Approche	Méthodologie	Données
Retraits standards	Mawani et Paquettes (2011)	1991-2003	Lissage de la consommation	- Régressions Moindres Carrés Généralisés - Régression logistiquie	DAL
	Leblanc (2002)	1982-1999	Lissage de la consommation	- Correction pour échantillon auto-sélectionné (Heckman Two-Step) - Différence-en-différence	Canadian Department of Finance
	Messacar (2016)	2002-2009	Élasticité des retraits rapport aux taux de retenu à la source	- Modèles de discontinuité non-paramétrique	Enquête dépense sur les ménages
Aucune distinction	Giles et Maser (2004)	1993-2001	Lissage de la consommation	Descriptive	Statistique Canada
Défauts du RAP	Steele (2011)	1995-2005	Épargne	Descriptive	Statistique Canada, Agence de revenu du Canada
	Mawani et Paquette (2007)	1991-2003	Lissage de consommation	Descriptive	DAL
Contributions	Veall (2001)	1988	Élasticité des contributions par rapport aux TEMI	- Différences-en-différences	DAL
	Miligan (2002)	1982-1984-1986 -1990-1992	Lissage de consommation	- Probit - 2SLS	Enquête des dépenses sur les ménages
	LeBlanc (2002)	1982-1999	Lissage de consommation	- Tobit 2 limites - OLS	Canadian Department of Finance

## 2.2 L'utilisation des techniques d'apprentissage statistique en économétrie

Nous observons, depuis les quelques dernières années, une démocratisation des techniques d'apprentissages statistique au travers des sciences naturelles comme sociales. Mullainathan et Spiess, dans leur revue de littérature « Machine Learning : An Applied Econometric Approach » couvrent l'utilisation des techniques d'apprentissage statistique appliquées au domaine de l'économétrie (MULLAINATHAN et SPIESS, 2017). Les auteurs divisent l'utilisation des techniques d'apprentissage statistique en quatre type d'applications possibles : nouvelles données, prédiction en matière de lois, la validation de théories et la prédiction au service de l'estimation.

- **Nouvelles données** fait référence aux nouveaux types de données accessibles aux économistes, comme l'imagerie et le langage, qui, malgré leur forte dimensionnalité, sont désormais traitables grâce à l'apprentissage statistique. Ces techniques ont d'ailleurs été utilisées pour estimer l'activité économique d'un pays grâce à des images satellites. (HENDERSON, STOREYGARD et WEIL, 2012).
- **Prédiction en matière de lois** désigne les problèmes économiques pour lesquels une prédiction peut augmenter l'efficacité de politiques publiques. Mullainathan cite l'exemple où un juge doit décider si un accusé doit attendre sa peine en prison ou à son domicile. Il montre que des modèles d'apprentissage statistiques détiennent la capacité d'affecter les prisonniers au mode d'attente de plus optimal (KLEINBERG et al., 2015). L'utilisation de ce type d'algorithme par le système judiciaire permettrait de réduire les taux de récidivisme et les coûts du système carcéral.
- **Validation de théories** inclue les cas où l'on teste si une théorie basée sur la prédiction est empiriquement observable. Ce type d'utilisation est souvent retrouvé lorsqu'on tente de prouver l'existence d'une corrélation entre deux variables.
- **Prédiction au service de l'estimation** englobe les cas où une technique d'apprentissage statistique est utilisée afin de retrouver un coefficient d'intérêt ou de valider l'utilisation d'une technique économétrique. À titre d'exemple, les auteurs mentionnent le cas où l'on utilise des techniques de régularisation comme LASSO ou Ridge lors d'une première étape d'une régression 2SLS pour subvenir au problème de surapprentissage qui survient avec de petits échantillons. Les auteurs énoncent également le cas où l'on utilise l'apprentissage statistique non supervisé pour déterminer le niveau de similarité entre un groupe contrôle et un groupe traitement.

Dans le présent mémoire, nous développons une méthodologie qui appartient au groupe *prédiction au service de l'estimation*. La classification d'un retrait à l'appartenance d'un programme permet de subdiviser l'échantillon et finalement d'obtenir des coefficients qui reflètent les dynamiques propres à chaque type de retraits. James J. Feigenbaum, dans son article « A Machine Learning Approach to Census Record Linking » utilise une méthodologie similaire qui combine à la fois techniques d'apprentissage statistique et économétriques (FEIGENBAUM, 2016). L'auteur lie des individus au travers différentes bases de données pour ensuite mesurer la mobilité sociale. Son utilisation des outils d'apprentissage statistique est similaire à la nôtre puisque l'erreur de prédiction de l'auteur se traduit par la présence d'individus faux positifs et l'absence d'individus faux négatifs dans l'échantillon final.

Certains chercheurs ont utilisé des techniques similaires pour imputer des variables à un échantillon. Yang et al., dans leur article « Mind the Gap : Accounting for Measurement Error and Misclassification in Variables Generated Via Mining », adaptent un modèle proposé par Küchenhoff et al. (KÜCHENHOFF, MWALILI et LESAFFRE, 2006) pour corriger les coefficients associés à une variable prédite qui contiennent une erreur de mesure classique ou non-classique (YANG et al., 2018). Les auteurs démontrent l'efficacité de leur méthode grâce à trois applications concrètes du domaine des technologies de l'information. Dans une de leurs applications, les auteurs tentent de déterminer si le sentiment détecté (positif ou négatif) d'une publication sur Facebook influence le nombre de commentaires sur cette publication. Les auteurs obtiennent la matrice de confusion suivante sur la portion test de leur base de données :  $(M_{00}, M_{10}, M_{01}, M_{11}) = (0.92, 0.19, 0.08, 0.81)$  où  $M_{00}$  présente la classification de sentiment positif correctement classifié et  $M_{11}$  présente la classification de sentiment négatif correctement classifié. Yang et al. retrouvent des coefficients similaires aux coefficients générés par les valeurs réelles (YANG et al., 2018). L'analyse de Yang et al., nous indique qu'il est possible d'obtenir des coefficients représentatifs grâce aux techniques d'apprentissage statistique. Contrairement à nous, les auteurs tentent d'ajuster les coefficients à leurs vraies valeurs grâce à la technique d'extrapolation MC-SIMEX.

## 2.3 Synthèse

La présente revue de littérature nous permet de formuler quelques constats sur l'utilisation des REER chez la population canadienne. Dans un premier temps, la recherche existante s'accorde pour désigner les événements de la vie (divorce, séparation, décès d'un conjoint, perte d'emplois) comme éléments instigateurs aux retraits et défauts. De plus, Mawani et Paquette (2007) ainsi que LeBlanc (2002) identifient les baisses de revenus comme facteurs

clés pour expliquer retrait et défauts du RAP. Ces auteurs observent des relations mitigées entre TEMI et retraits. Ils démontrent que les individus ne cherchent pas nécessairement à maximiser leur « profil d'imposition temporel ». En ce qui a trait aux contributions, Veall (2001) ne distingue aucune incidence significative entre contributions et TEMI, tandis que Miligan (2002) trouve une relation significative, mais toutefois faible avec une méthode 2SLS. LeBlanc (2002) ne parvient pas à reproduire ce résultat.

En ce qui a trait au RAP, Steele (2007) observe une forte prévalence de défauts parmi les individus à faible revenu. Mawani et Paquette (2007) observent que les individus qui retirent des REER et les individus qui font défaut du RAP sont foncièrement différents sur le plan sociodémographique. Les auteurs émettent l'hypothèse que ces derniers sont probablement plus contraints en termes de liquidités.

Notre recherche se distingue de la littérature courante sur les retraits préretraites de différente manière. Tout en utilisant le modèle théorique de Daly (1981), nous étudions cette problématique grâce à une approche méthodologique novatrice. Non seulement nous utilisons des modèles économétriques qui nous permettent d'analyser le comportement d'épargne des contribuables sur un horizon de temps inégalé, mais nous étendons également la portée notre étude sur l'ensemble de la population canadienne. De plus, les données de l'ELIA et des indicateurs économiques ou financiers rattachés nous permettent d'étudier le phénomène avec nouvelle profondeur. Nous sommes par exemple en mesure d'inclure des variables sociodémographiques telles que l'éducation dans nos analyses statistiques.

Bien entendu, une faiblesse apparente de notre étude consiste à ne pas observer les limites de contributions aux REER comme sont en mesure de le faire Mawani et Paquette. En effet, l'ELIA et les documents administratifs qui lui sont rattachés ne nous fournissent aucune information sur le montant exact accumulé dans les REER au cours de la vie d'un contribuable. Mawani et Paquette sont ainsi en mesure de tester si les retraits préretraites sont négativement corrélés au pourcentage de contribution utilisé (MAWANI et PAQUETTE, 2011). Selon le modèle de Daly, l'atteinte de la limite de contribution contraint les consommateurs d'atteindre leur profil de taux marginal d'imposition optimal (DALY, 1981). Finalement, les formes fonctionnelles employées pour le présent mémoire ne nous permettent pas d'analyser les effets de la volatilité du revenu sur les habitudes d'épargne.

Dans un second temps, les techniques d'apprentissage statistique sont utilisées d'une panoplie de manières par les économètres. Nous présentons un cas d'imputation où les prédictions d'un algorithme sont utilisées pour subdiviser un échantillon. Cette démarche nous permet d'analyser retraits et défauts du RAP de manière distincte.



# Chapitre 3

## Modèles

Ce chapitre vise à expliquer les fondements théoriques de notre approche ainsi qu'à familiariser le lecteur avec les modèles statistiques utilisés pour notre analyse.

### 3.1 Modèle théorique et hypothèses

Daly (1981) formule un modèle de cycle de vie qui intègre l'avantage fiscal intrinsèque des REER pour modéliser le comportement optimal d'épargne des individus. Comme Mawani et Paquette, nous utilisons ce modèle pour formuler certaines hypothèses que nous évaluerons par la suite grâce à certains modèles économétriques conçu pour exploiter la forme des données longitudinales. Dans cette section, nous décrirons ce modèle de cycle de vie et les hypothèses qui en découlent.

#### 3.1.1 Modèle de cycle de vie

Un individu planifie sa consommation et ses contributions aux REER au cours de sa vie d'une durée de  $D$  années afin de maximiser la fonction d'utilité :

$$\int_0^D U(C(t))e^{-\rho t} dt + e^{-\rho D} \psi(W_d) \quad (3.1)$$

où  $C(t)$  est la consommation,  $W_d$  son héritage à la fin de la durée  $D$  de sa vie. Les fonctions  $U$  et  $\psi$  sont croissantes, strictement concaves et différentiables,  $\rho$  correspond au taux d'actualisation.

L'individu peut emprunter ou prêter à un taux d'intérêt  $i$  et reçoit un flux de revenu  $Y(t)$ , de la période 0 jusqu'à sa mort  $D$ . Il reçoit également un héritage  $W_0$ . Son accumulation de

richesse dans les REER est  $V(t)$ . L'individu fait face à un régime d'imposition progressif  $T(Z(t))$  où  $T' > 0, T'' > 0$  et  $T \leq T' \leq 1$ , où  $Z(t)$  dénote le revenu imposable :  $Z(t) = Y(t) + iW(t) - R(t)$ , où  $R(t)$  est la contribution aux REER qui est assujettie à la contrainte suivante :

$$R(t) \leq \min\{mY(t), M(t)\} \quad (3.2)$$

Nous obtenons les contraintes budgétaires de sa richesse hors REER :

$$W(t) = Y(t) + iW(t) - R(t) - T(Y(t) + iW(t) - R(t)) - C(t) \quad (3.3)$$

et sa richesse des REER :

$$V(t) = R(t) + iV(t). \quad (3.4)$$

Le consommateur maximise une fonction d'utilité 3.1 en fonction des contraintes posées par les équations 3.2, 3.3 et 3.4 par rapport à  $C(t)$  et  $R(t)$ . Daly traduit ce modèle par l'Hamiltonien  $\bar{H}$  :

$$\begin{aligned} \bar{H} = e^{-\rho t} \{ & U + \alpha[Y + iW - R - T(\cdot) - C] + \beta[R + iV] \\ & + \mu[\min\{mY, M\} - R] \} \end{aligned} \quad (3.5)$$

où  $\alpha, \beta$  et  $\mu$  sont des variables de coûts, et où :

$$H = e^{-\rho t} \{ U + \alpha[Y + iW - R - T(Y + iW - R) - C] + \beta[R + iV] \} \quad (3.6)$$

Daly suppose un taux d'intérêt différent pour les actifs REER  $j$  et les actifs hors REER  $i$ . Après avoir dérivé les conditions de premier ordre par rapport au temps et avoir fait les substitutions nécessaires, Daly dérive les équations canoniques des variables de coûts :

$$\delta(C) \frac{\dot{C}}{C} = j(1 - T') - \rho \quad (3.7)$$

et

$$\delta(C) \frac{\dot{C}}{C} = i - \rho - \frac{\dot{T}'}{1 - T'} \quad (3.8)$$

ce qui nous donne le chemin optimal suivant :

$$\dot{T} = [i - j(1 - T')](1 - T') \quad (3.9)$$

L'équation 3.9 implique que l'épargne incrémentale associée à contribuer un dollar additionnel en un REER peut être séparée en deux composantes (DALY, 1981). La première, la *composante différée*,  $-\dot{T}'$ , est l'épargne fiscale associée à la différence entre la charge fiscale à un âge  $t$  à  $t + dt$ . La deuxième composante, que Daly désigne comme la composante d'*intérêt accru*,  $[i - j(1 - T')](1 - T')$ , provient des revenus d'intérêts des REER exempts d'impôts sur la même période  $t$  à  $t + dt$ .

Selon Daly, il existe donc un avantage intrinsèque aux actifs en REER comparativement aux actifs hors REER : l'épargne marginale sur un dollar,  $iT'$ , imposée lors du retrait, permet une épargne fiscale nette de  $[i - j(1 - T')](1 - T')$ . Étant donné que  $[i - j(1 - T')](1 - T')$  est positif, un individu devrait choisir ses contributions et retraits en différant sa charge fiscale lors de périodes de faible taux d'imposition. Le profil d'imposition marginal d'un individu est optimisé lorsque le bénéfice marginal de contribuer un dollar additionnel au REER,  $[i - j(1 - T')](1 - T')$  est égal à son coût marginal  $\dot{T}'$ .

Nous reprenons donc certaines hypothèses formulées par Mawani et Paquette (2011). Toutes choses étant égales par ailleurs :

**Hypothèse 1 :** les retraits des REER avant l'âge de la retraite sont négativement associés au TEMI

**Hypothèse 2 :** les retraits des REER avant l'âge de la retraite sont négativement associés au revenu et aux changements de revenus

**Hypothèse 3 :** les retraits des REER avant l'âge de la retraite sont négativement associés au niveau de revenu d'investissement

**Hypothèse 4 :** les retraits des REER avant l'âge de la retraite sont positivement associés à certains événements de la vie tels que la perte d'emploi, le divorce, la séparation, la perte d'un époux, ou le mariage

Les hypothèses 1 et 2 proviennent de la théorie de Daly qui stipule que les individus maximisent leur profil d'imposition en contribuant à leur REER en période de TEMI faible et en retirant de leur REER en période de TEMI élevé. Suit l'hypothèse 3, qui projette que les périodes de faible revenu d'investissement contraignent les contribuables au niveau de leurs liquidités et poussent ceux-ci à désinvestir. Nous anticipons que les signes des coefficients associés aux variables de revenus d'investissements seront négatifs. Finalement, l'hypothèse 4 est justifiée par un mécanisme de lissage de revenu. Nos hypothèses diffèrent quelque peu

de celles amenées par Mawani et Paquette puisque nous les évaluons grâce à des modèles économétriques qui évaluent les coefficients en prenant en considération les effets fixes par individu.

## 3.2 Modèles économétriques

### 3.2.1 Wooldridge 1995

Wooldridge introduit en 1995 une méthode pour corriger le biais de sélection pour les modèles appliqués aux données longitudinales. Le modèle étudié est une variation du modèle à effets fixes. Le modèle permet une corrélation à travers le temps de facteurs non observés et ne suppose également aucune forme de l'effet non observé. Le modèle suppose une distribution normale des erreurs de l'équation de sélection. L'on permet également à ces erreurs d'être corrélées à travers le temps de manière arbitraire et l'on permet la présence hétéroscédasticité. La procédure pour estimer le modèle est simple. Pour chaque période nous estimons une forme probit ou tobit du modèle. Nous estimons ensuite le modèle grâce à une régression linéaire multivariée à effets fixes en incluant l'inverse du ratio de Mills. Nous pouvons écrire le modèle comme suit :

$$y_{it} = \alpha_i + x_{it}\beta + \epsilon_{it}; \quad i = 1, \dots, N; t = 1, \dots, T \quad (3.10)$$

$$d_{it}^* = \eta_i + z_{it}\gamma + u_{it}; \quad d_{it} = 1[d_{it}^* > 0] \quad (3.11)$$

où  $1[\cdot]$  est une fonction indicatrice qui est égale à 1 si l'argument est vrai et 0 dans le cas contraire.  $\beta$  et  $\gamma$  sont les paramètres de notre modèle;  $x_{it}$  et  $z_{it}$  sont des variables explicatives qui contiennent possiblement des éléments communs.  $\alpha_i$  et  $\eta_i$  sont des effets fixes inobservés spécifiques aux individus.  $\epsilon_{it}$  et  $u_{it}$  sont des termes d'erreurs non-observés.

Le vecteur de l'indicateur de sélection pour chaque observation  $i$  est dénoté par  $s_i = (s_{i1}, \dots, s_{iT})'$ . Nous posons l'hypothèse que  $(x_{it}, y_{it})$  sont seulement observés si  $s_{it} = 1$ . On définit les paramètres  $\hat{\beta}$  du modèle à effet fixe comme suit :

$$\hat{\beta} = \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} x_{it}^* x_{it}^*{}' \right)^{-1} \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} x_{it}^* y_{it}^* \right) \quad (3.12)$$

où, pour  $s_{it} = 1$  :

$$x_{it}^* \equiv x_{it} - T_i^{-1} \sum_{r=1}^T s_{ir} x_{ir}$$

$$y_{it}^* \equiv y_{it} - T_i^{-1} \sum_{r=1}^T s_{ir} y_{ir}$$

$$T = \sum_{t=1}^T s_{it}$$

Wooldridge affirme que l'équation 3.12 est suffisante pour montrer que les paramètres sont consistants et convergents asymptotiquement lorsque N tend vers l'infini pour un modèle à effets fixes.

Wooldridge pose un ensemble d'hypothèses pour construire le modèle suggéré. D'abord, il suppose que le processus de sélection est strictement exogène conditionnellement à  $a_i$  et  $x_i$  :

*Hypothèse 1 :*

$$E(\epsilon_{it} | a_i, x_i, s_i) = 0, \quad t = 1, 2, 3, \dots, T. \quad (3.13)$$

Pour que la matrice variance-covariance et que l'inférence soit valide, Wooldridge spécifie une hypothèse additionnelle :

*Hypothèse 1.1 :*

$$E(\epsilon_i, \epsilon_i' | a_i, x_i, s_i) = \sigma^2 I_t. \quad (3.14)$$

Il est possible de corriger la matrice variance-covariance pour obtenir une forme robuste et maintenir la validité de l'hypothèse.

À partir de l'hypothèse 1 (équation 3.13), Wooldridge suggère une méthodologie pour tester la présence d'un biais de sélection. Une forme alternative simple de l'hypothèse 1 nous permet d'identifier le biais de sélection :

$$E(\epsilon_{it} | a_i, x_i, s_i) = E(\epsilon_{it} | u_{it}) = \rho u_{it}, \quad t = 1, 2, 3, \dots, T. \quad (3.15)$$

Wooldridge nous fait remarquer que  $u_{it}$  est indépendant conditionnellement à la moyenne de  $(\eta_i, x_i, v_{i1}, \dots, v_{i,t-1}, v_{i,t+1}, \dots, v_T)$ . De plus, la régression de  $\epsilon_{it}$  sur  $v_{it}$  est linéaire et constante pour tous les  $t$ . L'alternative (3.15) nous fournit le résultat suivant :

$$E(y_{it} | a_i, x_i, u_i, s_i) = E(y_{it} | a_i, x_i, u_i) = a_i + x_{it} \beta + \rho u_{it}. \quad (3.16)$$

L'équation 3.16 nous permet de formuler une méthodologie suffisante pour tester la présence d'un biais de sélection. Nous posons d'abord un mécanisme de sélection pour chaque  $T$ , que nous estimons grâce à un modèle Probit ou Tobit :

$$d_{it}^* = z_t \gamma + u_{it} \quad (3.17)$$

À chaque période  $T$ , nous évaluons  $u_{it}$  grâce aux résidus de notre modèle :  $\hat{u}_{it} = d_{it} - z_t \gamma_t$ . On estime ensuite l'équation suivante par régression groupée sur les observations où  $s_{it} = 1$  :

$$y_{it}^* = x_{it}^* \beta + \rho v_{it}^* + e_{it} \quad (3.18)$$

Nous testons finalement pour  $H_0 : \rho = 0$  où  $H_0$  forme l'hypothèse que le biais de sélection est nul. Rejeter  $H_0$  nous indique qu'il faut corriger pour le biais de sélection.

