

PROJET STATISTIQUE DE 2^{ème} ANNEE Année scolaire 2021-2022

SUJET N°14

Modélisation de risque de transition sur les TPE/PME

Réalisé par :

Amadou Tidjani SIRADJI MANZO
Pape Tidiane GUEYE
Tahar AMADOU HAMADOU

Tuteurs :

Mme. Céline MERCADO
M. Cyril REGNIER

Coach :

Mme. Dominique LEGOUBEY

Avril 2022

REMERCIEMENTS

Nous tenons tout d'abord à remercier notre tutrice, Madame Céline Mercado pour l'encadrement et l'attention portée à cette étude. Nous remercions également Monsieur Cyril Regnier, notre co-tuteur pour ses précieux conseils et orientations dans la réalisation de ce travail.

Nos remerciements vont aussi à l'endroit de Madame Dominique Legoubey pour ses nombreux conseils dans l'organisation du travail et la communication entre les membres du groupe. Nous remercions nos enseignants pour leur conseil sur les aspects techniques.

Nous adressons enfin nos remerciements à toutes les personnes qui, de près ou de loin, ont apporté leur soutien à ce projet.

ABSTRACT

In the face of the climate emergency, the need to think about reducing the human contribution to environmental degradation has become a major issue that decision makers must address. Rising seas levels, increasing temperatures and seasonal irregularities are among the reasons why radical decisions are made to raise awareness and act on this emergency. These policies are not without consequences on economic actors, especially financial institutions that must justify to what extent their investments take into account environmental aspect. These consequences, qualified as transition risks, are the object of much interest of the companies which must estimate it in order to better manage it.

Our study, focused on Very Small Companies, is part of this framework with an uncommon approach, an evaluation method using little information, the TOP-DOWN. This is an approach that estimates the GHG emissions of companies using available information from their respective sectors. Mostly used in a simplified way as did ADEME [2], the Top-Down developed in this paper uses statistical model with a ratio combination as well as a machine learning model sector level data. In order to grasp the error of the Top-Down approach, classical models (linear regression, Elastic Net, decision tree regression, random forest and XGboost) using companies information were also developed as reference. The data used for this work comes from various sources. However, the reference sample is the Ademe database on the GHG emissions balance of companies.

Through this study several results have been found. First, after segmentation of companies according to the INSEE definition, we found that there were only 16.5% of very small companies in our sample. This shows how difficult it is for very small companies to make their GHG evaluation. Second, the results showed that the Top-Down model with ratios was proved to be better than the Top-Down approach at sector level. A comparison of the classical models with the TOP-DOWN approaches developed here, according to the prediction error, allowed us to show that, statistically, our approach is as efficient as the classical approaches. These results, which are very satisfactory in view of the very small amount of information used in the TOP/DOWN approach, make TOP/DOWN a very innovative approach, especially for small companies that do not often have huge amounts of information. It is even more innovative for banks in their strategy to decarbonize their portfolio.

Key words : Transition risks, GHG emissions, Top-Down model, climate risk, Machine learning.

SOMMAIRE

| | |
|---|-------|
| Remerciements | I |
| Abstract | II |
| Sommaire | III |
| Sigles et abréviations | IV |
| Liste des graphiques | IV |
| Liste des tableaux | VI |
| Introduction | 1 |
| 1 Cadre conceptuel et revue de la littérature | 3 |
| 2 Cadre méthodologique | 9 |
| 3 Evaluation du risque de transition | 17 |
| Conclusion | 29 |
| Bibliographie | i |
| A Annexe | ii |
| Tables de matières | xxiii |

SIGLES ET ABRÉVIATIONS

| | | |
|---------------|---|--|
| ACPR | : | Autorité de Contrôle Prudentiel et de Résolution |
| ACP | : | Analyse en Composantes Principales |
| ACT | : | Assessing Low-Carbon Transition |
| ADEME | : | Agence de l'environnement et de la maîtrise de l'énergie |
| ASD | : | Fusion des bases Ademe, Siren et Données entreprises |
| ASDES | : | Fusion des bases Ademe, Siren, Données entreprises, Esane et Secten |
| BIS | : | Bank of International Settlements |
| BCE | : | Banque Centrale Européenne |
| Citepa | : | Centre interprofessionnel technique d'études de la pollution atmosphérique |
| COP | : | Conference Of Parties |
| DE | : | Données Entreprises |
| ESANE | : | Elaboration des statistiques annuelles des entreprises |
| GE | : | Grande Entreprise |
| GES | : | Gaz à Effet de Serre |
| INSEE | : | Institut National de la Statistique et des Etudes Economiques |
| LTECV | : | Loi sur la Transition Ecologique pour le Croissance Verte |
| ODD | : | Objectifs de Développement Durable |
| PME | : | Petite et Moyenne Entreprise |
| SIREN | : | Système d'Identification du Répertoire des Entreprises) |
| TCFD | : | Task Force on Climate-related Financial Disclosures |
| TPE | : | Très Petite Entreprise |

TABLE DES FIGURES

| | | |
|------|--|------|
| 2.1 | Création des bases de données et processing | 11 |
| 3.2 | Répartition des entreprises par catégorie | 18 |
| 3.3 | Distribution des émissions | 18 |
| 3.4 | Distribution des émissions (en tonnes de CO2 équivalent | 19 |
| 3.5 | Moyenne des émissions GES par département (en tonnes) | 19 |
| 3.6 | Analyse de correspondances des variables issues filtres Forward | 20 |
| 3.7 | Évolution des émissions de GES par secteur (en Mt de CO2e) | 21 |
| 3.8 | Transformations de la variable cible | 22 |
| 3.9 | Erreurs quadratiques moyennes calculée sur l'échantillon test sur les différents modèles | 23 |
| 3.10 | Comparaisons du niveau de surapprentissage entre la Régression Linéaire et le Ran- dom Forest | 23 |
| 3.11 | Arbre de régression | 27 |
| A.12 | Démarche de travail | iv |
| A.13 | Illustration de l'algorithme Forward | v |
| A.14 | Illustration d'un arbre de régression | v |
| A.15 | Illustration de l'algorithme Random Forest | vi |
| A.16 | Comparaison des GE et TPE | vii |
| A.17 | Histogramme des valeurs propres | xii |
| A.18 | Positionnement des TPE/PME et GE par filière en termes d'émission de GES . . . | xiii |
| A.19 | Emission par scope et par filière | xiii |
| A.20 | Répartition du chiffre d'affaire | xiv |
| A.21 | Nuage de point entre variables cibles et les variables quantitatives les plus corrélées | xv |
| A.22 | Distribution de la variable cible dans les différentes classes des variables qualitatives | xvii |

| | |
|---|-------|
| A.23 Résidus Studentisé vs Index | xviii |
| A.24 Résidu vs Quantile loi normale centrée réduite | xviii |
| A.25 Résidus Studentisé vs Valeurs ajustées | xix |
| A.26 Régression Linéaire : Qualité d'ajustement et courbe d'apprentissage | xix |
| A.27 Elastic Net : Qualité d'ajustement et courbe d'apprentissage | xx |
| A.28 Arbre de régression : Qualité d'ajustement et courbe d'apprentissage | xx |
| A.29 Random Forest : Qualité d'ajustement et courbe d'apprentissage | xxi |
| A.30 XG Boost : Qualité d'ajustement et courbe d'apprentissage | xxi |
| A.31 Matrice de corrélation- variables filière | xxii |
| A.32 Régression rigde | xxii |
| A.33 Arbre de régression | xxii |
| A.34 Random Forest pour la régression | xxii |
| A.35 Courbes d'apprentissage modèles ratio 2 | xxiii |
| A.36 Courbes d'apprentissage modèles ratio 2 | xxiii |

LISTE DES TABLEAUX

| | | |
|-----|---|-------|
| 3.1 | Description des transformations | 22 |
| 3.2 | Résultat du modèle de régression linéaire | 25 |
| 3.3 | Résultat des modèles ratio | 26 |
| 3.4 | Résultat des modèles filière | 26 |
| 3.5 | Erreurs quadratiques moyennes des modèles | 28 |
| 3.6 | Test de différences de moyennes | 28 |
| A.1 | Rapprochement des secteurs Citepa aux filières APE | iii |
| A.2 | Résultat du test de différence de moyennes | viii |
| A.3 | Liste des variables significativement corrélées aux émissions | ix |
| A.4 | Test de liaison entre les explicatives qualitatives et la variable dépendante | x |
| A.5 | Variables issues de la sélection Forward selon les critères R^2 ajusté puis l'AIC | xi |
| A.6 | Une dernière filtre des variables pour la régression linéaire | xi |
| A.7 | Description des émissions par catégorie | xii |
| A.8 | Variables explicatives sélectionnées pour les modèles ratio | xxiii |
| A.9 | Matrice de corrélation- variables ratios | xxiii |

INTRODUCTION

Responsable des changements d'événements météorologiques extrêmes (sécheresses, tempêtes, inondations) et de crises alimentaires à travers le monde, le changement climatique est au centre des préoccupations du siècle. Selon le rapport de 2019 de l'ONU sur les ODD (Objectifs de Développement Durable), la concentration atmosphérique en CO₂ a augmenté de 146% en 2017 relativement aux niveaux préindustriels. L'année suivante, la température moyenne mondiale s'est élevée d'1°C et continue son cours haussier avec l'année 2019 étant la deuxième année la plus chaude jamais enregistrée. C'est ainsi que la communauté internationale, à travers les Conférences des Parties (COP), a nourri une volonté accrue d'endiguer ce phénomène et d'amoindrir ces impacts socio-économiques, notamment avec l'accord de Paris (2015), le fond vert pour le climat, la taxe carbone, le principe du pollueur payeur, et entre autres décisions.

Aujourd'hui, dans le cadre de l'atteinte des objectifs de l'accord de Paris, les institutions financières sont reconnues comme étant des leviers de taille. L'intégration de la problématique climat dans les stratégies microprudentielles des institutions financières est plus que jamais un enjeu considérable, notamment dans le cadre du risque de transition. Non seulement, une politique de transition abrupte pourrait fragiliser l'activité de certaines entreprises et, par ricochet, le système financier, mais aussi l'ignorance de ce risque conduit à un risque juridique conformément à la "loi sur la transition énergétique et pour la croissance verte"(LTECV), plus précisément son article 173 qui donne un cadre sur les informations extra-financières que doivent fournir ces institutions. Pour ces dernières, la prise en compte des risques financiers liés au climat consistera en une évaluation, un suivi et une atténuation des risques physiques, de transitions et de réputation (ACPR¹ – Les groupes bancaires français face au risque climatique, 2019). Ceci recentre le problème sur la capacité des institutions financières à démontrer en quoi leur choix d'investissement tient compte de l'aspect environnemental, ce qui se résume à une connaissance du niveau de pollution des entreprises présentes dans leurs portefeuilles. Si ces informations sont plus faciles à obtenir pour les grandes entreprises, il en est moins pour les TPE/PME.

Malgré de nombreux travaux de recherche sur les risques financiers liés au climat, la question sur les petites et moyennes entreprises reste en suspension, notamment à cause de la difficulté du traçage de leurs empreintes carbone. En effet, cette catégorie d'entreprises n'est pas soumise aux mêmes obligations que les grandes entreprises et les entreprises de taille intermédiaire et ne dispose pas des outils suffisants pour mesurer les émissions carbone de leur activité et travailler sur leurs stratégies de décarbonisation.

Dès lors, prédire ou modéliser les émissions de gaz à effet de serre de ces entreprises serait un input important non seulement pour les institutions financières dans la mesure où elles peuvent intégrer l'impact de ces risques dans leur politique interne de gestion de risque mais aussi pour les autorités de surveillance qui veillent sur le changement climatique dans le pilotage d'une politique bas-carbone. Ainsi, cette étude cherchera à répondre à la question : **«comment identifier et mesurer le risque**

1. Autorité de Contrôle Prudentiel et de Résolution, créé le 21 janvier 2010, a pour objectif le contrôle des banques et assurances françaises.

de transition sur les TPE/PME dans un contexte d'insuffisance des données ?».

A travers cette étude, l'objectif général est d'évaluer le risque de transition sur les TPE/PME. De manière spécifique, il s'agit de :

- développer une méthode pour évaluer le risque de transition au niveau filière puis l'extrapoler au niveau entreprise ;
- évaluer le risque de transition directement avec les données disponibles au niveau entreprise ;
- comparer les deux approches afin d'appréhender l'ampleur des erreurs de la méthode développée.

Pour mener à bien cette étude, nous avons structuré le travail en trois grands chapitres. Le premier chapitre porte sur le cadre conceptuel et la revue de la littérature sur l'évaluation de risque de transition. Ensuite, le deuxième chapitre aborde la présentation des données et de la méthodologie utilisée pour répondre à la problématique de l'étude. Enfin, le dernier chapitre présente les résultats de l'étude. Dans ce chapitre, une première partie fait ressortir l'analyse descriptive (analyse uni et multi-variée) le profil des entreprises présentes dans l'étude, puis les résultats de la modélisation, leurs discussions et la conclusion.

CHAPITRE 1

CADRE CONCEPTUEL ET REVUE DE LA LITTÉRATURE

Ce chapitre est composé de deux sections. La première section aborde la définition des concepts en apportant un éclairage sur les différentes terminologies liées à la thématique. Dans la deuxième section, nous présentons différentes méthodes développées dans le cadre de l'évaluation du risque de transition.

1.1 Définitions des concepts

Il s'agit dans cette section de définir les concepts clés liés à ce travail de recherche.

1.1.1 TPE/PME

La segmentation des entreprises utilisée dans cette étude suit la définition officielle donnée par l'INSEE. Ainsi, selon cette définition, une Très Petite Entreprise (TPE) ou microentreprise est une entreprise dont l'effectif est inférieur à 10 personnes et dont le chiffre d'affaires ou le total du bilan annuel n'excède pas 2 millions d'euros. Une Petite et Moyenne Entreprise est quant à elle une entreprise dont l'effectif est inférieur à 250 personnes et dont le chiffre d'affaires annuel n'excède pas 50 millions d'euros ou dont le total de bilan n'excède pas 43 millions d'euros.

1.1.2 Grande Entreprise (GE)

Toujours selon la définition de l'INSEE, sont classées GE les entreprises qui occupent plus de 5 000 employés et ayant un chiffre d'affaire excédant 1 500 millions d'euros ou un total bilan de plus de 2 000 millions.

1.1.3 Risque climatique (BIS, 2021)[3]

Les risques climatiques regroupent l'ensemble des risques liés au climat. Cela regroupe les aléas météo-marins comme les tempêtes et les cyclones, ainsi que les inondations ou les glissements de terrain

lorsqu'ils sont liées aux précipitations, mais aussi les épisodes de sécheresse (mais pas les tsunamis qui sont d'origine tectonique).

Les risques climatiques sont une constante de l'histoire de l'humanité et ont donné lieu à des formes d'adaptation extrêmement variées. Toutefois, aujourd'hui la notion de « risque climatique » se définit plutôt comme l'exposition d'un agent économique aux impacts climatiques et aux répercussions de ceux-ci sur son activité.

En septembre 2015, dans un discours qui fait date, Marc Carney, le gouverneur de la banque d'Angleterre alerte les investisseurs sur une classe de risque d'un nouveau genre : le risque climatique qui est un risque systémique pour l'ensemble du secteur financier.

Il décrit ainsi 3 types de risques climatiques qui s'ajoutent aux risques financiers traditionnels : il s'agit des risques physique, de transition et juridique.

Risque physique

Il s'agit de facteurs de risques dit aigüe dans le sens où leur survenue peut être brusque et inattendue, comme notamment les inondations qui deviennent très fréquentes dans certaines régions du monde, ou encore des facteurs de risque dit chroniques qui, quant à eux se manifestent par une dégradation dans le temps de conditions climatiques dont le plus d'actualité est l'augmentation de la température.

Risque de transition

Le risque de transition correspond aux conséquences économiques entraînées par la mise en place d'un modèle économique bas-carbone. Parmi ces facteurs de risque, il y a les politiques gouvernementales qui sont de nature à contraindre les acteurs économiques dans leurs activités, ce qui peut se traduire par des taxes, des amendes ou encore des subventions. Les facteurs de risque de transition sont aussi le changement d'habitude que pourraient effectuer les investisseurs dans leur choix d'investissement, les épargnants, et même les consommateurs dans leurs choix quotidiens.

Risque juridique

Le risque juridique correspond aux compensations qu'une personne morale pourrait être conduite à payer au cas où elle serait jugée responsable de dommages causés par les conséquences du changement climatique.

1.1.4 Gaz à effet de serre (GES)

La BIS[3] définit les gaz à effet de serre (GES) comme étant les constituants gazeux de l'atmosphère, tant naturels qu'anthropiques, qui absorbent et émettent des rayonnements à des longueurs d'onde spécifiques dans le spectre du rayonnement infrarouge thermique émis par la surface de la terre, par l'atmosphère elle-même et par les nuages. Cette propriété est à l'origine de l'effet de serre. La vapeur d'eau (H_2O), le dioxyde de carbone (CO_2), l'oxyde nitreux (NO_2), le méthane (CH_4) et l'ozone (O_3) sont les principaux gaz à effet de serre présents dans l'atmosphère terrestre.

1.1.5 Vision par scope des émissions de GES : (Ademe, 2014)[2]

La vision par scope des émissions fait ressortir 3 scopes : les scopes 1, 2 et 3. Le scope 1 est composé des émissions de GES qui ont eu lieu directement au niveau de l'entreprise (procédés de production,

combustion de ressources fossiles, etc.). Le scope 2 comprend les émissions indirectes liées à l'énergie. Il s'agit principalement des émissions liées à l'électricité qui sont émises au moment de sa production (e.g. la combustion d'une centrale à gaz). Les émissions du scope 2 sont estimées en attribuant les émissions du secteur énergie à chaque agent économique en fonction de ses achats auprès de ce secteur. Enfin, le scope 3 regroupe toutes autres les émissions indirectes. Ces émissions sont comptabilisées en prenant en compte l'utilisation de produits fabriqués, les déchets d'activité, le transport des marchandises, etc. L'inconvénient du scope 3 est qu'elle génère de nombreux double compte entre entreprises, ce qui rend difficile son intégration dans les études.

1.2 Revue sur l'évaluation de risque de transition

Aujourd'hui, la recherche de méthodologies efficaces pour l'évaluation et le suivi des risques financiers liés au climat est à son paroxysme. Dans le cadre du risque de transition la plupart des méthodes d'évaluations, telles que présentées ci-dessous, partent des émissions de GES des unités concernées. Elles mettent en relation les chocs sur les contreparties des banques et les risques financiers classiques. Par exemple une transition désordonnée et imprévue conduirait à un risque de crédit émanant des pertes probables enregistrés par les emprunteurs. De même, elle pourrait être vecteur de problèmes de liquidité résultant des changements dans les habitudes d'épargne et d'investissement.

L'état de l'existant indique, d'une part, l'utilisation de différents indicateurs permettant la mesure, la cartographie et le suivi des expositions des banques. D'autre part, des méthodes plus sophistiquées sont mises en place pour quantifier l'effet d'une transition sur un portefeuille bancaire. En parallèle, ces méthodes peuvent être intégrées dans une approche "Bottom up" ou "Top down" en fonction de la granularité des données.

1.2.1 Principaux indicateurs de mesure d'exposition

Selon le comité de Bâle sur la supervision bancaire (Bank for International Settlements (BIS), 2021) et l'Autorité de Contrôle Prudentiel et de Résolution (ACPR, 2019), la première étape de l'évaluation du risque de transition est une identification des secteurs les plus vulnérables et les mécanismes de transmission. Dans cette optique, les indicateurs utilisés permettent de cartographier, mesurer et suivre les expositions. Les indicateurs fréquemment utilisés par les banques, selon le comité de Bâle, la TCFD¹ et la Banque Centrale Européenne, sont présentés ci-dessous.

a. Intensité carbone des contreparties

Elle se définit comme étant le volume d'émission de CO2 par unité de chiffre d'affaires des contreparties présentes dans un portefeuille. La TCFD, dans son rapport de 2017[15], propose la méthodologie suivante pour le calcul de cet indicateur :

$$Intensite\ carbone_i = \sum \frac{EncoursCredit_i}{EncoursCredit_{total\ Portefeuille}} * \frac{EmissionsGES_i}{TotalChiffreAffaire_i}.$$

EmissionGES_i : emission scope 1 et 2 de GES de la contrepartie

1. La TCFD (Task Force on Climate-related Financial Disclosures) est un groupe de travail international créé en décembre 2015 dont l'objectif est de fournir des recommandations sur la manière de reporter les risques et opportunités liés aux changements climatiques

Cette mesure est exigé par la Banque Centrale Européenne dans le cadre du stress test (Banque Centrale Européenne, 2022 [10]). Cette dernière propose également la prise en compte du scope 3 dans le calcul de l'indicateur. La construction d'un tel indicateur permet d'identifier les portefeuilles les plus vulnérables et de construire des cartes thermiques pour le suivi des actifs intensifs en carbone (Bank for International Settlements (BIS), 2021 [4]).

b. Actifs liés au carbone

Il s'agit du niveau des investissements bancaire placé dans les secteurs les plus sensibles au carbone. Afin d'éviter toute ambiguïté, la TCFD définit les actifs liés au carbone comme étant ceux liés au secteur de l'énergie et des services publics. Le rapport du comité de Bâle [3] souligne que dans le cadre de l'évaluation du risque de transition, certaines banques mesurent les actifs liés au carbone qu'elles détiennent comme proxy du risque de transition. L'utilisation de cet indicateur a été recommandée par la TCFD qui demande aux banques de divulguer les concentrations importantes d'actifs liés au carbone. Dans son rapport final [15], elle fournit également une méthodologie de mesure de cet indicateur qui est défini au niveau portefeuille comme étant la part des actifs liés aux secteurs forts en pollution sur la valeur totale du portefeuille :

$$\text{Actifs liés au carbone du portefeuille} = \sum \frac{\text{Investissement total actifs liés au carbone}}{\text{Valeur total du portefeuille}} * 100.$$

Dans le cadre de l'exercice du test de résilience [10] cet indicateur est utilisé pour comprendre la sensibilité du business model de la banque face aux secteurs intensifs en carbone. Il est souvent accompagné d'un autre qui va dans le même sens général mais a l'avantage d'être plus granulaire. Il s'agit des émissions financées.

c. Emissions financées

Les émissions financées donnent une mesure de l'empreinte carbone des actifs bancaires. En d'autres termes, il s'agit de la contribution des banques à la pollution des contreparties. Elle se calcule par une moyenne des ratios émissions/total _capitaux pondérée par le poids de la contrepartie dans le portefeuille de la banque (TCFD, 2017 [15]) :

$$\text{Emissions financées d'une portefeuille} = \frac{1}{\text{ValeurTotalPortefeuille}} \sum \frac{\text{EncoursCredit}_i}{\text{TotalAction}_i} * \text{EmissionGES}_i$$

Contrairement à l'identification des actifs liés au carbone, cet indicateur nécessite la connaissance, au niveau contrepartie, des niveaux de pollution et d'engagement. Dans le cadre du test de résilience, il est demandé aux banques de fournir des mesures de cet indicateur pour les scopes 1 et 2, d'une part, et pour le total des émissions, d'autre part [10].

d. Scoring par méthodologie ACT (OID, 2020) [13]

L'ACT (Assessign Low-Carbon Transition) est une méthodologie d'évaluation de la maturité d'une entreprise dans sa transition vers une économie bas-carbone (ADEME, 2015). Cette méthodologie est

conçue pour les entreprises afin d'évaluer elles-mêmes leur adaptation à la transition. L'ACT va permettre de construire plusieurs indicateurs pour une entreprise et un poids est attribué à chaque indicateur en fonction des caractéristiques du secteur auquel l'entreprise appartient. Ainsi, une combinaison de ces indicateurs peut permettre d'évaluer le risque. Le calcul des derniers se basent sur les réponses aux différentes questions méthodologiques liées au changement climatique auxquelles les entreprises sont censées pouvoir répondre.

