



Institut des
Politiques Publiques

ANNEXE METHODOLOGIQUE – NOTE IPP N° 102 - Août 2024

Les émissions de CO₂ de l'industrie française et le « ciblage carbone » des politiques publiques

Laurent Bach

Paul Dutronc-Postel

Arthur Guillouzouic

Clément Malgouyres

Rachel Paya

A large, stylized version of the IPP logo is positioned in the bottom right corner of the page. It consists of the lowercase letters 'ipp' in a bold, sans-serif font, centered between two curved lines that form a partial circle above and below the text. The logo is rendered in a dark teal color against the teal background of the page.

ipp



L'Institut des politiques publiques (IPP) a été créé par PSE et est développé dans le cadre d'un partenariat scientifique entre PSE-École d'Économie de Paris et le Groupe des écoles nationales d'économie et de statistique (GENES). L'IPP vise à promouvoir l'analyse et l'évaluation quantitatives des politiques publiques en s'appuyant sur les méthodes les plus récentes de la recherche en économie.

www.ipp.eu



ANNEXE METHODOLOGIQUE – NOTE IPP N° 102 -
Août 2024

Les émissions de CO₂ de l'industrie française et le « ciblage carbone » des politiques publiques

Laurent Bach
Paul Dutronc-Postel
Arthur Guillouzouic
Clément Malgouyres
Rachel Paya

SOMMAIRE

Introduction	9
1 Estimation des émissions carbonées des entreprises industrielles	11
1.1 Données utilisées	11
1.1.1 Consommations d'énergie	11
1.1.2 Caractéristiques des établissements	12
1.1.3 Facteurs d'émissions	12
1.2 Méthodologie détaillée	15
1.2.1 Champ de l'estimation	15
1.2.2 Calcul des émissions des établissements à partir des consommations d'énergie observées	17
1.2.3 Imputations des émissions de CO ₂ des établissements non observés dans les enquêtes	17
1.2.4 Agrégation des émissions de CO ₂ au niveau entreprise	19
1.2.5 Imputation des émissions de procédés	20
1.3 Sélection du modèle de prédiction	21
1.3.1 Régressions linéaires	21
1.3.2 Régressions LASSO	22
1.3.3 Forêts aléatoires	22
1.3.4 Gradient boosting	23
1.3.5 Comparaison des modèles	24
1.4 Résultats	25
1.4.1 Estimations	25
1.4.2 Contrôle de cohérence	26
2 Micro-simulation de la tarification effective du carbone	31
2.1 Les taxes intérieures de consommation	31
2.2 Méthodologie détaillée	31
2.2.1 Estimation des consommations d'énergie	32
2.2.2 Identification des exonérations	33
2.2.3 Calcul du coût carbone effectif pour les entreprises industrielles	36

Liste des tableaux	37
Liste des figures	39

INTRODUCTION

La [note IPP n°102](#), réalisée dans le cadre du projet GreenFirms, financé par l'Ademe, documente les émissions carbone des entreprises industrielles françaises et évalue le « ciblage carbone » des politiques publiques. Afin de garantir la plus large couverture possible, les résultats reposent sur une estimation des émissions de CO₂ des entreprises à partir des données de consommation d'énergie. En effet, les données existantes d'émissions carbone des entreprises sont à ce jour principalement issues d'auto-déclarations et ne couvrent que très partiellement le tissu productif. Au vu de l'intérêt grandissant des acteurs publics pour une évaluation systématique de l'impact environnemental des politiques publiques, une meilleure couverture de ces émissions apparaît nécessaire. Ces données sont également essentielles pour estimer au mieux la tarification effective du carbone, qui constitue le principal outil de politique publique pour décarboner l'économie. La stratégie adoptée pour la construction d'une base de données sur les émissions carbone des entreprises industrielles françaises repose en majeure partie sur l'exploitation des données de l'Enquête Annuelle sur les Consommations d'Énergie dans l'Industrie (EACEI) réalisée par l'Insee. Cette enquête recense les consommations des différentes énergies d'établissements industriels français. L'objectif étant d'obtenir une mesure des émissions carbone au niveau de l'entreprise, un certain nombre d'enjeux méthodologiques se posent, liés à l'agrégation des données établissement au niveau entreprise. Le chapitre 1 résume ces enjeux et présente les solutions choisies ainsi que les résultats obtenus. Le chapitre 2 détaille la méthodologie pour la micro-simulation de la tarification effective du carbone.

CHAPITRE 1

ESTIMATION DES ÉMISSIONS CARBONES DES ENTREPRISES INDUSTRIELLES

1.1 Données utilisées

1.1.1 Consommations d'énergie

Cette étude mobilise les données de l'Enquête Annuelle sur la Consommation des Établissements Industriels (EACEI), de l'Enquête sur les Consommations d'Énergie dans le Tertiaire (ECET) et de l'Enquête sur la Consommation d'Énergie dans le secteur de la Construction (ECEC) mises à disposition par l'Insee. L'EACEI est une enquête annuelle sur la consommation énergétique des installations industrielles de plus de 20 employés. Chaque année, environ 8500 établissements du secteur industriel sont sondés, les établissements de plus de 250 employés étant systématiquement enquêtés, tandis que les établissements plus petits sont sélectionnés de manière aléatoire parmi l'ensemble de la population des établissements. Les secteurs couverts par l'EACEI sont les divisions 07 à 33 de la NAF et le groupe 383, hors division 19, hors groupe 091 et hors sous classes 1013B, 1071B, 1071C, 1071D, 2013A et 2446Z. L'ECET et l'ECEC sont quant à elles des enquêtes ponctuelles qui

n'ont à ce jour été réalisées qu'une seule fois, en 2011 pour l'ECET et en 2015 pour l'ECEC. L'ECET recense la consommation énergétique de 20000 établissements des secteurs marchands du tertiaire tandis que l'ECEC couvre 7500 établissements appartenant aux secteurs de la construction. Ces trois enquêtes fournissent une ventilation détaillée des quantités d'énergie consommées par les établissements par type d'énergie, tel que l'électricité, le gaz et la houille. La base PEFA du Ministère de la Transition Ecologique, qui contient des données sectorielles de consommation finale par type d'énergie, est également utilisée afin de calculer les taux de croissance annuels de la consommation des différentes énergies par secteur entre 2005 et 2021 permettant de simuler les consommations des établissements de l'ECET sur l'ensemble de la période.

1.1.2 Caractéristiques des établissements

Les caractéristiques économiques des établissements sont issues de la Base Tous Salariés¹ (Insee), qui couvre l'ensemble des employeurs et des salariés au niveau du poste de travail (salarié × établissement). Ces données sont agrégées au niveau établissement puis appariées avec les données de consommation d'énergie sur la base du SIREN. Cet appariement permet de connaître l'ensemble des établissements au sein des entreprises dont au moins un établissement est présent dans l'EACEI ou l'ECET, ainsi que certaines de leurs caractéristiques, telles que la masse salariale, la proportion d'ouvriers parmi les salariés et le secteur d'activité, entre autres.

1.1.3 Facteurs d'émissions

Les facteurs d'émissions permettant de calculer les émissions de CO₂eq liées aux consommations d'énergies recensées dans les différentes enquêtes proviennent de la Base Carbone de l'ADEME. Les facteurs d'émissions retenus sont résumés dans le

1. anciennement connue sous le nom de DADS, pour Déclaration Annuelle de Données Sociales.

tableau 1.1 et les choix méthodologiques faits sont détaillés ci-après pour chaque énergie.

