

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

POUVOIR PRÉDICTIF DES RÉSIDUS GRANULAIRES DANS DES
MODÈLES MACROÉCONOMIQUES

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉCONOMIQUE

PAR
JEAN BEAUDET

JUIN 2019

TABLE DES MATIÈRES

TABLE DES FIGURES	iv
LISTE DES TABLEAUX	v
RÉSUMÉ	vi
INTRODUCTION	1
CHAPITRE I	
CONSTRUCTION DE L'INDICE GRANULAIRE	9
1.1 Analyse de la base de données de Compustat	9
1.1.1 Types d'entreprises	9
1.1.2 Localisation et secteurs d'activités	12
1.2 Indice granulaire	13
1.2.1 Plafond aux chocs idiosyncratiques	17
1.2.2 Les indices granulaires américains et canadiens	20
1.2.3 Secteurs composant les bases de données et les indices granulaires	24
1.2.4 Comparaison de l'indice granulaire obtenu avec celui de Gabaix .	25
1.2.5 Autres données utilisées	26
CHAPITRE II	
POUVOIR PRÉDICTIF DE L'INDICE GRANULAIRE	28
2.1 Les modèles de prédiction économique	28
2.1.1 Modèle AR(p)	29
2.1.2 Modèle ADL avec indice granulaire	30
2.1.3 Modèle ADL avec une variable supplémentaire	31
2.1.4 Exercice de prévision hors échantillon	32
2.2 Résultats des prévisions	33
2.2.1 Résultats avec les données américaines	33
2.2.2 Résultats avec les données canadiennes	36

2.3	Limitations du modèle	38
2.3.1	Performance hors échantillon	38
2.3.2	Période de prévision	40
	CONCLUSION	43
	RÉFÉRENCES	45

TABLE DES FIGURES

Figure		Page
1.1	Comparaison de deux indices de granularité calculés avec deux différentes valeurs moyennes de productivité	16
1.2	Comparaison de trois indices granulaires calculés avec des plafonds et des planchers à 50%, 20% et 10% et une situation sans limites . . .	18
1.3	Comparaison de l'indice granulaire américain avec l'indice canadien .	21
1.4	Comparaison de l'indice granulaire américain par industrie avec l'indice granulaire canadien par industrie	22
2.1	Illustration comparative du taux de croissance du PIB réel américain avec trois modèles de prévision sur un horizon d'un an	35
2.2	Illustration comparative du taux de croissance du PIB réel américain avec trois modèles de prévision sur un horizon de deux ans	36
2.3	Illustration comparative du taux de croissance du PIB réel canadien avec trois modèles de prévision sur un horizon d'un an	38
2.4	Illustration comparative du taux de croissance du PIB réel canadien avec trois modèles de prévision sur un horizon de deux ans	39

LISTE DES TABLEAUX

Tableau		Page
1.1	Taux de corrélation entre des indices granulaires américains et canadiens avec différentes limites et les taux de croissance de leur PIB respectif	19
1.2	Taux de corrélation entre les différents indices granulaires américains et canadiens	23
1.3	Taux de corrélation entre les différents indices granulaires américains et canadiens et les taux de croissance des PIB	24
1.4	Variance des indices granulaires	24
1.5	Répartition en pourcentage par secteur des entreprises	27
2.1	Comparaison entre différents modèles de prévision du taux de croissance du PIB réel américain et un modèle étalon autorégressif de 1996 à 2015	34
2.2	Comparaison entre différents modèles de prévision du taux de croissance du PIB réel canadien et un modèle étalon autorégressif de 1996 à 2015	37
2.3	Comparaison avec différentes années de départ entre des modèles de prévision du PIB réel canadien et un modèle étalon autorégressif . . .	41

RÉSUMÉ

Les modèles de prévisions économiques se basent majoritairement sur des données macroéconomiques et financières. Le présent travail cherche à construire et à tester un indicateur microéconomique dans des modèles de prédictions. Les économies en général sont composées d'une multitude de petites entreprises et d'un nombre restreint de très grandes entreprises. Par exemple, les 500 plus grandes entreprises contribuaient à près de 73% du PIB américain en 2013. Des chocs spécifiques à ces grandes entreprises peuvent avoir des effets sur les agrégats économiques tels que l'emploi, la production industrielle, les exportations ou le PIB. En se basant sur les travaux de Gabaix (2011) qui consistent à isoler ces chocs spécifiques et à produire un indice granulaire, ce travail calcule un indice granulaire américain et étend la recherche à l'économie canadienne en construisant un indice granulaire canadien. Les indices granulaires sont ensuite utilisés dans des modèles de prédictions macroéconomiques de l'économie américaine et canadienne. Dans un contexte de prévision hors échantillon, le pouvoir de prévision des indices granulaires est faible. Il est toutefois un peu plus élevé pour l'économie canadienne que pour l'économie américaine.

Mots clés : indice granulaire, prévision hors échantillon, prévision du PIB, indicateur microéconomique, Gabaix.

INTRODUCTION

Les économistes sont constamment à la recherche de variables économiques ayant un pouvoir prédictif sur l'économie. Il est ainsi possible d'anticiper les problèmes futurs et possiblement d'agir avant qu'ils se concrétisent ou, à tout le moins, d'en atténuer les effets négatifs. C'est pourquoi un grand nombre d'économistes construisent des modèles de prévisions économiques et recherchent le plus d'information possible sur l'état de l'économie afin de l'utiliser dans leurs modèles. Souvent, cette information consiste en des variables macroéconomiques. Toutefois, certains éléments microéconomiques peuvent également avoir un poids assez important pour contribuer aux prévisions des chercheurs. Lorsqu'on utilise des agrégats macroéconomiques, l'histoire des données microéconomiques qui composent l'agrégat est alors perdue. On peut définir un agrégat macroéconomique comme étant la somme d'un grand nombre de petites unités microéconomiques. Grâce au théorème central limite, une telle somme possède des propriétés statistiques intéressantes. Il est ainsi possible d'ignorer les chocs individuels parce que l'on suppose qu'en moyenne les chocs idiosyncratiques vont s'annuler et avoir un effet négligeable. Par ailleurs, cette approche ne permet pas d'agréger les chocs uniques aux entreprises et d'en estimer les effets sur l'économie (Gabaix, 2011).

Plusieurs études cherchent à identifier la source de la variabilité observée au niveau des variables macroéconomiques, tel le PIB. En trouvant la source de la variabilité, il est alors plus facile d'en prévoir les fluctuations. Par exemple, Foerster et al. (2011) ont décomposé la production industrielle et ont conclu que la moitié de la variabilité

provenait de chocs idiosyncratiques. Ça confirme l'intérêt d'étudier ces chocs et de créer des variables pour les représenter.

Une étude réalisée par (Horvath, 1998) établit que dans un système où les transferts économiques sont lents entre les secteurs, les chocs idiosyncratiques propres aux secteurs sont significatifs et peuvent expliquer jusqu'à 80% des variations de la production d'une économie. Alors que les chocs idiosyncratiques au niveau sectoriel sont significatifs, il est intéressant d'étudier si ces chocs sont également significatifs à un niveau encore plus désagrégé, c'est-à-dire au niveau des entreprises. De plus, les chocs microéconomiques expliquent en grande partie la volatilité macroéconomique, (Carvalho et Gabaix, 2013) confirmant ainsi la pertinence de s'attarder aux données microéconomiques dans les modèles de prévisions macroéconomiques.

La production de quelques grandes entreprises peut représenter un pourcentage considérable du produit intérieur brut (PIB) d'une province, d'un état ou d'un pays. Par exemple, il est facile d'imaginer que Bombardier représente une part significative du PIB québécois. Ce dernier était d'environ 320 G\$ en 2016 alors que les revenus de Bombardier se chiffraient de 16,3 G\$ cette même année. En 2002, Walmart représentait 2,16 % du PIB américain (Gabaix, 2011). En Corée du Sud, il est estimé que Samsung et Hyundai représentent 22 % du PIB du pays. Les grandes entreprises peuvent avoir un impact important non seulement sur le PIB, mais également sur d'autres variables, tel les exportations. Eaton et al. (2012) ont estimé qu'en France les cent plus grandes entreprises sont responsables de près de la moitié des exportations du pays.

Ces grandes firmes vont connaître des variations liées à l'état général de l'économie ainsi qu'à leurs industries respectives. En plus de ces variations communes, elles vont

connaître des variations propres à elles-mêmes qui ne suivront pas l'économie du pays ou de l'industrie dans laquelle elles sont actives. Par exemple, lors d'une période de croissance économique la tendance générale de la production sera à la hausse, mais une entreprise pourrait connaître une grève générale qui paralyserait en grande partie sa production pendant plusieurs mois. Si cette entreprise a un poids assez important dans l'économie du pays, cet événement isolé aura des répercussions sur la production de son secteur industriel. Il pourrait même avoir des répercussions sur le PIB du pays. Ce dernier pourrait ainsi connaître une croissance plus lente ou une décroissance due à un choc idiosyncratique subit par une des grandes entreprises composant son économie. De tels exemples ont été observés par le passé, ils peuvent être positifs ou négatifs. Un choc positif pourrait être une innovation augmentant fortement les ventes. Par exemple, aux États-Unis, Ford a connu beaucoup de succès avec son modèle Pinto en 1972, les ventes de cette voiture ont explosé et il a été estimé que cette hausse des ventes aurait représenté une croissance de 0,22 % sur le PIB américain (Gabaix, 2011). Un choc négatif peut être une grève comme mentionnée plus haut ou encore une catastrophe naturelle, on peut penser aux feux de forêt en Alberta en 2016 qui ont fortement affecté l'industrie pétrolière.

L'idée que les chocs idiosyncratiques ont un impact sur l'économie générale d'un pays est renforcée par l'effet de cascade observé auprès des fournisseurs et clients d'une entreprise ayant subi un choc. C'est un argument qui a été amené en 2008 dans la décision de sauver l'industrie automobile aux États-Unis et au Canada (Acemoglu *et al.*, 2012). Au Québec, c'est l'une des raisons qui a incité le gouvernement à s'impliquer dans le sauvetage de Bombardier en 2015. En effet, des dizaines d'entreprises ont des liens économiques avec Bombardier, un choc à sa production d'avions aurait éventuellement eu des répercussions sur toute la grappe industrielle reliée au

secteur de l'aéronautique. (Carvalho *et al.*, 2016) s'est penché sur ce type d'effet de cascade en étudiant l'effet d'un tremblement de terre au Japon sur la chaîne de production et l'impact sur le PIB du pays. Il faut toutefois faire attention à l'effort qu'on met à protéger les grandes entreprises parce que leur stabilité ne se traduit pas nécessairement par une croissance économique supérieure (Fogel *et al.*, 2006).