1.2.2 Méthodes de quantification du risque de transition

Après la cartographie de l'exposition, des méthodologies sont développées pour quantifier les risques associées à une probable transition. Ces méthodes d'évaluation utilisées par les banques et les autorités de surveillance sont l'analyse de scénarios, le stress test et l'analyse de la sensibilité (Bank for International Settlements (BIS), 2021 [4]).

a. Analyse de scénarios

Les analyses de scénarios évaluent l'impact d'un ensemble de facteurs de risques sur le portefeuille financier. Selon le BIS (2021), ces analyses sont réalisées en quatre (4) étapes :

- 1 identifier les scénarios de risques ;
- 2 relier les impacts des scénarios aux risques financiers ;
- 3 évaluer les sensibilités des contreparties et/ou des secteurs aux risques ;
- 4 extrapoler les impacts de ces sensibilités pour calculer une mesure globale de l'exposition et des pertes potentielles.

Les méthodes narratives permettent de développer ces scénarios climatiques (ADEME, 2019). Ces méthodes vont permettre de mettre en exergue la façon dont les scénarios sont appréhendés. Il s'agit de quatre (4) approches narratives : approche prospective (ou normative), approche énumérative, approche historique et approche hypothétique. Ces approches diffèrent selon la connaissance en amont du scénario et le déterminant du narratif.

b. Stress test

Le stress test est un sous-ensemble spécifique de l'analyse de scénarios, généralement utilisé pour évaluer la résilience à court terme d'une institution financière face aux chocs économiques, souvent par le biais d'un objectif d'adéquation des fonds propres. En général, lorsqu'on examine la solvabilité, il existe deux types de tests de résistance : les tests macroprudentiels, qui mesurent la façon dont les chocs financiers affectent un système financier et peuvent déclencher un risque systémique, et les tests microprudentiels, qui évaluent la solvabilité d'une institution financière individuelle compte tenu des risques de son portefeuille.

c. Analyse de la sensibilité

L'analyse de la sensibilité est un sous-ensemble spécifique de l'analyse de scénarios. C'est une approche explicative de mesure de risque qui est souvent utilisée pour évaluer l'impact d'une variable climatique sur les agrégats économiques et financiers à travers des modèles économétriques (BIS, 2021). Comme le stress test ou l'analyse de scénarios, l'analyse de sensibilité est une méthodologie utilisée par les banques et les autorités de surveillance pour mesurer le risque. Le **Three-ME** (OFCE-ADEME, 2008) est un exemple d'analyse de sensibilité qui évalue l'impact d'une politique liée au climat au niveau national.

1.2.3 Approches Bottom-up et Top-Down

Depuis l'intéressement des entreprises à la gestion des risques auxquelles elles s'exposent, deux approches opposées d'appréhender ces risques ont vu le jour : les approches TOP-Down et Bottom-Up. L'approche Top-Down, utilisée le plus souvent par les organismes de régulation financière et les banques centrales, est une approche qui consiste à mesurer le risque au niveau d'un ensemble consolidé contenant les individus cibles de l'étude puis d'en descendre en affectant à ces individus une part du risque global. Cette approche, même si trop simpliste est efficace pour des études contenant un grand échantillon d'individus dont le traitement individuel nécessiterait la mise en oeuvre d'énormes moyens. Plus concrètement, elle est utilisée par l'ADEME pour la quantification des émissions financées². De plus, la BCE [10] suggère l'utilisation du Top-Down pour estimer le scope 3, via des ratios basés sur la consommation intermédiaire.

Elle pourrait également être utilisée dans un scénario de faillite d'un fournisseur de matières premières d'une industrie donnée pour en déduire le risque de faillite des entreprises constituant le secteur en ajoutant aux risques individuels de faillite de ces entreprises un risque de faillite correspond au risque du fournisseur pondéré de leur dépendance vis-à-vis de ce fournisseur.

La mise en oeuvre de cette approche diffère selon le problème posé, mais se traduit le plus souvent sous la forme de ratio.

Au contraire, une approche Bottom-Up évalue le risque au niveau des composants. Ces mesures individuelles du risque sont agrégées pour fournir une vue intégrée du risque. Par exemple, l'approche Bottom-Up de cartographie des expositions peut être utilisée pour évaluer l'exposition relative au risque au niveau de l'actif (par exemple, un prêt) ou de la contrepartie (ou groupes d'actifs ou de contreparties présentant des caractéristiques substantiellement similaires) en additionnant les expositions afin de fournir une vue consolidée des expositions du portefeuille (ou secteur bancaire). Cette approche est le plus souvent utilisée en interne au sein des entreprises pour l'évaluation des risques, en partant des risques individuels liés à leurs différents activités ou services.

Les estimations de l'exposition et du risque font un choix entre les approches Top down et Bottom Up. Notre étude portera une attention particulière à l'approche Top-Down qui semble être la mieux adaptée compte tenu des informations dont nous disposons.

Conclusion partielle

En résumé, ces méthodologies visent à quantifier les impacts financiers potentiels auxquels les banques ou le système financier peuvent être confrontés en comparant un scénario de base à des scénarios reflétant divers degrés de risque découlant du changement climatique. L'enjeu majeur pour les banques est d'intégrer le risque climatique comme paramètre d'évaluation de risque financier. Jusqu'ici la plupart des méthodologies d'évaluation de risque de transition sont utilisées en milieu bancaire et intégrées dans l'évaluation de risque bancaire. L'objectif de notre étude étant la modélisation du risque de transition sur les TPE/PME, elle est une première dans ce sens. Néanmoins, ces approches utilisées par les institutions financières en vue d'évaluer le risque climatique seront contextualisées aux TPE/PME pour répondre à la problématique de notre étude.

2. Une étude de cas d'une méthodologie « top-down » de quantification des émissions financées, ADEME 2014

CHAPITRE 2

CADRE MÉTHODOLOGIQUE

Ce chapitre expose la méthodologie qui sera adoptée pour répondre à notre problématique. Il présente d'abord les données utilisées, ensuite la procédure de construction de l'échantillon de travail, la construction de la variable cible, la procédure de sélection des variables pour la modélisation et enfin les méthodes d'apprentissage supervisé ainsi que l'approche "Top-down".

2.1 Présentation des données et pré-traitement

2.1.1 Sources de données

Diverses sources de données ont été nécessaires à cette étude. Il s'agit des données sur les émissions de GES, les caractéristiques financières des TPE/PME et les filières auxquelles elles appartiennent. Nous disposons principalement de cinq sources de données : bilan GES de l'ADEME, base SIREN, Données caractéristiques des entreprises, base Citepa et données ESANE.

Bilan des émissions de gaz à effet de serre au niveau entreprise

L'échantillon de base pour cette étude est constitué de 1860 entreprises, collectivités et associations qui ont présenté un bilan de leurs émissions de gaz à effet de serre (ADEME, 2022). Les informations contenues dans cet échantillon sont entre autres les émissions par scope de l'entreprise, le nombre d'employés, l'année à laquelle le bilan a été fait, les numéros SIREN et SIRET, etc.

Inventaire des émissions de GES des secteurs en France

Les données sur les émissions de GES concernent les secteurs (filières) auxquels les TPE/PME appartiennent. Ces données proviennent des travaux d'inventaires nationaux de 2021 réalisés par Citepa¹. Cette édition de 2021 présente les émissions depuis 1990 jusqu'à 2019, ainsi qu'une estimation provisoire pour l'année 2020. Elles fournissent également les émissions de plus de **31 substances et de gaz à effet de serre** par secteur d'émission. Les principaux secteurs concernés sont au nombre de huit (8) :

1. Citepa (Centre interprofessionnel technique d'études de la pollution atmosphérique) est un centre technique interprofessionnel avec un statut d'association loi de 1901.

l'industrie de l'énergie, l'industrie manufacturière et construction, les déchets (centralisés), le secteur Résidentiel-Tertiaire, l'agriculture, le transport, l'UTCATF (Utilisation des Terres, Changement d'Affectation des Terres et Foresterie), et enfin le secteur naturel. Les GES pris en compte sont les GES directs qui entrent dans le "panier" de Kyoto² : CO₂, CH₄, N₂O, la famille des HFC, la famille des PFC, le SF₆ et le NF₃. Ces différents types de GES sont donnés en CO₂ équivalent et mesurés en kilotonne (Kt) ou mégatonne (Mt).

Données Ésane

Les données sur l'élaboration des statistiques annuelles d'entreprises (Ésane) utilisées concernent les caractéristiques comptables, financières et d'emploi des entreprises au niveau branche d'activités³ selon le code APE. Ces données sont publiées annuellement par l'INSEE et concernent toutes les entreprises françaises. Elles sont disponibles pour les années 2009 à 2019 au moment de cette étude, et portent essentiellement sur le chiffre d'affaires, le nombre de salariés en temps plein, l'investissement, la capacité de financement des filières.

Données SIREN

Les données SIREN ont été collectées sur le site de l'INSEE à partir des numéros disponibles de SIREN dans l'échantillon de base (bilan GES de l'ADEME). Elles fournissent des caractéristiques géographiques (code postal), le code APE, la catégorie d'entreprise, etc. Cette base de données permettra la fusion des informations niveau entreprise et filière.

Données Entreprises

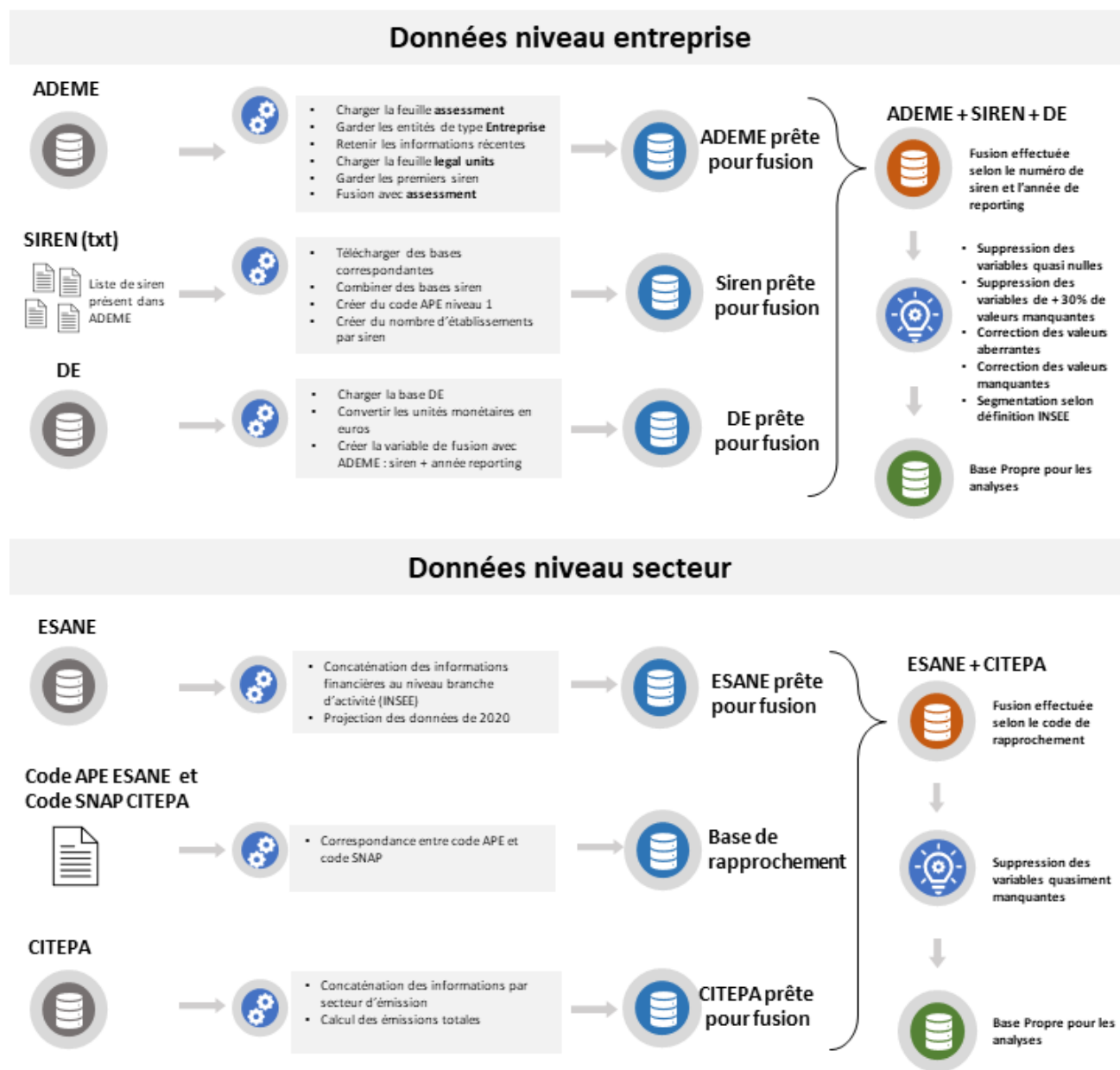
Les données SIREN sont complétées par d'autres données portant sur les caractéristiques financières des entreprises que l'on appellera DE (Données Entreprises). Les DE contiennent des informations annuelles sur les entreprises identifiées par numéro SIREN.

2.1.2 Différents pré-traitements

Afin de rendre les bases utilisables pour les analyses, différents traitements ont été nécessaires. La connaissance de ces derniers rendra possible la reproductibilité du travail. Ces pré-traitements sont les suivants.

-
2. "panier" de Kyoto : liste des polluants considérés comme GES depuis le protocole de Kyoto en 1997 lors de la COP3.
 3. Branches d'activités : elles sont données en annexe A.1.

GRAPHIQUE 2.1 – Création des bases de données et processing



2.2 Choix d'échantillon de travail : TPE/PME sans ou avec les GE

Dans le but de trouver le meilleur modèle permettant de prédire les émissions de GES des TPE/PME, une question de taille fait surface : faut-il intégrer les GE dans l'échantillon ? Ces dernières ayant un niveau d'activité fortement plus élevé que les TPE/PME, nous allons essayer de vérifier l'homogénéité des deux groupes en s'appuyant sur des ratios. Cette vérification se fera par des tests de différence de moyennes entre ces deux groupes. L'usage des ratios fait du sens dans la mesure où ils permettent de neutraliser l'effet taille et sont souvent utilisés dans la mesure de la santé économique d'une entreprise (ratio de liquidité, ratio de solvabilité, etc.). Au cas où les résultats suggèrent une homogénéité, nous utiliserons tout l'échantillon avec les ratios comme variables explicatives. Les résultats de ces tests sont en annexe A.2 et A.16.

2.3 Construction de la variable cible : les émissions de GES des scopes 1 et 2, un proxy du risque de transition

Dans la détermination du niveau d'exposition d'une entreprise au risque de transition, nous avons vu que l'ensemble des méthodes présentées dans la littérature partent des émissions de GES des entreprises (avec, par exemple, le calcul de l'intensité carbone qui n'intègre que les scopes 1 et 2). D'abord, ceci motive l'utilisation des scopes 1 et 2 comme indicateur du niveau d'exposition au risque de transition. Ensuite, ces scopes couvrent les émissions venant de l'activité de l'entreprise (par le biais du scope 1) et de sa consommation en énergie (scope 2). Ces deux instances de pollution constituent principalement la source de pollution des entreprises présentent dans notre échantillon. Enfin, les émissions du scope 3 sont sujettes à plusieurs doubles comptages rendant difficile leur intégration dans l'analyse.

Cette variable cible ainsi construite sera modélisée par deux approches : une modélisation classique au niveau entreprise ; et une modélisation Top Down partant des informations au niveau secteur.

2.4 Méthodologie de modélisation

2.4.1 Modèles d'apprentissage supervisé

Rappelons que l'objectif de l'étude est de prédire les émissions de GES comme indicateur du risque de transition. La variable cible à prédire étant quantitative, nous proposons de tester les modèles suivants : **Régression Linéaire**, **Elastic Net**, **l'Arbre de Régression**, **le Random Forest** et **le XG Boost**.

Régression Linéaire

Le modèle de régression linéaire est une modélisation qui postule une relation linéaire entre la variable cible et les variables explicatives. Cette relation se spécifie comme suit :

$$Y_i = \beta_0 + \beta_1.X_{i1} + \beta_2.X_{i2} + \dots + \beta_p.X_{ip} + \epsilon_i.$$

Ce modèle suppose plusieurs hypothèses telles que :

- la linéarité de la relation liant la variable dépendante (Y) aux variables explicatives ;

- l'absence de multicollinéarités (liaison linéaire) entre variables explicatives ;
- l'homocédasticité des résidus, c'est à dire que les résidus sont décorrélés et de variance constante ;
- ainsi que la normalité des résidus qui donne du sens aux intervalles de confiance et p-value.

L'estimation des coefficients $\beta_0, \beta_1, \dots, \beta_p$ se fait par minimisation de la somme du carré des résidus :

$$(\beta_0, \beta_1, \beta_2, \dots, \beta_p) = \underset{\beta}{argmin} ||X\beta - Y||_2^2 \quad (2.1)$$

Avec :

- X : la matrice de données avec une colonne de 1
- p : le nombre de paramètres
- Y : la variable cible

Ces coefficients ainsi obtenus permettent de voir l'effet marginal de chaque variable explicative sur la variable cible. Cependant, du fait de l'hypothèse de multicollinéarité, on est souvent amené à retirer certaines variables du modèle. Une meilleure gestion de cet aspect pourrait être effectuée avec un autre modèle tel que le modèle Elastic Net.

Elastic Net

Le modèle Elastic Net ne diffère de la Régression Linéaire que dans le postulat d'estimation des coefficients (équation 2.2). La technique Elastic Net est, en effet, un mixte de la Régression Rigde⁴ et la Régression Lasso⁵, en ce sens qu'il pénalise le critère d'optimisation à la fois avec la norme L1 et la norme L2. Le problème d'optimisation de ce modèle est la suivante :

$$\underset{\beta}{min} \frac{||X\beta - y||_2^2}{2n} + \alpha_1 ||\beta||_1 + \alpha_2 ||\beta||_2^2 \quad (2.2)$$

Cette pénalisation permet de gérer l'impact de la multicollinéarité sur l'estimation des coefficients.

Arbre de régression

L'arbre de régression est une technique d'apprentissage supervisé. C'est l'arbre de décision dans le cas où la variable cible est quantitative. Il existe plusieurs implémentations d'arbre de régression dont le CHAID, CART, C.5, etc. Nous utilisons dans cette étude l'algorithme CART qui est le plus utilisé.

Il s'agit d'un algorithme de moyennage local par partition. Comme illustré en annexe A.14, par une séquence de division binaire, l'algorithme arrive à prédire la valeur de la variable cible par la moyenne obtenue au niveau des feuilles de l'arbre (dernier niveau de partition). Avant chaque division le modèle cherche la variable permettant de réduire l'impureté d'un noeud, ici mesurée par la variance de la variable dépendante. Chaque division cherche donc à maximiser la variation d'impureté :

$$division* = argmax s^2(t) - \frac{n(t_g)}{n(t)} s^2(t_g) - \frac{n(t_d)}{n(t)} s^2(t_d) \quad (2.3)$$

Avec :

- $s^2(t)$: variance dans le noeud t qu'on cherche à segmenter ;

4. Ridge : pénalisation des coefficients avec la norme L2

5. Lasso : pénalisation des coefficients avec la norme L1

- $s^2(t_g)$: la variance dans le noeud fille de gauche ;
- $s^2(t_d)$: la variance dans le noeud fille de droite ;
- $n(t)$, $n(t_g)$, $n(t_d)$: taille d'échantillon respective.

