UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRÉVISION RÉTROSPECTIVE DES ÉMISSIONS DES GAZ À EFFET DE SERRE AU CANADA

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN ÉCONOMIQUE

PAR

JORGE EDUARDO MANCIO LÓPEZ

AOÛT 2024

REMERCIEMENTS

Je tiens à remercier mes directeurs de recherche, MM. Charles Séguin et Dalibor Stevanovic, pour leurs conseils tout au long du processus du mémoire, pour leur patience et pour le soutien financier qu'ils m'ont accordé. Le thème du mémoire a grandement aidé ma candidature lors du processus d'embauche du poste que j'occupe en ce moment, alors je leur serai toujours reconnaissant de m'avoir accordé la possibilité de réaliser l'étude de cette question de recherche.

D'autre part, je profite de l'occasion que cet ouvrage m'accorde pour souligner les efforts et les sacrifices que mes parents ont eu à faire depuis notre arrivée au pays. Étant eux aussi diplômés universitaires, les dernières années de mon cheminement académique m'ont permis de mieux valoriser leur décision de rester au Canada malgré le fait que leur scolarité et leurs compétences professionnelles n'étaient pas reconnues. Je ne pourrai jamais les remercier assez pour l'amour inconditionnel qu'ils ont accordé à leurs deux fils et pour toujours prioriser notre bien avant le leur. De plus, je veux prendre un instant pour mentionner que je suis reconnaissant envers mon frère pour ses conseils et son soutien absolu; grâce à lui je ne serai plus jamais seul dans ce monde. Finalement, à la suite de l'aboutissement de mon parcours académique, j'ai une pensée pour mes grands-parents qui m'ont transmis les valeurs qui guident mes actions et pour avoir été en partie responsables d'élever la personne que je suis aujourd'hui.

TABLE DES MATIÈRES

REMERCIEMENTS ii

LISTE DES FIGURES vi

LISTE DES TABLEAUX vii

RÉSUMÉ viii

INTRODUCTION 1

CHAPITRE 1 REVUE DE LA LITTÉRATURE 3

1.1 Premier article de référence : le modèle structurel à facteurs dynamiques 4

1.2 Deuxième article de référence : l'apprentissage automatique comme modèle de prévision 6

1.3 Contribution du mémoire à la littérature 6

CHAPITRE 2 CADRE ÉCONOMÉTRIQUE 8

2.1 L'enjeu de la prévision rétrospective 8

2.2 Le modèle de référence 9

2.3 Modèle autorégressif à retards échelonnés (ADL) 9

2.4 Modèle autorégressif augmenté de facteurs latents (ARFL) 10

2.5 Critères d'évaluation des prévisions 12

CHAPITRE 3 LES DONNÉES 16

3.1 La variable d'intérêt 16

3.2 Le panel contenant les variables des spécifications du modèle ADL 18

3.3 Les bases de données pour l'estimation des facteurs latents 19

3.4 Les transformations de la variable d'intérêt 21

CHAPITRE 4 RÉSULTATS 24

4.1 Méthodologie 24

4.2 Résultats de l'exercice pseudo hors échantillon récursif 25

4.2.1 Le modèle de référence – Le modèle AR(1) 25
4.2.2 Le modèle ADL – Le choix logique du PIB réel 26
4.2.3 Le modèle ADL – Spécification exploitant la consommation de pétrole 28
4.2.4 Modèle autorégressif augmenté de facteurs latents – La spécification aux sept premières composantes principales 29
4.2.5 Discussion sur les résultats obtenus avec le test de robustesse 31

4.3 Spécification retenue – La consommation de pétrole (ADL) 33

4.4 Prévision rétrospective pour le niveau des émissions en 2022 34

CONCLUSION 37

ANNEXE A EFFET DES RÉVISIONS MÉTHODOLOGIQUES SUR LA SÉRIE DES ÉMISSIONS DES GES 39

ANNEXE B VARIABLES PROVENANT DES BASES DE DONNÉES DE LA CHAIRE EN MACROÉCONOMIE ET PRÉVISIONS 40

ANNEXE C MODÈLES UNIVARIÉS : RÉSULTATS DE L'EXERCICE HORS ÉCHANTILLON 44

C.1 Modèle AR(1) : 45

C.2 Modèle AR(2) : 46

C.3 Modèle ARMA(1,1) : 46

C.4 Choix du modèle de référence : 47

ANNEXE D RÉSULTATS DES DEUX SPÉCIFICATIONS DU MODÈLE ADL 48

D.1 Spécification du modèle ADL – PIB réel 49

D.2 Spécification du modèle ADL – Consommation de pétrole au Canada 52

ANNEXE E RÉSULTATS DES SPÉCIFICATIONS DU MODÈLE ARFL 55

E.1 Le modèle autorégressif augmenté de facteurs latents (ARFL) – Base de données mensuelles 56

E.2 Le modèle autorégressif augmenté de facteurs latents (ARFL) – Base de données trimestrielles 61

E.3 Conclusion de l'Annexe 66

BIBLIOGRAPHIE.....68

LISTE DES FIGURES

Figure 3.1 Niveau des émissions des GES au Canada 17

Figure 3.2 Série chronologique de la variable d'intérêt et ses trois transformations 22

Figure 4.1 Prévisions obtenues lors de l'exercice pseudo hors échantillon - Modèle AR(1) 25

Figure 4.2 Prévisions obtenues lors de l'exercice pseudo hors échantillon - PIB réel 26

Figure 4.3 Prévisions obtenues lors de l'exercice pseudo hors échantillon - Consommation de pétrole .. 29

Figure 4.4 Prévisions obtenues lors de l'exercice pseudo hors échantillon - Modèle autorégressif augmenté de sept composantes principales 30

Figure 4.5 Comparaison des prévisions rétrospectives pour l'année 2022 35

Figure A.1 Comparaison des tendances d'émissions (RIN 2023 par rapport au RIN 2022) 39

Figure C.1 Prévisions obtenues lors de l'exercice pseudo hors échantillon 44

Figure D.1 Prévisions obtenues avec les spécifications du modèle ADL exploitant le PIB réel 49

Figure D.2 Prévisions obtenues avec les spécifications du modèle ADL exploitant la consommation de pétrole 52

Figure E.1 Prévisions obtenues avec les spécifcations du modèle ARFL – Variables en niveau du panel LCDMA-M 58

Figure E.2 Prévisions obtenues avec les spécifcations du modèle ARFL – Taux de croissance des variables du panel LCDMA-M 60

Figure E.3 Prévisions obtenues avec les spécifcations du modèle ARFL – Variables en niveau du panel LCDMA-Q 63

Figure E.4 Prévisions obtenues avec les spécifcations du modèle ARFL – Taux de croissance des variables du panel LCDMA-Q 65

LISTE DES TABLEAUX

Tableau 3.1 Liste des variables prédictives considérées pour les spécifications du modèle ADL 19

Tableau 4.1 Valeurs obtenues avec les critères de sélection 26

Tableau 4.2 Résultats du test Diebold-Mariano (DM) : comparaison avec le modèle AR(1)..... 27

Tableau 4.3 Résultats du test Diebold-Mariano (DM) : comparaisons avec les modèles univariés 32

Tableau B.1 Liste des variables mensuelles 40

Tableau B.2 Liste des variables trimestrielles 42

Tableau C.1 Comparaison des métriques obtenues avec les modèles univariés (Mt d'équivalent CO_2).... 45

Tableau D.1 Ratios relatifs au modèle de référence AR(1) – PIB réel..... 50

Tableau D.2 Résultats du test Diebold-Mariano – Spécifications utilisant le PIB réel 50

Tableau D.3 Ratios relatifs au modèle de référence AR(1) – Consommation de pétrole..... 53

Tableau D.4 Résultats du test Diebold-Mariano - Spécifications utilisant la consommation de pétrole ... 53

Tableau E.1 Ratios relatifs au modèle de référence AR(1) – Variables en niveau du panel LCDMA-M..... 58

Tableau E.2 Résultats du test Diebold-Mariano – Variables en niveau du panel LCDMA-M..... 59

Tableau E.3 Ratios relatifs au modèle de référence AR(1) – Taux de croissance des variables du panel LCDMA-M 61

Tableau E.4 Résultats du test Diebold-Mariano – Taux de croissance des variables du panel LCDMA-M . 61

Tableau E.5 Ratios relatifs au modèle de référence AR(1) – Variables en niveau du panel LCDMA-Q 63

Tableau E.6 Résultats du test Diebold-Mariano – Variables en niveau du panel LCDMA-Q..... 64

Tableau E.7 Ratios relatifs au modèle de référence AR(1) – Taux de croissance des variables du panel LCDMA-Q 66

Tableau E.8 Résultats du test Diebold-Mariano – Taux de croissance des variables du panel LCDMA-Q.. 66

RÉSUMÉ

L'estimation annuelle des émissions des gaz à effet de serre (GES) au Canada est publiée avec seize mois de retard. Ceci a pour effet de ralentir le suivi des objectifs de réduction mis en place par le gouvernement fédéral. C'est ainsi que le mémoire étudie l'utilisation de spécifications du modèle autorégressif à retards échelonnés (ADL) et du modèle autorégressif augmenté de facteurs latents (ARFL) pour la réalisation de prévisions rétrospectives des GES au Canada. Un panel de quatorze variables, inspiré des résultats présentés dans des articles de recherche se penchant sur la même question aux États-Unis, a été construit pour contenir l'ensemble de candidats pour la composante exogène du modèle ADL. D'autre part, les bases de données macroéconomiques de la Chaire en macroéconomie et prévisions de l'ESG UQAM ont servi à estimer par composantes principales les facteurs latents du modèle ARFL. Suite aux résultats obtenus avec un exercice de prévision, le mémoire propose la spécification du modèle ADL incluant la consommation de pétrole canadienne pour réaliser des prévisions rétrospectives du niveau des émissions des GES au Canada. En exploitant les plus récentes données, cette spécification prévoit que le niveau des émissions devrait s'établir à 687,5 Mt d'équivalent de dioxyde de carbone (CO_2) en 2022, ce qui correspond à une hausse de 2,54 % par rapport à la dernière observation publiée.

Mots clés :

Prévision, émissions des gaz à effet de serre, modèle autorégressif à retards échelonnés, Canada.

INTRODUCTION

En 1992, dans le cadre du Sommet de la Terre de Rio, le Canada a ratifié la Convention-cadre des Nations unies sur les changements climatiques (CCNUCC). Depuis, l'une des exigences que les signataires doivent satisfaire annuellement est la confection d'un inventaire national sur leurs émissions des gaz à effet de serre (GES) et doivent le remettre à la CCNUCC au plus tard le 15 avril de chaque année (Environnement et Changement climatique Canada, 2023a). En plus de l'analyse portant sur l'inventaire national, le rapport contient également des données et des séries chronologiques sur les émissions des GES. Celles-ci permettent de suivre la progression du niveau des émissions des GES qui sont réjetés dans l'atmosphère par les signataires de la CCNUCC.

D'autre part, la hausse de la température moyenne planétaire lors des dernières décennies a motivé la communauté internationale à prendre des engagements afin de limiter cette tendance. C'est ainsi qu'en 2015 le Canada et 194 autres pays ont signé l'Accord de Paris, dont l'objectif principal est de coordonner les efforts internationaux pour limiter la hausse de la température moyenne mondiale sous le niveau de 2 °C et de la maintenir à 1,5 °C si possible (Environnement et Changement climatique Canada, 2016). C'est dans le cadre de cet accord que la Loi canadienne sur la responsabilité en matière de carboneutralité (projet de loi C-12) a été ratifiée en 2021. Celle-ci oblige le gouvernement du Canada à « atteindre la carboneutralité d'ici 2050 et à mettre en place un processus rigoureux d'objectifs, de plans et de rapports provisoires. Elle obligera le gouvernement du Canada à fixer des objectifs pour chaque étape de cinq ans entre 2030 et 2050 » (Environnement et ressources naturelles, 2024). La série chronologique sur les émissions nationales est alors utilisée comme baromètre pour l'évaluation de la progression du plan quinquennal et des objectifs établis par le gouvernement fédéral.

Toutefois, la fréquence et la vitesse de publication de ces données rendent cette démarche difficile. En effet, au moment d'écrire ces lignes, le dernier Rapport d'inventaire national (RIN) des gaz à effet de serre a été publié en avril 2023, mais la période couverte par ses séries chronologiques débute en 1990 et prend fin en 2021. Donc, bien que l'année 2022 se soit achevée un trimestre avant cette publication, les données pour 2022 seront seulement connues lors de la publication suivante, c'est-à-dire en avril 2024. Par conséquent, ce mémoire est motivé par le besoin d'avoir plus rapidement des données sur les GES au Canada. Pour répondre à ce besoin, un exercice de prévisions rétrospectives sera realisé avec différents modèles, dont le modèle autorégressif à retards échelonnés (ADL) et le modèle autorégressif augmenté

de facteurs latents (ARFL). La spécification retenue, parmi un pléthore de possibilités offertes par ces deux types de modèles, sera celle ayant obtenu l'un des meilleurs niveaux de précision relativement au modèle de référence AR(1) lors de la réalisation d'un exercice de prévision hors échantillon. La contribution du mémoire correspond aux deux éléments différencient cette étude de la littérature existante. Premièrement, le type de modèle mis de l'avant comporte un niveau de parcimonie supérieur au modèle structurel à facteurs dynamiques et au modèle d'apprentissage automatique à réseaux de neurones proposés dans d'autres articles de recherche. Deuxièmement, la variable d'intérêt porte sur le Canada au lieu des États-Unis. Le corps de ce mémoire sera divisé comme suit. Le chapitre 1 contient la revue de littérature portant sur les prévisions immédiates des émissions des GES aux États-Unis. Ensuite, le chapitre 2 établit le cadre économétrique du mémoire en présentant le modèle de référence, les types de modèles considérés, ainsi que les critères de sélection et le test de robustesse utilisés dans l'évaluation des différentes spécifications. La discussion portant sur la variable d'intérêt et sur les variables exogènes a lieu dans le chapitre 3. Enfin, en plus d'analyser les résultats obtenus dans le cadre de l'exercice pseudo hors échantillon, le chapitre 4 contient également la prévision du niveau des émissions en 2022 qui a été obtenue avec la spécification retenue.

CHAPITRE 1
REVUE DE LA LITTÉRATURE

L'intérêt pour les prévisions des émissions des GES ne date pas d'hier. À la fin du dernier siècle, certains chercheurs proposaient déjà l'utilisation de variables portant sur le revenu et sur la production pour faire des projections des émissions de CO_2 (Schmalensee *et al.*, 1998). Les modèles structurels du type « *integrated assessment models* » (IAMs) étaient souvent mis de l'avant à cette époque et avec le temps ils sont devenus le point de référence (Bennedsen *et al.*, 2021). Toutefois, dans le troisième chapitre de sa thèse de doctorat, Jensen (2021) indique que les modèles IAMs sont caractérisés par des inconvénients importants : la réalisation d'un grand nombre de calculs pour pouvoir spécifier le modèle, l'imposition de beaucoup d'hypothèses sur les paramètres et le fait que les IAMs sont surtout utilisés pour faire des prévisions des émissions sur le long terme, parfois s'étalant jusqu'à une centaine d'années. De plus, même si ces défis techniques sont surmontés, ce type de modèles est également caractérisé par un inconvénient implicite. Pindyck (2017) discute de la mauvaise utilisation de modèles économiques dans le cadre de politiques climatiques et argumente que moins d'importance doit être accordée aux estimations obtenues avec les modèles IAMs. Cet article indique que bien des intrants du modèle, dont les formes fonctionnelles et les paramètres, sont arbitraires. Donc, le jugement et les décisions de l'utilisateur du modèle ont un impact démesuré sur les prévisions produites par celui-ci. À l'aide d'un exemple, Pindyck (2017) démontre que des changements mineurs effectués au taux d'escompte utilisé dans le modèle causent l'obtention de résultats significativement différents. Ainsi, la manipulation des modèles IAMs afin d'obtenir des résultats souhaités est relativement facile. Par conséquent, l'article conclu que dans le cadre de politiques climatiques, la meilleure option disponible en ce moment repose sur la combinaison des avis des experts avec des modèles relativement simples et transparents.

Les nombreuses faiblesses des modèles IAMs et les engagements entrepris récemment pour la réduction des émissions des GES ont créé de l'intérêt pour des modèles effectuant des prévisions de court terme. C'est pour cette raison que des articles se penchant sur ce besoin sont publiés dernièrement. Les prochaines sous-sections présentent deux articles portant sur les émissions des GES aux États-Unis. Ensuite, le chapitre prend fin avec une discussion sur la contribution du mémoire à cette littérature.

1.1 Premier article de référence : le modèle structurel à facteurs dynamiques

En 2021, Bennedsen *et al.* ont publié un article de recherche qui propose l'utilisation d'un modèle structurel à facteurs dynamiques, « *dynamic factor model* » (DFM) en anglais, pour effectuer des prévisions immédiates des émissions de dioxyde de carbone (CO_2) aux États-Unis. Les auteurs considèrent un panel de 226 variables macroéconomiques comme prédicteurs potentiels. De ce nombre, 132 proviennent de la base de données à fréquence mensuelle FRED-MD de la Réserve fédérale de Saint-Louis, tandis que le reste des séries chronologiques, de fréquence annuelle, sont obtenues auprès de la Banque Mondiale et portent sur la production agricole (18 variables) et sur le commerce international (76 variables) des États-Unis. Comme la fréquence de la variable d'intérêt est annuelle, les auteurs utilisent tout simplement les moyennes annuelles des variables mensuelles de FRED-MD. Ensuite, pour obtenir des spécifications simples, Bennedsen *et al.* (2021) considèrent l'utilisation de méthodes de sélection telles que LASSO[1] (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1996), Elastic Net (Zou et Hastie, 2005) et le critère d'information d'Akaike (AIC) afin de choisir les variables prédictives à inclure dans les spécifications analysées. Toutefois, les auteurs mentionnent que ces méthodes engendrent des spécifications peu parcimonieuses. Ainsi, ils optent finalement pour la méthode de sélection automatisée dénominée AutoMetrics (Doornik et Hendry, 2015), car celle-ci permet d'inclure une contrainte sur le nombre de variables prédictives qui peuvent être incluses dans le modèle. Enfin de compte, l'article de recherche a retenu un modèle avec six régresseurs :

$$y_t = \alpha + \sum_{i=1}^{2} \beta_i x_{i,t} + \sum_{j=1}^{4} \gamma_j f_{j,t} + \delta_1 z_{1970} + \delta_2 z_{1990} + \varepsilon_t \,. \qquad (1)$$

La variable y_t correspond au taux de croissance des émissions de CO_2 par habitant aux États-Unis. D'autre part, $x_{i,t}$ représente deux variables exogènes, dont la première correspond à un indice sur le niveau général de la production industrielle américaine et la deuxième est un indice sur le secteur des services publics aux États-Unis. Les auteurs indiquent qu'il n'est pas étonant que cette combinaison explique sept dixièmes de la variation du taux de croissance des émissions par habitant, car la première variable représente la portion industrielle de l'économie tandis que la deuxième correspond au secteur résidentiel

[1] LASSO est une méthode d'estimation de modèles linéaires qui minimise la somme des carrés des résidus à condition que la somme des valeurs absolues des coefficients soit inférieure à la valeur d'une constante donnée (Tibshirani, 1996). Cette technique de régression pénalisée est utilisée en statistique et en apprentissage automatique pour réaliser la sélection de variables dans les modèles.

du pays. Le troisième terme inclut quatre facteurs latents notés $f_{j,t}$. Ceux-ci sont tirés du panel contenant les 226 variables macroéconomiques. Ensuite, les deux dernières variables prédictives sont des variables dichotomiques correspondant aux années 1970 et 1990[2], car des valeurs aberrantes de la variable d'intérêt sont observées lors de ces deux années. Ainsi, ces variables dichotomiques sont incluses afin d'ajuster les prévisions du modèle pour ces deux années. Enfin, le terme d'erreur ε_t est supposé iid(0,σ_ε^2). Bennedsen *et al.* (2021) indique que les valeurs à la période *t* des variables prédictives, dont les deux variables exogènes et celles du panel dont les facteurs latents sont tirés, sont publiées plus rapidement que celle de la variable d'intérêt. Ceci rend alors possible la réalisation de *nowcasts*[3] au temps *t* de la variable d'intérêt avec la spécification (1) dès que les valeurs contemporaines des variables explicatives sont disponibles car les coefficients sont estimés avec les données historiques.

Suite à l'analyse des résultats de l'exercice pseudo hors échantillon, les auteurs concluent que le gain marginal obtenu avec l'inclusion des quatre facteurs latents n'est pas assez significatif pour justifier le travail qui doit être accompli pour les estimer. Par conséquent, la conclusion de l'article propose plutôt l'utilisation d'une version simplifiée de la spécification, c'est-à-dire celle excluant les facteurs latents

$$y_t = \alpha + \sum_{i=1}^{2} \beta_i x_{i,t} + \delta_1 z_{1970} + \delta_2 z_{1990} + \epsilon_t \,. \qquad (2)$$

Enfin, les auteurs indiquent que cette spécification obtient de meilleurs résultats hors échantillon que le modèle à croissance constante, que des modèles univariés et multivariés de type ARMA, et qu'un modèle à vecteur autorégressif (VAR).