Un indicateur a été créé par (Gabaix, 2011) pour capter les chocs idiosyncratiques qui affectent les plus grandes entreprises d'une économie. Dans un premier temps, cette étude cherchera à reproduire l'indicateur de Gabaix en utilisant la même méthodologie. Son indicateur a été bâti en utilisant les données annuelles américaines. Ce travail reprendra les données américaines et créera également un indicateur granulaire basé sur les données canadiennes. Il sera intéressant de comparer l'efficacité de l'indicateur dans ces deux marchés de types et de grandeurs différents.

En général, dans une économie où les grandes entreprises occupent beaucoup de place, c'est-à-dire avec un indice Gini élevé, on pourrait s'attendre à ce que le pouvoir prédictif de l'indice de granularité soit plus significatif. En effet, si la majorité de la production est concentrée entre quelques grosses firmes, il est facile d'imaginer qu'un choc les affectant aura plus d'effet que si quelques petites firmes subissent le même choc. Les États-Unis ont généralement un indice Gini plus élevé que le Canada.

Par ailleurs, dans les petites économies, les grandes firmes occupent souvent plus de place que dans les plus grandes économies. Ainsi, la volatilité des variables macroéconomiques est plus élevée dans les petites économies et provient en partie des chocs idiosyncratiques subis par les grandes entreprises (di Giovanni et Levchenko, 2012). Selon cette approche, on peut s'attendre à un pouvoir prédictif de l'indice granulaire plus important au Canada qu'aux États-Unis.

En général, on considère que les entreprises d'une économie sont distribuées selon une loi normale, mais ici, on va considérer qu'elles sont distribuées asymétriquement ou « fat-tailed ». Alors si un élément avec un poids assez grand se situe vers une extrémité, sa variabilité sera non négligeable. La diversification des entreprises ne suit pas $\frac{1}{\sqrt{N}}$, mais bien $\frac{1}{\ln(X)}$. Même si le nombre de firmes est très grand, le poids des firmes ne s'estompe pas aussi rapidement lorsque leur diversification suit $\frac{1}{\ln(x)}$, ainsi la variabilité reste non négligeable.

Tandis que Gabaix (2011) se limite au PIB américain et à une analyse prédictive à l'intérieur de l'échantillon, cette étude s'étendra à l'économie canadienne et testera le pouvoir prédictif de l'indicateur hors échantillon.

Les données utilisées sont extraites de la base de données Compustat. Il s'agit d'une base de données américaine qui couvre les firmes à capitalisation boursière et qui fournit une grande quantité de données provenant notamment des états financiers de ces firmes. De ce fait, la base de données contient presque seulement les firmes cotées en bourse.

Dans la base de Compustat, les données annuelles sont disponibles à partir de 1950, alors que les données trimestrielles débutent en 1962. Par ailleurs, les données trimestrielles nécessaires à la réalisation de cette étude ne sont disponibles qu'à partir des années 2000. Les données nécessaires sont le montant total des ventes et le nombre d'employés, elles sont utilisées pour calculer une productivité et construire l'indicateur granulaire. Pour une période donnée, seulement les entreprises dont on retrouve à la fois l'information sur les ventes et le nombre d'employés peuvent être retenues. Ainsi, le bassin d'entreprises à l'étude change d'une période à une autre dépendamment des années d'activités des entreprises et de la disponibilité des don-

nées. De plus, les entreprises qui ne sont pas cotées en bourse ne se retrouvent pas dans la base de données de Compustat. Cependant, puisque la majorité des entreprises classées dans le top cent américain et canadien se retrouve en bourse et que les entreprises dont les données sont manquantes sont rares, l'impact de l'absence de données sera négligeable.

Pour l'économie américaine, les données sur le PIB seront obtenues à partir du *Bureau of Economic Analysis* (BEA), il s'agit de la même source que Gabaix a utilisée. Dans le cas de l'économie canadienne, la base de données de Statistique Canada sera utilisée pour obtenir la valeur du PIB en dollars courants et celle de l'OCDE pour le PIB en dollars constants.

Dans l'étude de Gabaix (2011), les industries financières et de l'énergie ont été exclues. Les industries financières ont des chiffres de ventes très volatiles qui ne sont pas nécessairement en lien avec l'état de l'économie. On n'a qu'à penser aux fonds d'investissement qui vont acheter et vendre des parts de diverses entreprises, ces transactions vont apparaître sur les états financiers de la firme d'investissement et influencer les données, alors que leur impact sur l'économie n'est pas bien défini. Ce n'est pas le type d'effet que l'on désire capter avec l'indice granulaire.

Considérant que les ventes sont utilisées, les acquisitions devront également être prises en compte. Par exemple, si deux firmes de grosseur égale fusionnent, les ventes de la firme fusionnée seront du double. Ainsi, d'une année à l'autre la croissance de l'entreprise fusionnée serait d'environ 100 %, alors qu'en ventes nettes au niveau macroéconomique le résultat est le même. Un plafond et un plancher seront utilisés pour limiter les trop fortes fluctuations de croissance et éviter que les fusions ou les acquisitions ne viennent fausser les résultats. Le plafond et le plancher retenus sont

les mêmes que ceux de Gabaix (2011), soit 20 %.

Les modèles de prévisions vont généralement utiliser des données trimestrielles. Dans la présente étude, les données annuelles seront utilisées puisque les données trimestrielles de la base Compustat sont trop récentes pour donner des résultats robustes. Les données annuelles sont quant à elles disponibles depuis les années soixante.

Ainsi, pour une année donnée, la productivité de chaque entreprise est calculée en utilisant les ventes divisées par le nombre d'employés. La moyenne de productivité des cent plus grandes entreprises est produite pour chaque année. Ensuite, la moyenne générale est soustraite de la productivité de chaque entreprise pour isoler les chocs idiosyncratiques propres à chacune. Ces chocs sont ensuite pondérés selon le poids des ventes de l'entreprise par rapport au PIB de l'année précédente du pays en question. L'entreprise ayant généré le plus de ventes aura un poids plus important que celle en ayant généré moins.

Une fois un indicateur granulaire robuste obtenu, il sera intégré dans des modèles de prévisions économiques afin de tester son pouvoir prédictif. Ainsi, différents modèles de prévisions seront comparés pour établir s'il y a une valeur ajoutée à les bonifier de l'indice granulaire et pour voir si l'indice granulaire est un indicateur prévisionnel utile. L'étude comparera trois modèles de prévisions sur deux horizons de prévision, soit des horizons d'un et de deux ans.

L'étude de Gabaix a été publiée en 2011, ce qui en fait un indicateur prévisionnel très récent. Il n'est pas encore très répandu dans la communauté scientifique, d'où la pertinence de l'étudier et de tester son efficacité. Gabaix (2011) obtient des résultats très intéressants en termes de relation avec le PIB. Il obtient un R^2 de 34,6 % lorsqu'il régresse l'indicateur avec le taux de croissance du PIB américain. Il

existe donc un lien entre les chocs idiosyncratiques et le PIB américain, autrement le R^2 se rapprocherait de 0. Gabaix pousse plus loin son indicateur en contrôlant pour les divers types d'industries. Les chocs idiosyncratiques sont alors calculés par secteur d'activités industrielles. Par exemple, la productivité d'une entreprise de télécommunication sera comparée avec la moyenne des autres entreprises en télécommunication. Dans ce cas le R^2 obtenu est de 47,7 % toujours en relation avec la croissance du PIB américain. Par la suite, Gabaix utilise l'indicateur granulaire comme outil prévisionnel à l'intérieur de l'échantillon. Il obtient des résultats intéressants en comparant entre autres des prévisions réalisées avec l'indicateur granulaire et des prévisions réalisées avec le « spread » entre les bons du trésor de 5 ans et de 3 mois. Cette étude vise à créer un indice granulaire à partir de données canadiennes et à approfondir les pouvoirs prédictifs de l'indicateur granulaire.

Le pouvoir prédictif des indices granulaires calculé dans cette étude est faible, mais significatif. L'indice granulaire canadien produit des résultats meilleurs que pour l'indice granulaire américain. Dans un contexte de prévision hors échantillon annuel, ces résultats sont intéressants surtout pour l'indice granulaire canadien.

CHAPITRE I

CONSTRUCTION DE L'INDICE GRANULAIRE

1.1 Analyse de la base de données de Compustat

Dans son article, Gabaix (2011) fournit beaucoup d'informations sur la construction d'un indice granulaire. Par ailleurs, ses codes de programmation et sa méthodologie complète ne sont pas disponibles. Cette étude tentera de reproduire l'indice granulaire de Gabaix le plus fidèlement possible, mais il est probable que certaines disparités apparaissent entre l'indice de Gabaix et celui de cette étude. Les principales caractéristiques resteront toutefois les mêmes, un des buts du présent travail est de s'approcher le plus possible de l'indice original.

Les variables utilisées proviennent de la base de données Compustat. Compustat est une division de Standard & Poor qui a vu le jour en 1962. La base de données rassemble des données financières, statistiques et de marché sur des milliers d'entreprises à capitalisation boursière provenant d'un peu partout dans le monde, mais principalement des États-Unis.

1.1.1 Types d'entreprises

Certaines grosses entreprises ne sont pas publiques et leurs informations ne font pas partie de la base de données de Compustat. Le magazine Forbes (Murphy, 2016) a

publié en 2016 une liste de 223 grandes entreprises privées présentes aux États-Unis. Ces dernières représenteraient 1,62 trillion de dollars et emploieraient 4,5 millions de personnes. Quelques-unes de ces entreprises seront étudiées afin de mieux comprendre l'effet de leur non inclusion dans l'indice. Certaines de ces entreprises ont des volumes de production parmi les plus élevés aux États-Unis. N'ayant pas de compte à rendre à des actionnaires, les entreprises privées n'ont pas à rendre publics leurs états financiers. C'est ce qui empêche les bases de données comme Compustat d'en intégrer les données financières.