### 3.2.2 Probit à effets aléatoires

Nous souhaitons comprendre les facteurs expliquant le choix du retrait tout en exploitant la forme longitudinale des données de l'ELIA. Le modèle probit à effets aléatoire est développé pour le cas où la variable dépendante est formée par une série de résultats binaires corrélées. Ces réponses sont obtenues comme le produit d'un processus longitudinal où un individu est classifié de manière répétée par une variable dichotomique (GIBBONS et HEDEKER, 1994). Cette variable peut prendre une forme dichotomique en fonction de la présence du retrait d'un REER ( $[0,1]$ ) ou catégorique si le retrait d'un REER est classifié en fonction d'intervalles de montant (eg  $y = 1$  si  $\in [0 : 2000[, [2000 : 4000[,$  etc. ). Nous utilisons la forme ordonnée du probit aléatoire dans ce dernier cas. Le modèle impose que la corrélation entre les termes d'erreurs pour un même individu soit une constante (ARULAMPALAM, 1999).

Nous considérons le modèle suivant :

$$y_{it}^* = x_{it}' \beta + \alpha_i + u_{it} \quad (3.19)$$

où  $y_{it} = 1$  si  $y_{it}^* > 0$  et sinon = 0.  $\alpha_i$  dénote l'effet individuel spécifique inobservable et  $u_{it}$  est l'erreur aléatoire. Nous assumons que ce dernier terme suit une distribution normale de moyenne 0 et de variance  $\sigma_u^2$ . Nous posons également l'hypothèse que, conditionnellement à  $x_{it}$ ,  $\alpha_i$  suit également une distribution  $N(0, \sigma_a^2)$  qui est orthogonale aux  $u_{it}$  et  $x_{it}$ .

Heckman (1981) montre que les paramètres de ce modèle peuvent être facilement estimés lorsque l'on note que la distribution de  $y_{it}^*$  par rapport à  $\alpha_i$  est indépendante et normale. Nous avons :

$$Prob(y_{it}|\alpha_i, x_{it}) = Prob\left(\frac{u_{it}}{\sigma_u} > \frac{-x'_{it}\beta - \alpha_i}{\sigma_u}\right) = \Phi(z_{it}) \quad (3.20)$$

où  $z_{it} = (-x'_{it}\beta + \alpha_i)/\sigma_u$  et  $\Phi$  la fonction de distribution normale. Nous dérivons la fonction de vraisemblance appropriée en fonction de  $\alpha$ , donnée par :

$$\prod_i \left\{ \int_{-\infty}^{\infty} \prod_{t=1}^T [1 - \Phi(x'_{it}\beta^* + \sqrt{\frac{\rho}{1-\rho}}\alpha^*)]^{1-y_{it}} [\Phi(x'_{it}\beta^* + \sqrt{\frac{\rho}{1-\rho}}\alpha^*)]^{y_{it}} \phi(\alpha^*) d\alpha^* \right\} \quad (3.21)$$

où  $\beta^* = \beta/\sigma_u$ ,  $\alpha^* = \alpha/\sigma_a$  et où  $\rho = corr(v_{it}, v_{it-1}) = \sigma_a^2/(\sigma_a^2 + \sigma_u^2)$ .

Les objets d'intérêts sont les effets marginaux moyens des coefficients calculés. Ils permettent d'évaluer l'effet du changement d'une variable  $x$  sur la probabilité estimée. Dans le cas du probit à effets aléatoires, l'on obtient :

$$\frac{\partial[Prob(y_{it} = 1|x_{it})]}{\partial x_{jit}} = \phi(x'_{it}\beta)(\beta_j) = \phi(x'_{it}\beta^* \sqrt{1-\rho})(\sqrt{1-\rho}\beta_j^*) \quad (3.22)$$

### 3.3 Modèle d'apprentissage statistique

L'apprentissage statistique supervisé consiste à estimer une fonction  $f(x)$  qui prédit une variable  $y$  à partir d'un ensemble de variables explicatives  $x = \{x_1, \dots, x_n\}$ . Le processus est appelé « supervisé » puisque les paramètres de la fonction sont ajustés pour optimiser la précision de la prédiction. De manière générale, l'objectif de ce type d'apprentissage est de minimiser l'erreur de prédiction sur des données qui n'ont préalablement pas été utilisées lors de la formation de l'algorithme.

Le processus d'apprentissage est donc formé de deux étapes. Lors de l'étape d'entraînement, on estime une fonction prédictive  $F(x)$  sur une portion des données (ensemble de formation) de façon à ce que  $F(x)$  minimise la valeur de l'espérance d'une fonction de perte choisie  $L(y, F(x))$ . On valide ensuite la performance réelle de la fonction grâce à la portion restante de nos données (ensemble test). Dans le cas présent d'un algorithme de classification binaire, nous utilisons une fonction de perte qui est la forme négative du log-vraisemblance (entropie croisée).

$$L(y, F(x)) = \log(1 + e^{-2yF}) \quad (3.23)$$

La fonction de perte en apprentissage statistique est une condition d'optimisation. Par exemple, une régression linéaire minimise l'erreur au carrée moyenne. Dans le cas d'un problème de classification, l'entropie croisée est la forme la plus utilisée. Nous évaluons la performance de notre modèle en fonction de l'erreur absolue en termes de pourcentage. Un algorithme qui obtient une erreur semblable sur l'ensemble d'entraînement et l'ensemble test est en mesure de « généraliser » sa prédiction.

Pour ce mémoire, nous utilisons une méthode d'estimation « Gradient Boosting » développée par Jerome H. Friedman. Faisant partie de la famille des méthodes d'ensemble, l'algorithme Gradient Boosting est généralement constitué de centaines d'arbres de décisions peu profonds (4 à 12 nœuds). Chacun de ces arbres est optimisé grâce à une méthode de descente du gradient. Cette méthode permet à l'algorithme d'ajuster ses paramètres sur plusieurs itérations consécutives en fonction de la valeur de la fonction de perte donnée à l'itération précédente. Ce processus itératif d'optimisation se nomme stratégie « glouton ». Cette méthode d'optimisation permet à l'algorithme Gradient Boosting d'être généralement plus performant qu'un algorithme de forêt aléatoire. En effet, d'un côté, l'algorithme de forêt aléatoire agrège les prédictions de centaines d'arbres peu profonds et minimise ainsi la variance. De l'autre, l'algorithme de Gradient Boosting est constitué de centaines d'arbres qui s'ajustent à chaque itération, réduisant à la fois le biais et la variance. Bien entendu, cette méthode rend l'algorithme de Gradient Boosting plus susceptible de surapprendre l'échantillon d'entraînement. Il faut porter une attention particulière à la méthodologie que nous employons lors de son optimisation. Nous présentons à la section 4.2 un ensemble d'hyperparamètres qui permettent d'éviter ce problème. Ces hyperparamètres imposent des contraintes à la minimisation de cette fonction de perte lors de l'entraînement du modèle.

L'algorithme Gradient Boosting vise à optimiser l'espace numérique plutôt que l'espace paramétrique d'une fonction (FRIEDMAN, 2001). En général, un modèle  $F(x; P)$  qui vise l'optimisation d'un paramètre est donné par :

$$P^* = \arg \min_P \theta(P) \quad (3.24)$$

où

$$\theta(P) = E_{y,x} L(y, F(x; P)) \quad (3.25)$$

Dans le cas d'une technique d'optimisation numérique, nous optimisons 3.24 grâce à une solution de la forme suivante :

$$P^* = \sum_{m=0}^M p_m \quad (3.26)$$

où  $p_0$  est une valeur aléatoire à la première itération et  $\{p_m\}_1^M$  sont des incréments successifs de "boosts", tous basée sur leur incréments précédents. Ces boosts permettent ainsi aux paramètres de converger jusqu'à ce que la fonction de atteigne un minimum local ou global. Nous obtenons  $p_m$  selon une fonction définie d'optimisation.

Donc, dans le cas d'un algorithme de Gradient Boosting, nous développons une fonction de forme additive :

$$F(\mathbf{x}; \{B_m, \mathbf{a}_m\}_1^M) = \sum_{m=1}^M B_m h(\mathbf{x}; \mathbf{a}_m) \quad (3.27)$$

où la fonction  $h(\mathbf{x}; a)$  est un arbre de classification paramétrée des variables explicatives  $\mathbf{x}$ , caractérisée par les paramètres  $\mathbf{a} = \{a_1, a_2, \dots\}$  et  $B$  est la meilleure descente possible du gradient. La meilleure descente du gradient donne le meilleur ajustement possible aux paramètres pour minimiser la fonction de perte. On estime la distribution conjointe de  $(y, \mathbf{x})$  estimée par un ensemble de données finies  $\{y_i, \mathbf{x}_i\}_1^N$ . Pour ce faire, on estime les paramètres qui nous permettent de minimiser la fonction de perte suivante :

$$\{B_m, \mathbf{a}_m\}_1^M = \arg \min_{\{B'_m, \mathbf{a}'_m\}_1^M} \sum_{i=1}^N L(y_i, \sum_{m=1}^M B'_m h(\mathbf{x}_i; \mathbf{a}'_m)) \quad (3.28)$$

Nous appliquons une stratégie « glouton ». Pour  $m = 1, 2, \dots, M$ , où  $M$  est un nombre d'itérations prédéfinies :

$$(B_m, \mathbf{a}_m) = \arg \min_{B, \mathbf{a}} \sum_{i=1}^N L(y_i, F_{m-1}(\mathbf{x}_i) + B h(\mathbf{x}_i; \mathbf{a})) \quad (3.29)$$

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + B_m h(\mathbf{x}; \mathbf{a}) \quad (3.30)$$

À la première itération, les paramètres  $a$  de départ sont fixés à des valeurs proches équivalente à 0. Les paramètres du modèle s'ajustent ensuite à chaque itération  $m$  en fonction de la perte précédente. Ce processus s'appelle « boosting » dans le domaine de l'apprentissage statistique.

La valeur  $B_m$ , qui est la meilleure descente possible du gradient à chaque étape, est donnée par :

$$B_m = \arg \min_B \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + Bh(x_i; a_m)) \quad (3.31)$$

L'algorithme Gradient-Boosting permet donc de remplacer le difficile problème de minimisation de (3.28) par de multiples minimisations de la fonction de perte, suivit par une optimisation de paramètre basé sur le critère original (3.31).

Il est possible d'interpréter l'importance relative des variables explicatives de l'algorithme de Gradient Boosting. Nous mesurons l'importance relative  $I_j$  d'une variable  $x_j$  en fonction de la variation qu'elle apporte sur  $\hat{F}(\mathbf{x})$ . Nous pouvons caractériser cette mesure comme suit :

$$I_j = \left( E_x \left[ \frac{\partial \hat{F}(\mathbf{x})}{\partial x_j} \right]^2 * var_x[x_j] \right)^{1/2} \quad (3.32)$$

## Chapitre 4

# Apprentissage statistique : distinction entre retraits et défauts

Comme le suggère certaines publications de la littérature, il est important de distinguer les retraits des défauts du RAP pour analyser les facteurs expliquant les retraits (MAWANI et PAQUETTE, 2011). Effectivement, ces deux types de retraits sont assujettis à des contraintes distinctes et surviennent lors de circonstances qui leur sont propres. Par exemple, les défauts du RAP ne peuvent dépasser un montant de 1,330 dollars puisque le montant maximal emprunté au RAP est de 20,000 dollars payable sur 15 ans. Les défauts du RAP surviennent également lorsqu'un contribuable ne contribue pas en son REER. Il est envisageable que les contribuables soient moins sensibles au TEMI lors du défaut puisque ce dernier suit un processus plus inconscient ou involontaire que le retrait.

Puisque la base de données de l'ELIA confond retraits et défauts sous une même colonne, nous utilisons la base de données du DAL pour former un modèle d'apprentissage statistique qui nous permet d'inférer si un montant appartient à la catégorie défaut ou la catégorie retrait. Cette démarche est possible puisque les deux échantillons proviennent des mêmes populations (sont indépendamment et identiquement distribués) et puisque des deux bases de données ont des dizaines de variables provenant des fichiers administratifs en commun. Nous pouvons ainsi prédire hors échantillon en l'ELIA tout en assumant un niveau de performance similaire avec l'échantillon test du DAL.

En cette section, nous détaillons d'abord les étapes entreprises pour obtenir notre échantillon d'entraînement et de test dans la base de données du LAD. Nous exposons par la même occasion les distributions des montants retirés et défauts en notre échantillon final. Avant de finalement présenter les performances de notre modèle, nous offrons au lecteur un aperçu de la méthodologie d'entraînement utilisée.

## 4.1 Données : La Banque de données administratives longitudinales

La base de données du DAL contient près de 20 % des payeurs d'impôts canadiens, soit plus de 7 millions d'observations par année. Pour des fins d'efficacité et à cause de contraintes de puissance de calcul, nous sélectionnons un échantillon aléatoire de 200,000 d'observations par année. Nous conservons les contribuables au travers des années consécutives si leur observation est répétée pour conserver la forme longitudinale des données. Nous rajoutons également de nouveaux contribuables pour simuler l'arrivée de nouveaux participants dans l'enquête. Nous éliminons ensuite de notre échantillon les contribuables qui n'ont pas retiré des REER ou qui n'ont pas fait défaut du RAP. Nous conservons les valeurs nominales des montants exprimés en dollars : comme nous le constatons en cette section, les catégories de retraits sont associés à certaines valeurs nominales de retraits. Nous entraînons donc notre algorithme « gradient boosting » à partir de 210,765 observations (75%) de la base de données et nous testons sa performance sur 70,256 autres observations (25%).

Nous portons un intérêt particulier aux distributions des montants de retraits des REER et défauts du RAP en la base de données. Nous utilisons pour les figures 4.1 et 4.2 la densité comme mesure de distribution. La densité est donnée par la formule suivante :

$$h(x) = \frac{f_i}{n(c_i - c_{i-1})} \quad (4.1)$$

Où  $c_i$  donne les limites supérieures et inférieures des classes de TEMI,  $f$  donne la fréquence de la classe dans l'échantillon, et  $n$  la taille de l'échantillon. La densité permet de comparer des distributions qui ont des tailles d'observations différentes : l'aire sous la courbe totalise 1 pour chaque distribution présentée.

Nous remarquons d'abord à la figure 4.1 un phénomène qu'observe également Messacar dans sa publication « Trends in RRSP Contributions and Pre-retirement Withdrawals, 2000 to 2013 » : les retraits sont regroupés aux valeurs nominales de 5, 000 et 10, 000 et finalement 15, 000 dollars. Nous observons également une forte densité de retrait autour de 1, 350 dollars dû à la prévalence des défauts du RAP, qui représentent 47% des retraits des REER dans le base de donnée du LAD.

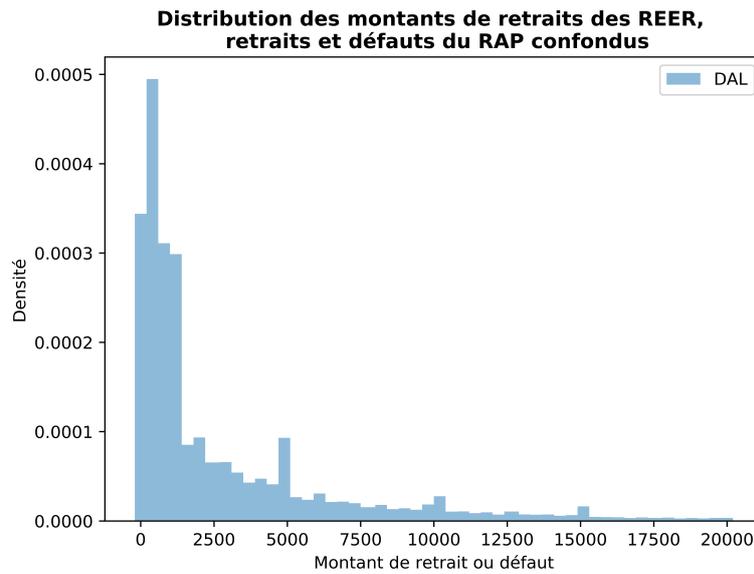


FIGURE 4.1: Distribution des retraits et défauts confondus dans la base de donnée du DAL

Distinguer entre retraits et défauts nous permet de constater un changement important à la distribution des retraits à la figure 4.2a. Quoiqu’une importante proportion (59.1%) des retraits se trouvent sous la barre des 1, 400 dollars, cette proportion est nettement moins importante lorsque l’on retranche les défauts du RAP de l’échantillon (29.5%). Nous remarquons, grâce à la figure 4.2b, que la plupart des contribuables remboursent partiellement leur montant dû en leur REER, mais qu’une importante proportion (10.3%) ne rembourse aucunement la somme attendue.

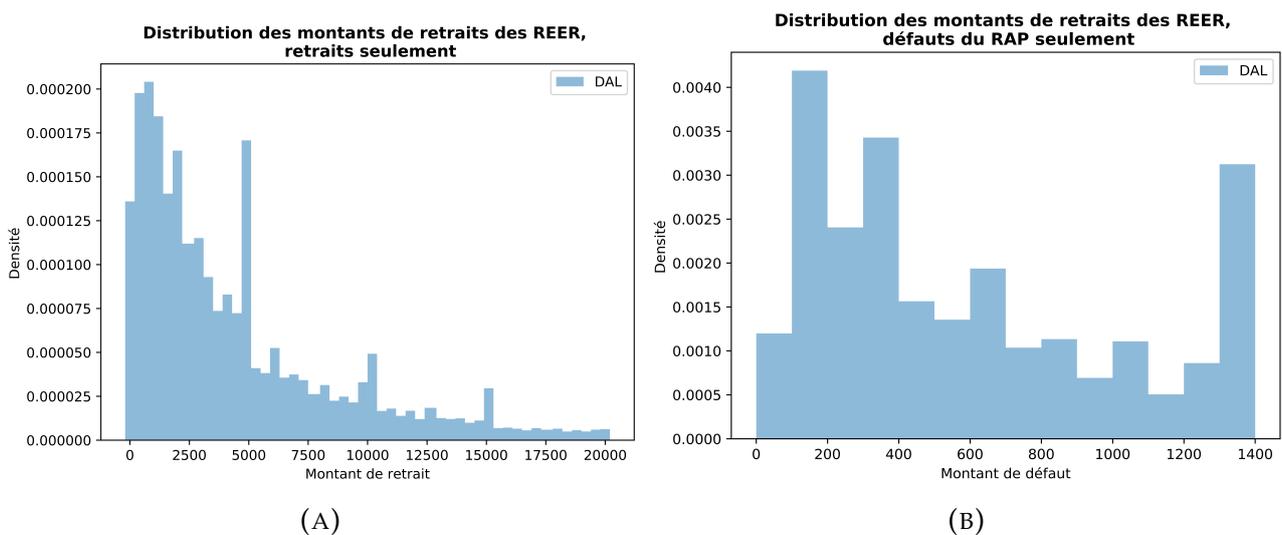


FIGURE 4.2: Distributions des retraits des REER et défauts du RAP dans la base de donnée du DAL

## 4.2 Apprentissage statistique

### 4.2.1 Méthodologie

Afin qu'il soit possible d'utiliser l'algorithme entraîné pour prédire le type des retraits des individus présents dans l'ELIA, nous entraînons notre modèle à partir des variables communes entre le DAL et l'ELIA. Nous rajoutons les valeurs passées de certaines variables (*Montant du retrait, Montant de contribution, Revenu d'emploi, etc.*) ainsi que certaines variables démographiques, telle que la population de la municipalité administrative, aux bases de données pour accroître la précision du modèle. Un total de soixante-quatorze variables sont donc utilisées pour former le modèle. Pour optimiser la performance de l'algorithme, nous utilisons une méthode de recherche quadrillée qui consiste à tester la performance du modèle en fonction de différentes combinaisons d'hypers paramètres. Nous choisissons le modèle qui minimise l'erreur-test (l'erreur de prédiction sur les données non observées lors de l'entraînement). Nous décrivons ici chacun des hypers paramètres utilisés pour la construction du modèle.

#### **RA (Rythme d'apprentissage)**

Le *rythme d'apprentissage* est le taux auquel les coefficients convergent à chaque itération d'entraînement pour minimiser la fonction de perte lors de la descente stochastique du gradient. La taille du RA influence notamment la capacité du modèle d'atteindre des minimums locaux : un faible RA pourrait faire « stagner » le gradient alors qu'un RA élevé pourrait empêcher aux paramètres de converger vers un minimum local.

#### **TSE (Taux de sous-échantillonnage)**

Le *taux de sous-échantillonnage* permet au modèle de mieux généraliser ses prédictions hors de l'échantillon d'entraînement. À chaque itération, le modèle s'entraîne (observe) seulement sur une portion aléatoire de l'échantillon, ce qui prévient le modèle de s'adapter excessivement aux données d'entraînement.