[2] Les auteurs indiquent que d'autres articles de recherche ont aussi detecté une valeur aberrante autour de l'année 1970. Voir Auffhammer et Steinhauser (2012) et Schmalensee et *al.* (1998). Les deux indiquent que la valeur aberrante dans les émissions de CO_2 est une conséquence du choc pétrolier du début des années 1970. D'autre part, selon Bennedsen et *al.* (2021), l'interprétation de l'année 1990 n'est pas aussi évidente. Les auteurs mentionnent que le coefficient δ_2 dans (1). Ceci indique alors que les émissions en 1990 sont supérieures à ce qui est supposé par les deux autres variables exogènes.

[3] Méthode qui consiste à estimer la valeur actuelle de la variable d'intérêt sur les données disponibles jusqu'à la date *t*.

1.2 Deuxième article de référence : l'apprentissage automatique comme modèle de prévision

Jensen (2021) propose plutôt l'utilisation de l'apprentissage automatique afin de formuler les modèles de prévision pour les émissions de CO_2 aux États-Unis. Tout comme l'article précédent, l'auteur se sert de 126 variables mensuelles provenant de la base FRED-MD. Par contre, l'un des éléments différentiateurs est l'utilisation de 135 variables trimestrielles provenant de la base FRED-QD. Ces variables sont fournies à des modèles d'apprentissage par arbre de décision et des modèles de réseaux de neurones. En se basant sur les erreurs quadratiques moyennes (EQM) obtenues avec les différentes spécifications, le modèle d'apprentissage automatique à réseaux de neurones composé d'une seule couche et de 32 unités est choisi.

De plus Jensen (2021) mentionne que l'attrait principal de l'utilisation de l'apprentissage automatique est que celui-ci permet l'exploitation d'information disponible dans des données à plus haute fréquence. Par conséquent, la production de prévisions immédiates pour l'année actuelle est possible plusieurs fois au cours de celle-ci. En plus de diverger au niveau des modèles étudiés, l'auteur indique que les résultats dans Bennedsen *et al.* (2021) n'étaient pas comparables. Effectivement, Bennedsen *et al.* (2021) fait abstraction des défis qui surviennent lorsque la fréquence de la variable d'intérêt diffère de celles des régresseurs. Les auteurs de cet article utilisent tout simplement les moyennes annuelles des variables à fréquence mensuelle. Donc, en plus de proposer un modèle pour la prévision immédiate des émissions de CO_2, Jensen (2021) offre également une solution aux défis qui surviennent à cause des différentes fréquences des variables.

1.3 Contribution du mémoire à la littérature

Ce mémoire s'inscrit dans la même littérature que les deux articles de recherche présentés précédemment, mais à cause de leur grande divergence méthodologique, ultimement seulement l'un des deux est choisi comme ligne directrice. D'une part, en plus d'être l'une de ses principales caractéristiques, le niveau de parcimonie de la spécification proposée par Bennedsen *et al.* (2021) est l'un de ses plus grands attributs. Effectivement, le peu de prédicteurs inclus facilite l'estimation des paramètres et rend son utilisation accessible à un plus grand public. De plus, cet article exploite indirectement l'information présente dans des données à plus haute fréquence en se servant de leur moyenne annuelle pour effectuer des prévisions immédiates du niveau des émissions des GES. D'autre part, Jensen (2021) met de l'avant un modèle complexe et qui est tiré de l'avant-garde de la littérature statistique et économétrique. Même si l'un des avantages de cet article est l'utilisation de variables avec des fréquences différentes, la contribution

principale de celui-ci demeure la possibilité de réaliser des prévisions immédiates pour l'année en cours et ce bien avant son aboutissement. Par contre, l'utilisation d'un modèle à boîte noire rend difficile l'identification des variables ayant le plus d'importance prédictive pour le niveau des émissions des GES. En définitive, le mémoire s'inspire davantage des attributs et de la méthodologie du premier article (Bennedsen *et al.*, 2021). Effectivement, l'étude présentée ultérieurement propose l'utilisation d'un modèle autorégressif à retards échelonnés (ADL) exploitant la consommation de pétrole comme modèle de prévision rétrospective des émissions des GES au Canada. Par conséquent, le mémoire diffère des deux autres articles en proposant une spécification provenant d'une classe de modèles différente, tout en offrant une solution à la problématique liée au retard de publication de la série canadienne plutôt que celle des États-Unis.

CHAPITRE 2
CADRE ÉCONOMÉTRIQUE

L'exercice de prévision dont il est question dans ce mémoire diffère légèrement des cadres habituels, c'est pourquoi ce chapitre débute par une brève explication de l'enjeu entourant cette prévision. Ensuite, le modèle qui servira comme référence est précisé. Subséquemment, les deux modèles économétriques qui ont été étudiés sont présentés. Enfin, une discussion sur les critères d'évaluation et sur le test de robustesse bouclera le chapitre.

2.1 L'enjeu de la prévision rétrospective

Le niveau des émissions des GES, mesuré en mégatonnes (Mt) d'équivalent CO_2[4] dans ce mémoire, est la variable d'intérêt des modèles de prévision. De plus, cette variable est publiée avec une année de retard. Alors, posons y_t comme étant la représentation de sa valeur au temps t. Étant donné le retard qui la caractérise, la valeur de y_t est seulement publiée lors de la période suivante, c'est-à-dire en $t+1$. D'autre part, supposons que x_t représente une variable exogène dont la valeur est disponible au temps t. Ainsi, le retard caractérisant les publications de la variable d'intérêt permet l'utilisation de la valeur contemporaine de x_t comme variable prédictive de l'observation inconnue de y_t. De plus, ce retard permet de considérer des régresseurs dont la fréquence est plus élevée que celle des émissions des GES, car il est possible de calculer leur moyenne annuelle étant donné que toutes les observations de l'année correspondant à la période t sont disponibles avant la publication de y_t. Par conséquent, le retard de publication de y_t fait en sorte que l'exercice de prévision du mémoire ne porte pas sur la prédiction d'une valeur future de la variable d'intérêt, mais plutôt sur sa valeur présente. Le mémoire considère le problème de prévision immédiate comme étant donné par la relation suivante

$$\hat{y}_{\{t|t-1\}} = E[y_t|\Omega_{t-1}]\,, \tag{3}$$

[4]« L'équivalent dioxyde de carbone (CO_2) est une mesure métrique utilisée pour comparer les émissions de divers gaz à effet de serre sur la base de leur potentiel de réchauffement global (PRG), en convertissant les quantités des divers gaz émis en la quantité équivalente de dioxyde de carbone ayant le même potentiel de réchauffement planétaire » (Eurostat, 2023).

où l'espérance de y_t est conditionnelle à l'ensemble d'information $\Omega_{\{t-1\}}$. Ce dernier contient tous les retards de la variable d'intérêt, ainsi que les valeurs passées et la valeur contemporaine (t) des variables exogènes considèrées.

2.2 Le modèle de référence

Les deux articles de recherche dans la revue de littérature choisissent le modèle à croissance constante comme modèle de référence pour leurs exercices de prévision respectifs. Dans ce mémoire, un modèle univarié joue plutôt ce rôle. Cette décision est basée sur la simplicité qui le caractérise, mais surtout par la réputation qui précède ce type de modèle quand il est temps d'effectuer des prévisions. De plus, la stabilisation de la tendance de la série chronologique depuis le tournant du siècle est favorable pour ce type de modèle, voir Figure 3.1 dans le chapitre suivant. Des trois modèles univariés considérés au départ, c'est le modèle autorégressif d'ordre un (AR1) qui est finalement choisi comme modèle de référence, car il est celui qui engendre les plus petites valeurs auprès des critères de racine carrée de l'erreur quadratique moyenne (REQM) et de l'erreur absolue moyenne (EAM) lors de l'exercice pseudo hors échantillon dans l'Annexe C. Dans le cadre de prévision rétrospective, ce modèle prend la forme suivante :

$$y_t = c + \phi_1 \, y_{t-1} + \varepsilon_t \, , \tag{4}$$

où c représente la constante du modèle. Le symbole ϕ_1 est le paramètre autorégressif et il est estimé par moindres carrés ordinaires. Le symbole ε_t est le terme d'erreur et il est considéré comme étant un bruit blanc (BB).

2.3 Modèle autorégressif à retards échelonnés (ADL)

La précision des prévisions du modèle de référence repose entièrement sur l'information disponible dans les retards de la variable d'intérêt. Étant donné qu'il s'agit d'une méthode basée sur de l'information passée, celle-ci peut engendrer des résultats acceptables lorsque la progression de la variable d'intérêt est exemptée de variations exogènes significatives. La faiblesse du modèle AR(1) est clairement illustrée dans la Figure 4.1. Celui-ci est incapable de bien prévoir la chute des émissions des GES en 2020. Ainsi, les spécifications du modèle de cette section représentent une solution aux lacunes du modèle de référence.

Dans les modèles autorégressifs à retards échelonnés (ADL), une ou plusieures variables exogènes sont ajoutées au modèle autorégressif. Cette inclusion a comme objectif la réduction de l'erreur de prévision

lors des mouvements cycliques de la variable d'intérêt. Le modèle ADL considéré dans ce mémoire se présente comme suit :

$$y_t = c + \Phi_1 \, y_{t-1} \; + \; \beta(L) \, x_t \; + \; \varepsilon_t \, , \tag{5}$$

où $\beta(L)$ est un polynôme de retards L et x_t représente la variable exogène ajoutée au terme autorégressif. Une panoplie de spécifications de ce type de modèle, se différenciant chacune par l'ordre du polynôme $\beta(L)$ et par la variable exogène exploitée, est considérée au départ du processus de prévision. Toutefois, seulement celles ayant obtenu les meilleurs résultats dans l'exercice pseudo hors échantillon sont présentées et analysées au Chapitre 4.

2.4 Modèle autorégressif augmenté de facteurs latents (ARFL)

Le nombre de variables prédictives qui peuvent être incluses dans les spécifications du modèle ADL est limité par les 31 observations qui forment la série chronologique des émissions des GES. La hausse du nombre de variables exploitées cause une perte de degrés de liberté considérable et ceci affecte le niveau de précision des spécifications estimées. Cet effet est encore plus important dans le cadre de l'exercice pseudo hors échantillon où le processus d'estimation est effectué avec seulement une vingtaine d'observations lors des premières années de la période d'évaluation. Une possible solution à cette problématique est le remplacement des variables individuelles par des facteurs latents tirés de bases de données macroéconomiques.

La démarche suivie pour l'obtention des prévisions avec des facteurs latents se base sur la procédure présentée dans Stock et Watson (2022a,b). Ce processus comporte principalement les deux étapes suivantes. D'abord, les facteurs latents doivent être estimés à partir d'un panel contenant un grand nombre de variables. Ensuite, les paramètres multipliant ces facteurs dans le modèle de prévision doivent être estimés par les moindres carrés ordinaires.

Soit X_t un ensemble de *N* séries temporelles observées au temps t. Le modèle à facteurs suivant, avec r facteurs latents F_t est supposé :

$$X_t = \Lambda \, F_t + u_t \, , \tag{6}$$

où Λ désigne une matrice de poids, de taille *N* x r , qui multiplie le vecteur des facteurs latents F_t de taille r x 1 et qui est invariable au temps. Le vecteur u_t de taille *N* x 1 représente les perturbations idiosyncratiques. Il est possible que les perturbations soient autocorrélées et légèrement corrélées entre elles (Chamberlain et Rothschild, 1983).

Ensuite, les facteurs latents F_t sont estimés par composantes principales. Considérons le problème des moindres carrés suivant :

$$\min_{(F_1,\ldots,F_t,\Lambda)} V_r\,(\Lambda, F) = (NT)^{-1} \sum_{t=1}^{T} (X_t - \Lambda F_t)\,'(X_t - \Lambda F_t)\,, \tag{7}$$

où la période couverte par les séries temporelles de l'ensemble X_t correspond à T. Lorsqu'on minimise cette fonction objectif et on isole par rapport à F_t, un problème de minimisation ayant Λ comme seule variable de contrôle est obtenue :

$$\min_{\Lambda} T^{-1} \sum_{t=1}^{T} X_t{}'[I - \Lambda(\Lambda'\Lambda)^{-1}\Lambda]X_t\,. \tag{8}$$

Ce problème est équivalent au problème de maximisation

$$\max_{\Lambda} tr\left\{\left((\Lambda'\Lambda)^{-1/2}\right)'\Lambda'\left(T^{-1}\sum_{t=1}^{T} X_t X_t{}'\right)(\Lambda'\Lambda)^{-1/2}\right\}. \tag{9}$$

Ce dernier est sujet à la normalisation $N^{-1}\Lambda'\Lambda = I_r$, où tr correspond à la trace de la matrice. La parenthèse au centre du problème correspond à la matrice variance–covariance et ceci nous permet de réécrire le problème comme étant

$$\max_{\Lambda} tr\{\Lambda'\hat{\Sigma}_{XX}\Lambda\}\,. \tag{10}$$

Finalement, ce problème est résolu quand la valeur de Λ est égale aux vecteurs propres de $\hat{\Sigma}_{XX}$ et ceux-ci correspondent aux r valeurs propres les plus élevées. Comme il avait été supposé que $N^{-1}\Lambda'\Lambda = I_r$, l'estimé par composantes principales de F_t est alors

$$\hat{F}_t = \hat{F}_t\left(N^{-1}\hat{\Lambda}\right) = N^{-1}\hat{\Lambda}'X_t \,. \tag{11}$$

Bien sûr, lorsqu'un nombre r de facteurs latents est inclus dans l'une des spécifications, ces facteurs correspondent aux r valeurs propres absolues les plus élevées de la matrice $N^{-1}\hat{\Lambda}X_t$. Une fois les facteurs latents estimés, un certain nombre de ceux-ci sont ajoutés au modèle de référence. La représentation algébrique du modèle autorégressif augmenté de facteurs latents correspond à

$$y_t = c + \beta_1 y_{t-1} + \theta_1 \hat{F}_t + e_t \,, \tag{12}$$

où $\hat{F}_t$ symbolise le vecteur de r facteurs latents estimés par composantes principales à partir de l'ensemble de variables exogènes X_t

2.5 Critères d'évaluation des prévisions

Dans ce mémoire, deux spécifications du modèle ADL et une spécification du modèle ARFL sont présentées et discutées. Toutefois, il est à noter que ces spécifications ont été choisies parmi un grand nombre de candidats considérés pour la prévision des émissions des GES. En plus de la différence au niveau des variables explicatives, ces spécifications diffèrent par le nombre de régresseurs exploités, par les différentes combinaisons possibles de ceux-ci, et par l'utilisation des valeurs contemporaines et/ou retards des variables prédictives. Les différentes combinaisons de toutes ces caractéristiques permettent de confectionner des centaines de spécifications uniques. Donc, la méthode d'évaluation choisie permet de comparer équitablement ces spécifications malgré toutes leurs différences à l'aide d'un exercice hors échantillon récursif réalisé avec des centaines d'entre elles. Les erreurs de prévisions sont obtenues à partir de la différence entre la prévision de la spécification et la valeur réalisée de la variable d'intérêt,

$$e_{m,t} = y_t - \hat{y}_{\{m,t|t-1\}} \,,$$

où $e_{m,t}$ symbolise l'erreur de prévision obtenue avec la spécification m lors de l'année t.

Après avoir complété les démarches nécessaires pour l'obtention des prévisions et des erreurs de prévision, la prochaine étape correspond au calcul des critères permettant de comparer chacune des spécifications. Les deux critères choisis pour cet exercice hors échantillon sont la racine carrée de l'erreur quadratique moyenne (REQM) et l'erreur absolue moyenne (EAM). Voici la forme algébrique des deux mesures :

$$\widehat{REQM}_m = \sqrt{\frac{1}{\rho}\sum_{t=1}^{\rho}\left(y_t - \hat{y}_{\{m,t|t-1\}}\right)^2} \tag{13}$$

$$\widehat{EAM}_m = \frac{1}{\rho}\sum_{t=1}^{\rho}\left|y_t - \hat{y}_{\{m,t|t-1\}}\right| , \tag{14}$$

où le symbole ρ correspond au nombre total de prévisions effectuées lors de la période d'évaluation de l'exercice pseudo hors échantillon. Le choix d'avoir opté pour l'application de la racine carrée à l'erreur quadratique moyenne (EQM) s'explique par le fait que ceci permet d'obtenir une métrique dont l'unité de mesure est égale à celle de la variable d'intérêt. Ceci facilite la comparaison et l'interprétation des résultats obtenus. De plus, la racine carrée n'a aucune incidence sur l'ordre relatif des résultats obtenus avec l'EQM (Hodson, 2022). Une fois que les REQM et les EAM ont été calculés pour chacune des spécifications, ces mesures sont comparées avec les résultats obtenus avec le modèle de référence pour déterminer si la spécification en question fait relativement mieux. Le ratio entre la spécification en question et le modèle AR(1) est le critère utilisé pour classer les différentes spécifications. Voici la formule des ratios :

$$Ratio_{REQM,m} = \frac{\widehat{REQM}_m}{\widehat{REQM}_{AR(1)}} \tag{15}$$

$$Ratio_{EAM,m} = \frac{\widehat{EAM}_m}{\widehat{EAM}_{AR(1)}} \ . \tag{16}$$

Les deux critères d'évaluation précédents permettent d'identifier les meilleures spécifications des modèles ADL et ARFL. D'autre part, leurs ratios relatifs permettent de quantifier l'amélioration du niveau de précision obtenu par rapport au modèle de référence. Toutefois, à cette étape, il est incertain si le niveau de précision des spécifications est statistiquement différent que celui obtenu avec le modèle

autorégressif d'ordre un (AR1). En conséquence, les erreurs de prévision obtenues avec les spécifications lors de l'exercice pseudo hors échantillon sont utilisées dans le cadre du test présenté dans Diebold et Mariano (1995) afin de pouvoir affirmer que la précision des prévisions des spécifications est statistiquement différente que celle du modèle de référence.

Le principe régissant le test présenté en 1995 par Diebold et Mariano (DM) est la différence entre les fonctions de coût de deux modèles comparés, où la fonction de coût $g(e_{m,t})$ dépend directement de l'erreur de prévision $e_{m,t}$. Dans ce mémoire, la fonction de coût est quadratique, alors la différence des fonctions de coût correspond à

$$d_t = e_{m,t}^2 - e_{j,t}^2$$

$$e_{m,t} = y_t - \hat{y}_{\{m,t|t-1\}} .$$

L'hypothèse nulle et l'hypothèse alternative du test DM sont alors

$$H_0 : E[d_t] = 0 \; \forall \; t$$

$$H_1 : E[d_t] \neq 0 \; \forall \; t ,$$

où l'hypothèse nulle suggère que la précision des deux modèles est égale, alors que le rejet de celle-ci indique que le niveau de précision de l'un des modèles est différent de celui de l'autre modèle. Sous l'hypothèse nulle, la statistique du test correspond à ceci

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi f_d(0)}{T}}} \sim N(0,1) \qquad (17)$$

où

$$\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t ,$$

est la moyenne de l'échantillon des différences des fonctions de coûts. La densité spectrale de fréquence nulle $f_d(0)$ de la différence des fonctions de coût est égale à

$$f_d(0) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau) \,,$$

où $\gamma_d(\tau)$ est l'autocovariance de la différence des fonctions de coût avec la valeur d'écart égale à τ et μ la moyenne de la population des différences des fonctions de coûts

$$\gamma_d(\tau) = E[(d_t - \mu) - (d_{t-\tau} - \mu)] \,.$$

Afin de comprendre la densité spectrale, celle-ci peut être interpretée comme étant « l'analogue de la fonction de densité de probabilité de variables aléatoires continues » (Shumway et Stoffer, 2017). Bref, la covariance d'une série temporelle peut être representée par une fonction connue sous le nom de densité spectrale.

Enfin, il est à noter que l'implémentation du test Diebold et Mariano n'a pas comme objectif la validation de la spécification du modèle, mais plutôt la confirmation statistique que la différence entre les prévisions de deux modèles prédictifs est significative. D'autre part, le non-rejet de l'hypothèse nulle du test DM justifie une discussion portant sur la valeur ajoutée de l'inclusion de variables exogènes ou facteurs latents dans la spécification considérée.

CHAPITRE 3
LES DONNÉES

La série des émissions des gaz à effet de serre au Canada qui est mise à jour avec chacune des publications du Rapport d'inventaire national (Environnement et Changement climatique Canada, 2023a) représente la variable d'intérêt de l'exercice de prévision du mémoire. D'autre part, les variables explicatives qui ont été considérées avec les diverses spécifications des modèles de prévision proviennent de deux types de sources différents. La majorité de ces variables sont tirées des deux bases de données macroéconomiques élaborées par la Chaire en macroéconomie et prévisions de l'École des sciences de la gestion (ESG). Les autres variables indépendantes sont issues de bases de données d'organisations nationales et internationales comme l'Organisation de coopération et de développement économiques (OCDE), la Réserve fédérale de Saint-Louis (FRED) et Statistiques Canada, entre autres. Les prochaines sections du chapitre discutent plus en détail les caractéristiques de la variable dépendante et des variables prédictives.

