STATISTIQUES APPLIQUÉES
RAPPORT FINAL

Performances environnementales et performances sociales dans les entreprises, analyse d'une base de données inédite

Matéo Moisan, Pierre Perrigault, Aymeric Perrin
Encadré par Patricia Crifo et Bertrand Garbinti

REMERCIEMENTS

Les travaux de recherche de ce rapport ont été réalisés grâce aux données « ESG indicators sourced from company reports » fournies dans le cadre de l'Équipe Pladifes (ANR-21-ESRE-0036) hébergée par l'Institut Louis Bachelier que nous remercions.

Merci également à Patricia Crifo et Bertrand Garbinti pour leur accompagnement et leurs conseils tout au long de ce travail ainsi qu'à Yuanzhe Tang pour les pistes de progrès indiquées.

INTRODUCTION

Avec la croissance des préoccupations sociales et environnementales dans le débat public qui se reflète notamment dans l'augmentation de la réglementation européenne en matière ESG (Environnement, Social et Gouvernance) [1] , les entreprises semblent intégrer de plus en plus les objectifs RSE (responsabilité sociale des entreprises) dans leur stratégie.

Une littérature économique récente semble montrer que ces préoccupations et la pression sociale qui l'accompagne ont un impact sur les décisions économiques et les choix des entreprises. Dans Austin, Bobek et Harris (2021), les auteurs montrent que l'information sur l'écart salarial entre les sexes influence le comportement des investisseurs. Dans Blundell (2020), les auteurs démontrent qu'une politique obligeant les employeurs à rendre public l'écart salarial entre les sexes a conduit à sa diminution. Ils l'expliquent par une pression croissante exercée par les travailleurs et les consommateurs. De la même manière, Flammer (2013), utilise une étude autour des annonces liées à l'environnement faites par les entreprises. Il montre que les entreprises réputées responsables vis-à-vis de l'environnement connaissent une augmentation significative du cours de leurs actions, tandis que celles agissant de manière irresponsable subissent une diminution significative.

Une question persistante concerne le lien entre les préoccupations environnementales et sociales au sein des entreprises. Par exemple, les recherches de Burkhardt, Nguyen et Poincelot (2020) indiquent que la présence de diversité de genre dans les postes de direction a un impact positif sur la performance environnementale des entreprises, ce qui semble indiquer qu'une relation existe. Cependant, le nombre d'études sur ce sujet est limité par le manque de données agrégées reliant les différentes composantes de l'action RSE des entreprises.

Ce projet vise à répondre à cette question en exploitant une base de données inédite, la base de données Pladifes, qui fournit des informations quantitatives sur les préoccupations environnementales et sociales. L'idée est de vérifier si la réduction de l'empreinte écologique est complètement dissociée de toute recherche de réduction des inégalités sociales ou à l'inverse, si le modèle d'entreprises "vertueuses" - "non-vertueuses" est adapté i.e. les entreprises sont progressistes sur l'ensemble des aspects de la Responsabilité sociétale des entreprises (RSE) ou sur aucun.

Ce projet ne prétend pas poursuivre une analyse causale mais cherche plutôt à mesurer les corrélations entre les différentes préoccupations au sein de notre échantillon d'entreprises, toutes choses égales par ailleurs. Ainsi, les estimations présentés dans ce projet cherchent à mesurer des corrélations partielles entre les variables d'intérêt et non des effets causaux.

Le rapport est organisé comme suit : la Section 1 fournit des informations contextuelles sur le processus de construction des données et les données résultantes. La Section 2 détaille les informations que l'on peut obtenir sur les préoccupations environnementales des entreprises. La Section 3 décrit les informations disponibles sur les préoccupations sociales. La Section 4 établit un lien entre ces deux dimensions. La Section 5 propose des pistes d'approfondissement. La Section 6 revient sur les approches qui se sont révélées infructueuses. Enfin, la Section 7 conclut et explicite les implications pour la construction d'indicateurs ECG.

1. https ://www.amf-france.org/fr/actualites-publications/actualites/la-nouvelle-directive-csrd-sur-le-reporting-de-durabilite-des-societes

1 SOURCE ET CARACTÉRISTIQUES DE LA BASE PLADIFES

Nombre d'observations	40554
Nombre d'entreprises	13518
Nombre moyen d'employés	13142
Années incluses	(2018, 2019, 2020)

Table 1 – Informations sur la base de données

La base de données Pladifes [2] à laquelle nous avons accès est une base de données originale et inédite qui permet d'explorer la relation entre les résultats financiers et extra-financiers des grandes entreprises figurant parmi les plus importantes en termes de capitalisation boursière. Cette base de données de panel regroupe plus de 13 000 multinationales sur une période de trois ans (2018, 2019, 2020).

Elle est élaborée à partir des rapports des multinationales étudiées (rapports annuels, rapports de durabilité, etc.). Les données sont initialement extraites automatiquement à l'aide d'un algorithme de reconnaissance, puis des vérifications et des corrections manuelles sont réalisées. La création de cette base de données répond à un besoin crucial des chercheurs et des investisseurs : bien que de nombreuses grandes entreprises incluent des informations sur leurs initiatives RSE dans leurs rapports annuels, ces données sont souvent dispersées et difficiles à comparer. La base de données Pladifes représente l'une des premières initiatives visant à consolider un vaste éventail d'informations RSE sur plusieurs années et dans différents pays.

Région	**% des Observations**
Asie / Pacifique	59.89
dont Chine	31.14
États-Unis et Canada	19.75
dont États-Unis	17.46
Europe	13.73
Afrique / Moyen-Orient	3.49
Amérique latine et Caraïbes	3.14

Table 2 – Pourcentage d'observations non-manquantes par région

Les multinationales étudiées se concentrent principalement en Chine, en Europe et aux Etats-Unis, mais il y a des observations sur tous les continents. Ces résultats invitent à modifier la catégorisation régionale afin de séparer la Chine du reste des pays de la region Asie-Pacifique, celle-ci représentant plus de la moitié des observations. De manière similaire, la base de données regroupe des entreprises opérant dans un large éventail de secteurs d'activité. La nomenclature américaine GICS (Global Industry Classification Standard), utilisée dans la base de données, répartit les entreprises multinationales en 12 secteurs industriels. Parmi ceux-ci, les secteurs prédominants dans la base de données sont les technologies de l'information, l'industrie manufacturière et l'industrie des matériaux (qui inclut les matières premières et l'industrie chimique). Une classification en 28 sous-secteurs (voir Table A.2) montre également l'importance des industries chimiques, pharmaceutiques et des équipements technologiques.