TABLEAU 1.1 – Facteurs d'émissions

Energie	FE	unités
Bois	0	kgCO ₂ eq/kg
Butane/Propane	2,965	kgCO ₂ eq/kg
Coke de houille	3,03	kgCO ₂ eq/kg
Coke de pétrole	3,1	kgCO ₂ eq/kg
Electricité 2005	0,0502	kgCO ₂ eq/kWh
Electricité 2006	0,0502	kgCO ₂ eq/kWh
Electricité 2007	0,0502	kgCO ₂ eq/kWh
Electricité 2008	0,0502	kgCO ₂ eq/kWh
Electricité 2009	0,0502	kgCO ₂ eq/kWh
Electricité 2010	0,0502	kgCO ₂ eq/kWh
Electricité 2011	0,0502	kgCO ₂ eq/kWh
Electricité 2012	0,0518	kgCO ₂ eq/kWh
Electricité 2013	0,0715	kgCO ₂ eq/kWh
Electricité 2014	0,0677	kgCO ₂ eq/kWh
Electricité 2015	0,0575	kgCO ₂ eq/kWh
Electricité 2016	0,0542	kgCO ₂ eq/kWh
Electricité 2017	0,0532	kgCO ₂ eq/kWh
Electricité 2018	0,0553	kgCO ₂ eq/kWh
Electricité 2019	0,0587	kgCO ₂ eq/kWh
Electricité 2020	0,0583	kgCO ₂ eq/kWh
Electricité 2021	0,0556	kgCO ₂ eq/kWh
Fioul domestique	2,68	kgCO ₂ eq/L
Fioul lourd	3,14	kgCO ₂ eq/kg
Gaz naturel	0,169	kgCO ₂ eq/kWhPCS
Gaz non-naturels	0,469	kgCO ₂ eq/kWhPCI
Houille	2,49	kgCO ₂ eq/kg
Lignite	1,72	kgCO ₂ eq/kg
Liqueur noire	0	kgCO ₂ eq/kg
Vapeur	0,112	kgCO ₂ eq/kg

LECTURE : 1kg de coke de houille consommé correspond à 3,03kg de CO₂eq émis.

NOTES : Le tableau répertorie les facteurs d'émissions liés à la consommation des différentes énergie utilisés dans le cadre de cette étude.

SOURCE : Base Carbone, ADEME.

Choix méthodologiques concernant les facteurs d'émissions :

- **Bois** : Par convention, le facteur d'émissions retenu pour les produits issus de la biomasse, dont le bois, est 0.
- **Butane/propane** : Le facteur d'émission retenu correspond à la moyenne des

facteurs d'émissions du propane et du butane pour la combustion uniquement.

- **Coke de houille, coke de pétrole, fioul domestique, fioul lourd, lignite :** Le facteur d'émission retenu correspond aux émissions liées à la combustion uniquement.
- **Electricité :** Le contenu en CO₂ d'un kWh d'électricité livré sur le réseau varie fortement en France en fonction du mix énergétique utilisé pour le produire, qui change en fonction de la demande, qui elle-même dépend principalement des conditions climatiques. La méthode retenue pour le calcul du facteur d'émission de l'électricité est la méthode "moyenne mensualisée par usage" qui permet de refléter le contenu carbone moyen de l'électricité en fonction à la fois du mix énergétique et du profil d'usage. L'usage utilisé pour cet exercice est l'usage industriel. Le facteur d'émissions retenu comprend les émissions en amont de l'utilisation ainsi que le transport.
- **Gaz naturel :** Les factures de gaz naturel étant généralement exprimée en kWh PCS (et non PCI comme pour les autres combustibles), le facteur d'émission utilisé convertit 1 kWh PCS de gaz en kg de CO₂ équivalent. Seulement les émissions liées à la combustion sont prises en compte.
- **Gaz non-naturels :** L'EACEI renseigne sur la consommation totale de gaz non-naturels, dits de première famille : gaz de cokerie, gaz de haut-fourneau, gaz d'aciérie, gaz d'air propané et gaz d'air butané. En l'absence de détail sur la consommation de chaque gaz, un facteur d'émission unique pour l'ensemble des gaz non-naturels, correspondant à la moyenne des facteurs d'émission des cinq gaz précédemment cités, est utilisé.
- **Houille :** Le facteur d'émission retenu est celui du charbon à coke pour les émissions liées à la combustion uniquement. Ce choix s'explique par le fait que le charbon à coke est la forme de houille utilisée par les hauts fourneaux,

qui constituent les principaux consommateurs de houille dans l'EACEI.

- **Liqueur noire** : S'agissant d'une substance issue de la biomasse, le facteur d'émissions carbone de la liqueur noire est 0 par convention.
- **Vapeur** : Le facteur d'émissions de la vapeur consommée dépend des combustibles utilisés pour la produire (charbon, fioul lourd, gaz, et ordures ménagères). La base carbone répertorie les facteurs d'émissions associés aux différents producteurs de vapeur en France. On retient ici la moyenne de ces facteurs d'émissions.
- **Autres produits pétroliers** : Cette catégorie regroupe dans l'EACEI
 - les gaz de raffinerie
 - le pétrole lampant
 - les résidus lourds de raffinage
 - les goudrons
 - l'éthane
 - le fioul lourd non commercial et le fioul gaz

La base carbone ne répertoriant pas les facteurs d'émissions de ces différents produits, le choix est fait de retenir le facteur d'émission du produit pétrolier le plus proche soit le fioul lourd.

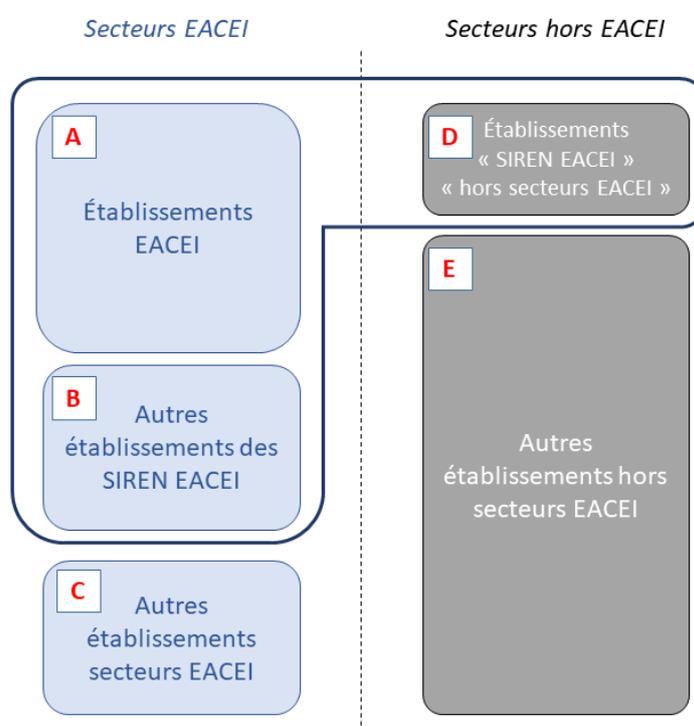
1.2 Méthodologie détaillée

1.2.1 Champ de l'estimation

La méthodologie utilisée pour le calcul des émissions directes de GES des entreprises repose principalement sur les consommations d'énergie des établissements industriels renseignées dans l'EACEI. De ce fait, le champ des émissions estimées est celui des émissions directes des entreprises possédant au moins un établissement

ayant répondu à l'EACEI. Ce champ est représenté dans la figure 1.1. La période couverte s'étend de 2005 à 2021. À noter que pour l'année 2015, les émissions directes des entreprises de la construction dont au moins un établissement a répondu à l'EACEI sont également estimées. Les trois étapes principales du calcul des émissions directes de GES des entreprises sont décrites dans les sections qui suivent.