#### **TEV (Taux d'échantillonnage de variables)**

À chaque itération, un sous-échantillon de variables est choisi pour construire le modèle. Ce processus empêche certains coefficients de gagner trop d'importance relative, et donc de prévenir le surapprentissage.

#### **Profondeur**

La *Profondeur* est calculée par le nombre de noeuds (nombre de règles qui classifient l'échantillon  $[0,1]$ ) d'un arbre de décision. Augmenter la profondeur amenée à augmenter la complexité de l'arbre, et ainsi la probabilité que ce dernier engendre des prédictions non généralisables.

## Ridge

La méthode *Ridge* est une méthode de régularisation des coefficients qui ajoute une pénalité lors de la descente stochastique du gradient. Cette pénalité est donnée par le deuxième terme de l'équation suivante :

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (4.2)$$

où  $\lambda$  est un hyper paramètre que nous fixons à 0.01. La méthode Ridge pénalise donc les coefficients qui prennent des valeurs élevées, ce qui permet au modèle de généraliser ses prédictions hors échantillon. Dans notre cas, cette méthode régularise l'espace numérique du modèle : un  $\lambda$  élevé pénalise les paramètres  $a$  de l'équation 3.27 qui obtiennent des valeurs élevées.

## Validation croisée $k$

Les techniques de validation croisée permettent d'augmenter la précision du modèle hors de l'échantillon d'entraînement. À chaque itération, nous divisons le modèle en  $k$  sous-échantillon aléatoire. Nous utilisons  $k - 1$  échantillons pour entraîner le modèle et prédisons sur l'échantillon restant. Cette méthode fournit donc une idée vraisemblable de la performance réelle du modèle. La validation croisée est utilisée pour trouver la combinaison d'hyper paramètres optimale lors de la recherche quadrillée. Une fois cette combinaison trouvée, nous ré-entraînons un modèle sur l'échantillon d'entraînement complet et testons finalement sa performance sur l'échantillon test.

### 4.2.2 Résultats de l'entraînement

La figure 4.3 présente les résultats de l'entraînement avec validation croisée  $k$  pour 16 modèles testés à chaque itération d'entraînement, sur l'échantillon d'entraînement. À chaque itération d'entraînement, l'erreur est mesurée sur les échantillons d'entraînement et de test. Cette méthode nous permet de constater si l'ajustement des arbres de décisions en fonction du gradient permet d'augmenter la performance de l'algorithme. Bien entendu, le gain marginal

de l'ajustement de la fonction sur la performance du modèle est généralement marginalement décroissant.

Grâce à la figure 4.3a, nous constatons que les modèles dont le rythme d'apprentissage a été fixé à 0.05 convergent plus rapidement que les modèles où le rythme d'apprentissage de 0.01. On observe également que les modèles 4, 5, 6 et 7 avec une profondeur de 12 noeuds et un RA de 0.05 convergent rapidement, et apprennent excessivement des données de l'échantillon d'entraînement : leurs erreurs de généralisation (la différence entre l'erreur d'entraînement et l'erreur test) excèdent celles des autres modèles.

On constate par contre que les modèles 0, 1, 2 et 3 obtiennent des taux de précision sur l'échantillon d'entraînement plus faibles que ces derniers modèles, mais parviennent mieux à généraliser sur l'échantillon test à chaque itération.

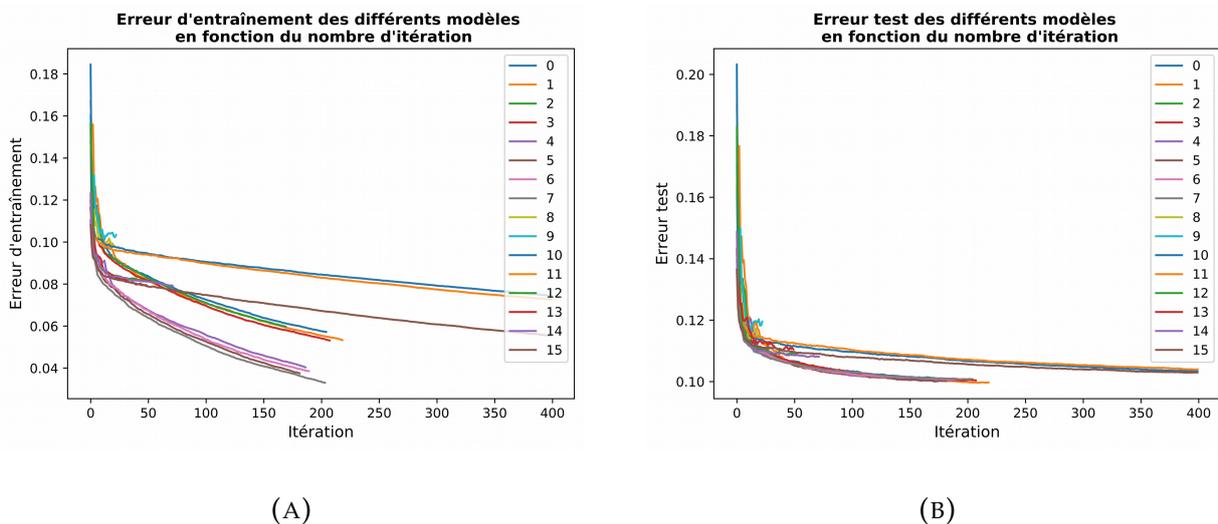


FIGURE 4.3: Résultats de l'entraînement du modèle

La table 4.1 présente les résultats finaux de la recherche quadrillée. Le modèle 1 avec un TSE de 0.7, un TEV de 0.6 et une profondeur de 10 noeuds obtient des erreurs test et d'entraînement inférieures aux autres modèles. C'est ce modèle que nous sélectionnons pour prédire le type de retrait sur la base de données de l'ELIA.

TABLEAU 4.1: Résultats de la recherche quadrillée

Modèle	Erreur-test	Erreur-entraînement	TSE	TEV	Profondeur	RA	Ridge
0	0.1007145	0.057196	0.6	0.6	10	0.05	1
<b>1</b>	<b>0.0997205</b>	<b>0.0532835</b>	<b>0.7</b>	<b>0.6</b>	<b>10</b>	<b>0.05</b>	<b>1</b>
2	0.101147	0.0596865	0.6	0.8	10	0.05	1
3	0.1003625	0.053131	0.7	0.8	10	0.05	1
4	0.1003905	0.040305	0.6	0.6	12	0.05	1
5	0.1002145	0.0374855	0.7	0.6	12	0.05	1
6	0.100733	0.038541	0.6	0.8	12	0.05	1
7	0.100795	0.0330075	0.7	0.8	12	0.05	1
8	0.114334	0.0979375	0.6	0.6	10	0.01	1
9	0.1195165	0.103367	0.7	0.6	10	0.01	1
10	0.103281	0.0743865	0.6	0.8	10	0.01	1
11	0.1039375	0.072651	0.7	0.8	10	0.01	1
12	0.108801	0.081156	0.6	0.6	12	0.01	1
13	0.110455	0.0790975	0.7	0.6	12	0.01	1
14	0.108073	0.0791875	0.6	0.8	12	0.01	1
15	0.102953	0.0552655	0.7	0.8	12	0.01	1

Finalement, pour nous assurer de la fiabilité de notre modèle, nous testons la performance de notre modèle sur une portion *validation* de notre échantillon. Présentant 25% de l'échantillon utilisé, les paramètres de la fonction ne se sont jamais ajustés en fonction de ces données. Ce test nous permet d'obtenir la matrice de confusion suivante :

$$\Pi = \begin{pmatrix} VP : 0.9092 & FP : 0.0907 \\ FN : 0.1039 & VN : 0.8960 \end{pmatrix}$$

Le taux de vrais positifs (*VP*) exprime le taux d'observations qui ont été classifiées comme étant un défaut du RAP et qui s'avéraient des défauts du RAP. Le taux de faux positifs (*FP*) exprime le taux d'observations qui ont été classifiées comme des défauts du RAP alors qu'ils étaient des retraits. Le taux de faux négatif (*FN*) exprime le taux d'observations classifiées comme des retraits alors qu'elles s'avéraient des défauts du RAP. Le taux de vrais négatif (*VN*) donne le taux d'observations correctement classifiées comme des retraits des REER.

Nous obtenons un taux de classification global de 90.34%. Ce taux représente la plus fiable estimation de la précision que nous obtenons sur les données de l'ELIA. Nous posons ainsi l'hypothèse que ce taux de réussite est équivalent entre les deux bases de données. Il nous est

possible de formuler cette hypothèse puisque les deux échantillons proviennent de la même population et sont indépendants et identiquement distribués. Ce taux de réussite est fiable d'autant plus que la différence entre l'erreur-test et l'erreur de validation est minimale : notre modèle détient une forte capacité à généraliser.

La courbe ROC (*Receiver operating characteristic*) permet de visualiser la performance des modèles de classification d'apprentissage statistique supervisé. Le taux de Vrais Positifs (Réceptivité) est tracé en fonction du taux de Faux Positifs (Spécificité). La courbe nous informe sur la capacité d'un modèle à distinguer entre deux classes en fonction du seuil de classification. Un modèle qui distingue parfaitement entre deux classes aurait une courbe ROC qui atteint le coin supérieur gauche du graphique. L'aire sous la courbe (statistique  $c$  de concordance) nous donne ainsi la probabilité que le modèle prédise à une observation positive une probabilité supérieure à une observation aléatoire négative. Nous obtenons pour le présent modèle une statistique  $c$  de 96,21. Le modèle sépare donc adéquatement les observations positives (défaut) des observations négative (retrait).

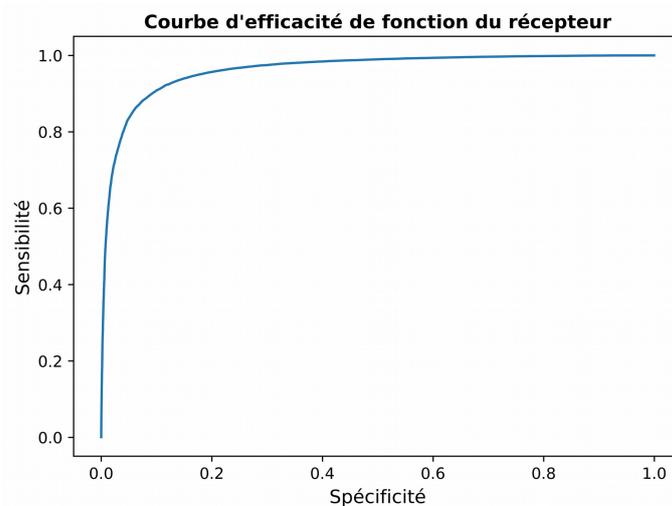


FIGURE 4.4: Performance du modèle d'apprentissage statistique

Nous tenons également à comprendre la fonction qui sous-tend les prédictions de notre algorithme. La figure 4.5 montre l'importance relative des trente variables plus importantes de notre algorithme gradient boosting. L'importance relative des coefficients est calculée selon la formule 3.32. On remarque que parmi les cinq variables contribuant le plus à la variance de la prédiction, quatre variables représentent les montants et les valeurs passées des montants des retraits des REER.

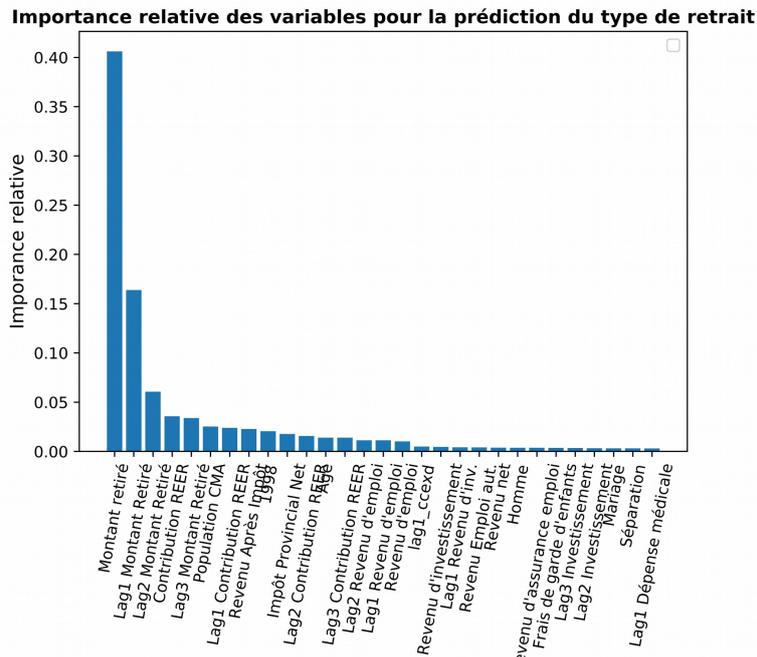


FIGURE 4.5: Interprétation du modèle d'apprentissage statistique

Finalement, nous prédisons le type de retrait des contribuables avec les données de l'ELIA grâce aux mêmes variables utilisées lors la formation de l'algorithme. Nous formons cette prédiction sur les données où un retrait est observé. Nous prédisions ainsi près de 4900 retraits dans les REER et 4800 défauts du RAP à travers notre base de 80100 individus/années. Ces proportions sont semblables aux distributions trouvées dans le LAD, qui représentent, pour les retraits et défauts, respectivement 53 % et 47% en cette base de données.

Nous sommes confiants que notre modèle conserve son niveau performance en l'ELIA grâce à sa capacité de répliquer les distributions des retraits du LAD. Nous observons ce phénomène grâce aux graphiques 4.6, 4.7a, 4.7b. Ces résultats semblent valider notre hypothèse d'indépendance et d'identité des distributions qui garantit des performances similaires de l'algorithme entre les deux échantillons. Nous constatons que de discerner retraits et défauts à l'aide de méthodes d'apprentissage statistique nous permet de distinguer entre deux distributions qui se chevauchent pour les montants inférieurs à 1400\$. Par exemple, définir un seuil inférieur à 1, 400 dollars pour analyser les retraits nous aurait empêché de considérer une portion substantielle de retraits.

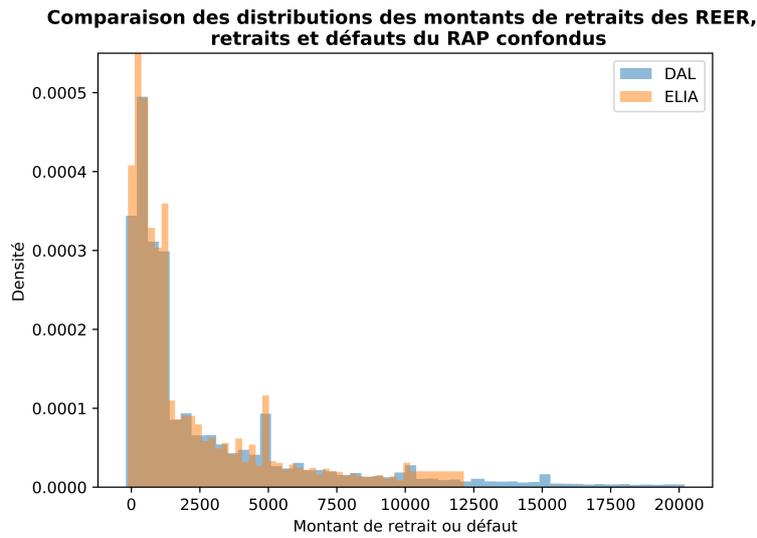


FIGURE 4.6: Distribution des montants retirés dans la base de donnée du DAL et dans l'ELIA

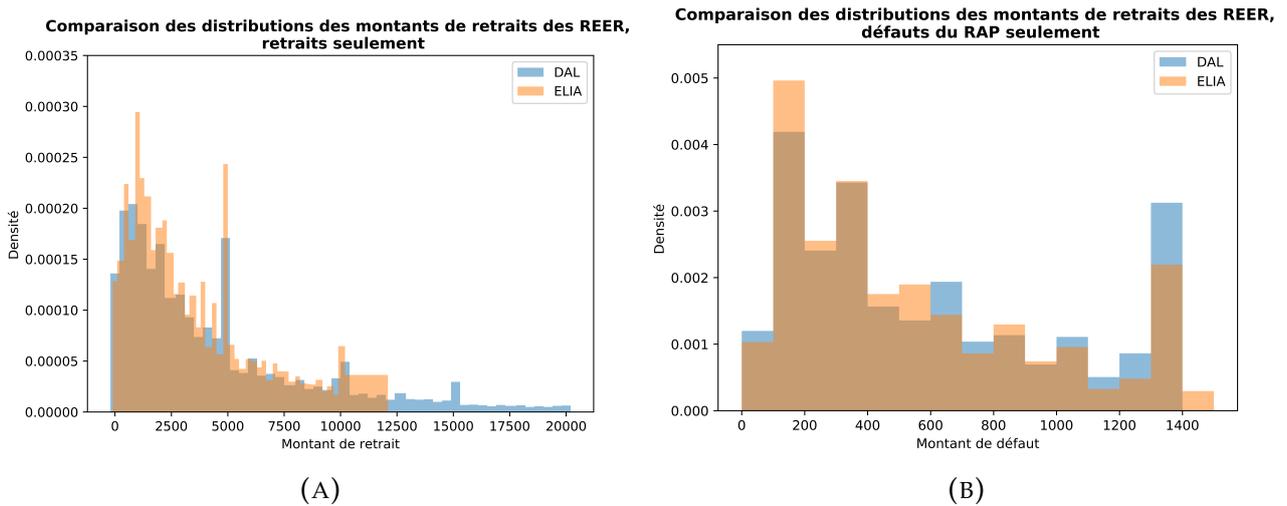


FIGURE 4.7: Distribution des montants retirés dans la base de donnée du DAL et prédits dans l'ELIA

Ainsi, au cours des chapitres Analyse univariée (5) et Analyse économétrique (6), nous ferons référence à différents sous-échantillons qui ont été subdivisés grâce à la prédiction de l'algorithme d'apprentissage statistique :

- *Aucun retrait* : rassemble tous les contribuables qui n'ont pas fait défaut ou qui n'ont pas retiré de leurs REER.
- *Aucune distinction* : confond les contribuables qui ont retiré de leur REER ou qui ont fait défaut au RAP.

- *Retrait* : regroupe tous les contribuables identifiés par notre modèle statistique comme ayant retiré de leur REER.
- *Défaut RAP* : regroupe tous les contribuables identifiés par notre modèle statistique comme ayant fait défaut au RAP.



## Chapitre 5

# Analyse statistique

En ce chapitre, nous détaillons les données de l'ELIA et la méthodologie employée pour sélectionner le sous-échantillon utilisé. Par la suite, nous détaillons le calcul des TEMI des contribuables. Nous affichons une régression et des graphiques de distribution pour évaluer la robustesse de notre estimé. Nous exposons dans les prochaines sous-sections une suite de statistiques descriptives et tests statistiques qui comparent les sous-échantillons *Aucun retrait*, *Aucune distinction*, *Retrait* et *Défaut*.

### 5.1 Données : l'Étude longitudinale et internationale des adultes

Mené par Statistique Canada au nom d'Emploi et Développement social Canada (EDSC), L'Étude longitudinale et internationale des adultes (ELIA) s'intéresse aux parcours de vie des Canadiens. Les participants sont questionnés à propos d'une multitude de sujets tels que l'éducation et la formation, la famille, la santé, le revenu et l'emploi. Les données de l'ELIA sont jumelées aux données administratives des répondants. Pour chaque contribuable, il nous est permis d'observer l'ensemble des données des formulaires T4 et T1, sans toutefois avoir accès à leurs annexes.

Nous conservons en la base de données les contribuables de 30 à 50 ans qui ont contribué en leur REER. L'ensemble des variables continues présentées dans les chapitre d'analyse statistique et économétrique sont transformées pour refléter leur valeur réelle en 2013. Sur près de 270,000 observations uniques présentes dans l'ELIA, nous obtenons un échantillon final de 80,100 observations uniques. De cet échantillon, 42.4% des 8300 contribuables uniques ont retiré ou fait défaut au RAP au cours de la période observée. Chaque année, ce sont 12.4 % des propriétaires d'un REER qui retirent un montant moyen de 2200 \$ lorsque nous

confondons retraits et défauts. Comme mentionné à la section précédente, l'algorithme que nous développons permet de d'identifier 4800 défauts et 4900 retraits en notre échantillon, pour un total de 9700 événements confondus.

## 5.2 Statistiques descriptives

Cette section se divise comme suit. D'abord, nous présentons la méthodologie employée pour estimer les TEMI par contribuable. Nous fournissons ensuite des statistiques agrégées par province et année pour offrir une perspective nationale du phénomène. Puis, nous présentons un aperçu de la fréquence des défauts et retraits chez les Canadiens. Dans la sous-section Analyse univariée, nous comparons les contribuables qui ont décaissé des REER ou fait défaut au RAP au reste des contribuables des années 1999 à 2013. Nous dressons finalement un portrait des caractéristiques socio-démographiques des contribuables en fonction du type de retrait identifié par le modèle d'apprentissage statistique.