Les indicateurs qui nous intéressent particulièrement sont ceux relatifs à la mesure de la pollution de l'entreprise et à son engagement social. En ce qui concerne les préoccupations sociales, nous nous concentrons

2. https ://pladifes.institutlouisbachelier.org/

Secteur	% des Observations
Technologies de l'information	12.86
Biens industriels	12.58
Matériaux	11.58
Biens non-essentiels	10.95
Secteur financier	10.93
Santé	9.72
Biens de consommation courante	8.48
Commerce, transport	7.66
Services de communication	4.44
Immobilier	3.94
Énergie	3.48
Services publics	3.39

Table 3 – Pourcentage d'observations non-manquantes par secteur

principalement sur l'écart de rémunération entre hommes et femmes, un indicateur largement utilisé dans la littérature (Voir Austin, Bobek et Harris (2021)). Pour ce qui est des préoccupations environnementales, notre attention se porte principalement sur les émissions directes et indirectes de gaz à effet de serre (GES), mesurées selon les "scopes" définis par la méthodologie internationale de comptabilité carbone GHG Protocol. La Figure 1 détaille les différentes émissions prises en compte théoriquement dans les scopes.

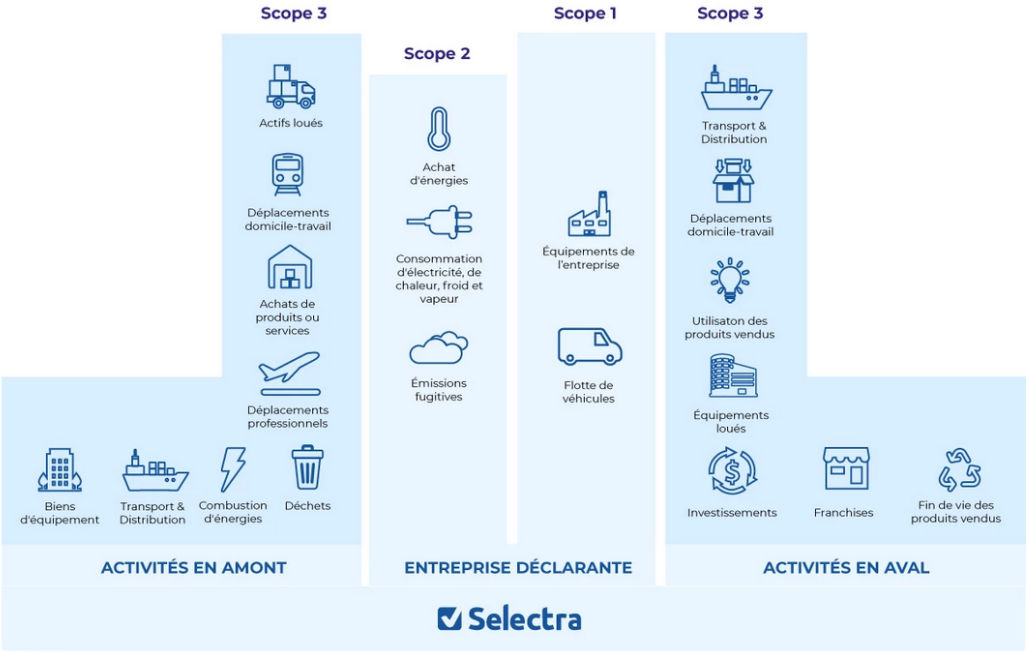

Figure 1 – Émissions au sein des scopes, source : Selectra

Afin de résumer ce que représentent les différents scopes, on peut garder en tête que le scope 1 représente les émissions directes de GES produits par l'entreprise, le scope 2 correspond aux émissions indirectes liées à l'énergie, mais qui ne se produisent pas directement sur le site de l'entreprise, et le scope 3 prend en compte toutes les autres émissions de la chaîne de valeur, qu'il s'agisse des fournisseur comme des clients. Il faut noter que le reporting des émissions scope 3 est souvent compliqué à effectuer précisément. La délimitation précise du scope 3 peut être sujette à débat ; en pratique, beaucoup d'entreprises minimisent leurs émissions sur ce périmètre. Les émissions scope 3 étant tellement larges (dans le sens où beaucoup d'activités, parfois

très éloignées de l'entreprise, doivent être considérées), il est aisé de laisser passer un grand nombre de celles-ci pour une entreprise. Dans les années à venir, des normes et des obligations permettront sûrement de pouvoir travailler avec des données de meilleure qualité et plus comparable entre elles.

Nom des colonnes	**Description**	**Valeurs manquantes (%)**
scope_1	Les émissions scope 1 (directes) de l'entreprise (en T/CO2e)	66%
scope_2	Les émissions scope 2 (indirectes, détenues) de l'entreprise (en T/CO2e)	66%
scope_3	Les émissions scope 3 (indirectes, non détenues) de l'entreprise (en T/CO2e)	69%
waste_production	La production annuelle de déchets produits (en T)	84%
waste_recycling	La quantité annuelle de déchets recyclés (en T)	89%
water_consumption	La quantité annuelle d'eau consommée (en T)	72%
energy_consumption	La consommation d'énergie électrique annuelle (en KWH)	82%
hours_of_training	Le nombre d'heures de formations dispensées pour tous les employés sur l'année (en H)	87%
gender_pay_gap	L'écart relatif entre les salaires horaires moyen masculin et féminin (en %)	96%
legal_costs_paid_for_controversies	Le montant total des coûts liés à des procédures légales (en $)	59%
independant_board_members	Le pourcentage de membres du board de l'entreprise indépendants (en %)	75%

Table 4 – Description des indicateurs environnementaux et de responsabilité sociale des entreprises

Comme l'indique la Table 4, ces indicateurs sont malheureusement fortement affectés par un problème de données manquantes. 96% des observations ne fournissent pas d'informations sur l'écart salarial entre hommes et femmes, et 66% ne mentionnent pas le niveau des émissions.

