

N° 11-633-X au catalogue — N° 027
ISSN 2371-3437
ISBN 978-0-660-35714-0

Études analytiques : méthodes et références

Indices expérimentaux de l'activité économique des provinces et territoires au Canada : mesures expérimentales fondées sur des combinaisons de séries chronologiques mensuelles

par Nada Habli, Ryan Macdonald et Jesse Tweedle

Date de diffusion : le 19 août 2020



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

Programme des services de dépôt

- | | |
|-----------------------------|----------------|
| • Service de renseignements | 1-800-635-7943 |
| • Télécopieur | 1-800-565-7757 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2020

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Indices expérimentaux de l'activité économique des provinces et territoires au Canada : mesures expérimentales fondées sur des combinaisons de séries chronologiques mensuelles

par

Nada Habli, Ryan Macdonald et Jesse Tweedle

Division de l'analyse économique
Statistique Canada

11-633- N° 027

2019027

ISSN 2371-3437

ISBN 978-0-660-35714-0

Août 2020

Études analytiques : méthodes et références

Les documents de cette série traitent des méthodes utilisées pour produire des données qui seront employées pour effectuer des études analytiques à Statistique Canada sur l'économie, la santé et la société. Ils ont pour but de renseigner les lecteurs sur les méthodes statistiques, les normes et les définitions utilisées pour élaborer des bases de données à des fins de recherche. Tous les documents de la série ont fait l'objet d'un examen par les pairs et d'une révision institutionnelle, afin de veiller à ce qu'ils soient conformes au mandat de Statistique Canada et qu'ils respectent les normes généralement reconnues régissant les bonnes pratiques professionnelles.

Les documents peuvent être téléchargés gratuitement de www.statcan.gc.ca.

Remerciements

Les auteurs tiennent à remercier Étienne Saint-Pierre, Yvan Clermont, Danny Leung et Brenda Bugge pour leur soutien et leurs conseils tout au long du projet. Ils aimeraient aussi remercier Steve Matthews, Fred Picard et Michel Ferland pour leurs commentaires et leurs contributions sur la modélisation. Ils souhaitent également remercier Philip Smith et Karen Wilson pour leurs conseils, leurs suggestions et leur soutien.

Table des matières

Résumé.....	5
1 Introduction.....	6
2 Ensemble de données d'entrée	9
3 Estimation	11
3.1 Stratégie d'estimation.....	13
3.1.1 Modèle simple.....	15
3.1.2 Analyse des composantes principales.....	16
3.1.3 Moindre contraction absolue et opérateur de sélection.....	18
4 Évaluation mensuelle de l'indice	19
5 Conclusion	23
Références	25

Résumé

Le présent document porte sur les méthodes de création d'un indicateur mensuel de l'activité économique pour les provinces et les territoires. Il commence par la création d'un ensemble de données pour les provinces et les territoires, composé de séries mensuelles sur : la population active, y compris les salaires et l'emploi; le commerce international; les mesures de production telles que les ventes du secteur de la fabrication ou la production d'électricité; et les prix (consommation, logement et électricité). Lorsque cela est nécessaire, les séries sont désaisonnalisées, couplées et déflatées pour créer des séries chronologiques continues allant de janvier 2002 à avril 2020. Des méthodes de réduction des variables sont ensuite appliquées à l'ensemble de données provincial et territorial mensuel pour créer les indices expérimentaux mensuels de l'activité économique des provinces et territoires. Trois méthodes sont examinées : l'analyse en composantes principales (ACP), la moindre contraction absolue et opérateur de sélection (LASSO) et un indice simple composé de trois séries prédéterminées (emploi total, exportations totales et ventes au détail). Une moyenne pondérée de l'indice simple et de l'indice ACP est également calculée. En général, les indices produits donnent une assez bonne image de l'activité économique provinciale, suivant les mouvements cycliques des économies provinciales et territoriales. Toutefois, la configuration pour les modèles n'est pas idéale, car des données annuelles sont utilisées pour produire les paramètres des modèles, ce qui crée de l'incertitude quant au rendement du modèle. Par conséquent, plusieurs indices sont fournis. Une évaluation de la qualité donne une indication des forces et des limites des indices en ce qui concerne les différentes utilisations.

1 Introduction

Des mesures actuelles de l'activité économique sont essentielles pour comprendre le rendement des économies, et pour orienter les réponses stratégiques aux fluctuations macroéconomiques. Le début de la pandémie attribuable à l'émergence du virus SRAS-CoV-2 a mis l'accent sur ce point, ainsi que sur la nécessité de mesures spécifiques à la région géographique. À l'heure actuelle, le Canada dispose d'un solide système de production de mesures d'activité à jour, comme le produit intérieur brut (PIB) réel, au niveau national. En ce qui concerne les économies provinciales et territoriales, des renseignements mensuels sur les marchés du travail ou sur des activités particulières comme la fabrication ou le commerce international sont disponibles, mais on ne dispose pas d'une mesure mensuelle de l'activité économique agrégée.

Dans des circonstances normales, la production d'un nouvel ensemble d'indicateurs économiques agrégés pour les provinces et les territoires exigerait la création de mesures exploratoires, et demanderait peut-être le lancement de nouvelles enquêtes ou l'expansion des activités de collecte de statistiques existantes, ainsi que la création de l'infrastructure nécessaire à la production et à la diffusion des indicateurs de manière continue. La mise en œuvre de ces modifications nécessite du temps. Dans le contexte actuel, où la pandémie du SRAS-CoV-2 accentue les besoins en indicateurs économiques régionaux mensuels, le temps nécessaire pour mettre sur pied un nouveau programme statistique afin de répondre aux exigences actuelles signifie que cette approche est irréalisable.

Une approche plus rapide consiste à adopter une stratégie fondée sur des modèles statistiques afin de créer rapidement des mesures exploratoires pour l'activité économique provinciale. Bien que cette approche introduise une nouvelle mesure de l'activité économique en temps opportun, le compromis est que les méthodes employées ne sont pas utilisées dans les situations pour lesquelles elles sont idéales, que les entrées pour les modèles utilisent des séries actuellement disponibles sans pouvoir adapter leurs utilisations à la création d'un indicateur, et que les modèles reposent peu sur la théorie. Autrement dit, les modèles recherchent des corrélations dans les données plutôt que d'utiliser la théorie économique pour guider leur construction. Enfin, les modèles employés ici ont généralement un ensemble différent d'intrants pour chaque province ou territoire. Par conséquent, l'uniformité des structures du modèle ne peut être maintenue pour chaque province et territoire, ce qui peut avoir une incidence sur les comparaisons entre les secteurs de compétence¹.

Pour estimer les indices de l'activité, une stratégie à deux volets est appliquée. En premier lieu, un ensemble de données de panel équilibré composé de séries chronologiques provinciales et territoriales mensuelles est constitué à partir de tableaux de l'Entrepôt commun de données de sortie (ECDS) qui sont accessibles au public. Cet ensemble de données s'étend de janvier 2002 aux plus récents points de données disponibles (actuellement avril 2020). Deuxièmement, trois

1. On ne sait pas exactement dans quelle mesure les différences dans le nombre des données d'entrée et la diversité des données d'entrée peuvent avoir une incidence sur les comparaisons. Les plus grandes provinces disposent d'un ensemble de données plus riches dans lequel elles peuvent puiser pour modéliser leur économie. De plus, les enquêtes qui recueillent des données provinciales et territoriales comportent généralement davantage d'observations et/ou produisent des données qui présentent moins de variabilité par rapport aux mouvements tendanciels dans les plus grandes provinces. Les économies plus petites sont également moins complexes et peuvent nécessiter un moins grand nombre de séries pour résumer leurs activités.

Cette situation est normale lorsqu'on fait des comparaisons entre les économies. Par exemple, les économies plus importantes de l'Organisation de coopération et de développement économiques (OCDE), comme les États-Unis, l'Allemagne, le Royaume-Uni ou le Canada, peuvent avoir des systèmes statistiques plus importants que les économies plus modestes de l'OCDE. Cela ne veut pas dire que les données de ces économies sont incomparables. Au contraire, les études sont entreprises en tenant compte du fait que les données sont aptes à faire des comparaisons puisqu'elles reflètent ce qui se passe au sein des secteurs de compétences sur lesquels elles sont censées faire rapport, même si la quantité de renseignements statistiques disponibles diffère. La même notion devrait s'appliquer aux indices expérimentaux de l'activité économique.

méthodes sont appliquées à l'ensemble de données pour transformer la série mensuelle en un indicateur de l'activité économique provinciale et territoriale : un modèle simple, l'analyse en composantes principales (ACP) et la moindre contraction absolue et opérateur de sélection (LASSO). Un quatrième indice qui combine l'indice du modèle simple et l'indice basé sur l'ACP est également produit. Les résultats de ces approches forment les mesures exploratoires de l'activité économique provinciale agrégée qui sont présentées ici.