Il est également possible de combiner plusieurs arbres de régression grâce aux méthodes ensemblistes tels que le Random Forest et le XG Boost.

Random Forest : "sagesse des foules" (Chloé-Agathe Azencott, 2018)

Le Random Forest fait partie de la classe de méthodes dites ensemblistes (ensemble method en anglais). Il a été proposé par Brieman en 2001. Le terme "ensembliste" fait référence au fait que ces classes d'algorithmes combinent de nombreux apprenants faibles afin d'obtenir un puissant modèle. Dans le cas du Random Forest, comme présenté sur la figure en annexe A.15, sont combinés plusieurs sous arbres peu profonds qui sont ensuite entraînés en parallèle sur le jeu de données. La prédiction du modèle final est enfin obtenu en faisant la moyenne des prédictions de ces faibles prédicteurs.

XG Boost : (Tianqi Chen et al., 2016)

Le XG Boost (Extreme Gradient Boosting) est un algorithme ensembliste qui combine la descente de gradient⁶ et la méthode Boosting⁷. Il est réputé être un algorithme performant. Dans le processus d'optimisation, l'algorithme met en jeu une fonction de coût à minimiser (somme des carrés des résidus dans le cas de la régression) et une fonction pour pénaliser la complexité du modèle pour éviter le sur-apprentissage. Supposons que nous ayons K arbres à entraîner de manière séquentielle et à l'itération t , la valeur prédite est donnée par : $\hat{y}_i^t = \sum_{k=1}^t f_k(X_i)$, où X_i désignent l'ensemble des variables explicatives et $f_k()$ la fonction de régression obtenu à l'itération k . La fonction à optimiser par l'algorithme est donnée par :

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^t) + \Omega(f_t(X_i)) = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(X_i)) + \Omega(f_t(X_i)) \quad (2.4)$$

Avec $\Omega(f_t(X_i)) = \gamma.T + \lambda.||f_t(X_i)||_2^2$ est la fonction qui pénalise l'arbre. La partie $\gamma.T$, où T représente le nombre de feuilles de l'arbre, permet d'éviter le sur-apprentissage dû à un arbre trop profond. Le deuxième terme ($\lambda.||f_t(X_i)||_2^2$) permet d'éviter des prédictions trop grandes. Ces paramètres γ et λ seront déterminés en validation.

La différence du XG Boost par rapport aux autres méthodes de Gradient Boosting réside dans le fait que la fonction $f_t(X_i)$ présente dans l'équation 2.4 est approchée par un développement de Taylor à l'ordre 2, ce qui permet une convergence plus rapide. Ainsi par une succession de minimisation de proche en proche, l'algorithme parvient à trouver la meilleure fonction de prédiction permettant de minimiser la fonction de coût global.

6. Technique itérative permettant d'approcher la solution d'un problème d'optimisation

7. Processus d'amélioration en proche en proche d'une séquence de modèle

2.4.2 Approche top-down

L'approche Top-Down consiste à estimer les émissions des entreprises à partir des émissions de GES de leurs filières respectives. Le choix de cette approche dans le cadre de notre étude est justifié par la nature des données disponibles. En effet, les TPE/PME ne disposant pas de moyens suffisants pour évaluer leurs risques de transition, et étant donné la disponibilité de cette information au niveau filière, l'approche Top-Down va nous permettre d'estimer le risque de transition de ces entreprises en partant de leur secteur.

Modèle ratio

Dans la littérature (cf. [2]), nous avons remarqué qu'une seule variable est utilisée pour faire le top-down, comme utilisé par l'ADEME pour calculer les émissions financées cité plus haut dans la section revue de la littérature. Par exemple, en supposant que le chiffre d'affaire explique bien les émissions de GES, on aura le modèle suivant :

$$\frac{EmissionGES_{entreprise}}{EmissionGES_{filiere}} = \frac{ChiffreAffaires_{entreprise}}{ChiffreAffaires_{filiere}}$$

où la variable à expliquer est l'émission des GES de l'entreprise. Pour éviter une erreur de prédiction élevée, nous allons utiliser plusieurs combinaisons de variables. Nous aurons ainsi une spécification qui respecte la définition du modèle ci-après.

Considérons $s \in \{1, \dots, S\}$ les filières des TPE/PME, et s_i la TPE/PME i de la filière s . Le modèle de prédiction top-down par les ratios est donné par :

$$\frac{Y_{s_i}}{Y_s} = \sum_{j=1}^p \alpha_j \frac{X_{j,s_i}}{X_{j,s}} \quad (2.5)$$

avec $j \in \{1, \dots, p\}$, $0 \leq \alpha_j \leq 1$, et $\sum_{j=1}^p \alpha_j = 1$

p est le nombre de variables explicatives

X_{j,s_i} est une variable caractéristique de l'entreprise s_i qui peut être le chiffre d'affaires, l'investissement, etc. $X_{j,s}$ est la même variable caractéristique mais pour la filière ; Y_s est l'émission de la filière s

L'objectif ici est de prédire Y_{s_i} qui est la quantité émise de gaz à effet de serre par l'entreprise i du secteur s .

Tout le problème se résume donc à trouver les bonnes combinaisons de α_j qui respectent les contraintes du modèle top-down théorique, à l'instar du modèle linéaire.

Formalisation

Rappelons que la méthode top-down dans le domaine de risque climatique est toujours dans une étape exploratoire. Ainsi, les formalisations que nous allons proposer sont aussi à titre exploratoire.

Cependant, étant donné que nous ne connaissons pas, a priori la bonne combinaison des prédicteurs des émissions des GES, nous avons utilisé plusieurs variables de notre base de données pour faire le modèle. Finalement, ceci reviendra à faire un modèle prédictif sur les variables ratios, comme spécifié plus haut sans la contrainte sur les α_j . Ainsi, à cet effet, nous proposons de tester certains modèles décrits dans la section 2.4.1 : régression linéaire, régression ridge, régression Lasso et Elastic Net. Les modèles linéaires vont être sans constante pour respecter la définition du Top-Down.

Pour résumer, l'approche top-down va se réduire à des modèles de ratios "valeurs entreprise sur valeurs filière" où la variable à expliquer est le ratio émissions de GES de TPE/PME et émissions des filières correspondantes.

Nous réalisons aussi un modèle au niveau filière, pour approcher le comportement en terme d'émissions de GES des TPE/PME à travers celles de leurs filières.

Modèle niveau filière

L'idée de cette méthode est d'approcher le comportement en émission des TPE/PME par le comportement moyen observé au niveau filière. Plus précisément, il revient à entraîner un modèle sur les filières des TPE/PME normalisées par le nombre d'entreprises, puis utiliser ce modèle pour prédire les émissions et évaluer l'ampleur des erreurs. Ce modèle, dans le cas d'une régression linéaire se spécifierait comme suit :

$$\bar{Y}_s = \beta_0 + \beta_1 \bar{X}_{1s} + \dots + \beta_p \bar{X}_{ps} + \epsilon_s \quad (2.6)$$

Avec :

- \bar{Y}_s : les émissions moyennes par filière ;
- \bar{X}_s : les caractéristiques moyennes par filière.

Tout comme une imputation par la moyenne, modéliser le comportement individuel par le comportement moyen peut engendrer de grandes erreurs, raison pour laquelle nous utiliserons les modèles classiques pour mesurer l'ampleur de cette erreur.

Conclusion partielle

En résumé, nous disposons de diverses sources de données avec lesquelles nous avons construit notre échantillon de travail et notre indicateur du niveau de risque de transition. Les deux approches de modélisation sus-citées seront comparées à l'aide des erreurs quadratiques moyennes. En effet, la variable cible étant quantitative, notre critère d'évaluation est l'erreur quadratique moyenne. En plus, nous appuyons l'évaluation par l'étude de surapprentissage à l'aide de la courbe d'apprentissage construite par validation croisée. Pour comparer le meilleur modèle classique au modèle top-down, nous étudions l'intervalle de confiance des erreurs quadratiques moyennes et élaborons un test de différence des EQM. Les résultats issus de ces analyses sont présentées ci-dessous.

CHAPITRE 3

EVALUATION DU RISQUE DE TRANSITION

Ce chapitre fait ressortir les résultats de l'étude. Il présente d'abord le profil des entreprises en termes de pollution. Ensuite, il ressort les résultats des modèles classiques et du Top-Down et compare ces deux approches. Enfin, une discussion des résultats globaux est faite en dernière position.

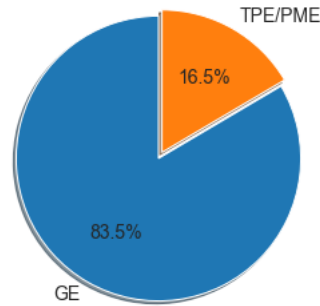
3.1 Répartition de l'échantillon par catégorie d'entreprises et analyse des émissions

3.1.1 Un échantillon dominé par les Grandes entreprises

Les entreprises constituant notre échantillon de travail sont, pour une grande part, à hauteur de **83,5%** des grandes entreprises (**GE**), soit **1026** entreprises contre **203 TPE/PME**. Cette répartition loin d'être étonnante viendrait de l'obligation qu'ont les grandes entreprises au regard de la loi française¹ de disposer d'une comptabilité de leurs émissions de gaz à effet de serres, alors que les plus petites entreprises n'y sont pas soumises. Avec un très grand nombre de GE dans le secteur tertiaire (442 GE contre 121 TPE/PME), ainsi que celui de l'industrie manufacturière (484 GE contre 38 TPE/PME), notre échantillon est beaucoup dominé par ces secteurs qui représentent à eux deux 75% des entreprises.

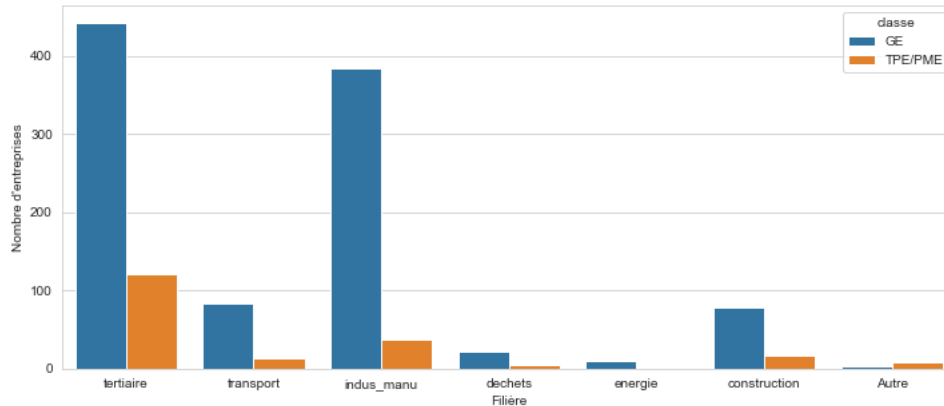
1. Loi portant Engagement National pour l'Environnement du 12 juillet 2010, article L.229-25

GRAPHIQUE 3.2 – Répartition des entreprises par catégorie



Source : Données ASDES, calculs des auteurs

GRAPHIQUE 3.3 – Distribution des émissions



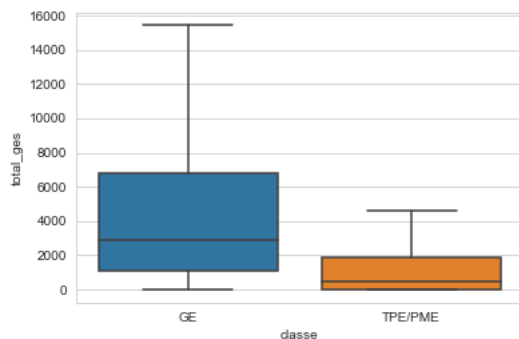
Source : Données ASDES, calculs des auteurs

3.1.2 Distribution des émissions de GES selon la catégorie d'entreprises

L'analyse de la distribution des émissions de gaz à effet de serres fait ressortir que les grandes entreprises sont les plus polluées. Les émissions considérées ici sont celles des scopes directs 1 et 2. Ce choix est conforté par le graphique A.19 en annexe faisant ressortir la distribution par scope des émissions de GES, qui montre que les émissions du scope 3 sont très faibles dans chaque secteur. Il ressort également du graphique 3.7 ci-dessous que les 37% des GE qui émettent le moins de gaz à effet de serre (37e centile) émettent autant que les 75% des entreprises qui émettent le moins au sein des TPE/PME (75e centile). Enfin, les 1% des grandes entreprises qui émettent le moins de gaz à effet de serre en émettent autant que les 23% des entreprises TPE/PME qui en émettent le moins.

Un autre angle de vision, celui du marché sur lequel opèrent les entreprises, illustré par le graphique ?? en annexe corrobore nos précédentes analyses. En effet, les grandes entreprises ont tendance à s'implanter sur des marchés de grandes tailles, à l'échelle nationale ou internationale, et donc ces marchés sont naturellement ceux qui émettent le plus de gaz à effet de serres soit environ 5000 Mégatonnes de CO_2 équivalent pour chacun des deux marchés. Les entreprises qui font leurs activités sur des marchés locaux, en émettent dans une moindre mesure, ce qui s'explique par la proportion plus élevée de TPE/PME dans ce marché, comme le décrit le graphique ?? en annexe.

GRAPHIQUE 3.4 – Distribution des émissions (en tonnes de CO2 équivalent)

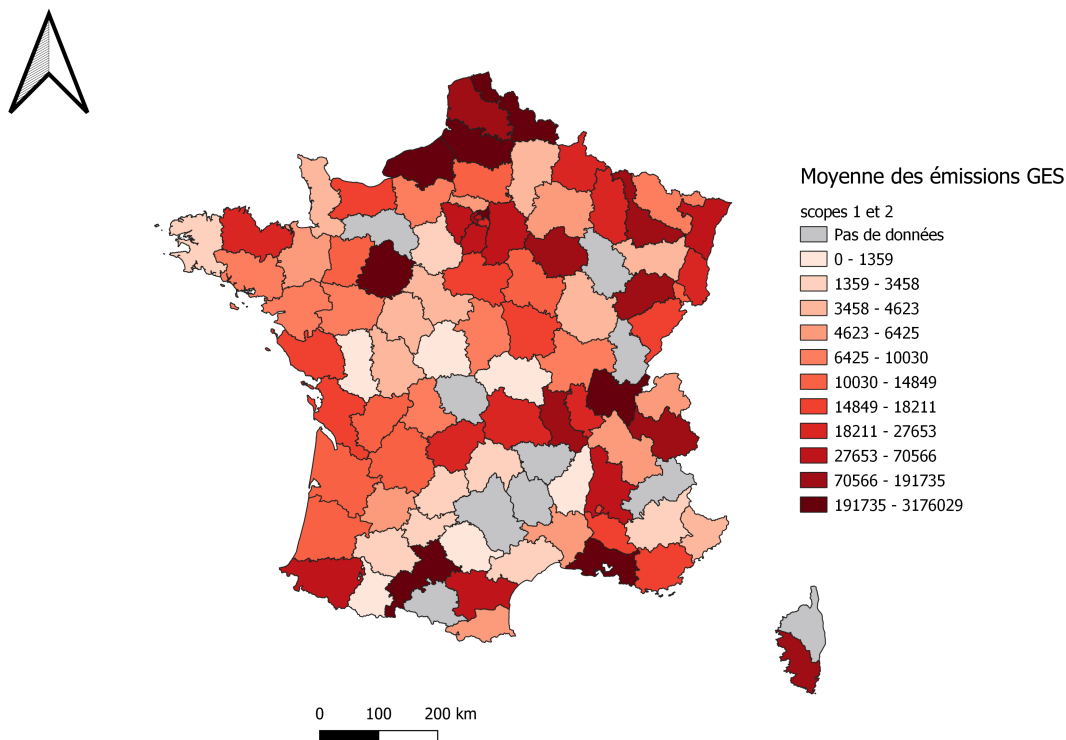


Source : Données ASDES, calculs des auteurs

3.1.3 Emissions des TPE/PME selon le département

L'analyse géographique des émissions des scope 1 et 2 par département montre que les TPE/PME des grandes villes françaises émettent beaucoup plus de gaz à effet de serre. En effet, la représentation 3.5 montre une forte concentration des émissions au nord, notamment au département du nord de Hauts-de-France et en Ile-de-France qui concentrent plus de 190 000 tonnes d'équivalent CO2. Au centre-Ouest et au Sud les émissions sont relativement plus basses et atteignent à peine 4600 tonnes de CO2 équivalent.

GRAPHIQUE 3.5 – Moyenne des émissions GES par département (en tonnes)



Source : Données ASDES, calculs des auteurs

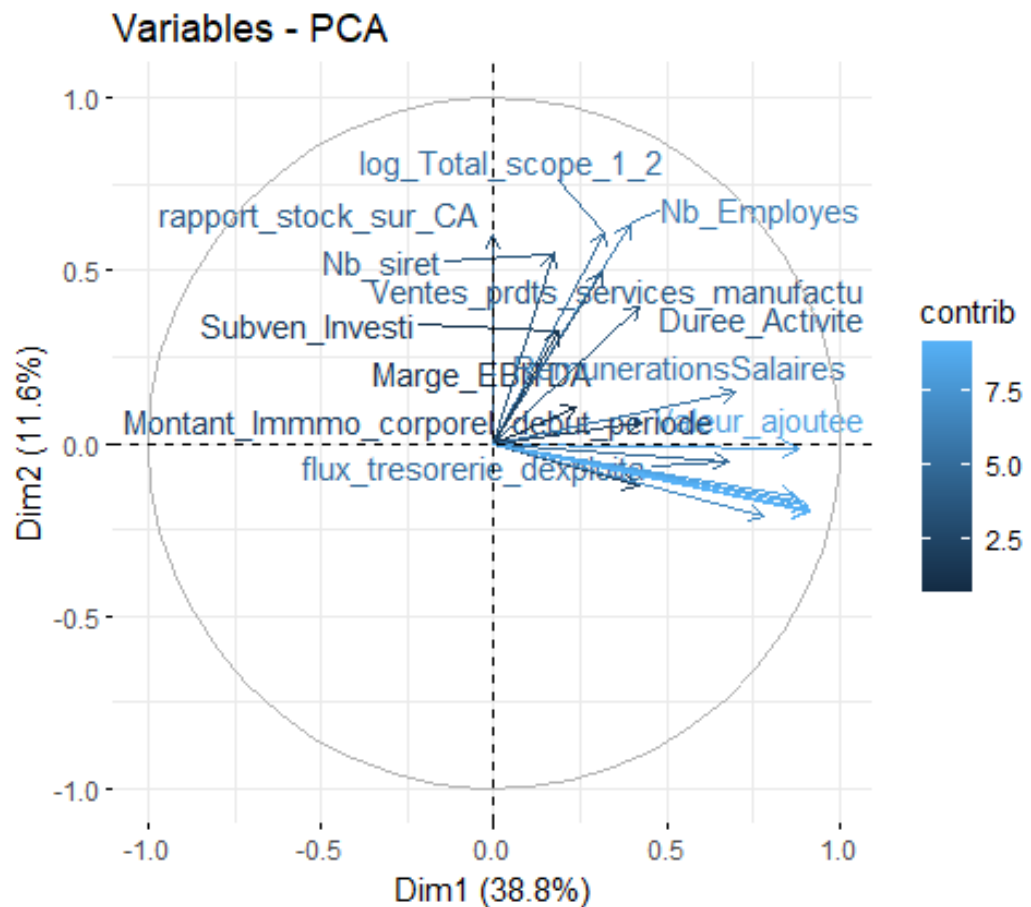
3.1.4 Analyse de liaison entre les variables

L'idée de cette partie est de faire ressortir les corrélations entre les variables disponibles à travers une analyse en composantes principales. Ainsi, l'ACP nous montre deux catégories de variables :

- variables sur l'intensité de l'activité de l'entreprise, très corrélées entre elles et avec la variable cible ;
- variables concernant les flux financiers, très corrélées entre elles mais moyennement corrélées avec la variable cible.

De plus, cette analyse multivariée va nous permettre une meilleure sélection de nos régresseurs. Les variables utilisées pour l'ACP rentrent dans le cadre d'un processus de filtrage des informations détaillé en annexe A.4, A.5 et A.6, passant d'abord par des tests de corrélation pour les variables quantitatives ainsi que des Anova les qualitatives, puis une sélection pas à pas basée sur le critère du R2 ajusté et de l'AIC. Ces filtrages permettent d'éviter la violation de certaines hypothèses dans la modélisation notamment celle de la multicolinéarité dans le cas des modèles linéaires.

GRAPHIQUE 3.6 – Analyse de correspondances des variables issues filtres Forward



Source : Base ASDES, calcul des auteurs.

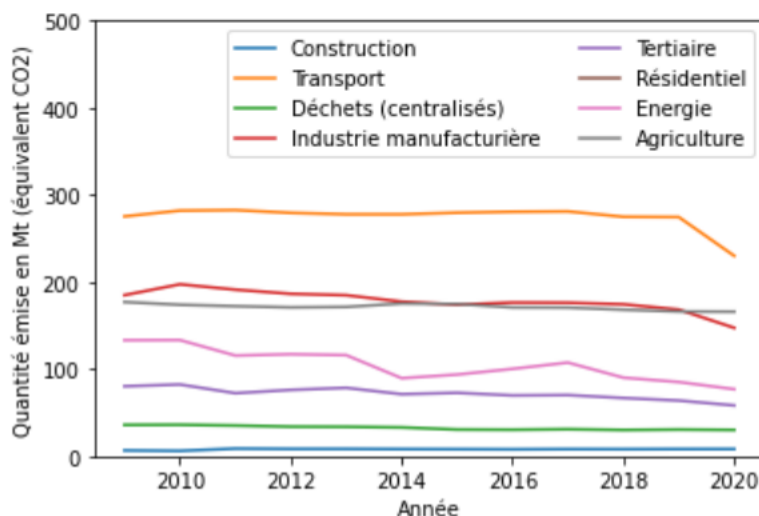
Les analyses descriptives précédentes sont principalement faites à partir des informations sur les entreprises. Pour s'intéresser aux bases filière, nous réalisons l'analyse de la section suivante.