FIGURE 1.1 – Représentation du champ des établissements considérés



LECTURE : Les établissements dans le champ considéré sont ceux qui sont présents dans l'EACEI et ceux qui appartiennent à une entreprise dont au moins un établissement est présent dans l'EACEI, qu'ils soient dans un secteur couvert par l'EACEI ou non.

NOTE : La figure représente le champ des établissements considérés selon qu'ils appartiennent ou non à un secteur couvert par l'EACEI et font ou non partie d'une entreprise dont au moins un des établissements est présent dans l'EACEI.

SOURCES : Auteurs.

TABEAU 1.2 – Statistiques descriptives des groupes d'établissements

	A	B	C	D	E	X
Nombre d'établissements	7517	6286	98612	4157	2193642	18
Nombre de salariés (en milliers)	2066,5	459,8	1894,3	148,6	36740,5	7,2
Heures travaillées (en millions)	2593,7	572,5	2151,7	188,2	35506,9	9,2
Salaires versés (en milliard €)	59,2	15,5	42,7	5,6	700,9	0,3

NOTES : A : établissements dont le secteur est couvert par l'EACEI et ayant répondu à l'enquête. B : établissements dont le secteur est couvert par l'EACEI et appartenant à une entreprise dont au moins un établissement a répondu à l'enquête. C : établissements dont le secteur est couvert par l'EACEI. D : établissements dont le secteur n'est pas couvert par l'EACEI mais appartenant à une entreprise dont au moins un établissement a répondu à l'enquête. E : établissements dont le secteur n'est pas couvert par l'EACEI, n'ayant pas répondu à l'enquête et n'appartenant pas à une entreprise dont au moins un établissement a répondu à l'enquête. X : établissements ayant répondu à l'enquête mais dont le secteur n'est théoriquement pas couvert par l'EACEI.

SOURCES : EACEI ET BASE TOUS SALARIÉS, INSEE.

1.2.2 Calcul des émissions des établissements à partir des consommations d'énergie observées

Dans un premier temps, il s'agit de calculer les émissions directes de CO₂ des établissements industriels observés dans l'EACEI (qui correspondent au groupe A dans la figure 1.1). Pour ce faire, les données individuelles de consommation de chaque type d'énergie (électricité, fioul lourd, fioul domestique, bois, charbon, etc) des établissements enquêtés sont multipliées par les facteurs correspondant aux émissions carbonées engendrées par la consommation de ces différents types d'énergie calculés par l'Ademe et présentés dans le tableau 1.1. La même méthode est appliquée pour calculer les émissions directes de CO₂ des établissements observés dans l'ECET en 2015.

1.2.3 Imputations des émissions de CO₂ des établissements non observés dans les enquêtes

1.2.3.1 Établissements manquants dans un secteur couvert par l'EACEI

Une fois les émissions carbone calculées et additionnées pour chaque établissement observé, la seconde étape consiste à compléter la base de données avec les établissements non enquêtés dans l'EACEI appartenant aux entreprises dont au moins un établissement a répondu à l'EACEI (groupe B et D dans la figure 1.1). Afin de connaître l'ensemble des établissements pour chaque entreprise possédant au moins un établissement dans le groupe A, les données EACEI d'une année donnée sont appariées avec la Base Tous Salariés de l'année correspondante, préalablement agrégée au niveau établissement, grâce à leur variable commune d'identification des entreprises, le SIREN. La Base Tous Salariés permet également d'obtenir des informations sur les caractéristiques de l'ensemble des établissements telles que leur secteur d'activité, leur localisation au niveau départemental et la composition de leur masse salariale en terme de catégories socio-professionnelles. L'estimation des données de consommation d'énergie et d'émissions carbone manquantes au niveau établissement est ensuite réalisée grâce à un algorithme de gradient boosting² entraîné sur les données EACEI pour la prédiction des établissements du groupe B et sur les données de l'Enquête sur les Consommations des Établissements du Tertiaire (ECET) pour les établissements tertiaires du groupe D. Dans les deux cas, les variables explicatives utilisées dans le modèle de prédiction de la consommation d'énergie et des émissions carbone sont les caractéristiques des établissements de la Base Tous Salariés (effectif, part de salariés dans les différentes catégories socio-professionnelles, nombre d'heures travaillées, salaires, secteur d'activité, département), qui sont connues pour l'ensemble des établissements, qu'ils soient observés ou non dans une des enquêtes.

2. Cet algorithme a été sélectionné pour sa meilleure performance en terme de réduction de l'erreur carrée moyenne suite à un exercice de validation croisée présenté dans la section suivante.

1.2.3.2 Etablissements manquants dans le secteur de la construction

De la même manière que précédemment mais uniquement pour l'année 2015, les données ECET sont appariées avec la Base Tous Salariés afin d'identifier l'ensemble des établissements appartenant à une entreprise dont au moins un établissement est observé dans l'ECET. Un algorithme identique mais entraîné sur les données ECET est ensuite utilisé pour prédire les émissions de CO₂ des établissements non observés dans l'ECET appartenant au secteur de la construction. Les émissions de CO₂ des établissements hors construction ne sont pas estimés.

1.2.3.3 Etablissements manquants dans le secteur tertiaire

Pour la prédiction des émissions des établissements du secteur tertiaire appartenant à une entreprise avec au moins un établissement dans l'EACEI, l'ECET, qui n'a été réalisée que pour l'année 2011, est préalablement "vieillie". Cela consiste à multiplier les données individuelles de consommation des différentes énergies par les taux de croissance de la consommation des énergies correspondantes au niveau national, calculés à partir des données PEFA sur la consommation finale d'énergie par secteur. Le modèle de prédiction est ensuite entraîné pour chaque année sur l'échantillon d'établissements observés dans l'ECET puis utilisé pour estimer les données de consommation d'énergie et d'émissions CO₂ non observées de l'année correspondante.

1.2.4 Agrégation des émissions de CO₂ au niveau entreprise

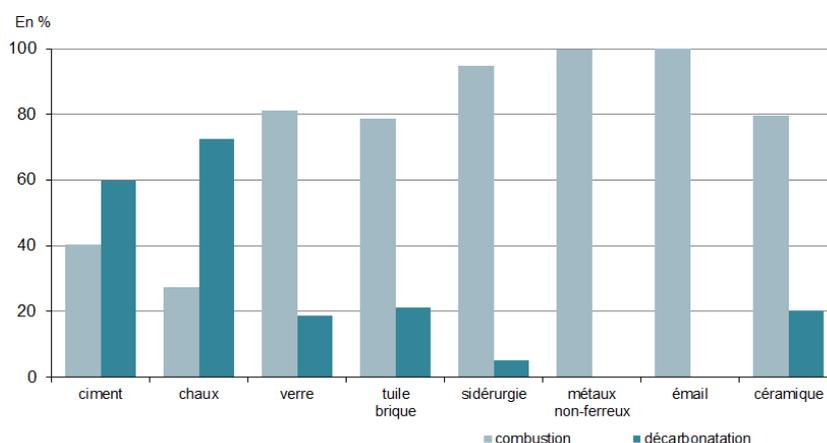
La dernière étape est l'agrégation des émissions de CO₂ de l'ensemble des établissements appartenant à une même entreprise qui consiste simplement à sommer les émissions de CO₂ de ces établissements.