Les indices expérimentaux sont basés sur l'utilisation de modèles statistiques avec un ensemble de données d'entrée qui contient une série d'approximations et d'hypothèses. Parmi les hypothèses pour l'ensemble de données d'entrée, on compte notamment : l'utilisation de déflateurs nationaux pour produire de vraies séries provinciales lorsque les déflateurs provinciaux n'existent pas, l'hypothèse que les taux de croissance pour toutes les séries sont stationnaires en matière de covariance, et l'hypothèse que la winsorisation des données aux 5^e et 95^e percentiles est appropriée.

Pour les modèles, la croissance annuelle du PIB réel dans l'ensemble des provinces et des territoires sert de mesure de l'activité provinciale agrégée par rapport à laquelle les mesures dérivées de l'indice de l'activité économique sont comparées. Il en résulte une situation où un petit nombre d'observations annuelles sont utilisées pour estimer les paramètres du modèle qui sont utilisés pour déduire les fluctuations mensuelles. Le petit nombre d'observations a une incidence sur la capacité des modèles à fournir des inférences suffisantes, et l'utilisation de données annuelles peut masquer d'importantes différences dans le calendrier mensuel des variations des prix, de la production et de l'emploi.

Afin de permettre la possibilité que l'utilisation des données annuelles puisse avoir une incidence sur le rendement et l'inférence du modèle, le modèle simple suppose qu'un ensemble de variables est approprié et utilise les moindres carrés ordinaires (MCO) pour déterminer leur contribution relative. Dans le cas de l'ACP, on utilise un R carré ajusté et maximisé à partir des régressions par MCO pour choisir lesquels des 10 premiers composants principaux doivent être inclus. Dans ces deux situations, les régressions par MCO incluent des variables qui sont statistiquement non significatives. Pour le LASSO, la signification statistique des variables d'entrée potentielles détermine le modèle final. Toutefois, en ce qui concerne le LASSO, pour certaines économies, aucune variable d'entrée n'est sélectionnée. Dans ces cas, on utilise plutôt la méthode elastic-net ou une stratégie de modélisation générale à spécifique.

Comme l'ensemble de données d'entrée contient un certain nombre d'hypothèses et d'approximations et que les modèles statistiques sont utilisés dans un contexte imparfait, les indices expérimentaux qui sont créés doivent être considérés comme des approximations de l'activité économique agrégée plutôt que comme des mesures exactes. Les indices d'activité ont tendance à présenter une volatilité mensuelle considérable, et lorsqu'on les compare aux estimations du PIB réel produites par le gouvernement du Québec ou le gouvernement de l'Ontario, ils peuvent montrer une plus grande volatilité cyclique.

Dans l'ensemble des mesures, les modèles simples et les modèles LASSO tendent à s'appuyer davantage sur les séries d'emplois comme données d'entrée. L'indice simple a l'avantage d'être facile à comprendre et est le plus comparable pour toutes les provinces et tous les territoires puisqu'il maintient la constante des variables utilisées dans l'indice d'activité, peu importe la province ou le territoire. Les indices de l'ACP semblent saisir davantage de fluctuations liées à d'autres aspects de l'activité globale (p. ex. ventes ou exportations), mais ils présentent aussi une plus grande variabilité. La combinaison pondérée des indices simples définis par les chercheurs et des indices ACP a une meilleure corrélation avec la croissance du PIB réel que les indices constitutifs.

À l'heure actuelle, les indices pondérés ou les indices basés sur le LASSO semblent offrir le meilleur compromis entre les signaux présents dans les données et la variabilité des séries mensuelles. Cette évaluation est fondée sur la mesure dans laquelle les modèles semblent se conformer à la configuration utilisée pour l'estimation ainsi qu'au comportement des indices. Dans certains cas, comme l'indice fondé sur l'ACP pour Terre-Neuve-et-Labrador, il y a une anomalie qui remet en question la véracité de l'indice. Lorsque ce genre de situation se produit, les données sont considérées comme impropres à l'utilisation à l'heure actuelle, et les estimations pour cet indice ne sont pas fournies. Les évaluations sont en cours et peuvent entraîner des changements dans les utilisations ou les indices recommandés, à mesure que l'on étudie la possibilité d'élaborer davantage l'ensemble de données d'entrée, d'affiner les modèles ou d'autres stratégies de modèles.

Le reste du présent document est structuré comme suit. La section 2 traite de la création de l'ensemble de données d'entrée ainsi que des hypothèses utilisées pour filtrer et transformer les données avant la modélisation. La section 3 décrit les modèles employés, les hypothèses intégrées aux modèles, ainsi que leurs forces, leurs limites et leur application pour créer des indices mensuels. La section 4 présente une analyse du rendement du modèle et illustre les indices qui en résultent. La section 5 présente la conclusion.

2 Ensemble de données d'entrée

L'ensemble de données d'entrée est composé de mesures propres à une province ou à un territoire pour l'activité économique et les déflateurs économiques au niveau du Canada, sauf dans les cas où des déflateurs propres à une province ou à un territoire sont disponibles. Les séries mensuelles de données d'entrée sont composées d'enquêtes mensuelles sur la main-d'œuvre, les produits et les prix (tableau 1). Dans certains cas, les tableaux actifs ne contiennent pas de renseignements continus allant de janvier 2002 à ce jour. Dans ce cas, les tableaux historiques sont utilisés pour rétropoler les tableaux actifs.

Tableau 1

Tableaux des données d'entrée des séries chronologiques provinciales et territoriales

Numéro de tableau	Titre du tableau
12100099	Importations et exportations de marchandises, base douanière, selon le Système harmonisé de désignation et de codification des marchandises (SH), Canada, provinces et territoires, États-Unis, États
12100119	Commerce international de marchandises par province, produit et principaux partenaires commerciaux
14100036	Heures effectivement travaillées selon l'industrie, données mensuelles non désaisonnalisées
14100201	Emploi selon l'industrie, données mensuelles non désaisonnalisées
14100222	hebdomadaires moyennes pour l'ensemble des industries excluant les entreprises non classifiées, données mensuelles désaisonnalisées
14100287	Caractéristiques de la population active, données mensuelles désaisonnalisées et tendance-cycle, cinq derniers mois
14100292	Caractéristiques de la population active selon le territoire, moyennes mobiles de trois mois, données désaisonnalisées et non désaisonnalisées, 5 derniers mois
14100355	Emploi selon l'industrie, données mensuelles désaisonnalisées et non désaisonnalisées, et tendance-cycle, cinq derniers mois
16100048	Ventes pour les industries manufacturières selon l'industrie et province, données mensuelles (dollars sauf indication contraire)
18100004	Indice des prix à la consommation mensuel, non désaisonné
18100204	Indice des prix de vente de l'énergie électrique, mensuel
18100205	Indice des prix des logements neufs, mensuel
20100008	Ventes de commerce de détail par province et territoire
20100074	Commerce de gros, ventes
21100019	Enquête mensuelle sur les services de restauration et les débits de boissons
24100002	Nombre de véhicules voyageant entre le Canada et les États-Unis
25100001	Statistiques de l'énergie électrique, avec des données pour les années 1950 à 2007
25100015	Production de l'énergie électrique, production mensuelle selon le type d'électricité
34100003	Permis de bâtir, valeur selon le secteur d'activité
34100066	Permis de bâtir, par type de structure et type de travaux
34100158	Société canadienne d'hypothèques et de logement, logements mis en chantier, toutes les régions, pour Canada et les provinces, désaisonnalisées au taux annuel, mensuel

Note : SH : Système harmonisé.

Source : Statistique Canada, compilation des auteurs.

Les déflateurs sont principalement recueillis dans le cadre de programmes d'enquête au niveau du Canada pour les mesures de l'activité économique (tableau 2). À l'heure actuelle, Statistique Canada ne produit pas de déflateurs propres aux provinces et aux territoires pour les mesures en dollars courants du commerce international, des ventes du secteur de la fabrication, des ventes en gros, des ventes au détail ou de l'Enquête mensuelle sur les services de restauration.

Pour recueillir les déflateurs, les indices de prix au niveau du Canada sont tirés d'enquêtes lorsqu'ils sont disponibles. Dans le cas de la fabrication, l'indice implicite de prix est calculé en tant que ratio entre la valeur nominale et la valeur réelle. Il existe des déflateurs pour les séries nominales depuis janvier 2002, sauf pour la fabrication. Pour la fabrication, l'Indice des prix des produits industriels (IPPI) par industrie, l'IPPI par groupe de produits et le ratio entre le PIB nominal et le PIB mensuel réel sont utilisés comme base de projection pour les déflateurs de l'industrie manufacturière.