### 5.2.1 Taux effectif d'imposition marginal

Nous utilisons le simulateur *Canadian Tax and Credit Simulator* (CTACS) afin de calculer les taux effectifs d'imposition marginaux (TEMI) (MILLIGAN, 2013). Élaboré par Kevin Milligan, professeur à l'Université de la Colombie-Britannique, le CTACS est un simulateur libre d'exploitation créé à partir de STATA. Le CTACS assigne les charges fiscales et crédits d'impôt pour les contribuables ou ménages résidants dans l'ensemble des provinces canadiennes, des années 1962 à 2016 en fonction des lois canadiennes. Le simulateur prend en intrant les micro données administratives des contribuables, leur statut matrimonial ainsi que leur province de résidence et forme en extrant une estimation des TEMI par contribuable. Le CTACS prend également en considération les données administratives du conjoint pour ajouter de la précision à ses estimations.

Nous fournissons en annexe A, au tableau A.1, une description de l'ensemble des variables utilisées pour la formation de cette estimation. Nous fournissons par la même occasion une liste des variables manquantes qui auraient pu nous permettre d'obtenir un niveau de précision supplémentaire à notre estimation des TEMI. Nous calculons le TEMI grâce à un montant marginal de retrait de 1000\$. Les valeurs nominales de l'ensemble des variables sont conservées. Nous sommes en mesure de joindre les données des conjoints grâce aux identifiants uniques des ménages fournis dans la base de données de l'ELIA. Nous excluons

les montants de retrait et contribution des REER du calcul puisque nous désirons calculer le TEMI théorique des contribuables.

Nous tenons à valider la robustesse de notre estimation du taux effectif d'imposition marginal. Nous évaluons d'abord une régression linéaire pour déterminer comment la province, l'année, les revenus et certaines variables sociodémographiques sont liés aux TEMI. Les résultats de cette régression sont présentés au tableau 5.1.

Tout comme Bancquaert et al., dans leur cahier de recherche « Quel impact la fiscalité québécoise a-t-elle sur les incitations au travail ? », nous observons des TEMI plus faibles pour les contribuables mariés (BLANCQUAERT et al., 2015). De plus, le coefficient associé à la variable dichotomique de la province de Québec présente un fort effet fixe pour les contribuables québécois, qui sont assujettis à des TEMI plus élevés que l'ensemble des provinces canadiennes. Nous observons les signes des coefficients attendus pour les variables *Revenus d'emploi*, *Emploi autonome*, *Assurance emploi*, *Gain de capital* et autres. Par ailleurs, nous observons, comme anticipé, de forts effets négatifs des variables *Revenu d'assistance sociale* et *Frais scolarité*. La régression linéaire détient un  $R^2$  ajusté de 0.33.

TABLEAU 5.1: Régression linéaire : TEMI en fonctions des revenus, provinces, années et autres variables socio-démographiques

Variable	Coefficient
Constante	<b>0.294***</b> (0.003)
Assurance emp.	<b>0.0***</b> (0.0)
Indemnisation du travailleur	<b>-0.0***</b> (0.0)
Contribution au RCP	<b>-0.0***</b> (0.0)
Gain Capital	<b>0.0***</b> (0.0)
Immigrant	<b>-0.011***</b> (0.001)
Marrié	<b>-0.004***</b> (0.001)
Célibataire	<b>-0.002***</b> (0.001)
Montant au soignant	<b>0.0***</b> (0.0)

Mâle	<b>0.011***</b> (0.001)
Rev. emploi autonome	<b>0.0***</b> (0.0)
Revenu d'emploi	<b>0.0***</b> (0.0)
(Revenu d'emploi) <sup>2</sup>	<b>-0.0***</b> (0.0)
Revenu div.	<b>-0.0***</b> (0.0)
Revenu inv.	<b>0.0***</b> (0.0)
Rev. de prestations pour enfants	<b>-0.0***</b> (0.0)
Revenu d'assistance sociale	<b>-0.0***</b> (0.0)
Frais scolarité	<b>-0.0***</b> (0.0)
Âge	<b>-0.0***</b> (0.0)
Colombie Britannique	<b>0.006***</b> (0.001)
I-P-E	<b>0.044***</b> (0.002)
Manitoba	<b>0.047***</b> (0.002)
Nouvelle-Écosse	<b>0.042***</b> (0.002)
Ontario	<b>0.015***</b> (0.001)
Québec	<b>0.079***</b> (0.001)
Saskatchewan	<b>0.02***</b> (0.002)
T-N-L	<b>0.055***</b> (0.002)
2000	<b>-0.011***</b> (0.002)

2001	<b>-0.037***</b> (0.002)
2002	<b>-0.042***</b> (0.002)
2003	<b>-0.039***</b> (0.002)
2004	<b>-0.043***</b> (0.002)
2005	<b>-0.04***</b> (0.002)
2006	<b>-0.041***</b> (0.002)
2007	<b>-0.049***</b> (0.002)
2008	<b>-0.053***</b> (0.002)
2009	<b>-0.06***</b> (0.002)
2010	<b>-0.059***</b> (0.002)
2011	<b>-0.056***</b> (0.002)
2012	<b>-0.061***</b> (0.002)
2013	<b>-0.054***</b> (0.002)
<hr/>	
F-statistic	999.6
P-value	2.22e-16
R-Squared	0.33
N	80700

Les graphiques des figures 5.1 et 5.2 nous permettent de visualiser les densités (distribution) des TEMI des canadiens âgés de 30 à 50 ans. Les taux d'imposition atteignent un minimum de -1.09 et au maximum de 0.82. Nous discernons une première concentration des TEMI autour de zéro et une seconde entre 0.25 et 0.5. Pour ce qui est de la différence des TEMI en fonction du nombre d'enfants, le tableau 5.2a montre que les contribuables ayant deux enfants à leur charge atteignent de plus faibles TEMI. L'on observe également que les contribuables ayant un enfant sont en plus forte concentration autour des valeurs de TEMI de zéro. La figure

5.2b affiche les distributions de TEMI en fonction du statut matrimonial. Les contribuables célibataires atteignent des TEMI supérieurs à 0.5 en plus forte proportion que les contribuables mariés, qui bénéficient également en plus grande proportions de taux marginaux d'imposition négatifs.

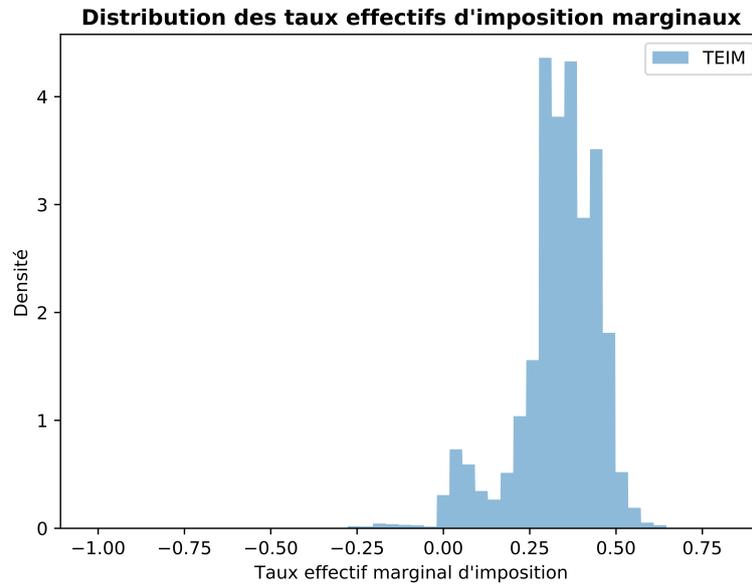


FIGURE 5.1: Distribution des TEMI

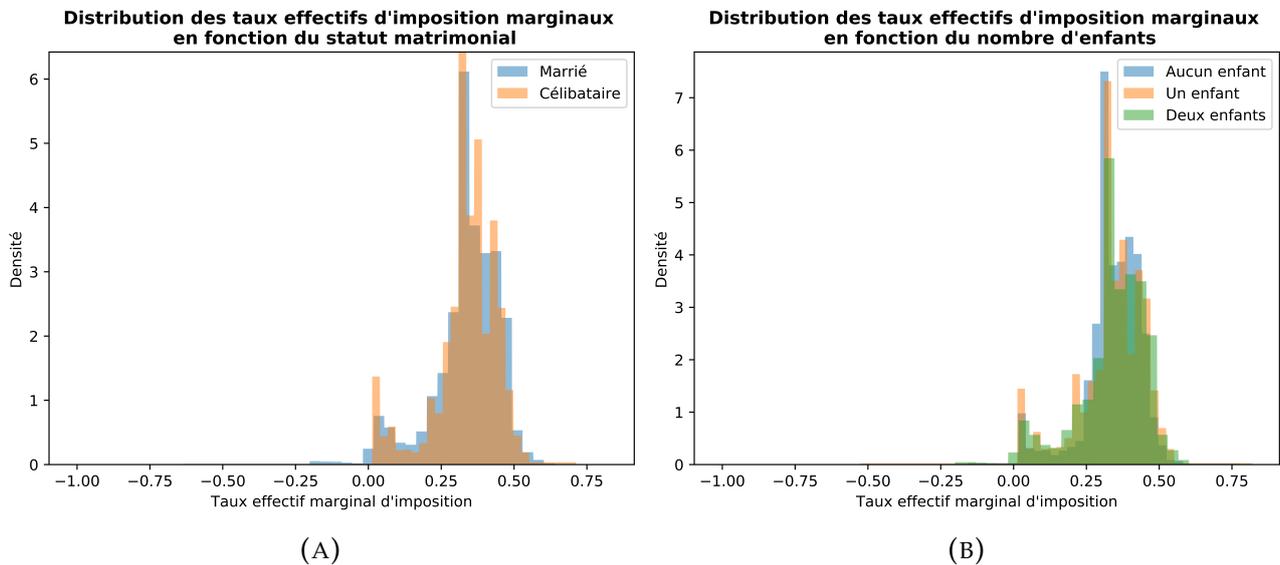


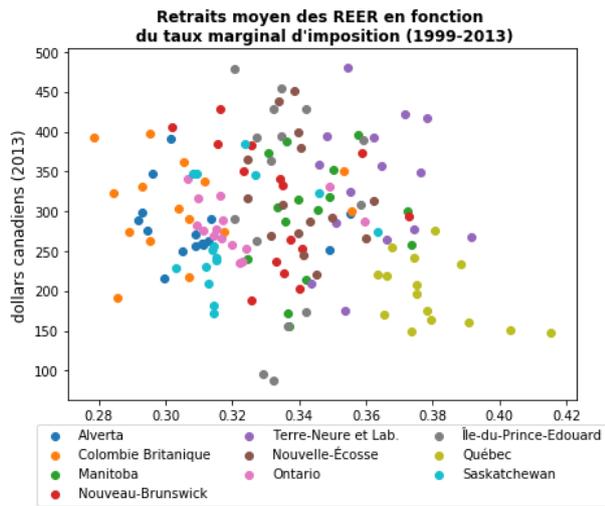
FIGURE 5.2: Comparaison des TEMI en fonction de différentes variables sociodémographiques

Quoiqu'il soit ardu de former des constats robustes sur les TEMI en fonction des caractéristiques socio-démographiques des contribuables à l'aide des graphiques présentés ci-dessus, la régression linéaire présentée au tableau 5.1 présente les effets attendus de différentes variables sur le montant d'imposition. Ces résultats confirment que le CTACS nous permet d'obtenir un estimé convenable des TEMI pour les contribuables canadiens.

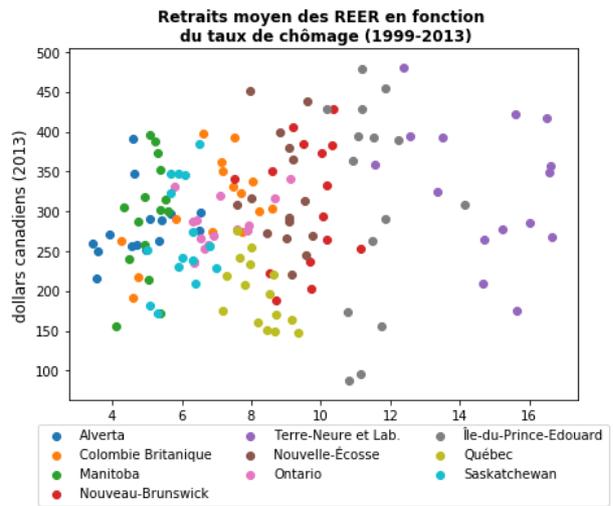
## 5.2.2 Tendances entre les provinces

Nous poursuivons notre travail de vulgarisation des données en présentant des statistiques agrégées par province et année. Les graphiques de la figure 5.3 nous permettent de distinguer certaines tendances propres aux provinces, des années 1999 à 2013, en fonction des retraits et défauts confondus. D'abord, la figure 5.3a nous permet de constater une corrélation négative entre montants moyens retirés et TEMI par province/année. Certaines provinces affichent une élasticité-prix faible comme Terre-Neuve-et-Labrador et l'Île-du-Prince-Édouard. La province de Québec détient les TEMI moyens les plus élevés et également les montants de retraits les plus faibles. On observe un coefficient de corrélation entre les TEMI et les montants des retraits de  $R^2 = -0.15$ , et un  $R^2 = -0.36$  si l'on exclut Terre-Neuve-et-Labrador, l'Île-du-Prince-Édouard et le Nouveau-Brunswick.

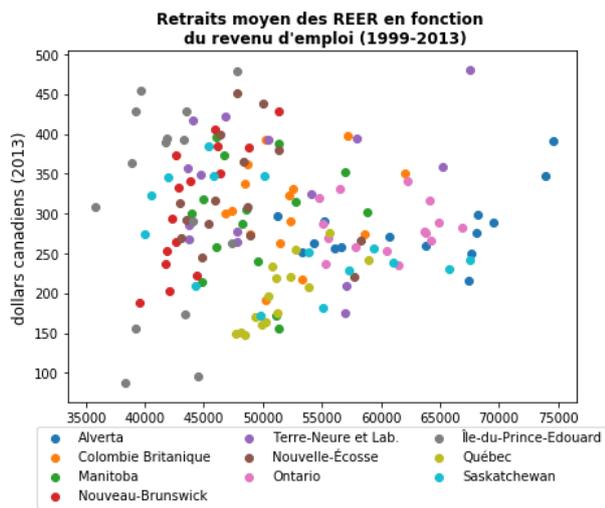
La figure 5.3b indique une absence de corrélation entre le taux de chômage et les montants retirés des REER. On observe tout de même une augmentation de la variance des montants retirés pour les provinces de l'Île-du-Prince-Édouard et Terre-Neuve-et-Labrador en fonction du taux de chômage. Les graphiques 5.3c, 5.3d, 5.3e et 5.3f expriment les montants retirés des REER confondus aux défauts au RAP en fonction de différents types de revenus. On observe des corrélations particulièrement faibles pour les cas des revenus en gain de capital et revenus de dividendes.



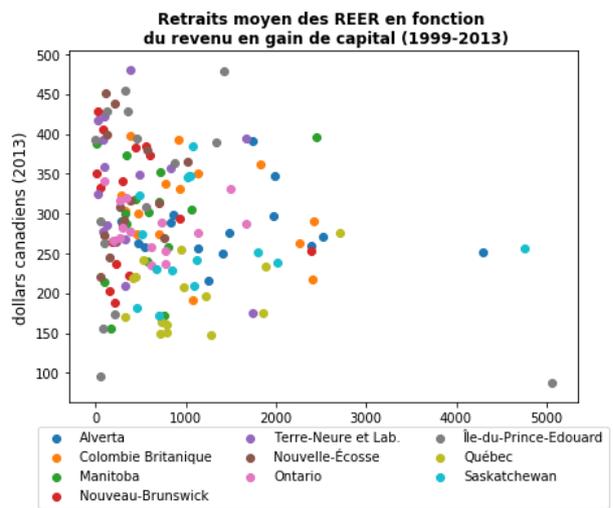
(A)



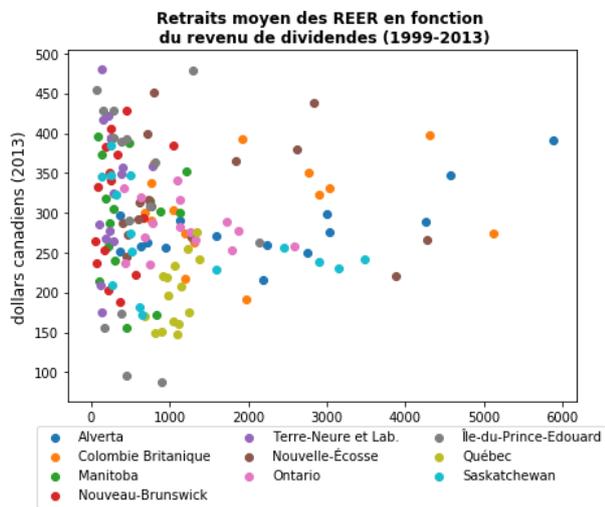
(B)



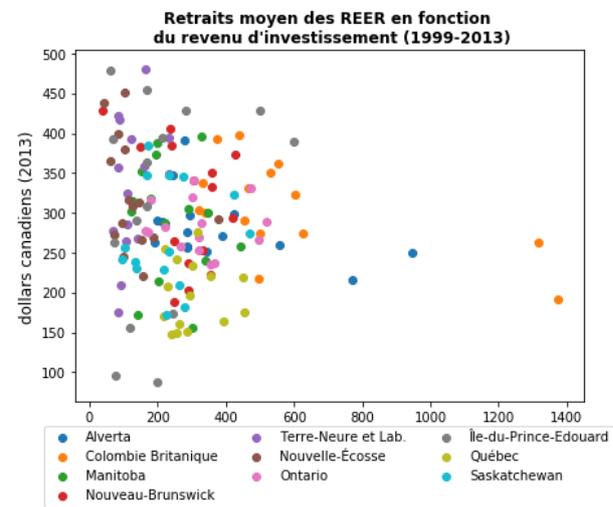
(C)



(D)



(E)



(F)

FIGURE 5.3: Retraits et défauts agrégés par provinces et années

### 5.2.3 Utilisation des REER

La figure 5.4 fournit un aperçu de la distribution de la fréquence des retraits et défauts des contribuables canadiens observés en l'ELIA. Les contribuables font généralement défaut plus fréquemment qu'ils retirent. Selon nos résultats de classification, 27.8 % des contribuables ayant fait défaut du RAP à au moins une reprise ont fait défaut 6 fois ou plus. Cette proportion n'atteint que 4 % pour les contribuables qui ont retiré de leur REER au moins une fois. Ce résultat confirme les observations de Mawani et Paquette (2011) qui remarquent un scénario « tout-ou-rien » : les auteurs observent que certains participants au RAP ne réinvestissent jamais les montants requis dans leur REER.