1.2.5 Imputation des émissions de procédés

Pour certains secteurs (notamment des minéraux métalliques et de la chimie), les émissions de procédés représentent une part importante des émissions. Les émissions de procédés ne sont pas directement observables dans l'EACEI. Cependant, puisque l'objectif de notre exercice est d'aboutir à une mesure d'intensité CO₂ de la valeur ajoutée qui soit commensurable à la comptabilité nationale, nous imputons ces émissions de procédés, secteur par secteur.

Pour ce faire, nous obtenons les parts que représentent les émissions de procédés dans les émissions totales de chaque secteur à partir des données du rapport SECTEN. Ces parts sont reproduites dans la Figure 1.2 ci-dessous :

FIGURE 1.2 – Répartition des émissions de CO₂ entre émissions de procédés et de combustion en 2018



LECTURE : Les émissions de procédés représentent 60% des émissions de CO₂ liées à la production de ciment.

NOTES : Les valeurs présentées dans ce tableau correspondent aux parts respectives des émissions de procédés et de combustion pour différentes activités manufacturières.

SOURCE : CITEPA, Rapport SECTEN 2020.

Pour chaque établissement d'un secteur donné, on multiplie ses émissions de combustion de l'inverse du ratio des émissions de combustion dans les émissions totales, pour obtenir une estimation de ses émissions totales. Ceci fait implicitement l'hypothèse qu'à secteur donné, les émissions de procédés sont proportionnelles aux

émissions de combustion.

Pour l'exercice de calcul de la tarification implicite du CO₂ détaillé dans le chapitre suivant, on ne retient que les émissions de combustion (de la même manière que dans les méthodologies de l'OCDE et du CGDD).

1.3 Sélection du modèle de prédiction

Afin de sélectionner le modèle de prédiction le plus fiable, un exercice de validation croisée est réalisé. Il consiste à comparer les performances de plusieurs algorithmes sur cinq couples d'échantillons d'entraînement et de test différents, sélectionnés au hasard parmi l'échantillon d'établissements observés. L'exercice de validation a été réalisé à partir des données de l'année 2019. Les huit algorithmes testés ainsi que les résultats de l'exercice sont présentés ci-dessous.

1.3.1 Régressions linéaires

Deux régressions linéaires différentes sont testées. La première correspond à l'équation suivante :

$$\begin{aligned} \text{Log}(CO_2) = & Nb_heures + Salaires + Nb_salaries \\ & + Part_PCS23 + Part_PCS4 + Part_PCS5 \\ & + Part_PCS6 + EF_secteur + EF_departement \end{aligned} \quad (1.1)$$

La seconde régression linéaire inclut les carrés de certaines variables indépen-

dantes continues afin de capturer d'éventuelles non-linéarités :

$$\begin{aligned} \text{Log}(CO_2) = & Nb_heures + Salaires + Nb_salaries \\ & + Nb_heures^2 + Salaires^2 + Nb_salaries^2 \\ & + Part_PCS23 + Part_PCS4 + Part_PCS5 \\ & + Part_PCS6 + EF_secteur + EF_departement \end{aligned} \tag{1.2}$$

1.3.2 Régressions LASSO

Pour les régressions de type LASSO, la valeur du paramètre de pénalité λ qui minimise l'erreur quadratique moyenne (λ_{min}) est déterminée à l'aide d'une validation croisée à cinq volets. Le modèle (3) consiste en une régression LASSO simple tandis que le modèle (4) est une régression de type Elastic Net avec un paramètre de mélange α égal à 0,1. Les variables des deux modèles sont identiques à celles du modèle linéaire (2).

1.3.3 Forêts aléatoires

Les forêts aléatoires ont de nombreux paramètres pouvant être ajustés, notamment le nombre d'arbres dans la forêt, le nombre de variables sélectionnées pour la division à chaque nœud, et la taille minimale d'un nœud, déterminant la profondeur de l'arbre. Le nombre d'arbres doit être suffisamment élevé pour obtenir une estimation précise, mais pas trop élevé pour rester réalisable sur le plan computationnel. Dans le cadre de cette étude, le nombre d'arbres est fixé à 500 pour les deux modèles de forêt aléatoire, en ligne avec la valeur par défaut de la plupart des logiciels. De même, la taille minimale d'un nœud est 5, conformément à la valeur par défaut recommandée pour les modèles de régression (Breiman et al. 2009). En ce qui concerne le nombre de variables sélectionnées pour la division à chaque nœud, Breiman et al. (2009) suggère de choisir une

valeur par défaut égale au nombre de variables indépendantes divisé par 3. Les variables indépendantes utilisées pour les deux modèles sont les mêmes que dans l'équation (1). Les variables catégorielles étant transformées en variables indicatrices, le nombre de variables indépendantes dans les forêts aléatoires est de 361. En utilisant la règle précédente, la valeur par défaut pour le nombre de variables sélectionnées pour la division est dans ce cas de 120. Toutes les variables indicatrices n'étant probablement pas pertinentes pour la prédiction et par souci de gestion du temps de calcul, un second modèle est testé avec un nombre de variables sélectionnées pour la division fixé à 40. Ainsi, les deux modèles de forêts aléatoires ont les paramètres suivants :

- Modèle (5) :
 - Nombre d'arbres = 500
 - Taille minimale d'un nœud = 5
 - Nombre de variables sélectionnées pour la division = 120
- Modèle (6) :
 - Nombre d'arbres = 500
 - Taille minimale d'un nœud = 5
 - Nombre de variables sélectionnées pour la division = 40

1.3.4 Gradient boosting

Le gradient boosting est un algorithme d'apprentissage supervisé qui consiste en un ensemble d'arbres de décisions aux performances faibles, appelés "weak learners", qui améliorent itérativement les prédictions des weak learners précédents, permettant d'obtenir un modèle de prédiction plus performant. Les modèles de gradient boosting, à l'instar des forêts aléatoires, ont de nombreux paramètres qui peuvent être ajustés pour améliorer la performance du modèle. Entre autres, le nombre maximal d'itérations peut être déterminé par une validation croisée. Ce

nombre est fixé pour les deux modèles de gradient boosting au nombre d'itérations minimisant la racine de l'erreur quadratique moyenne. La profondeur maximale d'un arbre est quant à elle fixée à 3 pour le modèle (7) et à 5 pour le modèle (8). Les autres paramètres prennent les valeurs par défaut de la fonction `R xgboost`.

1.3.5 Comparaison des modèles

Dans le but de comparer les performances des différents modèles, cinq rééchantillonnages aléatoires des ensembles d'entraînement et de test sont réalisés. Les valeurs de R^2 , de la racine de l'erreur quadratique moyenne (RMSE) et de l'erreur absolue moyenne (MAE) de chaque modèle sont enregistrées pour chaque rééchantillonnage. Le tableau 1.3 présente les moyennes de chaque métrique de performance calculées sur les cinq rééchantillonnages pour chacun des huit modèles de prédiction.

TABLEAU 1.3 – Performance des modèles

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R^2	0,545	0,585	0,574	0,572	0,558	0,578	0,656	0,644
RMSE	1,411	1,348	1,354	1,358	1,384	1,359	1,216	1,238
MAE	1,064	1,010	1,016	1,019	1,054	1,036	0,903	0,920

LECTURE : Le modèle (1) a un R^2 égal à 0,545, une racine de l'erreur quadratique moyenne (RMSE) de 1,411 et une erreur absolue moyenne (MAE) de 1,064. Il est donc moins performant que le modèle (2) qui a un R^2 plus élevé et une RMSE et une MAE plus faible.

NOTES : Les valeurs présentées dans ce tableau correspondent aux moyennes calculées sur cinq rééchantillonnages aléatoires des ensembles d'entraînement et de test pour chacun des modèles.

SOURCE : Auteurs.