Tableau 2

Tableaux de données pour les séries chronologiques des déflateurs

Numéro de tableau	Titre du tableau
12100128	Commerce international de marchandises par classification des produits, indices de prix et de volume, mensuel
16100013	Valeur réelle des ventes, des commandes, des stocks possédés et le ratio des stocks aux ventes des industries manufacturières, en dollars de 2012, désaisonnalisées
16100047	Stocks, ventes, commandes et rapport des stocks sur les ventes pour les industries manufacturières, selon l'industrie (dollars sauf indication contraire)
18100004	Indice des prix à la consommation mensuel, non désaisonné
18100029	Indice des prix des produits industriels, par principaux groupes de produits, mensuel
18100032	Indice des prix des produits industriels, par industrie, mensuel
20100003	Ventes de grossistes, prix et volume, selon l'industrie, désaisonnalisées
20100038	Commerce de détail, ventes, dollars enchaînés et indices de prix, inactif
20100051	Commerce de gros, ventes, dollars enchaînés et indices de prix, inactif
20100078	Ventes au détail, prix et volume, désaisonné
36100434	Produit intérieur brut (PIB) aux prix de base, par industriel, mensuel

Note : SH : Système harmonisé.

Source : Statistique Canada, compilation des auteurs.

Pour combiner les données provinciales et les données des déflateurs afin de créer l'ensemble de données d'entrée, il y a quatre étapes :

1. Assembler les données. Les tableaux de l'ECDS sont filtrés pour sélectionner les données désirées. Seules les variables avec des données continues sont sélectionnées. Les séries pouvant faire l'objet de suppression sont exclues, mais les séries comportant des valeurs nulles sont incluses. Les séries pouvant faire l'objet de suppression sont généralement des séries aux valeurs plus petites, ce qui signifie qu'elles contiennent moins de renseignements sur les fluctuations économiques agrégées. Bien qu'il existe des méthodes pour interpoler ces points de données, si la suppression se produit au cours du dernier mois, une prévision serait nécessaire pour remplir le point de données supprimé. Étant donné que l'indice est en cours de construction pour fournir des renseignements sur le plus grand choc ayant frappé l'économie canadienne depuis la Seconde Guerre mondiale, on juge qu'il n'est pas souhaitable d'inclure des séries comportant des valeurs prévues.
2. Désaisonnaliser les séries de données. Ce ne sont pas toutes les séries qui sont fournies sur une base désaisonnée. Il y a au total 966 séries qui sont désaisonnalisées pour être utilisées dans l'indicateur. Étant donné le nombre élevé de séries, les options automatiques de l'algorithme ARIMA-SEATS du progiciel R relatif à la saisonnalité sont utilisées pour supprimer la saisonnalité. Les paramètres saisonniers sont déterminés en fonction des séries chronologiques mensuelles disponibles jusqu'en décembre 2019. Dans le cas où les séries chronologiques ne couvrent pas toute la période, comme les séries interrompues, les données jusqu'à la période la plus récente sont utilisées pour déterminer les options de désaisonnalisation.

Afin de s'assurer que les séries désaisonnalisées sont de bonne qualité, elles ont été validées par diverses mesures de qualité. Cela inclut notamment vérifier la présence de saisonnalité, la quantité de saisonnalité stable présente par rapport à la quantité de saisonnalité évolutive (M7), l'absence d'effets saisonniers dans la composante irrégulière, l'uniformité de la série désaisonnalisée par rapport à sa forme brute et de contrôler le nombre de valeurs aberrantes détectées automatiquement par ARIMA-SEATS jusqu'à un maximum de cinq. Les séries qui se sont avérées non saisonnières (125 séries) sont conservées sous leur forme brute. Les séries désaisonnalisées dont la qualité est médiocre (183 séries) sont éliminées de l'ensemble de données avant l'estimation.

3. Coupler les données. Après la désaisonnalisation, les données sont couplées au besoin. Lorsqu'il existe des périodes de chevauchement, des liens sont établis en traitant les données comme des indices et en les enchaînant rétrospectivement sur les périodes historiques. S'il n'existe pas de périodes de chevauchement, les valeurs de niveau sont jointes « telles quelles ».
4. Appliquer les déflateurs. Si nécessaire, des déflateurs sont appliqués aux séries en dollars courants ou aux séries de prix pour produire des variables relatives des prix.

L'ensemble de données complet avant le filtrage contient des séries désaisonnalisées et non désaisonnalisées, des séries nominales et des séries déflatées propres à la province et au territoire. Pour modéliser l'activité économique provinciale et territoriale, des séries en unités naturelles (p. ex. employés, heures travaillées), des séries déflatées, des taux (p. ex. le taux de chômage) et les prix relatifs sont sélectionnés.

3 Estimation

L'objectif est d'estimer les séries chronologiques mensuelles pour l'activité économique agrégée dans les provinces et les territoires en tant que fonction des séries chronologiques économiques mensuelles provinciales et territoriales disponibles :

$$y_{\text{mensuel},t} = f(\text{mensuel}/\text{période}/\text{série})$$

La mesure la plus couramment citée de l'activité économique agrégée est la mesure du PIB réel décrite dans le Système de comptabilité nationale de 2008 (Nations Unies, [2010](#))². Au niveau de l'ensemble du Canada, la fonction pour la transformation des séries de données d'entrée en PIB réel mensuel est fondée sur des méthodologies et des points de repère propres à l'industrie qui ont été élaborés et enrichis au fil du temps. Les méthodologies utilisent diverses sources de données pour estimer les variations de la production brute qui servent d'approximation pour le PIB réel. Les approximations sont combinées à des points de repère annuels du PIB réel pour produire la série mensuelle du PIB réel. Dans de nombreux cas, une mesure directe de la production brute est disponible. Toutefois, dans certaines industries, aucune mesure directe de la production n'est disponible et les estimations sont produites au moyen d'autres sources de données, comme l'emploi (Statistique Canada, [2020b](#))³. Cette méthodologie forme la fonction

-
2. Le Système canadien des comptes macroéconomiques produit des mesures annuelles, trimestrielles et mensuelles du PIB pour le Canada. Les mesures sont intégrées et les mesures mensuelles du PIB réel sont calculées en se basant sur une méthodologie qui utilise des pseudo-indices à une fréquence mensuelle. Ces indices sont pondérés ensemble au moyen de formules de calcul d'indices pour former la série agrégée du PIB réel mensuel. Les indices sont étalonnés chaque année aux tableaux annuels des ressources et des emplois, ce qui permet de corriger les différences qui peuvent découler de l'utilisation de pseudo-indices au lieu des séries à valeur ajoutée à double déflation produites dans les tableaux des ressources et des emplois.
 3. Des méthodes détaillées pour les industries individuelles ainsi que des discussions sur les pseudo-indices et leur utilisation sont disponibles à Canada ([2020a](#)).

$f(.)$ dans laquelle les séries mensuelles sont placées pour produire le PIB réel mensuel du Canada.

La difficulté en ce qui concerne mesurer l'activité économique agrégée des provinces et des territoires est le fait que la fonction $f(.)$ pour les provinces et les territoires est inconnue et que la série désirée, $y_{mensuel,t}$, est également inconnue.

Puisque le but est d'estimer $y_{mensuel,t}$, si un proche substitut existait, il pourrait servir d'instrument pour le vrai $y_{mensuel,t}$. La fonction $f(.)$ pour combiner des renseignements mensuels afin de produire une mesure agrégée de l'activité économique pourrait alors être approximative. Bien qu'il n'existe aucun substitut mensuel proche pour $y_{mensuel,t}$, le Système canadien des comptes macroéconomiques produit une mesure du PIB provincial et territorial à une fréquence annuelle. L'approche adoptée ici suppose donc que les données annuelles peuvent servir d'instrument pour aider à éclairer la structure de $f(.)$ lorsque la moyenne des taux de croissance mensuels de la série d'entrées disponibles est calculée pour une année civile et utilisée pour estimer les valeurs des paramètres. Les paramètres de $f(.)$ et les caractéristiques de variance de la série d'entrées mensuelles sont ensuite ajustées pour tenir compte de la différence de périodicité. Les indices mensuels de l'activité économique agrégée sont alors élaborés à partir des valeurs mensuelles estimées de $y_{mensuel,t}$.

L'utilisation d'une variable de fréquence plus basse comme instrument pour l'activité économique mensuelle réduit le nombre de degrés de liberté et introduit des problèmes liés au rythme des fluctuations mensuelles par rapport aux fluctuations annuelles. Ces problèmes auront des conséquences sur la capacité des modèles à produire une estimation mensuelle de l'activité économique agrégée. Les petits degrés de liberté et la covariance autour de la récession de 2008 auront tendance à produire des paramètres non statistiquement significatifs pour $f(.)$, si les variables explicatives ne sont pas touchées de façon importante par les fluctuations du cycle économique. En outre, les modèles peuvent avoir tendance à sélectionner un nombre plus faible d'entrées à une fréquence annuelle qu'il n'est nécessaire pour expliquer la variation mensuelle, car d'importantes fluctuations mensuelles peuvent être masquées par une agrégation à une fréquence plus faible. De plus, les changements liés aux prix, aux ventes/à la production et à l'emploi se produiront simultanément à une fréquence annuelle. Toutefois, à une fréquence mensuelle, ces fluctuations peuvent ne pas s'aligner.