Selon le magazine Forbes, la plus grande entreprise privée aux États-Unis se nomme Cargill. Fondée en 1865, Cargill est la propriété de la famille MacMillan depuis plus d'une centaine d'années. Les ventes de l'entreprise ont atteint 120 milliards de dollars en 2015, dépassant celles, entre autres, de Costco, de General Electric ou d'Amazon en 2016. C'est donc une entreprise qui se classerait facilement parmi les cent plus grandes entreprises aux États-Unis. Cargill est active principalement dans le secteur des produits alimentaires.

La deuxième entreprise privée en importance aux États-Unis est Koch. Cette dernière est active principalement dans le secteur des produits pétroliers et des produits chimiques. Ses ventes ont atteint 100 milliards en 2015, assez pour faire partie de l'indice granulaire. Cette entreprise est la propriété de la famille Koch reconnue pour son implication politique aux États-Unis.

À noter la présence d'autres entreprises d'envergure dans le classement de Forbes comme Dell, Ernst & Young et PricewaterhouseCoopers. On constate que les ventes totales diminuent rapidement dans le classement de Forbes. Les trois firmes mentionnées plus haut ne feraient pas nécessairement partie de l'indice granulaire. Il y

a toutefois une perte d'information associée à l'exclusion de ces entreprises faute de données. Toutefois, leurs poids dans l'indice seraient relativement faibles étant donné leurs importances relativement modestes par rapport aux grosses entreprises capitalisées en bourse.

Une inspection cas par cas démontre déjà que peu d'entreprises privées feraient partie de l'indice si leurs données étaient disponibles. On peut également se poser la question par rapport à leur vulnérabilité face aux chocs idiosyncratiques. N'étant pas soumises aux pressions des marchés boursiers et au besoin de satisfaire les actionnaires, on peut imaginer que ces entreprises sont généralement plus stables que celles présentes sur les marchés. Cette hypothèse est confirmée par les résultats obtenus par Davis et al. (2006) qui ont observé des taux de volatilité plus faibles pour les entreprises privées que pour les entreprises publiques. De plus, une étude récente démontre que la privatisation d'une entreprise est associée à une diminution de la croissance de ses ventes et à une augmentation de ses profits (Aslan et Kumar, 2011).

Les entreprises privées sont susceptibles de connaître des chocs négatifs et positifs tout comme les entreprises publiques. Par exemple, historiquement, certaines usines et certains groupes de travailleurs de Cargill ont fait des grèves importantes.

Au Canada, l'exercice est beaucoup plus compliqué à réaliser. Plusieurs entreprises sont considérées comme privées parce qu'elles sont des branches de multinationales américaines ou d'autres pays. Par exemple, en 2012, Walmart Canada a réalisé des ventes de 22 milliards, mais elle n'apparaît pas dans la base de données de Compustat pour le Canada. Ces ventes placeraient Walmart parmi les vingt meilleures entreprises au niveau des ventes. C'est le cas aussi dans le domaine de l'industrie automobile où les compagnies mères sont basées ailleurs qu'au Canada, mais dont

la production et les ventes occupent une place très importante dans l'économie canadienne.

Les multinationales peuvent amener un biais au niveau de l'indice granulaire américain. Certaines entreprises multinationales américaines rapportent des données qui ne se limitent pas à l'économie américaine. Ainsi, un choc en Chine pourrait potentiellement avoir des répercussions sur l'indice granulaire américain. Il est difficile d'évaluer s'il est juste d'inclure de tels chocs dans l'indice granulaire. Il est possible que ces chocs sur les multinationales se répercutent sur le PIB américain.

Les entreprises gouvernementales ne sont pas non plus répertoriées dans la base de données Compustat. Elles sont plus présentes au Canada qu'aux États-Unis. Au Canada, elles peuvent représenter une part importante de l'économie. Ces entreprises devraient normalement être beaucoup plus stables que les entreprises composant l'indice granulaire. L'impact de leur non inclusion est atténué par la plus grande stabilité de ces entreprises par rapport aux entreprises publiques incluses dans l'indice granulaire. N'empêche qu'elles sont vulnérables également aux chocs idiosyncratiques, comme les grèves.

1.1.2 Localisation et secteurs d'activités

Grâce aux variables « loc » et « fic » de la base de données, il est possible de trier facilement les entreprises américaines et les entreprises canadiennes. Certaines firmes canadiennes ont publié leurs résultats en dollars américains. Ces données sont converties en dollars canadiens grâce au taux de change effectif lors de l'année de déclaration. Les taux de change sont fournis dans la base de données de Compustat. Il s'agit du taux de change effectif le jour de la publication des résultats.

Comme mentionné dans l'introduction de cette étude, des entreprises œuvrant dans certains secteurs d'activité ont été exclues de l'étude. Le tri des entreprises est basé sur les codes SIC. Ainsi les secteurs suivants ont été exclus :

- industrie financière (SIC 6000 à 7000) ;
- raffinerie et vente de produits pétroliers (SIC 2911 et 5172) ;
- l'extraction du pétrole et du gaz naturel (SIC 1311 et 1389) ;
- les services associés à l'énergie (SIC 4900 à 4941).

Ce tri représente environ 24 % des observations aux États-Unis et 29 % des observations au Canada. Le pourcentage plus élevé au Canada s'explique par le fait que les secteurs des finances et de l'énergie sont deux secteurs très importants de l'économie canadienne.

1.2 Indice granulaire

Une fois la base de données traitée, l'étape suivante consiste à calculer une productivité annuelle pour chaque entreprise. Pour ce faire, une nouvelle variable représentant la productivité est créée à partir du logarithme des revenus annuels divisés par le nombre d'employés.

$$\rho_{it} = \ln \frac{\textit{ventes}}{\textit{travailleurs}} \quad (1.1)$$

Par la suite, le taux de croissance de la productivité est calculé en soustrayant le logarithme de la productivité de la période précédente.

$$g_{it} = \rho_{it} - \rho_{it-1} \quad (1.2)$$

Le taux de croissance ainsi obtenu est composé des chocs idiosyncratiques et des fluctuations communes de l'économie. Les prochaines manipulations vont chercher à retirer les variations attribuables à l'état général de l'économie.

L'indice granulaire est construit à partir de taux de croissance de la productivité. On cherche à estimer la part de la variation de la productivité d'une entreprise attribuable à des chocs par rapport à l'année précédente. Le fait de travailler en taux de croissance exige d'avoir deux données de disponibles, soit l'année d'intérêt et l'année qui la précède. Cette exigence restreint un peu la base de données. En effet, si une année est manquante dans le continuum des revenus d'une entreprise alors deux années seront manquantes dans le calcul de l'indice de granularité. Ce n'est pas significatif, mais c'est une limite de la base de données de Compustat. Certaines données sont effectivement manquantes dans le continuum des déclarations d'une entreprise. Quelques fois, il s'agit d'une seule variable comme l'emploi alors que les revenus et les autres données sont présents. Bien que ce problème ait été observé, il n'est pas assez fréquent pour exiger d'envisager un traitement à la base de données. Gabaix (2011) ne mentionne pas cet enjeu dans son étude.

$$\Gamma_t = \sum_{i=1}^{100} \frac{S_{i,t-1}}{Y_{t-1}} (g_{it} - \bar{g}_t) \quad (1.3)$$

Une fois le taux de croissance de la productivité obtenu, les chocs idiosyncratiques seront extraits de la variation totale grâce à la formule 1.3. La valeur de \bar{g}_t peut être obtenue de deux façons, soit en calculant la moyenne de tous les taux de croissance pour une même année dans la base de données de Compustat, soit en allant chercher une valeur de productivité d'une autre source. Dans le premier cas, il s'agit de

calculer pour une même année la moyenne annuelle des taux de croissance de la productivité, on peut inclure toutes les entreprises ou en restreindre le nombre. Gabaix (2011) propose de prendre la moyenne des 100 ou des 1000 plus grandes. Dans ce calcul, chaque entreprise, peu importe sa taille, a le même poids sur le résultat. La moyenne est limitée par les caractéristiques de la base de données de Compustat. C'est un avantage parce que la valeur ainsi obtenue sera comparée avec d'autres valeurs en provenance de la même base de données. C'est toutefois un désavantage parce que ce n'est pas une valeur reconnue à l'échelle macroéconomique. La moyenne des cent plus grandes entreprises a été retenue dans l'indice granulaire de cette étude.

Une comparaison a été effectuée entre l'utilisation d'une productivité calculée à partir des données de Compustat et une productivité en provenance de la Conference Board Total Economy Database. On constate que la variance de l'indice granulaire basé sur la productivité externe est plus élevée. Considérant que la seule différence entre les deux calculs est le taux de croissance de la productivité moyenne utilisé. On aurait pu s'attendre à une plus grande similitude entre les deux indices. Le taux de corrélation entre les deux indices granulaires est négatif à -0,15. Le taux de corrélation avec le taux de croissance du PIB est de 0,22 pour l'indice de Compustat et de -0,14 pour celui basé sur le taux de croissance de productivité externe. Le taux de croissance de la productivité provenant de Compustat sera retenu étant donné sa corrélation plus forte avec le taux de croissance du PIB et parce que c'est la méthode de calcul utilisée par Gabaix.

Une fois les chocs idiosyncratiques obtenus, les valeurs sont agrégées par année. Pour une année t , l'agrégation des données est calculée selon le poids des ventes au temps $t-1$ de chaque entreprise par rapport au PIB total du pays au temps $t-1$ également.