La valeur de R^2 rapportée est celle calculée par le package *caret* dans R et correspond au carré de la corrélation entre les valeurs observées et prédites. La RMSE est la racine carrée de la somme des carrés des différences entre les valeurs prédites et observées, divisée par le nombre d'observations, tandis que la MAE est la somme des différences absolues entre les valeurs prédites et observées, également divisée par le nombre d'observations. Ces deux dernières mesures de performance sont similaires, mais les erreurs importantes ont un poids plus élevé dans la RMSE

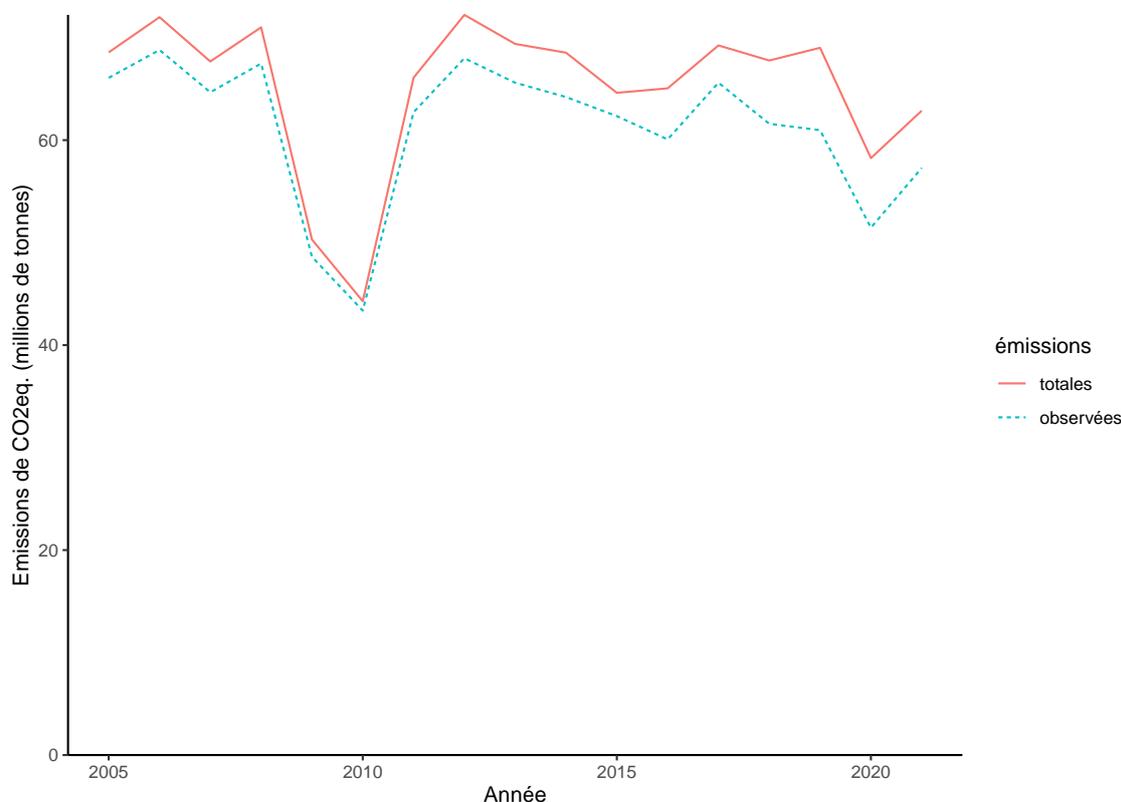
que dans la MAE. La RMSE constitue donc un indicateur de performance plus pertinent dans le cadre de cet exercice. Les modèles de gradient boosting performant significativement mieux que les autres modèles sur les trois métriques. Aucun des modèles de type LASSO ou forêts aléatoires ne permet d'améliorer le modèle linéaire incluant des variables au carré. Cela peut être le signe que les relations entre les variables sont non-linéaires et complexes. Le modèle le plus performant est le modèle (7) et c'est donc celui-ci qui est retenu pour la prédiction des émissions de CO₂ des établissements non observés.

1.4 Résultats

1.4.1 Estimations

Les figures 1.3, 1.4 et 1.5 présentent les résultats agrégés des estimations d'émissions de CO₂. La figure 1.3 montre l'évolution entre 2005 et 2021 des émissions carbone calculées, à la fois sur la base de données observées et d'estimation, dans le champ considéré. La part, en pourcentage du total d'émissions carbonées calculé, d'émissions observées et estimées est représentée dans la figure 1.4 pour chaque année de la période à l'étude. La figure 1.5 décompose les émissions calculées observées et estimées par secteur A38 pour l'année 2019.

FIGURE 1.3 – Evolution des émissions de CO₂ totales et estimées entre 2005 et 2021



LECTURE : EN 2019, LE TOTAL DES ÉMISSIONS DE CO₂EQ CALCULÉES S'ÉLÈVE À 69MT, DONT 61MT DE CO₂EQ SONT OBSERVÉES, SOIT SEULEMENT 8MT DE CO₂EQ SONT ESTIMÉES.

NOTES : Les émissions totales calculées correspondent à la somme des émissions calculées à partir des consommations d'énergie observées dans l'EACEI et des émissions estimées.

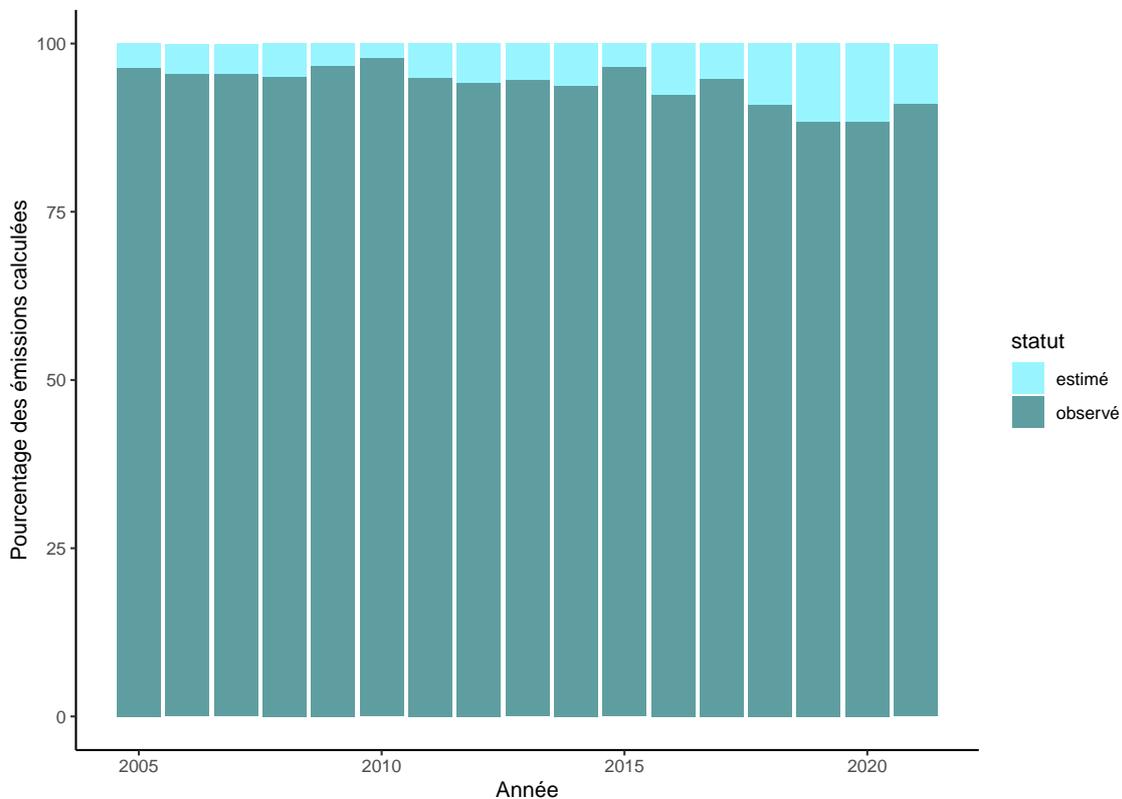
SOURCE : Auteurs.