Étant donné qu'une conversion de la fréquence annuelle à une fréquence mensuelle est nécessaire pour générer les valeurs estimées désirées, la stratégie de modélisation comprend certaines approches qui préfèrent capturer davantage de variation dans les données plutôt que de se concentrer uniquement sur la parcimonie des modèles. Cela ne veut pas dire que les modèles sont produits de manière ponctuelle. Au contraire, les critères de sélection, comme la maximisation d'un R carré ajusté, sont utilisés en parallèle avec des stratégies de modélisation de type général ou spécifique plus traditionnelles.

En résumé, parce que la forme fonctionnelle pour la transformation des données mensuelles en mesures agrégées de l'activité économique est inconnue, et parce que les valeurs réelles de la série $y_{mensuel,t}$ sont également inconnues, la meilleure chose possible est d'approcher le vrai $y_{mensuel,t}$. Cela signifie que la série $\hat{y}_{mensuel,t}$ ressemblera fortement à une série sur le PIB réel, mais ne sera pas une véritable mesure du PIB mensuel réel. Elle sera plutôt une estimation de l'indice d'activité économique qui correspond aux conditions macroéconomiques dans les provinces et les territoires.

3.1 Stratégie d'estimation

La fonction $f(\cdot)$ est un ensemble d'instructions pour transformer un grand nombre de données d'entrée en une seule série. Dans cet article, on suppose que la fonction peut être basée de façon approximative sur une combinaison linéaire d'entrées et que les entrées peuvent être sélectionnées soit en sélectionnant un sous-ensemble des données disponibles, soit en créant une combinaison de toutes les séries de données d'entrée. Ci-dessous, trois approches sont explorées : 1) un modèle simple; 2) l'ACP; et 3) le LASSO. Le modèle simple et le modèle LASSO relèvent de la première catégorie, tandis que l'ACP appartient à la deuxième catégorie.

Les hypothèses et la mise en œuvre des modèles sont examinées en détail ci-après. Dans toutes les stratégies de modélisation, les étapes suivantes sont suivies pour estimer les valeurs des indices dans tous les cas :

- 1 : Préparer les données d'entrée.

L'ensemble de données d'entrée comporte 1 341 séries qui sont distribuées inégalement à travers les provinces et les territoires (tableau 3). Toutes les séries n'ont pas la même utilité pour modéliser l'activité économique agrégée. Dans les cas où la désaisonnalisation a échoué, les séries sont supprimées. De même, les séries comportant des valeurs nulles sont supprimées. Ces séries ont tendance à être des séries où les valeurs nulles sont entrecoupées par des valeurs nominales. Dans ces cas, les valeurs désaisonnalisées peuvent être négatives, le taux de croissance ou les transformations des différences logarithmiques peuvent ne pas fonctionner, et les séries ont une valeur douteuse en ce qui concerne leur utilité comme indicateur permanent de l'activité économique. Dans l'ensemble, 198 variables sont abandonnées pour ces raisons.

Tableau 3
Nombre de variables d'entrée

	Vecteurs de départ	Avec 0	Échec de la DS	Non renouvelé	principa ux	principa ux	pour le LASSO	Vecteurs pour l'ACP
Terre-Neuve-et-Labrador	108	3	8	8	15	25	83	73
Île-du-Prince-Édouard	101	12	15	20	12	20	65	57
Nouvelle-Écosse	115	2	14	14	15	25	84	74
Nouveau-Brunswick	115	1	12	13	15	24	81	72
Québec	131	1	10	11	18	29	97	86
Ontario	133	0	11	11	18	30	99	87
Manitoba	120	1	6	6	17	28	94	83
Saskatchewan	118	0	11	11	16	26	85	75
Alberta	122	0	8	8	17	27	91	81
Colombie-Britannique	122	0	8	8	17	28	92	81
Yukon	63	22	20	26	6	10	31	27
Territoires du Nord-Ouest	59	27	29	31	5	7	23	21
Nunavut	46	29	29	32	2	4	11	9
Total	1 353	98	181	199	173	283	936	826

Note : LASSO : moindre contraction absolue et opérateur de sélection; ACP : analyse en composantes principales; DS : désaisonnalisation.

Source : Statistique Canada, compilation des auteurs.

L'ensemble de données d'entrée filtré contient alors 1 143 séries. Cependant, les séries sont généralement rapportées en niveaux (p. ex. heures travaillées ou ventes manufacturières en dollars enchaînés) et présentent de fortes tendances au cours de la période d'échantillonnage. Pour tenir compte des tendances, les séries sont transformées en taux de croissance d'un mois à l'autre. Ces taux de croissance seront finalement comparés aux mesures de la croissance du

PIB réel, et ils ont l'avantage d'être liés par -100 % pour le déclin maximum⁴. Les taux de croissance pour la série sont censés être stationnaires en matière de covariance.

Pour utiliser la série dans l'estimation, toutes les séries sont centrées et mises à l'échelle pour avoir une variance au niveau des unités à une fréquence mensuelle. Ce processus de normalisation est appliqué aux variables afin d'éviter que des variables dont les valeurs unitaires sont naturellement plus élevées n'affectent les résultats. Étant donné que les séries chronologiques mensuelles peuvent avoir une forte variabilité, et parce que les périodes de chocs économiques comme les récessions peuvent produire des points de données aberrants, toutes les séries sont « winsorisées » (ou tronquées des valeurs extrêmes supérieures et inférieures) avant l'estimation. Cela empêche les points de données extrêmes d'affecter les résultats. Pour créer les indices mensuels, les valeurs des paramètres provenant des modèles basés sur les données obtenues par winsorisation sont combinées avec des données non-winsorisées, ce qui permet aux valeurs plus importantes d'avoir toute leur influence lorsque de grands chocs surviennent.

Enfin, les séries les plus bruyantes sont supprimées. Dans le cas de l'ACP, la tranche supérieure de 25 % des séries selon la variance est éliminée par province, tandis que pour le LASSO, la tranche supérieure de 15 % des séries selon la variance est éliminée. Ces seuils sont arbitraires, mais leur imposition a permis d'améliorer la capacité des modèles d'éclairer sur l'activité agrégée (c.-à-d. améliorer le signal par rapport au bruit) et d'améliorer la cohérence des résultats entre les méthodes. Après ajustement pour les séries à variation élevée, il y a un total de 928 variables d'entrée pour le LASSO et 820 pour l'ACP. Le Nunavut possède le plus petit nombre de séries disponibles, tandis que l'Ontario et le Québec en comptent le plus grand nombre.

- 2 : En utilisant les taux de croissance winsorisés, calculer la moyenne annuelle des taux de croissance mensuels qui seront utilisés dans les modèles.
- 3 : Estimer les paramètres du modèle.

Cette estimation est fondée sur la combinaison de la croissance du PIB réel et des moyennes annuelles des séries mensuelles de données d'entrée pour les années 2002 à 2018. La croissance réelle du PIB n'est ni mise à l'échelle ni centrée. En utilisant la croissance réelle du PIB comme variable cible et les moyennes annuelles des séries mensuelles comme variables d'entrée, la forme fonctionnelle de $f(\cdot)$ est estimée. Dans tous les modèles employés, on suppose qu'une combinaison linéaire de variables d'entrée peut être utilisée pour transformer la multiplicité des variables d'entrée en une mesure unique de l'activité agrégée. Dans le cas du modèle simple et du LASSO, un sous-ensemble des variables est utilisé directement. Dans le cas de l'ACP, les 10 premières composantes principales sont utilisées comme point de départ. On présume aussi que les MCO peuvent être utilisés pour générer des contributions permettant de combiner des variables d'entrée. En utilisant des méthodes de régression pour combiner les données d'entrée, on impose une hypothèse supplémentaire comme quoi les structures économiques sont, en moyenne, les mêmes pendant toute la période d'échantillonnage.

- 4 : Utiliser le modèle pour estimer les taux de croissance mensuels.

Étant donné que toutes les approches peuvent être considérées comme une régression des MCO avec des données d'entrée centrées, l'interception peut être interprétée comme le taux de croissance annuel moyen du PIB réel entre 2002 et 2018. Les données d'entrée sélectionnées (séries chronologiques mensuelles ou composantes principales) produisent des fluctuations pour

4. Il est plus traditionnel d'utiliser les différences logarithmiques dans l'analyse des séries chronologiques. Dans ce cas, les taux de croissance sont utilisés parce que le taux de changement est limité à -100 %, ce qui n'est pas le cas avec les différences logarithmiques. La mise en œuvre d'un examen approprié des propriétés de stationnarité de la série des données d'entrée demeure une priorité pour les travaux futurs.

ce taux de croissance moyen. Pour 2019 et 2020, on suppose que le taux de croissance moyen entre 2002 et 2018 est représentatif de la croissance sous-jacente.