2 PRÉOCCUPATIONS ENVIRONNEMENTALES

Dans cette section, nous nous intéressons à comprendre les différentes dimensions de l'action environnementales des entreprises. En particulier, nous cherchons à savoir s'il existe des entreprises "vertueuses", c'est-à-dire qui sont exemplaires (au sein de leur secteur ou région) sur différents indicateurs environnementaux. Une entreprise qui se démarquerait seulement sur un indicateur ne pourrait effectivement pas être considérée comme vertueuse, cette performance pouvant être attribuée à un facteur d'opportunité ou à une envie de paraître "verte".

Nous étudions dans un premier temps les relations qu'il peut exister entre les émissions selon les différents scopes, afin de montrer que justement certaines entreprises ont des niveaux d'émissions très différents selon le

scope. Ensuite, nous réfléchissons au moyen d'identifier les entreprises vertueuses, et nous mettons en avant leur présence ou absence au sein de la base de données.

2.1 Normalisation des variables

Dans cette partie et dans la suite de notre étude, nous allons comparer les entreprises entre elles sur des variables continues comme les émissions ou la consommation d'énergie. Ces variables sont caractérisées par le fait qu'elles augmentent avec la taille de l'entreprise, cette taille étant mesurée par le nombre d'employés, le chiffre d'affaire, ou la capitalisation boursière. Une entreprise de plus grande taille aura, par rapport à une autre entreprise présentant, sinon la taille, les mêmes caractéristiques, des émissions supérieures du fait de son activité plus importante. Afin de ne pas perturber l'analyse avec ce biais de taille, il nous faut normaliser les variables continues. Nous avons fait le choix de travailler avec les variables ramenées à leur nombre d'employés.

Tout d'abord, la capitalisation boursière ne nous semble pas une option avantageuse, celle-ci peut varier rapidement et les bourses régionales peuvent valoriser des critères différents, amenant un biais supplémentaire. Pour le chiffre d'affaires, il nous semble qu'il constitue un indicateur de taille plus directement corrélé aux autres variables que le nombre d'employés. En effet, une entreprise faisant des efforts de réduction d'émissions va faire varier son chiffre d'affaires, car elle va gagner en efficacité ou bien réduire certaines activités. Mais son nombre d'employés risque de rester, au moins temporairement, constant.

Si l'on veut capter les ambitions vertueuses ou non des entreprises, il paraît plus pertinent d'utiliser le nombre d'employés comme indicateur de taille, ce que nous ferons. Dans la suite de notre étude, nous mentionnerons le fait qu'une variable ait été normalisée ou non.

2.2 Niveaux des émissions

Dans un premier temps, nous nous intéressons aux corrélations entre les émissions selon les différents scopes pour une même entreprise. Il est question de savoir si une entreprise plus émettrice sur un scope l'est forcément sur les autres ou si au contraire un scope supérieur à la moyenne implique de faible émissions pour les autres scopes, ce qui voudrait dire que les entreprises ne sont pas vertueuses mais "délocalisent" leurs émissions, ou à l'inverse polluent énormément pour des entreprises plus "propres".

En prenant les émissions divisées par leur moyenne par scope, on obtient la matrice de corrélation suivante.

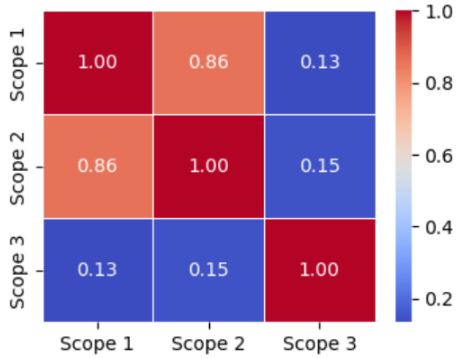

Figure 2 – Matrice de corrélation entre les différentes scopes, ramenées à leur moyenne

On obtient avec combinaison de scope des corrélations positives, avec une corrélation particulièrement élevée pour la relation scope 1 - scope 2 (0.86), et plus faibles entre scope 3 et les autres (0.13 entre scope 1 - scope 3).

La première relation découle des définitions des deux premières scopes, les activités émettrices en premier scope étant également particulièrement intenses en énergie [3].

Les corrélations de scope 3 avec les autres scopes d'émissions étant plus faibles, on peut penser qu'il existe des mécanismes plus complexes et surtout que cela peut cacher une diversité importante de cas. Deux phénomènes principaux nous semblent être principalement à l'oeuvre. Premièrement, une dynamique de corrélation positive : les entreprises faisant des efforts sur le plan environnemental auront tendance à avoir des émissions plus faibles au sein de leur secteur, et ce selon toutes les scopes, et inversement pour les entreprises moins exemplaires au sein de leur secteur. Deuxièmement, une dynamique de corrélation négative : l'activité d'une entreprise peut être émettrice au premier scope (et au deuxième) tandis que le reste de la chaîne de valeur ne l'est pas, alors le scope 3 sera faible, et inversement, on peut parler d'entreprises "exportatrices" ou "délocalisatrices" d'émissions scope 1.

Pour se rendre compte de la distribution de ces entreprises "exportatrices", "délocalisatrices" ou autre, nous représentons les classements d'émissions intra-sectorielle : par secteur, une entreprise est classée selon son niveau d'émission. Elle est classée première si c'est l'entreprise la moins émettrice, et dernière si c'est la plus émettrice. Ces graphes font apparaître qu'il existe effectivement des dynamiques diverses.

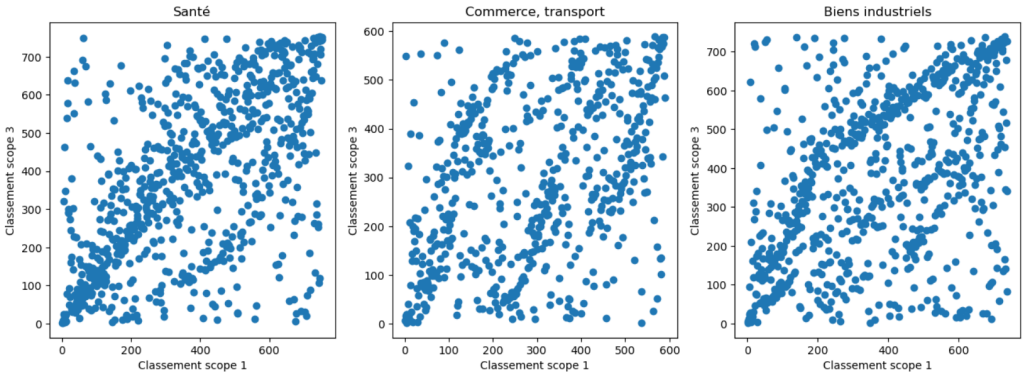

Figure 3 – Relation entre classement scope 1 et 3 intra-sectorielle pour 3 secteurs particuliers