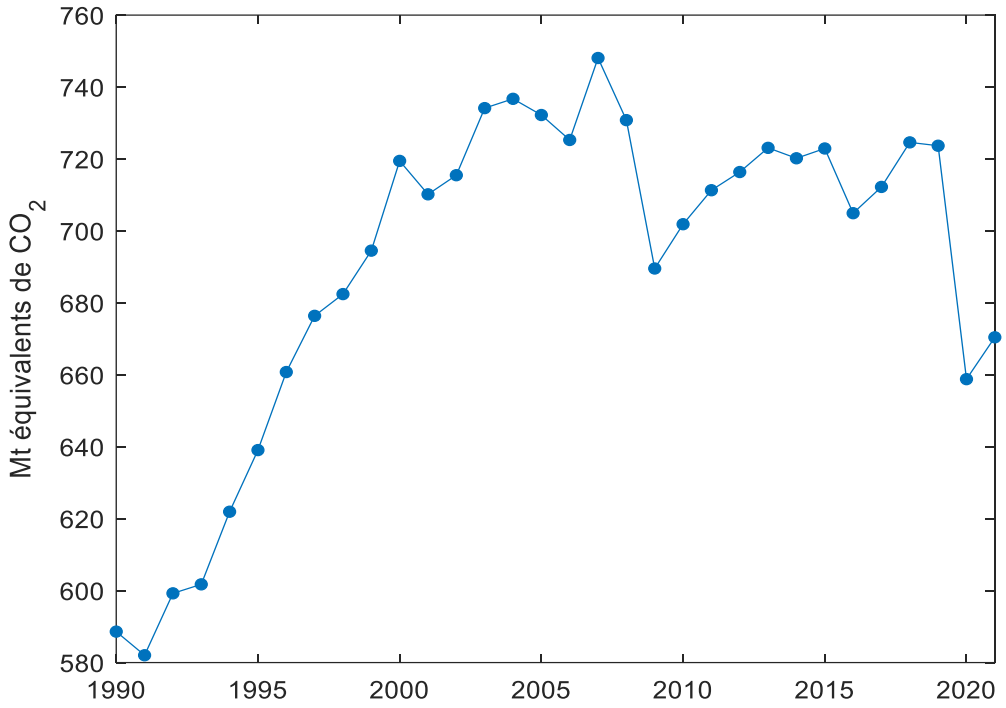
3.1 La variable d'intérêt

Les données qui accompagnent la publication annuelle du RIN ne sont pas des observations brutes recueillies avec des instruments de mesure. Elles correspondent plutôt à des estimations. Ces dernières sont obtenues à partir de modèles conçus par Environnement et Changement climatique Canada. Ces modèles se basent sur les méthodes proposées par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) (Environnement et Changement climatique Canada, 2023a). Ce processus débute par la réalisation d'estimations pour chacun des types de source d'émissions. Ensuite, les estimations sont incluses dans l'un des secteurs d'activité définis par le GIEC. Les cinq secteurs d'activité sont : l'énergie, les procédés industriels et l'utilisation des produits (PIUP), l'agriculture, l'affectation des terres, changement d'affectation des terres et foresterie (ATCATF), et les déchets (Environnement et Changement climatique Canada, 2023c). Ensuite, les résultats sont regroupés par province et territoire. Finalement, le tout est additionné afin d'obtenir une estimation annuelle des émissions pour l'ensemble du Canada (Environnement et Changement climatique Canada, 2023b).

Une fois que les détails de la confection de la série chronologique des émissions des GES au Canada ont été discutés, la prochaine étape consiste à présenter ses caractéristiques statistiques. Comme le RIN est produit annuellement, la fréquence de la série chronologique en question est également annuelle. Au moment d'écrire ces lignes, la plus récente version a été publiée en avril 2023 et elle contient 32

observations, dont la première correspond à l'année 1990 et la dernière à l'année 2021. L'unité de mesure utilisée par Environnement Canada est les kilotonnes d'équivalent de dioxyde de carbone (kt d'éq. CO_2), à l'exception des figures où les mégatonnes d'éq. CO_2 sont plutôt utilisées afin d'alléger leur présentation. De plus, la série agrégée peut être segmentée par source d'émission, par secteur d'activité économique ou par province et territoire si la question de recherche le requiert. Le graphique suivant illustre la série chronologique de la variable d'intérêt.

Figure 3.1 Niveau des émissions des GES au Canada



Une caractéristique particulière de la publication mérite d'être mentionnée. Les méthodes et les modèles d'estimation utilisés par Environnement et Changement climatique Canada sont revus et modifiés constamment. Il est alors possible que la valeur des estimations diffère d'une publication à une autre. À titre d'exemple, lors de la publication de 2023, des modifications ont été apportées aux estimations « des émissions attribuables aux sites d'enfouissement de déchets et aux transports, entre autres, de même que l'inclusion d'une nouvelle source d'émissions fugitives en aval du compteur » (Environnement et Changement climatique Canada, 2023a). Ces modifications méthodologiques ont eu une répercussion non négligeable sur les estimations publiées. Effectivement, la valeur des données a diminué sur toute la

période. Toutefois, ces révisions ont eu un impact homogène sur le niveau des estimations. Sur un graphique, l'effet correspond à une translation vers le bas de la série (Annexe A). Donc, les valeurs des observations ont changé, mais la dynamique de la série est restée inchangée. Les implications que ces révisions ont sur la question de recherche de cet article seront discutées dans la conclusion.

3.2 Le panel contenant les variables des spécifications du modèle ADL

Comme discuté dans le chapitre précédent, l'inclusion de variables exogènes dans le modèle ADL a comme objectif de combler les lacunes du modèle AR(1). Il est assumé que la variation de ces variables permet de mieux prévoir les mouvements exogènes de la variable d'intérêt. Le choix des variables prédictives se base sur les variables proposées dans l'article de recherche de Bennedsen *et al.* (2021) et dans le troisième chapitre de thèse de doctorat de Jensen (2021). Étant donné que ces deux articles s'intéressent aux émissions américaines, des variables similaires portant sur le Canada sont sélectionnées. Au final, le panel des régresseurs exogènes, utilisés avec les spécifications du modèle ADL, contient quatorze variables. Ces variables sont recueillies auprès de l'agence nationale de statistique du pays, mais également auprès de bases de données d'institutions reconnues comme la Réserve fédérale de Saint-Louis (FRED) ou l'OCDE. Les séries chronologiques de ces variables sont de fréquence mensuelle, trimestrielle ou annuelle. Par conséquent, lorsque la fréquence est mensuelle ou trimestrielle, la moyenne annuelle est calculée et c'est cette dernière qui est utilisée dans l'analyse. Le Tableau 3.1 dresse la liste des variables considérées pour les diverses spécifications du modèle ADL.

Tableau 3.1 Liste des variables prédictives considérées pour les spécifications du modèle ADL

Variable	Description	Fréquence	Source
AIR_PASSENGER	Nombre de passagers (en milliers)	Mensuelle	Statistique Canada
AIR_PASSENGER_KM	Nombre de kilometres parcourus par les passagers (en milliers de km)	Mensuelle	Statistique Canada
AIR_FUEL_CONS	Quantité d'essence consommée (milliers de litres)	Mensuelle	Statistique Canada
CLI_CAN	Composite Leading Indicator: Série construite pour signaler des points d'inflexion du cycle économique.	Mensuelle	OCDE
CONSTRUCTION_RES	Mises en chantier d'immobilier résidentiel au Canada	Annuelle	FRED
ELECTRICITY_GEN	Production totale d'électricité (Twh)	Annuelle	Energy Institute - Statistical Review of World Energy
IND_PROD_TOTAL	Production industrielle totale au Canada (le secteur de la construction exclu)	Annuelle	FRED
INV_GFCF	Formation brute de capital fixe	Trimestrielle	OCDE
INVENTORY_CHANGES	Variation réelle du stock au Canada	Trimestrielle	FRED
MANU_PRODUCTION	Production totale dans le secteur manufacturier	Trimestrielle	FRED
OIL_PROD_EI	Production totale de pétrole (en millions de tonnes)	Annuelle	Energy Institute - Statistical Review of World Energy
OIL_CONS_EI	Consommation totale de pétrole (en millions de tonnes)	Annuelle	Energy Institute - Statistical Review of World Energy
PIB_R	PIB réel obtenu en divisant le PIB nominal par l'indice des prix implicites	Annuelle	Statistique Canada
RETAIL_SALES	Ventes au détail totales	Trimestrielle	FRED
TRADE_GOODS	Total des biens échangés à partir du commerce international (en millions de USD)	Trimestrielle	OCDE

3.3 Les bases de données pour l'estimation des facteurs latents

Les deux articles de recherche discutés dans la revue de littérature utilisent la base de données présentée dans McCraken et Ng (2016) pour estimer leurs modèles de prévision. L'attribut principal de cette base de données est son large éventail de séries macroéconomiques portant sur les États-Unis, mais surtout le fait qu'elle est mise à jour périodiquement par la Réserve fédérale de Saint-Louis. Étant donné que le sujet du mémoire porte sur un enjeu canadien, cette base de données n'est pas adéquate pour l'exercice de prévision des émissions des GES canadiennes. Par conséquent, le mémoire utilise une autre base de données pour estimer les facteurs latents et les paramètres des spécifications du modèle ARFL. La Chaire

en macroéconomie et prévisions de l'ESG rend accessible sur son site web la base de données macroéconomiques canadiennes désaisonnalisées LCDMA construite par Fortin-Gagnon *et al.* (2020). Les variables de cette base de données peuvent être regroupées dans huit catégories : la production, le marché du travail, le marché immobilier, les stocks des manufacturiers, la monnaie et le crédit, le commerce international et les flux financiers, les indices de prix et le marché boursier.

Sur le site web de la Chaire en macroéconomie et prévisions, deux bases de données macroéconomiques canadiennes sont disponibles. Toutes les deux regroupent un large nombre de variables nationales et provinciales. L'élément qui les différencie est la fréquence des données. L'une d'entre elles contient des données mensuelles et l'autre regroupe des données trimestrielles. Dans ce mémoire, on fait référence à la base de données mensuelles comme étant LCDMA-M et le terme LCDMA-Q est utilisé pour la base de données trimestrielles. Deux versions sont disponibles pour chacune des bases de données, la première version correspond aux « données brutes recueillies auprès de Statistiques Canada depuis 1914 » et l'autre version est un « panel stationnaire et équilibré » dont la première observation correspond au mois/trimestre de 1981[5]. Le panel stationnaire est la version utilisée dans ce mémoire.

Une nouvelle publication de la base de données LCDMA-M est disponible à chaque mois, alors que la base de données LCDMA-Q l'est à tous les trois mois. Pour des raisons de reproductibilité, il est à noter que la version de LCDMA-M utilisée dans cette analyse correspond à la publication du mois de mai 2023. Le panel téléchargé contient 405 séries chronologiques, dont la première observation de chaque série correspond à janvier 1981 et la dernière à avril 2023. Les séries ont alors un total de 508 observations mensuelles. Comme l'exercice de prévision s'intéresse au niveau national des émissions des GES, les séries provinciales du panel sont exclues. Après cette modification, le panel final regroupe 115 séries canadiennes. Le Tableau B.1 dans l'Annexe B contient la liste énumérant toutes les variables mensuelles retenues.

Tout comme pour la base de données mensuelle, le panel téléchargé pour LCDMA-Q est celui correspondant à la publication de mai 2023. Ce panel contient initialement 551 variables nationales et provinciales. De ce nombre, la majorité des séries sont des moyennes trimestrielles des séries mensuelles dans LCDMA-M. Lorsqu'on fait abstraction de ces dernières, les séries restantes portent sur le produit intérieur brut et ses composantes. Des modifications sont alors apportées au panel afin de retirer les séries provinciales et les séries correspondant à une moyenne des séries dans LCDMA-M. Au final, le panel

[5] Pour plus de détails consulter https://chairemacro.esg.uqam.ca/donnees/base-de-donnees-canadiennes/

trimestriel utilisé regroupe 71 séries trimestrielles. Les données de ces dernières débutent au premier trimestre de 1981 et prennent fin au premier trimestre de 2023 pour un total de 169 observations. La liste complète des variables retenues se trouve dans le Tableau B.2 de l'Annexe B.

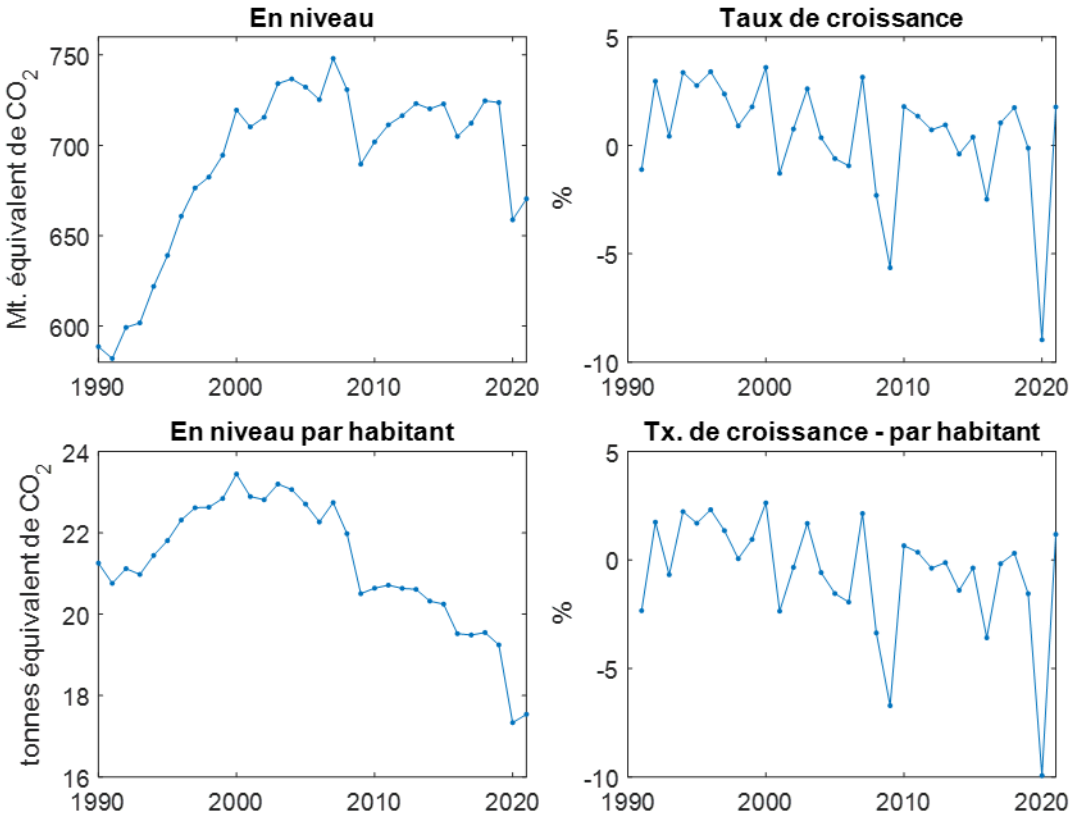
3.4 Les transformations de la variable d'intérêt

La revue de littérature a présenté deux articles dont la question de recherche porte sur l'élaboration de modèles de prévision pour les émissions de CO_2 aux États-Unis. La variable d'intérêt des modèles semble alors bien évidente, mais les auteurs de ces articles ont décidé d'appliquer une transformation au niveau des émissions avant de débuter le processus d'estimation. L'article de Bennedsen *et al.* (2021) divise les émissions de CO_2 par la population des États-Unis pour obtenir le niveau des émissions par habitant. Ensuite, les auteurs calculent la log-différence de cette dernière afin d'approximer le taux de croissance des émissions par habitant. L'explication fournie par les auteurs fait allusion aux conclusions tirées des tests de racines unitaires et de stationnarité appliqués. Selon les tests KPSS et Augmented Dickey-Fuller (ADF) « une racine unitaire est présente dans la série en niveau et dans sa version logarithmique, mais les différences premières de celles-ci semblent être stationnaires » (Bennedsen *et al.*, 2021). D'autre part, le troisième chapitre de la thèse de doctorat de Jensen (2021) s'intéresse à la même question de recherche et il se sert également de la log-différence des émissions de CO_2 comme variable d'intérêt. Ainsi, on remarque que le choix de l'utilisation du taux de croissance est justifié avec les tests de stationnarité, mais aucun des deux articles n'offre une explication pour la division de la variable d'intérêt par la population nationale.

Les deux articles de recherche mentionnés précédemment choisissent la log-différence des émissions de GES par habitant aux États-Unis comme variable d'intérêt. Ainsi, les démarches suivies dans ces articles portent sur la prévision d'un taux de croissance et non d'une variable en niveau comme dans la question de recherche de ce mémoire. Étant donné cette différence, le processus dans ce mémoire diverge légèrement et ne suit pas à la lettre les démarches entreprises par les auteurs de ces deux articles de recherche. Pour vérifier si la transformation du niveau des émissions des GES permet d'obtenir des prévisions plus précises, l'exercice d'estimation est effectué à quatre reprises avec chacune des spécifications. La première fois, l'estimation est réalisée avec la variable d'intérêt originale et les trois autres avec des transformations de celle-ci. La première transformation est tout simplement le taux de croissance des émissions des GES. Ensuite, le niveau des émissions des GES est divisé par la population canadienne afin d'obtenir la deuxième transformation. Enfin, la troisième transformation correspond à la

variable d'intérêt des deux articles mentionnés auparavant, c'est-à-dire au taux de croissance du niveau des émissions par habitant.

Figure 3.2 Série chronologique de la variable d'intérêt et ses trois transformations



En observant les quatre graphiques dans la Figure 3.2, on s'aperçoit que les deux transformations correspondant à des taux de croissance sont quasiment identiques. Ceci peut insinuer que le niveau de la population n'a pas d'incidence sur la variation du niveau des émissions des GES. Suite à cette remarque, il est difficile de croire que la dernière transformation peut améliorer la précision des différentes spécifications. Donc, elle peut être écartée pour accélérer le processus, mais ceci n'a pas été effectué étant donné qu'elle correspond à la variable d'intérêt dans les deux articles présentés précédemment.

D'ailleurs, lorsque l'analyse est effectuée avec l'une des transformations, une étape additionnelle doit être réalisée avant de pouvoir comparer les résultats avec ceux obtenus avec les variables en niveau. Effectivement, les prévisions obtenues dans le cadre de ces trois transformations correspondent au taux de croissance des émissions, au niveau des émissions par habitant et au taux de croissance du niveau par habitant. Donc, pour obtenir le niveau des émissions de GES correspondant à ces prévisions, il faut tout

simplement leur appliquer en ordre inverse les étapes suivies pour l'obtention des transformations. Par exemple, lorsque l'analyse est effectuée dans le cadre de la deuxième transformation, la prévision obtenue est multipliée par le niveau des émissions de l'année précédente afin d'obtenir le niveau correspondant à la prévision du taux de croissance. Finalement, il faut mentionner que lors de l'estimation des paramètres, les variables prédictives subissent les mêmes transformations que la variable d'intérêt. Cette procédure est implémentée afin d'avoir la même échelle des deux côtés de l'équation de la spécification estimée.

CHAPITRE 4
RÉSULTATS

4.1 Méthodologie

Des centaines de spécifications des modèles ADL et ARFL sont évaluées avec un exercice pseudo hors échantillon récursif, mais seulement trois d'entre elles sont discutées dans ce chapitre. En plus de faire partie des spécifications engendrant les meilleurs résultats auprès des critères REQM et EAM, celles-ci se démarquent par leur parcimonie ou par le fait qu'elles sont similaires au modèle proposé dans l'article de Bennedsen *et al*. (2021). Avant de présenter les spécifications retenues, une discussion portant sur la démarche suivie avec l'exercice pseudo hors échantillon doit avoir lieu.

Comme indiqué dans le chapitre précédent, la série originale de la variable d'intérêt comporte 32 observations, dont la première correspond à l'année 1990 et la dernière à 2021. Cette période est divisée en deux parties différentes : les dix dernières observations, correspondant plus ou moins au tiers de l'échantillon, sont réservées pour la période d'évaluation hors échantillon, tandis que le reste de la période sert à estimer les paramètres des spécifications. L'exercice hors échantillon commence avec la prévision de l'année 2012. La période servant à l'estimation des paramètres par moindres carrés ordinaires est alors le sous-échantillon contenant les observations des années 1990 à 2011. Ainsi, l'estimation est réalisée avec 21 observations, car une observation est perdue à cause de la nature du processus d'estimation du terme autorégressif dans les spécifications. D'autre part, quand le processus est effectué dans le cadre des deux transformations correspondant à des taux de croissance, l'estimation des paramètres est alors réalisée avec un échantillon de 20 observations. Une fois que les coefficients estimés sont obtenus, une prévision pour l'année 2012 est réalisée et l'erreur de prévision correspondante est calculée. Par la suite, l'observation de l'année 2012 est ajoutée à la période d'estimation et les paramètres sont estimés à nouveau. La prévision pour l'année 2013 est effectuée et l'erreur de prévision correspondante est obtenue. Cette procédure est répétée récursivement jusqu'à la réalisation de la prévision pour l'année 2021. Ainsi, à la fin de l'exercice pseudo hors échantillon, un vecteur contenant 10 prévisions et un vecteur regroupant 10 erreurs de prévision sont obtenus pour chacune des spécifications. Finalement, les erreurs de prévision sont assujetties aux équations (13) à (16) du Chapitre 2 pour obtenir les valeurs des critères REQM et EAM, tout comme les ratios relatifs au modèle de référence.

4.2 Résultats de l'exercice pseudo hors échantillon récursif

4.2.1 Le modèle de référence – Le modèle AR(1)

Le modèle fait bien lors des années avec peu de volatilité, mais il est incapable de prédire les variations déviant significativement de la tendance (voir Figure 4.1). Ceci est un défaut indésirable qui va à l'encontre de l'objectif de la question de recherche. Effectivement, les métriques obtenues lors de l'exercice hors échantillon avec les critères REQM et EAM peuvent être considérées comme étant acceptables. La valeur obtenue auprès du REQM est égale à 21,71 Mt d'équivalent CO_2 et de 12,39 Mt dans le cas du EAM. Ces valeurs correspondent respectivement à 3,2 % et à 1,8 % des Mt d'équivalent CO_2 totales émises au pays en 2021.