1.4.2 Contrôle de cohérence

Afin de contrôler la qualité de nos estimations des intensités carbonées, nous comparons,

- d'une part, les estimations obtenues sur la base des données individuelles d'entreprises, c'est-à-dire le calcul des émissions au niveau de chaque unité légale (incluant les imputations décrites dans les sections précédentes), et la valeur ajoutée déclarée dans les liasses fiscales déposées par l'unité légale.
- d'autre part, les ratios équivalents obtenus sur la base des agrégats de comp-

FIGURE 1.4 – Part des émissions de CO₂ observées et estimées par année



LECTURE : En 2019, environ 88% des émissions de CO₂ calculées sont observées et 12% sont estimées.

NOTES : Les émissions observées correspondent aux émissions calculées à partir des consommations d'énergie observées dans l'EACEI tandis que les émissions estimées sont celles prédites grâce à un algorithme de gradient boosting sur la base des caractéristiques des établissements.

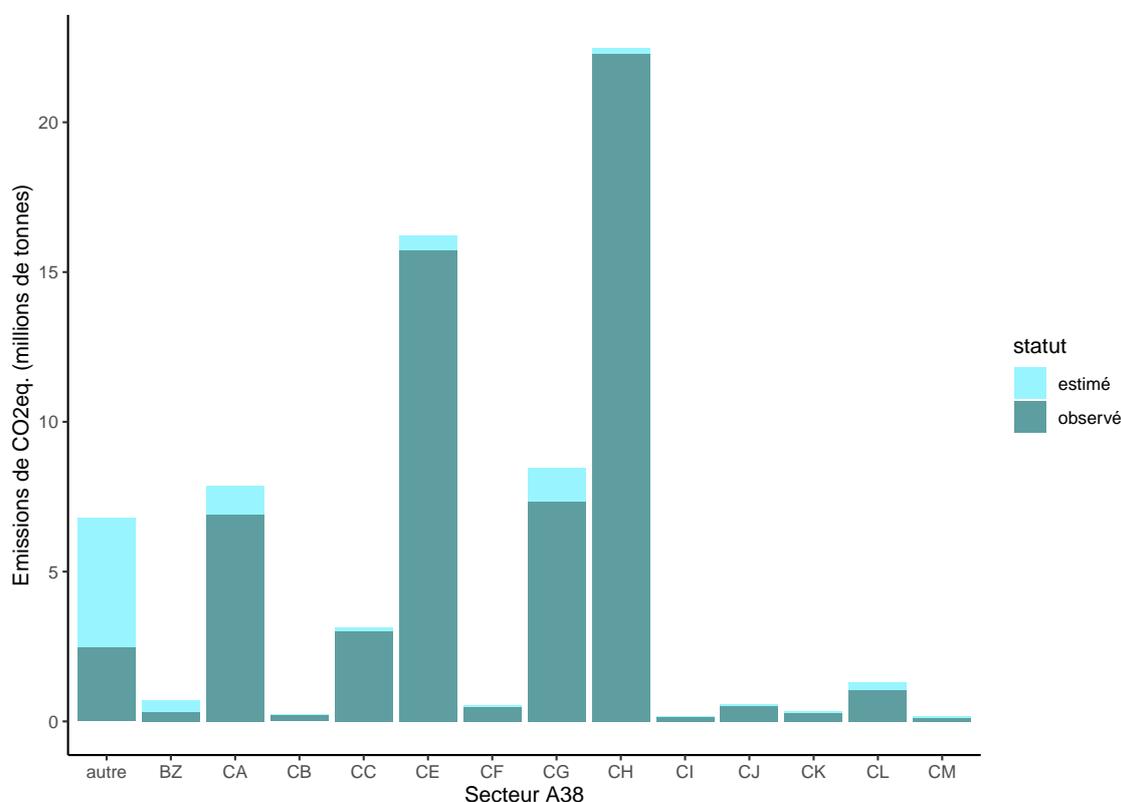
SOURCE : Auteurs et EACEI, Insee.

tabilité nationale disponibles. On retient ici au numérateur les émissions par secteurs issues de l'inventaire national, publiées dans le rapport SECTEN du CITEPA (et incluant donc les émissions de procédés) ; et au dénominateur, la valeur ajoutée par secteur A88 publiée dans la comptabilité nationale.

C'est ce que représente la Figure 1.6 ci-dessous. Globalement, les résultats sont cohérents entre estimations au niveau individuel et niveau agrégé.

Au moins deux autres sources de divergence peuvent influencer sur cette comparaison : la nature des données, et le calcul de la valeur ajoutée. Les données sur lesquelles sont basées les intensités carbone "IPP" représentées en jaune sur la Fi-

FIGURE 1.5 – Emissions de CO₂eq observées et estimées par secteur en 2019



LECTURE : Dans le secteur de la fabrication de produits en caoutchouc et en plastique ainsi que d'autres produits non métalliques (CG), environ 7Mt de CO₂eq sont observées et 1Mt de CO₂eq sont estimées en 2019.

NOTES : Les émissions observées correspondent aux émissions calculées à partir des consommations d'énergie observées dans l'EACEI tandis que les émissions estimées sont celles prédites grâce à un algorithme de gradient boosting sur la base des caractéristiques des établissements.

SOURCE : Auteurs et EACEI, Insee.

gure 1.6 ne sont pas pondérées par les poids d'enquête renseignés dans l'enquête EACEI; en effet, l'étape d'imputation des émissions des établissements non observés (cf. supra) et l'agrégation au niveau unité légale pour pouvoir rapporter les émissions à la valeur ajoutée rendent ces pondérations, utiles au niveau établissement, inopérantes. Une part des petites divergences observées sur la Figure peuvent donc venir de l'échantillonnage de l'EACEI. De plus, le concept de valeur ajoutée retenue dans la comptabilité nationale est légèrement différent du concept retenu dans les liasses fiscales des entreprises, ce qui engendre des petites différences au