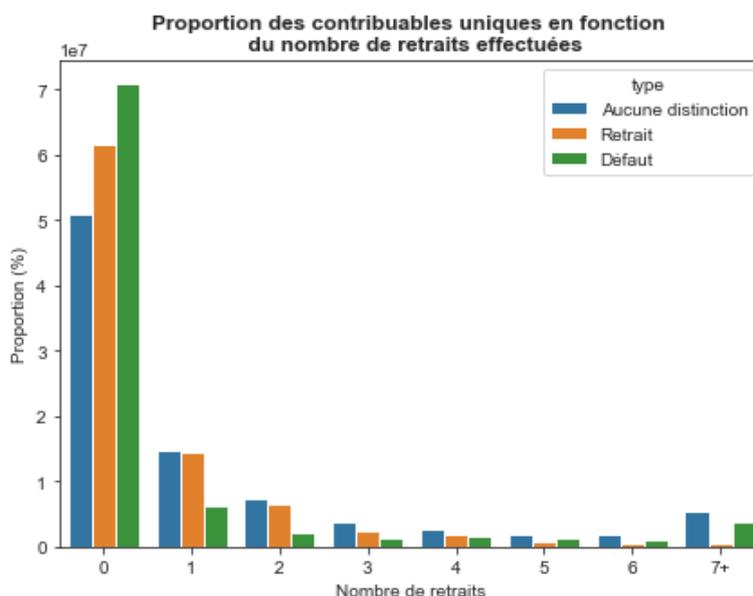
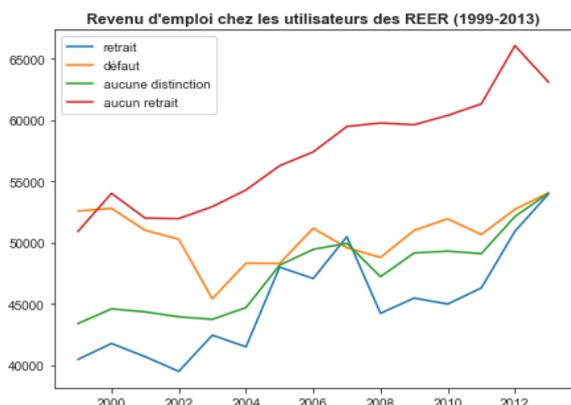


FIGURE 5.4: Fréquences des retraits par contribuable

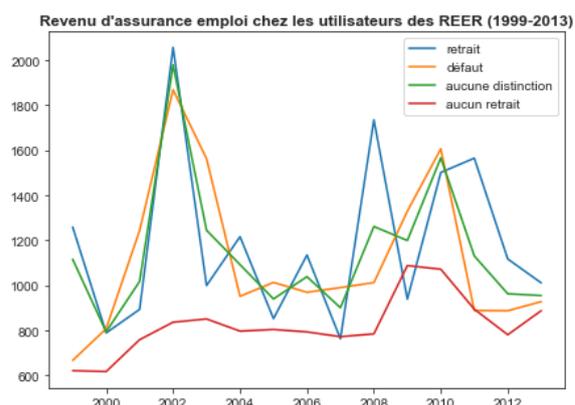
### 5.2.4 Analyse univariée

Nous souhaitons discerner la façon dont les contribuables se différencient selon leurs sources de revenus et autres variables socio-démographiques. Nous analysons ceux-ci en fonction des groupes identifiés au chapitre 4. D'abord, comme constaté par Mawani et Paquette (2011), et comme le démontrent les graphiques de la figure 5.5, une distinction importante subsiste entre les revenus d'emploi des contribuables de l'échantillon *Aucun retrait* et le reste des contribuables canadiens. Nous remarquons des tendances similaires en ce qui a trait aux montants moyens de revenus d'assurance emploi et d'investissement : ces mesures sont toutes deux « avantageuses » pour les contribuables de l'échantillon *Aucun retrait*. Finalement,

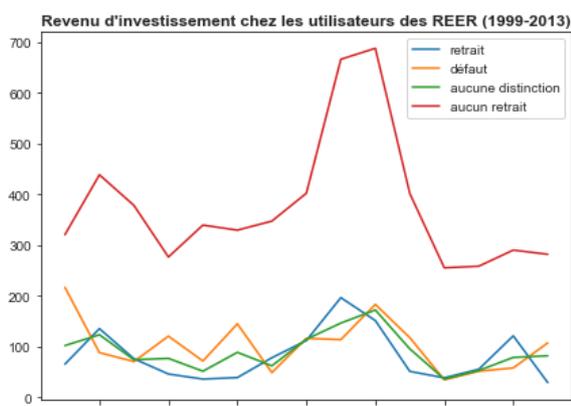
les contribuables de l'échantillon *Retrait* semblent généralement bénéficier de revenus en capital et en dividendes nettement plus faibles que les individus de de l'échantillon *Défaut RAP*. Le montant cumulé de revenu en dividendes est 3.78 fois supérieur pour ce dernier groupe pour la période étudiée.



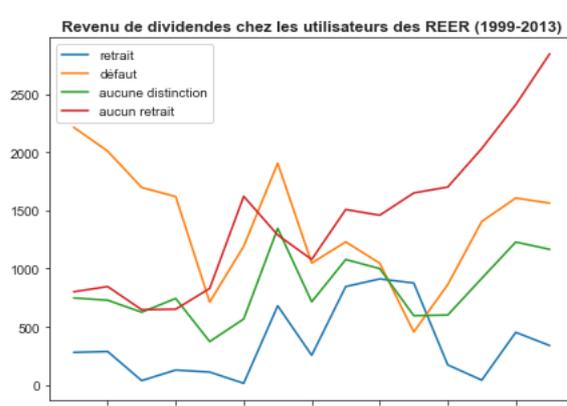
(A)



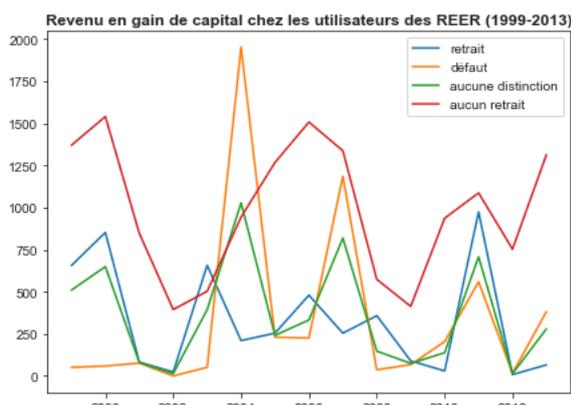
(B)



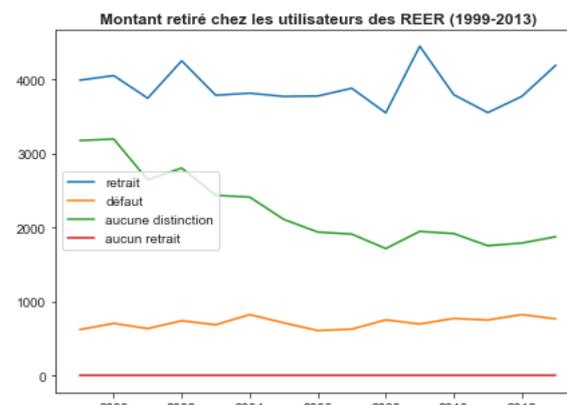
(C)



(D)



(E)



(F)

FIGURE 5.5: Tendances de revenus et retraits entre les contribuables

Le tableau 5.2 affiche les moyennes de variables continues pour les participants aux REER âgés de 30 à 50 ans par échantillon. Les tests *T* sont calculés entre l'échantillon *Aucun retrait* et les autres sous-échantillons. Les résultats des tests effectués pour les variables qui expriment des types de revenus sont statistiquement significatifs au seuil de 0.01 % et avantagent les contribuables qui appartiennent à l'échantillon *Aucun retrait*. En ce qui a trait aux TEMI, les Canadiens qui ont retiré sont assujettis à des taux effectifs d'imposition marginaux moyens plus faibles que les Canadiens du groupe *Aucun retrait*. Cette différence absolue n'est pas contre significative lorsque l'on compare l'échantillon *Aucun retrait* et *Défaut*. Une fois que l'on différencie entre les contribuables selon le type de retrait identifié, on remarque que les Canadiens appartenant au groupe *Aucune distinction* ont généré des revenus supérieurs aux Canadiens qui ont déboursé de leur REER. Cette différence est particulièrement importante pour les revenus de dividendes.

TABLEAU 5.2: Analyse statistique : variables continues

Variable	Groupe	Moyenne	Écart-type	Stat. T	Stat. P
<b>Revenu d'emploi</b>	Aucun retrait	57048.29	353.65		
	Aucune distinction	48071.32	476.03	-15.131	2.20E-16
	Retrait	44874.56	675.07	-15.969	2.20E-16
	Défaut	50679.04	663.25	-8.473	2.20E-16
<b>Revenu d'assurance emploi</b>	Aucun retrait	820.9	14.23		
	Aucune distinction	1148.34	43.73	7.119	1.10E-12
	Retrait	1186.98	63.69	5.609	2.04E-08
	Défaut	1116.81	60.05	4.794	1.64E-06
<b>Revenu de dividendes</b>	Aucun retrait	1387.06	78.21		
	Aucune distinction	845.28	86.74	-4.638	3.52E-06
	Retrait	343.96	62.87	-10.395	2.20E-16
	Défaut	1254.22	148.61	-0.791	0.4289
<b>Âge</b>	Aucun retrait	41.05	0.03		
	Aucune distinction	40.47	0.07	-7.89	3.06E-15
	Retrait	41.18	0.1	1.199	0.2307
	Défaut	39.89	0.09	-12.437	2.20E-16
<b>Revenu d'investissement</b>	Aucun retrait	380.02	16.2		
	Aucune distinction	89.22	7.62	-16.246	2.20E-16
	Retrait	80.26	10.76	-15.416	2.20E-16
	Défaut	96.53	10.69	-14.608	2.20E-16
<b>TEMI</b>	Aucun retrait	0.335	0.001		
	Aucune distinction	0.331	0.001	-2.875	0.004048
	Retrait	0.328	0.002	-2.904	0.003687
	Défaut	0.332	0.165	-1.389	0.1647
<b>Gain en Capital</b>	Aucun retrait	984.21	70.4		
	Aucune distinction	350.59	60.05	-6.847	7.59E-12
	Retrait	343.57	90.28	-5.596	2.20E-08
	Défaut	356.32	80.41	-5.875	4.25E-09
<b>Retrait</b>	Aucun retrait	0	0		
	Aucune distinction	2151.95	33.58	64.091	2.20E-16
	Retrait	3895.09	55.88	69.708	2.20E-16
	Défaut	730.01	18.09	40.35	2.20E-16

En ce qui a trait aux événements de la vie, l'échantillon de contribuables qui ont déboursé ou fait défaut ont généralement vécu des séparations ou pertes d'emploi dans de plus

importantes proportions que les individus du groupe *Aucun retrait*. La figure 5.6 illustre les taux de retraits des années 1999-2013 en fonction de ces deux types d'événements. Nous remarquons, par exemple, qu'en 2004, 5% des contribuables qui ont retiré avaient également vécu une séparation alors que cette proportion était près d'un pourcent pour les contribuables de l'échantillon *Aucun retrait*. De plus, la figure 5.6b montre que les individus du groupe *Retrait* semblent subir des pertes d'emplois dans de plus importantes proportions lors de la crise économique de 2008.

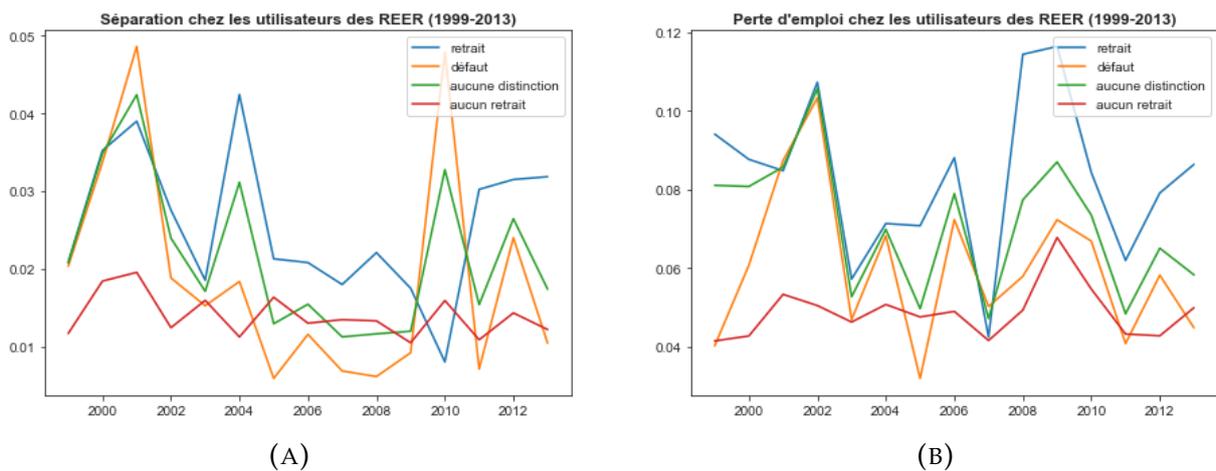


FIGURE 5.6: Tendances des événements de la vie des contribuables

Le tableau 5.3 détaille les proportions observées de diverses variables catégoriques en fonction de leur sous-échantillon. La colonne « Échantillon total » expose les répartitions de la population en fonction des diverses catégories dans notre échantillon. La colonne « Ont retiré », comme aux graphiques 5.6, fournit la distribution des individus qui ont décaissé de leur REER ou fait défaut en fonction de la variable catégorique. La colonne « Taux de retrait » donne la proportion des individus qui ont retiré ou fait défaut parmi la catégorie. Les statistiques T ( $\chi^2$ ) et P comparent ces échantillons. Par exemple, pour la variable *Perte d'emploi*, nous observons que 5.12 % des contribuables ont perdu leur emploi au cours de la période observée. De ce 5.12 %, 6.94 % ont retiré ou fait défaut au RAP. 12.12 % des contribuables qui ont conservé leur emploi ont retiré. Cette proportion est de 16.76 % pour les individus qui ont subi une perte d'emploi. Cette différence se traduit en une valeur T de 6.7 et une valeur  $P \approx 0$ .

Nous discernons des tendances de retraits et défauts distinctes entre les provinces. La Saskatchewan est la province qui détient le taux de retrait par individu le plus faible au pays (10.66%), suivit par le Nouveau-Brunswick (11.61 %) et le Québec (11.92%). Pour ce qui est

de l'éducation, les catégories pour lesquelles nous observons le plus de retraits regroupent les Canadiens qui possèdent un certificat d'apprenti (15.06%) et un diplôme de cegep (14.94%). Le groupe qui a retiré dans la plus faible proportion est composé des contribuables qui détiennent un premier diplôme professionnel (6.78%).

Le tableau 5.4 décrit les caractéristiques des individus en fonction du type de retrait identifié par notre algorithme. Le seul évènement de la vie qui s'avère statistiquement significatif pour les contribuables qui font défaut au RAP est la perte d'emploi. Dans le même sens que Mawani et Paquette, nous trouvons que les femmes ont été moins susceptibles de faire défaut et plus susceptibles de retirer de leur REER. Nous trouvons également que les contribuables qui ont fait défaut au RAP sont généralement moins âgés que les individus qui ont retiré des REER. Ces résultats sont également appuyés par Messacar (2017).

TABLEAU 5.3: Analyse statistique : sans distinction entre retraits et défauts

Variable	Réponse	Échantillon total	Ont retiré	Taux de retrait	Stat. T ( $\chi^2$ )	Stat. P
<b>Séparation</b>	Non	98.52%	97.92%	12.29%	3.21	0.01
	Oui	1.48%	2.08%	17.37%		
<b>Divorce</b>	Non	99.07%	98.78%	12.33%	2.04	0.041
	Oui	0.93%	1.22%	16.17%		
<b>Perte d'emploi</b>	Non	94.88%	93.06%	12.12%	6.70	2.094e-11
	Oui	5.12%	6.94%	16.76%		
<b>Mâle</b>	Non	47.21%	45.94%	12.03%	2.175	0.03
	Oui	52.79%	54.06%	12.66%		
<b>Province</b>	Alberta	11.58%	11.40%	12.17%	27.06	0.0001
	Colombie-Britannique	12.80%	13.09%	12.65%		
	Manitoba	3.48%	3.81%	13.56%		
	Nouveau-Brunswick	1.66%	1.56%	11.61%		
	Terre-Neuve	1.25%	1.26%	12.43%		
	Nouvelle-Écosse	2.44%	2.94%	14.86%		
	Ontario	38.68%	39.09%	12.49%		
	IPE	0.35%	0.36%	12.92%		
	Québec	25.01%	24.12%	11.92%		
	Saskatchewan	2.76%	2.38%	10.66%		
	<b>Groupe d'âge</b>	[30 :35[	21.35%	22.88%		
[35 :40[		24.68%	27.10%	13.58%		
[40 :45[		27.53%	28.16%	12.64%		
[45 :50]		26.45%	21.87%	10.22%		
<b>Éducation</b>	Certificat d'apprenti	3.07%	3.73%	15.06%	221.35	2.2e-16
	Baccalauréat	21.84%	19.26%	10.92%		
	Diplôme de cegep	5.55%	6.70%	14.94%		
	Premier diplôme professionnel	2.73%	1.50%	6.78%		
	Études secondaire	19.29%	21.38%	13.72%		
	Moins qu'études secondaires	6.68%	6.92%	12.82%		
	Maîtrise	7.67%	6.12%	9.87%		
	Certificat d'études non-universitaire	18.89%	20.32%	13.32%		
	Doctorat	1.32%	0.99%	9.25%		
	Certificat professionnel	10.35%	10.56%	12.63%		
	D.E.S.S.	2.61%	2.53%	12.02%		

TABLEAU 5.4: Analyse statistique : distinction entre retraits et défauts

Variable	Réponse	Ont Retiré		Taux retrait		Stat. T ( $\chi^2$ )		Stat. P	
		Retrait	Défaut	R	D	R	D	R	D
<b>Séparation</b>	Non	97.42%	98.32%	5.89%	7.19%	3.59	1.04	0.0003	0.295
	Oui	2.58%	1.68%	10.46%	8.54%				
<b>Divorce</b>	Non	98.51%	98.06%	5.93%	7.14%	2.39	0.53	0.017	0.596
	Oui	1.49%	1.94%	9.59%	14.51%				
<b>Perte d'emploi</b>	Non	91.73%	94.14%	5.76%	7.14%	6.40	2.50	1.5e-12	0.012
	Oui	8.27%	5.86%	9.73%	8.56%				
<b>Mâle</b>	Non	49.77%	42.82%	6.24%	6.56%	2.50	5.24	0.012	1.5e-09
	Oui	50.23%	57.18%	5.71%	7.79%				
<b>Province</b>	Alberta	13.04%	10.07%	6.64%	6.31%	178.6	108.49	2.2e-16	2.2e-16
	C-B	15.52%	11.12%	7.16%	6.34%				
	Manitoba	4.48%	3.27%	7.65%	6.89%				
	N-B	2.20%	1.03%	7.71%	4.56%				
	T-N-L	1.70%	0.90%	7.93%	5.30%				
	N-E	3.15%	2.77%	7.75%	8.30%				
	Ontario	38.68%	39.42%	5.97%	7.35%				
	IPE	0.54%	0.22%	9.04%	4.68%				
	Québec	17.47%	29.53%	4.22%	8.37%				
	Saskatchewan	3.23%	1.69%	6.78%	4.45%				
	<b>Groupe d'âge</b>	[30 :35[	20.61%	24.72%	5.82%				
[35 :40[		23.81%	29.79%	5.84%	8.68%				
[40 :45[		28.16%	28.16%	6.11%	7.38%				
[45 :50[		27.42%	17.34%	6.03%	4.74%				
<b>Éducation</b>	Certificat apprenti	3.28%	3.83%	6.77%	9.47%	431.78	138.02	2.2e-16	6.2e-15
	Baccalauréat	14.30%	21.91%	4.07%	7.42%				
	Diplôme de cegep	5.20%	7.44%	5.98%	10.08%				
	Premier diplôme professionnel	1.36%	1.50%	3.00%	4.02%				
	Études secondaires	24.62%	17.23%	7.86%	6.86%				
	Moins qu'études secondaires	7.78%	5.72%	7.16%	6.54%				
	Maîtrise	3.68%	7.65%	2.98%	7.31%				
	Certificat d'études non-universitaire	22.37%	17.21%	7.31%	6.96%				
	Doctorat	0.74%	1.12%	3.42%	6.24%				
	Certificat professionnel	11.19%	9.30%	6.66%	6.82%				
D.E.S.S.	1.16%	3.47%	2.83%	9.71%					



## Chapitre 6

# Analyse économétrique

Nous poursuivons notre analyse des retraits des REER grâce aux données de l'ELIA. Nous examinons d'abord la relation entre retraits et variables explicatives grâce à une régression linéaire et une régression logistique. Nous évaluons ensuite les résultats de modèles avec variable dépendante dichotomique qui évaluent la probabilité pour un individu de retirer des ou faire défaut du RAP. Nous évaluons par la suite les résultats d'un modèle probit ordonné à effets aléatoires. Nous nous penchons finalement sur les résultats du modèle avec variable dépendante continue proposé par Wooldrige (1995) afin d'évaluer l'incidence de certaines variables de revenus et sociodémographiques sur les montants retirés. Au cours des sections 6.2, 6.3 et 6.4, nous utilisons deux formes fonctionnelles distinctes. La seconde forme fonctionnelle nous permet de contrôler pour les effets temporels et évaluent l'incidence des TEMI en seuils.

### 6.1 Régression linéaire et régression logistique

Nous trouvons en annexe, aux tableaux A.2 et A.3, les résultats de régressions linéaires et logistiques de la forme :

$$\text{Retrait} = \beta_a A + \beta_{temi} \text{TEMI} + \beta_y Y + \beta_x X + \beta_{prov} \text{Provinces} + \beta_{annees} \text{Annees} + \epsilon \quad (6.1)$$

où  $A$  est une constante,  $Y$  sont une séries de variables continues représentant le revenu,  $X$  une série de variables à traits sociodémographiques.  $Provinces$  et  $Annees$  sont une série de variable dichotomiques.  $Retrait$  prend une forme dichotomique pour la forme logistique :  $Retrait$  est égal à 1 s'il y a présence d'un retrait et 0 dans le cas contraire.

Pour les formes linéaires et logistiques, nous trouvons une relation positive entre le TEMI et  $Retrait$  qui vient à l'encontre de nos hypothèses initiales. Le choix des modèle économétriques

et des formes fonctionnelles ici présentées détiennent quelques faiblesses qui expliquent potentiellement cette la forme de cette relation. D'abord, une régression linéaire régulière ne permet d'estimer les coefficients en fonction d'effets fixes par individu. Il est nécessaire d'exploiter la forme longitudinale des données pour interpréter les effets des variables comme une différence par rapport à leur moyenne respective par contribuable. La régression par effets fixes serait donc plus appropriée pour capturer l'effet du profil d'imposition temporel.