L'incapacité du modèle à prédire les variations significatives de la série est expliquée par la nature même du modèle. Par exemple, lors de la diminution significative du niveau des émissions en 2020, le modèle prédit seulement la baisse lors de l'année suivante. Par rapport à la valeur totale des émissions des GES, les valeurs obtenues avec les critères REQM et EAM sont tout de même petites. Toutefois, l'objectif du mémoire est d'identifier un modèle capable d'indiquer les mouvements de la série de façon contemporaine. Ceci est encore plus important lors des années caractérisées par des variations significatives. Par conséquent, l'incorporation de variables exogènes permettant d'exploiter de l'information contemporaine devrait combler les lacunes du modèle de référence.

Figure 4.1 Prévisions obtenues lors de l'exercice pseudo hors échantillon - Modèle AR(1)

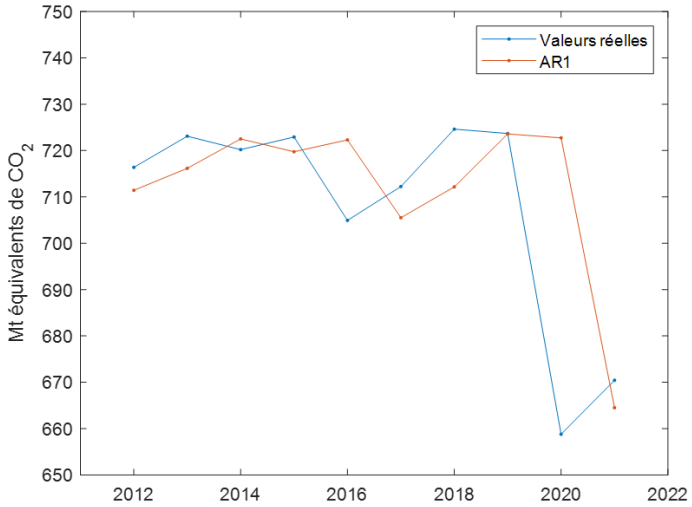


Tableau 4.1 Valeurs obtenues avec les critères de sélection

Modèle	REQM	EAM
AR1	21,71	12,39
ADL_PIB	0,54	0,74
ADL_CONS	0,32	0,41
ARFL_7PCs	0,62	0,81

Les valeurs de la rangée du modèle de référence AR(1) correspondent à des Mt d'équivalent de CO_2. Les chiffres dans les rangées subséquentes représentent les ratios relatifs au modèle de référence AR(1). Ces derniers sont obtenus avec les équations (15) et (16) du Chapitre 2. Lorsque la valeur du ratio est inférieure à 1, cela signifie que le degré de précision atteint avec la spécification en question est plus élevé que celui du modèle AR(1).

4.2.2 Le modèle ADL – Le choix logique du PIB réel

Le PIB réel est ajouté au modèle autorégressif AR(1) de trois façons différentes, mais la spécification retenue est celle incluant son retard et sa valeur contemporaine (Annexe D). Comme dans le cas du modèle AR(1), cette spécification engendre des erreurs de prévision relativement petites au début de la période d'évaluation de l'exercice pseudo hors échantillon. Toutefois, ce qui la différencie du modèle de référence est l'amélioration du niveau de précision de ses prévisions lors des années 2020 et 2021. Effectivement, la Figure 4.2 montre que lors de ces deux années la spécification prévoit correctement une diminution du niveau des émissions des GES, suivie par un rebond. Donc, l'addition de la valeur contemporaine et du retard du PIB réel a l'effet escompté.

Figure 4.2 Prévisions obtenues lors de l'exercice pseudo hors échantillon - PIB réel

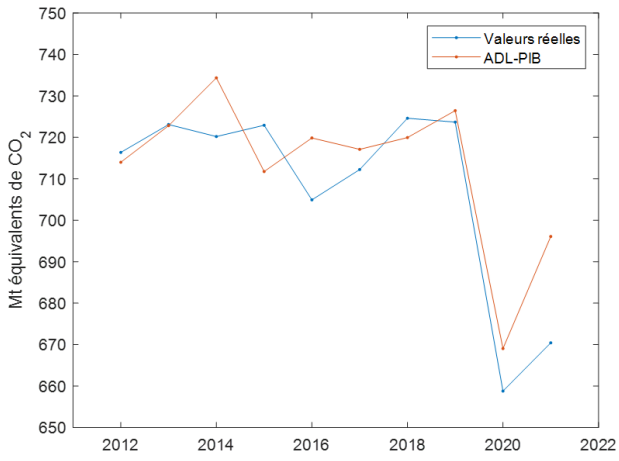


Tableau 4.2 Résultats du test Diebold-Mariano (DM) : comparaison avec le modèle AR(1)

	ADL_PIB	**ADL_CONS**	**ARFL_7PCs**
REQM relatif	0,54	0,32	0,62
valeur-p	*0,44*	*0,32*	*0,39*

La première rangée correspond au ratio relatif du REQM du modèle en question vis-à-vis la valeur du REQM obtunue avec le modèle de référence AR(1). Les erreurs de prévision obtenues avec chacune des spécifications ont été comparées à celles engendrées par le modèle de référence AR(1) afin de produire la valeur-p du test DM.

L'amélioration du niveau de précision est tout aussi visible avec les résultats des critères REQM et EAM illustrés dans la deuxième rangée du Tableau 4.1. Les ratios relatifs s'établissent à 0,54 et à 0,74 selon le REQM et l'EAM respectivement. Ces résultats indiquent une amélioration non négligeable vis-à-vis du modèle AR(1), dont le critère REQM indique que cette spécification est quasiment deux fois plus précise que le modèle de référence. Pour sa part, le test DM indique que le niveau de précision correspondant aux prévisions produites par cette spécification n'est pas statistiquement différent à celui du modèle AR(1). La statistique de test (0,81) est bien en dessous de la valeur critique supérieure du test (1,96), ce qui ne permet pas de rejeter l'hypothèse nulle. Toutefois, l'amélioration visible dans la Figure 4.2 et les valeurs inférieures pour les ratios relatifs justifient les étapes additionnelles qui doivent être effectuées lors de l'estimation des deux termes exogènes de cette spécification.

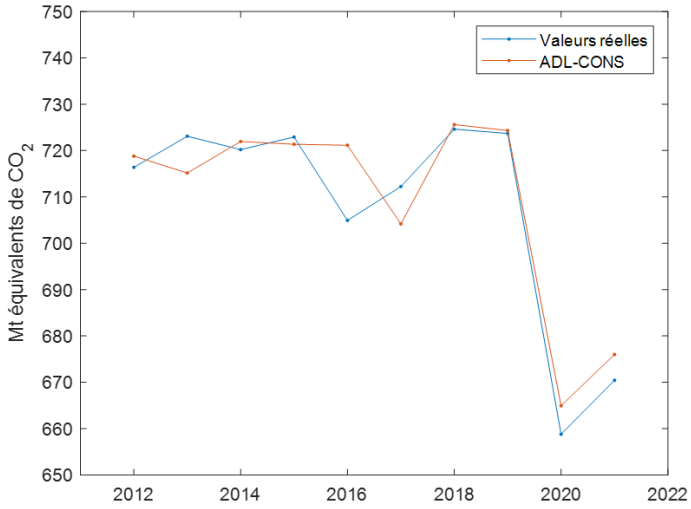
L'amélioration des résultats obtenus avec cette spécification peut être expliquée par l'exploitation d'information provenant de la variation de la série du PIB réel canadien. À titre d'exemple, les émissions des GES ont diminué significativement en 2020 à cause du ralentissement économique provoqué par les mesures sanitaires imposées au pays en raison des conséquences négatives de la pandémie du COVID-19. La baisse des émissions causée par le ralentissement économique fut captée par le modèle AR(1), mais avec une année de retard. La série du PIB réel est un indice agrégé de l'activité économique au pays. Par conséquent, la spécification peut prévoir de façon contemporaine la baisse des émissions qui allait suivre le ralentissement économique grâce à l'information tirée de la diminution du PIB réel. Donc, selon la REQM et l'EAM, l'inclusion de variables exogènes partageant un lien implicite avec la variable d'intérêt permet d'améliorer les prévisions obtenues avec le modèle de référence. Toutefois, la conclusion tirée du test DM indique que l'amélioration n'est pas statistiquement significative. Par conséquent, d'autres spécifications doivent être évaluées.

4.2.3 Le modèle ADL – Spécification exploitant la consommation de pétrole

La variable exogène exploitée dans la spécification du modèle ADL dont il est question dans cette sous-section est la consommation totale de pétrole au pays. Des trois variantes du modèle ADL testées avec l'inclusion de cette variable (Annexe D), c'est la spécification qui combine le premier retard et la valeur contemporaine en niveau qui engendre les plus petites valeurs auprès des critères REQM et EAM. La Figure 4.3 illustre les prévisions obtenues lors de l'exercice hors échantillon avec cette spécification. Comparativement aux prévisions de la spécification du PIB réel, la série dans cette figure diverge moins par rapport aux valeurs réelles de la variable d'intérêt et les erreurs de prévisions semblent plus petites. La dernière remarque est encore plus visible à la fin de la période d'évaluation. Effectivement, une nette amélioration est visible lors des années caractérisées par des variations considérables. La spécification incluant la consommation totale de pétrole prévoit quasiment parfaitement la diminution des émissions des GES en 2020 et le rebond qui s'en suit en 2021.

Dans le Tableau 4.1, le mnémonique représentant cette spécification est « ADL_CONS » et il indique que les valeurs obtenues avec celle-ci sont plus petites que celles obtenues avec la spécification exploitant le PIB réel. Par rapport au modèle de référence, le ratio de la REQM était de 0,32 et de 0,41 dans le cas de l'EAM. Ceci indique que le niveau de précision de cette spécification est 68 % et 59 % plus élevé que le modèle AR(1) selon ces deux critères. Dans le cadre du test DM, la valeur obtenue pour la statistique de test s'établit à 1,05. Cette valeur est plus élevée que la mesure obtenue avec la spécification discutée précédemment, mais elle est bien en dessous de la valeur critique de 1,96. Donc, le niveau de précision n'est pas statistiquement différent que celui du modèle de référence. Par contre, l'excellent niveau de précision atteint lors des dernières années de l'exercice hors échantillon justifie l'addition du retard et de la valeur contemporaine de la consommation de pétrole au terme autorégressif de la spécification.

Figure 4.3 Prévisions obtenues lors de l'exercice pseudo hors échantillon - Consommation de pétrole



Comme c'est le cas avec la spécification incorporant le PIB réel, l'amélioration des résultats obtenus avec la spécification de la consommation de pétrole repose sur l'exploitation de la relation entre la consommation de pétrole et les émissions des GES de nature anthropique. La publication de 2023 du RIN indique que 73 % des émissions des GES couvertes par le rapport sont une conséquence de l'activité économique de secteurs liés à l'énergie, tels que : l'extraction de pétrole et de gaz, l'industrie du transport, et les industries manufacturières. Par conséquent, la consommation de pétrole est une variable qui est liée directement ou indirectement à ces secteurs. Cette relation fait de cette variable un bon prédicteur pour le niveau total des émissions des GES anthropiques.

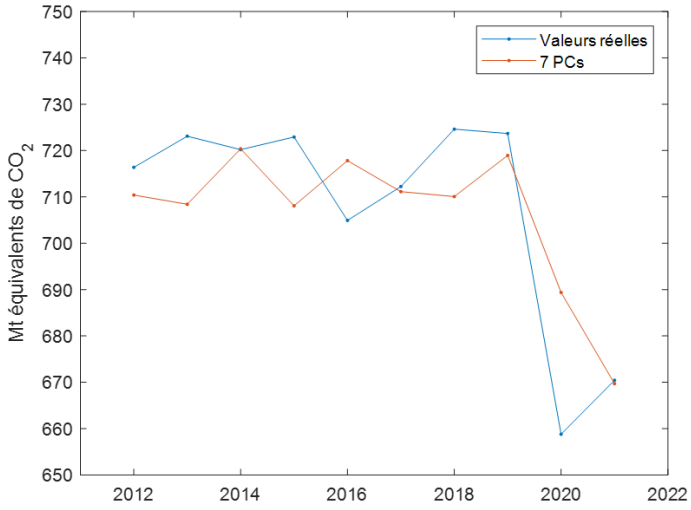
4.2.4 Modèle autorégressif augmenté de facteurs latents – La spécification aux sept premières composantes principales

Parmi les différentes spécifications de ce type de modèle, celle incluant les valeurs contemporaines des sept premières composantes principales, estimées à partir de la base de données LCDMA-M, engendre les meilleurs résultats lors de l'exercice hors échantillon selon les critères REQM et EAM (Annexe E). Les valeurs obtenues résultant de cet exercice sont égales à 13,45 Mt d'équivalent CO_2 dans le cas de la REQM et de 10,03 Mt pour l'EAM. En comparaison aux résultats du modèle de référence, le niveau de précision de cette spécification est 38 % et 19 % plus élevé selon les critères REQM et EAM respectivement.

À première vue, ces résultats semblent satisfaisants, mais lorsqu'ils sont comparés avec les résultats obtenus avec les modèles discutés précédemment, il est clair que cette spécification n'est pas excellente.

En fait, les valeurs des critères de sélection des deux spécifications du modèle ADL sont deux fois plus petites que celles de la spécification dont il est question dans cette sous-section. La faiblesse comparative des résultats peut être observée à l'œil nu dans la Figure 4.4. Lors de la majorité des années, la spécification se trompe sur le signe de la variation de la variable d'intérêt; lorsqu'elle prévoit une diminution du niveau des émissions, la variable d'intérêt augmente et vice versa. Ceci a comme conséquence des erreurs de prévision plus importantes que celles correspondant aux deux spécifications du modèle ADL. De plus, lors de la diminution significative des émissions des GES en 2020, la spécification prévoit correctement celle-ci, mais elle ne parvient pas à indiquer correctement l'ampleur de la diminution. Lors de l'année suivante, le modèle prédit une autre diminution alors que les émissions des GES ont en réalité augmenté avec la reprise économique. Enfin, c'est sans surprise que l'hypothèse nulle du test de robustesse n'est pas rejetée avec cette spécification.

Figure 4.4 Prévisions obtenues lors de l'exercice pseudo hors échantillon - Modèle autorégressif augmenté de sept composantes principales



Tout comme le modèle AR(1), la spécification à sept composantes principales prévoit de façon acceptable le niveau général des émissions des GES, mais elle ne prévoit pas convenablement la variance de la variable d'intérêt. La composante autorégressive du modèle permet à la spécification de réaliser des prévisions autour de la tendance de la variable d'intérêt. D'autre part, les sept termes correspondant aux composantes principales augmentent le niveau de précision des prévisions vis-à-vis la variance des émissions des GES. Ceci est seulement vrai si l'information contenue dans la variation des facteurs est pertinente pour la prévision de la dynamique des émissions des GES. Ainsi, l'interprétation des résultats

obtenus est que l'information tirée des composantes principales est moins adéquate que celle tirée à partir des variables exogènes des spécifications du modèle ADL précédentes. En fait, si on accorde plus d'importance à la parcimonie de la spécification, seulement le premier facteur est pris en considération. L'Annexe E indique que cette spécification engendre des résultats moins bons que celle exploitant les sept premiers facteurs latents, mais elle fait tout de même mieux lors de l'exercice hors échantillon que les spécifications incluant les trois et les cinq premiers facteurs. Donc, parmi les spécifications du modèle ARFL, celle exploitant les sept premières composantes principales obtient les plus petites valeurs avec les critères de sélection, car il s'agit de la spécification avec le nombre plus élevé de termes utilisés, huit termes plus spécifiquement. Ainsi, les résultats acceptables du modèle reposent sur la combinaison du pouvoir explicatif du retard de la variable d'intérêt et du grand nombre de facteurs latents inclus. Étant donné cette réflexion, le grand nombre de termes exploités par cette spécification ne justifie pas le niveau de précision obtenu et c'est pour cette raison qu'elle n'est pas retenue.

4.2.5 Discussion sur les résultats obtenus avec le test de robustesse

Avant de procéder à la présentation de la spécification sélectionnée, il faut aborder un aspect controversé des résultats obtenus avec le test Diebold-Mariano (DM). À première vue, les résultats présentés sont inquiétants lorsqu'on remarque que le niveau de précision d'aucune des spécifications mentionnées est statistiquement différent de celui du modèle de référence. Effectivement, l'hypothèse nulle du test DM n'a pas pu être rejetée à aucun des trois niveaux de significativité (1 %, 5 % et 10 %). Une possible explication pour ce dénouement est le modèle de référence choisi. Étant donné que la différence des fonctions de coût dépend de la précision des deux modèles comparés, si le niveau de précision du modèle de référence est assez élevé, ceci réduit la marge de manœuvre des spécifications considérées. Comme mentionné dans la sous-section 4.2, les erreurs de prévision du modèle de référence sont relativement petites lors des années caractérisées par une légère variance de la série des émissions des GES. Par contre, ses erreurs de prévision augmentent significativement lors des grandes variations cycliques de la série, notamment en 2016 et en 2020. C'est pourquoi l'attrait principal des spécifications du modèle ADL est le fait que les variables exogènes permettent de pallier aux lacunes du modèle autorégressif lorsque des variations de cette magnitude surviennent. Ainsi, les spécifications considérées peuvent seulement se démarquer lors de deux années sur un total de dix. Malgré la magnitude des variations lors de ces deux années, il est clair que ce n'est pas assez pour contrebalancer l'effet des huit autres années de la période d'évaluation de l'exercice pseudo hors échantillon.

Pour augmenter les probabilités de refuser l'hypothèse nulle du test DM, deux éléments sont nécessaires. Premièrement, si la série chronologique de la variable d'intérêt était plus longue, la période d'évaluation de l'exercice pseudo hors échantillon aurait été plus longue. Ceci aurait permis d'inclure plus d'années caractérisées par une variance significative, comme les années correspondant à la crise économique de 2008-09 par exemple (voir Figure 3.1). L'ajout des erreurs de prévision qui se seraient produites lors de ces années aurait sûrement augmenté la différence du niveau de précision entre le modèle de référence et les spécifications considérées.

Tableau 4.3 Résultats du test Diebold-Mariano (DM) : comparaisons avec les modèles univariés

	Statistique	**AR(1)**	**AR(2)**	**ARMA(1,1)**
ADL_PIB	REQM relatif	0,54	0,51	0,39
	valeur-p	*0,42*	*0,30*	*0,12*
ADL_CONS	REQM relatif	0,32	0,30	0,23
	valeur-p	*0,29*	*0,20*	*0,11*
ARFL_7PCs	REQM relatif	0,62	0,59	0,45
	valeur-p	*0,36*	*0,25*	*0,15*

La première rangée correspond au ratio relatif du REQM du modèle en question vis-à-vis la valeur du REQM obtenue avec le modèle de référence AR(1). Les erreurs de prévision obtenues avec chacune des trois spécifications ont été comparées à tour de rôle avec celles engendrées par les trois modèles univariés afin de produire les statistiques du test.

Deuxièmement, le rejet de l'hypothèse nulle du test DM dépend aussi de la pertinence des deux modèles comparés. Le choix du modèle AR(1) comme référence se base sur la comparaison avec les résultats obtenus avec les deux autres modèles univariés. Le modèle autorégressif d'ordre un fait mieux dans l'exercice hors échantillon que les deux autres candidats et c'est pour cela qu'il est choisi. Ainsi, le choix du modèle de référence diminue la probabilité de rejeter l'hypothèse nulle avec les trois spécifications discutées. Effectivement, le Tableau 4.3 illustre cet enjeu. Dans les trois cas, les valeurs-p sont plus élevées lorsque les spécifications sont comparées au modèle AR(1). Lorsque le test est effectué avec le modèle ARMA(1,1), qui est celui qui a fait le moins bien des trois modèles univariés, les valeurs des statistiques sont plus petites et l'hypothèse nulle est quasiment rejetée au niveau de confiance de 10 % avec la spécification exploitant la consommation totale de pétrole. D'autre part, si les spécifications discutées étaient moins performantes, alors il serait possible de rejeter l'hypothèse nulle de cette façon également. Par exemple, certaines spécifications présentées dans les Annexes D et E font tellement mal dans l'exercice hors échantillon que l'hypothèse nulle du test DM est rejetée lorsqu'elles sont comparées au modèle AR(1).

Lorsqu'on observe les graphiques illustrant leurs prévisions, on se rend compte rapidement que ces spécifications ont un faible niveau de précision et que leurs prévisions n'ont aucune valeur. La différence du niveau de précision entre ces spécifications et le modèle de référence est alors statistiquement significative.

4.3 Spécification retenue – La consommation de pétrole (ADL)

Si le choix de la spécification qui est proposée dans le mémoire dépend seulement des deux critères de sélection, le choix est évident. Comme illustré dans le Tableau 4.1, la spécification du modèle ADL exploitant le retard et la valeur contemporaine de la consommation de pétrole est celle qui a obtenu les plus petites métriques auprès des critères REQM et EAM. Toutefois, parmi les centaines de spécifications évaluées, certaines aboutissent à des valeurs similaires ou inférieures auprès de ces deux critères. Pourtant, elles n'ont même pas été mentionnées dans le corps du texte. Ceci est dû au fait qu'il y a d'autres éléments à considérer lors du choix du modèle proposé.