Pour produire des estimations mensuelles, il est nécessaire d'ajuster les estimations des paramètres ou les séries mensuelles pour tenir compte de la différence à l'égard de la périodicité. La constante du modèle est ajustée à une fréquence mensuelle basée sur le taux de croissance mensuel composé qui est équivalent à l'estimation annuelle :

$$\hat{\beta}_{0,mensuel} = (\hat{\beta}_{0,annuel} + 1)^{1/12} - 1$$

Pour estimer les fluctuations mensuelles autour du taux de croissance tendanciel, les séries de données d'entrées brutes sont utilisées. Cela permet aux fluctuations importantes dans la série chronologique de présenter toutes leurs répercussions lorsque les chocs économiques, comme les récessions ou les cycles de prix des produits de base, ont des répercussions sur les économies provinciales et territoriales. Les variations des données d'entrée mensuelles sont ajustées pour correspondre à l'écart annuel avant utilisation. Les estimations de régression des MCO sont basées sur le ratio $\text{cov}(x,y)/\text{var}(x)$. Dans le contexte actuel, l'agrégation au fil du temps réduit la variance de la matrice X. Pour tenir compte de cela, la variance des données mensuelles est rééchelonnée de façon à correspondre à la variance des données annuelles pour chaque série avec l'équation suivante :

$$\sigma_{mensuel,ajusté} = \left(\frac{(x - \mu_x)}{\sigma_{x,mensuel}} \right) * \sigma_{x,annuel} + \mu_x$$

- 5 : Générer les indices de niveau

Les valeurs ajustées provenant des modèles sont des estimations de la croissance mensuelle de l'activité économique. Elles peuvent être transformées en indices en ajoutant 1 pour créer une valeur de liaison pour un indice en chaîne. Le niveau de l'indice est ensuite calculé en enchaînant les valeurs à partir de janvier 2002.

Les estimations du taux de croissance sont associées à un intervalle de confiance, car il existe une incertitude quantifiable qui découle du modèle. Il existe également une incertitude non quantifiable qui découle d'une possible erreur de spécification du modèle. Pour produire un indice de niveau, il est nécessaire de supposer que les estimations du taux de croissance sont suffisamment précises pour pouvoir être utilisées pour l'enchaînement, même si les erreurs sont aggravées au fil du temps. Il s'agit là d'une hypothèse solide, mais elle est compatible avec la façon dont les valeurs moyennes des données d'enquête sur les prix, les valeurs et les quantités sont combinées pour produire des indices de quantité enchaînée ou de prix enchaînés.

L'utilisation de plusieurs modèles à l'étape 3 conduira finalement à la présentation de différents types de l'indice d'activité. Dans le contexte actuel, où les modèles statistiques sont utilisés pour donner des renseignements sur l'activité économique dans un environnement où ils ne peuvent être mis en œuvre de manière optimale, la création de plusieurs versions de l'indice de l'activité joue un rôle important. Parce que la vraie valeur pour $y_{monthly,t}$ est inconnue, valider $\hat{f}(\cdot)$ et $\hat{y}_{monthly,t}$ est difficile. Les données de sortie des différentes méthodes fournissent une forme naturelle de confrontation des données qui aide à évaluer l'acceptabilité et la généralisation des estimations.

3.1.1 Modèle simple

Le modèle simple part du principe que l'emploi total, les exportations totales et les ventes totales au détail contiennent les renseignements appropriés pour comprendre les fluctuations économiques agrégées. Cette hypothèse est probablement trop hardie, car plus de trois données

d'entrée sont nécessaires pour saisir pleinement les complexités de l'activité économique agrégée. Toutefois, le modèle est cohérent dans l'ensemble des provinces et des territoires, et il est facile de le comprendre. Il est donc utile en tant que base contre laquelle des méthodes plus complexes peuvent être évaluées. L'approche simple représente également une méthode qui peut être considérée comme compatible avec les types de bases de projection utilisées pour déduire les mouvements de la production brute qui sont utilisés comme données d'entrée pour le PIB mensuel du Canada (Statistique Canada, [2020b](#)).

Comme les séries n'ont pas de structure naturelle d'agrégation pour les combiner, les régressions sont utilisées pour déterminer les contributions relatives des variables plutôt qu'une formule de calcul de l'indice. Étant donné que les données d'entrée sont censées être les données d'entrée nécessaires, les trois séries sont incluses indépendamment de leur importance statistique en matière de régressions.

3.1.2 Analyse des composantes principales

L'ACP est une technique de réduction variable qui vise à expliquer la variance d'un ensemble de données particulier en utilisant un plus petit nombre de composantes principales (OCDE [2008](#), Jolliffe [2002](#)). Un ensemble de données avec des variables p $X = [x_1, x_1, \dots, x_p]$, peut être transformé pour produire les composantes principales p :

$$Z = AX$$

où

$$z_1 = a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,p}x_p$$

$$z_2 = a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,p}x_p$$

$$z_p = a_{p,1}x_1 + a_{p,2}x_2 + \dots + a_{p,p}x_p$$

Les composantes principales sont calculées comme des combinaisons linéaires des variables d'entrée (la série mensuelle). La première composante principale explique la plus importante proportion de la variance des variables d'entrée. C'est le vecteur propre associé à la plus grande valeur propre. La deuxième composante principale est orthogonale (sans corrélation) avec la première composante principale et est le vecteur propre associé à la deuxième plus grande valeur propre. Elle explique la deuxième composante la plus importante des variables d'entrée. On peut donc dire qu'elle mesure une différente dimension statistique des séries disponibles. La troisième composante principale est orthogonale aux deux premières et mesure la troisième plus grande proportion de variance dans les données. Et ainsi de suite jusqu'à la p^e composante principale.

Pour mettre en œuvre l'ACP ici, les principales composantes utilisées pour l'estimation du modèle et les chargements sont déterminées en employant la série chronologique mensuelle de la variance winsorisée et centrée au niveau des unités. Les chargements sont ensuite appliqués à la série brute non winsorisée pour produire les composantes principales brutes utilisées pour prédire les taux de croissance mensuels.

Lorsque l'ACP fonctionne bien, une grande partie de la variance d'un ensemble de données peut être expliquée par les premières composantes principales, et seules les premières composantes principales sont utilisées pour l'analyse. Malheureusement, dans le cas de l'ensemble de données d'entrée pour les économies provinciales et territoriales, l'ACP ne fonctionne pas bien pour réduire la portée des renseignements contenus dans l'ensemble de données (tableau 4). La première composante principale représente généralement moins de 10 % de la variation dans l'ensemble de données. De plus, lorsqu'on établit une moyenne pour produire une estimation de

la fréquence annuelle, la première composante principale ne correspond pas bien à la croissance annuelle du PIB réel pour la plupart des provinces et des territoires (tableau 5).

Tableau 4
Pourcentage de variation par composante principale

	Composantes principales									
	1	2	3	4	5	6	7	8	9	10
Terre-Neuve-et-Labrador	10,5	16,2	21,0	25,4	29,7	33,3	36,6	39,7	42,6	45,3
Île-du-Prince-Édouard	12,1	19,5	25,7	31,2	35,2	39,1	42,7	46,0	49,2	52,2
Nouvelle-Écosse	9,0	15,2	20,6	24,9	28,5	32,0	35,2	38,3	41,1	43,8
Nouveau-Brunswick	9,6	16,4	21,3	25,8	29,8	33,6	37,3	40,4	43,4	46,3
Québec	7,0	13,0	17,8	21,9	25,5	28,7	31,6	34,3	37,0	39,7
Ontario	8,9	14,6	19,3	23,1	26,4	29,6	32,5	35,3	37,8	40,2
Manitoba	7,9	13,6	19,2	23,8	27,3	30,7	34,0	37,2	39,9	42,6
Saskatchewan	8,6	14,5	19,7	24,4	28,7	32,0	35,1	38,0	40,9	43,6
Alberta	10,5	16,4	21,5	25,9	29,5	33,0	36,1	39,0	41,7	44,3
Colombie-Britannique	8,6	14,7	20,2	25,3	29,4	32,8	35,8	38,6	41,3	44,0
Yukon	14,2	24,8	33,8	39,8	45,1	50,2	54,9	59,3	63,4	67,5
Territoires du Nord-Ouest	17,8	30,6	37,6	44,3	50,3	56,1	61,4	66,6	71,4	76,1
Nunavut	35,8	49,4	62,6	73,6	83,7	92,5	98,1	99,6	100,0	...

... sans objet.

Source : Statistique Canada, compilation des auteurs.

Les corrélations indiquent qu'en dehors de l'Alberta, de la Colombie-Britannique et de l'Ontario, le fait d'utiliser uniquement la première composante principale ne produira pas un indice d'activité qui fournit une mesure adéquate pour déterminer le rendement des provinces et des territoires en fonction des fluctuations mensuelles agrégées. Par conséquent, un indice d'activité basé uniquement sur la première composante principale, comme celui produit par le la Federal Reserve Board of Chicago (Federal Reserve Board of Chicago [2020](#), Brave et Butters [2010](#), Evans et Pham-Kanter [2002](#)), n'est pas poursuivi ici.