3.1.5 Le transport, secteur très pollueur

Les données filière contiennent 216 observations faisant état des caractéristiques des différents secteurs sur une période de 12 ans.

L'analyse des émissions de GES des différents secteurs d'émissions montre que, sur 12 ans, le secteur du **transport** est celui qui émet le plus de GES, avec plus de 250 Mégatonnes de CO₂ équivalent. De même, on note une légère tendance à la baisse des émissions de ce secteur en 2020. Le secteur de l'industrie manufacturière et celui de l'agriculture sont respectivement le deuxième et le troisième secteur les plus émetteurs de GES, avec une quantité émise en équivalent CO₂ évaluée à près de 200 Mégatonnes. Toutefois, le secteur de la construction reste le secteur qui en émet le moins, d'après les résultats du graphique 3.7.

GRAPHIQUE 3.7 – Évolution des émissions de GES par secteur (en Mt de CO₂e)



Source : Données Citepa, calculs des auteurs

Une analyse des corrélations est également effectuée sur la base filière. Les résultats de cette analyse en annexe A.31 montrent que les variables caractéristiques financières du secteur (Chiffre d'affaires, Excédent brut d'exploitation, Salariés en temps plein, etc.) sont fortement corrélées entre elles mais faiblement avec la variable cible (émissions de GES de la filière). Ceci donne d'ores et déjà une première piste des potentiels problèmes d'estimation pour le modèle filière, qui risquerait d'avoir peu de variables explicatives significatives.

3.2 Résultats des modèles classiques

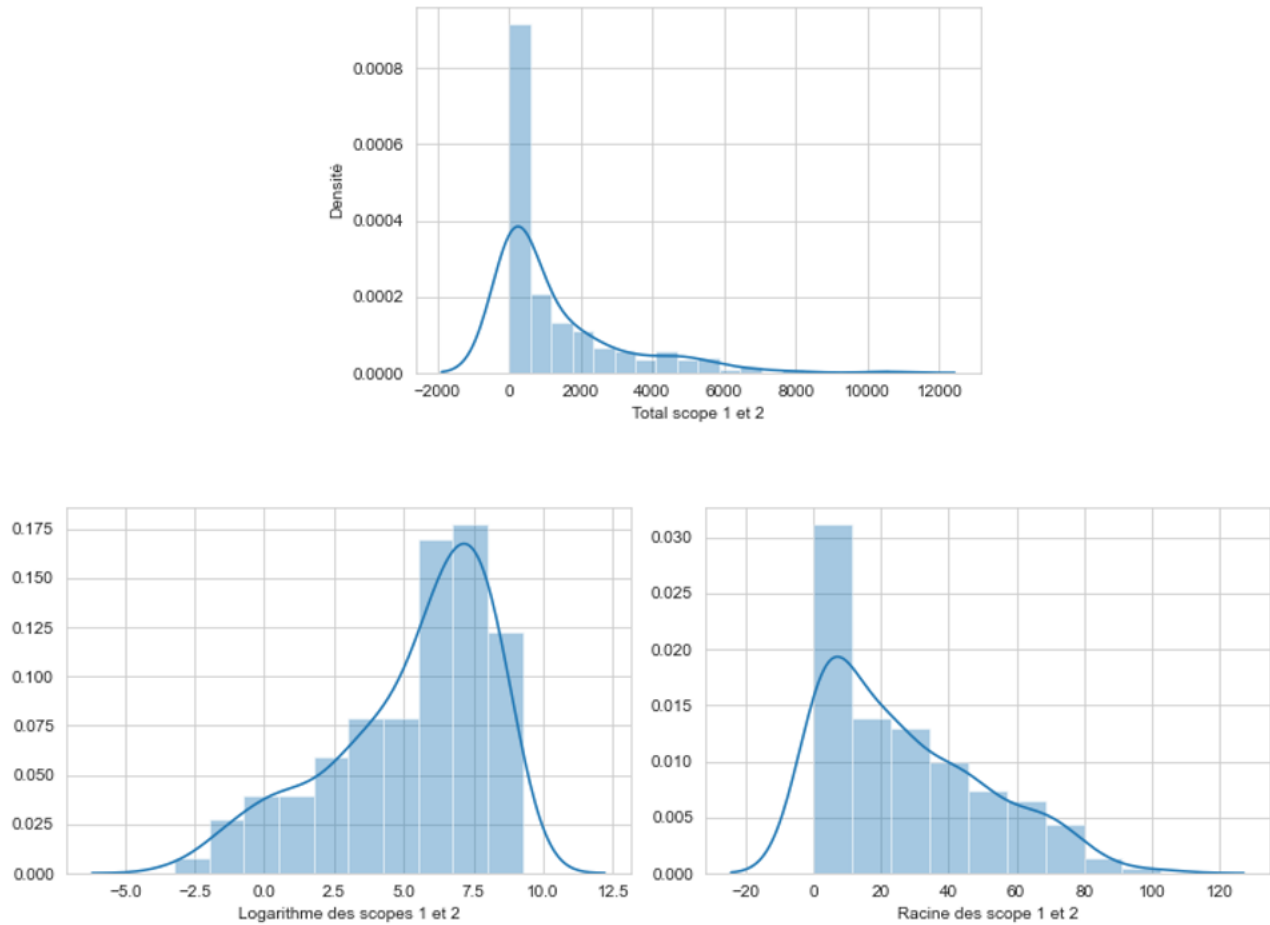
Dans cette section, nous présentons la transformation de notre variable cible ainsi que les résultats des modèles entraînés sur l'échantillon des TPE/PME (choix justifié ici A.16). La section présente également en détail les résultats du modèle final choisi.

3.2.1 Transformation de la variable cible

Afin d'éviter l'effet de variance trop élevée dans le modèle, nous avons procédé à deux transformations de notre variable à expliquer. Comme l'illustre le graphique ci-dessous, la transformation logarithmique

est plus élargie et a une forme qui se rapproche le plus de la courbe en cloche d'une loi normale. Sur le tableau contenant les descriptions des trois formes, nous constatons que cette transformation à une variance beaucoup plus faible que les autres. Ainsi nous garderons cette transformation dans la suite pour les modèles.

GRAPHIQUE 3.8 – Transformations de la variable cible



Source : Bases ASDES, calculs des auteurs

TABLEAU 3.1 – Description des transformations

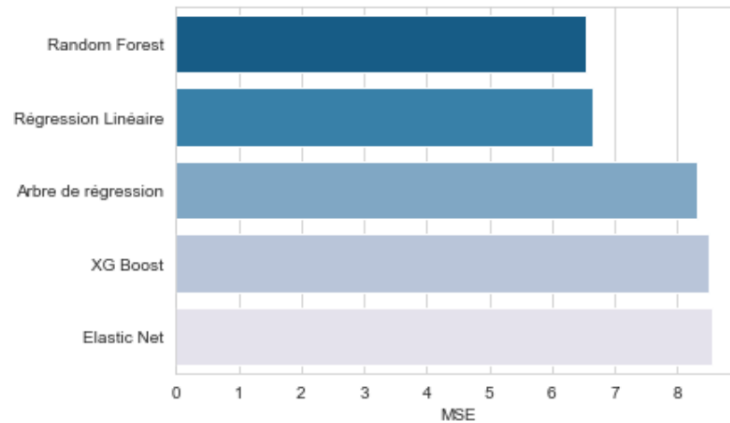
| Statistiques | Total scope 1 et 2 | Log.(Total scope 1 et 2) | Racine carrée(scope 1 et 2) |
|--------------|--------------------|--------------------------|-----------------------------|
| count | 203 | 203 | 203 |
| mean | 1289,070 | 5,287 | 27,043 |
| std | 1818,533 | 2,854 | 23,675 |
| min | 0,040 | -3,227 | 0,199 |
| 25% | 40,689 | 3,703 | 6,374 |
| 50% | 483,874 | 6,182 | 21,997 |
| 75% | 1861,211 | 7,528 | 43,139 |
| max | 10546,622 | 9,263 | 102,697 |

Source : Bases ASDES, calculs des auteurs

3.2.2 Choix du meilleur modèle

Dans l'approche de modélisation classique, nous avons comme variable cible le logarithme des émissions scope 1 et 2 et comme variables explicatives celles issues des procédures de sélection (voire annexe A.3). Sur ces données, nous avons appliqué différents modèles que sont la régression linéaire, la régression pénalisée Elastic Net, l'arbre de régression, le Random Forest (Forêt aléatoire) et le XG Boost. Comme le montre le graphique ci-dessous, la Random Forest est sortie comme étant le meilleur modèle au sens de l'erreur quadratique moyenne suivi de la régression linéaire.

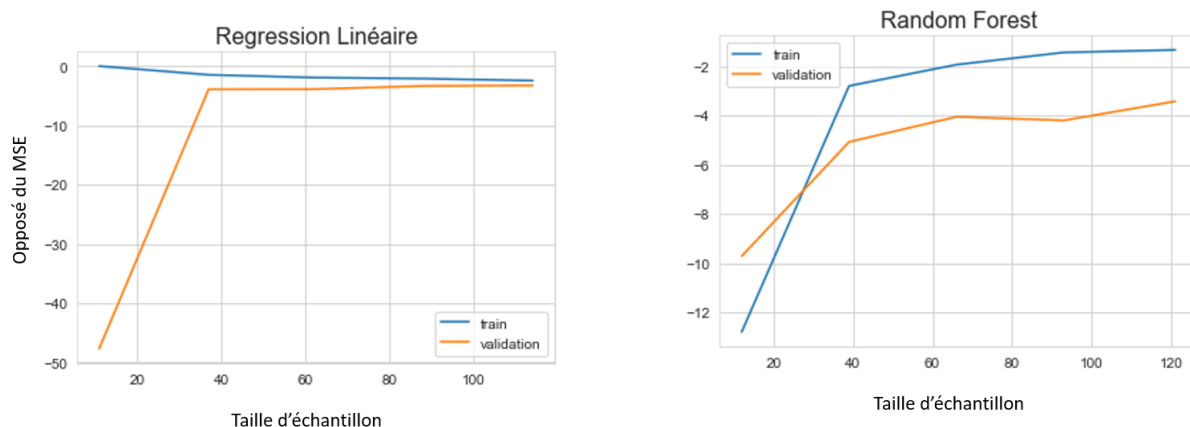
GRAPHIQUE 3.9 – Erreurs quadratiques moyennes calculée sur l'échantillon test sur les différents modèles



Source : Bases ASDES, calculs des auteurs

Cependant, lorsque nous tenons compte du niveau de surapprentissage des modèles (annexe A.6.1) nous voyons que le modèle de régression linéaire est moins sensible à ce phénomène. En effet, comme illustré ci-dessous, nous remarquons que les performances de la régression en validation et apprentissage sont assez proches contrairement au Random Forest. Ainsi pour la classe de modèle classique nous avons retenu la régression linéaire.

GRAPHIQUE 3.10 – Comparaisons du niveau de surapprentissage entre la Régression Linéaire et le Random Forest



Source : Bases ASDES, calculs des auteurs

3.2.3 Résultat de la régression linéaire

Le modèle de régression retenu se spécifie de la façon suivante :

$$\text{Log}(\text{EmissionsGES})_i = \beta_0 + \beta_1 \text{NombreEmployes}_i + \beta_2 \text{DureeActivite}_i + \beta_3 \text{StockSurCA}_i + \beta_4 \text{Pays_FR}_i + \beta_5 \text{FiliereIndusManu}_i + \beta_6 \text{FiliereTransport}_i + \beta_7 (\text{DureeActivite})_i^2 + \epsilon_i \quad (3.1)$$

Le modèle de régression linéaire choisi explique 64,1% de la distribution de la variable cible ($R^2 = 64,1\%$). Les tests de normalité des résidus et d'hétéroscédasticité n'ont décelé aucune violation des hypothèses (voire annexe A.23). L'analyse du tableau des coefficients ci-dessous permet de voir les déterminants des émissions des scopes 1 et 2 chez les TPE/PME.

D'abord, concernant le nombre d'employés, en moyenne chez une TPE/PME le recrutement d'un nouvel employé augmente de 0,53% les émissions annuelles de l'entreprise (toute chose égal par ailleurs). Cela pourrait se justifier par le fait qu'un nouvel employé entraîne une hausse des consommations en énergie et également un niveau d'activité plus élevé d'où l'augmentation des émissions en scope 1 et 2.

De surcroît, nous constatons que la durée d'activité de l'entreprise a un effet positif (6,42%, IC = [0,79% ; 12,04%]) mais cet effet diminue lorsque l'entreprise gagne plus de maturité (6,42%, - 0,07%*DuréeActivité). Autrement dit, en moyenne une entreprise ayant une durée d'activité de 10 ans en 2022 augmenterait son niveau de pollution annuelle de 5,72% en 2023, tandis que pour une entreprise datant de 20 ans en 2022, cette hausse serait de 5,02 points de pourcentage (toute chose égale par ailleurs). Ceci pourrait être dû au fait que les entreprises avec plus de maturité sur le marché prennent mieux en compte les aspects environnementaux dans leur processus de production, notamment avec la réduction de papiers et l'usage d'énergie verte.

Le modèle met également en exergue un effet significatif au seuil de 10% du ratio stock sur chiffre d'affaires. Une croissance unitaire de ce ratio génère en moyenne une hausse de 0,86% du niveau de pollution annuelle de l'entreprise, toute chose égale par ailleurs.

De plus, nous constatons que les entreprises françaises polluent en moyenne 2,5 fois plus que les autres. Les entreprises du secteur industrie manufacturière émettent également 2,5 fois plus que les entreprises des autres secteurs hors secteur transport (i.e. énergie, déchets centralisés, construction, agriculture, tertiaire et résidentiel). Enfin, ceux du secteur transport émettent en moyenne 2,9 fois plus de GES que les entreprises des autres secteurs hors industrie manufacturière (i.e. énergie, déchets centralisés, construction, agriculture, tertiaire et résidentiel).

Ce modèle nous servira ainsi de référence de comparaison pour apprécier l'ampleur des erreurs de l'approche Top Down.

TABLEAU 3.2 – Résultat du modèle de régression linéaire

| variables | Coef, | Std,Error | t | P> t | IC(95%) |
|----------------------------------|-----------|-----------|---------|--------|--------------------|
| const | 1,3300*** | 0,3891 | 3,4183 | 0,0008 | [0,5609 ; 2,0991] |
| Nombre d'employés | 0,0053*** | 0,0007 | 7,8383 | 0,0000 | [0,0039 ; 0,0066] |
| Durée d'activité | 0,0642** | 0,0284 | 2,2554 | 0,0256 | [0,0079 ; 0,1204] |
| Stock sur CA | 0,0086* | 0,0051 | 1,6810 | 0,0949 | [-0,0015 ; 0,0188] |
| Pays_FR | 1,5708*** | 0,2812 | 5,5858 | 0,0000 | [1,0149 ; 2,1266] |
| Filiere industrie manufacturière | 1,4939*** | 0,3684 | 4,0548 | 0,0001 | [0,7656 ; 2,2221] |
| Filiere transport | 1,8865*** | 0,5144 | 3,6675 | 0,0003 | [0,8697 ; 2,9033] |
| (Duree d'activite) ² | -0,0007* | 0,0004 | -1,6778 | 0,0956 | [-0,0015 ; 0,0001] |
| Significativité : | * :10% ; | **5% ; | *** :1% | | |

| | | | |
|----------------------|-------------------------------|----------------------|----------|
| Modèle : | <i>OLS</i> | R^2 ajusté : | 0,623 |
| Dependent Variable : | <i>Log Total scope 1 et 2</i> | AIC : | 565,4280 |
| No, Observations : | 151 | Log-Likelihood : | -274,71 |
| Df Model : | 7 | F-statistic : | 36,48 |
| Df Residuals : | 143 | Prob (F-statistic) : | 7,16e-29 |
| R-squared : | 0,641 | Scale : | 2,3517 |

Source : Bases ASDES, calculs des auteurs

3.3 Résultats des modèles Top-Down

Dans l'approche top-down, comme évoquer dans la méthodologie, nous avons deux approches de modélisations : un modèle sur les variables ratios et un modèle avec les données filière.

3.3.1 Modèle ratio

Le modèle ratio proposé est la prédiction du ratio émissions de GES de l'entreprise sur les émissions de la filière à partir des ratios d'autres variables financières. Pour la modélisation, le logarithme de la variable dépendante a été utilisé. Les coefficients du modèle vont donc s'interpréter en terme de semi-élasticité. De plus, l'interprétation du modèle se fera en terme de variation de la part de l'entreprise dans sa filière. La méthode pas à pas (à l'aide du critère AIC) et l'analyse de la corrélation nous ont permis de sélectionner les meilleurs prédicteurs qui sont au nombre de six (6) : le ratio Chiffre d'affaires, le ratio nombre d'employés, le ratio chiffre d'affaires à l'exportation, l'appartenance à la filière Construction, à la filière déchets centralisés et la filière industrie manufacturière (A.8).

Nous avons réalisé des modèles linéaires (régression linéaire, Elastic Net, régression ridge et régression Lasso). Les erreurs de prédiction par le calcul de l'EQM sont présentées dans le tableau 3.3. Ainsi, la régression ridge a l'EQM relativement la plus faible (9,524), suivi par l'Elastic Net et la régression Lasso (9,549) et enfin la régression linéaire (9,913). En combinant le critère basé sur l'EQM et les courbes d'apprentissage des figures A.35 et A.36, le meilleur modèle est la régression ridge : faible EQM et moindre surapprentissage.

TABLEAU 3.3 – Résultat des modèles ratio

| <i>Modèle</i> | MSE |
|---------------------|-------|
| Régression linéaire | 9,913 |
| Régression Lasso | 9,549 |
| Régression ridge | 9,524 |
| Elastic Net | 9,549 |

Source : Bases ASDES, calculs des auteurs

En considérant le modèle choisi (ridge), les coefficients associés aux prédicteurs sont présentés dans le tableau A.8 en annexe. En utilisant les coefficients, la prédiction du ratio des émissions de GES sera donnée par :

$$\logarithme(\hat{eges_rate})_i = -0,15*CA_rate - 0,18*NBEmpl_rate - 0,10*CA_export_rate - 12*Construction - 9,36*Dechets - 14,47*Indus_manu \quad (3.2)$$

3.3.2 Modèles filière

Les modèles au niveau filière ont été réalisés dans le but de comparer le comportement des filières en terme d'émissions de gaz à effet à celui des entreprises, en l'occurrence pour faire le "top-down". Pour rendre encore plus possible cette comparaison, chaque donnée au niveau filière a été divisée par le nombre d'entreprises de la filière. Les valeurs seront ainsi interprétées comme des moyennes des entreprises.

La liste des variables utilisées se trouve en annexe A.31. Notons également que, contrairement aux modèles classiques et modèles ratio, aucune transformation de la variable cible n'a été effectuée pour les modèles filière.

Trois modèles sont retenus pour la modélisation au niveau filière. Il s'agit de la régression ridge, l'arbre de décision et le Random Forest.

A l'issu de modèle de régression ridge, de l'arbre de régression et de la forêt aléatoire pour la régression, nous avons les résultats suivants (tableau 3.4) sur l'erreur de prédiction calculées à partir de l'échantillon test. Il ressort que l'arbre de régression a l'erreur quadratique moyenne la plus faible (EQM égale à 649 634,04). De plus, l'analyse de sur-apprentissage de ces modèles (voire graphiques A.32, A.33 et A.34) montre toujours le modèle de l'arbre de régression est le meilleur où les EQM en validation et en apprentissage sont faibles et convergent plus vite avec l'augmentation de la taille de la base.