Un aspect important à considérer est la taille des modèles et le nombre de paramètres qui doit être estimé. Dans le cas des spécifications exploitant le PIB réel et la consommation, trois paramètres doivent être estimés : celui correspondant au retard de la variable d'intérêt, ainsi que le retard et la valeur contemporaine de la variable exogène. Le peu de spécifications qui font mieux que celles-ci dans l'exercice hors échantillon nécessitent l'estimation de sept paramètres ou plus. Comme le gain en précision de ces spécifications est marginal, cette hausse n'est pas assez significative pour justifier le nombre supplémentaire de paramètres qui doit être estimé. De plus, un nombre important de ces spécifications incluent également la consommation de pétrole. En effet, le gain marginal de ces spécifications est seulement la conséquence de l'ajout d'une ou de deux variables exogènes supplémentaires à la spécification exploitant la consommation de pétrole. Ainsi, l'amélioration des résultats obtenus repose tout simplement sur le fait que plus de régresseurs sont inclus dans la spécification. Donc, sur la base de parcimonie, la spécification du modèle ADL exploitant l'information disponible dans la variable de la consommation de pétrole est préférable.

Un autre aspect à considérer est la logique derrière les éléments liant les variables exogènes à la variable d'intérêt. La relation exploitée par une spécification dont le régresseur est une variable représentant l'activité économique au pays, comme le PIB réel, est plus facile à comprendre que dans le cas où la spécification est un modèle à facteurs latents par exemple. Effectivement, l'activité économique du pays

devrait être cointégrée avec la quantité de GES rejetés dans l'atmosphère. Dans le même ordre d'idées, on s'attend à ce que la consommation de l'un des intrants les plus polluants au sein de l'activité économique ait un niveau de pouvoir prédictif élevé pour les émissions des GES. Ainsi, la relation logique et théorique qui caractérise la spécification exploitant la consommation de pétrole est un autre attribut en faveur de celle-ci.

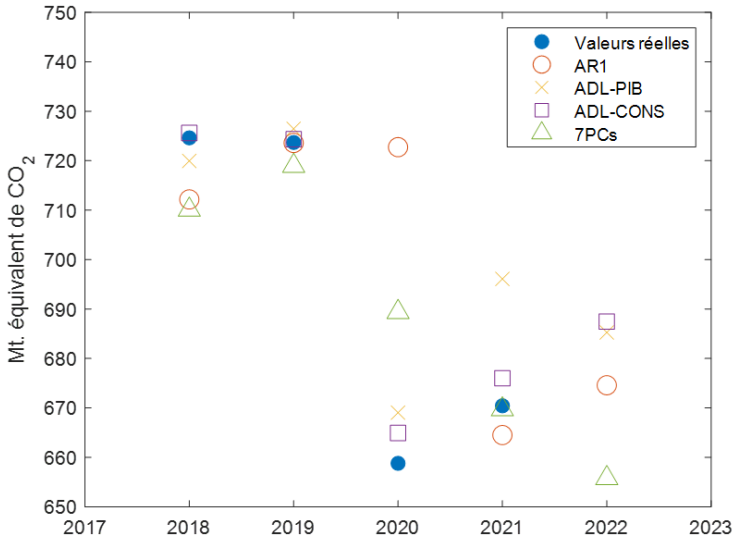
Enfin, le dernier élément à considérer est la possibilité d'utiliser la spécification retenue pour réaliser des prévisions futures du niveau des émissions des GES au lieu de prévisions immédiates. Le niveau de précision bien plus que satisfaisant obtenu avec la spécification exploitant la consommation répond parfaitement aux besoins qui causent la raison d'être de ce mémoire. Toutefois, cette spécification et ce mémoire seraient davantage appréciés s'il est également possible d'effectuer des prévisions futures de court terme. En fait, une fois que les paramètres sont estimés, une prévision future est possible en alimentant la spécification avec des prévisions de la valeur future de la variable exogène. Cet exercice est plus accessible avec la spécification exploitant le PIB réel. Effectivement, à cause de son importance économique, une multitude de prévisions de cette variable provenant du secteur public et du secteur privé sont disponibles. Ceci n'est pas le cas avec la variable portant sur la consommation de pétrole. Par conséquent, la spécification exploitant le PIB réel est préférable selon ce critère de sélection et c'est pour cela qu'elle est présentée et discutée dans cet ouvrage. Toutefois, la réalisation d'une étude portant sur la relation entre l'activité économique et la consommation de pétrole au pays peut rendre possible l'utilisation des prévisions du PIB réel pour obtenir des prévisions futures de la consommation de pétrole. Donc, ceci peut être une solution possible à l'une des faiblesses de la spécification proposée dans le mémoire.

4.4 Prévision rétrospective pour le niveau des émissions en 2022

Une fois le choix de la spécification du modèle ADL annoncé, il est temps de passer à l'application d'intérêt du mémoire. La première étape consiste à ajouter l'observation de la consommation de pétrole de l'année 2022 au panel de variables exogènes. Ensuite, les paramètres de la spécification du modèle ADL exploitant la consommation de pétrole sont estimés par moindres carrés ordinaires. Enfin, la prévision rétrospective pour l'année 2022 est obtenue à l'aide de la valeur des émissions des GES en 2021 et des valeurs de 2021 et 2022 de la consommation de pétrole. Celle-ci indique que le niveau des émissions devrait s'établir à 687,5 Mt d'équivalent de dioxyde de carbone (CO_2) en 2022, ce qui correspond à une hausse de 2,54 % par rapport à la dernière observation publiée. La Figure 4.5, permet de faire la comparaison de cette

prévision avec celles obtenues avec le modèle de référence ainsi qu'avec les deux autres spécifications discutées dans le mémoire (les prévisions des années 2018 à 2021 proviennent de l'exercice pseudo hors échantillon).

Figure 4.5 Comparaison des prévisions rétrospectives pour l'année 2022



Le graphique précédent montre que la spécification choisie prévoit une hausse des émissions des GES en 2022. Ainsi, le rebond qui a débuté en 2021 devrait se poursuivre l'année suivante. Par contre, malgré le fait que la reprise économique a été vigoureuse en 2022, cette spécification indique que le niveau des émissions devrait rester en dessous des niveaux atteints avant la pandémie de 2020.

Un élément rassurant cette prévision est le fait que la spécification exploitant le PIB réel prévoit un niveau des émissions des GES très similaire. Les prévisions de ces deux spécifications ont tendance à être similaires et peu distantes de l'observation réelle. L'année 2021 représente l'une des seules exceptions, car cette année est caractérisée par un écart inhabituel entre les prévisions de ces deux spécifications. Le fait que cet écart soit disparu en 2022 est un élément bienvenu et il augmente le niveau de confiance dans la prévision de la spécification choisie. De plus, le modèle de référence coïncide dans la direction de la prochaine observation de la variable d'intérêt. Effectivement, la prévision du modèle AR(1) indique que les émissions des GES devraient augmenter en 2022. D'autre part, la spécification du modèle autorégressif augmenté de facteurs latents est la seule à prévoir une diminution de la variable d'intérêt. Cette prévision ne fait que s'inscrire dans la tendance baissière des prévisions obtenues avec cette spécification depuis la

crise sanitaire de 2020. Comme cette spécification a obtenu les plus grandes valeurs auprès des critères REQM et EAM parmi les trois spécifications discutées et que la tendance mentionnée précédemment semble peu probable, peu d'importance doit être accordée à cette prévision.

Finalement, un bémol qui vaut la peine d'être rappelé est le fait que la méthodologie et les intrants utilisés pour la production de la série des émissions des GES sont revus à chaque nouvelle publication. Par conséquent, la série des émissions des GES au Canada subit une révision à chaque année. Par conséquent, ceci est un risque pour la précision de la prévision discutée précédemment. Toutefois, le modèle de référence et les deux spécifications du modèle ADL indiquent que la variable d'intérêt devrait augmenter en 2022, alors le niveau de confiance envers cet aspect de la prévision est élevé.

CONCLUSION

Le retard indésiré qui caractérise la publication du RIN ralentit le suivi des émissions des GES et complique l'évaluation des politiques de réduction mises en place par les pays signataires de la Convention-cadre des Nations unies sur les changements climatiques. Compte tenu de l'importance croissante de cet enjeu politique et environnemental, des articles de recherche publiés récemment se sont intéressés à la prévision immédiate des émissions des GES aux États-Unis. À défaut d'avoir des études se penchant sur cet enjeu au Canada, ce mémoire a la particularité de plutôt s'intéresser à la prévision rétrospective du niveau des émissions canadiennes.

L'étude menant à la sélection d'une spécification a commencé par le choix d'un modèle de référence. Parmi les trois modèles univariés considérés, c'est le modèle AR(1) qui a été retenu, car il a obtenu les plus petites valeurs auprès des critères REQM et EAM. Ensuite, des centaines de spécifications différentes des modèles ADL et ARFL ont été évaluées. Chacune d'entre elles a été soumise à un exercice pseudo hors échantillon et les erreurs de prévision obtenues ont servi à calculer les valeurs pour les critères REQM et MAE correspondants. Parmi les spécifications qui se sont démarquées, celle jumelant le retard de la variable d'intérêt avec le retard et la valeur contemporaine de la consommation de pétrole totale au Canada a été choisie. Enfin, cette spécification prévoit qu'en 2022 le niveau des émissions de GES au Canada devrait se situer à 687,5 Mt d'équivalent CO_2.

Ce travail comporte toutefois certaines limites. En plus d'être publiée avec une année de retard, la série des émissions des GES au Canada est à nouveau estimée lors de chaque publication du RIN. Ceci est une conséquence des améliorations méthodologiques effectuées par Environnement et Changement climatique Canada (ECCC) afin d'accroître l'exactitude des estimations. Par exemple, lors de la dernière édition de l'inventaire des GES, les revisions réalisées ont causé une diminution générale de la série (Annexe A), dont « une diminution des émissions de 0,9 à 1,3 pour cent (6 à 9 Mt) des émissions pour la période de 1990 à 2015, et à une diminution légèrement plus grande de 1,4 à 2 pour cent (10 à 14 Mt), pour la période de 2016 à 2020 » (Environnement et Changement climatique Canada, 2023a). Comme l'effet principal a été une translation de la série, ce type de révision n'a pas un effet significatif sur la capacité de la spécification pour effectuer des prévisions satisfaisantes. Toutefois, des améliorations méthodologiques futures pourraient entraîner des mouvements asymétriques de la série, ce qui représente un risque pour la précision du modèle sur le long terme.

Par ailleurs, une limite du processus suivi dans ce mémoire est la méthode de sélection des possibles combinaisons de prédicteurs pour les différentes spécifications considérées. La sélection s'est surtout basée sur les conclusions tirées de la revue de littérature. Effectivement, les types de variables proposés dans les deux articles de recherche ont grandement influencé la liste de variables exogènes considérées dans le mémoire. Même si cette stratégie a engendré des résultats plus que satisfaisants, des méthodes plus rigoureuses telles que LASSO (Tibshirani, 1996), Elastic Net (Zou et Hastie, 2005) et AutoMetrics (Doornik et Hendry, 2015) auraient pu être suivies. Ces procédures auraient permis de mettre des contraintes sur le nombre de variables prédictives à inclure dans les spécifications pour avoir un meilleur contrôle sur le niveau de parcimonie des spécifications. En outre, plusieurs centaines de spécifications ont été évaluées avec le processus utilisé, mais ces méthodes de sélection auraient permis de considérer davantage de spécifications et auraient diminué le risque d'ignorer des combinaisons ayant un niveau de précision élevé.

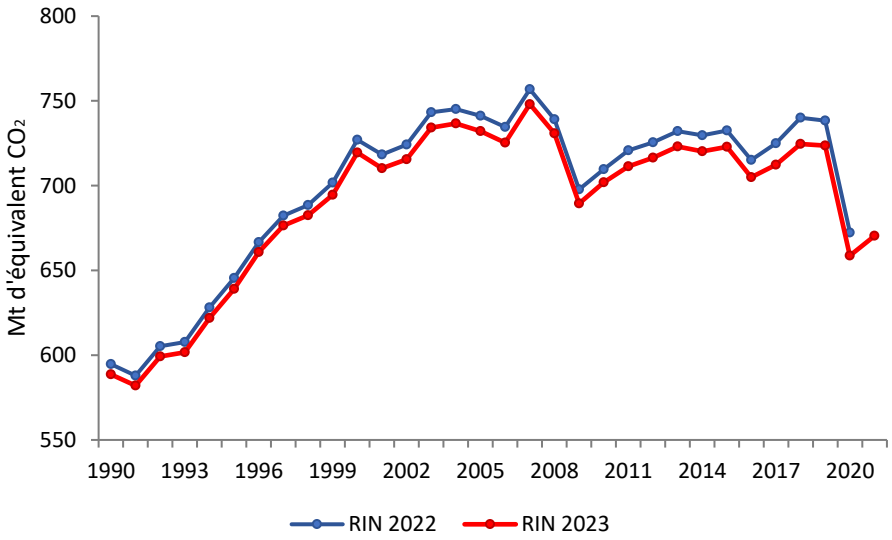
Une amélioration qui pourrait être réalisée est le fait d'utiliser des lois exactes. Comme discuté dans le sous-chapitre 3.1, la variable d'intérêt contient seulement 32 observations. Ceci représente un petit échantillon. Donc, l'utilisation de lois asymptotiques pour les estimateurs, ou bien de statistiques telles que celle de DM, fait partie des limites du mémoire. Ainsi, une possible extension de celui-ci serait de revoir les hypothèses effectuées sur ces lois pour travailler plutôt avec des lois exactes.

Enfin, l'intérêt du mémoire portait sur la prévision rétrospective du niveau national des émissions des GES afin de dynamiser le suivi de cette série. Pourtant, la conception de spécifications se focalisant sur les émissions provinciales et sectorielles permettrait d'obtenir un portrait plus complet et approfondi de la variable d'intérêt ainsi que des enjeux environnementaux et politiques qui en découlent. De plus, ces spécifications pourraient offrir des éléments utiles pour l'amélioration du modèle retenu dans le mémoire.

ANNEXE A

EFFET DES RÉVISIONS MÉTHODOLOGIQUES SUR LA SÉRIE DES ÉMISSIONS DES GES

Figure A.1 Comparaison des tendances d'émissions (RIN 2023 par rapport au RIN 2022)



Source :

Environnement et Changement climatique Canada. (2023a). *Rapport d'inventaire national 1990-2021 : Sources et puits de gaz à effet de serre au Canada (Partie 1)*. Gouvernement du Canada. https://publications.gc.ca/collections/collection_2023/eccc/En81-4-2021-1-fra.pdf

ANNEXE B

VARIABLES PROVENANT DES BASES DE DONNÉES DE LA CHAIRE EN MACROÉCONOMIE ET PRÉVISIONS

Le Tableau B.1 ci-dessous énumère les 115 variables macroéconomiques à fréquence mensuelle qui proviennent de la base de données LCDMA-M. D'autre part, le Tableau B.2 contient la liste des 71 variables trimestrielles de la base de données LCDMA-Q. Les noms et les descriptions des variables correspondent à celles utilisées dans Fortin-Gagnon *et al.* (2020). Les colonnes Vecteur (1) et Vecteur (2) indiquent le numéro de vecteur dans le répertoire de Statistique Canada. Certaines variables doivent être construites à partir de deux séries chronologiques différentes et la colonne Vecteur (1) correspond à la plus récente série des deux. La dernière colonne, intitulée « code », correspond à la transformation appliquée aux séries afin de les rendre stationnaires. Voici la signification de chacune des quatre catégories : (1) pas de transformation, (2) première différence, (4) le logarithme, et (5) première différence des logarithmes.

Tableau B.1 Liste des variables mensuelles

No	Variable	Description	Vecteur (1)	Vecteur (2)	Code
1	GDP_new	GDP total	v41881478	v65201483	5
2	BSI_new	GDP business	v41881479	v65201486	5
3	GPI_new	GDP goods	v41881485	v65201484	5
4	SPI_new	GDP services	v41881486	v65201485	5
5	IP_new	GDP industrial production	v41881487	v65201492	5
6	NDM_new	GDP non durable goods	v41881488	v65201493	5
7	DM_new	GDP durables	v41881489	v65201494	5
8	OILP_new	GDP mining, petrol and gas	v41881501	v65201509	5
9	CON_new	GDP construction	v41881523	v65201531	5
10	RT_new	GDP retail trade	v41881688	v65201641	5
11	WT_new	GDP wholesale trade	v41881689	v65201631	5
12	PA_new	GDP public administration	v41881775	v65201749	5
13	FIN_new	GDP finance and insurance	v41881725	v65201680	5
14	OIL_CAN_new	Crude oil production (Cubic metres)	v17948	v107757044	5
15	EMP_CAN	Employment total	v2057603		5
16	EMP_SERV_CAN	Employment services	v2057610		5
17	EMP_FOR_OIL_CAN	Employment forestry, fishing, mining, oil and gas	v2057606		5
18	EMP_CONS_CAN	Employment construction	v2057608		5
19	EMP_SALES_CAN	Employment sales (wholesale and retail trade)	v2057611		5
20	EMP_FIN_CAN	Employment finance, insurance and real estate	v2057613		5
21	EMP_MANU_CAN	Employment manufacturing	v2057609		5
22	EMP_PART_CAN	Employment part time	v2062813		5
23	UNEMP_CAN	Unemployment rate	LRUNTTTTCAM156S (FRED)	v2062815	2
24	UNEMP_DURA_1.4_CAN	Unemployment duration (1-4 weeks)	v1078667742		5
25	UNEMP_DURA_5.13_CAN	Unemployment duration (5-13 weeks)	v1078667850		5
26	UNEMP_DURA_14.25_CAN	Unemployment duration (14-24 weeks)	v1078667958		5
27	UNEMP_DURA_27._CAN	Unemployment duration (27+ weeks)	v1078668066		5
28	UNEMP_DURAvg_CAN_new	Unemployment average duration	v3433887	v1078668391	5
29	CLAIMS_CAN	Employment insurance initial claims, Allowed	v383942		1
30	TOT_HRS_CAN	Hours worked total	v4391505		5
31	GOOD_HRS_CAN	Hours worked goods	v4391507		5
32	GOOD_OVT_HRS_CAN	Average overtime hours of all employees, Goods-producing sector	v3595115		2
33	NHOUSE_P_CAN	New housing price index, Total (house and land)	v111955442		5
34	hstart_CAN_new	Housing starts (units)	v730413	v52300157	5
35	build_Total_CAN_new	Building permits (tous)	v42061	v121293395	5
36	build_Ind_CAN_new	Building permits (industries)	v42064	v121301795	5
37	build_Comm_CAN_new	Building permits (commerce)	v42065	v121304915	5