Un second défaut des régressions présentées réside dans la forme fonctionnelle utilisée. Comme Leblanc (2002) et Mawani et Paquette (2011) le mentionnent, il subsiste une relation non-linéaire entre TEMI et retraits ou montants retirés. Il est ainsi nécessaire de décomposer les TEMI en différents seuils pour capturer cette relation. Nous adressons ces problématiques grâce aux modèles et formes fonctionnelles adéquats dans les sous-sections suivantes.

## 6.2 Modèle de probabilité linéaire à effets fixes

Nous utilisons un modèle de probabilité linéaire à effets fixes pour évaluer les facteurs qui influencent la probabilité d'observer un retrait ou un défaut. Nous utilisons les deux formes fonctionnelles pour évaluer les coefficients d'intérêts pour les trois échantillons *Aucune Distinction*, *Retrait* et *Défaut*.

### Échantillon *Aucune distinction*

À partir des résultats de l'échantillon complet, nous ne sommes pas en mesure de confirmer l'hypothèse que les retraits des REER ou défauts du RAP soient négativement associés aux TEMI et aux changements de TEMI. En prenant comme variable de référence  $0.35 \leq TEMI < 0.4$ , les coefficients des variables dichotomiques qui présentent des seuils supérieurs à 0 et inférieurs à 0.35 sont négatifs et significatif. Les signes de ces coefficients nous indiquent que la probabilité pour les individus de retirer des REER ou faire défaut au RAP diminue en dessous du point de référence. Évidemment, ces résultats contredisent l'hypothèse de Daly (1981) qui stipule que les agents retirent lors des périodes où leurs TEMI sont faibles. Nous nous serions plutôt attendus à observer une relation strictement négative.

Les événements de la vie comme la *Perte d'emploi*, *Séparation* et *Divorce* sont tous associés à une augmentation significative de la probabilité qu'un individu retire du REER ou fasse défaut au RAP.

Les coefficients associés aux revenus confirment l'hypothèse que les individus lissent leurs revenus grâce aux REER. Les contribuables retirent ou ne remboursent leur RAP lors de

périodes de faibles revenus. Le coefficient négatif de la variable *Revenu Inv.* appuie l'hypothèse qu'une diminution des revenus d'investissement est liée à une augmentation de la probabilité qu'un individu débourse du REER.

### Échantillon *Retrait*

En excluant les retraits identifiés comme défauts de l'échantillon, nous sommes incapables d'identifier que les contribuables adoptent un comportement qui leur permettrait de maximiser leur profil d'imposition marginal. Effectivement, les signes des seuils de TEMI présentent plutôt une relation strictement positive entre TEMI et probabilité de retrait. Les résultats que nous obtenons sont similaires sous certains aspects à ceux de Mawani et Paquette (2011). Effectivement, en utilisant des variables dichotomiques pour des seuils de TEMI analogues, les auteurs observent également une augmentation de la probabilité de retrait pour des seuils de TEMI supérieurs à  $0.35 \leq TEMI < 0.4$ . Les auteurs justifient ces résultats en formulant l'hypothèse que les contribuables canadiens sont probablement assujettis à des contraintes de liquidités et/ou ne sont pas conscients de la valeur de leur TEMI pour la période donnée.

La sélection du sous-échantillon nous permet de cerner quelques différences par rapport aux résultats donnés par l'échantillon complet. On observe par exemple que les contribuables qui ont retiré sont susceptibles d'avoir subi une baisse de revenus de dividendes pour les deux formes fonctionnelles.

### Echantillon *Defaut RAP*

Nous observons que les TEMI ne détiennent que peu d'incidence sur les défauts du RAP. Pour les deux formes fonctionnelles, nous constatons qu'aucun des signes associés aux taux marginaux d'imposition ne sont significatifs. En fait, seuls les coefficients de la variable de revenu d'emploi, son carré, l'âge et l'année 2000 détiennent des valeurs  $p$  sous le seuil de 0.05. Ces résultats ne sont pas surprenants étant donné qu'une vaste proportion des contribuables ne remboursent tout simplement jamais leur montant dû. Ce comportement ne peut être capturé par cette forme fonctionnelle ou nos données.

TABLEAU 6.1: Modèle de probabilité linéaire à effet fixe

	Complet		Retraits		Défauts RAP	
	(1)	(2)	(1)	(2)	(1)	(2)
Séparation	<b>0.027*</b> (0.011)	<b>0.028**</b> (0.011)	<b>0.027**</b> (0.01)	<b>0.027**</b> (0.01)	0.004 (0.007)	0.005 (0.007)
Divorce	<b>0.022.</b> (0.012)	<b>0.023.</b> (0.012)	<b>0.021.</b> (0.012)	<b>0.021.</b> (0.012)	0.009 (0.007)	0.009 (0.007)
Veuf	0.029	0.03	0.047	0.047	-0.026	-0.025

	(0.039)	(0.039)	(0.038)	(0.038)	(0.017)	(0.017)
Perte d'emploi	<b>0.015*</b>	<b>0.015*</b>	<b>0.02***</b>	<b>0.02***</b>	-0.003	-0.003
	(0.006)	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)
Assurance emp.	<b>0.0*</b>	<b>0.0*</b>	<b>0.0**</b>	<b>0.0**</b>	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Indemnisation	0.0	0.0	0.0	0.0	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Gain Capital	-0.0	-0.0	-0.0	-0.0	<b>0.0.</b>	<b>0.0.</b>
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Rev. emploi autonome	-0.0	-0.0	<b>-0.0***</b>	<b>-0.0***</b>	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Revenu d'emploi	<b>-0.0***</b>	<b>-0.0***</b>	<b>-0.0***</b>	<b>-0.0***</b>	<b>-0.0***</b>	<b>-0.0***</b>
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
(Revenu d'emploi)2	<b>0.0***</b>	<b>0.0***</b>	<b>0.0***</b>	<b>0.0***</b>	<b>0.0***</b>	<b>0.0***</b>
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Revenu div.	-0.0	-0.0	<b>-0.0.</b>	<b>-0.0.</b>	-0.0	-0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Revenu inv.	<b>-0.0.</b>	<b>-0.0.</b>	-0.0	-0.0	-0.0	-0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
TEMI	<b>0.09***</b>		<b>0.1***</b>		-0.002	
	(0.016)		(0.013)		(0.011)	
TEMI < 0		-0.024		<b>-0.026.</b>		-0.005
		(0.016)		(0.014)		(0.011)
0 <= TEMI < 0.1		<b>-0.037***</b>		<b>-0.034***</b>		-0.005
		(0.007)		(0.006)		(0.005)
0.1 <= TEMI < 0.2		<b>-0.02**</b>		<b>-0.024***</b>		0.003
		(0.007)		(0.006)		(0.005)
0.2 <= TEMI < 0.25		<b>-0.027***</b>		<b>-0.027***</b>		-0.002
		(0.007)		(0.006)		(0.005)
0.25 <= TEMI < 0.3		<b>-0.011*</b>		<b>-0.013*</b>		0.001
		(0.006)		(0.005)		(0.004)
0.3 <= TEMI < 0.35		<b>-0.016***</b>		<b>-0.015***</b>		-0.001
		(0.004)		(0.003)		(0.003)
0.4 <= TEMI < 0.5		0.004		<b>0.007*</b>		-0.002
		(0.004)		(0.003)		(0.003)
0.5 <= TEMI		0.012		<b>0.02**</b>		-0.005
		(0.008)		(0.007)		(0.006)
2000		<b>-0.013*</b>		-0.007		<b>-0.009*</b>

		(0.006)		(0.005)		(0.004)
2001		<b>-0.025**</b>		<b>-0.017**</b>		-0.012
		(0.009)		(0.006)		(0.008)
2002		<b>-0.035**</b>		<b>-0.025**</b>		-0.017
		(0.013)		(0.008)		(0.012)
2003		<b>-0.043*</b>		<b>-0.032**</b>		-0.021
		(0.017)		(0.01)		(0.015)
2004		<b>-0.052*</b>		<b>-0.04**</b>		-0.026
		(0.021)		(0.013)		(0.019)
2005		<b>-0.065*</b>		<b>-0.054***</b>		-0.029
		(0.025)		(0.015)		(0.023)
2006		<b>-0.08**</b>		<b>-0.067***</b>		-0.032
		(0.029)		(0.017)		(0.027)
2007		<b>-0.091**</b>		<b>-0.077***</b>		-0.036
		(0.033)		(0.02)		(0.03)
2008		<b>-0.103**</b>		<b>-0.091***</b>		-0.037
		(0.038)		(0.022)		(0.034)
2009		<b>-0.105*</b>		<b>-0.096***</b>		-0.037
		(0.042)		(0.024)		(0.038)
2010		<b>-0.119*</b>		<b>-0.1***</b>		-0.05
		(0.046)		(0.027)		(0.042)
2011		<b>-0.136**</b>		<b>-0.11***</b>		-0.061
		(0.05)		(0.029)		(0.046)
2012		<b>-0.148**</b>		<b>-0.119***</b>		-0.066
		(0.054)		(0.031)		(0.049)
2013		<b>-0.159**</b>		<b>-0.13***</b>		-0.071
		(0.058)		(0.034)		(0.053)
Âge	<b>0.006***</b>	<b>0.017***</b>	<b>0.002***</b>	<b>0.011***</b>	<b>0.005***</b>	<b>0.009*</b>
	(0.0)	(0.004)	(0.0)	(0.002)	(0.0)	(0.004)
Constante	<b>-0.129***</b>	<b>-0.475**</b>	<b>-0.029*</b>	<b>-0.324***</b>	<b>-0.129***</b>	<b>-0.282*</b>
	(0.019)	(0.145)	(0.014)	(0.085)	(0.015)	(0.133)
$\rho$	0.433	0.514	0.349	0.412	0.559	0.595
$\sigma_e$	0.258	0.258	0.217	0.216	0.17	0.17
$\sigma_u$	0.226	0.266	0.159	0.181	0.191	0.206
F	.	7.67	.	.	0.003	8.57
$R^2$	0.00	0.004	0.003	0.0002	0.003	0.01
N	80700	80700	75900	75800	75800	75900

### 6.3 Probit ordonné à effets aléatoires

Le modèle probit ordonné à effets aléatoires permet d'évaluer les effets marginaux des variables dépendantes sur la probabilité de retirer des REER pour différents groupes de montants de retraits. Les tableaux 6.2 et 6.3 détaillent les résultats des effets marginaux évalués à la moyenne pour l'échantillon complet et pour l'échantillon où l'on exclut les défauts du RAP.

Pour les retraits sous 2000 \$ et pour les échantillons *Aucune distinction* et *Retrait*, les résultats du modèle pointent vers une relation strictement décroissante. En ce qui a trait aux retraits supérieur à 2000 \$, les signes des coefficients pointent plutôt vers une relation strictement positive entre TEMI et probabilité d'observer un retrait. L'hypothèse de Daly est ainsi seulement validée pour le premier groupe de retrait.

Différentes hypothèses pourraient soutenir ces résultats. Il serait possible que les contribuables à faible revenu (TEMI faible) retirent généralement des montants peu élevés, faute de liquidités. Les contribuables à fort revenu (TEMI élevé) seraient également moins sensibles à l'effet prix des TEMI et en mesure de retirer des montants élevés. Nous pourrions également expliquer ces résultats par un phénomène de « sur-épargne » des contribuables. Comme nous ne pouvons observer la date du retrait ni des contributions, nous ne pouvons distinguer si les contribuables retirent suite à une somme annuelle de contributions qui excèdent leur seuil maximal. Les contribuables assujettis à des TEMI élevés seraient plus susceptibles de se trouver dans une pareille situation. Une fois de plus, ces résultats vont à l'encontre des hypothèses qui découlent du modèle théorique utilisé ou en sortent du cadre.

TABLEAU 6.2: Effets marginaux partiels évalués à la moyenne évalués aux groupes de montants : Échantillon Complet

	Groupe				
	X < 2000	2000 < X < 4000	4000 < X < 6000	6000 < X < 8000	8000 < X
TEMI < 0	<b>0.043***</b> (0.012)	<b>-0.024***</b> (0.007)	<b>-0.005***</b> (0.002)	<b>-0.004***</b> (0.001)	<b>-0.003***</b> (0.001)
0 <= TEMI < 0.1	<b>0.049***</b> (0.006)	<b>-0.027***</b> (0.003)	<b>-0.006***</b> (0.001)	<b>-0.005***</b> (0.001)	<b>-0.004***</b> (0.0)
0.1 <= TEMI < 0.2	<b>0.035***</b> (0.006)	<b>-0.019***</b> (0.004)	<b>-0.004***</b> (0.001)	<b>-0.004***</b> (0.001)	<b>-0.003***</b> (0.0)
0.2 <= TEMI < 0.25	<b>0.033***</b> (0.006)	<b>-0.018***</b> (0.003)	<b>-0.004***</b> (0.001)	<b>-0.003***</b> (0.001)	<b>-0.002***</b> (0.0)
0.25 <= TEMI < 0.3	<b>0.022***</b>	<b>-0.012***</b>	<b>-0.003***</b>	<b>-0.002***</b>	<b>-0.002***</b>

	(0.005)	(0.003)	(0.001)	(0.0)	(0.0)
0.3 <= TEMI < 0.35	<b>0.02***</b>	<b>-0.011***</b>	<b>-0.003***</b>	<b>-0.002***</b>	<b>-0.001***</b>
	(0.003)	(0.002)	(0.0)	(0.0)	(0.0)
0.4 <= TEMI < 0.5	<b>-0.008*</b>	<b>0.004*</b>	<b>0.001*</b>	<b>0.001*</b>	<b>0.001*</b>
	(0.003)	(0.002)	(0.0)	(0.0)	(0.0)
0.5 <= TEMI	<b>-0.011.</b>	<b>0.006.</b>	<b>0.001.</b>	<b>0.001.</b>	<b>0.001.</b>
	(0.007)	(0.004)	(0.001)	(0.001)	(0.0)

Signif. codes : 0 : '\*\*\*', 0.001 : '\*\*', 0.01 : '\*' 0.05 : '.', 0.1 : ''

TABLEAU 6.3: Effets marginaux évalués à la moyenne évalués aux groupes de montants : Échantillon *Retrait*

	Groupe				
	X < 2000	2000 < X < 4000	4000 < X < 6000	6000 < X < 8000	8000 < X
TEMI < 0	<b>0.034**</b>	<b>-0.01**</b>	<b>-0.008**</b>	<b>-0.006**</b>	<b>-0.004**</b>
	(0.012)	(0.003)	(0.003)	(0.002)	(0.001)
0 <= TEMI < 0.1	<b>0.035***</b>	<b>-0.01***</b>	<b>-0.008***</b>	<b>-0.006***</b>	<b>-0.004***</b>
	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)
0.1 <= TEMI < 0.2	<b>0.029***</b>	<b>-0.009***</b>	<b>-0.007***</b>	<b>-0.005***</b>	<b>-0.003***</b>
	(0.006)	(0.002)	(0.001)	(0.001)	(0.001)
0.2 <= TEMI < 0.25	<b>0.025***</b>	<b>-0.007***</b>	<b>-0.006***</b>	<b>-0.004***</b>	<b>-0.003***</b>
	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)
0.25 <= TEMI < 0.3	<b>0.014***</b>	<b>-0.004***</b>	<b>-0.003***</b>	<b>-0.003***</b>	<b>-0.002***</b>
	(0.004)	(0.001)	(0.001)	(0.001)	(0.0)
0.3 <= TEMI < 0.35	<b>0.014***</b>	<b>-0.004***</b>	<b>-0.003***</b>	<b>-0.003***</b>	<b>-0.002***</b>
	(0.003)	(0.001)	(0.001)	(0.0)	(0.0)
0.4 <= TEMI < 0.5	<b>-0.007*</b>	<b>0.002*</b>	<b>0.001*</b>	<b>0.001*</b>	<b>0.001*</b>
	(0.003)	(0.001)	(0.001)	(0.0)	(0.0)
0.5 <= TEMI	<b>-0.013*</b>	<b>0.004*</b>	<b>0.003*</b>	<b>0.002*</b>	<b>0.001*</b>
	(0.006)	(0.002)	(0.001)	(0.001)	(0.001)

Signif. codes : 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1

## 6.4 Wooldridge 1995 : modèle à variable dépendante continue

Pour estimer l'incidence de différentes variables sur le montant retiré des REER, nous utilisons une méthode développée par Wooldridge (1995) qui permet de tenir compte d'un potentiel

effet de sélection. Nous posons l'hypothèse que la fonction qui détermine le montant retiré des contribuables est distincte de la fonction qui détermine la décision du retrait. Les spécificités du modèle sont détaillées à la section 3.2.1. En bref, cette méthode consiste à calculer l'inverse du ratio de Mills grâce à l'utilisation d'une régression *probit* pour chacune des années de l'échantillon. Nous jumelons ensuite tous les sous-échantillons annuels pour en former un seul qui contient l'ensemble des années observées. Nous utilisons alors un modèle à effets fixes en incluant l'inverse du ratio de Mills.

Nous utilisons l'équation suivante comme équation de sélection :

$$d\_Retrait = \beta_{sep}Separation + \beta_{div}Divorce + \beta_{veuf}Veuf + \beta_{perte}Perte\_emploi + \beta_{temi_{pos}}\Delta TEMI(> 0.5) + \beta_{temi_{neg}}\Delta TEMI(< -0.5) \quad (6.2)$$

où les variables *Séparations*, *Divorce*, *Divorce*, *Veuf* expriment toutes des changements de statuts matrimoniaux à une période spécifique, où *Perte\_emploi* signale la perte involontaire d'un emploi à une période spécifique et où  $\Delta TEMI$  sont des variables dichotomiques qui indiquent des changements absolus du taux effectif marginal d'imposition de plus ou moins 5 %. Lors de l'étape de régression linéaire à effets fixes, la seconde forme fonctionnelle spécifie les TEMI en seuils et contrôle pour les années de la période étudiée. Le calcul des variations de TEMI génèrent des variables manquantes pour les contribuables qui n'ont pas des observations répétées au travers de l'ensemble des années. Ces observations sont éliminées des sous-échantillons finaux.

### **Échantillon *Aucune distinction***

En ce qui concerne les résultats des TEMI, on note que tout en contrôlant pour le revenu, et avec  $0.35 \leq TEMI < 0.4$  comme variable de référence, il subsiste une relation positive entre les TEMI et le montant retiré.

En ce qui a trait aux variables en lien au Revenu d'emploi, les coefficients des régressions nous permettent de discerner que les Canadiens utilisent les REER pour compenser à leurs baisses de revenus. Pour les deux formes fonctionnelles, les coefficients associée aux niveaux de revenu d'emploi présentent une relation négative avec le montant retiré.

Pour ce qui est de l'inverse du ratio de Mills, les coefficients ne sont pas significatifs dans les deux formes fonctionnelles. Nous rejetons alors l'hypothèse qu'il existe un effet de sélection distinct du processus qui détermine la valeur du montant retiré. Nous constatons qu'il existe

un lien négatif entre le montant retiré et le montant d'assurance emploi reçu à la période courante.