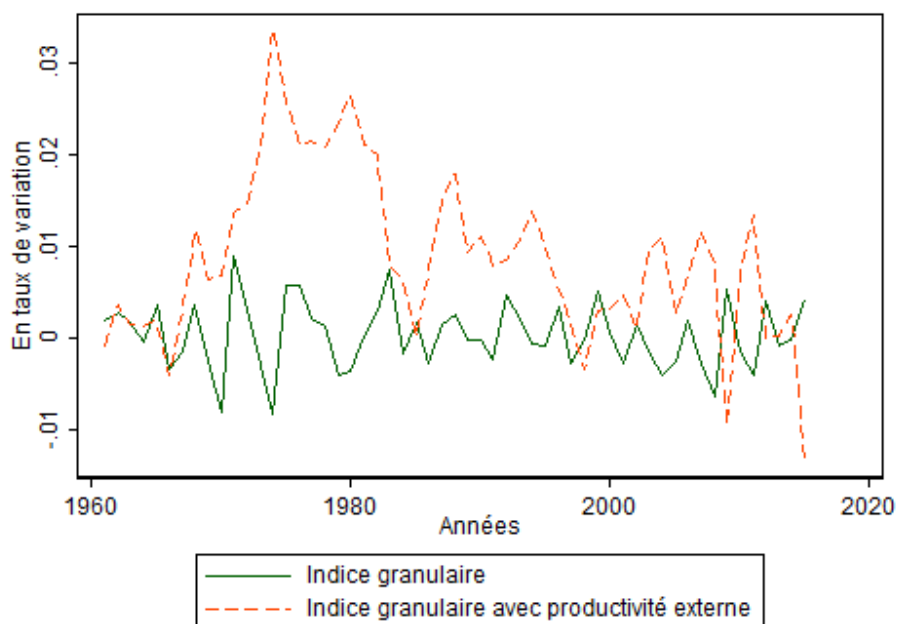


Figure 1.1 Comparaison de deux indices de granularité calculés avec deux différentes valeurs moyennes de productivité

Note : L'indice granulaire est calculé à partir de productivité provenant de données de la base de données Comstat tel qu'estimé par Gabaix. L'indice granulaire avec productivité externe est calculé à partir de productivité provenant de la Conference Board Total Economy Database.

Cette méthode d'agrégation suit généralement les principes d'agrégation (Zelenyuk, 2011).

Comme il a déjà été discuté dans la section précédente, la littérature s'est déjà penchée sur les chocs par industrie. Une option pour la construction de l'indice granulaire est de calculer les moyennes par secteur au lieu d'une moyenne globale pour tous les secteurs. Ainsi, les chocs sectoriels devraient être pris en compte et l'indice serait alors plus centré sur les chocs individuels.

Pour regrouper les entreprises par secteur, un tri en utilisant les codes SIC a été effectué. Le niveau de précision du tri s'arrête aux deux premiers chiffres du code SIC.

Ainsi, la formule suivante pour calculer un indice granulaire est obtenue. Quelques variations sont possibles selon les critères déjà discutés.

1.2.1 Plafond aux chocs idiosyncratiques

La source des chocs est inconnue à moins de procéder à une étude exhaustive de l'historique financier de chacune des entreprises incluses dans l'indice granulaire. Certains chocs ne sont pas pertinents et représentent un bruit nuisible à l'indice. Une source de bruit provient des fusions ou acquisitions entre les entreprises. Par exemple, lorsque deux entreprises fusionnent, leurs ventes s'additionnent ce qui peut générer une grande variation au niveau de la productivité, et ce, même si le nombre d'employés s'additionne également.

Un exemple répertorié par cette étude est celui de Lyondell Chemical, une entreprise américaine active internationalement. L'entreprise s'est établie en 1985 aux États-Unis, Lyondell Chemical fait son apparition dans le top 100 des ventes en 2005 à la suite de l'acquisition de Millenium Chemical. Les ventes passent alors de 5,9 milliards à 18,6 milliards, une augmentation de 315 %. La principale cause de cette augmentation est l'addition des ventes de Millenium Chemical à celles de Lyondell Chemical. Les ventes de l'entreprise fusionnée se sont accrues en 2005, tandis que le nombre d'employés de l'entreprise a été ajusté en 2004, passant de 3350 à 9930. Ainsi en 2004, l'entreprise a connu une forte diminution de sa productivité étant donné que les ventes étaient seulement celles de Lyondell tandis que les employés des deux entreprises étaient comptabilisés. En 2005, l'entreprise a connu une forte hausse de sa productivité alors que les ventes des deux entreprises ont été comptabilisées sous Lyondell. Deux ans plus tard, les données s'arrêtent pour Lyondell Chemical alors que l'entreprise devient LyondellBasel.

Gabaix (2011) a introduit un plafond et un plancher afin d'éviter que ces chocs aient une trop grande influence sur l'indice. Il établit le plafond à 20% de fluctuations de la productivité. La détermination de cette valeur semble être arbitraire. Des tests sur la variabilité ont été effectués avec différentes limites. Quatre tests sont répertoriés dans ce travail, soit des situations avec des plafonds et planchers à 50%, 20% et 10% et une situation sans limites aux fluctuations de la productivité.

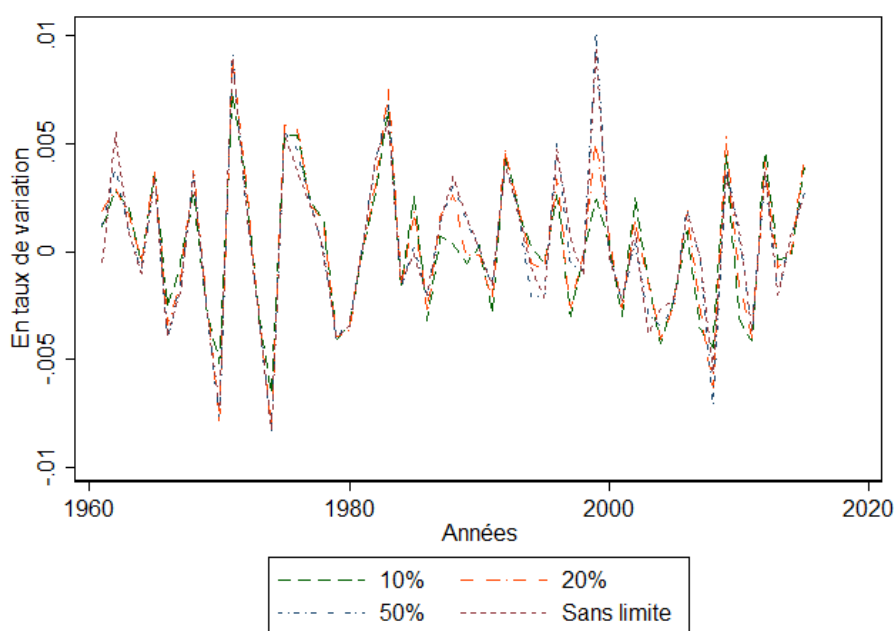


Figure 1.2 Comparaison de trois indices granulaires calculés avec des plafonds et des planchers à 50%, 20% et 10% et une situation sans limites

Note : Les plafonds et les planchers sont appliqués individuellement sur la productivité des 100 plus grandes entreprises annuellement.

Les tracés des quatre indices granulaires sont très fortement corrélés. Toutefois, on constate que l'amplitude de l'indice sans limites est plus forte certaines années. Les taux de corrélation entre les indices et le taux de croissance du PIB diminuent avec l'augmentation des limites, c'est-à-dire que le taux de corrélation est le plus élevé pour l'indice dont les fluctuations de productivité sont limitées à 20 %. La

différence entre les taux de corrélation est faible. On observe la même dynamique avec l'indice granulaire canadien, mais avec un taux de corrélation négatif avec le taux de croissance du PIB canadien.

Indices limites	10%	20%	50%	Sans
Américains	0,066	0,070	0,058	0,053
Canadiens	-0,101	-0,078	-0,044	0,010

Tableau 1.1 Taux de corrélation entre des indices granulaires américains et canadiens avec différentes limites et les taux de croissance de leur PIB respectif

Note : Les données relatives aux PIB proviennent du Bureau of Economics Analysis pour les États-Unis et de Statistique Canada pour le Canada. De 1960 à 2015, les indices granulaires sont calculés à partir de la productivité annuelle des 100 plus grandes entreprises provenant des données de Compustat

Le nombre d'entreprises à être intégré dans l'indice est une variable qui mérite d'être étudiée de plus près. Gabaix (2011) a utilisé les 100 et les 1000 plus grandes entreprises sans observer de gain significatif. Le modèle a été testé avec les 50, les 100 et les 1000 plus grandes entreprises. Il en résulte que la correspondance entre l'indice granulaire et le taux de croissance du PIB est légèrement meilleure pour le modèle à 1000 entreprises avec les données américaines. Au Canada, le manque de données ne permet pas de tester le modèle avec 1000 plus grandes entreprises.

Un autre élément à considérer est l'étendue de la période sur laquelle on désire calculer l'indice granulaire. En effet, il y avait beaucoup moins d'entreprises cotées en bourse au début des années soixante que dans les années 2000. Par exemple, pour l'année 1960, seulement 1 167 observations sont disponibles pour le nombre d'employés aux États-Unis tandis le nombre d'observations grimpe à 8 156 en 2000. Ainsi, c'est presque l'échantillon au complet qui est utilisé dans les années soixante

lorsqu'on utilise le modèle avec les 1000 premières entreprises.

Le peu d'entreprises disponibles pour le Canada entre 1960 et 1980 sera étudié dans la prochaine section de cette étude. Il s'agit d'un problème propre au Canada, il n'a donc pas été considéré par Gabaix dans son étude. L'indice granulaire canadien sera quand même généré de 1960 à 2015 pour permettre de le comparer avec l'indice granulaire américain sur la plus longue période possible.

1.2.2 Les indices granulaires américains et canadiens

Un graphique a été tracé pour comparer les indices granulaires américain et canadien. Les fluctuations sont très importantes tant pour le Canada que pour les États-Unis. On constate que pour certaines périodes les fluctuations suivent les mêmes tendances, mais que pour d'autres périodes les indices suivent des tendances opposées. On aurait pu s'attendre à des fluctuations plus synchronisées, comme c'est le cas pour les corrélations observées entre les fluctuations du PIB des deux pays. La proximité des deux économies nous a habitué à observer des tendances similaires entre les variables économiques des deux pays. Par ailleurs, étant donné que les indices granulaires sont composés de chocs idiosyncratique, il est normal de ne pas observer de corrélation importante entre les deux indices.

On observe à la figure 1.4 le graphique pour les indices granulaires par industrie. On observe les mêmes tendances entre les indices granulaires par industrie qu'entre les indices granulaires généraux.

En plus d'une étude graphique des indices, les taux de corrélation ont été calculés entre les quatre indices et sont présentés dans le tableau 1.2.