Tableau 5
Corrélation entre la première composante principale et la croissance annuelle du produit intérieur brut (PIB) réel

	Indice winsorisé	Indice non winsorisé
Terre-Neuve-et-Labrador	-0,349	-0,256
Île-du-Prince-Édouard	-0,199	-0,307
Nouvelle-Écosse	0,531	0,491
Nouveau-Brunswick	-0,709	-0,673
Québec	0,657	0,661
Ontario	0,775	0,778
Manitoba	0,562	0,567
Saskatchewan	0,477	0,338
Alberta	0,937	0,940
Colombie-Britannique	0,840	0,848
Yukon	0,177	0,251
Territoires du Nord-Ouest	0,423	0,422
Nunavut	0,355	0,444

Source : Statistique Canada, compilation des auteurs.

Bien que la première composante principale ait des difficultés à corrélérer avec les fluctuations annuelles du PIB réel, cela ne signifie pas qu'il n'y a aucun renseignement dans les premières composantes principales pour expliquer la croissance du PIB réel. Par conséquent, pour générer un modèle basé sur les composantes principales, les régressions sont effectuées sur toutes les

combinaisons des 10 premières composantes principales en tant que variables explicatives pour expliquer la croissance du PIB réel. La régression qui maximise le R carré ajusté est ensuite sélectionnée en tant que modèle préféré. Cela produit 13 modèles qui fonctionnent raisonnablement bien pour expliquer la croissance du PIB réel. De plus, les modèles fonctionnent généralement bien pour expliquer la récession de 2008 et d'autres fluctuations propres à une province ou à un territoire.

3.1.3 Moindre contraction absolue et opérateur de sélection

Le LASSO est la solution à un problème d'optimisation restreint semblable aux MCO. Sous une régression linéaire classique, $X = [x_1, x_2, \dots, x_p]$ est une matrice $n \times p$ contenant les variables prédictives utilisées pour expliquer la variation dans un vecteur cible y de longueur n . Les coefficients pour la régression $\hat{\beta} = [\hat{\beta}_0, \dots, \hat{\beta}_p]$ sont alors la solution au problème qui cherche à réduire la somme des erreurs quadratiques entre y et une combinaison linéaire des variables dans X :

$$\hat{\beta}_{OLS} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$$

Le LASSO (Tibshirani [1996](#)) fait partie d'une classe d'estimateurs qui cherche à pénaliser l'estimateur des MCO pour ajustement excessif (c.-à-d. inclure trop de variables) au moyen de son paramètre de régulation λ . On pourrait comparer cela à l'utilisation d'un R carré ajusté ou d'un critère de renseignement pour pénaliser l'inclusion d'un trop grand nombre de variables explicatives. Cependant, cela va plus loin que de pénaliser à cause de variables explicatives supplémentaires quand on regarde la qualité du modèle. Cela sélectionne des variables pertinentes. Le LASSO est la solution pour :

$$\hat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2 + \lambda \sum_{j=1}^p |\beta_j| \right]$$

Le paramètre $\lambda \geq 0$ contrôle la force de la pénalité; plus la valeur de lambda est grande, plus la valeur du rétrécissement est importante. L'algorithme LASSO peut uniquement inclure des valeurs pour β_j jusqu'à un total absolu précis. Par conséquent, le LASSO fixe à 0 les coefficients variables moins importants. On peut considérer cela comme semblable au type de résultat trouvé en employant une stratégie de modélisation générale à spécifique/guidée par descriptions approximatives, mais qui est applicable à plus grande échelle. Le résultat est une méthode pour traiter de grands ensembles de données où un grand nombre de prédicteurs peuvent être inclus, et l'algorithme sélectionnera ceux dont les propriétés de covariance sont les plus importantes pour prédire la variable cible (y).

Le LASSO a lui aussi des limites. Dans les cas où les groupes de variables prédictives sont fortement corrélés les uns aux autres, le LASSO tend à garder une variable de chaque groupe et à contracter le coefficient des autres variables à zéro. Et dans d'autres cas, lorsque l'ensemble de données a un petit n et un grand p , le LASSO sélectionne au maximum n variables avant d'être saturé. Cependant, il peut y avoir plus de variables n avec un coefficient qui n'est pas égal à zéro dans le vrai modèle.

La méthode elastic-net (Zou et Hastie, [2005](#)) est une extension du LASSO. En contrôlant le poids de pénalité α , le modèle elastic-net stabilise la sélection des variables à partir d'un groupe de variables corrélées et supprime la limite au nombre de variables sélectionnées. Les coefficients sont estimés comme suit :

$$\hat{\beta}_{EN} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2 + \lambda \sum_{j=1}^p \left[\frac{1}{2} (1 - \alpha) \beta_j^2 + \alpha |\beta_j| \right] \right\}$$

Où $0 \leq \alpha \leq 1$ est la pondération de pénalité. Comme α est égal à 1, le modèle elastic-net est le même que pour le modèle LASSO et, comme α est proche de 1, le modèle elastic-net se comporte de façon semblable au LASSO, mais élimine le comportement problématique causé par de fortes corrélations entre les variables.

Les données de sortie du LASSO en matière du nombre de variables sélectionnées et de leur importance statistique ont été soigneusement étudiées. Dans presque tous les cas où le LASSO fonctionnait, le LASSO semble inclure dans le modèle des variables qui ne sont pas statistiquement importantes. Pour s'assurer que la relation entre le y cible et les variables explicatives est justifiable avec un meilleur résultat statistique, une analyse graduelle par régression avec sélection descendante est utilisée sur les variables sélectionnées par le LASSO pour enlever les variables non significatives du modèle.

L'analyse graduelle par régression est une méthode qui examine l'importance statistique de chaque variable indépendante au sein du modèle. Elle construit un modèle en ajoutant successivement (sélection ascendante) ou en supprimant (sélection descendante) des variables en fonction des statistiques t de leurs coefficients estimés. La méthode d'élimination descendante commence par l'inclusion de toutes les variables dans le modèle, puis chaque variable est supprimée une par une, pour tester son importance. Les variables qui ne sont pas statistiquement significatives sont enlevées du modèle.

Le modèle LASSO n'a pas choisi de variables pour le Nouveau-Brunswick, la Nouvelle-Écosse, l'Ontario et les Territoires du Nord-Ouest. La méthode elastic-net est utilisée à sa place pour ces provinces et territoires. Et pour les deux autres territoires, le Yukon et le Nunavut, une analyse graduelle par régression manuelle est effectuée.

Dans les deux méthodes, LASSO et elastic-net, la validation croisée du progiciel caret R est utilisée pour ajuster les paramètres lambda et alpha. La validation croisée utilise une technique de prévision continue de l'origine (Hyndman et Athanasopoulos, [2014](#)) au lieu de l'échantillonnage aléatoire simple. Cette technique est propre aux ensembles de données des séries chronologiques.

4 Évaluation mensuelle de l'indice

Les trois approches présentent en outre des points forts et des points faibles qui leur sont propres, et qui ont une incidence sur leur utilisation (tableau 6). Les indices simples et le LASSO ont pour avantage que leurs modèles sont parcimonieux et que les indices qu'ils produisent sont moins bruyants que les indices fondés sur l'ACP. Cependant, ces indices sont basés sur un ensemble de variables très réduit, qui sont souvent non statistiquement significatives pour les indices simples dans les régressions annuelles. Ces indices tendent aussi à mettre l'accent sur les séries relatives à l'emploi plutôt que sur un large éventail d'activités économiques et peuvent, par conséquent, ne pas présenter de valeurs explicatives idéales des fluctuations mensuelles de l'activité, si les variations de la production ne rendent pas compte au même moment des variables d'emploi.

L'indice ACP a pour avantage que la méthodologie est saine et bien comprise. Il fonctionne pour toutes les provinces et tous les territoires. Cependant, il produit les indices d'activité les plus bruyants, rendant leur interprétation difficile; de plus, dans certains cas (p. ex. T.-N.), l'indice peut diminuer brusquement. Les indices ACP sont aussi combinés en fonction de la maximisation du

R carré ajusté d'une régression à l'autre. Cela produit une combinaison linéaire des composantes principales qui sont statistiquement significatives et non significatives. Ces inclusions préfèrent ajouter des renseignements supplémentaires qui incluent un certain bruit, car il n'est pas clair que les données à une fréquence annuelle représentent la variabilité d'un mois à l'autre.