TABLEAU 6.4: Variable dépendante continue : Wooldrige 1995

	Complet		Retraits		Défauts RAP	
	(1)	(2)	(1)	(2)	(1)	(2)
Divorce	311.27 (332.69)	334.1 (332.43)	-180.75 (674.52)	-275.6 (677.16)	<b>854.37***</b> (228.79)	<b>825.33***</b> (236.49)
Séparation	364.63 (249.58)	368.66 (249.35)	135.11 (452.71)	187.77 (454.22)	<b>651.24***</b> (175.98)	<b>647.77***</b> (179.47)
Veuf	-34.576 (1118.2)	13.009 (1117.0)	831.56 (2301.6)	189.79 (2309.5)	-16.125 (694.51)	2.943 (708.03)
Perte d'emploi	192.53 (138.49)	215.69 (139.23)	145.54 (277.21)	174.77 (281.43)	5.237 (92.628)	0.919 (97.215)
Assurance emp.	<b>-0.022.</b> (0.013)	<b>-0.022.</b> (0.013)	-0.043 (0.03)	-0.038 (0.031)	0.005 (0.007)	0.004 (0.008)
Indemnisation	0.021 (0.015)	0.021 (0.015)	-0.037 (0.039)	-0.022 (0.039)	<b>0.021*</b> (0.01)	<b>0.021*</b> (0.01)
Gain Capital	-0.002 (0.018)	-0.003 (0.018)	0.004 (0.031)	0.002 (0.031)	0.0 (0.002)	0.0 (0.002)
Rev. emploi autonome	<b>-0.027**</b> (0.009)	<b>-0.024**</b> (0.009)	<b>-0.093***</b> (0.022)	<b>-0.084***</b> (0.022)	-0.003 (0.003)	-0.002 (0.003)
Revenu d'emploi	<b>-0.021***</b> (0.005)	<b>-0.018***</b> (0.005)	<b>-0.049***</b> (0.011)	<b>-0.039**</b> (0.012)	<b>-0.007*</b> (0.003)	<b>-0.007*</b> (0.003)
(Revenu d'emploi) <sup>2</sup>	<b>0.005*</b> (0.002)	0.003 (0.002)	<b>0.015*</b> (0.006)	<b>0.011.</b> (0.006)	0.002 (0.001)	0.002 (0.002)
Revenu div.	-0.009 (0.007)	-0.01 (0.007)	<b>-0.091*</b> (0.045)	<b>-0.094*</b> (0.045)	-0.001 (0.004)	-0.001 (0.004)
Revenu inv.	-0.091 (0.098)	-0.058 (0.098)	0.203 (0.234)	0.246 (0.235)	-0.028 (0.04)	-0.029 (0.04)
TEMI	<b>3731.0***</b> (572.45)		<b>7287.7***</b> (1274.1)		<b>888.4**</b> (314.88)	
TEMI < 0		-386.5 (541.54)		-569.08 (1589.4)		-182.43 (220.69)
0.0 <= TEMI < 0.1		-187.53 (324.96)		-606.43 (711.21)		<b>-273.76.</b> (163.53)
0.1 <= TEMI < 0.2		-216.07 (383.42)		-280.93 (707.95)		-49.897 (225.3)
0.2 <= TEMI < 0.25		-82.527		522.59		-239.75

		(348.54)		(833.08)		(166.92)
0.25 <= TEMI < 0.3		-236.84		-342.64		-37.388
		(145.68)		(278.42)		(95.582)
0.35 <= TEMI < 0.4		<b>383.28***</b>		<b>574.21**</b>		77.694
		(102.84)		(200.37)		(67.76)
0.4 <= TEMI < 0.5		<b>771.49***</b>		<b>1411.6***</b>		131.22
		(127.93)		(256.04)		(79.769)
TEMI >= 0.5		<b>1507.3***</b>		<b>3073.6***</b>		245.05
		(283.87)		(592.72)		(198.71)
IMR	194.19	195.29	-361.31	-278.83	<b>440.1.</b>	427.67
	(146.42)	(157.09)	(244.56)	(255.91)	(247.99)	(339.21)
2000		-121.64		-560.28		435.31
		(1103.5)		(1972.6)		(859.75)
2001		-12.097		-1306.4		555.21
		(2159.4)		(3868.6)		(1693.7)
2002		-310.02		-2089.0		871.98
		(3224.9)		(5796.4)		(2527.3)
2003		-791.78		-3258.9		981.85
		(4284.0)		(7707.7)		(3356.7)
2004		-1382.4		-4476.5		1281.3
		(5352.9)		(9637.4)		(4193.6)
2005		-1634.8		-5234.3		1401.4
		(6420.4)		(11558.0)		(5031.3)
2006		-1917.9		-6077.4		1548.0
		(7488.0)		(13482.0)		(5868.6)
2007		-2381.5		-7354.7		1779.7
		(8557.7)		(15414.0)		(6706.1)
2008		-2766.6		-8492.2		2125.5
		(9625.6)		(17338.0)		(7544.1)
2009		-2811.1		-8696.5		2324.8
		(10694.0)		(19258.0)		(8382.3)
2010		-3002.6		-9963.6		2523.6
		(11760.0)		(21171.0)		(9219.0)
2011		-3616.7		-11449.0		2723.7
		(12828.0)		(23091.0)		(10056.0)
2012		-3784.6		-12373.0		2994.1
		(13896.0)		(25012.0)		(10894.0)
2013		-4192.1		-13491.0		3124.3

		(14964.0)		(26937.0)		(11732.0)
Âge	<b>49.084***</b>	375.75	<b>99.699***</b>	1084.7	<b>41.836**</b>	-174.86
	(13.367)	(1068.4)	(27.573)	(1921.8)	(12.97)	(837.85)
F-statistic	5.27	3.44	4.77	4.097	3.362	1.65
P-value	1.355e-10	2.2636e-11	4.3314e-09	1.1813e-08	1.153e-05	0.0087976
R <sup>2</sup>	0.02	0.033	0.052	0.077	0.017	0.022
N	6100	6100	3000	3000	4000	4000
Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

### Échantillon *Retraits*

Lorsque nous excluons les défauts du RAP, nous observons des coefficients de signes similaires au modèle *Sans distinction*, sauf pour quelques exceptions. D'abord, nous observons que coefficient du revenu d'emploi au carré est significatif au seuil de 5% pour les deux formes fonctionnelles. Cette dernière relation avait d'ailleurs été mise en évidence par LeBlanc (2002) et Mawani et Paquette (2011), qui affirment qu'une forme non linéaire du revenu était nécessaire pour expliquer les montants retirés. Nous constatons également que les coefficients associés aux revenus de dividendes deviennent significatifs.

### Échantillon *Défauts*

Pour ce sous-échantillon, les coefficients des variables *Divorce* et *Séparation* deviennent significatives. Nous constatons également que les modèles (1) et (2) ont des faibles pouvoirs explicatifs. Ce résultat appuie l'hypothèse « tout ou rien » formulée par Mawani et Paquette : certains contribuables utilisent le RAP pour réunir le montant nécessaire à une mise de fonds et ne réinvestissent jamais la somme retirée en leur REER. Nos données ou formes fonctionnelles ne pourraient refléter ce type de comportement.

## 6.5 Synthèse

### 6.5.1 Modèles à variable dépendante dichotomique

Les résultats des modèles probit à effets aléatoires et de probabilité linéaire montrent généralement une relation positive entre les TEMI et la probabilité d'observer un retrait. Ce lien entre TEMI élevés et retrait est également observé par Mawani et Paquette (2011). Tout comme ces auteurs, nous expliquons la forme de cette relation par de possibles contraintes de

liquidité ou une ignorance générale des taux marginaux d'imposition des contribuables. Cette contrainte de liquidité pourrait entre autres survenir suite à une « sur-épargne » des agents. Effectivement, il est commun que les Canadiens retirent lors de la même année fiscale qu'ils contribuent en leur REER. En notre échantillon, 48.3 % des individus qui retirent contribuent lors de la même année. Cette tendance est aussi observée par Messacar (MESSACAR, 2017). Il est par contre impossible pour nous de discerner si les Canadiens retirent avant ou après la cotisation. Cette information nous permettrait de confirmer ou d'infirmer cette hypothèse. Finalement, nous appuyons les résultats empiriques observés dans la littérature : les baisses de revenus d'emplois ainsi que certains événements de la vie ont une incidence importante sur le retrait.

La séparation des échantillons nous permet d'observer des tendances distinctes pour les défauts du RAP et retraits des REER dans le cas du modèle de probabilité linéaire. Le peu de coefficients significatifs que nous obtenons dans le cas des défauts nous indique que nous ne sommes en mesure d'expliquer le comportement de ces contribuables.

## 6.5.2 Modèles à variable dépendante continue

Le modèle de Wooldridge 1995 affiche une relation positive entre les REER et les TEMI pour l'échantillon *Aucune distinction* et l'échantillon *Retrait*. Cette relation n'est pas observée pour l'échantillon *Défaut*. Comme mentionné auparavant pour les modèles à variable dépendante dichotomique, plusieurs éléments inobservés pourraient expliquer cette relation. Nous observons par contre les relations anticipées entre retraits ou défauts et revenus d'emplois, revenus d'investissement et perte d'emploi.

La distinction des échantillons *Retrait* et *Défaut* permet d'observer des relations spécifiques retraits ou défauts et variables explicatives. Nous observons que les variables en lien aux revenus détiennent une forte incidence pour le premier échantillon. Nous constatons que les variables des événements de la vie comme *Divorce* et *Séparation* deviennent significative dans le cas du groupe *Défaut*.

## Chapitre 7

# Conclusion

Amené par le gouvernement fédéral de Louis St-Laurent en 1957, le régime enregistré d'épargne retraite s'est taillé une place importante dans l'écosystème fiscal canadien au fil du temps. En cours des années 2000, c'était plus du quart des contribuables canadiens qui utilisaient ce régime d'épargne. La popularité du régime d'épargne s'explique par les avantages fiscaux qu'il prodigue jumelé aux faibles contraintes qui sont rattachées à l'épargne investie. Effectivement, d'une part, le REER permet aux contribuables de réduire leur charge fiscale. Tout montant investi au REER permet au contribuable de déduire ce montant à son revenu imposable. Cette loi vise à fournir les incitatifs nécessaires aux Canadiens et Canadiennes à épargner lors de périodes de revenus élevés. D'une autre, le REER est un régime d'épargne flexible. Les contribuables sont libres de retirer de leur épargne à tout moment. Lors du retrait des REER, le montant retiré est ajouté au revenu imposable. Aucun coût supplémentaire n'est ajouté : seul le taux effectif marginal d'imposition (TEMI) donne une mesure du coût associé au retrait. Des montants sont retenus à la source, mais sont redistribués lors du remboursement d'impôt.

Or, il est primordial de comprendre les motifs qui poussent les contribuables à retirer de leur REER. Aiguiser notre compréhension collective de l'utilisation de ce véhicule d'épargne pourrait ultimement permettre aux législateurs d'instaurer des lois accommodées à la réalité des Canadiens et Canadiennes. Ainsi, nous tenons à gagner une compréhension profonde des profils sociodémographiques des contribuables qui retirent des REER et font défaut au Régime d'accession à la propriété. Nous tenons à déterminer si les événements de la vie détiennent une incidence sur les retraits. Nous analysons également comment les niveaux de revenus et de TEMI exercent une influence sur ces décisions. Nous encadrons notre recherche à l'aide d'un modèle théorique édifié par Daly (1981).

Daly, dans son papier *The role of registered retirement savings plans in a life-cycle model*, (1981), modélise les choix d'un consommateur qui distribue ses revenus entre consommation, épargne

en REER et épargne hors-REER. Les équations dérivées de son modèle nous permettent de formuler quatre hypothèses. Nous testons empiriquement celles-ci grâce aux méthodes présentées dans le chapitre 3.

Nous offrons ci-dessous un rappel des quatre hypothèses au lecteur. Nous commentons celles-ci en offrant un récapitulatif des résultats de notre étude. Le tableau 7.1 présente en outre un sommaire de l'ensemble des résultats divulgués.

**Hypothèse 1 :** *les retraits des REER avant l'âge de la retraite sont négativement associés au TEMI*

- Les résultats provenant des différentes méthodes utilisées ne nous permettent pas d'accepter cette hypothèse. Seuls les résultats de l'analyse univariée, qui ne contrôle pour le revenu et autres variables importantes, montrent que les niveaux de TEMI sont plus faibles pour les contribuables qui retirent ou font défaut. Les modèles multivariés probit à effets aléatoires et de probabilité linéaire à effets fixes indiquent que la probabilité de retirer diminue pour les TEMI faibles. Lorsque nous évaluons le retrait en tant que variable dépendante continue avec le modèle de Wooldridge, les résultats pointent plutôt vers une augmentation du montant retiré pour des TEMI élevés.

**Hypothèse 2 :** *les retraits des REER avant l'âge de la retraite sont négativement associés au revenu et aux changements de revenus*

- L'ensemble des résultats empiriques appuie cette hypothèse. La relation entre revenu d'emploi et retrait, formulée en tant que variable dichotomique ou continue, est négative et significative pour des seuils de probabilités élevés.

**Hypothèse 3 :** *les retraits des REER avant l'âge de la retraite sont négativement associés au niveau de revenu d'investissement*

- Quoique les tests statistiques univariés semblent tous pointer vers une corrélation négative entre retraits et montants d'investissements, cette relation semble moins prononcée pour le modèle de probabilité linéaire à effets fixes qui présente certains coefficients non significatifs. Cette relation est d'autant plus atténuée pour le modèle à variable dépendante continue.

**Hypothèse 4 :** *les retraits des REER avant l'âge de la retraite sont positivement associés à certains événements de la vie tels que la perte d'emploi, le divorce ou la séparation*

- Les événements de la vie détiennent une incidence positive sur les retraits des REER. Seul le modèle à variable dépendante continue présente des effets nuls pour les événements de *Perte d'emploi*, *Divorce* et *Séparation* pour l'ensemble des formes fonctionnelles présentées.

Bien que nous ne sommes pas en mesure d'accepter l'ensemble des hypothèses posées, nous obtenons toutefois des résultats similaires à ceux observés dans la littérature. Effectivement, tout comme Mawani et Paquette, nous ne sommes pas en mesure de constater une relation strictement négative entre TEMI et revenus. Tout comme les auteurs, nous observons les signes attendus pour les variables de revenu d'emploi, d'investissement et des événements de la vie.

Sur un autre ordre d'idées, il ne va pas sans dire que notre mémoire apporte des résultats auparavant inexploré dans la littérature économique canadienne sur les retraits des REER avant l'âge de la retraite, entre autres grâce à la panoplie de variables présentes dans l'ELIA. Par exemple, nous observons que les retraits sont plus fréquents chez les contribuables qui ont des niveaux d'éducatifs *Certificat d'apprenti* ou *Diplôme cegep* alors qu'ils sont moins fréquent chez les détenteur de diplôme de cycles supérieurs. Nous observons également certaines tendances spécifiques aux provinces en fonction de leur taux moyen de TEMI au travers des années 1999 à 2013. Ces résultats pourront sans doute inspirer quelconques recherches futures.

De surcroît, une importante particularité de notre étude réside dans l'utilisation de techniques d'apprentissages statistiques dans une optique de « prédiction au service de l'estimation », terme amené par Mullainathan et Spiess (2017). En effet, nous développons un algorithme « Gradient Boosting » à partir des données du DAL pour nous permettre de différencier les retraits des REER des défauts du RAP dans la base de données de l'ELIA. Les fortes performances de cet algorithme sur un échantillon test de la base de données du DAL nous permettent d'extrapoler ces résultats sur notre échantillon de l'ELIA. Nous sommes ainsi en mesure de former des sous-échantillons libres de retraits des REER ou défaut du RAP en l'ELIA.

Somme toute, les relations entre TEMI et retraits nous portent à croire que certaines informations supplémentaires pourraient nous permettre d'expliquer cette relation avec plus de précision. Comme mentionné au chapitre 5, il est impossible pour nous de distinguer si les contribuables retirent suite à une contribution en leur REER. Cette information nous permettrait de distinguer entre contribuables qui retirent pour lisser leur revenu et contribuable qui ont « sur-épargner ». Par ailleurs, les résultats que nous obtenons par rapport aux niveaux d'éducatifs et retraits nous pointent vers une problématique importante. Effectivement,

dans de recherches ultérieures, il serait intéressant de comprendre comment l'éducation peut agir en tant que variable d'approximation pour le niveau de littératie financière. Il serait également pertinent de comprendre comment les contribuables sont en mesure d'estimer leur propre TEMI en fonction de ce niveau de littératie. Ces informations permettraient entre autres de déterminer si la relation non négative observée dans notre étude pourrait être expliquée par une simple ignorance des TEMI des contribuables.

TABLEAU 7.1: Sommaire des résultats

Échantillon	Hypothèse	Variable	Analyse univariée	Modèle probabilité linéaire	Probit ordonné à effets aléatoires	Wooldridge
Aucune Distinction	H1	TEMI	Négative	Positive	Négative/ Positive	Positive
	H2	Revenu d'emploi	Négative	Négative	N/A	Négative
	H3	Dividendes	Négative	0	N/A	0
		Investissement	Négative	Négative	N/A	0
	H4	Perte emploi	Positive	Positive	N/A	0
Divorce Séparation		Positive Positive	Positive Positive	N/A N/A	0 0	
Retrait	H1	TEMI	Positive	Positive	Positive	Positive
	H2	Revenu d'emploi	Négative	Négative	N/A	Négative
	H3	Dividendes	Négative	Négative	N/A	Négative
		Investissement	Négative	0	N/A	0
	H4	Perte emploi	Positive	Positive	N/A	0
Divorce Séparation		Positive Positive	Positive Positive	N/A N/A	0 0	
Défaut	H1	TEMI	0	0	N/A	Positive
	H2	Revenu d'emploi	Négative	Négative	N/A	Négative
	H3	Dividendes	0	0	N/A	0
		Investissement	Négative	0	N/A	0
	H4	Perte emploi	Positive	0	N/A	0
Divorce Séparation		0 0	0 0	N/A N/A	Positive Positive	

N/A : Non-applicable

Positive : Relation positive

**Négative** : Relation négative

**(1)** : Une seule forme fonctionnelle présente un coefficient significatif

**0** : Relation non-significative au seuil de 10%



# Bibliographie

- ARULAMPALAM, Wiji (1999). « A note on estimated coefficients in random effects probit models ». Dans : *Oxford Bulletin of Economics and Statistics* 61.4, p. 597–602.
- BLANCQUAERT, Arnaud et al. (2015). *Quel impact la fiscalité québécoise at-elle sur les incitations au travail ?* Chaire de recherche Industrielle Alliance sur les enjeux économiques des changements démographiques.
- DALY, Michael J (1981). « The role of registered retirement savings plans in a life-cycle model ». Dans : *Canadian Journal of Economics*, p. 409–421.
- FEIGENBAUM, James J (2016). « A machine learning approach to census record linking ». Dans : *Retrieved March 28*, p. 2016.
- FRIEDMAN, Jerome H (2001). « Greedy function approximation : a gradient boosting machine ». Dans : *Annals of statistics*, p. 1189–1232.
- GIBBONS, Robert D et Donald HEDEKER (1994). « Application of random-effects probit regression models. » Dans : *Journal of consulting and clinical psychology* 62.2, p. 285.
- GILES, Philip et Karen MASER (2004). « Using RRSPs before retirement ». Dans : *Canadian Economic Observer*.
- HENDERSON, J Vernon, Adam STOREYGARD et David N WEIL (2012). « Measuring economic growth from outer space ». Dans : *American economic review* 102.2, p. 994–1028.
- KLEINBERG, Jon et al. (2015). « Prediction policy problems ». Dans : *American Economic Review* 105.5, p. 491–95.
- KLEVEN, Henrik J et Mazhar WASEEM (2013). « Using notches to uncover optimization frictions and structural elasticities : Theory and evidence from Pakistan ». Dans : *The Quarterly Journal of Economics* 128.2, p. 669–723.
- KÜCHENHOFF, Helmut, Samuel M MWALILI et Emmanuel LESAFFRE (2006). « A general method for dealing with misclassification in regression : The misclassification SIMEX ». Dans : *Biometrics* 62.1, p. 85–96.
- LEBLANC, Pierre (2002). « RRSPs and pre-retirement earnings replacement ». Dans : *Chapter in PhD dissertation, "Essays on Tax-Deferred Savings in Canada."* Harvard University, Cambridge, MA.

- MAWANI, A et S PAQUETTE (2011). « Pre-retirement RRSP withdrawals ». Dans : *Canadian Tax Journal* 59.2, p. 183–219.
- MESSACAR, Derek (2016). « Three Essays on the Behavioural Economics of Saving for Retirement ». Thèse de doct.
- (2017). « Trends in RRSP Contributions and Pre-retirement Withdrawals, 2000 to 2013 ». Dans : *Statistics Canada*.
- MILLIGAN, Kevin (2002). « Tax-preferred savings accounts and marginal tax rates : evidence on RRSP participation ». Dans : *Canadian Journal of Economics/Revue canadienne d'économie* 35.3, p. 436–456.
- (2013). « Canadian Tax and Credit Simulator, CTaCS ». Dans : *Version 1*, p. 1980–2005.
- MULLAINATHAN, Sendhil et Jann SPIESS (2017). « Machine learning : an applied econometric approach ». Dans : *Journal of Economic Perspectives* 31.2, p. 87–106.
- RAGAN, Christopher (1994). « Progressive income taxes and the substitution effect of RRSPs ». Dans : *Canadian Journal of Economics*, p. 43–57.
- SAEZ, Emmanuel (2010). « Do taxpayers bunch at kink points ? » Dans : *American economic Journal : economic policy* 2.3, p. 180–212.
- VEALL, Michael R (2001). « Did tax flattening affect RRSP contributions ? » Dans : *Canadian Journal of Economics/Revue canadienne d'économie* 34.1, p. 120–131.
- YANG, Mochen et al. (2018). « Mind the gap : Accounting for measurement error and misclassification in variables generated via data mining ». Dans : *Information Systems Research* 29.1, p. 4–24.

## Annexe A

# Taux effectif d'imposition marginal

### A.1 Calcul du TEMI

Nous présentons ci-dessous les variables utilisées pour le calculs des TEMI. La colonne « Obligatoire » du tableau A.1 nous renseigne si la variable est indispensable pour le CTACS dans son calcul. La colonne « Valeur assumée » nous renseigne sur la valeur que le CTACS octroie aux contribuables si la variable est manquante dans la base de donnée. Finalement, la colonne « Présente ELIA » nous indique si cette variable est présente dans notre base de donnée. Nous utilisons l'ensemble des variables présentes dans l'ELIA comme intrants au calculateur CTACS.