Pour un même pays, les indices sont fortement corrélés entre eux. On aurait pu

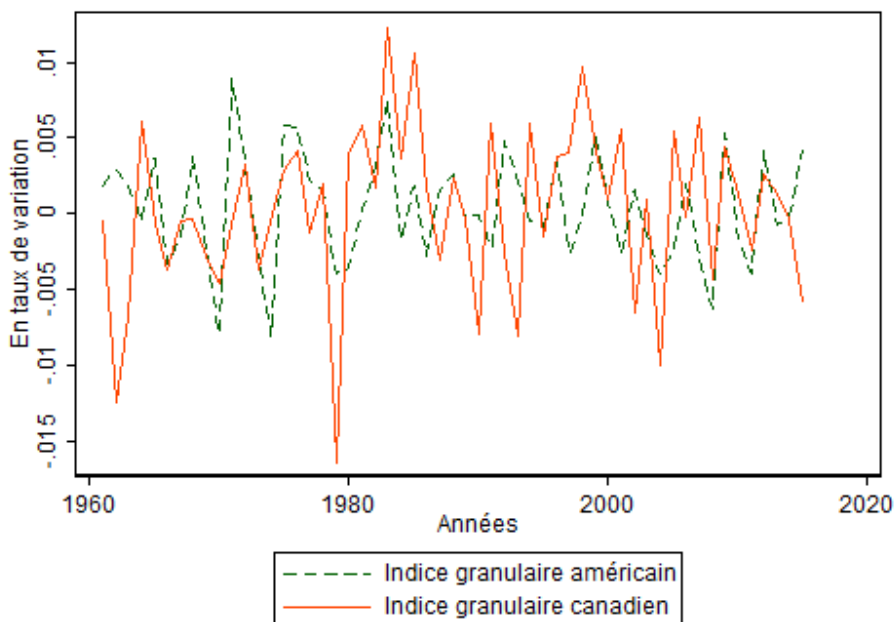


Figure 1.3 Comparaison de l'indice granulaire américain avec l'indice canadien

Note : De 1960 à 2015, les indices granulaires sont calculés à partir de la productivité annuelle des 100 plus grandes entreprises provenant des données de Compustat.

s'attendre à un taux de corrélation pour le Canada plus proche du taux de 0,851 tel qu'observé entre l'indice américain et l'indice américain par industrie. En considérant que le taux de corrélation entre le taux de croissance du PIB américain et le taux de croissance du PIB canadien est de 0,836, on aurait pu s'attendre à une corrélation plus importante entre les indices des deux pays. La corrélation est toutefois plus forte entre les indices granulaires globaux que ceux par industrie soient 0,206 pour le premier comparativement à 0,043 pour le second.

Il n'y a aucune corrélation significativement élevée entre les indices granulaires et les taux de variation des PIB. On aurait pu s'attendre à des taux de corrélation plus élevés. En effet, l'indice granulaire en tant qu'instrument prédictif devrait varier de façon corrélée avec le PIB. Les indices granulaires peuvent quand même avoir un

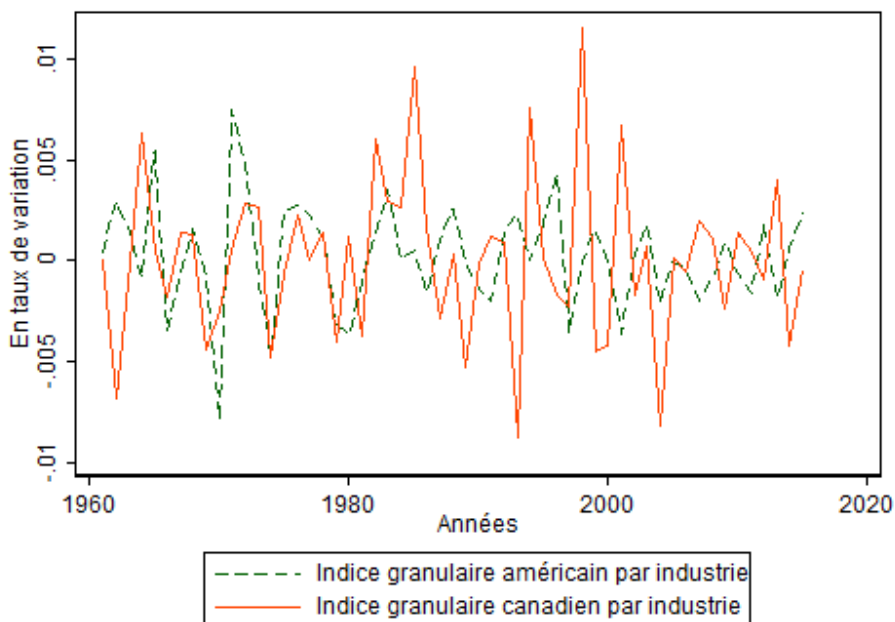


Figure 1.4 Comparaison de l'indice granulaire américain par industrie avec l'indice granulaire canadien par industrie

Note : De 1960 à 2015, les indices granulaires sont calculés à partir de la productivité annuelle des 100 plus grandes entreprises provenant des données de Compustat.

excellent pouvoir prédictif malgré ces taux de corrélation faibles. Par exemple, il est possible que l'indice granulaire suive des tendances similaires à celles du PIB même si la corrélation entre les deux variables est faible. La relation entre les deux variables pourrait ne pas être linéaire, expliquant ainsi la faiblesse des taux de corrélation.

Un test de stationnarité de type Dickey Fuller a été effectué sur l'indice granulaire. L'indice américain et l'indice canadien sont significativement stationnaires.

On constate également que les variances des indices granulaires canadiens sont plus élevées que celles des indices granulaires américains. Ceci pourrait s'expliquer par le fait que les ressources naturelles représentent une grande part de l'économie canadienne, particulièrement en termes d'entreprises inscrites à la bourse. Ce type

Indices	Américain	Américain par industrie	Canadien	Canadien par industrie
Américain	1			
Américain par industrie	0,851 (0,000)	1		
Canadien	0,205 (0133)	0,097 (0,481)	1	
Canadien par industrie	0,071 (0,607)	0,045 (0,744)	0,611 (0,000)	1

Tableau 1.2 Taux de corrélation entre les différents indices granulaires américains et canadiens

Note : De 1960 à 2015, les indices granulaires sont calculés à partir de la productivité annuelle des 100 plus grandes entreprises provenant des données de Compustat.

d'industrie est très volatile étant donné les variations importantes du prix des ressources naturelles. De plus, les entreprises canadiennes qui composent l'indice sont relativement plus petites que les entreprises qui composent l'indice américain. Ainsi, elles sont possiblement plus sujettes à des fluctuations importantes. Rappelons toutefois que certains secteurs économiques ont été retirés du calcul des indices. Seuls les secteurs utilisés par Gabaix ont été conservés. Ils ont donc été établis en fonction de l'économie américaine et non de l'économie canadienne. Une étude plus approfondie des entreprises composant les indices granulaires canadiens pourrait potentiellement mener à un ajustement des secteurs économiques retirés.

Indices\PIB	Américain	Canadien
Américain	0,068 (0,621)	-0,164 (0,232)
Américain par industrie	0,122 (0,377)	-0,069 (0,616)
Canadien	-0,014 (0,920)	-0,079 (0,567)
Canadien par industrie	-0,032 (0,818)	-0,084 (0,541)

Tableau 1.3 Taux de corrélation entre les différents indices granulaires américains et canadiens et les taux de croissance des PIB

Note : Les données relatives aux PIB proviennent du Bureau of Economics Analysis pour les États-Unis et de Statistique Canada pour le Canada. De 1960 à 2015, les indices granulaires sont calculés à partir de la productivité annuelle des 100 plus grandes entreprises provenant des données de Compustat

Indices	Variance
Américain	$1,34e^{-05}$
Américain par industrie	$7,08e^{-06}$
Canadien	$3,03e^{-05}$
Canadien par industrie	$1,60e^{-05}$

Tableau 1.4 Variance des indices granulaires

Note : De 1960 à 2015, les indices granulaires sont calculés à partir de la productivité annuelle des 100 plus grandes entreprises provenant des données de Compustat

1.2.3 Secteurs composant les bases de données et les indices granulaires

En observant les distributions des données par secteurs d'activité, on constate que certains secteurs sont surreprésentés dans les indices granulaires. Les statistiques

descriptives des répartitions par secteur sont présentées au tableau 1.5. Pour les États-Unis et le Canada, les secteurs de la production des biens de consommation et de la vente au détail sont surreprésentés dans l'indice granulaire. Il est intéressant de voir que les mêmes secteurs sont surreprésentés et sous-représentés pour les deux pays. Toutefois, on constate que le secteur des mines et de l'énergie est beaucoup plus important au Canada qu'aux États-Unis et que le contraire est observable pour le secteur des biens industriels. L'étude par secteur permet de constater que la répartition des plus grandes entreprises d'une économie ne suit pas la même répartition que celle de la totalité des entreprises. Il est normal que le secteur de la finances soit à 0 % étant donné que les entreprises provenant de ce secteur ont été retirées de la base de données.

1.2.4 Comparaison de l'indice granulaire obtenu avec celui de Gabaix

L'indice granulaire américain obtenu est presque pareil à celui obtenu par Gabaix. En effet, l'indice granulaire est corrélé à 99,1 % avec l'indice granulaire de Gabaix. L'indice granulaire par industrie de cette étude est pour sa part corrélé à 96,4 % avec l'indice granulaire par industrie de Gabaix. Ces taux de corrélation sont très élevés et permettent de conclure que les indices granulaires de Gabaix ont été reproduits avec succès.

Il est possible que les indices de Gabaix n'ont pu être reproduits avec un taux de corrélation de 100 % parce que les données de Compustat n'ont pas été extraites au même moment. Compustat doit constamment apporter des ajustements à sa base de données et doit de temps en temps réviser des données passées. L'extraction pour la réalisation de cette étude a été effectuée en 2017 tandis que Gabaix a probablement extrait les données avant 2011.

1.2.5 Autres données utilisées

En plus des données nécessaires à la construction de l'indice granulaire, d'autres données seront utilisées dans cette étude. Cette section présente les différentes données utilisées.

Le PIB américain qui servira aux prévisions provient de la base de données du Bureau of Economics Analysis (BEA). Les valeurs du PIB sont présentées en dollars américains de 2016 et sont disponibles selon des fréquences trimestrielles et annuelles. Gabaix utilise également le BEA comme source pour ses valeurs du PIB américain.

Le PIB canadien qui servira aux prévisions provient de la série de Statistique Canada sur le PIB en termes de dépenses. Les données sont disponibles en termes trimestriels non désaisonnalisés, les valeurs annuelles ont été obtenues en additionnant les valeurs trimestrielles.