Tableau 6
Caractéristiques des méthodes d'estimation d'indice

Critère	Indice simple	Indice ACP	Indice pondéré	Indice LASSO
Données d'entrée cohérentes pour tous les niveaux géographiques	Oui	Non	Non	Non
Modèle-types cohérents pour tous les niveaux géographiques	Oui	Oui	Oui	Non
Spécification du modèle	Trois données d'entrée, quelques variables non significatives	Nombre variable de composantes principales. Quelques variables non significatives	Combinaison des indices simples et ACP	Sélection de données d'entrée variable
Adéquation du modèle	Adéquation du modèle peut varier entre les provinces et les	Adéquation généralement bonne pour l'échantillon	Adéquation pour l'échantillon améliorée par rapport aux indices	Adéquation généralement bonne pour l'échantillon
Intelligibilité	Données d'entrée et contributions faciles à comprendre	Difficile de comprendre ce qui contribue aux variations Difficile d'interpréter les composantes principales	Difficile de comprendre ce qui contribue aux variations	Données d'entrée fondées sur des corrélations Contributions pouvant être interprétées Indice à faible variance
Adéquation du modèle	Ces modèles peuvent ne pas être efficaces selon l'importance statistique Les données d'entrée reflètent les attentes pour les variables importantes.	Ces modèles peuvent ne pas être efficaces selon l'importance statistique Utilisation exhaustive des données d'entrée	Hérite des propriétés des indices de données d'entrée	Approche de modélisation non adaptée à la configuration actuelle

Note : ACP : analyse en composantes principales; LASSO : moindre contraction absolue et opérateur de sélection.

Source : Statistique Canada, compilation des auteurs.

La combinaison des indices fournit une méthode supplémentaire pour leur utilisation. Puisque l'indice simple est relativement stable, mais se concentre sur un nombre limité de séries fondamentales, et que l'indice ACP est plus variable, mais comprend des combinaisons linéaires de toutes les entrées, ces séries sont combinées pour fournir un indice pondéré présentant de meilleures caractéristiques que les composantes. Pour toutes les méthodes, les indices produits génèrent généralement des types similaires de renseignements sur les cycles économiques et les principaux chocs économiques dans les provinces et les territoires. Comme pour les coefficients de régression, la croissance annuelle du PIB réel est utilisée comme comparaison, car elle constitue la principale source d'activité économique agrégée qui est disponible pour les provinces et les territoires. Pour combiner l'indice ACP et l'indice simple, les valeurs de degré de liberté entre 1 % et 100 % sont utilisées pour créer des indices pondérés comme suit :

$$\text{indice_pondéré} = (1 - v) * \text{indice_simple} + v * \text{indice_d'activité}$$

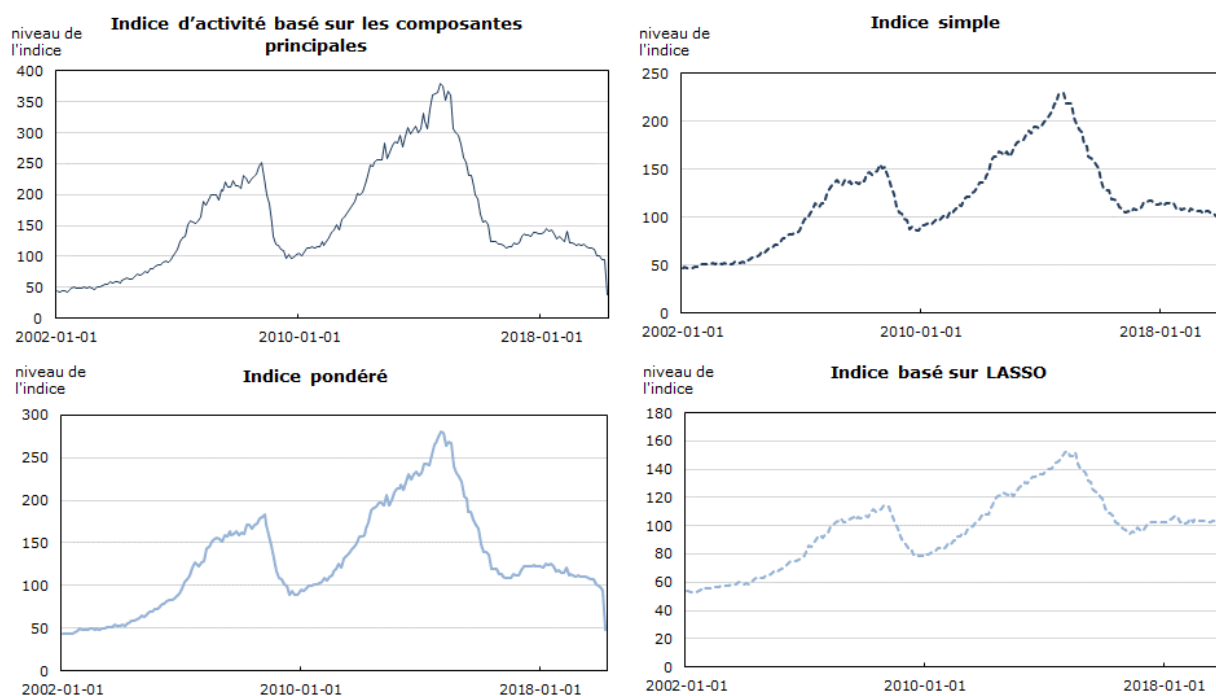
Le degré de liberté correspondant à l'indice pondéré qui a la plus forte corrélation avec la croissance du PIB réel est alors sélectionné.

Les méthodes utilisées pour produire des indices mensuels donnent généralement des types semblables de renseignements sur les cycles économiques et les chocs économiques majeurs dans les provinces et les territoires. À titre d'exemples, les indices pour l'Alberta (panel 1) et Terre-Neuve-et-Labrador (panel 2) sont présentés ci-dessous.

Dans le cas d'une économie plus vaste, comme celle de l'Alberta, toutes les approches donnent des renseignements similaires sur les périodes de croissance ou de déclin, mais l'ampleur des cycles peut varier selon la méthodologie. En général, les indices basés sur l'ACP ont la plus grande variabilité, alors que l'indice simple a le moins de variabilité. Dans certains cas, comme l'indice ACP pour Terre-Neuve-et-Labrador, le modèle ne produit pas un résultat raisonnable. Dans ces cas, l'indice ne sera pas rendu disponible et est jugé impropre à l'utilisation. Néanmoins, lorsque les indices semblent présenter les caractéristiques appropriées, il y a une forte corrélation entre les mesures de l'activité économique implicite et le mouvement des indices dans le temps correspond à ce que l'on sait sur la performance économique des provinces et des territoires.

De plus, les comparaisons avec les estimations sous-annuelles du PIB réel pour l'Ontario et le Québec montrent que les taux de croissance d'une année à l'autre sont fortement corrélés, mais que les cycles économiques peuvent être accentués dans les indices d'activité. Les indices semblent donc saisir des renseignements pertinents pour les cycles économiques, les périodes de croissance plus forte ou plus faible et pour comprendre la performance économique. Ils ne correspondent toutefois pas directement au PIB réel et ne doivent pas être interprétés comme une mesure directe du PIB réel.

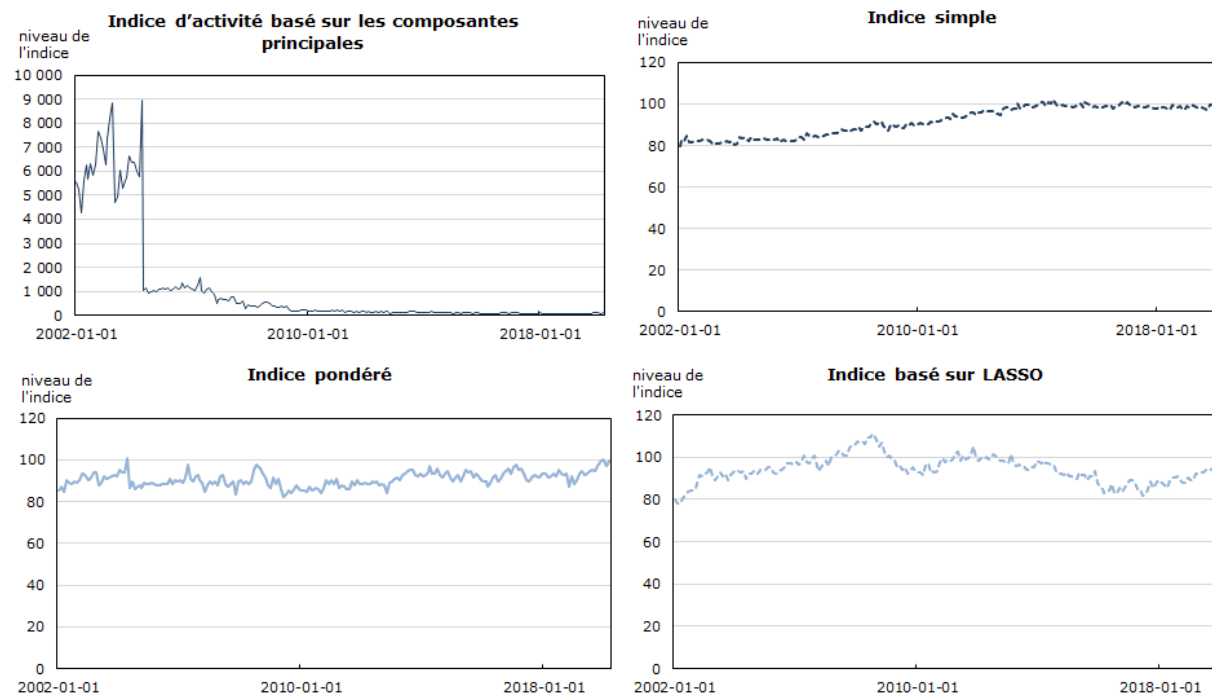
Graphique 1
Alberta : indices d'activité; degré de liberté = 0,39



Note : LASSO = moindre contraction absolue et opérateur de sélection.
Source : Statistique Canada, calculs des auteurs.