No	Variable	Description	Vecteur (1)	Vecteur (2)	Code
38	MANU_N_ORD_new	Manufacturing new orders (total)	v723019	v800913	5
39	MANU_UNFIL_new	Manufacturing unfilled orders (total)	v723313	v803189	5
40	MANU_TOT_INV_new	Manufacturing inventories (total)	v724933	v803227	5
41	MANU_INV_RAT_new	Manufacturing inventories to shipments ratio (total)	v725059	v803313	1
42	N_DUR_INV_RAT_new	Manufacturing inventories to shipments ratio (durables)	v725060	v803314	1
43	DUR_N_ORD_new	Manufacturing new orders (durables)	v723034	v800926	5
44	DUR_UNFIL_new	Manufacturing unfilled orders (durables)	v723328	v803202	5
45	DUR_TOT_INV_new	Manufacturing inventories (durables)	v724948	v803240	5
46	DUR_INV_RAT_new	Manufacturing inventories to shipments ratio (durables)	v725074	v803326	1
47	M3	M3 (gross)	v41552794		5
48	M2p	M2+ (gross)	v41552798		5
49	M_BASE1	Monetary base	v37145		5
50	CRED_BUS_cb	Business loans, Chartered banks only	v36716		5
51	CRED_HOUS_cb	Personal loans, Chartered banks only	v36717		5
52	CRED_MORT_HOUSE_cb	Residential mortgages, total, Chartered banks only	v36724		5
53	CRED_T_cb	Total loans, Chartered banks only	v36723		5
54	CRED_HOUS_non_MORT	Credit liabilities of households, Non-mortgage loans	v1231415611		5
55	CRED_HOUS_MORT	Credit liabilities of households, Residential mortgages	v1231415621		5
56	CRED_HOUS	Credit liabilities of households, Total credit liabilities of households	v1231415625		5
57	CRED_BUS	Credit liabilities of private non-financial corporations, Total credit	v1231415669		5
58	BANK_RATE_L	Bank rate	v122530		2
59	GOV_AVG_1_3Y	Governmental bonds (average rate) (1-3 years)	v122558		2
60	GOV_AVG_3_5Y	Governmental bonds (average rate) (3-5 years)	v122485		2
61	GOV_AVG_5_10Y	Governmental bonds (average rate) (5-10 years)	v122486		2
62	GOV_AVG_10pY	Governmental bonds (average rate) (10+ years)	v122487		2
63	MORTG_1Y	Mortgage rate (1 year)	v122520	BoC (V80691333)	2
64	MORTG_5Y	Mortgage rate (5 years)	v122521	BoC (V80691335)	2
65	TBILL_3M	Treasury bills (3 months)	v122541		2
66	TBILL_6M	Treasury bills (6 months)	v122552		2
67	G_AVG_1.3.Bank_rate	Government bonds (1-3 years) - Bank rate	Difference		1
68	G_AVG_3.5.Bank_rate	Government bonds (3-5 years) - Bank rate	Difference		1
69	G_AVG_5.10.Bank_rate	Government bonds (5-10 years) - Bank rate	Difference		1
70	TBILL_6M.Bank_rate	Treasury bond (6 months) - Bank rate	Difference		1
71	G_AVG_10p.TBILL_3M	Government Bonds (10+ years) - Treasury Bond (3 months)	Difference		1
72	RES_TOT	Total Canada's official international reserves	v122396		5
73	RES_USD	Canadian USD reserves	v122398		5
74	RES_IMF	Canadian reserve position at the IMF	v122401		5
75	Imp_BP_new	Imports total	v183406	v1001826653	5
76	IOIL_BP_new	Imports oil	v183426	v1001826667	5
77	Exp_BP_new	Exports total	v191490	v1001827265	5
78	EOIL_BP_new	Exports oil	v191516	v1001827279	5
79	EX_ENER_BP_new	Export energy products	v191516	v1001827278	5
80	EX_MINER_BP_new	Exports non-metallic ores	v191511	v1001827292	5
81	EX_METAL_BP_new	Exports metal and other mineral products	v191522	v1001827303	5
82	EX_IND_EQUIP_BP_new	Exports industrial machinery, pieces and equipment	v191545	v1001827350	5
83	EX_TRANSP_BP_new	Exports motor vehicules and parts	v191550	v1001827369	5
84	EX_CONS_BP_new	Exports consumption goods	v191492	v1001827385	5
85	IMP_METAL_BP_new	Imports metal and other mineral products	v183446	v1001826691	5
86	IMP_IND_EQUIP_BP_new	Imports industrial machinery, pieces and equipment	v183450	v1001826738	5
87	IMP_TRANSP_BP_new	Imports motor vehicules and parts	v183469	v1001826757	5
88	IMP_CONS_BP_new	Imports consumption goods	v183457	v1001826773	5
89	USDCAD_new	Exchange rate CADUSD	v37426	v111666275	5
90	JPYCAD_new	Exchange rate CADJPY	v37456	v111666258	5
91	GBPCAD_new	Exchange rate CADGBP	v37430	v111666274	5
92	CAN_EQTY_NETFLOW	Canadian equity and investment fund shares, net flows	v61916203		1
93	CAN_SEC_NETFLOW	Canadian securities, Net flows	v61915649		1
94	FOR_SEC_NETFLOW	Foreign securities, Net flows	v61915715		1
95	CAN_US_SEC_NETFLOW	Canadian securities, United States, Net flows	v61915862		1
96	CPI_ALL_CAN	Consumption price index (CPI) (all)	v41690973		5
97	CPI_SHEL_CAN	CPI (shelter)	v41691050		5
98	CPI_CLOT_CAN	CPI (clothing and footwear)	v41691108		5
99	CPI_HEA_CAN	CPI (health and personal care)	v41691153		5
100	CPI_MINUS_FOO_CAN	CPI (all minus food)	v41691232		5
101	CPI_MINUS_FEN_CAN	CPI (all minus food and energy)	v41691233		5
102	CPI_GOO_CAN	CPI (durable goods)	v41691223		5
103	CPI_DUR_CAN	CPI (goods)	v41691222		5
104	CPI_SERV_CAN	CPI (services)	v41691230		5
105	IPPI_CAN	Industrial production price index (IPPI) (all)	v1230995983		5
106	IPPI_ENER_CAN	IPPI (energy)	v1230995996		5
107	IPPI_WOOD_CAN	IPPI (wood)	v1230995994		5
108	IPPI_METAL_CAN	IPPI (metal and construction materials)	v1230995999		5
109	IPPI_MOTOR_CAN	IPPI (motor vehicles and parts)	v1230996000		5
110	IPPI_MACH_CAN	IPPI (industrial machinery and equipment)	v1230996001		5

No	Variable	Description	Vecteur (1)	Vecteur (2)	Code
111	WTISPLC	Petroleum price Western Intermediate Select (WTI) (FRED)	WTISPLC		5
112	TSX_HI	Toronto Stock Exchange (high)	v122618		5
113	TSX_LO	Toronto Stock Exchange (low)	v122619		5
114	TSX_CLO	Toronto Stock Exchange (close)	v122620		5
115	OIL_ALB_new	Crude oil production (ALB) (Cubic metres)	v18050	v107757710	5

Tableau B.2 Liste des variables trimestrielles

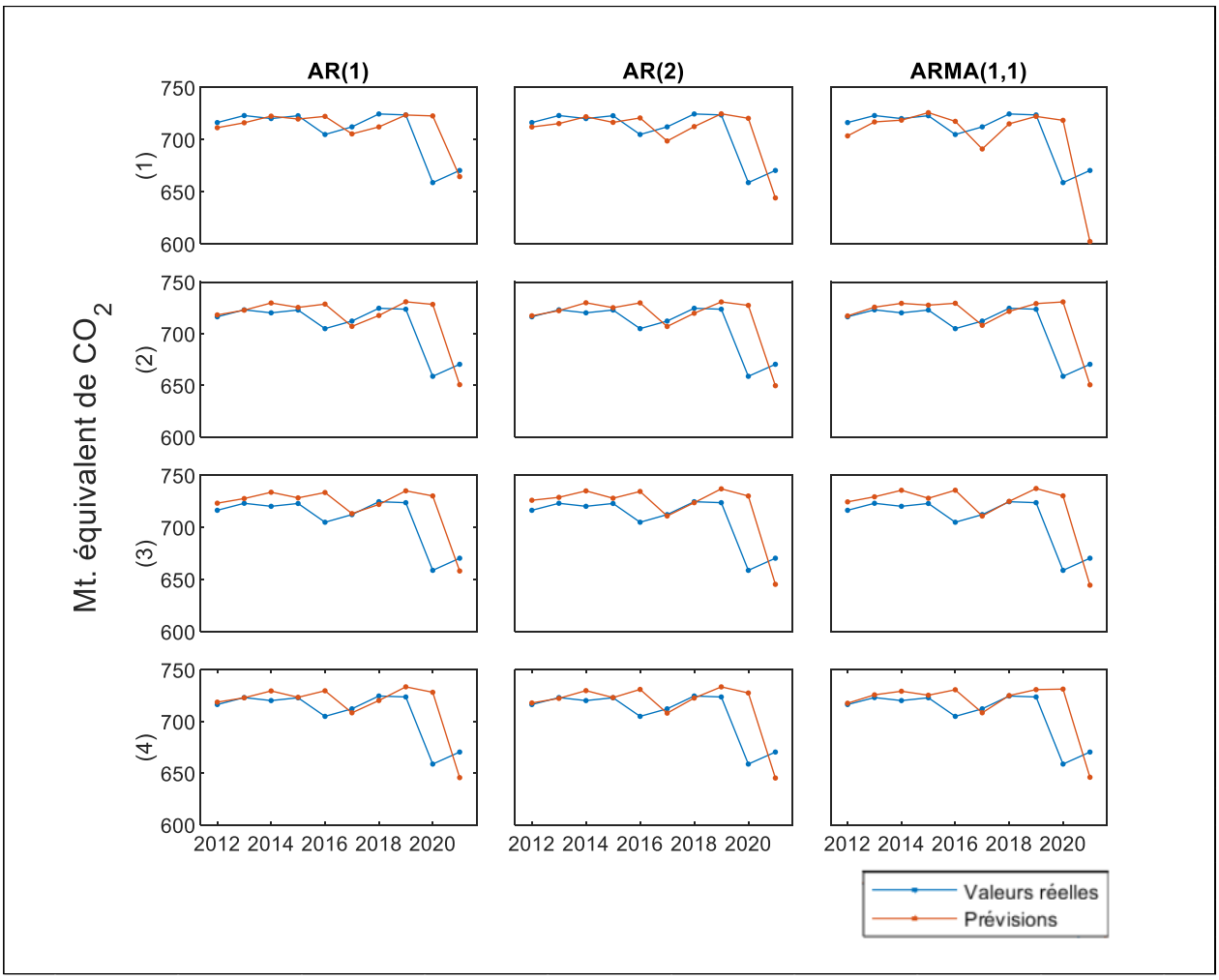
No	Variable	Description	Vecteur (1)	Code
1	REAL_C	Real Final consumption expenditure, chained (2012) dollars	v62305723	5
2	REAL_C_GOODS	Real Final consumption expenditure,Goods, chained (2012) dollars	v62305725	5
3	REAL_C_DUR_GOODS	Real Final consumption expenditure, Durable goods, chained (2012) dollars	v62305726	5
4	REAL_C_SERV	Real Final consumption expenditure, Services, chained (2012) dollars	v62305729	5
5	REAL_C_G	Real General governments final consumption expenditure,, chained (2012) dollars	v62305731	5
6	REAL_I	Real Gross fixed capital formation, chained (2012) dollars	v62305732	5
7	REAL_I_BUS	Real Business gross fixed capital formation, chained (2012) dollars	v62305733	5
8	REAL_I_RES	Real Gross fixed capital formation, Residential structures, chained (2012) dollars	v62305734	5
9	REAL_I_NRES	Real Gross fixed capital formation, Non-residential structures, chained (2012) dollars	v62305736	5
10	REAL_I_ME	Real Gross fixed capital formation, Machinery and equipment, chained (2012) dollars	v62305737	5
11	REAL_I_G	Real Gross fixed capital formation, General governments gross fixed capital formation, chained (2012) dollars	v62305740	5
12	INVEST_INV	Investment in inventories, chained (2012) dollars	v62305741	6
13	REAL_EXP	Real Exports of goods and services, chained (2012) dollars	v62305745	5
14	REAL_EXP_SERV	Real Exports of services, chained (2012) dollars	v62305747	5
15	REAL_IMP	Real Imports of goods and services, chained (2012) dollars	v62305748	5
16	REAL_IMP_SERV	Real Imports of services, chained (2012) dollars	v62305750	5
17	REAL_GDP	Real Gross domestic product , chained (2012) dollars	v62305752	5
18	C_PRICE	Implicit price index : Final consumption expenditure, 2012 = 100	v62307258	5
19	C_PRICE_GOODS	Implicit price index : Final consumption expenditure,Goods, 2012 = 100	v62307260	5
20	C_PRICE_DUR_GOODS	Implicit price index : Final consumption expenditure, Durable goods,2012 = 100	v62307261	5
21	C_PRICE_SERV	Implicit price index : Final consumption expenditure, Services, 2012 = 100	v62307264	5
22	G_PRICE	Implicit price index : General governments final consumption expenditure, 2012 = 100	v62307266	5
23	I_PRICE	Implicit price index : Gross fixed capital formation, 2012 = 100	v62307267	5
24	I_PRICE_BUS	Implicit price index : Business gross fixed capital formation, 2012 = 100	v62307268	5
25	I_PRICE_RES	Implicit price index : Gross fixed capital formation, Residential structures, 2012 = 100	v62307269	5
26	I_PRICE_NRES	Implicit price index : Gross fixed capital formation, Non-residential structures, 2012 = 100	v62307271	5
27	I_PRICE_ME	Implicit price index : Gross fixed capital formation, Machinery and equipment, 2012 = 100	v62307272	5
28	I_PRICE_G	Implicit price index : Gross fixed capital formation, General governments gross fixed capital formation, 2012 = 100	v62307275	5
29	EXP_PRICE	Implicit price index : Exports of goods and services, 2012 = 100	v62307276	5
30	EXP_PRICE_SERV	Implicit price index : Exports of services, 2012 = 100	v62307278	5
31	IMP_PRICE	Implicit price index : Imports of goods and services, 2012 = 100	v62307279	5
32	IMP_PRICE_SERV	Implicit price index : Imports of services, 2012 = 100	v62307281	5
33	GDP_PRICE_IND	Implicit price index : Gross domestic product, 2012 = 100	v62307282	5
34	indus_capu_tot	Industrial capacity utilization rates, Total industrial	v4331081	2
35	indus_capu_for	Industrial capacity utilization rates, Forestry and logging [113]	v4331082	2
36	indus_capu_mine_oil	Industrial capacity utilization rates, Mining and oil and gas extraction [21]	v4331083	2
37	indus_capu_con	Industrial capacity utilization rates ,Construction [23]	v4331087	2
38	indus_capu_manu	Industrial capacity utilization rates, Manufacturing [31-33]	v4331088	2
39	indus_capu_oil	Industrial capacity utilization rates, Petroleum and coal products manufacturing [324]	v4331101	2
40	indus_capu_chem	Industrial capacity utilization rates, Chemical manufacturing [325]	v4331102	2
41	indus_capu_mach	Industrial capacity utilization rates, Machinery manufacturing [333]	v4331109	2
42	JOB_BI	Business Index, 2012 = 100, Total number of jobs	v1409157	5
43	HOURS_AV_BI	Business Index, 2012 = 100, Average hours worked	v1409156	5
44	HOURS_TOT_BI	Business Index, 2012 = 100, Hours worked	v1409155	5
45	LPROD_BI	Business Index, 2012 = 100, Labour productivity	v1409153	5
46	COMP_HOUP_BI	Business Index, 2012 = 100, Total compensation per hour worked	v1409158	5
47	U_LCOST_BI	Business Index, 2012 = 100, Unit labour cost	v1409159	5
48	U_NLABOR_PAY_BI	Business Index, 2012 = 100, Unit non-labour payments	v20805659	5
49	LABOR_SH_BI	Business Index, 2012 = 100, Labour share	v20805661	5
50	COMP_TOT_BI	Business Index, 2012 = 100, Total compensation	v20805663	5
51	NLABOR_PAY_BI	Business Index, 2012 = 100, Non-labour payments	v20805664	5
52	EMP_COMP	Compensation of employees - households	v62468791	5
53	TOT_WAGE_SAL	Wages and salaries	v62468795	5
54	nat_dispo_income	National disposable income	v62305868	5
55	hhold_dispo_income	Households disposable income	v62305869	5
56	corp_dispo_income	Corporations disposable income	v62305871	6

No	Variable	Description	Vecteur (1)	Code
57	nat_saving	National net saving	v62305876	6
58	hhold_saving	Households net saving	v62305877	5
59	corp_saving	Corporations net saving	v62305879	6
60	corp_profit	Corporation profit before taxes	v66462652	5
61	hstart_CAN	House Starts (units)	v730416	5
62	BUILD_IND_NRES	Building construction price indexes, 2017 = 100, 11 census metropolitan area, Non-residential buildings [2362]	v122667915	5
63	BUILD_IND_COMM	Building construction price indexes, 2017 = 100, 11 census metropolitan area, Commercial buildings [23622]	v122667916	5
64	BUILD_IND_WARE	Building construction price indexes, 2017 = 100, 11 census metropolitan area, Warehouse	v122667918	5
65	BUILD_IND_INDUS	Building construction price indexes, 2017 = 100, 11 census metropolitan area, Industrial buildings [23621]	v122667920	5
66	BUILD_IND_FAC	Building construction price indexes, 2017 = 100, 11 census metropolitan area, Factory	v122667921	5
67	LOANS_NBUS_CAN	Non-mortgage loans, to individuals for non-business purposes, total personal loans	v52354203	5
68	LOANS_BUS_CAN	Non-mortgage loans, to individuals and others for business purposes, total loans to Canadian resident non-financial business	v52354207	5
69	MORT_RES_CAN	Mortgages, residential	v52354213	5
70	MORT_NRES_CAN	Mortgages, non-residential	v52354214	5
71	LIAB_ACCEPT_CAN	Customers' liability under acceptances	v52354215	5

ANNEXE C

MODÈLES UNIVARIÉS : RÉSULTATS DE L'EXERCICE HORS ÉCHANTILLON

Figure C.1 Prévisions obtenues lors de l'exercice pseudo hors échantillon



Les rangées correspondent au type de transformation des variables utilisé lors de l'exercice pseudo hors échantillon. Voici les codes des rangées/transformations : (1) Variables en niveau, (2) taux de croissance, (3) en niveau par habitant, et (4) taux de croissance du niveau par habitant.

Conformément aux articles de Bennedsen *et al.* (2021) et de Jensen (2021), le modèle de référence devait être de nature simple. C'est pour cette raison que les candidats étaient des modèles univariés. Le premier concurrent fut le modèle autorégressif d'ordre un AR(1), alors que le modèle AR(2) était le deuxième candidat. Le modèle autorégressif avec moyenne mobile d'ordre un ARMA(1,1) complétait la liste de concurrents. Le choix a été basé sur les métriques obtenues avec les critères REQM et EAM lors de l'exercice hors échantillon. Les paragraphes suivants discutent les résultats de chacun des trois candidats.

Tableau C.1 Comparaison des métriques obtenues avec les modèles univariés (Mt d'équivalent CO_2)

	Critère	AR(1)	AR(2)	ARMA(1,1)
Variables en niveau	REQM	21,71	22,79	30,20
	EAM	12,39	15,06	19,59
Transformation: Taux de croissance	REQM	24,53	24,40	25,19
	EAM	14,65	14,51	14,62
Transformation: Niveau par habitant	REQM	25,44	26,70	26,96
	EAM	15,73	17,65	17,80
Transformation: Taux de croissance du niveau par habitant	REQM	25,03	25,00	25,84
	EAM	14,92	14,80	14,95

C.1 Modèle AR(1) :

Le modèle fait bien lors des années avec peu de volatilité, mais il n'est pas en mesure de prédire les variations exogènes significatives. Les résultats obtenus lors de l'exercice hors échantillon sont acceptables. Le Tableau C.1 illustre que les plus petites valeurs ont été obtenues lorsque l'exercice a été effectué avec le niveau de variables. En d'autres termes, les transformations appliquées aux variables n'ont pas permis d'augmenter le niveau de précision du modèle. Ainsi, la valeur obtenue pour le critère REQM fut de 21,71 Mt d'équivalent CO_2 et pour l'EAM cela a été de 12,39 Mt. Ces métriques correspondent respectivement à 3,24 % et à 1,85 % du niveau des émissions de CO_2 émises au pays en 2021.

Le graphique dans le coin supérieur gauche de la Figure C.1 montre que le modèle performe bien au début de la période d'évaluation. En effet, le modèle AR(1) engendre des prévisions satisfaisantes lorsque la série des émissions des GES est peu volatile, mais il est incapable de prédire la diminution significative de la série en 2020. Étant donné la nature du modèle, celui-ci prédit la baisse avec une année de retard, ce qui va à l'encontre de l'objectif de cet exercice. Par rapport à la valeur totale des émissions des GES, les valeurs obtenues avec les critères REQM et EAM sont petites, mais l'objectif du mémoire est d'identifier un modèle capable d'indiquer les mouvements de la série de façon contemporaine. Ceci est encore plus important lors des années caractérisées par des variations significatives. Par conséquent, il faudrait incorporer au modèle AR(1) une ou plusieurs variables exogènes permettant de combler l'incapacité du modèle à capter les mouvements de façon contemporaine. Enfin, les autres graphiques de la colonne de gauche de la Figure C.1 correspondant aux trois transformations. Ces graphiques indiquent que les

prévisions obtenues avec les transformations sont semblables à celles résultant de la démarche effectuée avec le niveau des variables. Comme les valeurs obtenues pour les critères de sélection étaient plus élevées avec les trois transformations et leurs prévisions étaient incapables de mieux prédire les variations exogènes de manière contemporaine, seulement la version en niveau du modèle AR(1) a été retenue pour le reste de l'analyse.

C.2 Modèle AR(2) :

Lorsqu'on compare les graphiques de la colonne centrale de la Figure C.1 avec ceux de la colonne de gauche, on observe beaucoup de similarités. Malgré l'inclusion d'un retard additionnel, les prévisions de ce modèle partagent les mêmes caractéristiques que celles obtenues avec le modèle AR(1). Effectivement, peu importe si une transformation a été appliquée aux variables ou non, le modèle fait bien au début de l'échantillon d'évaluation, lorsque la volatilité du niveau des émissions de GES est raisonnable, mais le niveau de précision diminue significativement vers la fin de la période de l'échantillon. De plus, les valeurs obtenues avec les critères REQM et EAM sont plus élevées que celles obtenues avec le modèle AR(1) lorsque les variables sont en niveau ou en niveau par habitant. La seule instance où le modèle AR(2) a mieux fait que le modèle AR(1) fut quand l'exercice a été réalisé en taux de croissance, mais ces résultats étaient tout de même moins bons que ceux engendrés par le modèle AR(1) en niveau.

Enfin, l'addition d'un retard additionnel au modèle AR(1) n'a pas eu des effets positifs sur la performance du modèle. En plus d'obtenir des valeurs égales ou supérieures auprès des critères REQM et EAM, les graphiques des prévisions indiquent que le modèle AR(2) partage les mêmes forces et les mêmes faiblesses que celles du modèle AR(1), et ce peu importe l'utilisation ou non de l'une des transformations. Par conséquent, la conclusion tirée à partir des critères REQM et EAM, ainsi que le critère de parcimonie, indique que le modèle AR(1) sera préféré au modèle AR(2).

C.3 Modèle ARMA(1,1) :

Tout comme les deux modèles autorégressifs discutés précédemment, le modèle ARMA(1,1) prédit relativement bien le niveau des émissions des GES lors des périodes stables de la série. Par contre, à la fin de la période d'évaluation, le modèle est incapable de prédire les variations significatives de la variable d'intérêt. En plus de prédire ces déviations avec une année de retard, le modèle surestime la magnitude de la diminution du niveau de la variable d'intérêt. Ceci est reflété par les valeurs obtenues avec le REQM et l'EAM. Le Tableau C.1 indique que le modèle ARMA(1,1) est celui qui a le moins bien fait parmi les trois

candidats et ceci est le cas avec ou sans transformation des variables. Relativement au modèle AR(1), le modèle ARMA(1,1) performe le moins bien lorsque l'analyse est effectuée avec le niveau des variables. Effectivement, les valeurs obtenues étaient 39 % et 58 % plus élevées pour les critères REQM et EAM respectivement.