Pour interpréter ces graphes, il est nécessaire de comprendre que les entreprises présentes autour de la droite $x = y$ représentent la dynamique de corrélation positive évoquée au-dessus, et que les entreprises éloignées de cette droite la dynamique de corrélation négative. Les entreprises présentes sur la diagonale ont un aussi bon classement sectoriel pour les émissions scope 1 que scope 3, elles sont aussi exemplaires pour ces deux scopes d'émissions. A l'inverse, les entreprises éloignées de la droite $x = y$ sont soit exemplaires pour les émissions scope 1 mais mal classées pour les émissions scope 3 (toujours au sein d'un même secteur), soit exemplaires pour les émissions scope 3 mais mal classées pour les émissions scope 1. Ce sont justement les entreprises "délocalisatrices" et "exportatrices" que l'on évoquait précédemment.

Ces graphes laissent donc apparaître la présence de tous ces types d'entreprises. La relation entre scope 1 et scope 3 est effectivement plus complexe qu'entre scope 1 et scope 2. On se demande alors si ces différentes dynamiques ne dépendent pas seulement du sous-secteur : il y a des sous-secteurs plus délocalisateurs ou plus exportateurs, qui font que la distribution n'est pas concentrée autour de la droite $x = y$, mais au sein des sous-secteurs il y a uniquement des entreprises plus ou moins exemplaires vis à vis de leurs émissions. On retrouve effectivement cela au sein de certains secteurs, comme "Biens industriels".

On voit clairement que pour certains sous-secteurs, la distribution est assez claire, comme "Machines industrielles" ou "Construction et ingénierie". Au sein de ces sous-secteurs, les émissions scope 1 - émissions

3. Les deux secteurs les plus émetteurs en scope 1 et les plus consommateurs d'énergie sont les mêmes : Services Publics et Énergie

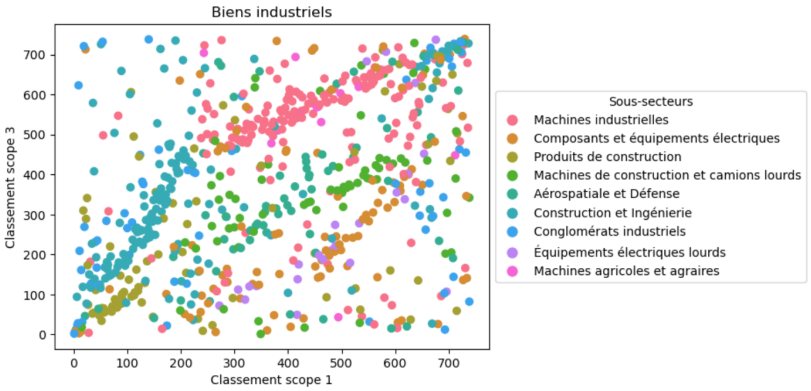

Figure 4 – Relation entre classement scope 1 et scope 3 intra-secteur pour "Biens industriels"

scope 3 sont très fortement corrélées. Il serait intéressant de pouvoir effectuer des régressions en contrôlant par sous-secteur, malheureusement le faible nombre de données disponibles ne nous l'a pas permis.

Afin de déterminer plus globalement quels sont les facteurs (régionaux, sectoriels ou nombre d'employés) qui font que les entreprises sont plutôt "exportatrices" ou "délocalisatrices", nous faisons une régression sur le fait d'appartenir ou non à un de ces deux groupes. Puisqu'une analyse secteur par secteur ne serait pas pertinente du fait du faible nombre d'entreprises concernées par secteur, nous proposons une analyse tous secteurs confondus. On classe les entreprises par déciles d'émissions scope 1 et déciles d'émissions scope 3. On définit ensuite deux groupes : le premier est composé des entreprises qui sont dans les trois premiers déciles de scope 1 et dans les trois derniers de scope 3 ; le deuxième est définit en inversant les scopes. Le groupe 1 représente donc les entreprises "délocalisatrices" car elles sont exemplaires pour les émissions scope 1 mais mauvaises pour les émissions scope 3, et le groupe 2 les entreprises "exportatrices" car elles sont exemplaires pour les émissions scope 3 mais mauvaises pour les émissions scope 1. Après cette catégorisation, nous avons 198 observations dans le groupe 1, et 253 observations dans le groupe 2.

En créant une variable binaire qui correspond au fait d'appartenir ou non au groupe 1, on peut caractériser l'effet des différentes variables sur le fait d'être dans un groupe ou un autre via des régressions Logit, Probit et linéaire, voir 7. Nous effectuons les différentes régressions afin de tester la robustesse de nos résultats, mais nous nous intéressons principalement aux résultats de la régression Logit.

Pour interpréter les effets marginaux de la table 7, on peut prendre l'exemple de l'effet moyen obtenu pour la variable "Afrique / Moyen-Orient" : le fait d'être implanté dans la région diminue en moyenne de 27% la probabilité d'être dans le groupe 1 plutôt que dans le groupe 2 (toute chose égales par ailleurs, et par rapport à la référence régionale "États-Unis et Canada").

On obtient pour chaque régression des résultats négatifs et significatifs pour les régions qui peuvent effectivement être "exportatrices" d'émissions scope 1, comme "Afrique / Moyen-Orient" ou "Asie / Pacifique" (Chine mise à part ou non). L'Europe est au contraire significativement plus représentée dans le groupe 1 des entreprises "délocalisatrices", ce qui semble normal dans le cadre actuel avec une Europe peu industrialisée et comptant plus que les États-Unis sur les usines dans les régions du monde évoquées au dessus.

Au sein des secteurs, les résultats correspondent également à ce que l'on pourrait attendre. Les secteurs "Biens industriels", "Commerce, transport" et "Services publics" amenèrent plus à être au sein du groupe des entreprises "exportatrices" (coefficients négatifs et significatifs). De plus, les secteurs qui paraissent plus propres en première approximation, comme "Services de communication" ou "Secteur financier", présentent des résultats positifs et significatifs pour toutes les régressions.