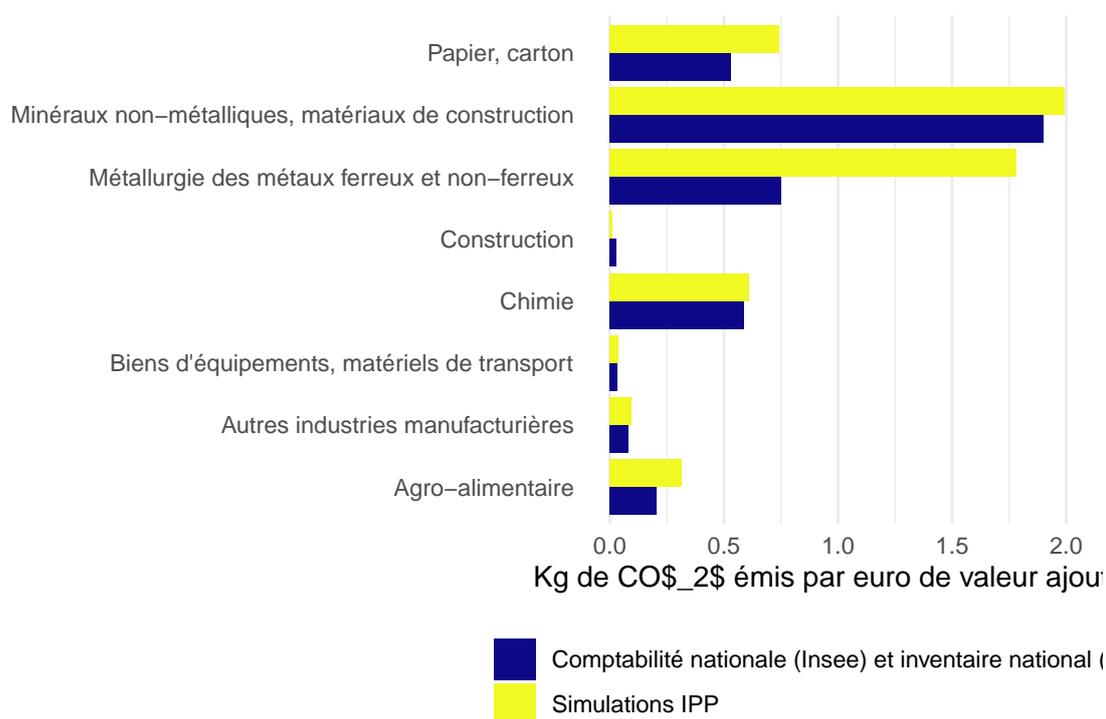
niveau sectoriel (pour une discussion des différentes sources de données sur la valeur ajoutée, cf. Bach, L. (2024) L'impôt sur les superprofits en quête de recettes (1/2) -Petite histoire d'une contribution plus basse que prévu).

Enfin, une divergence importante concerne nos estimations relatives au secteur de la métallurgie. Dans la méthodologie du SECTEN, les émissions de CO₂ liées à la combustion d'énergie primaire dans les hauts fourneaux (typiquement, charbons et gaz de houilles) ne sont pas comptabilisées dans le secteur de la métallurgie mais dans celui de l'énergie. Cependant, les établissements du secteur de la métallurgie concerné présentent des consommations d'énergies importantes qui vont au-delà de l'utilisation dans les hauts-fourneaux (laminoirs, etc). Si des données suffisamment détaillées sur les usages précis faits des différentes énergies par ces établissements peuvent permettre de distinguer les émissions, une difficulté particulière émerge quand il s'agit de rapporter les émissions de CO₂ à la valeur ajoutée. En effet, il est particulièrement délicat, aussi bien théoriquement qu'empiriquement, d'allouer une part de la valeur ajoutée d'un établissement à une partie séparée de son activité, et ce d'autant plus qu'il s'agit d'une partie aussi intégrée au reste de son activité qu'un haut-fourneau pour un métallurgiste ; or c'est ce qui devrait être fait pour conserver une définition cohérente au numérateur et au dénominateur du ratio d'intensité carbone³. C'est pourquoi, nous faisons le choix de laisser l'intégralité des émissions au numérateur des ratios d'intensité carbone pour les établissements relevant du secteur de la métallurgie, étant bien entendu que ceci obtient une mesure d'intensité carbone supérieure à ce qui serait obtenu par l'usage des valeurs agrégées disponibles issues de la comptabilité nationale⁴.

3. Même à supposer qu'une telle répartition soit théoriquement possible, les données fiscales renseignant la valeur ajoutée n'existe pas au niveau d'un établissement mais seulement au niveau d'une unité légale, ce qui pose des enjeux additionnels similaires.

4. Une alternative possible serait d'obtenir des données pour les émissions du secteur de l'énergie qui distinguent la part des hauts-fourneaux dans ces émissions, et de réintégrer ces émissions dans le calcul des indicateurs "agrégés".

FIGURE 1.6 – Comparaison des intensités carbonées par secteur



LECTURE : Dans le secteur de la métallurgie des métaux ferreux et non-ferreux, chaque euro de valeur ajoutée est associée à 0,75 kgCO₂ d'après la comptabilité nationale ; et 1,78 kgCO₂ d'après les estimations IPP.

NOTES : Intensités carbonées par secteur, telles que mesurées par la valeur moyenne prises dans les estimations IPP au niveau de chaque unité légale, d'une part, et les agrégats de comptabilité nationale (valeur ajoutée) et d'inventaire nationale (carbone) d'autre part.

SOURCE : Citepa, Insee, et auteurs (EACEI, ECET, ECEC, BIC RN/RSI, BTS, Base Carbone).

CHAPITRE 2

MICRO-SIMULATION DE LA TARIFICATION EFFECTIVE DU CARBONE

2.1 Les taxes intérieures de consommation

Les quatre taxes énergétiques considérées dans cet exercice de simulation sont les taxes intérieures de consommation sur le gaz naturel (TICGN), sur les produits énergétiques (TICPE), sur le charbon (TICC) et la contribution au service public de l'électricité (CSPE). Seules les trois premières font partie de la tarification effective du carbone selon la définition de l'OCDE¹ et sont simulées de manière plus exhaustive, en prenant en compte les différents tarifs et exemptions. La CSPE n'est quant à elle simulée que pour le tarif normal et sert uniquement dans les calculs d'éligibilité aux tarifs réduits des autres TICs.

2.2 Méthodologie détaillée

1. [Effective Carbon Rates 2018 : Pricing Carbon Emissions Through Taxes and Emissions Trading](#)

2.2.1 Estimation des consommations d'énergie

Afin de calculer le montant de taxes énergétiques dont doivent s'acquitter les entreprises, il est nécessaire de connaître la consommation totale d'énergie de chacun de leurs établissements ainsi que la répartition par type d'énergie et par usage de cette consommation totale. Ces informations sont disponibles pour les établissements observés dans l'EACEI. Pour les autres, une estimation est requise.

2.2.1.1 Consommation totale d'énergie et facture énergétique

La méthode employée pour estimer la consommation totale d'énergie des établissements non observés dans l'EACEI appartenant à une entreprise dont au moins un établissement est enquêté est la même que celle présentée dans le Chapitre 1 pour les émissions carbonées. Pour la facture énergétique, un prix de l'énergie est calculé pour les établissements observés en divisant la facture énergétique renseignée dans l'EACEI par la consommation totale. Le prix moyen payé par les établissements observés d'une même entreprise est ensuite imputé aux autres établissements de cette entreprise, puis multiplié par la consommation totale d'énergie estimée pour obtenir la facture énergétique de ces établissements.

2.2.1.2 Répartition de la consommation par type d'énergie

Pour connaître la répartition de la consommation par type d'énergie des établissements, la part de chaque énergie, en pourcentage de la consommation totale, est calculée pour les établissements observés dans l'EACEI. On calcule la part moyenne de chaque énergie sur l'ensemble des établissements observés appartenant à une même entreprise, que l'on impute ensuite aux établissements qui ne sont pas observés dans l'EACEI. Enfin, multiplier la part de chaque énergie par la consommation totale d'énergie précédemment imputée permet d'obtenir une estimation de la consommation en volume de chaque énergie par les établissements non observés.

2.2.1.3 Répartition de la consommation par usage

Le charbon et la gaz utilisés pour des usages autre que carburant ou combustible n'étant pas soumis aux taxes intérieures, il est important de connaître la répartition de la consommation de chaque énergie par usage pour l'ensemble des établissements. L'EACEI renseigne sur la part de chaque énergie utilisée pour divers usages, tels que la fabrication, la production d'électricité ou comme matière première. La répartition par usage pour les établissements non observés dans l'EACEI est estimée par la part moyenne de chaque usage calculée sur les établissements observés appartenant à la même entreprise, ou la part moyenne de chaque usage au niveau du secteur lorsque l'information au niveau entreprise n'est pas disponible.