TABLEAU A.1: Présentation des variables utilisées par CTACS pour le calcul du TEMI

Description	Obligatoire	Valeur assumée	Présente ELIA
Âge	Non	50	Oui
Personne soignante	Non	0	Oui
Revenu du soignant	Non	0	Oui
Gain en capitaux	Non	0	Oui
Revenu RPC/RPQ	Non	0	Oui
Dépenses de garderie	Non	0	Oui
Statut handicap	Non	0	Oui
Cotisations syndicales	Non	0	Oui
Revenues en dividendes	Non	0	Oui
Revenu d'emploi	Non	0	Oui
Revenu en intérêts	Non	0	Oui
Âge du plus jeune enfant	Non	Aucun enfant	Oui
Âge de deuxième plus jeune enfant	Non	Aucun enfant	Oui

Mâle	Non	homme	Oui
Status matrimonial	Non	Célibataire	Oui
Dépenses médicales	Non	0	Oui
Autres crédits provinciaux remboursables	Non	0	Oui
Autres sources de revenu	Non	0	Oui
Province de résidence	Oui		Oui
Contribution aux RPA	Non	0	Oui
Contributions aux REER	Non	0	Oui
Revenu des REER	Non	0	Oui
Revenu d'assistance sociale	Non	0	Oui
Revenu d'employeur autoNonme	Non	0	Oui
Frais de scolarités	Non	0	Oui
Revenu d'assurance emploi	Non	0	Oui
Compensation au travailleur	Non	0	Oui
Année	Oui		Oui
Pension, revenu des RPA	Non	0	Oui
Déductions additionnelles avant le revenu	Non	0	Non
Dépenses d'adoptions	Non	0	Non
Dépenses qualifiées en art pour les enfants	Non	0	Non
Dépenses qualifiées en sport pour les enfants	Non	0	Non
Dépenses médicales d'un dépendant	Non	0	Non
Dons à la charité	Non	0	Non
Revenu en dividendes Non-éligible	Non	0	Non
Revenu d'une personne déclarée équivalente à l'épouse	Non	0	Non
Autres dépendents	Non	0	Non
Supplément de revenu garanti	Non	0	Non
Personnes à charge	Non	0	Non
Revenu de la personne à charge	Non	0	Non
Âge de troisième plus jeune enfant	Non	Aucun enfant	Non
Âge de quatrième plus jeune enfant	Non	Aucun enfant	Non
Âge de cinquième plus jeune enfant	Non	Aucun enfant	Non
Âge de sixième plus jeune enfant	Non	Aucun enfant	Non
Crédits transférés du retour d'un enfant	Non	0	Non
Achat d'une maison l'année passée	Non	0	Non
Portion de l'année passée en territoires nordiques	Non	0	Non

Autres crédits fédéraux non-remboursables	Non	0	Non
Autres crédits provinciaux non-remboursables -ligne 428	Non	0	Non
Autres déduction avant revenu imposable	Non	0	Non
Nombre de mois étudiant à temps partiel	Non	0	Non
Nombres de mois étudiant temps-plein	Non	0	Non
Contributions politiques fédérales	Non	0	Non
Contributions politiques provinciales	Non	0	Non
Taxe de propriété pour crédit provincial	Non	0	Non
Dépenses en transport publique qualifiées	Non	0	Non
Éligibilité au plan d'assurances médicament du Québec	Non	0	Non
Québec autres déductions avant revenu imposable	Non	0	Non
Éligibilité aux programmes de résidents du Nord du Québec	Non	0	Non
Québec - autres déductions avant revenu imposable	Non	0	Non
Dépenses de rénovation	Non	0	Non
Paiements de loyer pour les crédits d'impôt foncier	Non	0	Non
Revenu de bourses	Non	0	Non
Intérêts sur prêts étudiants	Non	0	Non
Dépenses en fournitures d'enseignement	Non	0	Non
Pompier volontaire	Non	0	Non

## A.2 Régressions linéaires et logistiques

TABLEAU A.2: Régression linéaire : Retrait en fonctions des TEMI, revenus, provinces, années et autres variables sociodémographiques

Variable	Coefficient
Constante	<b>55.67***</b> (40.88)
TEMI	<b>495.8***</b> (44.12)
Rev. d'assistance sociale	<b>-0.011***</b> (0.008)
Frais de scolarité	<b>0.014***</b> (0.006)
Assurance emp.	<b>0.004***</b>

	(0.001)
Indemnisation travail	<b>0.008***</b>
	(0.002)
Contribution au RCP	<b>0.017***</b>
	(0.005)
Gain Capital	<b>-0.0***</b>
	(0.0)
Célibataire	<b>-10.43***</b>
	(17.45)
Marrié	<b>-71.42***</b>
	(11.45)
Montant au soignant	<b>0.012***</b>
	(0.008)
Rev. de prestations pour enfants	<b>0.005***</b>
	(0.004)
Mâle	<b>39.45***</b>
	(9.286)
Immigrant	<b>-15.32***</b>
	(11.68)
Rev. emploi autonome	<b>-0.002***</b>
	(0.0)
Revenu d'emploi	<b>-0.001***</b>
	(0.0)
(Revenu d'emploi) <sup>2</sup>	<b>0.0***</b>
	(0.0)
Revenu div.	<b>-0.0***</b>
	(0.0)
Revenu inv.	<b>-0.005***</b>
	(0.001)
Colombie Britannique	<b>36.2***</b>
	(17.88)
I-P-E	<b>20.64***</b>
	(31.29)
Manitoba	<b>-19.62***</b>
	(19.74)
Nouvelle-Écosse	<b>13.98***</b>
	(21.18)
Nouveau-Brunswick	<b>-7.469***</b>

	(23.05)
Ontario	<b>-1.971***</b>
	(15.15)
Québec	<b>-147.9***</b>
	(16.16)
T-N-L	<b>-0.131***</b>
	(24.37)
Saskatchewan	<b>-6.7***</b>
	(20.69)
2000	<b>17.01***</b>
	(22.57)
2001	<b>17.03***</b>
	(22.56)
2002	<b>31.37***</b>
	(22.62)
2003	<b>27.98***</b>
	(22.6)
2004	<b>49.95***</b>
	(22.64)
2005	<b>32.35***</b>
	(22.73)
2006	<b>25.39***</b>
	(22.85)
2007	<b>38.17***</b>
	(23.03)
2008	<b>23.38***</b>
	(23.2)
2009	<b>83.02***</b>
	(23.46)
2010	<b>81.97***</b>
	(23.7)
2011	<b>68.07***</b>
	(23.98)
2012	<b>79.12***</b>
	(24.34)
2013	<b>87.72***</b>
	(24.74)
Âge	<b>3.381***</b>

	(0.802)
F-statistic	12.71
P-value	2.22e-16
R-Squared	0.0066
N	80700

TABLEAU A.3: Régression logistique : Événement du retrait en fonctions des TEMI, revenus, provinces, années et autres variables sociodémographiques

Variable	Coefficient
Constante	<b>-1.861***</b> (0.107)
TEMI	<b>1.484***</b> (0.125) (0.029)
Rev. emploi autonome	<b>-0.0***</b> (0.0)
Revenu d'emploi	<b>-0.0***</b> (0.0)
(Revenu d'emploi) <sup>2</sup>	<b>0.0***</b> (0.0)
Revenu div.	<b>-0.0***</b> (0.0)
Revenu inv.	<b>-0.0***</b> (0.0)
Rev. de prestations pour enfants	<b>0.0***</b> (0.0)
Revenu d'assistance sociale	<b>-0.0***</b> (0.0)
Frais de scolarité	<b>0.0***</b> (0.0)
Assurance emp.	<b>0.0***</b> (0.0)
Indemnisation travail	<b>0.0***</b> (0.0)
Contribution au RCP	<b>0.0***</b> (0.0)
Gain Capital	<b>-0.0***</b>

	(0.0)
Montant au soignant	<b>0.0***</b>
	(0.0)
Célibataire	<b>0.125***</b>
	(0.042)(nan)
Marrié	<b>-0.162***</b>
Mâle	<b>0.231***</b>
	(0.024)
Immigrant	<b>-0.025***</b>
	(0.03)
Colombie Britannique	<b>0.09***</b>
	(0.046)
I-P-E	<b>0.03***</b>
	(0.078)
Manitoba	<b>-0.037***</b>
	(0.051)
Nouvelle-Écosse	<b>0.091***</b>
	(0.053)
Nouveau-brunswick	<b>-0.15***</b>
	(0.061)
Ontario	<b>0.043***</b>
	(0.039)
Québec	<b>-0.263***</b>
	(0.043)
Saskatchewan	<b>-0.218***</b>
	(0.056)
T-N-L	<b>-0.124***</b>
	(0.063)
Âge	<b>-0.018***</b>
	(0.002)
2000	<b>0.087***</b>
	(0.069)
2001	<b>0.158***</b>
	(0.068)
2002	<b>0.241***</b>
	(0.067)
2003	<b>0.358***</b>
	(0.066)

2004	<b>0.463***</b> (0.065)
2005	<b>0.524***</b> (0.065)
2006	<b>0.571***</b> (0.065)
2007	<b>0.665***</b> (0.064)
2008	<b>0.722***</b> (0.064)
2009	<b>0.834***</b> (0.064)
2010	<b>0.858***</b> (0.064)
2011	<b>0.859***</b> (0.065)
2012	<b>0.925***</b> (0.065)
2013	<b>0.958***</b> (0.066)
N	80700

### A.3 Résultats des modèles multivariés avec forme fonctionnelle sans revenu

TABLEAU A.4: Modèles de probabilité linéaire : spécification sans revenu

Variable	Complet	Retrait	Défauts
Constante	<b>-0.481**</b> (0.145)	<b>-0.324***</b> (0.085)	<b>-0.288*</b> (0.133)
Séparation	<b>0.028**</b> (0.011)	<b>0.027**</b> (0.01)	0.005 (0.007)
Divorce	<b>0.022.</b> (0.012)	<b>0.021.</b> (0.012)	0.009 (0.007)
Veuf	0.031 (0.039)	0.048 (0.038)	-0.025 (0.017)
Perte d'emploi	<b>0.024***</b>	<b>0.028***</b>	-0.001

	(0.005)	(0.005)	(0.004)
TEIM < 0	-0.012	-0.018	0.001
	(0.016)	(0.014)	(0.011)
0 <= TEIM < 0.1	<b>-0.025***</b>	<b>-0.026***</b>	0.0
	(0.007)	(0.006)	(0.005)
0.1 <=TEIM < 0.2	-0.012	<b>-0.018**</b>	0.007
	(0.007)	(0.006)	(0.005)
0.2 <= TEIM < 0.25	<b>-0.02**</b>	<b>-0.022***</b>	0.001
	(0.007)	(0.006)	(0.005)
0.25 <= TEIM < 0.3	-0.005	<b>-0.009.</b>	0.004
	(0.006)	(0.005)	(0.004)
0.3 <= TEIM < 0.35	<b>-0.013***</b>	<b>-0.014***</b>	0.0
	(0.004)	(0.003)	(0.003)
0.4 <= TEIM < 0.5	0.003	<b>0.006.</b>	-0.003
	(0.004)	(0.003)	(0.003)
0.5 <= TEIM	<b>0.015.</b>	<b>0.022**</b>	-0.004
	(0.008)	(0.007)	(0.006)
2000	<b>-0.013*</b>	-0.007	<b>-0.009*</b>
	(0.006)	(0.005)	(0.004)
2001	<b>-0.025**</b>	<b>-0.018**</b>	-0.013
	(0.009)	(0.006)	(0.008)
2002	<b>-0.035**</b>	<b>-0.025**</b>	-0.018
	(0.013)	(0.008)	(0.012)
2003	<b>-0.043*</b>	<b>-0.032**</b>	-0.022
	(0.017)	(0.01)	(0.015)
2004	<b>-0.053*</b>	<b>-0.04**</b>	-0.027
	(0.021)	(0.013)	(0.019)
2005	<b>-0.066**</b>	<b>-0.054***</b>	-0.03
	(0.025)	(0.015)	(0.023)
2006	<b>-0.082**</b>	<b>-0.067***</b>	-0.033
	(0.029)	(0.017)	(0.027)
2007	<b>-0.093**</b>	<b>-0.078***</b>	-0.038
	(0.033)	(0.019)	(0.03)
2008	<b>-0.105**</b>	<b>-0.092***</b>	-0.039
	(0.038)	(0.022)	(0.034)
2009	<b>-0.108*</b>	<b>-0.097***</b>	-0.039
	(0.042)	(0.024)	(0.038)
2010	<b>-0.121**</b>	<b>-0.1***</b>	-0.053

	(0.046)	(0.027)	(0.042)
2011	<b>-0.139**</b>	<b>-0.111***</b>	-0.064
	(0.05)	(0.029)	(0.046)
2012	<b>-0.152**</b>	<b>-0.121***</b>	-0.069
	(0.054)	(0.031)	(0.049)
2013	<b>-0.163**</b>	<b>-0.132***</b>	-0.074
	(0.058)	(0.034)	(0.053)
Âge	<b>0.016***</b>	<b>0.011***</b>	<b>0.009*</b>
	(0.004)	(0.002)	(0.004)
$\rho$	<b>0.515***</b>	<b>0.413***</b>	<b>0.596***</b>
	(0.0)	(0.0)	(0.0)
$\sigma_e$	<b>0.259***</b>	<b>0.217***</b>	<b>0.17***</b>
	(0.0)	(0.0)	(0.0)
$\sigma_u$	<b>0.266***</b>	<b>0.182***</b>	<b>0.206***</b>
	(0.0)	(0.0)	(0.0)
F	8.25	5.56	7.77
$R^2$	0.01	0	0.03
N	80700	75900	75800

TABLEAU A.5: Modèle à variable dépendante continue (Wooldrige 1995) : spécification sans revenu

Variable	Complet	Retrait	Défauts
Séparation	367.195 (250.16)	103.083 (456.428)	<b>643.339***</b> (179.592)
Divorce	369.087 (333.673)	-74.779 (681.016)	<b>796.149***</b> (236.706)
Perte d'emploi	<b>288.936*</b> (127.967)	382.829 (253.786)	59.62 (90.094)
Veuf	17.231 (1120.054)	24.146 (2328.182)	94.292 (708.196)
TEMI < 0	191.276 (528.719)	488.19 (1576.411)	-10.859 (212.32)
0 <= TEMI < 0.1	360.144 (306.236)	297.71 (673.86)	-77.117 (151.338)
0.1 <= TEMI < 0.2	106.714 (374.526)	533.571 (673.224)	121.441 (220.445)
0.2 <= TEMI < 0.25	131.175	1149.969	-113.628

	(345.803)	(823.094)	(162.509)
0.25 <= TEMI < 0.3	-156.548 (144.165)	-208.65 (275.945)	22.373 (93.849)
0.35 <= TEMI < 0.4	<b>315.109**</b> (102.582)	<b>485.094*</b> (200.288)	36.732 (66.81)
0.4 <= TEMI < 0.5	<b>630.081***</b> (125.967)	<b>1233.579***</b> (254.78)	73.373 (77.2)
TEMI < 0.5	<b>1471.231***</b> (278.547)	<b>2968.883***</b> (581.071)	230.867 (198.769)
Âge	213.478 (1071.94)	195.423 (1928.261)	-196.729 (839.145)
IMR	210.108 (157.669)	-236.372 (257.338)	411.72 (339.475)
2000	12.935 (1107.156)	315.87 (1978.561)	451.085 (861.077)
2001	286.878 (2166.544)	523.715 (3879.854)	579.465 (1696.319)
2002	117.696 (3235.606)	594.19 (5814.093)	905.796 (2531.178)
2003	-195.999 (4298.33)	258.194 (7733.684)	1050.638 (3361.894)
2004	-637.972 (5370.85)	-80.021 (9669.691)	1355.622 (4200.083)
2005	-730.318 (6441.901)	-1.601 (11598.316)	1492.92 (5039.118)
2006	-876.394 (7513.117)	66.584 (13527.089)	1654.072 (5877.723)
2007	-1217.576 (8586.42)	-372.545 (15466.861)	1889.342 (6716.573)
2008	-1470.469 (9657.937)	-691.039 (17397.092)	2262.506 (7555.87)
2009	-1356.835 (10729.781)	-47.072 (19323.989)	2472.577 (8395.363)
2010	-1384.226 (11799.584)	-380.113 (21242.872)	2699.306 (9233.379)
2011	-1840.217 (12871.182)	-1023.292 (23169.896)	2916.95 (10072.072)
2012	-1926.726	-1184.054	3176.882

	(13942.901)	(25098.175)	(10911.002)
2013	-2137.576	-1295.524	3338.322
	(15014.753)	(27027.977)	(11750.116)
F statistic	3.03	2.71	1.5
p-value	1.1813e-08	4.0455e-06	0.04
R <sup>2</sup>	0.02	0.06	0.01
N	6100	3000	4000

## A.4 Résultats des régressions probits à effets aléatoires

TABLEAU A.6: Résultats des régressions probits à effets aléatoires

Coefficient	1	2
Séparation	<b>0.256***</b>	<b>0.279***</b>
	(0.06)	(0.066)
Divorce	<b>0.2**</b>	<b>0.233**</b>
	(0.075)	(0.086)
Veuf	<b>0.434.</b>	<b>0.536*</b>
	(0.236)	(0.224)
TEIM < 0	<b>-0.335***</b>	<b>-0.364**</b>
	(0.095)	(0.125)
0 <= TEIM < 0.1	<b>-0.379***</b>	<b>-0.374***</b>
	(0.048)	(0.054)
0.1 <= TEIM < 0.2	<b>-0.273***</b>	<b>-0.31***</b>
	(0.05)	(0.061)
0.2 <= TEIM < 0.25	<b>-0.253***</b>	<b>-0.269***</b>
	(0.045)	(0.052)
0.25 <= TEIM < 0.3	<b>-0.172***</b>	<b>-0.154***</b>
	(0.037)	(0.042)
0.3 <= TEIM < 0.35	<b>-0.155***</b>	<b>-0.151***</b>
	(0.025)	(0.029)
0.4 <= TEIM < 0.5	<b>0.06*</b>	<b>0.069*</b>
	(0.026)	(0.029)
0.5 <= TEIM	<b>0.087.</b>	<b>0.138*</b>
	(0.052)	(0.06)
Assurance emp.	0.0	0.0
	(0.0)	(0.0)
Compensation du travailleur	<b>0.0.</b>	<b>0.0.</b>

	(0.0)	(0.0)
Perte d'emploi	<b>0.142***</b>	<b>0.193***</b>
	(0.034)	(0.039)
Rev. emploi autonome	<b>-0.0**</b>	<b>-0.0***</b>
	(0.0)	(0.0)
Revenu d'emploi	<b>-0.0***</b>	<b>-0.0***</b>
	(0.0)	(0.0)
(Revenu d'emploi) <sup>2</sup>	<b>0.0***</b>	<b>0.0***</b>
	(0.0)	(0.0)
Revenu div.	<b>-0.0**</b>	<b>-0.0***</b>
	(0.0)	(0.0)
Revenu inv.	<b>-0.0**</b>	<b>-0.0*</b>
	(0.0)	(0.0)
Âge	<b>-0.007**</b>	<b>0.008***</b>
	(0.002)	(0.002)
2000	0.037	0.032
	(0.042)	(0.042)
2001	0.068	0.015
	(0.043)	(0.044)
2002	<b>0.133**</b>	0.036
	(0.044)	(0.045)
2003	<b>0.187***</b>	0.055
	(0.044)	(0.046)
2004	<b>0.241***</b>	<b>0.088.</b>
	(0.044)	(0.046)
2005	<b>0.279***</b>	0.076
	(0.044)	(0.047)
2006	<b>0.284***</b>	0.046
	(0.045)	(0.048)
2007	<b>0.335***</b>	0.061
	(0.045)	(0.049)
2008	<b>0.357***</b>	0.008
	(0.045)	(0.051)
2009	<b>0.476***</b>	<b>0.106*</b>
	(0.046)	(0.051)
2010	<b>0.498***</b>	<b>0.157**</b>
	(0.047)	(0.051)
2011	<b>0.494***</b>	<b>0.127*</b>

	(0.047)	(0.051)
2012	<b>0.549***</b>	<b>0.16**</b>
	(0.048)	(0.053)
2013	<b>0.596***</b>	<b>0.178**</b>
	(0.049)	(0.055)
cut1	1.44	2.054
	(0.097)	(0.103)
cut2	2.247	2.349
	(0.098)	(0.103)
cut3	2.516	2.635
	(0.099)	(0.104)
cut4	2.801	2.941
	(0.1)	(0.106)
cut5	3.077	3.226
	(0.101)	(0.107)
$\sigma_u^2$	1.187	0.794
	(0.041)	(0.035)
$\chi^2$	478.99	309.12
N	80700	75900