Cette partie nous permet de voir qu'il est nécessaire de prendre en compte tous les niveaux d'émissions afin de déceler la présence ou non d'entreprises vertueuses, . En effet, une entreprise exemplaire en scope 1 ne l'est pas forcément en scope 3 et les dynamiques à l'oeuvre pour déterminer la relation entre scopes sont nombreuses (secteurs, sous-secteurs, régions). On ne peut se contenter de considérer qu'un seul niveau d'émissions.

2.3 Entreprises vertueuses

L'objectif de cette partie de notre étude est de mettre en évidence la présence ou non d'entreprises vertueuses au sein de la base de données, c'est-à-dire d'entreprises faisant attention à chacun des enjeux environnementaux, et donc étant exemplaire sur chaque variable. Nous allons procéder par un raisonnement similaire à la partie précédente, en assignant un groupe aux entreprises exemplaires sur les deux scopes 1 et 3, et en vérifiant que les entreprises de ce groupe sont aussi "meilleures" que les autres sur les autres variables environnementales, tout en contrôlant sur les secteurs, régions et années.

En reprenant notre catégorisation des entreprises par déciles d'émissions scope 1 et scope 3, on définit un groupe composé des entreprises qui sont dans les 3 premiers déciles pour scope 1 et scope 3. On retrouve donc dans ce groupe (que l'on nomme encore groupe 1, et les autres du groupe 2) les entreprises faisant attention à tous leurs niveaux d'émissions, et pas seulement "exportatrices" ou "délocalisatrices" comme vu précédemment. Les entreprises de ce groupe restent très diverses dans leurs activités, on peut y trouver des entreprises de services comme JPMorgan Chase & Co., d'alimentaire comme The Coca Cola Company ou bien d'électronique avec Samsung Electronics Co.. On régresse le fait d'appartenir ou non à ce groupe (variable binaire) sur les variables environnementales ramenées au nombre d'employés. (Voir tableau 8)

Pour les variables environnementales, les résultats obtenus sont systématiquement significatifs et sont cohérents entre les régressions Logit et Probit. Les coefficients de la régression linéaires diffèrent en termes d'ordre de grandeur et de significativité, ce qui peut venir du déséquilibre d'observations pour les deux groupes. Le modèle linéaire est plus facilement impacté par le fait qu'il y a des entreprises très différentes au sein du groupe 2.

En s'intéressant aux résultats des régressions Logit et Probit, on voit que toutes les variables environnementales excepté "Production de déchets" augmentent la probabilité d'être dans le groupe 1. Les entreprises du groupe 1 sont donc, toutes choses égales par ailleurs, plus attentives à leurs impacts environnementaux et pas seulement leurs émissions. Il est cependant nécessaire de noter que les ordres de grandeur sont très faibles, entre $10^{-2}\%$ et $10^{-5}\%$. Cela peut être dû au faible impact de ces variables, mais également au plus grand nombre d'observations dans le groupe 2 : la distribution est très étirée vers le groupe 2, et donc les effets marginaux moyens. Le résultat positif obtenu pour "Production de déchets" est difficilement interprétable, cela pourrait venir d'une trop grande corrélation avec les autres variables, en effet l'optimisation de la production de déchets est fortement corrélée à la baisse d'émission, via des changements de procédés ou de techniques de production.

Il semble donc qu'il existe bien des entreprises vertueuses dans le sens où elles s'intéressent à tous leurs impacts environnementaux. Il s'agit de prendre en compte cet aspect multivariable pour juger de la bonne volonté d'une entreprise.

Par ailleurs, il est également intéressant de noter qu'il semble y avoir un "effet Covid" : l'année 2020 a un effet positif significatif. Durant l'année 2020, une entreprise, toutes choses égales par ailleurs, a plus eu tendance à avoir des émissions scope 1 et scope 3 faibles.

3 PRÉOCCUPATIONS SOCIALES

Dans cette section, notre objectif est de comprendre les différentes dimensions de l'action sociale des entreprises. Pour ce faire, nous nous concentrons particulièrement sur l'analyse des écarts de rémunération entre hommes et femmes, qui représente l'indicateur principal des préoccupations sociales dans la littérature. Cet indicateur ne pouvant être étudié directement, compte tenu du taux très élevé de données manquantes, nous avons eu l'idée d'utiliser l'absence de cette information comme un moyen de capter la mauvaise performance d'une entreprise en matière d'égalité salariale. En effet, si l'on part du principe que l'égalité salariale est aujourd'hui perçue comme un objectif de RSE à atteindre par une majorité de la population (et des investisseurs), une entreprise n'aurait aucun intérêt à ne pas mettre en évidence cette information dans son rapport annuel si elle obtenait de bons résultats dans ce domaine.

3.1 Plausibilité de l'hypothèse d'identification

Il n'est pas possible de tester l'hypothèse selon laquelle l'écart de rémunération femmes-hommes est élevé pour les entreprises qui ne divulguent pas cette information, puisque par définition, cette information est manquante. Cependant, des tests de plausibilité de cette hypothèse sont envisageables.

Une première idée est de comparer les caractéristiques des entreprises avec un écart de rémunération femmes-hommes particulièrement élevé aux entreprises qui ne divulguent pas. Nous pourrions notamment envisager d'étudier la distribution sectorielle, géographique et la taille de ces entreprises. Cependant, le nombre restreint d'entreprises qui divulguent ne permet pas de définir un groupe d'entreprises avec un écart de rémunération femmes-hommes élevé et donc d'effectuer ce test.

	Écart de remunération (en %)
Échantillon d'entreprises qui divulguent	11.8 (0.57)
Référence mondiale	16
Référence UE	13

Écart type entre parenthèses

Table 5 – Écarts de rémunérations moyenne entre hommes et femmes

Une autre approche consiste à vérifier si le niveau d'écart salarial des entreprises qui divulguent est particulièrement faible par rapport à un référentiel international. Pour ce faire, nous calculons l'écart salarial moyen entre hommes et femmes sur les 405 observations qui divulguent et nous le comparons avec l'indicateur 8.5.1 des Objectifs de Développement Durable (ODD), qui mesure la différence de "rémunération horaire moyenne des salariés hommes et femmes" telle que rapportée dans ILO (2018). Dans la Table **??**, nous reportons ces estimations : L'écart salarial moyen entre hommes et femmes à l'échelle mondiale est d'environ 16% et de 13% dans l'UE, à partir de données d'enquêtes provenant de 73 pays représentant environ 80% des travailleurs du monde, contre 11.8% dans notre échantillon. Ce résultat semble confirmer notre hypothèse, mais il doit être pris avec prudence, car l'écart moyen varie considérablement selon le secteur d'activité et le pays, et notre échantillon d'entreprises n'est pas représentatif de la distribution mondiale des entreprises. Il est composé principalement de grandes entreprises situées dans des pays développés. Par conséquent, l'écart de salaire moyen de notre échantillon pourrait être inférieur à 16%, et pas seulement celui des entreprises qui divulguent.