Comme mentionné précédemment, des trois candidats, le modèle ARMA(1,1) a obtenu les moins bons résultats lors de l'exercice hors échantillon. La composante moyenne-mobile du modèle approxime l'inclusion d'un nombre de retards qui tend vers l'infini. En d'autres termes, cette composante permet d'estimer les résultats qui seraient obtenus en ajoutant un nombre grandissant de retards au modèle autorégressif. Les résultats obtenus avec les critères REQM et EAM, ainsi que les graphiques de la colonne de droite dans la Figure C.1 permettent de conclure que le modèle ARMA(1,1) est moins adéquat que les deux modèles autorégressifs discutés précédemment. Donc, l'ajout de retards additionnels au modèle autorégressif ne rend pas ses prévisions plus précises.

C.4 Choix du modèle de référence :

À la suite de l'évaluation des résultats obtenus avec les trois modèles lors de l'exercice hors échantillon, le modèle AR(1), dans un environnement sans transformations, a été choisi comme modèle de référence pour la suite du processus. En plus d'être le modèle le plus parcimonieux, il a obtenu les plus petites valeurs auprès des critères REQM et EAM. Par conséquent, les modèles des prochaines sections auront comme objectif la réalisation de prévisions plus précises que celles obtenues avec ce modèle univarié.

ANNEXE D

RÉSULTATS DES DEUX SPÉCIFICATIONS DU MODÈLE ADL

La section 2.3 du deuxième chapitre a présenté le modèle autorégressif à retards échelonnés (ADL). Voici un rappel de sa forme algébrique :

$$y_t = c + \Phi_1 \, y_{t-1} \; + \; \beta(L) \, x_t \; + \; \varepsilon_t \, . \tag{5}$$

Dans ce chapitre, il a également été indiqué que l'un des éléments différentiateurs des spécifications de ce modèle était le nombre de retards et de valeurs contemporaines de la variable exogène x_t. Trois différentes variantes ont été évaluées avec chacune des variables exogènes. Afin de faciliter la présentation des résultats, celles-ci ont été nommées ADL0, ADL1 et ADL2. Voici l'équation correspondant à chacune d'entre elles :

$$y_t = c + \Phi_1 \, y_{t-1} \; + \; \beta_1 \, x_t \; + \; \varepsilon_t \tag{$ADL0$}$$

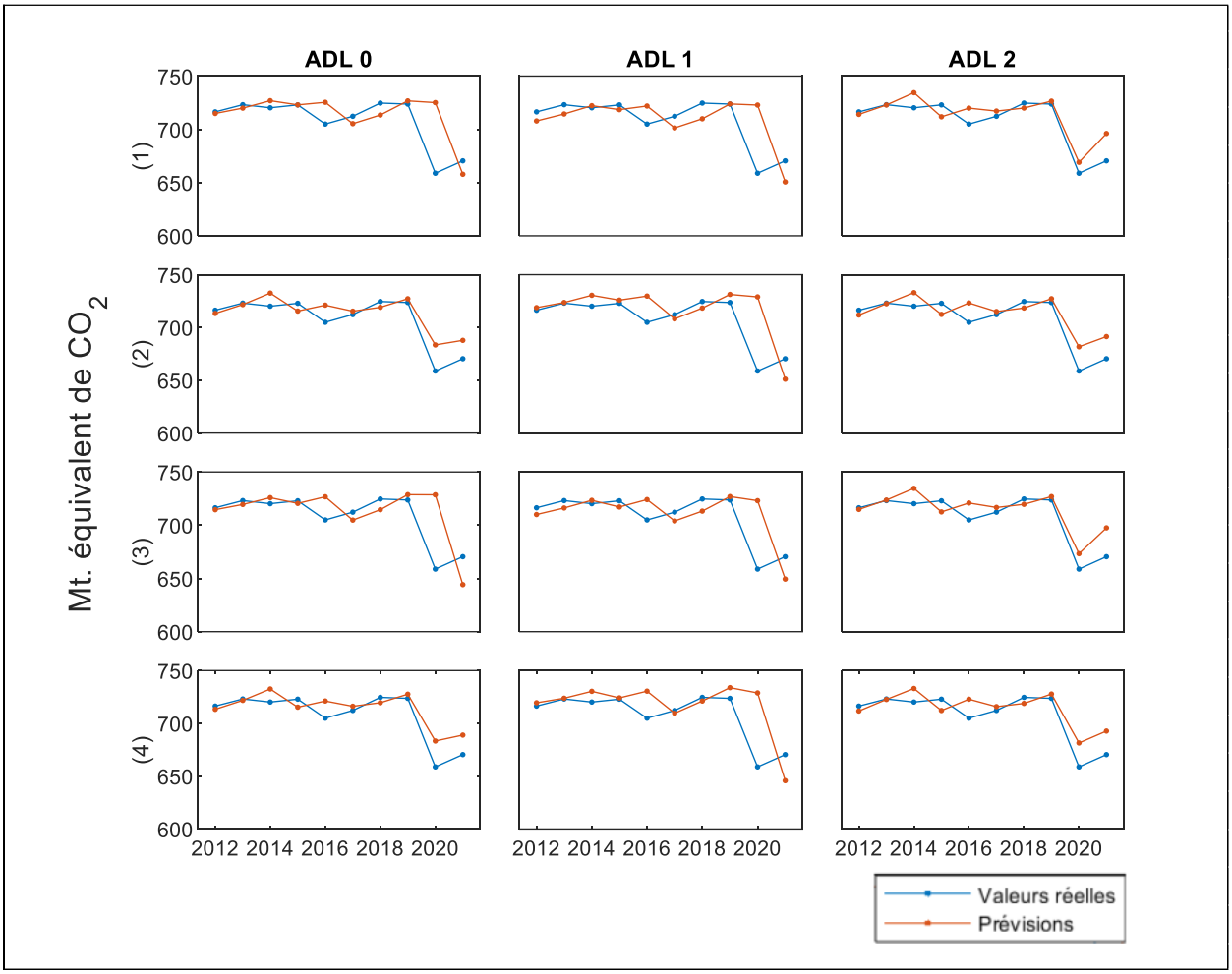
$$y_t = c + \Phi_1 \, y_{t-1} \; + \; \beta_1 x_{t-1} \; + \; \varepsilon_t \tag{$ADL1$}$$

$$y_t = c + \Phi_1 \, y_{t-1} \; + \; \beta_1 \, x_t + \; \beta_2 x_{t-1} \; + \; \varepsilon_t \, . \tag{$ADL2$}$$

Les sous-sections suivantes fournissent une présentation exhaustive des 12 résultats obtenus avec chacune des variables exogènes en question. Rappelons qu'en plus de réaliser l'exercice avec la variable en niveau, le processus a été effectué à trois reprises additionnelles avec des transformations du niveau de la variable.

D.1 Spécification du modèle ADL – PIB réel

Figure D.1 Prévisions obtenues avec les spécifications du modèle ADL exploitant le PIB réel



Les rangées correspondent au type de transformation des variables utilisé lors de l'exercice pseudo hors échantillon. Voici les codes des rangées/transformations : (1) Variables en niveau, (2) taux de croissance, (3) en niveau par habitant, et (4) taux de croissance du niveau par habitant.

Tableau D.1 Ratios relatifs au modèle de référence AR(1) – PIB réel

Transformation	Critère	ADL0	ADL1	ADL2
Variables en niveau	REQM	*1,05*	*1,06*	*0,54*
	EAM	*1,06*	*1,21*	*0,74*
Transformation: Taux de croissance	REQM	*0,55*	*1,14*	*0,59*
	EAM	*0,77*	*1,20*	*0,83*
Transformation: Niveau par habitant	REQM	*1,15*	*1,06*	*0,58*
	EAM	*1,24*	*1,21*	*0,78*
Transformation: Taux de croissance du niveau par habitant	REQM	*0,56*	*1,17*	*0,60*
	EAM	*0,78*	*1,22*	*0,85*

Les valeurs représentent les ratios relatifs au modèle de référence AR(1) obtenus à l'aide des équations (15) et (16) du Chapitre 2. Lorsque la valeur du ratio est inférieure à 1,00, cela signifie que le modèle en question a été plus précis que le modèle AR(1) lors de l'exercice pseudo hors échantillon.

Tableau D.2 Résultats du test Diebold-Mariano – Spécifications utilisant le PIB réel

Transformation	Statistique	ADL0	ADL1	ADL2
Variables en niveau	t	-1,48	-1,63	0,81
	Valeur-p	*0,17*	*0,14*	*0,44*
Transformation: Taux de croissance	t	0,93	-1,59	0,84
	Valeur-p	*0,38*	*0,15*	*0,42*
Transformation: Niveau par habitant	t	-1,62	-1,41	0,78
	Valeur-p	*0,14*	*0,19*	*0,46*
Transformation: Taux de croissance du niveau	t	0,91	-1,71	0,82
	Valeur-p	*0,39*	*0,12*	*0,43*

Les erreurs de prévision obtenues avec chacune des spécifications ont été comparées à celles engendrées par le modèle de référence AR(1) afin de produire les statistiques du test.

Le PIB réel a été ajouté au modèle autorégressif AR(1) de trois façons différentes. Ceci a permis de comparer trois variantes de la spécification du modèle ADL exploitant cette variable exogène. La spécification qui a été retenue fut celle incluant le retard et la valeur contemporaine du PIB réel, dénommée ADL2 dans le graphique et les tableaux précédents. Comme cela a été le cas avec le modèle de référence, cette spécification engendre des erreurs de prévision relativement petites au début de l'échantillon d'évaluation. Toutefois, ce qui le différencie du modèle de référence est l'amélioration de la

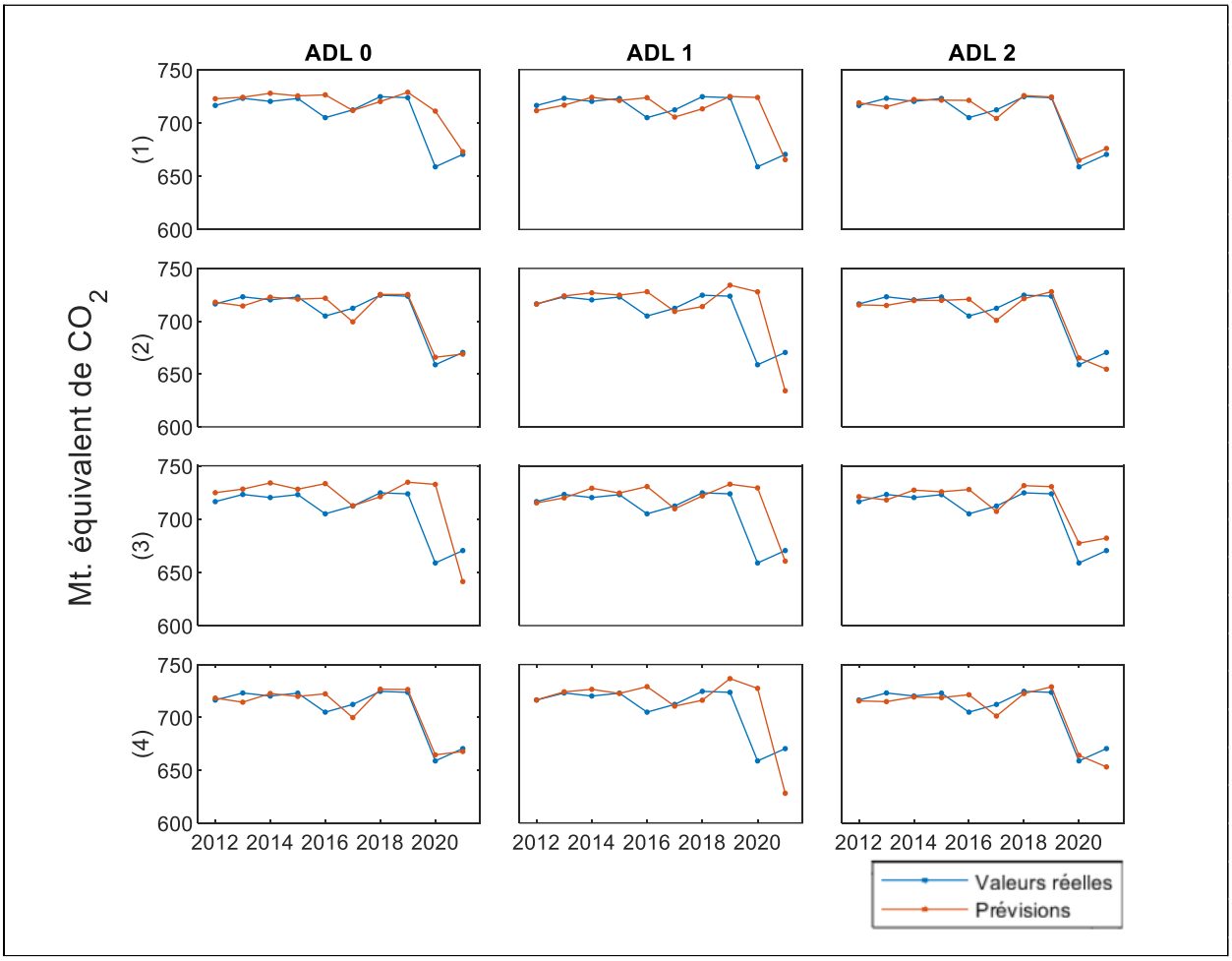
précision des prévisions lors des années 2020 et 2021. Effectivement, le graphique dans la première rangée et la troisième colonne de la Figure D.1 illustre que le modèle prévoit correctement une diminution du niveau des émissions de GES suivie par un rebond. L'amélioration de la précision est encore plus perceptible avec les résultats de l'exercice hors échantillon affichés dans le Tableau D.1. Les ratios relatifs au modèle de référence qui ont été obtenus avec la spécification ADL 2 (en niveau) se sont établis à 0,54 et 0,74 pour les critères REQM et EAM, respectivement. Ces résultats indiquent une nette amélioration, dont le ratio du critère REQM indique que la spécification retenue a été deux fois plus précise que le modèle de référence. Pour sa part, le test DM indique que les prévisions produites par le modèle ADL ne sont pas statistiquement différentes que celles du modèle AR(1). Le Tableau D.2 indique que la statistique de test (0,81) était bien en dessous de la valeur critique et c'est pour cela que l'hypothèse nulle ne peut même pas être rejetée à un niveau de confiance de 10 %.

L'amélioration du niveau de précision obtenu avec cette spécification peut être expliquée par l'exploitation d'information fournie par la variation de la série du PIB réel. À titre d'exemple, les émissions des GES ont diminué significativement en 2020 à cause du ralentissement économique provoqué par les mesures sanitaires imposées au pays en raison des conséquences négatives de la pandémie du COVID-19. La baisse des émissions causée par le ralentissement économique fut captée par le modèle AR(1), mais avec une année de retard. La série du PIB réel est un indice agrégé de l'activité économique au pays. Par conséquent, le modèle a pu prévoir de façon contemporaine la baisse des émissions qui allait suivre le ralentissement économique grâce à l'information tirée de la diminution du PIB réel. Ainsi, l'inclusion de variables exogènes ayant un lien implicite avec la variable d'intérêt permet d'améliorer les performances obtenues avec le modèle de référence selon les valeurs obtenues avec les critères REQM et EAM lors de l'exercice pseudo hors échantillon. Toutefois, la conclusion tirée du test DM indique que l'amélioration n'est pas statistiquement significative. Par conséquent, d'autres spécifications ont dû être étudiées.

Par sa nature, le PIB réel était un candidat logique lors du processus d'identification de potentielles variables exogènes, mais il était fort possible que d'autres variables engendrent des spécifications bien plus précises. La section suivante présente les résultats obtenus avec les différentes variantes/transformations de la spécification exploitant la consommation de pétrole nationale.

D.2 Spécification du modèle ADL – Consommation de pétrole au Canada

Figure D.2 Prévisions obtenues avec les spécifications du modèle ADL exploitant la consommation de pétrole



Les rangées correspondent au type de transformation des variables utilisé lors de l'exercice pseudo hors échantillon. Voici les codes des rangées/transformations : (1) Variables en niveau, (2) taux de croissance, (3) en niveau par habitant, et (4) taux de croissance du niveau par habitant.

Tableau D.3 Ratios relatifs au modèle de référence AR(1) – Consommation de pétrole

Transformation	Critère	ADL0	ADL1	ADL2
Variables en niveau (1)	REQM	*0,84*	*1,02*	*0,32*
	EAM	*0,84*	*1,01*	*0,41*
Transformation: Taux de croissance (2)	REQM	*0,35*	*1,21*	*0,41*
	EAM	*0,45*	*1,31*	*0,57*
Transformation: Niveau par habitant (3)	REQM	*1,26*	*1,12*	*0,51*
	EAM	*1,44*	*1,09*	*0,74*
Transformation: Taux de croissance du niveau par habitant (4)	REQM	*0,36*	*1,25*	*0,42*
	EAM	*0,48*	*1,34*	*0,58*

Les valeurs représentent les ratios relatifs au modèle de référence AR(1) obtenus à l'aide des équations (15) et (16) du Chapitre 2. Lorsque la valeur du ratio est inférieure à 1,00, cela signifie que le modèle en question a été plus précis que le modèle AR(1) lors de l'exercice pseudo hors échantillon.

Tableau D.4 Résultats du test Diebold-Mariano - Spécifications utilisant la consommation de pétrole

Transformation	Statistique	ADL0	ADL1	ADL2
Variables en niveau	t	0,99	-1,01	1,05
	Valeur-p	*0,35*	*0,34*	*0,32*
Transformation: Taux de croissance	t	1,02	-1,59	0,97
	Valeur-p	*0,33*	*0,15*	*0,36*
Transformation: Niveau par habitant	t	-1,88	-1,27	0,92
	Valeur-p	*0,09*	*0,24*	*0,38*
Transformation: Taux de croissance du niveau	t	1,01	-1,49	0,95
	Valeur-p	*0,34*	*0,17*	*0,37*

Les erreurs de prévision obtenues avec chacune des spécifications ont été comparées à celles engendrées par le modèle de référence AR(1) afin de produire les statistiques du test.

Trois spécifications se démarquent dans la Figure D.2 : la spécification en niveau du type ADL2, la spécification en taux de croissance de type ADL0, et la version de cette dernière en taux de croissance par habitant. On remarque que leurs graphiques indiquent que leurs prévisions ont été précises lors des deux dernières années de l'échantillon d'évaluation. À l'œil nu, il n'est pas clair laquelle de ces spécifications est légèrement meilleure que les deux autres. Le Tableau D.3 affiche les ratios relatifs de ces trois spécifications en comparaison du modèle de référence sur la base des critères REQM et EAM. Donc, c'est la spécification exploitant la valeur contemporaine et le retard (ADL2) du niveau de la consommation de

pétrole qui a obtenu les meilleurs résultats. Ces derniers peuvent être expliqués par le fait que la spécification exploite un terme de plus que les deux autres. Sur la base du critère de la parcimonie, les deux autres spécifications seraient alors préférables. Toutefois, l'application de transformations aux variables cause l'addition d'une ou de deux étapes supplémentaires lors de l'estimation de ces spécifications. Par conséquent, comme il est beaucoup plus simple de travailler avec le niveau des variables, la spécification en niveau ADL2 de la consommation de pétrole est retenue. Le corps du texte analyse plus amplement les résultats de celle-ci.

ANNEXE E

RÉSULTATS DES SPÉCIFICATIONS DU MODÈLE ARFL

Le corps principal du mémoire a seulement présenté et discuté les résultats obtenus avec la spécification exploitant les sept premiers facteurs latents tirés de la base de données mensuelles LCDMA-M. Cette Annexe présente les résultats obtenus avec d'autres spécifications, ainsi que les spécifications utilisant les facteurs latents tirés de la base de données trimestrielles LCDMA-Q.

La sous-section 2.4 du deuxième chapitre a présenté le processus d'estimation des facteurs latents par composantes principales et l'équation (12) illustrait la forme algébrique de la spécification du modèle autorégressif augmenté de facteurs latents (ARFL) dont les résultats ont été présentés dans le Chapitre 4. Afin d'être cohérents avec le processus suivi avec le modèle ADL, les mêmes types de spécifications ont été évalués dans le cas du modèle ARFL. Plus précisément, certaines spécifications ont inclus les retards des facteurs latents, tandis que d'autres ont exploité une combinaison des valeurs contemporaines et des retards des facteurs estimés. Afin d'alléger la notation dans la présentation des résultats, une nomenclature semblable à celle de l'Annexe D sera utilisée dans les graphiques et tableaux ci-dessous. Voici les formes algébriques générales de ces trois types de spécifications avec leurs nomenclatures entre parenthèses. :

$$y_t = c + \beta_1 y_{t-1} + \theta_1 \hat{F}_t + e_t \qquad (ARFL0)$$

$$y_t = c + \beta_1 y_{t-1} + \theta_1 \hat{F}_{t-1} + e_t \qquad (ARFL1)$$

$$y_t = c + \beta_1 y_{t-1} + \theta_1 \hat{F}_t + \theta_2 \hat{F}_{t-1} + e_t \,. \qquad (ARFL2)$$

Comme mentionné précédemment, les bases de données mensuelles et trimestrielles n'ont pas été jumelées, alors l'exercice a été réalisé à tour de rôle avec chacune d'entre elles. Quand la variable d'intérêt était en niveau, aucune transformation n'a été appliquée aux séries avant d'être standardisées. Lorsque la variable d'intérêt correspondait au taux de croissance des émissions des GES, les taux de croissance des variables ont été calculés et ceux-ci ont été standardisés par la suite. La même procédure a été suivie lorsque la variable d'intérêt correspondait au niveau des GES par habitant ou au taux de croissance de ce dernier. Il faut noter que la série sur la population canadienne est de fréquence trimestrielle, alors des

facteurs ont seulement pu être estimés avec la base de données LCDMA-Q lorsque la variable d'intérêt était en niveau par habitant ou en taux de croissance de ce dernier.