TABLEAU 3.4 – Résultat des modèles filière

| <i>Modèle</i> | Erreur quadratique moyenne |
|---------------------|----------------------------|
| Régression ridge | 29764502,12 |
| Arbre de régression | 649634,04 |
| Random Forest | 750613,91 |

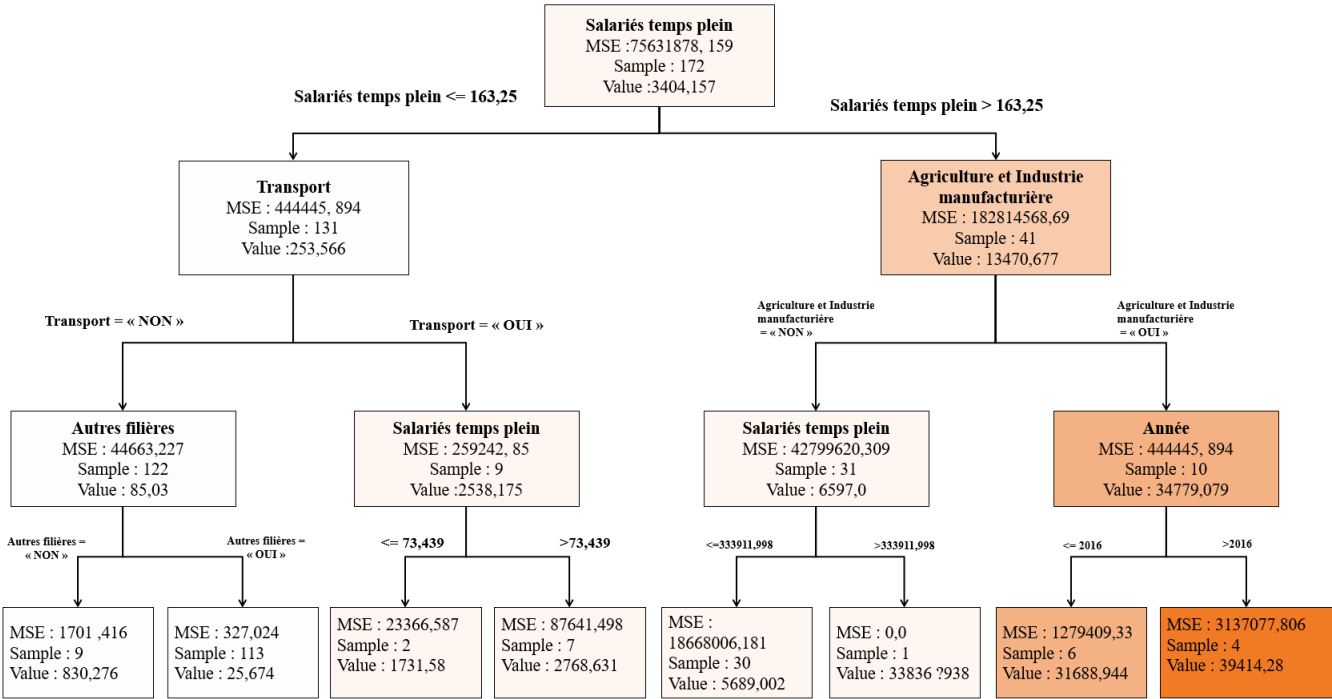
Source : Bases ASDES, calculs des auteurs

Description du modèle retenu

Le modèle ainsi retenu dans le cadre du modèle filière est l'arbre de régression. Cet arbre nous montre

que le premier déterminant des émissions des filières est le nombre de salariés en temps plein, où les secteurs ayant plus des salariés en temps plein polluent le plus. Les deuxièmes diviseurs de l'arbre sont respectivement la variable d'appartenance aux secteurs agricole et d'industrie manufacturière et la variable spécifiant le secteur du transport. En plus, les caractéristiques des gros pollueurs sont le regroupement des secteurs agricole et d'industrie manufacturière ayant en moyenne plus 163 salariés par entreprise au cours des années 2017, 2018, 2019 et 2020. Quant à celles des secteurs qui polluent le moins, ils ont en moyenne moins de 163 salariés en temps plein par entreprise et regroupe les secteurs de construction, de l'énergie, de la gestion des déchets et de l'activité tertiaire et résidentiel.

GRAPHIQUE 3.11 – Arbre de régression



Source : Bases Citepa & É sane, calculs des auteurs

3.4 Comparaison des modèles

Dans cette partie, nous allons comparer le modèle top-down à l'approche classique. Pour cela, nous nous basons sur la comparaison des erreurs de prédiction (EQM dans notre cas). Les erreurs quadratiques moyennes calculées sont basées sur la transformation logarithmique des émissions de GES. Ainsi, le modèle retenu dans l'approche classique (modèle linéaire) ressort une erreur quadratique moyenne de 6,64 tandis que le modèle retenu par l'approche top-down au niveau filière nous fournit une EQM de 10,83 inférieure à la première et celle du modèle ratio qui est de 9,52. Donc en terme absolu, l'approche classique peut être considérée comme la meilleure, qui minimise l'erreur de prédiction. Néanmoins, les intervalles de confiance de l'EQM de ces modèles laissent croire qu'ils peuvent être significativement les mêmes au seuil de 5%. En effet, l'EQM du modèle classique appartient à 95% de chance à l'intervalle 4,10 à 9,18, et celle du modèle top-down au niveau filière à l'intervalle 6,12 à 15,53. Pour le modèle ratio, l'EQM est à 95% de chance entre 5,32 et 13,74.

TABLEAU 3.5 – Erreurs quadratiques moyennes des modèles

| <i>Modèle</i> | MSE | IC à 95% |
|---------------------------|-------|----------------|
| Meilleur modèle classique | 6,64 | [4,10 ; 9,18] |
| Meilleur modèle filière | 10,83 | [6,12 ; 15,53] |
| Meilleur modèle ratio | 9,52 | [5,32 ; 13,74] |

Source : Bases ASDES, calculs des auteurs

En utilisant le test de différents de moyennes de variance égale et le test d'échantillons appariés, nous avons les résultats suivants :

TABLEAU 3.6 – Test de différences de moyennes

| <i>Comparaison de l'EQM des modèles</i> | Non-appariés | | Appariés | |
|---|--------------|---------|-------------|---------|
| | Statistique | p-value | Statistique | p-value |
| Modèle classique et modèle filière | -1,52 | 0,13 | — | |
| Modèle classique et modèle ratio | -1,13 | 0,26 | — | |
| Modèle ratio et modèle filière | -0,40 | 0,69 | -0,89 | 0,38 |
| * : 10% ; ** : 5% *** : 1% | | | | |

Source : Bases ASDES, calculs des auteurs

Ainsi, au seuil de 5%, on accepte l'égalité en terme d'erreur de prédiction (MSE). En effet, la p-valeur retournée par le test vaut 0,13 et est supérieure à 0,05. On ne peut pas donc rejeter l'hypothèse nulle du test qui est l'égalité des erreurs quadratiques moyennes.

Conclusion : les deux approches font significativement les mêmes erreurs de prédiction. Donc on peut considérer l'approche top-down comme une alternative pour estimer les émissions de GES des TPE/PME.

3.5 Discussions des résultats

L'approche Top-down telle que nous l'avons appréhendée permet d'obtenir des résultats très significatifs dans l'estimation des émissions de gaz à effet de serre. En effet, la comparaison statistique que nous avons effectuée confirme qu'il n'y a pas de différences significatives entre les modèles classiques et le modèle top-down en terme d'erreur de prédiction.

Au niveau filière, il ressort que les facteurs explicatifs des émissions de GES sont le nombre d'employés et l'indicatrice des filières. Une taille élevée en nombre d'employés augmente très significativement les émissions de ces filières. En plus, pour un nombre de salariés en temps plein élevé, la filière industrie manufacturière et la filière agriculture polluent le plus. La filière transport, malgré un faible effectif d'employés, est la grosse émettrice de gaz à effet de serre. C'est d'ailleurs le résultat montré par l'analyse descriptive. Les variables financières telles le chiffre d'affaires, l'investissement, etc. sont fortement corrélés entre elles mais faiblement avec les émissions de GES des filières, comme le montre le graphique A.31 en annexe. Par conséquent, ils n'ont pas significativement d'effets sur les émissions contrairement à ce qui est attendu. Ceci peut être expliqué par le fait que le niveau filière retenu dans cette étude par contrainte technique sont les principales branches d'activités d'économiques qui sont regroupés en huit (8) grands secteurs, donc inévitablement corrélés.

CONCLUSION

Au terme de cette étude, il apparaît que l'évaluation du risque de transition au niveau des TPE/PME, bien qu'étant un enjeu de taille dans une dynamique micro et macro prudentielle des institutions financières, est une tâche relativement complexe. Cette complexité est notamment accentuée par la difficulté du traçage de l'empreinte carbone de cette catégorie d'entreprises qui ne représente que 16,5% du bilan de l'ADEME. L'état de l'existant a montré que cette question essentielle n'avait pas, jusque là, été abordée. C'est ainsi que, disposant des données sur les bilans des émissions fournis par l'ADEME, des données entreprises ainsi que les données au niveau secteur (Citepa et ESANE), nous nous sommes fixés comme objectif la construction d'un modèle permettant de prédire les émissions de gaz à effet de serre des TPE/PME comme proxy de leur niveau d'exposition au risque de transition. Dans cette dynamique, deux approches ont été appliquées : une modélisation classique, utilisant les données niveau entreprise ; et une modélisation Top Down, partant du niveau secteur pour estimer les émissions niveau entreprise.

Dans la modélisation classique, la régression linéaire a été retenue et ses résultats font ressortir les principaux facteurs de vulnérabilité face au risque de transition. Ce modèle montre une grande disparité entre secteurs avec les entreprises du transport et de l'industrie manufacturière qui émettent en moyenne 2,5 fois plus de GES que les autres entreprises des autres secteurs. Il fait également ressortir le nombre d'employés comme facteur d'exposition au risque de transition (+0,53% sur les émissions annuelles). Ces résultats corroborent ceux de la modélisation Top Down qui dans ses deux approches soulignent l'effet positif du nombre d'employés sur les émissions.

Les résultats de cette étude montrent également que l'EQM du modèle linéaire classique n'est pas significativement différente de celle obtenue au niveau de la modélisation Top Down, au seuil d'erreur de 5%. Au niveau de l'échantillon, ce résultat voudrait dire que le comportement moyen en termes d'émission au niveau filière reflète le comportement individuel. Un tel résultat permet donc un gain de temps et de ressources dans l'étude du risque de transition du fait de l'accès facile aux données niveau filière. Cependant, quelques limites essentielles méritent d'être soulignées.

Bien que la procédure d'apprentissage supervisé permet de réduire certains biais, celui lié est à la représentativité persiste (seulement 203 TPE/PME pour un total de plus de 3 millions). Cette sous représentation, quoique logique, empêche la garantie de la robustesse des estimations ainsi que la validité externe des résultats. De surcroît, la totalité des émissions n'ont pas été prise en compte en raison du double comptage présent dans le scope 3. De même, au niveau filière, les résultats auraient pu être désagregés par scope si cette information était disponible.

En définitive, cette première tentative d'utilisation de l'approche Top down pourrait être améliorée par la prise en compte des différents scopes avec un dédoublement efficace du scope 3. Au niveau filière il serait aussi intéressant de faire les analyses par scope. Par ailleurs, la question du risque de transition sur les TPE/PME pourrait également être posée dans le cadre des ménages, ces derniers étant également importants dans un portefeuille bancaire. On se demanderait alors comment la transition énergétique affecterait les habitudes de consommation et d'épargne de ces derniers et quel risque pour le système financier.

BIBLIOGRAPHIE

- [1] Ademe (2021). *Stress-tests climatiques par scénarios : De l'analyse des risques à la modélisation*, Document de travail, 71p.
- [2] Ademe (2014) : Une étude de cas d'une méthodologie « top-down » de quantification des émissions financées Secteur financier- Tome 3, 46p.
- [3] Bank for International Settlements (2021) : Climate risk drivers and their transmission channel, 45(1) 1-33p.
- [4] Bank for International Settlements (2021) : Climate-related financial risks- measurement methodologies, 56p.
- [5] Brigitte G. Apprentissage supervisé, Polycopie de cours, 2eme année, ENSAI, 2021.
- [6] Chloé-Agathe Azencott : Introduction au Machine Learning , 46p.
- [7] Citepa, juillet 2021. Inventaire des émissions de polluants atmosphériques et de gaz à effet de serre en France- Format Secten, 16-19.
- [8] B. Warr et al.(2006), REXS : A forecasting model for assessing the impact of natural resource consumption and technological change on economic growth.
- [9] R. Marcus et al. (2021). *ACT Step-by-Step : Methodology*, version 1 - february 2021.*SSM stress test 2022*, 61p.
- [10] European Central Bank (2022), Climate risk stress test-*SSM stress test 2022*, 61p.
- [11] European Central Bank (2022), Occasional Paper Series-*ECB Economy-wide climate stress test : Methodology and results*, 61p.
- [12] GERARDI A. et al. (2015) : La quantification des émissions de gaz à effet de serre des institutions financières, 14p.
- [13] OID (2020), Décryptage : ACT - Assessing Low-Carbon Transition, *Evaluer la maturité des entreprises vis-à-vis de la transition bas-carbone*, 12p.
- [14] Site internet :<https://ichi.pro/fr/un-guide-sur-la-regression-des-donnees-de-panel-theorie-et-pratique>
msclkid=7bab155cb8ce11eca19854349b4a2e29
- [15] Task Force on Climate-related Financial Disclosures, Recommendations of the Task Force on Climate Disclosure, 2017
- [16] Tianqi Chen et Carlos Guestrin, 2016, XGBoost : A Scalable Tree Boosting System.

ANNEXE A

ANNEXE

A.1 Rapprochement des secteurs Citepa aux filières APE

A l'aide des données à notre disposition, nous avons d'abord construit la variable d'intérêt et l'échantillon de travail comme décrit plus haut dans la méthodologie. Ainsi, nous avons deux subdivisions de la base de données : une base regroupant les données ADEME, SIREN et Données entreprises, et une autre regroupant les données filière. La base filière comporte des informations agrégées sur les émissions (Inventaire des émissions de 1990 à 2020) et sur les filières (données Esane de 2009 à 2019). Ces deux sources ne disposent pas la même nomenclature pour les filières. En effet, Esane utilise le code APE pour identifier les filières, tandis que Citepa utilise la nomenclature SNAP des secteurs d'émissions de GES et de polluants atmosphériques. Ainsi nous nous sommes limité au niveau 1 du code APE pour pouvoir fusionner ces sources de données. L'affectation que nous avons effectuée à cet effet est décrite dans le tableau ci-dessous.

TABLEAU A.1 – Rapprochement des secteurs Citepa aux filières APE

| code | secteur | description du secteur | Branche d'Activités |
|------|--------------------------|--------------------------|---------------------|
| 1 | energie | Industrie de l'énergie | Section B |
| 1 | energie | Industrie de l'énergie | Section D |
| 2 | utcatf | UTCATF | |
| 3 | transport | Transports | Section H |
| 4 | déchets | Déchets (centralisés) | Section E |
| 6 | émissions naturelles | Emissions naturelles | |
| 5a | construction | Construction | Section F |
| 5b | industrie manufacturière | Autres | Section C |
| 7a | résidentiel | Résidentiels | Section T |
| 7b | tertiaire | Tertiaires | Section G |
| 7b | tertiaire | Tertiaires | Section I |
| 7b | tertiaire | Tertiaires | Section J |
| 7b | tertiaire | Tertiaires | Section K |
| 7b | tertiaire | Tertiaires | Section L |
| 7b | tertiaire | Tertiaires | Section M |
| 7b | tertiaire | Tertiaires | Section N |
| 7b | tertiaire | Tertiaires | Section O |
| 7b | tertiaire | Tertiaires | Section P |
| 7b | tertiaire | Tertiaires | Section Q |
| 7b | tertiaire | Tertiaires | Section R |
| 7b | tertiaire | Tertiaires | Section S |
| 8 | agriculture | Agriculture/sylviculture | Section A |
| 8 | agriculture | Agriculture/sylviculture | Section U |

Source : Données Secten et Esane, nos calculs

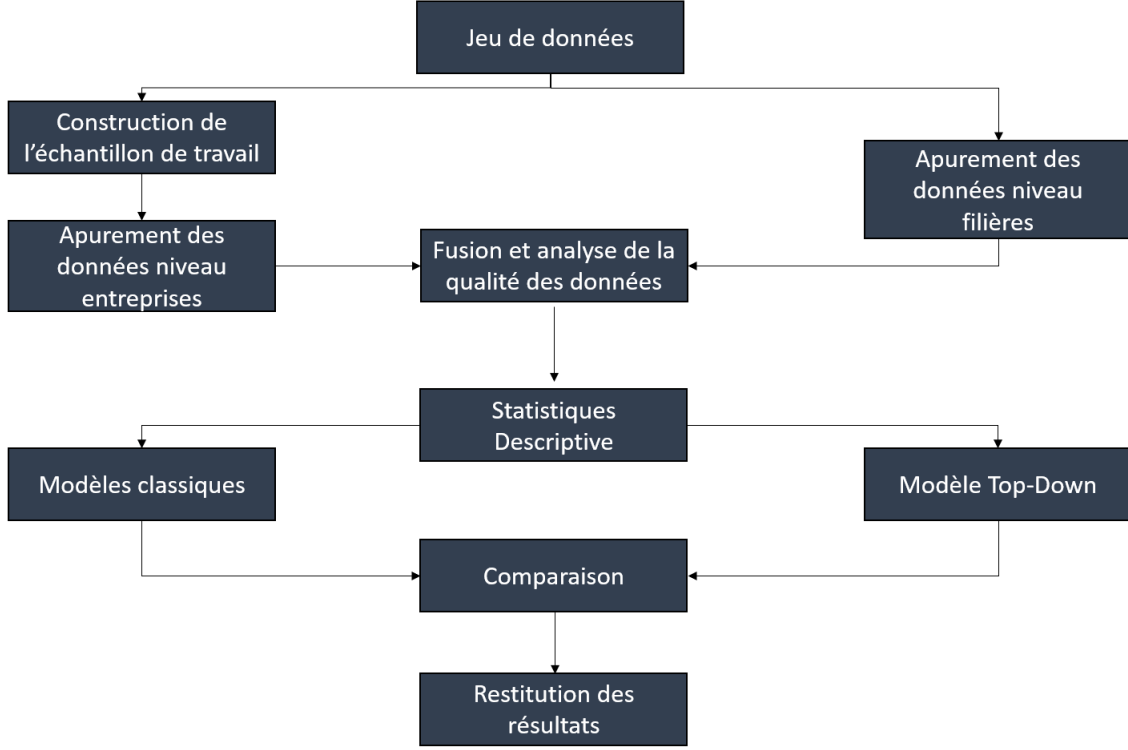
A.2 Complément méthodologique

A.2.1 Procédure de traitement des données

Après avoir construit nos deux bases de données, nous les avons apurées de manière séparée. La procédure d'apurement réalisée consiste d'abord à la suppression des variables ayant un taux de valeurs manquantes élevé (plus de 30%). Ce choix d'un seuil de 30% peut être expliqué le souhait de minimiser l'erreur pour les imputations des valeurs manquantes. Ensuite, pour les autres variables, nous avons utilisé une imputation par la méthode de k plus proches. Enfin, grâce à l'approche de boxplot, nous avons identifié et corrigé les valeurs extrêmes.

La démarche est décrite dans la figure A.12 ci-dessous.

GRAPHIQUE A.12 – Démarche de travail



Source : Auteurs

La qualité de la fusion des données est appréhendée par le taux de perdution qui est le ratio entre le nombre de valeurs manquantes après fusion et la taille de base de fusion. Les statistiques descriptives que nous allons réaliser consisteront d’abord au positionnement des TPE/PME, ensuite au profilage des entreprises et des filières en termes de émissions de GES, et enfin aux analyses multi-variées à travers une étude de corrélation et une analyse en composantes principales. Ces analyses descriptives nous permettront dès lors d’identifier les variables potentielles à la modélisation. Rappelons que nous avons deux approches à tester : l’approche classique pour estimer les émissions de GES et l’approche Top-Down qui consiste à utiliser les données filière pour estimer les emissions des entreprises. Dans la partie qui suit, nous allons décrire et justifier le choix sur les modèles d’apprentissage supervisé à utiliser et la formation du modèle top-down dans notre contexte.

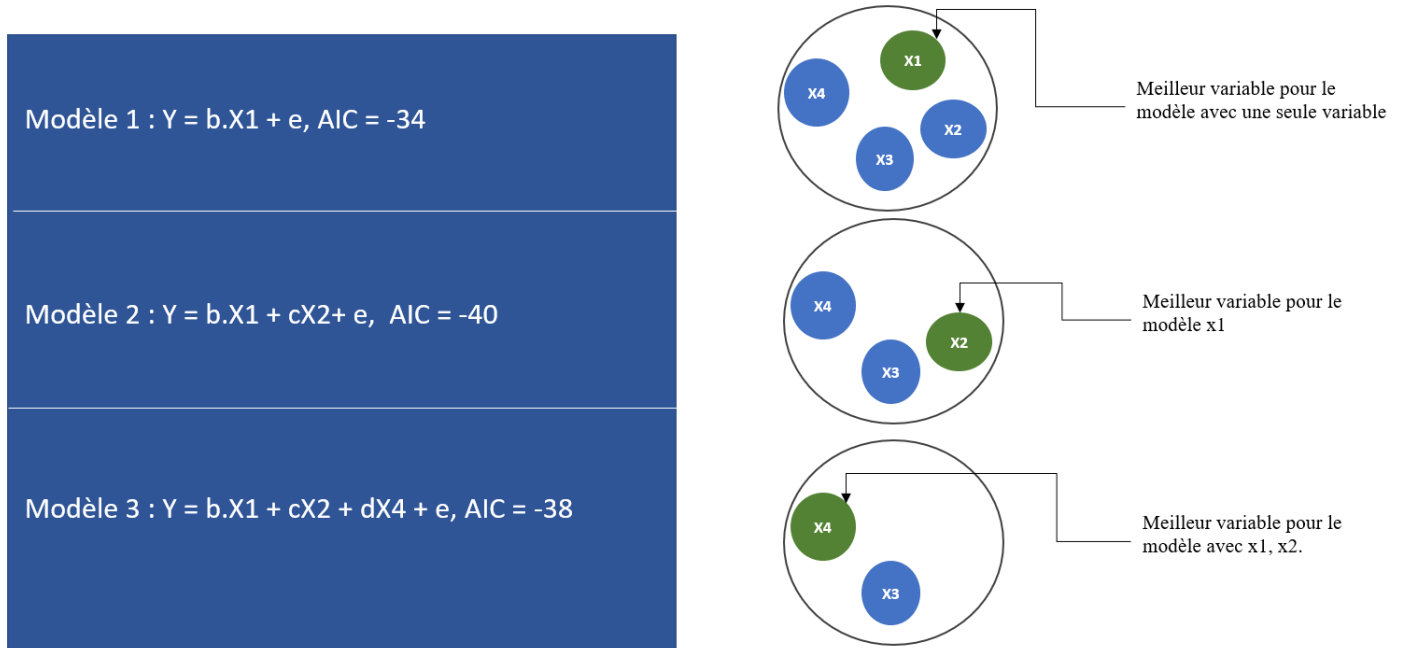
A.2.2 Sélection pas à pas : direction Forward

Les méthodes de sélection pas à pas sont des outils permettant de réduire la dimension des données en ne gardant que les variables les plus pertinentes. Dans notre étude nous avons utilisé l’approche dite Forward. Cette approche, comme illustré ci-dessous, choisit les variables une à une pour former des sous ensembles de variables qui optimisent à chaque fois le critère donné (AIC, BIC ou R^2_{ajust} . Dans notre étude nous avons utilisé l’AIC et le R^2 ajusté qui sont donnés par :

- $AIC = -2L_n(\beta) + 2p$, avec $L_n(\beta)$ étant la vraisemblance de l’échantillon et p le nombre de coefficients du modèles linéaires ;
- $R^2 \text{ ajusté} = 1 - \frac{n-1}{n-p-1} \frac{\text{Somme carrés résiduels}}{\text{Somme carrés totaux}}$

Ainsi le modèle qui maximise le R^2 ajusté ou minimise l’AIC donnerait un meilleur ajustement des données. Dans l’illustration ci-dessous, le modèle final retenu serait le deuxième.