2.2.2 Identification des exonérations

Il existe un certain nombre d'exemptions, partielles et totales, de TICs pour les entreprises. D'une part, les installations dites "grandes consommatrices d'énergie" bénéficient d'un tarif réduit pour la TICGN et la TICC lorsqu'elles sont soumises au marché du carbone européen (SCEQE-UE) ou exercent leurs activités dans un secteur à risque de fuite de carbone. D'après le décret n° 2014-913 du 18 août 2014 relatif aux modalités d'application de l'article 265 nonies du code des douanes², sont considérées comme grandes consommatrices d'énergie les installations respectant l'un des deux ratios suivants :

- "Leurs achats d'électricité, de chaleur et d'autres produits énergétiques représentent au moins 3% de la valeur de leur production"
- "Le montant total des taxes applicables à l'électricité et aux produits énergétiques afférents aux consommations de ces installations qui aurait été dû, sans application des exonérations, exemptions, réductions de taux et autres dispositions relatives au non-acquittement des taxes intérieures de consommation

2. Décret accessible ici : <https://www.legifrance.gouv.fr/>

prévues par le code des douanes, représente au moins 0,5% de leur valeur ajoutée."

D'autre part, certains usages spécifiques sont entièrement exemptés de taxe énergétique. Les principaux usages exemptés du gaz naturel et du charbon sont les usages autre que combustible, les usages doubles, la fabrication de produits énergétiques, la fabrication de produits minéraux non métalliques, la production d'électricité, et ceux pour leur propre production ou extraction. Le gaz naturel utilisé pour le biométhane est également exempté³. Pour les produits énergétiques visés par la TICPE, seuls les usages carburant et combustible sont soumis à la taxe, et l'usage pour production d'électricité est exempté.

2.2.2.1 Installations grandes consommatrices d'énergie

Dans cet exercice de simulation, la valeur de la production est calculée au niveau SIREN en sommant la production vendue de biens et de services, les subventions d'exploitations, les variations de stocks et en soustrayant les ventes de marchandises. La valeur ajoutée est également calculée au niveau SIREN comme la différence entre le chiffre d'affaire net et les achats de marchandises. Ces deux variables sont ensuite ventilées au niveau SIRET en utilisant la part des salaires bruts représenté par chaque établissement d'une même entreprise dans les données de la Base Tous Salariés (Insee) comme variable de pondération. Pour le calcul du premier ratio, la facture énergétique des établissements est divisée par leur valeur de production. Pour le second, les taux de base des différentes TICs sont appliqués aux consommations énergétiques correspondantes afin de calculer le montant de taxes qui serait payé par les établissements en l'absence d'exonération. Ce montant est ensuite divisé par la valeur ajoutée des établissements. Les établissements soumis au marché européen du carbone sont identifiés à partir des données de transactions du SCEQE-UE. Les secteurs à risque de fuite de carbone selon la Commission

3. [Article 266 quinquies du code des douanes](#).

européenne sont recensés dans l'annexe de la directive 2003/87/CE du Parlement européen et du Conseil pour la période 2015 à 2019 et dans celle de la décisions déléguée (UE) 2019/708 de la Commission pour la période 2021-2030⁴. Conformément à la définition du décret n° 2014-91, les établissements dont le premier ratio est supérieur ou égal à 3% ou le second ratio est supérieur ou égal à 0,5% et qui sont soumis au SCEQE ou qui exercent dans un secteur d'activité à risque de fuite de carbone sont considérés comme des installations grandes consommatrices d'énergie.

2.2.2.2 Usages exemptés

L'EACEI distingue cinq usages de la consommation des différentes énergies. Il s'agit de l'usage pour la fabrication, l'usage comme matière première (usage non énergétique), l'usage pour la production d'électricité, l'usage pour la production de vapeur et les autres usages. Parmi ces cinq usages, nous considérons l'usage comme matière première et l'usage pour la production d'électricité comme entièrement exemptés pour le gaz, le charbon et les produits pétroliers. Les établissements du secteur de la fabrication d'autres produits minéraux non métalliques (NAF 23) sont également considérés comme entièrement exemptés de TICGN, TICC et TICPE. Les données à notre disposition ne nous permettent pas de prendre en compte les exemptions pour les usages plus spécifiques, tels que le gaz utilisé pour des activités de déshydratation des légumes et plantes aromatiques ou les produits pétroliers utilisés comme carburant ou combustible à bord des navires de pêche.

4. Liste des secteurs à risque de fuite de carbone [2015-2019](#) et [2021-2030](#).

2.2.3 Calcul du coût carbone effectif pour les entreprises industrielles

Le calcul des différentes taxes payées est réalisé au niveau établissement en multipliant la quantité de chaque énergie consommée pour des usages non exonérés par le tarif correspondant à cette énergie et au statut de l'établissement. Pour les établissements soumis au marché du carbone européen, le coût des quotas est également calculé en multipliant leurs émissions vérifiées par le prix du quota carbone sur le marché. L'ensemble de ces coûts sont ensuite sommés et agrégés au niveau entreprise.

LISTE DES TABLEAUX

1.1	Facteurs d'émissions	13
1.2	Statistiques descriptives des groupes d'établissements	17
1.3	Performance des modèles	24

LISTE DES FIGURES

1.1	Représentation du champ des établissements considérés	16
1.2	Répartition des émissions de CO ₂ entre émissions de procédés et de combustion en 2018	20
1.3	Evolution des émissions de CO ₂ totales et estimées entre 2005 et 2021	26
1.4	Part des émissions de CO ₂ observées et estimées par année	27
1.5	Emissions de CO ₂ eq observées et estimées par secteur en 2019	28
1.6	Comparaison des intensités carbonées par secteur	30



L'Institut des politiques publiques (IPP) est développé dans le cadre d'un partenariat scientifique entre PSE-Ecole d'économie de Paris (PSE) et le Centre de Recherche en Économie et Statistique (CREST). L'IPP vise à promouvoir l'analyse et l'évaluation quantitatives des politiques publiques en s'appuyant sur les méthodes les plus récentes de la recherche en économie.

PSE a pour ambition de développer, au plus haut niveau international, la recherche en économie et la diffusion de ses résultats. Elle rassemble une communauté de près de 140 chercheurs et 200 doctorants, et offre des enseignements en Master, École d'été et Executive education à la pointe de la discipline économique. Fondée par le CNRS, l'EHESS, l'ENS, l'École des Ponts-ParisTech, l'INRA, et l'Université Paris 1 Panthéon Sorbonne, PSE associe à son projet des partenaires privés et institutionnels. Désormais solidement installée dans le paysage académique mondial, la fondation décloisonne ce qui doit l'être pour accomplir son ambition d'excellence : elle associe l'université et les grandes écoles, nourrit les échanges entre l'analyse économique et les autres sciences sociales, inscrit la recherche académique dans la société, et appuie les travaux de ses équipes sur de multiples partenariats. www.parisschoolofeconomics.eu



Le Groupe des écoles nationales d'économie et statistique (GENES) est un établissement public d'enseignement supérieur et de recherche. Au sein du GENES, le CREST est un centre de recherche interdisciplinaire spécialisé en méthodes quantitatives appliquées aux sciences sociales regroupant des chercheurs l'ENSAE Paris, de l'ENSAI, du département d'Économie de l'École polytechnique et du CNRS. <http://www.groupe-genes.fr/> – <http://crest.science>