États-Unis			
Secteurs	Répartition	Répartition des	Écart entre les
	globale	entreprises	
	des entreprises	composant	
		l'indice granulaire	
Agriculture, pêche, forêts	0,50	0,12	-0,38
Mines et énergie	3,21	1,02	-2,19
Biens de consommation	19,68	27,71	8,03
Biens industriels	35,20	31,23	-3,97
Transport	8,08	13,05	4,97
Ventes au détail	13,58	20,55	6,97
Finances	0	0	0
Services personnels	13,86	3,65	-10,21
Services industriels	4,49	0,40	-4,09
Affaires internationales	1,40	2,28	0,88
Canada			
Agriculture, pêche, forêts	0,34	0,04	-0,20
Mines et énergie	17,35	11,20	-6,15
Biens de consommation	22,84	29,45	6,61
Biens industriels	23,59	23,51	-0,08
Transport	10,27	13,40	3,13
Ventes au détail	12,25	16,78	4,53
Finances	0	0	0
Services personnels	9,87	2,44	-7,43
Services industriels	2,31	1,97	-0,34
Affaires internationales	1,18	1,21	0,03

Tableau 1.5 Répartition en pourcentage par secteur des entreprises

Note : Les données de 1960 et 2015 de la base de Compustat après le tri pour retirer certains secteurs dont celui finances sont utilisées pour calculer les répartitions par secteurs.

CHAPITRE II

POUVOIR PRÉDICTIF DE L'INDICE GRANULAIRE

Le précédent chapitre expliquait la provenance et les subtilités de l'indice granulaire. Cette section utilisera l'indice granulaire dans différentes prédictions macroéconomiques et cherchera à établir son pouvoir prédictif.

Dans un premier temps, plusieurs modèles de prédiction économique seront utilisés afin de vérifier le pouvoir prédictif de l'indice granulaire sur le PIB réel américain et le PIB réel canadien. Certains modèles incorporeront l'indice granulaire et d'autres non. Les modèles seront comparés afin d'établir si l'ajout de l'indice granulaire a eu un effet positif sur les prévisions.

Deux horizons de prévisions seront étudiés soit un horizon d'un an et un de deux ans. Étant donné qu'on travaille avec des données annuelles, un horizon de prévision plus long que deux ans serait difficile à justifier.

2.1 Les modèles de prédiction économique

Trois modèles seront utilisés pour obtenir les prévisions économiques les plus précises possible et pour les comparer.

2.1.1 Modèle AR(p)

Le premier modèle utilisé sera un modèle autorégressif. C'est un modèle relativement simple qui servira de modèle référence. Les autres modèles seront évalués en comparaison avec ce modèle simple de prévision. Selon ce modèle, les fluctuations d'une variable sont expliquées par le passé de cette même variable. La forme générale du modèle est définie par la formule suivante :

$$Y_t = \beta_0 + \beta_1 Y_{t-h} + u_t \quad (2.1)$$

Le nombre de retards est indiqué par p et représente l'horizon passé étudié. Comme l'étude utilise des variables annuelles, le nombre de retards ne dépassera pas deux ans. Comme une seule variable compose ce modèle, il ne sera pas possible d'y introduire l'indice de granularité.

Ainsi, pour cette étude le modèle suivant sera utilisé sur les horizons d'un et de deux ans pour estimer les PIB américains et canadiens :

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t \quad (2.2)$$

$$Y_t = \beta_0 + \beta_2 Y_{t-2} + u_t \quad (2.3)$$

2.1.2 Modèle ADL avec indice granulaire

Le deuxième modèle qui sera étudié permet d'ajouter des variables supplémentaires à la prévision du PIB :

$$Y_t = \beta_0 + \beta_1 Y_{t-h} + \delta_1 X_{t-h} + u_t \quad (2.4)$$

Il s'agit d'un modèle *autoregressive distributed lag* ou ADL. Le modèle AR(p) est limité par le nombre de variables explicatives. Il est intéressant d'utiliser un modèle qui permette l'ajout de variables explicatives supplémentaires afin d'expliquer les changements à partir de diverses sources. Ainsi, l'indice granulaire peut être incorporé aux prévisions en plus de la variable dépendante. Il sera alors possible de vérifier si cet ajout est positif en comparant les résultats des prévisions avec et sans l'indice granulaire.

Dans un premier temps, une construction simple de ce modèle est effectuée. Elle consiste à utiliser seulement le passé du PIB et l'indice granulaire pour prévoir la valeur future du PIB.

Cette version du modèle pour les deux horizons étudiés peut être exprimée ainsi :

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 \Gamma_{t-1} + u_t \quad (2.5)$$

$$Y_t = \beta_0 + \beta_1 Y_{t-2} + \beta_2 \Gamma_{t-2} + u_t \quad (2.6)$$

L'indice granulaire est représenté par la variable Γ

2.1.3 Modèle ADL avec une variable supplémentaire

Afin d'avoir une meilleure idée du pouvoir prédictif de l'indice de granularité, une autre variable explicative a été introduite dans le modèle, soit le taux de chômage annuel. Cette variable a été choisie notamment parce que les chocs idiosyncratiques auraient tendance à avoir un effet total nul au niveau de l'emploi (Karasik *et al.*, 2016), ce qui en fait une bonne variable pour compléter ce modèle. Le taux de chômage permet de capter les fluctuations de l'emploi tandis que l'indice granulaire capte d'autres éléments économiques. De plus, plusieurs recherches démontrent que les principaux créateurs d'emplois ne sont pas nécessairement les grandes entreprises, mais plutôt les nouvelles et souvent petites entreprises (Haltiwanger *et al.*, 2010, Criscuolo *et al.*, 2014).

Habituellement, les prévisions macroéconomiques sont réalisées sur des périodicités plus courtes qu'une année, soit sur des périodicités trimestrielles ou mensuelles. La longueur de la périodicité réduit les variables pouvant être utilisées. De plus, la variable choisie doit être disponible au Canada et aux États-Unis sur l'ensemble de l'horizon étudié, soit à partir de 1960. Le modèle avec cette variable supplémentaire est représenté par les formules suivantes :

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{t-1} + \beta_3 \Gamma_{t-1} + u_t \quad (2.7)$$

$$Y_t = \beta_0 + \beta_1 Y_{t-2} + \beta_2 X_{t-2} + \beta_3 \Gamma_{t-2} + u_t \quad (2.8)$$

Le taux de chômage est représenté par la variable X dans ces deux formules.

2.1.4 Exercice de prévision hors échantillon

Afin de comparer les modèles et d'évaluer le pouvoir prédictif de l'indice granulaire, une étude hors échantillon ou *pseudo-out-of-sample* a été réalisée. Dans son étude, Gabaix avait pour sa part réalisé un exercice comparatif à l'intérieur de l'échantillon *in-sample*.

Dans un premier temps, pour réaliser une étude hors échantillon, il faut calibrer les modèles sur une période donnée qui s'arrête quelques périodes avant la fin des mesures observées. C'est-à-dire qu'on estime les valeurs des β , δ et Λ dans les modèles à partir d'un échantillon restreint. Les valeurs ainsi estimées sont évaluées seulement en fonction de l'information disponible dans l'échantillon restreint. Avec ces variables, la prévision de la valeur du PIB au temps "t" est effectuée. Les valeurs prédites peuvent être comparées avec les vraies valeurs mesurées qui sont disponibles dans l'échantillon complet. Par la suite, l'échantillon est avancé d'une période et l'exercice de prévision est répété. On procède ainsi jusqu'à prédire la dernière valeur connue des données disponibles.

Dans le cadre de cette étude, une période de 20 années a été retenue pour effectuer l'étude hors échantillon, soit de 1996 à 2015. Par exemple, pour prédire la valeur du PIB de 1996, les données de 1961 à 1995 seront utilisées comme s'il s'agissait des seules données disponibles. Le point de départ de l'échantillon est maintenu stable tout au long de l'exercice, soit l'année 1961. La période choisie pour effectuer l'exercice hors échantillon présente l'avantage de couvrir différentes périodes économiques, incluant la crise économique de 2008.

2.2 Résultats des prévisions

Pour évaluer le pouvoir prédictif du modèle, les différentes prévisions ont été comparées. L'erreur quadratique moyenne (EQM) a été calculée entre chaque estimation et les valeurs réelles des taux de croissance du PIB (Y) correspondantes. Ainsi, une EQM faible signifie que les valeurs prédites du PIB sont proches des valeurs réelles des taux de croissance du PIB.

$$EQM = \sum_{i=1}^n (\hat{Y}_{it} - Y_{it})^2 \quad (2.9)$$

2.2.1 Résultats avec les données américaines

Dans un premier temps, l'exercice de prévision a été réalisé pour l'économie américaine avec les deux indices granulaires estimés pour cette économie. Le premier tableau compare les EQM obtenues pour les différentes prévisions du taux de croissance du PIB américain.

Le pouvoir prédictif de l'indice granulaire sur le PIB réel américain n'est pas concluant. Les EQM des modèles ADL avec l'indice granulaire seulement sont plus élevées que celles du modèle de référence AR. L'indice granulaire par industrie sur un horizon de deux ans y fait cependant exception, l'EQM y étant similaire à celle du modèle de référence. Sur un horizon d'un an, l'ajout du taux de chômage a un effet positif sur le pouvoir prédictif du modèle. En fait, le modèle ADL avec seulement le taux de chômage et le PIB passé comme variable explicative, enregistre l'EQM la plus faible.

Le taux de chômage semble perdre son pouvoir prédictif sur un horizon de deux

Modèle \ Horizon	1	2
AR	$4,80e^{-04}$	$5,05e^{-04}$
ADL avec l'indice granulaire	1,06 (0,049)	1,05 (0,718)
ADL avec l'indice granulaire par industrie	1,07 (0,063)	1,00 (0,806)
ADL avec l'indice granulaire et le taux de chômage	0,85 (0,026)	1,11 (0,489)
ADL avec l'indice granulaire par industrie et le taux de chômage	0,86 (0,026)	1,09 (0,528)
ADL avec le taux de chômage	0,83 (0,477)	1,08 (0,498)

Tableau 2.1 Comparaison entre différents modèles de prévision du taux de croissance du PIB réel américain et un modèle étalon autorégressif de 1996 à 2015

Note : Les résultats des modèles ADL sont présentés sous forme de fraction du résultat obtenu par le modèle AR. Les données de 1960 et 2015 de la base de Compustat ont été utilisées pour réaliser les prévisions. Les données relatives au PIB proviennent du Bureau of Economics Analysis. Les p-values sont présentés entre parenthèses sous les erreurs quadratiques moyennes.

ans. Le test de Diebold-Mariano a été effectué entre tous les modèles et le modèle de référence pour s'assurer que les résultats sont significativement différents. Les résultats sont présentés entre parenthèses sous les EQM. Tous les résultats sont significativement différents du modèle de référence. On peut toutefois constater que les résultats sont plus significativement différents pour les modèles avec un horizon d'un an que pour ceux avec un horizon de deux ans.