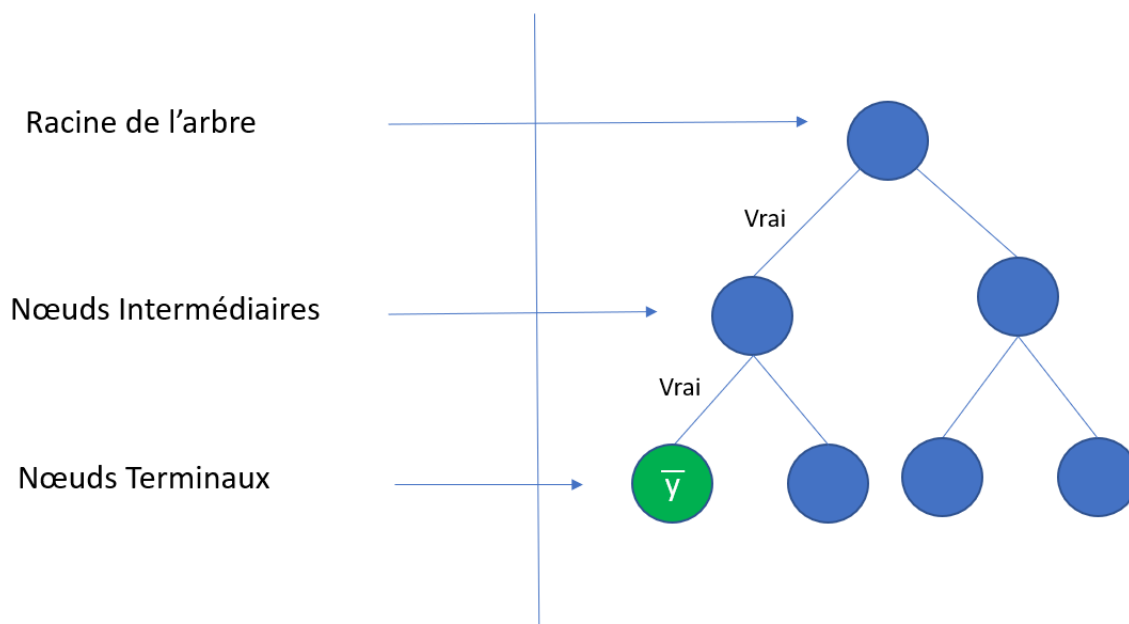
GRAPHIQUE A.13 – Illustration de l'algorithme Forward



A.2.3 Illustration des algorithmes de Machine Learning

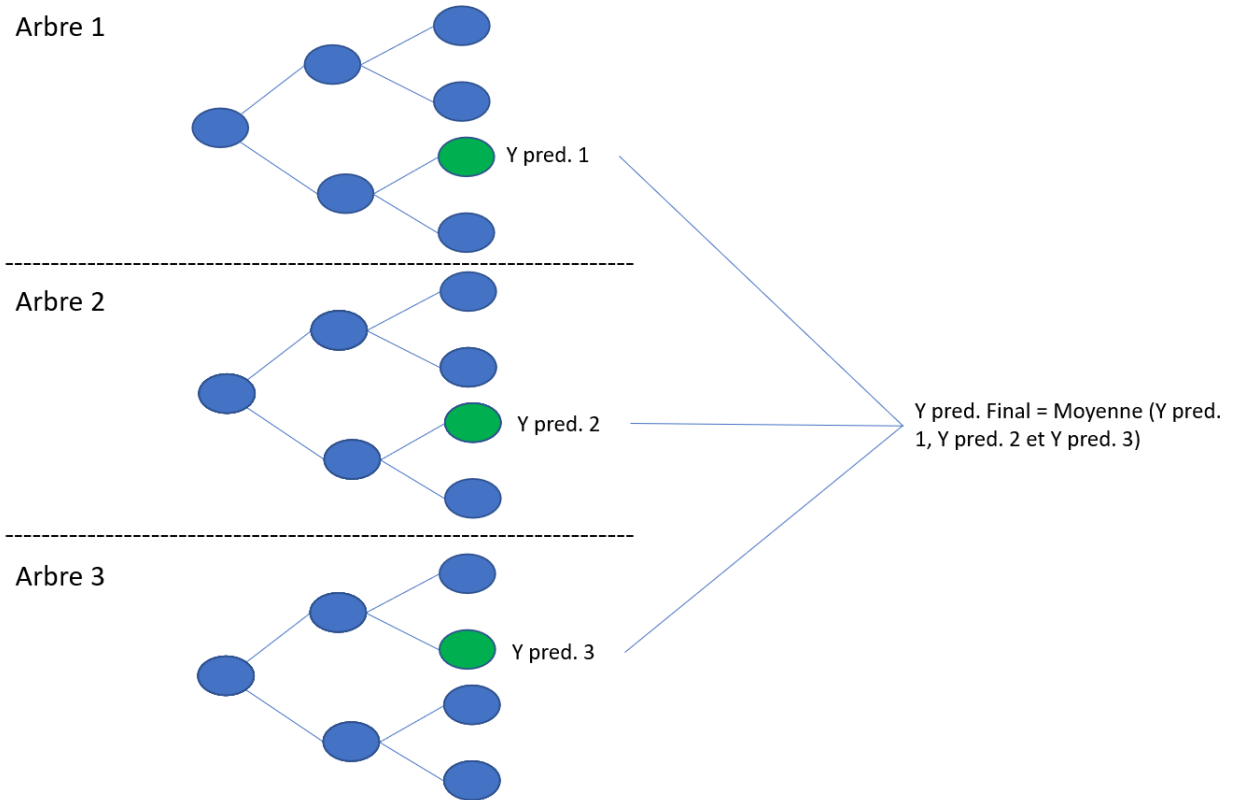
1. Arbre de régression

GRAPHIQUE A.14 – Illustration d'un arbre de régression



2. Random Forest

GRAPHIQUE A.15 – Illustration de l’algorithme Random Forest



A.2.4 Test de différence de moyennes utilisé dans l’étude

Dans les comparaisons d’échantillon effectuées dans ce document, nous avons appliqué le test de différence de moyenne bilatérale de Student. Ce test suppose la normalité de la variable de comparaison dans les deux échantillons. Etant donné deux échantillons indicés 1 et 2 et X la variable de comparaison, le test se spécifie comme suit :

- Sous $H_0 : \mu_{X_1} = \mu_{X_2}$;
- Sous $H_1 : \mu_{X_1} \neq \mu_{X_2}$.

La statistique du test est donnée par : $T = \frac{\mu_{X_1} - \mu_{X_2}}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})}}$ qui suit une loi de Student $t_{n_1+n_2-2}$ avec $s^2 =$

$\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}$. L’hypothèse H_0 est rejeté pour un seuil d’erreur $\alpha = 5\%$ si $|T| > t_{1-\alpha/2, n_1+n_2-2}^a$.

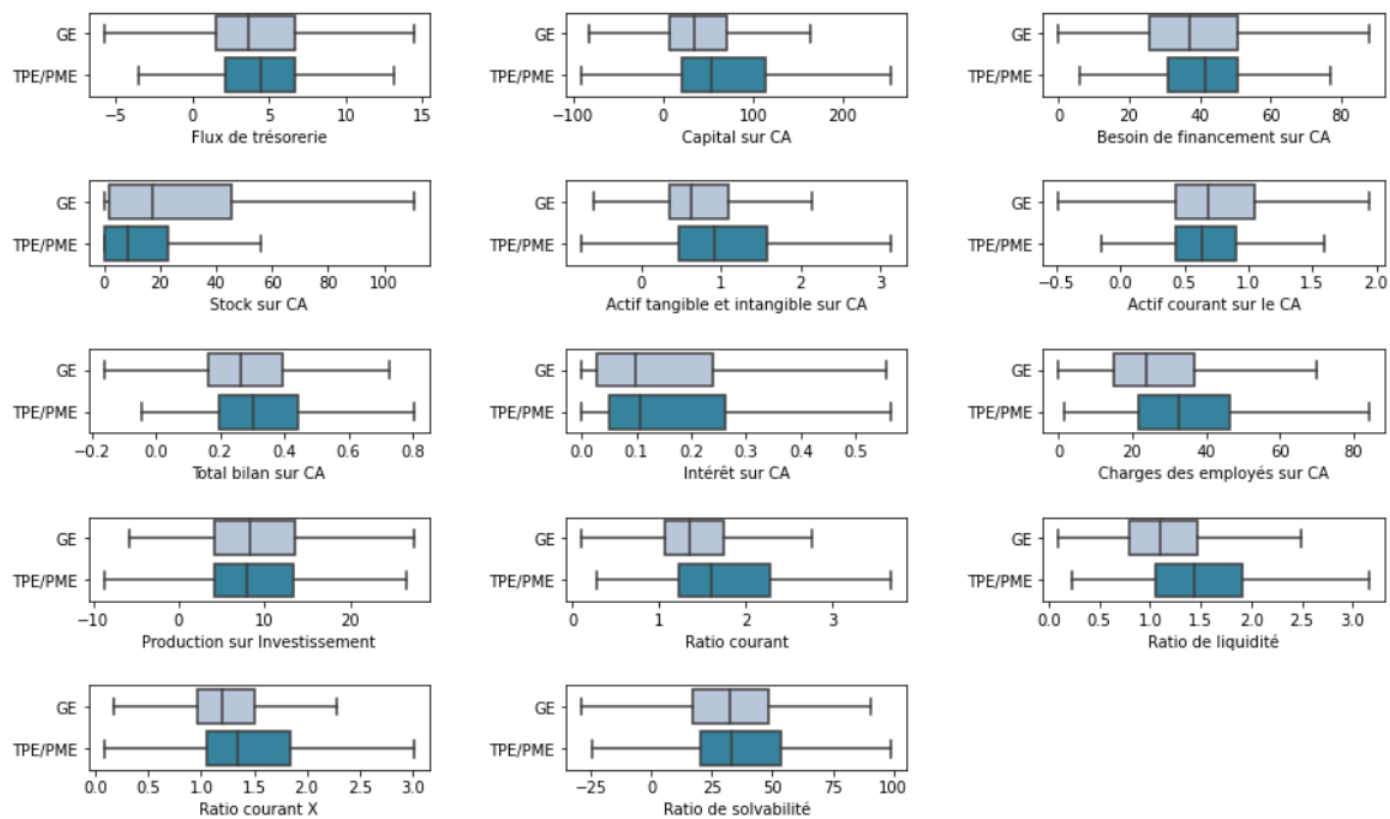
^a. Quantile d’ordre $1 - \alpha/2$ d’une loi de Student

A.3 Choix de l’échantillon de travail

La comparaison des deux catégories d’entreprises à travers les informations financières rapportées au chiffre d’affaire (CA) et au nombre d’employés n’a pas abouti à une conclusion d’homogénéité. Pour 7

ratios sur 10 ces deux sous population sont distinctes. Donc dans la suite de l'étude nous conserverons l'échantillon des TPE/PME pour la modélisation.

GRAPHIQUE A.16 – Comparaison des GE et TPE



Source : Base ASDES, calcul des auteurs.

TABLEAU A.2 – Résultat du test de différence de moyennes

| Ratios | f_value | p_value |
|-------------------------------------|----------------|----------------|
| Profit par employé | -0,388115 | 6,979981e-01 |
| Fonds des actionnaires par employé | 0,476716 | 6,336492e-01 |
| Intérêt sur CA | 0,716847 | 4,736049e-01 |
| Production sur Investissement | -0,872264 | 3,832348e-01 |
| Flux de trésorerie sur CA | 1,476304 | 1,401190e-01 |
| Besoin de financement sur cA | 1,704209 | 8,859518e-02 |
| ratio de solvabilité | 2,123506 | 3,391114e-02 |
| Capital par employé | 2,480559 | 1,325092e-02 |
| Actif courant sur le CA | -2,601141 | 9,403176e-03 |
| Chiffre d’affaire par employé | -2,965415 | 3,081186e-03 |
| Total bilan sur CA | 3,063068 | 2,238524e-03 |
| Total bilan par employé | 3,440070 | 6,010955e-04 |
| Stock sur CA | -5,928887 | 3,959580e-09 |
| Actif tangible et intangible sur CA | 5,947347 | 3,549178e-09 |
| Ratio courant X | 6,167345 | 9,411012e-10 |
| Charge des employés sur CA | 6,287013 | 4,489653e-10 |
| Ratio courant | 8,065714 | 1,718269e-15 |
| Capital sur CA | 8,133248 | 1,014515e-15 |
| Ratio de liquidité | 9,543650 | 7,155777e-21 |

Source : Bases ASDES, calculs des auteurs

A.3.1 Choix des variables pour la prédiction

Une première sélection des explicatives quantitatives

Le tableau ci-dessous présente les variables significativement corrélées avec la variable cible. Ici le score F value est obtenu en régressant la variable cible sur chaque variable et en calculant la statistique de Fisher (F value).

TABLEAU A.3 – Liste des variables significativement corrélées aux émissions

| variables | f_value | r_value | p_value_f |
|---|---------|---------|-----------|
| Nombre d'employés | 101,939 | 0,580 | 0,000 |
| Durée d'activité | 38,116 | 0,399 | 0,000 |
| Nombre d'établissements | 17,001 | 0,279 | 0,000 |
| Achat de matière première | 14,330 | 0,258 | 0,000 |
| Vente de produits | 11,735 | 0,235 | 0,001 |
| Charge des employés | 10,761 | -0,225 | 0,001 |
| Dépréciation | 10,112 | 0,219 | 0,002 |
| Actifs fixe net | 9,792 | 0,216 | 0,002 |
| Groupe_Entreprise par niveau | 9,620 | 0,214 | 0,002 |
| Banque | 9,241 | 0,210 | 0,003 |
| Part dans la capital | 8,992 | 0,207 | 0,003 |
| Total : total brut | 7,986 | 0,195 | 0,005 |
| Actif tangible et intangible | 7,724 | 0,192 | 0,006 |
| Ratio de liquidité | 7,650 | -0,191 | 0,006 |
| Compte net recevable | 7,522 | 0,190 | 0,007 |
| Total provision en début de période | 7,231 | 0,186 | 0,008 |
| Total passif et part du capital | 7,193 | 0,186 | 0,008 |
| Total Bilan | 7,193 | 0,186 | 0,008 |
| Fonds des actionnaires | 7,052 | 0,184 | 0,009 |
| Total fonds actionnaires I | 7,052 | 0,184 | 0,009 |
| Ratio stock sur chiffre d'affaire | 7,035 | 0,184 | 0,009 |
| Total charge | 6,850 | 0,182 | 0,010 |
| Valeur ajoutée | 6,765 | 0,180 | 0,010 |
| Passifs | 6,735 | 0,180 | 0,010 |
| Stock Net | 6,320 | 0,175 | 0,013 |
| EBITDA à la marge | 6,288 | 0,174 | 0,013 |
| Flux de trésorerie opérationnel | 6,212 | 0,173 | 0,014 |
| Ventes nettes | 5,923 | 0,169 | 0,016 |
| Chiffre d'affaire net | 5,923 | 0,169 | 0,016 |
| Investissements | 5,873 | 0,168 | 0,016 |
| Total revenu de fonctionnement | 5,727 | 0,166 | 0,018 |
| Total dépenses de fonctionnement | 5,665 | 0,166 | 0,018 |
| Ratio capital sur CA | 5,590 | -0,164 | 0,019 |
| Dettes financière | 5,557 | 0,164 | 0,019 |
| Actif courant net | 5,288 | 0,160 | 0,023 |
| Total revenu | 5,118 | 0,158 | 0,025 |
| Vente de service | 5,112 | 0,157 | 0,025 |
| Capital employé sur actifs fixes | 5,107 | -0,157 | 0,025 |
| Achat de matière première et autre | 4,956 | 0,155 | 0,027 |
| Nombre d'établissements secondaires | 4,950 | 0,155 | 0,027 |
| Chiffre d'affaire net lié à l'exportation | 4,948 | 0,155 | 0,027 |
| Total provision en fin de période | 4,592 | 0,149 | 0,033 |
| Perte de profit | 4,509 | 0,148 | 0,035 |
| Salaires | 4,436 | 0,147 | 0,036 |
| Chiffre d'affaire par employé | 4,157 | 0,142 | 0,043 |
| Chiffre d'affaire par employé en millier | 4,157 | 0,142 | 0,043 |
| Perte de profit net | 4,059 | -0,141 | 0,045 |
| Reserves | 3,984 | 0,139 | 0,047 |
| Stock Annuel | 3,953 | -0,139 | 0,048 |
| Stock courant | 3,953 | -0,139 | 0,048 |

Source : Bases ASDES, calculs des auteurs

A.4 Sélection des explicatives qualitatives

Les variables qualitatives présentes sont toutes significativement liées aux émissions.

TABLEAU A.4 – Test de liaison entre les explicatives qualitatives et la variable dépendante

| variables | f_value | p_value |
|-------------------|-----------|--------------|
| Pays | 62.479120 | 1.745365e-13 |
| Portée du marché | 10.254488 | 5.759336e-05 |
| Type d'entreprise | 6.218454 | 2.230242e-05 |
| Filière | 10.486340 | 6.140934e-09 |

Source : Base ASDES, calcul des auteurs.

Cette liste de variables sélectionnées sera utilisée pour le modèle Elastic net et les modèles de Machine Learning. Dans chaque modèle nous appliquons une grille de recherche qui permet au modèle de choisir les meilleurs prédicteurs parmi la liste ci-dessus. De plus les modèles de Machine Learning utilisés ici font appel aux arbres de régression qui à chaque étape cherche la meilleure variable pour les subdivisions.

Pour le cas particulier de la régression linéaire nous avons appliqué trois autres filtres afin d'avoir un modèle utile au sens métier et éviter la multicollinéarité. Ces filtres sont d'abord la méthode pas à pas "forward" pour minimiser le R^2 ajusté, puis un deuxième filtre forward pour minimiser l'AIC. Cette double sélection est résumé ci-dessous.

TABLEAU A.5 – Variables issues de la sélection Forward selon les critères R^2 ajusté puis l'AIC

| Critère R^2 ajusté | Critère AIC |
|--|-----------------------------------|
| Nombre d'employés | Nombre d'employés |
| Durée d'activité | Durée d'activité |
| Achat de matière première | Achat de matière première |
| Nombre d'établissements | Nombre d'établissements |
| Salaires | Salaires |
| Ratio stock sur chiffre d'affaire | Ratio stock sur chiffre d'affaire |
| Fonds des actionnaires | Fonds des actionnaires |
| Valeur ajoutée | Valeur ajoutée |
| Actif tangible et intangible | Actif tangible et intangible |
| Vente de service | Vente de service |
| Total charge | Total charge |
| EBITDA à la marge | EBITDA à la marge |
| Flux de trésorerie opérationnel | Flux de trésorerie opérationnel |
| Total revenu | Total revenu |
| Total Bilan | Total Bilan |
| Actif courant net | Actif courant net |
| Dette financière | |
| Actifs fixe net | |
| Ratio de liquidité | |
| Ratio capital sur CA | |
| Part dans la capital | |
| Total : total brut | |
| Dépréciation | |
| Investissements | |
| Total dépenses de fonctionnement | |
| Total revenu de fonctionnement | |

Source : Bases ASDES, calculs des auteurs

TABLEAU A.6 – Une dernière filtre des variables pour la régression linéaire

| Intensité de l'activité | Flux financier |
|------------------------------------|-----------------------|
| Nombre d'employées | EBITDA à la marge |
| Duree d'Activité | Total du bilan |
| Nombre d'établissement | Salaires |
| Vente de service | |
| Achat de matière première | |
| Ratio Stocks sur Chiffre d'affaire | |

Source : Base ASDES, calcul des auteurs.