3.2 Caractérisation des entreprises qui divulguent

En supposant que nous écartons les doutes que l'on peut avoir sur l'hypothèse, il est intéressant de comprendre ce qui distingue les entreprises qui divulguent de celles qui ne le font pas. Pour ce faire, nous examinons la corrélation entre la non-divulgation et les caractéristiques des entreprises telles que la localisation, la taille, la performance environnementale et sociale, et le secteur d'activité, afin de mettre en évidence les caractéristiques déterminantes.

Variable	**Corrélation avec la non-divulgation**
Royaume-Uni	-0.218913
Europe	-0.193395
Espagne	-0.102673
Chine	0.066475
Chine (région corrigée)	0.066475
Revenus	-0.062693
Consommation d'eau par employé	-0.060108
Nombre d'employés	-0.049331
France	-0.045801
Portugal	-0.041345

Table 6 – Corrélations principales entre la non-divulgation des écarts de salaires et les caractéristiques observables des entreprises

Les variables les plus corrélées avec la divulgation sont celles liées à la taille et à la localisation de l'entreprise. La Table 6 montre qu'au niveau de la localisation, les entreprises européennes (région Europe, Royaume-Uni, Espagne, France, Portugal) sont systématiquement associées à une proportion plus élevée d'entreprises qui divulguent, tandis que les pays de la région Asie-Pacifique, notamment la Chine, qui représente une part importante des entreprises, semblent être associés à une proportion plus faible d'entreprises qui divulguent. De même, le chiffre d'affaires et le nombre d'employés sont positivement corrélés avec la divulgation. En revanche, la variable du secteur d'activité, que nous imaginions pourtant déterminante, ne semble pas être particulièrement corrélée.

La Table A.1 parvient aux mêmes conclusions en examinant cette fois les corrélations partielles de nos variables identifiées comme fortement corrélées avec la non-divulgation. En contrôlant les effets sectoriels et la taille de l'entreprise et de l'année, le fait d'être en Chine par rapport aux États-Unis ou au Canada est associé à moins de divulgation, tandis que le fait d'être en Europe plutôt qu'aux États-Unis est associé à une plus grande divulgation, et ces corrélations partielles sont significatives au niveau de 1%. De même, la taille de l'entreprise, mesurée par le chiffre d'affaires (ou le nombre d'employés, variables corrélées à près de 70%), a un effet significatif à 1% et est associée à une divulgation plus élevée.

Les corrélations partielles sont identifiées via les effets marginaux d'un modèle Logit. Pour vérifier la robustesse de cette estimation, nous présentons également les résultats d'une régression Probit qui suppose une distribution normale des termes d'erreur, d'une régression linéaire estimée par les MCO qui ne suppose pas de distribution mais peut être biaisée si les valeurs prédites sont en dehors de [0,1] et d'une régression Logit dans laquelle les plus petites entreprises ont été exclues. En effet, la Table A.3 montre que la base contient quelques entreprises atypiques avec un nombre d'employés presque nul, ce qui est surprenant pour des multinationales. Les conclusions restent inchangées dans ces cas.

Une interprétation possible de ces résultats est que dans les pays occidentaux, en particulier en Europe où les préoccupations sociales prennent une plus grande importance, les populations exercent une forte pression sur les entreprises, notamment les plus grandes, pour qu'elles divulguent les écarts de salaires au sein de leur organisation. En revanche, les entreprises chinoises, qui ne sont pas soumises à la même pression sociale, peuvent se permettre de ne pas divulguer cette information et ont une performance moindre à cet égard. Ces résultats font écho aux engagements pris par certains pays développés, dont les pays européens, en matière de publication des salaires du secteur privé. En 2021, 18 pays de l'OCDE avaient déjà mis en place des réglementations contraignantes à cet égard (voir OECD (2021)). De même , le 24 avril 2023 le Conseil de l'UE a voté de nouvelles règles en matière de transparence des rémunérations. Cette décision stipule que les entreprises de l'UE comptant plus de 100 employés devront divulguer des informations sur les salaires et prendre des mesures si l'écart salarial entre les sexes dépasse 5% [4].

3.3 Cohérence des préoccupations sociales

Une autre question cruciale consiste à déterminer si les entreprises adoptent une approche cohérente dans les divers aspects des préoccupations sociales. Cette section tente de répondre à la question suivante : les entreprises affichant des écarts de salaire faibles (celles qui divulguent) sont-elles également celles qui mettent en place des politiques de formation et veillent au respect des droits de leurs employés ?

La Table 9 présente les corrélations partielles entre la divulgation et les autres variables sociales présentes dans la base de données, en contrôlant une nouvelle fois l'effet du secteur d'activité, de la taille des entreprises, de l'année et de la région. Le nombre d'heures de formation par employé, est corrélé positivement avec la non-divulgation des écarts de salaire à 1%. Le montant des procédures légales engagés par employés n'est pas corrélé aux écarts de salaires entre hommes et femmes. Enfin, une corrélation forte semble apparaître entre la proportion de membres indépendants dans le comité de direction de l'entreprise et un niveau plus faible d'inégalités femmes-hommes, et celle-ci est significative à 5%.

Les corrélations partielles sont identifiées via les effets marginaux d'un modèle Logit pour les mêmes raisons que précédemment. Nous avons choisi de normaliser le coût des procédures légales et le nombre d'heures de formation par le nombre d'employés afin d'avoir une mesure intensive. Pour garantir la robustesse des résultats, nous présentons également les résultats d'un Probit, d'une régression linéaire, et d'un Logit dans lequel nous avons winsorisé les variables sur le nombre d'heures travaillées par employé, le coût légal par employé et le pourcentage de membres indépendants du comité de direction au 5ème et au 95ème centile, afin de vérifier que des valeurs aberrantes ne conduisent pas nos résultats. Les conclusions restent inchangées dans ces cas.