Lorsqu'on étudie graphiquement les résultats, on constate que l'amplitude des fluc-

tuations du PIB est plus grande pour les données réelles que pour les modèles de prévision. On observe que plusieurs mouvements sont corrélés entre les données réelles et celles prédites avec les indices granulaires. Par exemple, en 2009, le modèle ADL avec l'indice granulaire a baissé de manière significative, comme le taux de croissance du PIB. Pour sa part, le modèle de référence AR a réagi avec un certain retard et de manière moins significative.

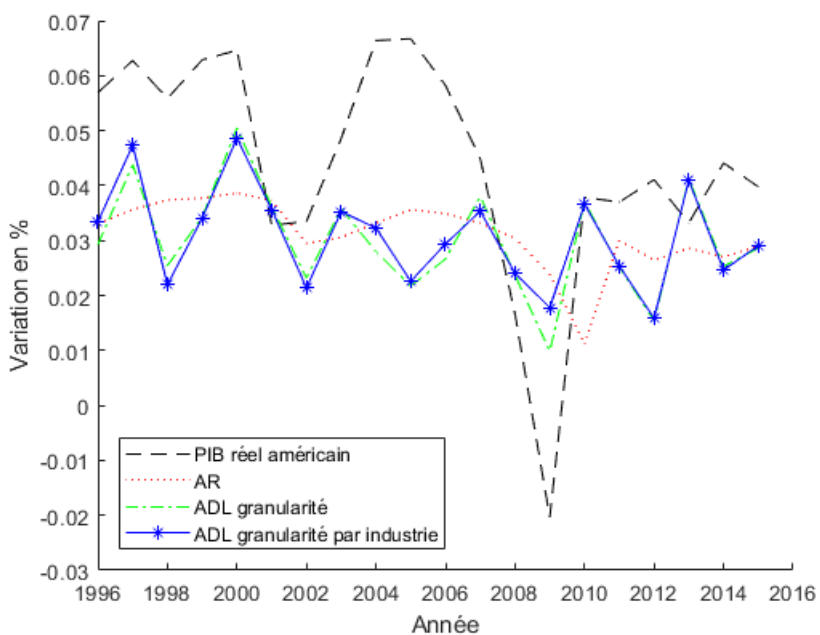


Figure 2.1 Illustration comparative du taux de croissance du PIB réel américain avec trois modèles de prévision sur un horizon d'un an

Note : Les données relatives au PIB proviennent du Bureau of Economics Analysis.

Le pouvoir prédictif de l'indice granulaire sur un horizon de deux ans est significativement moins important que sur un horizon d'un an. En effet, les deux courbes représentant les modèles avec les indices granulaires sont très rapprochées de la courbe du modèle de référence AR. Cette proximité suggère que l'indice granulaire a peu de poids dans la prévision du PIB sur un horizon de deux ans. Les valeurs prédites du taux de croissance du PIB se rapprochent fortement de la valeur moyenne

du taux de croissance du PIB sur cette période. L'EQM entre le modèle avec l'indice granulaire et la moyenne du PIB est de $1,06e^{-05}$ comparativement à $1,02e^{-04}$ pour l'horizon d'un an. Ça signifie que la prévision du modèle ADL avec l'indice granulaire se rapproche davantage de la moyenne du taux de croissance du PIB que de la valeur réelle.

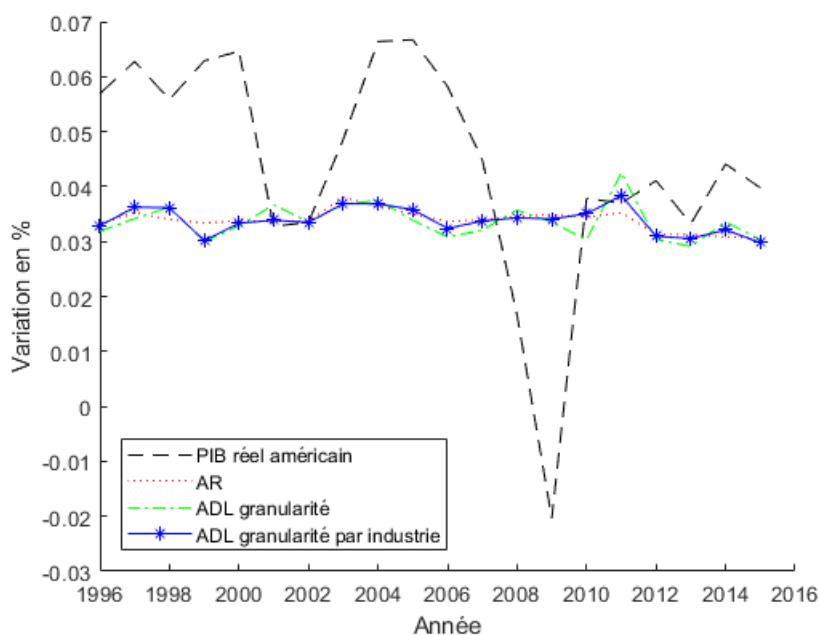


Figure 2.2 Illustration comparative du taux de croissance du PIB réel américain avec trois modèles de prévision sur un horizon de deux ans

Note : Les données relatives au PIB proviennent du Bureau of Economics Analysis.

2.2.2 Résultats avec les données canadiennes

Pour sa part, le tableau 2.2 compare les résultats des différentes prévisions pour le PIB canadien.

L'indice granulaire canadien semble pour sa part avoir un certain pouvoir prédictif sur le PIB canadien. Parmi les indices granulaires, l'indice général a un pouvoir prédictif plus grand que celui par industrie. En effet, les deux modèles avec seulement

Modèle \ Horizon	1	2
AR	$3,185e^{-04}$	$3,452e^{-04}$
ADL avec l'indice granulaire	0,93 (0,586)	0,99 (0,646)
ADL avec l'indice granulaire par industrie	1,00 (0,529)	1,05 (0,833)
ADL avec l'indice granulaire et le taux de chômage	1,05 (0,243)	1,08 (0,655)
ADL avec l'indice granulaire par industrie et le taux de chômage	1,13 (0,382)	1,15 (0,830)
ADL avec le taux de chômage	1,12 (0,458)	1,09 (0,961)

Tableau 2.2 Comparaison entre différents modèles de prévision du taux de croissance du PIB réel canadien et un modèle étalon autorégressif de 1996 à 2015

Note : Les résultats des modèles ADL sont présentés sous forme de fraction du résultat obtenu par le modèle AR. Les données de 1960 et 2015 de la base de Compustat ont été utilisées pour réaliser les prévisions. Les données relatives au PIB proviennent de Statistique Canada. Les p-values sont présentés entre parenthèses sous les erreurs quadratiques moyennes.

le PIB passé et l'indice granulaire général comme variables prédictives ont des EQM inférieures à celles du modèle de référence. Les écarts ne sont pas très grands, mais ils sont significativement différents de l'EQM du modèle de référence. Le modèle ADL avec l'indice granulaire global avec un horizon d'un an a obtenu le meilleur résultat de prévision du PIB canadien avec un ratio de 0,93 avec le modèle de référence. L'ajout de la variable du taux de chômage n'améliore pas les performances du modèle.

L'étude du graphique 2.3 montre que l'indice de granularité a un pouvoir prédictif

restreint sur les variations du PIB. En effet, les courbes représentant les deux estimations effectuées avec les indices granulaires sont à proximité de la courbe du modèle de référence AR(p). C'est ce qui nous amène à conclure que l'ajout des indices de granularité au modèle de prévision procure peu d'information permettant d'améliorer le modèle. On constate également que les réactions du modèle prédictif canadien sont souvent en retard par rapport aux fluctuations réelles du PIB canadien.

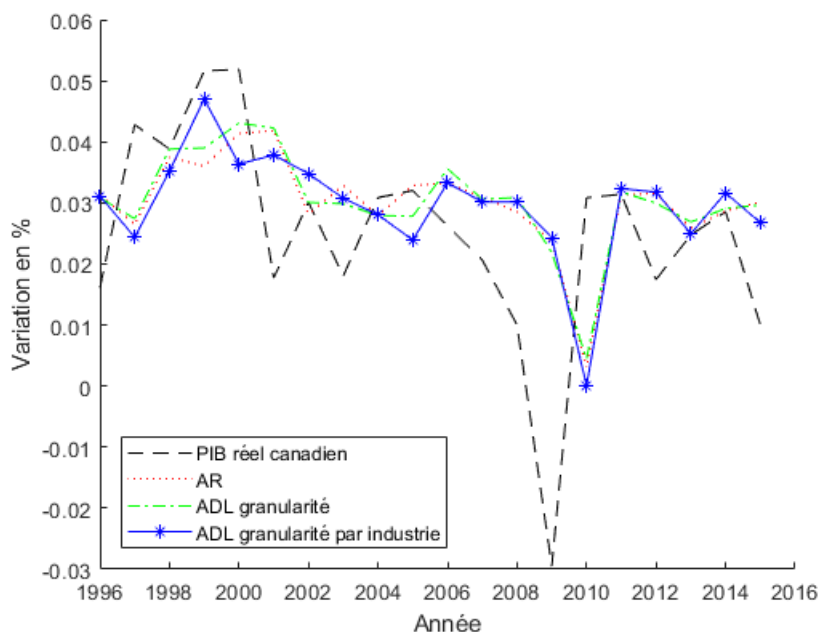


Figure 2.3 Illustration comparative du taux de croissance du PIB réel canadien avec trois modèles de prévision sur un horizon d'un an

Note : Les données relatives au PIB proviennent de Statistique Canada.