A.5 Quelques statistiques descriptives

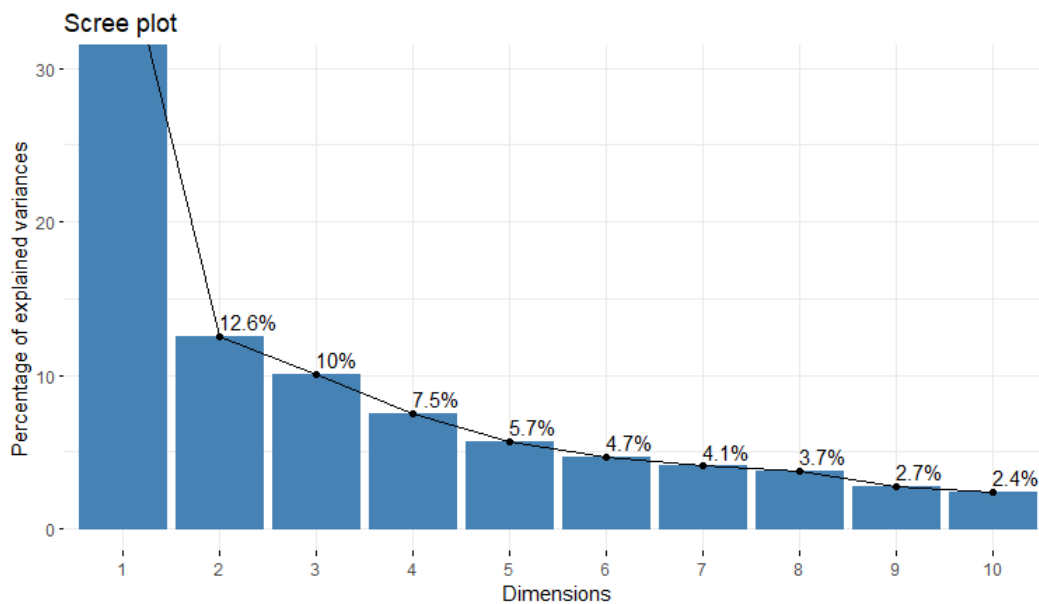
A.5.1 Distribution des émissions par catégorie

TABLEAU A.7 – Description des émissions par catégorie

| | TPE/PME | GE | Globale |
|-------|----------------|-----------|----------------|
| count | 203 | 1026 | 1229 |
| mean | 1289,069 | 5036,419 | 4417,451 |
| std | 1818,533 | 5768,520 | 5500,632 |
| min | 0,04 | 0,020 | 0,020 |
| 25% | 40,689 | 1064,629 | 676,066 |
| 50% | 483,874 | 2935 | 2274 |
| 75% | 1861,211 | 6821,277 | 5845,714 |
| max | 10546,622 | 28966 | 28966 |

Source : Base ASDES, calcul des auteurs.

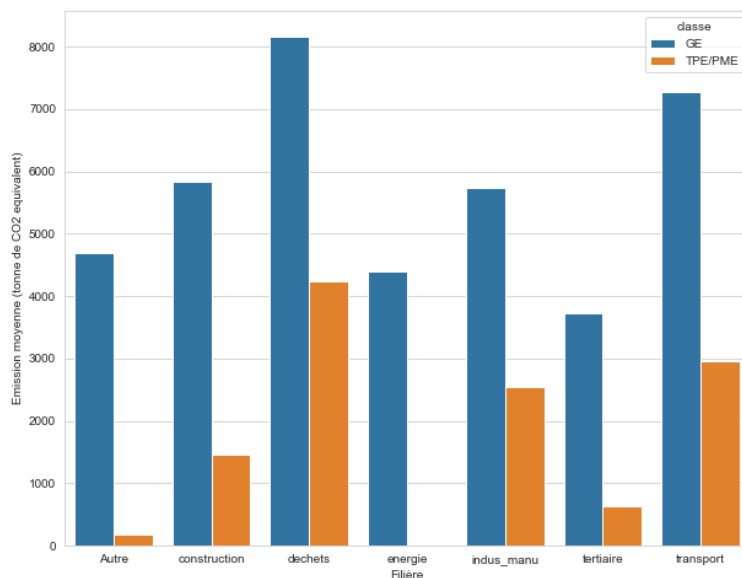
GRAPHIQUE A.17 – Histogramme des valeurs propres



Source : Base ASDES, calcul des auteurs.

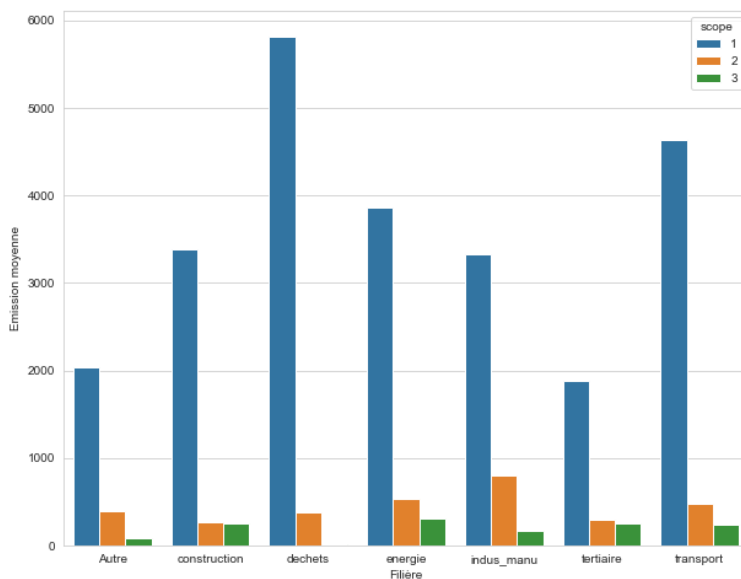
A.5.2 Analyse sectorielle

GRAPHIQUE A.18 – Positionnement des TPE/PME et GE par filière en termes d’émission de GES



Source : Base ASDES, calcul des auteurs.

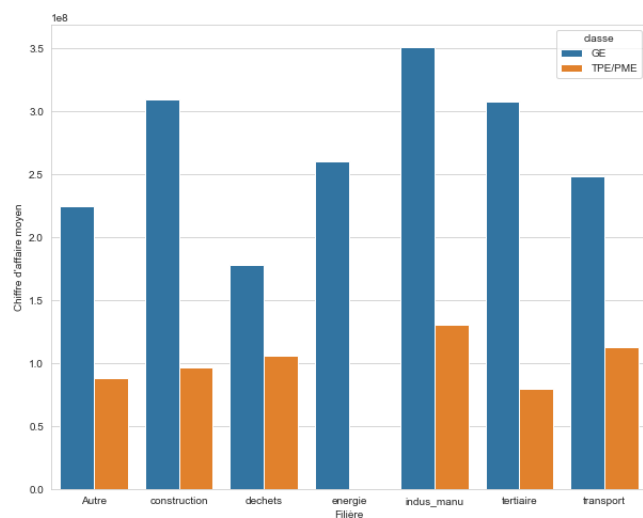
GRAPHIQUE A.19 – Emission par scope et par filière



Source : Base ASDES, calcul des auteurs.

A.5.3 Distribution selon le chiffre d'affaire

GRAPHIQUE A.20 – Répartition du chiffre d'affaire



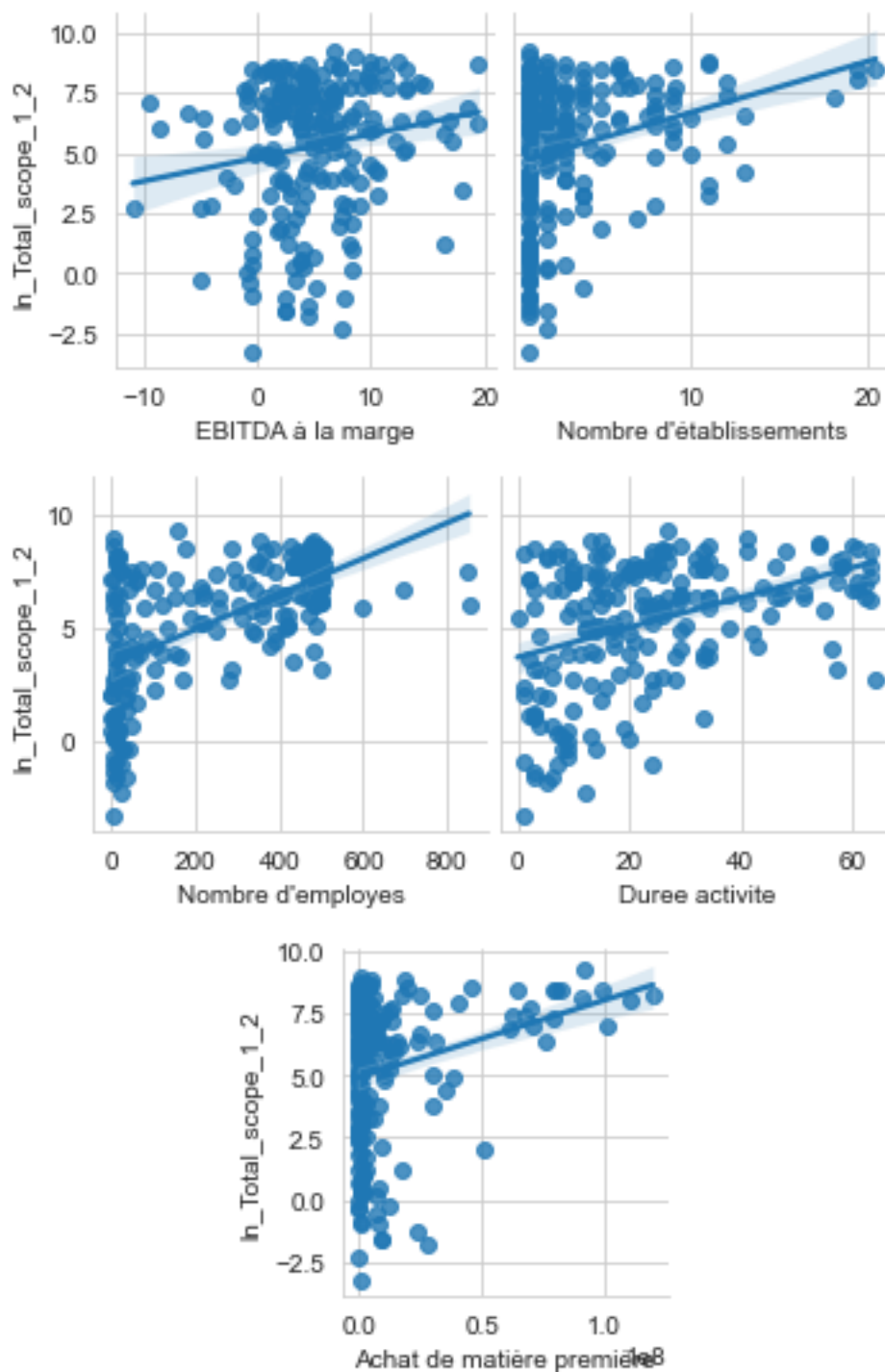
Source : Données ASDES, calculs des auteurs

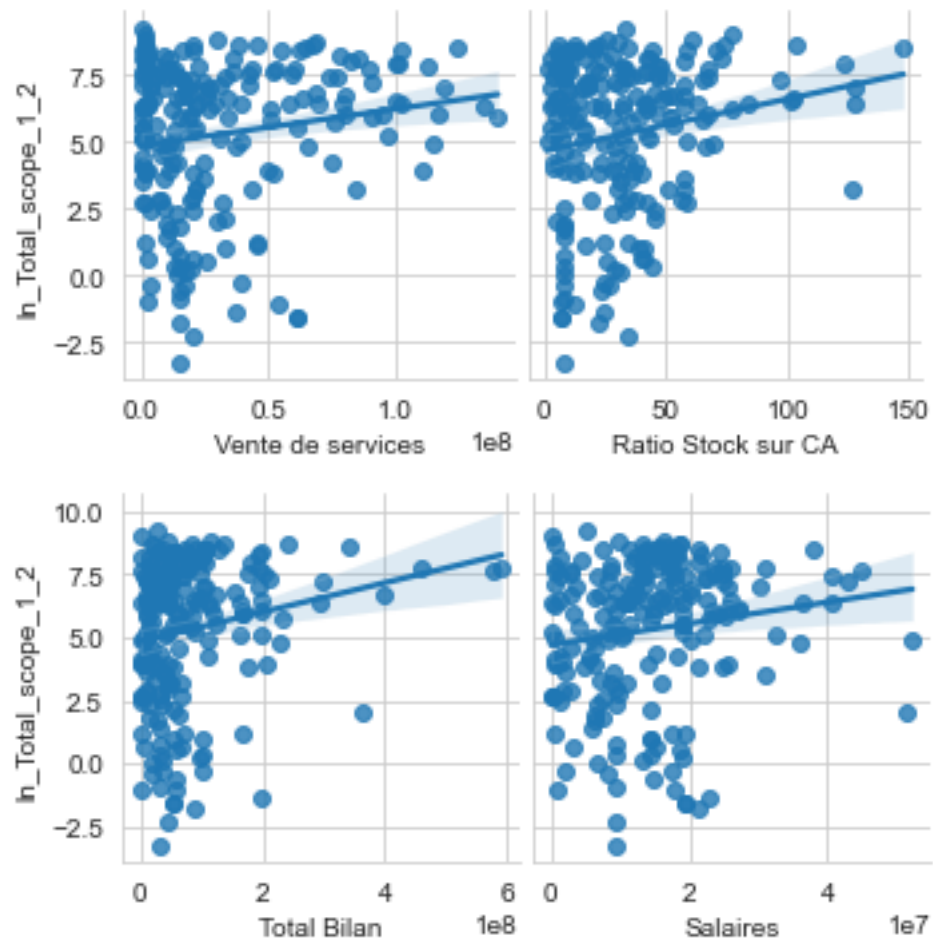
A.6 Résultats de la modélisation classique

A.6.1 Régression linéaire

Signes attendus sur les explicatives quantitatives

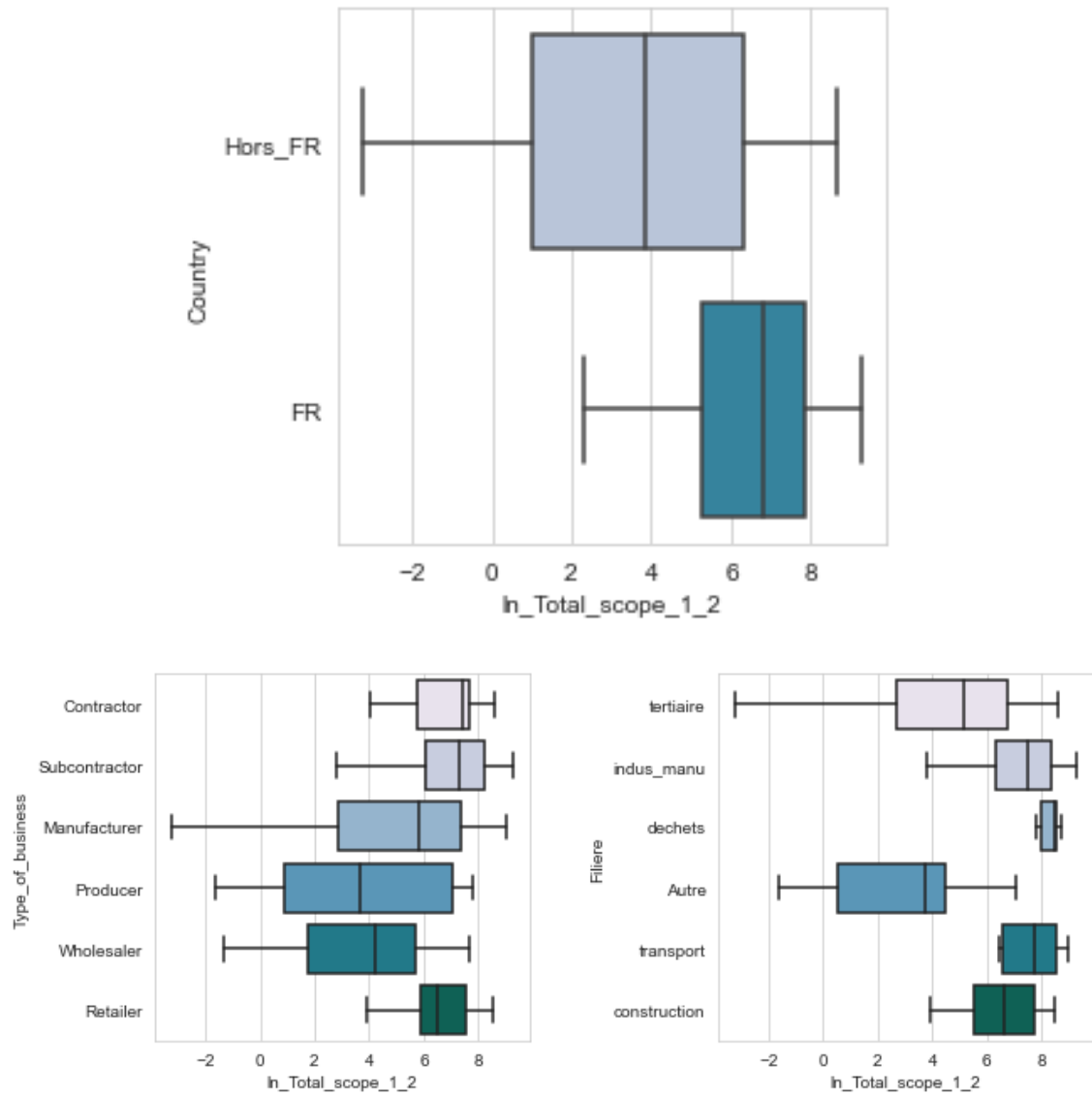
GRAPHIQUE A.21 – Nuage de point entre variables cibles et les variables quantitatives les plus corrélées





Source : Base ASDES, calcul des auteurs.

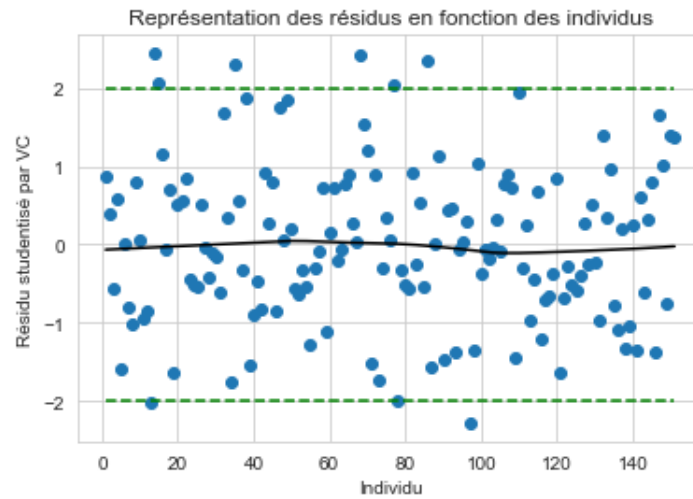
Relation avec les qualitatives



GRAPHIQUE A.22 – Distribution de la variable cible dans les différentes classes des variables qualitatives
Source : Base ASDES, calcul des auteurs.

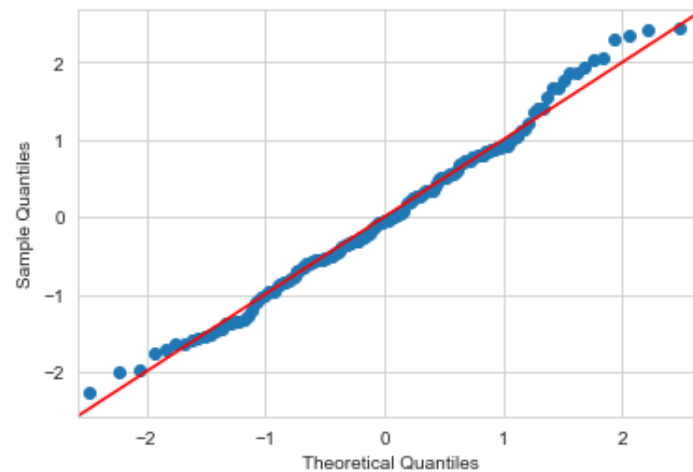
Résidu studentisé

GRAPHIQUE A.23 – Résidus Studentisé vs Index



Source : Base ASDES, calcul des auteurs.

Vérification de la Normalité

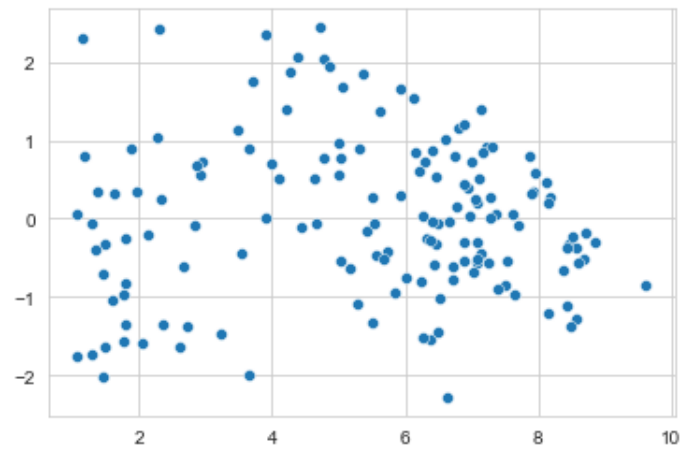


GRAPHIQUE A.24 – Résidu vs Quantile loi normale centrée réduite

Source : Base ASDES, calcul des auteurs.

Hétéroscédasticité

GRAPHIQUE A.25 – Résidus Studentisé vs Valeurs ajustées

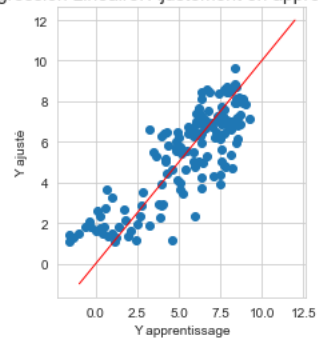


Source : Base ASDES, calcul des auteurs.