Graphique 2

Terre-Neuve-et-Labrador : indices d'activité; degré de liberté = 0,29



Note : LASSO = moindre contraction absolue et opérateur de sélection.
Source : Statistique Canada, calculs des auteurs.

5 Conclusion

Des mesures de l'activité économique agrégée pour les économies sont essentielles pour éclairer les décisions relatives aux politiques fiscales et financières, déterminer les caractéristiques des cycles économiques et examiner la performance économique. Dans cette étude, quatre indices d'activité économique provinciale et territoriale fondés sur différentes approches méthodologiques sont estimés et présentés. Celles-ci reposent sur un modèle simple; l'ACP; une combinaison pondérée des indices simples et de l'ACP; et le LASSO. Dans la plupart des cas, toutes les approches produisent des résultats à peu près semblables. Toutefois, le caractère cyclique et la variance des fluctuations d'un mois à l'autre peuvent être très différents. En règle générale, l'ACP produit la plus grande variabilité et les cycles les plus prononcés, alors que l'indice simple est le plus stable.

Selon les propriétés des méthodologies et de leurs données de sortie, l'indice simple est le plus uniforme entre les provinces et territoires. Il est aussi le plus facile à interpréter en termes de contributions de variables et de justification de l'inclusion de variable. Cependant, les valeurs des paramètres sont souvent non statistiquement significatives et les séries de données d'entrée sont choisies autant pour leur importance économique que pour leur présence dans toutes les provinces et tous les territoires. Par conséquent, ces modèles offrent une approche plus limitée pour l'examen de l'activité économique agrégée, mais ils constituent également une base pour les comparaisons avec des modèles plus complexes.

Les indices fondés sur l'ACP semblent donner une idée plus complète de l'évolution de l'activité au fil du temps, mais on ne sait pas encore comment les principaux éléments devraient être interprétés. À cause de cela, et parce que les indices de l'ACP présentent la plus grande variabilité, ils présentent un compromis entre l'utilisation globale des séries de données d'entrée et l'intelligibilité.

La pondération de l'indice simple et de l'indice ACP produit un résultat qui a une corrélation supérieure avec les fluctuations annuelles du PIB. La combinaison pondérée continue d'avoir plus de variabilité que l'indice simple. Étant donné que l'ACP est incluse, elle n'est pas aussi facile à interpréter que l'indice simple, mais elle fournit probablement une meilleure mesure de l'activité agrégée que ses composantes.

L'indice LASSO obtient de bons résultats en ce qui concerne le PIB réel annuel, mais la configuration du modèle n'est pas aussi adaptée à la situation rencontrée lors de l'estimation des indices d'activité. En particulier, le nombre relativement faible d'observations limite la capacité des algorithmes à effectuer une validation croisée. De plus, bien que les séries de données d'entrée constituent un sous-ensemble distinct de l'ensemble de données d'entrée et que leurs contributions puissent être générées de façon simple et directe, il n'y a aucune raison théorique pour expliquer pourquoi les variables sont importantes, ce qui limite l'intelligibilité du modèle.

Compte tenu des forces et des faiblesses présentes entre l'adéquation des modèles, leur rendement et l'examen de leurs données de sortie, les évaluations effectuées jusqu'à présent suggèrent que les indices simples ou les indices LASSO présentent des résultats liés à un ensemble de données d'entrée fondamentales (souvent fortement influencées par les séries sur l'emploi), que les indices ACP se rapportent davantage à une forme quelconque d'activité à court terme (mais le signal est bruyant) et que l'indice pondéré présente un compromis entre les deux.

Les indices tels qu'ils sont estimés actuellement sont corrélés aux mesures annuelles du PIB réel et aux mesures sous-annuelles du PIB réel pour l'Ontario et le Québec, mais ils ne doivent pas être interprétés comme une mesure du PIB réel. Les indices présentent une plus grande variabilité et un caractère cyclique plus prononcé que les mesures du PIB réel et sont constitués de mesures de la production brute, de l'emploi, de prix relatifs et de ratios importants, comme le

taux de chômage. Cela rend les indices appropriés pour comprendre l'activité économique, mais ils ne sont cependant pas équivalents au PIB réel. De plus, les indices ne donnent pas de renseignements sur les différents niveaux d'activité économique entre les provinces et les territoires.

Les indices sont également basés sur un ensemble de données d'entrée et des stratégies de modélisation qui ne sont pas idéales. De nombreuses hypothèses doivent être imposées pour produire les indices, et n'importe laquelle peut s'avérer être une source d'erreurs de mesure importantes. Par conséquent, les indices présentés ici doivent être considérés comme expérimentaux et peuvent être révisés ou remplacés à mesure que les recherches futures améliorent les processus et/ou testent les hypothèses pour déterminer leur validité.

À l'heure actuelle, les corrélations entre les différentes méthodes, leur corrélation positive avec les mesures du PIB réel infra-annuel produites à l'échelon provincial et les comparaisons des propriétés avec la performance économique provinciale et territoriale connue soutiennent leur utilisation en tant qu'indicateurs de cycles économiques, pour comprendre l'ampleur des chocs par rapport à l'histoire d'une province ou d'un territoire ainsi que l'évolution des économies régionales. Les comparaisons interprovinciales sont également soutenues, mais avec la mise en garde que l'efficacité du modèle est difficile à comprendre dans toutes les situations et que les comparaisons à divers niveaux entre les provinces ne sont pas possibles à partir des valeurs d'indice.

Références

Brave, Scott, et R. Andrew Butters. 2010. « Chicago Fed National Activity Index Turns Ten—Analyzing Its First Decade of Performance. » *Chicago Fed Letter*, n° 273 (avril). Federal Reserve Bank of Chicago. <https://www.chicagofed.org/~media/publications/chicago-fed-letter/2010/cflapril2010-273-pdf.pdf>.

Statistique Canada. 2020a. « Produit intérieur brut (PIB) aux prix de base, par industries, mensuel (36 100 434). » Statistique Canada. [Statistique Canada, le 24 janvier 2020. https://www150.statcan.gc.ca/n1/pub/13-607-x/2016001/230-fra.htm](https://www150.statcan.gc.ca/n1/pub/13-607-x/2016001/230-fra.htm) (consulté le 2 juin 2020)

Statistique Canada. 2020b. « Produit intérieur brut (PIB) aux prix de base, par industries, mensuel (36 100 434). » Statistique Canada. [Statistique Canada, le 31 juillet 2019. https://www.statcan.gc.ca/fra/programmes-statistiques/document/1301_D1_V3](https://www.statcan.gc.ca/fra/programmes-statistiques/document/1301_D1_V3) (consulté le 2 juin 2020)

Federal Reserve Board of Chicago. 2020. « Chicago Fed National Activity Index (CFNAI) Current Data. » Federal Reserve Board of Chicago, 22 juin 2020. <https://www.chicagofed.org/research/data/cfnai/current-data> (consulté le 2 juin 2020).

Evans, Liu, Charles L. et Genevieve Pham-Kanter. 2002. « The 2001 Recession and the Chicago Fed National Activity Index: Identifying Business Cycle Turning Points. » *Economic Perspectives* 26 (3). Federal Reserve Bank of Chicago : p. 26 à 43. <https://www.chicagofed.org/~media/publications/economic-perspectives/2002/3qepart2-pdf.pdf>.

Hyndman, R.J., et Athanasopoulos, G. 2018. *Forecasting: principles and practice*, 2^e édition, OTexts : Melbourne, Australie. OTexts.com/fpp2. Consulté le 2 juin 2020.

Jolliffe, I.T. 2002. *Principle Components Analysis Second Edition*. Springer-Verlag New York Inc., New York, NY.

Nations Unies (ONU), Commission européenne (CE). Fonds monétaire international (FMI), Organisation de coopération et de développement économiques (OCDE) et Banque mondiale (BM). 2009. *Système de comptabilité nationale, 2008*. New York : Nations Unies. Disponible à : <https://unstats.un.org/unsd/nationalaccount/docs/SNA2008FR.pdf> (consulté le 2 juin 2020).

Organisation de coopération et de développement économiques (OCDE). 2008. *Handbook on Constructing Composite Indicators Methodology and User Guide*. Organisation de développement économique. <https://www.oecd.org/sdd/42495745.pdf> (consulté le 2 juin 2020).

Tibshirani, Robert. 1996. « Regularization Shrinkage and Selection via the Lasso. » *Journal of Royal Statistical Society: Series B*.

Zou, Hui et Trevor Hastie. 2005. « Regularization and Variable Selection via the Elastic Net. » *Journal of Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–20. doi:[10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x).