2.3 Limitations du modèle

2.3.1 Performance hors échantillon

Les résultats obtenus sont moins significatifs que ceux obtenus par Gabaix (2011), ce résultat est normal étant donné que l'exercice de prévision est réalisé hors échantillon tandis que Gabaix (2011) procède à l'intérieur de l'échantillon. Il est toujours

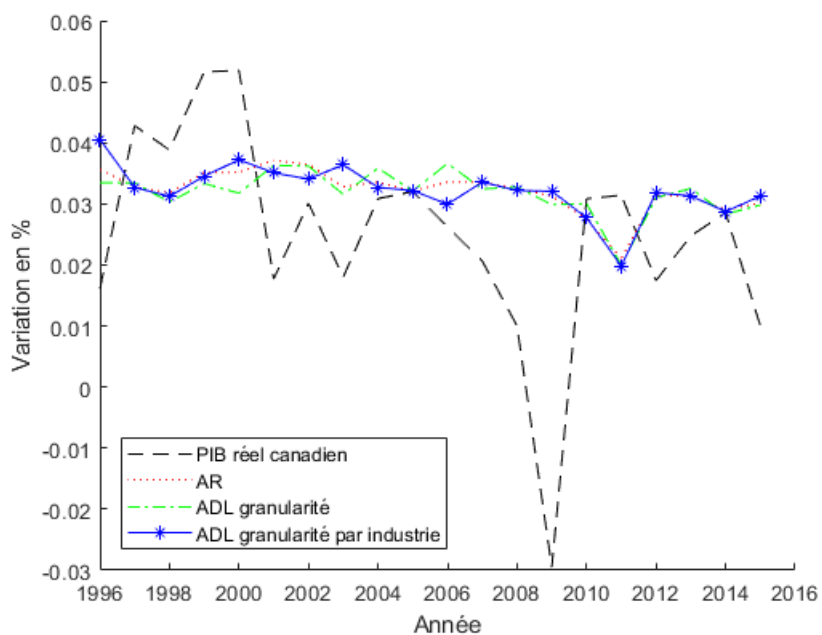


Figure 2.4 Illustration comparative du taux de croissance du PIB réel canadien avec trois modèles de prévision sur un horizon de deux ans

Note : Les données relatives au PIB proviennent de Statistique Canada.

plus ardu de réaliser un exercice prévisionnel hors échantillon qu'à l'intérieur de l'échantillon. En effet, l'exercice à l'intérieur de l'échantillon a l'avantage de pouvoir profiter de l'information disponible pour la même période que celle pour laquelle une prévision est effectuée.

Il est relativement normal que l'indice granulaire ait un pouvoir prédictif faible hors échantillon. Il est très difficile de trouver des variables ayant un pouvoir prédictif élevé sur une période d'une année. Le fait que les prévisions sont effectuées sur une base annuelle réduit le pouvoir prédictif du modèle. Un indice granulaire trimestriel aurait vraisemblablement un pouvoir prédictif supérieur par rapport à un indice granulaire annuel. Il n'était toutefois pas possible dans le cadre de ce travail de produire un indice granulaire trimestriel.

Une explication à la performance mitigée de l'indice en mode hors échantillon est le ciblage des chocs. En effet, il est possible que les chocs granulaires aient une influence sur le PIB lors de l'année où ils surviennent, mais que seulement certains types de chocs auront un impact qui se répercutera sur le PIB de l'année suivante.

2.3.2 Période de prévision

Au Canada, il y a moins de données disponibles sur les industries canadiennes pendant les années soixante. En effet, l'indice granulaire n'est pas composé des plus grandes entreprises pendant cette période, il est simplement composé des entreprises existantes dont les données sont disponibles. À partir des années soixante-dix, le nombre d'entreprises canadiennes augmente dans la base de données, cependant ce n'est qu'à partir de 1979 qu'il est possible de sélectionner les cent plus grandes entreprises. Ainsi, afin d'approfondir les effets que pourrait avoir le nombre plus restreint de données au Canada, différents horizons de prévisions ont été testés afin de déterminer si le nombre plus faible d'entreprises pendant les années soixante pouvait avoir un effet négatif sur le modèle de prévision. Deux étendues supplémentaires ont été utilisées pour réaliser les prévisions, soit à partir de 1970 et de 1980. L'étendue de l'exercice hors échantillon est restée la même, soit de 1996 à 2015.

En utilisant un point de départ plus tardif pour réaliser la prévision, le pouvoir prédictif des modèles augmente étant donné que les EQM estimées diminuent. Comme l'amélioration est observée également dans le modèle de référence AR, ce n'est pas seulement l'effet d'une amélioration de l'indice granulaire qui impacte le pouvoir prédictif du modèle. Il est possible que les fluctuations du PIB canadien à partir des années quatre-vingt expliquent mieux les fluctuations récentes du PIB. Durant les années soixante, la banque du Canada appliquait une politique monétaire différente

Modèle \ Année de départ	1961	1970	1980
AR	$3,18e^{-04}$	$2,99e^{-04}$	$2,98e^{-04}$
ADL avec l'indice granulaire	0,93 (0,586)	0,88 (0,456)	0,88 (0,486)
ADL avec l'indice granulaire par industrie	1,00 (0,529)	1,01 (0,584)	1,03 (0,674)
ADL avec l'indice granulaire et le taux de chômage	1,05 (0,243)	1,01 (0,335)	0,91 (0,441)
ADL avec l'indice granulaire par industrie et le taux de chômage	1,13 (0,382)	1,15 (0,589)	1,05 (0,712)
ADL avec le taux de chômage	1,12 (0,458)	1,09 (0,726)	1,04 (0,942)

Tableau 2.3 Comparaison avec différentes années de départ entre des modèles de prévision du PIB réel canadien et un modèle étalon autorégressif

Note : Les résultats des modèles ADL sont présentés sous forme de fraction du résultat obtenu par le modèle AR. Les données de 1960 et 2015 de la base de Compustat ont été utilisées pour réaliser les prévisions. Les données relatives au PIB proviennent de Statistique Canada. Les p-values sont présentés entre parenthèses sous les erreurs quadratiques moyennes.

de la politique actuelle, celle-ci aurait débuté au début des années quatre-vingt. Le pouvoir prédictif de l'indice granulaire général semble s'améliorer avec les années de départ plus tardives. En effet le ratio avec le modèle de référence passe de 0,93 à 0,88.

La même observation ne peut être faite dans le cas du modèle de prévision du PIB américain. En déplaçant le début de l'étendue des variables explicatives à 1970 et 1980, les EQM augmentent par rapport au modèle débutant en 1961. Dans le cas

américain, le manque de données pendant les années soixante ne constituait pas un enjeu.

CONCLUSION

Cette étude a permis de recréer un indice granulaire similaire à celui obtenu par Gabaix. La corrélation entre l'indice calculé dans cette étude et celui de Gabaix est très élevée. Des indices granulaires ont ainsi pu être calculés avec succès à partir des données canadiennes. Les taux de corrélation entre les indices granulaires des deux pays sont moins élevés qu'à ce à quoi on aurait pu s'attendre. Cette observation peut s'expliquer par le fait que les économies des deux pays ne sont pas concentrées dans les mêmes secteurs d'activités. Tandis que plusieurs des entreprises composant les indices granulaires canadiens proviennent du secteur des ressources naturelles, les indices granulaires américains sont composés d'une plus grande part d'entreprises œuvrant dans le secteur tertiaire de l'économie. Une autre explication est que les chocs sont généralement uniques à chaque pays et à une seule entreprise. En effet, une grève ou une catastrophe naturelle aura généralement lieu dans un des deux pays et non dans les deux pays simultanément.

Le pouvoir prédictif de l'indice granulaire sur le PIB dans une étude hors échantillon est faible, mais présent. On observe quand même qu'il détient un certain potentiel prédictif. Ce potentiel est observable davantage pour le PIB canadien que pour le PIB américain. Gabaix avait, pour sa part, réalisé ses prévisions basées sur un exercice de prévision à l'intérieur de l'échantillon. Les résultats obtenus sont donc complémentaires à ceux obtenus par Gabaix. Tout comme Gabaix, cet indice granulaire est calculé sur une base annuelle. Dans le cadre de prévisions hors échantillon, l'estimation de l'indice sur une base trimestrielle pourrait amener des résultats très

pertinents par rapport à son pouvoir prédictif. Il faudrait toutefois avoir accès à des données plus stables trimestriellement pour pouvoir réaliser une telle étude. Il serait intéressant d'effectuer cet exercice comparatif hors échantillon avec des données provenant d'autres pays afin de confirmer le potentiel prédictif de l'indice granulaire.

RÉFÉRENCES

- Acemoglu, D., Carvalho, V. M., Ozdaglar, A. et Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5), 1977–2016.
- Aslan, H. et Kumar, P. (2011). Lemons or cherries? growth opportunities and market temptations in going public and private. *Journal of Financial and Quantitative Analysis*, 46(2), 489–526.
- Carvalho, V. et Gabaix, X. (2013). The great diversification and its undoing. *American Economic Review*, 103(5), 1697–1727.
- Carvalho, V. M., Nirei, M., Saito, Y. et Tahbaz-Salehi, A. (2016). Supply chain disruptions : Evidence from the great East Japan earthquake. *Columbia Business School Research Paper*, 17(5).
- Criscuolo, C., Gal, P. N. et Menon, C. (2014). The dynamics of employment growth : New evidence from 18 countries. *OECD Science, Technology and Industry Policy Papers*, 14.
- di Giovanni, J. et Levchenko, A. A. (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy*, 120(6), 1083–1132.
- Fogel, K., Morck, R. et Yeung, B. (2006). Big business stability and economic growth : Is what’s good for General Motors good for America? *National Bureau of Economic Research*, 12394.
- Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79(3), 733–772.
- Haltiwanger, J. C., Jarmin, R. S. et Miranda, J. (2010). *Who creates jobs? Small vs. large vs. young*. Working Paper 16300, National Bureau of Economic Research.
- Horvath, M. (1998). Cyclicalité and sectoral linkages : Aggregate fluctuations from independent sectoral shocks. *Review of Economic Dynamics*, 1, 781–808.
- Karasik, L., Leung, D. et Tomlin, B. (2016). Firm-specific shocks and aggregate fluctuations. *Banque of Canada Staff Working Paper*, 51.

Murphy, A. (2016). America's largest private companies 2016. *Forbes*. consulté le : 2017-07-14. Récupéré de <https://www.forbes.com/sites/andreamurphy/2016/07/20/americas-largest-private-companies-2016/#1a0e3d76f0be>

Zelenyuk, V. (2011). Aggregation of economic growth rates and of its sources. *European Journal of Operational Research*, 212(1).