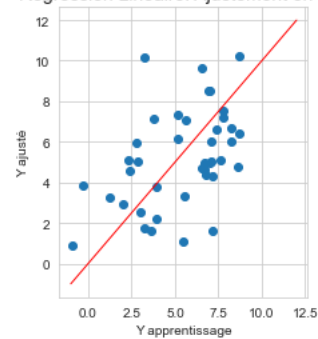
Qualité d'ajustement

GRAPHIQUE A.26 – Régression Linéaire : Qualité d'ajustement et courbe d'apprentissage

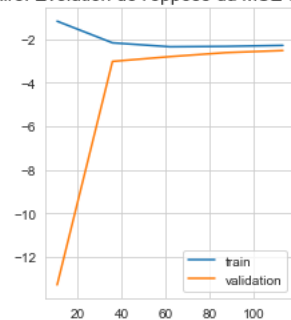
Regression Linéaire: Ajustement en apprentissage



Regression Linéaire: Ajustement en test



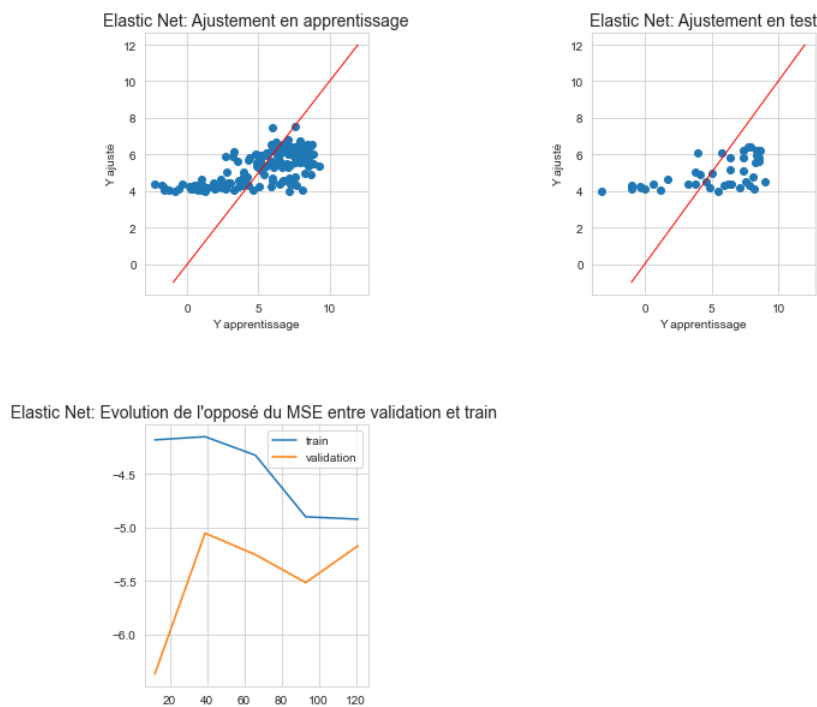
Regression Linéaire: Evolution de l'opposé du MSE entre validation et train



Source : Base ASDES, calcul des auteurs.

A.6.2 Elastic Net : qualité d'ajustement

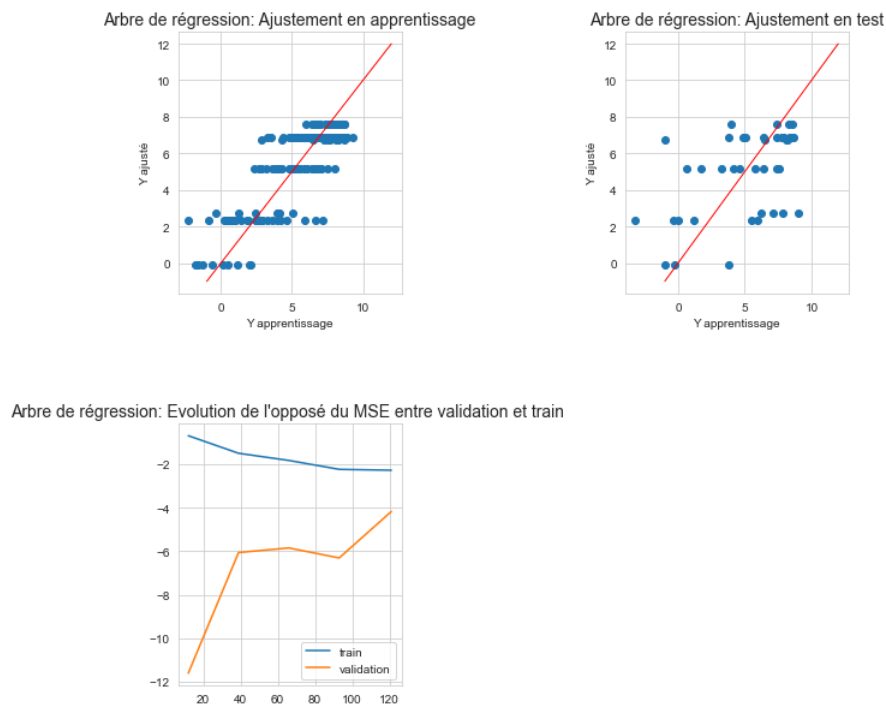
GRAPHIQUE A.27 – Elastic Net : Qualité d'ajustement et courbe d'apprentissage



Source : Base ASDES, calcul des auteurs.

A.6.3 Arbre de régression

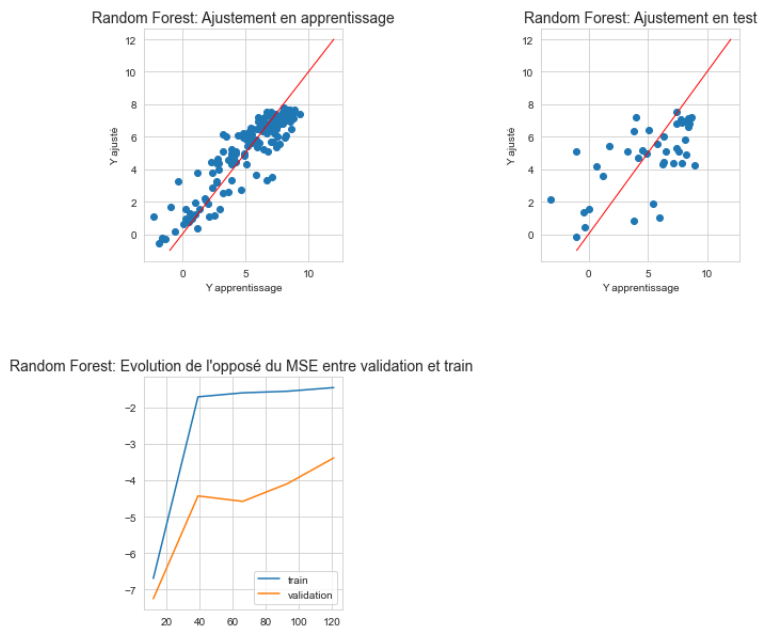
GRAPHIQUE A.28 – Arbre de régression : Qualité d'ajustement et courbe d'apprentissage



Source : Base ASDES, calcul des auteurs.

A.6.4 Random Forest

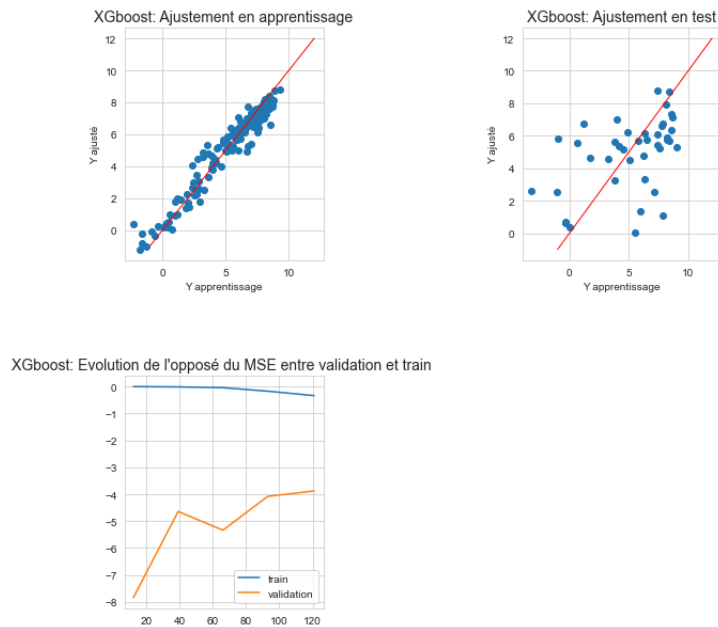
GRAPHIQUE A.29 – Random Forest : Qualité d'ajustement et courbe d'apprentissage



Source : Base ASDES, calcul des auteurs.

A.6.5 XG Boost

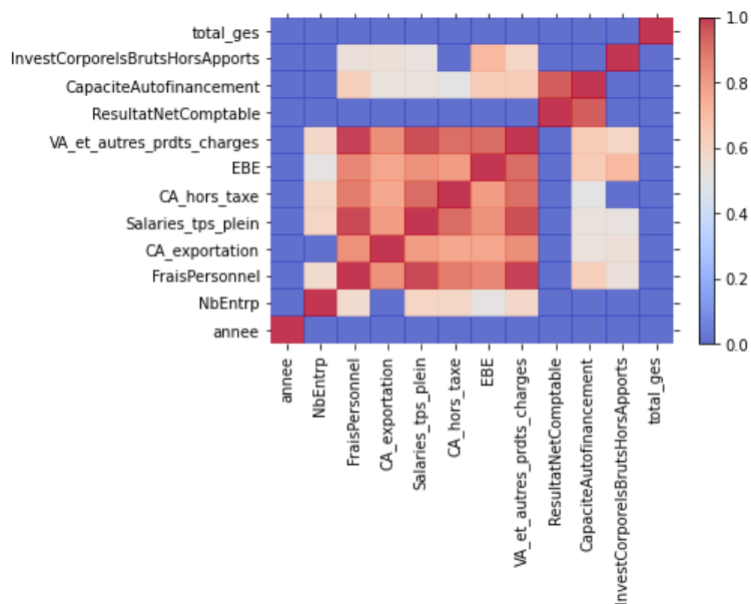
GRAPHIQUE A.30 – XG Boost : Qualité d'ajustement et courbe d'apprentissage



Source : Base ASDES, calcul des auteurs.

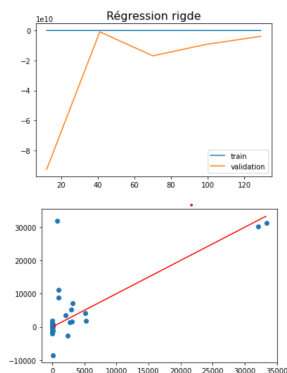
A.7 Résultats des modèles filière

GRAPHIQUE A.31 – Matrice de corrélation- variables filière

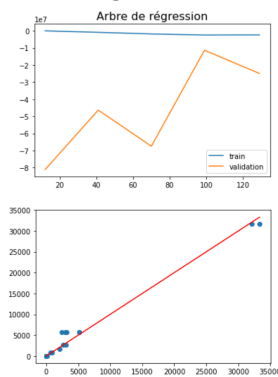


Source : Bases Citepa & Ésane, calcul des auteurs.

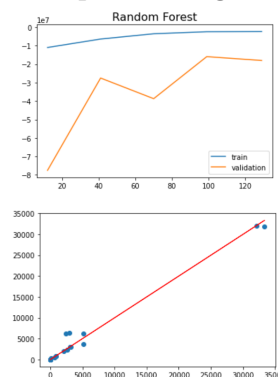
GRAPHIQUE A.32 – Régression rigde



GRAPHIQUE A.33 – Arbre de régression



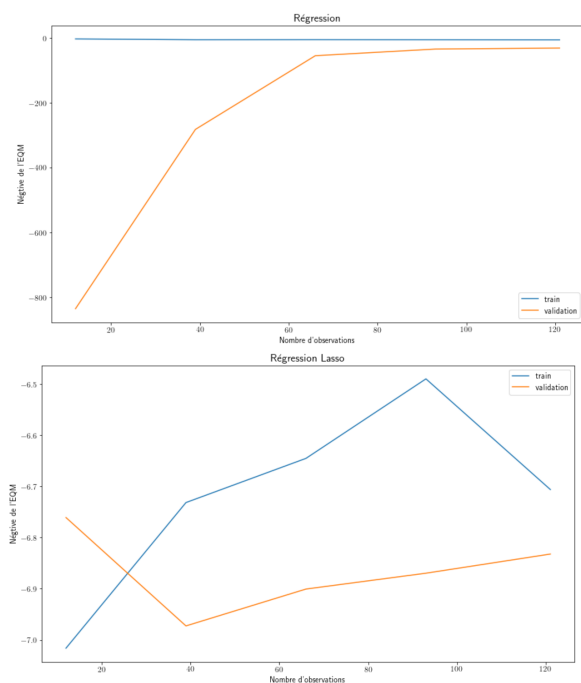
GRAPHIQUE A.34 – Random Forest pour la régression



Source : Base ASDES, calcul des auteurs.

A.8 Modèles ratio

GRAPHIQUE A.35 – Courbes d'apprentissage modèles ratio 2



GRAPHIQUE A.36 – Courbes d'apprentissage modèles ratio 2

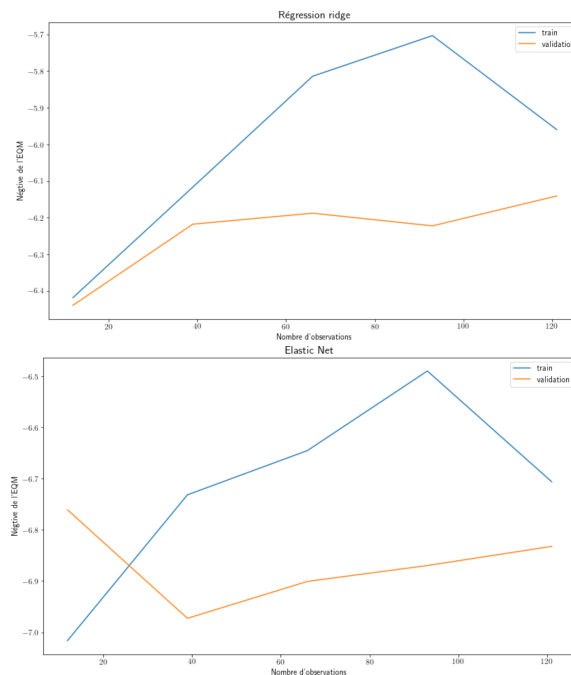


TABLEAU A.8 – Variables explicatives sélectionnées pour les modèles ratio

| Variables | Coefs associés au meilleur modèle (ridge) |
|--|---|
| Ratio Chiffre d'affaires | -0,15 |
| Ratio Nombre d'employés | -0,18 |
| Ratio Chiffre d'affaires à l'exportation | -0,10 |
| Appartenance à la filière Construction | -12,00 |
| Appartenance à la filière Déchets centralisés | -9,36 |
| Appartenance à la filière Industrie manufacturière | -14,47 |

TABLEAU A.9 – Matrice de corrélation- variables ratios

| | reporting_year | NbEntrp_secteur | CA_rate | VA_rate | CA_export_rate | EBE_rate | Bilan_rate | VAT_rate | NbEmploye_rate | Salaire_rate | invest_rate | ges_rate |
|-----------------|----------------|-----------------|---------|---------|----------------|----------|------------|----------|----------------|--------------|-------------|----------|
| reporting_year | 1,000 | 0,073 | -0,005 | 0,046 | 0,109 | 0,032 | 0,024 | -0,050 | -0,253 | 0,137 | -0,015 | -0,367 |
| NbEntrp_secteur | 0,073 | 1,000 | -0,034 | -0,124 | 0,079 | 0,037 | -0,523 | -0,369 | -0,118 | -0,121 | -0,142 | 0,132 |
| CA_rate | -0,005 | -0,034 | 1,000 | 0,860 | 0,238 | 0,901 | 0,542 | 0,570 | 0,427 | 0,866 | 0,175 | 0,191 |
| VA_rate | 0,046 | -0,124 | 0,860 | 1,000 | 0,159 | 0,791 | 0,592 | 0,582 | 0,476 | 0,828 | 0,179 | 0,128 |
| CA_export_rate | 0,109 | 0,079 | 0,238 | 0,159 | 1,000 | 0,293 | 0,139 | 0,160 | -0,187 | 0,209 | 0,078 | -0,216 |
| EBE_rate | 0,032 | 0,037 | 0,901 | 0,791 | 0,293 | 1,000 | 0,529 | 0,508 | 0,329 | 0,784 | 0,188 | 0,124 |
| Bilan_rate | 0,024 | -0,523 | 0,542 | 0,592 | 0,139 | 0,529 | 1,000 | 0,877 | 0,203 | 0,513 | 0,241 | -0,036 |
| VAT_rate | -0,050 | -0,369 | 0,570 | 0,582 | 0,160 | 0,508 | 0,877 | 1,000 | 0,243 | 0,482 | 0,308 | 0,026 |
| NbEmploye_rate | -0,253 | -0,118 | 0,427 | 0,476 | -0,187 | 0,329 | 0,203 | 0,243 | 1,000 | 0,432 | 0,053 | 0,539 |
| Salaire_rate | 0,137 | -0,121 | 0,866 | 0,828 | 0,209 | 0,784 | 0,513 | 0,482 | 0,432 | 1,000 | 0,116 | 0,129 |
| invest_rate | -0,015 | -0,142 | 0,175 | 0,179 | 0,078 | 0,188 | 0,241 | 0,308 | 0,053 | 0,116 | 1,000 | -0,033 |
| ges_rate | -0,367 | 0,132 | 0,191 | 0,128 | -0,216 | 0,124 | -0,036 | 0,026 | 0,539 | 0,129 | -0,033 | 1,000 |

Source : Base ASDES, calcul des auteurs.

TABLE DES MATIÈRES

| | |
|--|------------|
| Remerciements | I |
| Abstract | II |
| Sommaire | III |
| Sigles et abréviations | IV |
| Liste des graphiques | IV |
| Liste des tableaux | VI |
| Introduction | 1 |
| 1 Cadre conceptuel et revue de la littérature | 3 |
| 1.1 Définitions des concepts | 3 |
| 1.1.1 TPE/PME | 3 |
| 1.1.2 Grande Entreprise (GE) | 3 |
| 1.1.3 Risque climatique (BIS, 2021)[3] | 3 |
| 1.1.4 Gaz à effet de serre (GES) | 4 |
| 1.1.5 Vision par scope des émissions de GES : (Ademe, 2014)[2] | 4 |
| 1.2 Revue sur l'évaluation de risque de transition | 5 |
| 1.2.1 Principaux indicateurs de mesure d'exposition | 5 |
| 1.2.2 Méthodes de quantification du risque de transition | 7 |
| 1.2.3 Approches Bottom-up et Top-Down | 8 |

| | | |
|----------|---|-----------|
| 2 | Cadre méthodologique | 9 |
| 2.1 | Présentation des données et pré-traitement | 9 |
| 2.1.1 | Sources de données | 9 |
| 2.1.2 | Différents pré-traitements | 10 |
| 2.2 | Choix d'échantillon de travail : TPE/PME sans ou avec les GE | 12 |
| 2.3 | Construction de la variable cible : les émissions de GES des scopes 1 et 2, un proxy du risque de transition | 12 |
| 2.4 | Méthodologie de modélisation | 12 |
| 2.4.1 | Modèles d'apprentissage supervisé | 12 |
| 2.4.2 | Approche top-down | 15 |
| 3 | Evaluation du risque de transition | 17 |
| 3.1 | Répartition de l'échantillon par catégorie d'entreprises et analyse des émissions | 17 |
| 3.1.1 | Un échantillon dominé par les Grandes entreprises | 17 |
| 3.1.2 | Distribution des émissions de GES selon la catégorie d'entreprises | 18 |
| 3.1.3 | Emissions des TPE/PME selon le département | 19 |
| 3.1.4 | Analyse de liaison entre les variables | 20 |
| 3.1.5 | Le transport, secteur très pollueur | 21 |
| 3.2 | Résultats des modèles classiques | 21 |
| 3.2.1 | Transformation de la variable cible | 21 |
| 3.2.2 | Choix du meilleur modèle | 23 |
| 3.2.3 | Résultat de la régression linéaire | 24 |
| 3.3 | Résultats des modèles Top-Down | 25 |
| 3.3.1 | Modèle ratio | 25 |
| 3.3.2 | Modèles filière | 26 |
| 3.4 | Comparaison des modèles | 27 |
| 3.5 | Discussions des résultats | 28 |
| | Conclusion | 29 |
| | Bibliographie | i |
| A | Annexe | ii |
| A.1 | Rapprochement des secteurs Citepa aux filières APE | ii |
| A.2 | Complément méthodologique | iii |
| A.2.1 | Procédure de traitement des données | iii |
| A.2.2 | Sélection pas à pas : direction Forward | iv |
| A.2.3 | Illustration des algorithmes de Machine Learning | v |

| | | |
|-------|---|-------|
| A.2.4 | Test de différence de moyennes utilisé dans l'étude | vi |
| A.3 | Choix de l'échantillon de travail | vi |
| A.3.1 | Choix des variables pour la prédiction | viii |
| A.4 | Sélection des explicatives qualitatives | x |
| A.5 | Quelques statistiques descriptives | xii |
| A.5.1 | Distribution des émissions par catégorie | xii |
| A.5.2 | Analyse sectorielle | xiii |
| A.5.3 | Distribution selon le chiffre d'affaire | xiv |
| A.6 | Résultats de la modélisation classique | xv |
| A.6.1 | Régression linéaire | xv |
| A.6.2 | Elastic Net : qualité d'ajustement | xx |
| A.6.3 | Arbre de régression | xx |
| A.6.4 | Random Forest | xxi |
| A.6.5 | XG Boost | xxi |
| A.7 | Résultats des modèles filière | xxii |
| A.8 | Modèles ratio | xxiii |

Tables de matières

xxiii