E.1 Le modèle autorégressif augmenté de facteurs latents (ARFL) – Base de données mensuelles

La publication du mois de mai 2023 des deux bases de données de la Chaire en macroéconomie et prévisions a été utilisée pour estimer les facteurs. La première observation disponible des variables dans la base de données LCDMA-M correspondait à janvier 1981 et la dernière à avril 2023. Au total, chaque série avait un total de 508 observations. La base de données regroupe des variables nationales et provinciales, mais ces dernières ont été exclues du processus d'estimation des facteurs. Donc, le panel final contenait 115 séries. La première année de la période d'évaluation de l'exercice pseudo hors échantillon correspondait à 2012. Par conséquent, la période débutant en janvier 1981 et finissant en décembre 2011 a été utilisée dans l'estimation des facteurs par composantes principales[6]. Comme les estimations obtenues étaient de fréquence mensuelle, leur moyenne annuelle a été calculée. Enfin, comme la période couverte par la variable d'intérêt débute seulement en 1990, les moyennes annuelles des facteurs estimés de 1991 à 2011 ont été jumelées au retard de la variable d'intérêt afin d'estimer les paramètres des différentes spécifications du modèle ARFL. Finalement, comme indiqué précédemment, il n'a pas été possible d'estimer les facteurs lorsque l'exercice a été effectué en niveau par habitant ou en taux de croissance de celui-ci, car la série sur la population canadienne est de fréquence trimestrielle.

Lorsqu'on analyse les graphiques de la Figure E.1 et les résultats affichés dans le Tableau E.1, la raison pour laquelle seulement la spécification exploitant les sept premières composantes principales a été discutée dans le corps du texte est assez claire. En plus d'être celle ayant obtenu les plus petites valeurs auprès des critères REQM et EAM, elle est parmi les seules à prédire une diminution du niveau des émissions des GES en 2020. De toute évidence, la majorité des spécifications des types ARFL(1) et ARFL(2) n'ont pas prédit cette diminution et le font avec une année de retard. Cette observation avait également était fait avec les modèles univariés, alors l'ajout des facteurs latents dans ces cas n'a pas engendré les résultats recherchés. Effectivement, à quelques exceptions près, la majorité des spécifications ont moins bien fait dans l'exercice pseudo hors échantillon selon le Tableau E.1 comparativement au modèle de référence AR(1). Tout comme pour les spécifications qui ont obtenu de meilleurs résultats, le Tableau E.2 indique que le niveau

[6] Lorsque la variable d'intérêt correspondait au taux de croissance des émissions des GES, la matrice contenant les variables macroéconomiques a perdu une observation. La période utilisée pour l'estimation des facteurs était alors de février 1981 à décembre 2011.

de précision des prévisions des spécifications moins bonnes n'est tout de même pas différent statistiquement à celle du modèle AR(1). Donc, il faut que le niveau de précision soit encore plus précaire pour pouvoir rejeter l'hypothèse nulle.

D'autre part, parmi les spécifications qui ont obtenu des valeurs plus petites que 1,00 dans le Tableau E.1, la spécification exploitant seulement la première composante principale aurait pu être choisie pour la discussion du corps du mémoire au lieu de celle exploitant les sept premières composantes principales. En effet, si l'on accorde plus d'importance à la parcimonie des spécifications, celle-ci aurait dû être choisie. Dans ce cas, l'argument principal aurait été que le gain marginal de l'inclusion des six composantes principales suivantes ne justifie pas la prolifération des termes de la spécification. Le choix de plutôt retenir la spécification utilisant les sept composantes principales pour la discussion du corps du mémoire s'est basé sur les plus petites valeurs obtenues dans l'exercice hors échantillon, car bien des spécifications du modèle ADL ont engendré des valeurs petites que celles de la spécification à une seule composante principale. Ainsi, les meilleures spécifications du modèle ADL allaient être comparées à la meilleure spécification du modèle ARFL.

Figure E.1 Prévisions obtenues avec les spécifcations du modèle ARFL – Variables en niveau du panel LCDMA-M

Tableau E.1 Ratios relatifs au modèle de référence AR(1) – Variables en niveau du panel LCDMA-M

Modèle	Critère	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	REQM	*0,79*	*1,01*	*0,77*
	EAM	*0,90*	*1,05*	*0,86*
ARFL_3PCs	REQM	*0,86*	*1,06*	*1,00*
	EAM	*0,97*	*1,06*	*0,89*
ARFL_5PCs	REQM	*0,84*	*1,07*	*1,10*
	EAM	*1,01*	*1,09*	*1,06*
ARFL_7PCs	REQM	*0,62*	*1,10*	*0,91*
	EAM	*0,81*	*1,17*	*1,00*

Les valeurs représentent les ratios relatifs au modèle de référence AR(1) qui ont été obtenus avec les équations (15) et (16) du Chapitre 2. Lorsque la valeur du ratio est inférieure à 1,00, cela signifie que le modèle en question a été plus précis que le modèle AR(1) lors de l'exercice pseudo hors échantillon.

Tableau E.2 Résultats du test Diebold-Mariano – Variables en niveau du panel LCDMA-M

Modèle	Statistique	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	t	1,00	-0,18	1,09
	Valeur-p	*0,34*	*0,86*	*0,30*
ARFL_3PCs	t	0,81	-0,84	-0,08
	Valeur-p	*0,44*	*0,42*	*0,94*
ARFL_5PCs	t	0,80	-0,88	-0,84
	Valeur-p	*0,44*	*0,40*	*0,42*
ARFL_7PCs	t	0,91	-1,09	0,75
	Valeur-p	*0,39*	*0,31*	*0,47*

Les erreurs de prévision obtenues avec les spécifications énumerées dans la première colonne ont été comparées à celles engendrées par le modèle de référence AR(1) afin de produire les statistiques du test.

La Figure E.2 et les tableaux subséquents illustrent les résultats obtenus avec des spécifications du modèle ARFL exploitant des composantes principales estimées à partir des taux de croissance des variables du panel final de la base de données LCDMA-M. Contrairement à l'exercice effectué avec le niveau des variables, les spécifications utilisant la première composante sont celles qui ont obtenu les plus petites valeurs auprès des critères REQM et EAM selon le Tableau E.3. Par contre, ces valeurs sont très proches de 1,00 et ceci indique que l'addition de la première composante principale a seulement eu un effet très marginal sur le niveau de précision du modèle AR(1). Ceci est illustré par la Figure E.2 où on s'aperçoit que cet ensemble de spécifications est incapable de prédire de façon contemporaine les variations significatives de la variable d'intérêt. Pourtant, l'inclusion des facteurs latents devait aider à combler cette lacune caractérisant le modèle de référence. Par ailleurs, les résultats du test DM dans le Tableau E.4 montrent que certaines de ces spécifications ont fait significativement moins bien dans l'exercice hors échantillon. Certaines d'entre elles ont engendré des erreurs de prévision assez grandes pour rejeter pour une fois l'hypothèse nulle du test au niveau de confiance de 10 %. Enfin, après tous ces éléments de réflexion, il est clair qu'aucune de ces spécifications ne devait être retenue pour la discussion du mémoire.

Figure E.2 Prévisions obtenues avec les spécifcations du modèle ARFL – Taux de croissance des variables du panel LCDMA-M

Tableau E.3 Ratios relatifs au modèle de référence AR(1) – Taux de croissance des variables du panel LCDMA-M

Modèle	Critère	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	REQM	*0,98*	*0,99*	*0,95*
	EAM	*0,97*	*0,99*	*0,94*
ARFL_3PCs	REQM	*1,01*	*0,96*	*1,02*
	EAM	*1,01*	*0,98*	*1,00*
ARFL_5PCs	REQM	*1,06*	*1,02*	*1,20*
	EAM	*1,27*	*1,08*	*1,47*
ARFL_7PCs	REQM	*1,10*	*1,04*	*1,23*
	EAM	*1,35*	*1,11*	*1,42*

Les valeurs représentent les ratios relatifs au modèle de référence AR(1) qui ont été obtenus avec les équations (15) et (16) du Chapitre 2. Lorsque la valeur du ratio est inférieure à 1,00, cela signifie que le modèle en question a été plus précis que le modèle AR(1) lors de l'exercice pseudo hors échantillon.

Tableau E.4 Résultats du test Diebold-Mariano – Taux de croissance des variables du panel LCDMA-M

Modèle	Statistique	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	t	-1,44	-1,44	-1,43
	Valeur-p	*0,18*	*0,18*	*0,19*
ARFL_3PCs	t	-1,66	-1,60	-1,63
	Valeur-p	*0,13*	*0,14*	*0,14*
ARFL_5PCs	t	-1,86	-1,73	-2,27
	Valeur-p	*0,10*	*0,12*	*0,05*
ARFL_7PCs	t	-1,95	-1,70	-2,06
	Valeur-p	*0,08*	*0,12*	*0,07*

Les erreurs de prévision obtenues avec les spécifications énumerées dans la première colonne ont été comparées à celles engendrées par le modèle de référence AR(1) afin de produire les statistiques du test.

E.2 Le modèle autorégressif augmenté de facteurs latents (ARFL) – Base de données trimestrielles

La base de données LCDMA-Q contient des variables trimestrielles. Avant le processus d'estimation des facteurs latents, les séries provinciales ainsi que les séries correspondant à des moyennes trimestrielles de variables dans LCDMA-M ont été exclues. Le panel utilisé contient alors 71 variables, dont 169 observations pour chacune d'entre elles. Enfin, le processus suivi avec la base de données LCDMA-M a également été appliqué au panel final de la base de données trimestrielles. Contrairement aux spécifications obtenues avec LCDMA-M, la base de données trimestrielles permet de réaliser le processus avec le niveau par habitant et avec le taux de croissance de ce dernier.

Lorsque les facteurs latents ont été estimés à partir des niveaux des variables du panel, on s'aperçoit que les résultats obtenus sont très similaires à ceux obtenus avec les facteurs tirés du panel de variables mensuelles. Effectivement, selon le Tableau E.5, la spécification qui a obtenu les plus petites valeurs auprès des critères REQM et EAM a été celle incluant les valeurs contemporaines des sept premières composantes principales. Par contre, contrairement à la spécification exploitant les variables de la base de données LCDMA-M, celle exploitant les facteurs latents des variables trimestrielles a été capable de mieux prédire la dynamique de la variable d'intérêt lors des deux dernières années de la période d'évaluation de l'exercice pseudo hors échantillon. Le quatrième graphique de la première colonne dans la Figure E.3 illustre que cette spécification prévoit une diminution des émissions des GES en 2020, suivie d'un rébond en 2021. Ceci n'a pas été le cas avec la spécification tirée du panel de variables mensuelles, celle-ci a plutôt prévu une autre diminution en 2021. Toutefois, malgré cet élément positif, les résultats obtenus dans le Tableau E.5 indiquent que la spécification utilisant les composantes principales estimées avec LCDMA-Q a moins bien fait que celle estimée avec les variables mensuelles. De plus, la valeur obtenue avec le critère EAM est égale à 1,00. Ceci veut dire que la spécification trimestrielle n'a pas mieux fait que le niveau précision obtenu avec le modèle de référence selon ce critère. Cette remarque est applicable à toutes les autres spécifications considérées dans la figure et les tableaux ci-dessous. En effet, la majorité d'entre elles a obtenu au mieux des résultats aussi bons que le modèle AR(1) lors de l'exercice pseudo hors échantillon selon le Tableau E.5. Par conséquent, aucune des spécifications dans cette sous-section n'a été retenue pour la discussion dans le corps principal du mémoire.

Figure E.3 Prévisions obtenues avec les spécifcations du modèle ARFL – Variables en niveau du panel LCDMA-Q

Tableau E.5 Ratios relatifs au modèle de référence AR(1) – Variables en niveau du panel LCDMA-Q

Modèle	Critère	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	REQM	*0,95*	*1,00*	*0,87*
	EAM	*1,20*	*0,99*	*1,08*
ARFL_3PCs	REQM	*0,96*	*1,02*	*0,90*
	EAM	*1,18*	*0,99*	*0,95*
ARFL_5PCs	REQM	*0,76*	*1,01*	*0,80*
	EAM	*1,07*	*1,01*	*1,02*
ARFL_7PCs	REQM	*0,68*	*1,03*	*0,92*
	EAM	*1,00*	*1,09*	*1,25*

Les valeurs représentent les ratios relatifs au modèle de référence AR(1) qui ont été obtenus avec les équations (15) et (16) du Chapitre 2. Lorsque la valeur du ratio est inférieure à 1,00, cela signifie que le modèle en question a été plus précis que le modèle AR(1) lors de l'exercice pseudo hors échantillon.

Tableau E.6 Résultats du test Diebold-Mariano – Variables en niveau du panel LCDMA-Q

Modèle	Statistique	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	t	0,20	-0,06	0,55
	Valeur-p	*0,85*	*0,96*	*0,59*
ARFL_3PCs	t	0,19	-0,32	0,88
	Valeur-p	*0,85*	*0,75*	*0,40*
ARFL_5PCs	t	0,62	-0,12	0,72
	Valeur-p	*0,55*	*0,90*	*0,49*
ARFL_7PCs	t	0,65	-0,91	0,28
	Valeur-p	*0,54*	*0,39*	*0,79*

Les erreurs de prévision obtenues avec les spécifications énumerées dans la première colonne ont été comparées à celles engendrées par le modèle de référence AR(1) afin de produire les statistiques du test.

Lorsque la démarche a été réalisée avec le taux de croissance des variables de la base de données LCDMA-Q, les résultats ont été moins bons qu'avec le niveau des variables. À l'exception des spécifications utilisant la première composante principale, toutes les autres spécifications ont engendré des ratios relatifs au modèle de référence supérieurs à 1,00 auprès des critères REQM et EAM. Ceci indique clairement que cet ensemble de spécifications est moins bon que le modèle AR(1) pour prédire les émissions des GES au Canada. En fait, au lieu d'augmenter le niveau de précision atteint avec le modèle de référence, l'inclusion des composantes principales a empiré le pouvoir de prévision. Cet ensemble de prévisions a tellement mal fait lors de l'exercice pseudo hors échantillon, que le Tableau E.8 indique que l'hypothèse nulle a pu être rejetée à une couple d'occasions au niveau de confiance de 10 % et même au niveau de 5 % une fois. Par conséquent, aucune des spécifications de cet ensemble n'a été retenue pour la suite de l'analyse.

Figure E.4 Prévisions obtenues avec les spécifications du modèle ARFL – Taux de croissance des variables du panel LCDMA-Q

Tableau E.7 Ratios relatifs au modèle de référence AR(1) – Taux de croissance des variables du panel LCDMA-Q

Modèle	Critère	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	REQM	*0,99*	*0,99*	*0,98*
	EAM	*1,01*	*1,03*	*1,18*
ARFL_3PCs	REQM	*1,05*	*1,02*	*1,07*
	EAM	*1,17*	*1,03*	*1,18*
ARFL_5PCs	REQM	*1,06*	*1,15*	*1,22*
	EAM	*1,11*	*1,19*	*1,14*
ARFL_7PCs	REQM	*1,19*	*1,21*	*1,50*
	EAM	*1,47*	*1,19*	*1,87*

Les valeurs représentent les ratios relatifs au modèle de référence AR(1) qui ont été obtenus avec les équations (15) et (16) du Chapitre 2. Lorsque la valeur du ratio est inférieure à 1,00, cela signifie que le modèle en question a été plus précis que le modèle AR(1) lors de l'exercice pseudo hors échantillon.

Tableau E.8 Résultats du test Diebold-Mariano – Taux de croissance des variables du panel LCDMA-Q

Modèle	Statistique	ARFL(0)	ARFL(1)	ARFL(2)
ARFL_1PC	t	-1,75	-1,47	-1,65
	Valeur-p	*0,11*	*0,17*	*0,13*
ARFL_3PCs	t	-1,99	-1,61	-1,86
	Valeur-p	*0,08*	*0,14*	*0,10*
ARFL_5PCs	t	-1,67	-1,80	-1,31
	Valeur-p	*0,13*	*0,11*	*0,22*
ARFL_7PCs	t	-2,47	-1,73	-1,89
	Valeur-p	*0,04*	*0,12*	*0,09*

Les erreurs de prévision obtenues avec les spécifications énumerées dans la première colonne ont été comparées à celles engendrées par le modèle de référence AR(1) afin de produire les statistiques du test.

E.3 Conclusion de l'Annexe

En conclusion, l'analyse précédente démontre que la majorité des spécifications du modèle ARFL n'a pas engendré de bons résultats dans le cadre de l'exercice pseudo hors échantillon. En fait, seulement une couple de spécifications, dont les facteurs latents ont été estimés à partir des variables en niveau de la base de données LCDMA-M, ont réussi à faire légèrement mieux que le modèle AR(1). De plus, la hausse du niveau de précision de leurs prévisions ne justifie pas toutes les démarches additionnelles qui doivent être réalisées afin d'estimer les facteurs latents. Ceci est en ligne avec l'une des conclusions tirées par les auteurs de l'article de Bennedsen *et al.* (2021). Malgré le fait que l'article propose un modèle structurel à

facteurs dynamiques, « *structural augmented dynamic factor model* » (SADFM) en anglais, sa conclusion indique qu'un niveau de précision très semblable peut être obtenu sans l'utilisation de la composante dynamique de la spécification. Donc, seulement une spécification a été discutée dans le corps du texte afin de montrer que ce type de modèles a été étudié, mais qu'il ne devrait pas être proposé étant donné qu'il fait moins bien que les spécifications du modèle ADL.

BIBLIOGRAPHIE

Bennedsen, M., Hillebrand, E. et Koopman, S. (2021). Modeling, forecasting, and nowcasting U.S. CO_2 emissions using many macroeconomic predictors. *Energy economics*, *96*, Article 105118. https://doi.org/10.1016/j.eneco.2021.105118

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica, 51*(5), 1281-1304. https://doi.org/10.2307/1912275

Diebold, F. X. et Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253 – 263.

Doornik, J. et Hendry, D. (2015). Statistical model selection with "Big Data". *Cogent Economics & Finance*, *3*(1), Article 1045216. https://doi.org/10.1080/23322039.2015.1045216

Environnement et Changement climatique Canada. (2016, 6 janvier). *L'Accord de Paris*. Gouvernement du Canada. https://www.canada.ca/fr/environnement-changement-climatique/services/changements-climatiques/accord-paris.html

Environnement et Changement climatique Canada. (2023a). *Rapport d'inventaire national 1990-2021 : Sources et puits de gaz à effet de serre au Canada (Partie 1)*. Gouvernement du Canada. https://publications.gc.ca/collections/collection_2023/eccc/En81-4-2021-1-fra.pdf

Environnement et Changement climatique Canada. (2023b). R*apport d'inventaire national 1990-2021 : Sources et puits de gaz à effet de serre au Canada (Partie 2)*. Gouvernement du Canada. https://publications.gc.ca/collections/collection_2023/eccc/En81-4-2021-2-fra.pdf

Environnement et Changement climatique Canada. (2023c). *Rapport d'inventaire national 1990-2021 : Sources et puits de gaz à effet de serre au Canada (Partie 3).* Gouvernement du Canada. https://publications.gc.ca/collections/collection_2023/eccc/En81-4-2021-3-fra.pdf

Environnement et ressources naturelles. (2024, 14 février). *Un environnement sain et une économie saine*. Gouvernement du Canada. https://www.canada.ca/fr/services/environnement/meteo/changementsclimatiques/plan-climatique/survol-plan-climatique/environnement-sain-economie-saine.html

Eurostat – Statistics Explained (10 août 2023). *Glossaire: Équivalent dioxyde de carbone.* Eurostat. *https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Carbon_dioxide_equivalent&action=statexp-seat&lang=fr*

Fortin-Gagnon, O., Leroux, M., Stevanovic, D., & Surprenant, S. (2020). A large canadian database for macroeconomic analysis. https://chairemacro.esg.uqam.ca/donnees/base-de-donnees-canadiennes/

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, *Geoscientific Model development*, *15*(14), 5481-5487. https://doi.org/10.5194/gmd-15-5481-2022, 2022.

Jensen, S. (2021). *Use of machine learning in climate econometrics* [Thèse de doctorat, Aarhus University]. Creates. https://pure.au.dk/ws/portalfiles/portal/225863920/PhD_dissertation_Sebastian_Mathias_Jensen.pdf

McCracken, M. W. et Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economics Statistics*, *34*(4), 574-589. https://doi.org/10.1080/07350015.2015.1086655

Pindyck, R. S. (2017). The use and misuse of models for climate policy. *Review of Environmental Economics and Policy, 11*(1), 100-114 https://www.journals.uchicago.edu/doi/10.1093/reep/rew012

Schmalensee, R., Stoker, T. M., & Judson, R. A. (1998). World carbon dioxide emissions: 1950 – 2050. *Review of Economics and Statistics, 80*(1), 15-27. https://doi.org/10.1162/003465398557294

Shumway, R. H. et Stoffer, D. S. (2017). Time series analysis and applications using the R statistical package. *Free Dog Publishing*. https://swmprats.net/stoffer/tsa4/tsaEZ.pdf

Stock, J. H. et Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association, 97*(460), 1167-1179.

Stock, J. H. et Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics, 20*(2), 147-162. https://doi.org/10.1198/073500102317351921

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Zou, H. et Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 67*(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x