Ces résultats suggèrent que les entreprises ne suivent pas nécessairement une politique sociale cohérente. Les politiques d'égalité femmes-hommes sont corrélées négativement ou ne sont pas corrélées avec les politiques de formation, proxy par le nombre d'heures de formation, et avec la qualité des conditions de travail (si l'on interprète le montant des procédures judiciaires par employé comme un indicateur de mauvaises conditions de travail au sein de l'entreprise). En revanche, nos résultats suggèrent que la diversité au sein du comité de direction a un impact significatif sur les décisions prises en matière sociale. Ces résultats correspondent aux conclusions de la littérature, où l'on observe que les entreprises qui représentent mieux la société, en ayant une proportion plus importante de femmes par exemple (comme dans Burkhardt, Nguyen et Poincelot (2020)), ou qui comptent des membres indépendants dans leur comité de direction, tendent à mieux prendre en compte les préoccupations sociales dans leurs décisions.

4. https ://www.consilium.europa.eu/en/policies/pay-transparency/

4 RELATION ENTRE PRÉOCCUPATIONS ENVIRONNEMENTALES ET SOCIALES

Dans cette section, nous analysons la cohérence entre les préoccupations sociales des entreprises et leurs préoccupations environnementales. Pour ce faire, nous cherchons à relier les deux indicateurs principaux de la littérature sur ces sujets : le montant des écarts de rémunération femmes-hommes et le niveau d'émissions de CO2 des entreprises.

La Table 10 présente les corrélations partielles entre le niveau d'émissions par employé dans les différents scopes et la non-divulgation des écarts de salaire, en contrôlant une nouvelle fois l'effet du secteur d'activité, de la taille des entreprises, de l'année et de la région. Les scopes 1 et 2 ne sont pas corrélés avec la non-divulgation tandis que le scope 3 est positivement corrélé à un niveau 1%.

Les corrélations partielles sont identifiées via les effets marginaux d'un modèle Logit. Pour garantir la robustesse des résultats, nous présentons également les résultats d'un Probit et d'une régression linéaire pour les mêmes raisons que précédemment. Les conclusions restent inchangées dans ces cas, à l'exception de la corrélation avec les émissions scope 1, qui devient positive et significative à 1% dans la régression linéaire.

Ces résultats semblent indiquer un lien entre les différentes dimensions de la politique RSE, mais qui ne s'explique pas nécessairement par une volonté propre à l'entreprise. En effet, les politiques d'égalité femmes-hommes ne sont pas corrélées avec le niveau des émissions scopes 1 et 2, qui représentent les choix énergétiques des entreprises via leurs fournisseurs d'énergie et la qualité énergétique de leurs équipements par exemple. En revanche, ces politiques sont corrélés négativement avec les émissions du scope 3, qui sont davantage associés à l'environnement externe dans lequel s'insère l'entreprise. Ainsi, les entreprises qui ne divulguent pas leurs écarts salariaux ne s'intéressent pas à réduire leurs émissions plus que les autres, mais se trouvent dans une position où leurs fournisseurs et leurs clients sont plus respectueux de l'environnement, ce qui in fine amène des émissions plus faible dans le scope 3. Il peut s'agir d'une représentation de la théorie du signal : une entreprise ouvertement engagée (i.e qui divulgue ces données d'inégalités salariales) amène des clients et fournisseurs plus vertueux. Cette interprétation doit être relativisée au vu de la très faible magnitude de la corrélation partielle, qui pourrait simplement indiquer une absence de lien entre les préoccupations environnementales et sociales.

5 EXTENSION POSSIBLE : LA MODÉLISATION D'ÉVÉNEMENTS RARES

Plusieurs fois dans notre étude, nous faisons face à de l'estimation de paramètres dans un cadre que l'on pourrait qualifier d'événements rares. En effet, qu'il s'agisse de l'appartenance au groupe des entreprises exemplaires sur les émissions scope 1 et scope 3 en partie 3, ou à la divulgation des données d'inégalités salariales, nous travaillons avec des variables binaires qui sont très peu fréquemment égales à 1 par rapport au nombre d'observations total. Les modèles que nous avons utilisés pour représenter la relation entre ces variables binaires et les co-variables (Logit, Probit, Linéaire), ne sont pas forcément adaptés à la modélisation des événements rares. Nous avons pu en discuter, le grand nombre d'observations dans le groupe 0 (dans lequel la variable d'intérêt est égale à 0) peut avoir tendance à étaler la distribution, créant un biais d'atténuation qui minimise les coefficients obtenus pour la régression linéaire.

Dans ce cadre, on pourrait se pencher vers des modèles plus adaptés à cette situation, comme des lois d'extremums généralisées (GEV : Generalized Extreme Value). Cet ensemble de lois (Il y en a trois types différents, toutes pour des situations particulières.) permet de modéliser des phénomènes de valeurs extrêmes. Les grandes (ou petites) valeurs ont une probabilité plus grande d'apparaître que dans les distributions classiques. La pertinence de ces types de modèles dans le cas de variables d'intérêt binaires a déjà été discutée, par exemple dans un papier essayant d'estimer la probabilité qu'une entreprise adopte une technologie dont on lui aurait fait la promotion : Wang et Dey (2010). Dans cet article, les auteurs comparent les résultats de modèles Logit, Probit, Cloglog et GEV. Ils trouvent que pour l'estimation de la variable "L'entreprise adopte la technologie", qui arrive peu, le modèle GEV apporte de très bons résultats. Il serait intéressant d'étudier ce modèle dans le cadre de notre étude.

6 APPROCHES COMPLÉMENTAIRES

Dans cette section, nous présentons les approches que nous avons essayé de mettre en œuvre, mais qui se sont révélées infructueuses ou non-significatives en raison des données manquantes et de la particularité de l'année 2020. Les détails sur la construction des modèles et leurs résultats sont donnés en annexe 7.

6.1 Modèle Tobit

Pour poursuivre l'étude du lien entre les émissions et l'ampleur des écarts de salaire entre hommes et femmes, une approche intéressante serait d'identifier les corrélations avec l'étendue des écarts salariaux, nous nous adaptons pour cela des modèles Tobit développés dans l'étude de Cragg (1971).

Premier modèle : Données censurées

Dans un premier temps, nous considérons un modèle de données censurées tel que la variable observée est $Y_i = \min(s, Y_i^*)$ avec Y^* l'écart de rémunération théorique entre hommes et femmes, et s un seuil à partir duquel les entreprises ne révèlent plus l'écart de salaire. L'écart de rémunération théorique est alors :

$$Y_i^* = X_i'\beta_0 + \epsilon_i$$

Où l'on suppose que $\epsilon|Z$ suit une loi $N(0,1)$. Les paramètres de ce modèle peuvent être estimés par la méthode du maximum de vraisemblance.

Second modèle : Modèle de sélection

Dans un second temps, nous construisons un modèle de sélection où la variable observée est $Y_i = D_i Y_i^*$ avec Y^* l'écart de rémunération entre hommes et femmes théoriques et D_i une variable binaire qui vaut 1 si l'entreprise divulgue l'écart de rémunération femmes-hommes, tel que :

$$Y_i^* = X_i'\alpha_0 + \epsilon_i$$
$$D_i = 1_{Z_i'\gamma_0 + \nu_i < s}$$

Où l'on suppose que $\nu|Z$ suit une loi $N(0,1)$, que $E[\epsilon|\nu] = \delta_0\nu$, et que (ϵ, ν) est indépendant de (X, Z). Les paramètres de ce modèle peuvent être estimés via une procédure Heckit en deux étapes.

Identification du modèle pertinent

Nous n'avons pas de variable permettant de réaliser la procédure Heckit. Il reste possible de comparer la pertinence des deux modèles sous les hypothèses $Z_i = X_i$ et $\epsilon|X$ suit une loi $N(0,1)$, permettant de tester :

H_0 : le modèle est un modèle de données censurées.

H_1 : le modèle est un modèle de sélection.

Cet exercice de modélisation reste malheureusement théorique. En effet, le nombre trop important de données manquantes ne permet pas aux estimateurs Tobit d'être identifiables et donc d'être estimés.

6.2 Une possible exploitation de l'évolution temporelle pour identifier des effets causaux

Afin d'identifier l'effet causal de la variation des émissions scope 1 sur les émissions scope 3, nous avons recherché de possibles variables instrumentales. La plus probante est la capacité de l'entreprise à optimiser la consommation de sa production. En effet, si celle-ci est faible, il est plus difficile pour l'entreprise de réduire ses émissions scope 1 sans augmenter ses émissions scope 3 (condition de pertinence). Par ailleurs, la capacité d'optimisation de l'entreprise n'affecte pas celle de ses sous-traitants (condition d'exogénéité).

Cette variable instrumentale peut-être mesurée par l'élasticité de la consommation d'électricité au prix que l'on calcule avec la formule d'élasticité-prix simple ($e_{Q,P} = \frac{\Delta ln(Q)}{\Delta ln(P)}$) où Q est la consommation d'électricité et P son prix. Cette dernière information est accessible‘ sur le site de l'OCDE, où l'on trouve les prix de l'électricité par pays et par année.

La Table A.4 présente les résultats de la régression par variable instrumentale à deux étapes avec un contrôle par secteur. La première étape est l'estimation des variations des émissions scope 1 en fonction de l'élasticité, la seconde l'estimation des émissions scope 3 en fonction de la valeur estimée des émissions scope 1. Elle montre que la condition de pertinence est vérifiée : plus l'élasticité est grande, plus la variation des émissions scope 1 est importante. L'estimation de la corrélation entre les émissions montre qu'une variation positive d'émissions scope 1 est associée à une variation négative des émissions scope 3, toutes choses égales par ailleurs.

Cette approche demeure particulièrement fragile. En effet, nous sommes confrontés à un problème d'instrument faible, où la variation des émissions scope 1 est faiblement corrélée avec l'élasticité du prix de l'électricité. Cette faiblesse découle du fait que l'élasticité prix de l'électricité ne peut être calculée que pour la période antérieure à 2020. La pandémie de Covid-19 a en effet engendré une inversion de l'élasticité des prix entre 2019 et 2020, ainsi qu'une corrélation négative entre les émissions scope 1 et les émissions scope 3.

7 CONCLUSION

En résumé, nos analyses ne révèlent pas de relations claires entre les divers aspects de l'approche RSE adoptée par les entreprises. Tout d'abord, il semble que les entreprises ne cherchent pas à agir de manière coordonnée sur leurs préoccupations environnementales et sociales. En effet, les émissions scope 1 et scope 2 ne sont pas corrélées avec les écarts de rémunération femmes-hommes, soulignant ainsi une dissociation entre ces aspects de la RSE. Seules les émissions scope 3, qui reflètent davantage l'environnement global dans lequel opère l'entreprise, semblent jouer un rôle dans la relation entre action écologique et inégalités de revenus. De plus, lorsque nous examinons spécifiquement la cohérence entre les différentes dimensions des préoccupations sociales, nous observons une grande disparité dans les performances des entreprises selon les indicateurs. À ce titre, le nombre d'heures de formation par employé est corrélé négativement avec l'attention portée aux inégalités de revenus entre hommes et femmes. Enfin, la seule dimension où une certaine cohérence est observée est la dimension environnementale. Certaines entreprises se distinguent par leur engagement envers la réduction des émissions scope 1 et scope 3, ainsi que par leur gestion responsable de l'énergie et de l'eau, entre autres critères environnementaux. Ces conclusions doivent être considérées avec prudence en raison du nombre important de données manquantes qui limite la mise en œuvre de méthodes plus approfondies pour confirmer ces résultats. En particulier, nous aurions souhaité étudier plus en profondeur le lien entre les émissions et l'ampleur de l'écart salarial femmes-hommes, à l'aide de modèles Tobit.

Cette absence de liens souligne l'importance de réfléchir aux caractéristiques souhaitables des indicateurs ESG qui intègrent ces préoccupations. Il est essentiel de sanctionner l'absence de complémentarité pour encourager les entreprises à agir de manière concertée sur toutes les dimensions afin d'améliorer leur score global. À ce titre, l'utilisation de la moyenne géométrique plutôt que de la moyenne arithmétique pourrait être envisagée. Ce travail contribue donc à la réflexion sur la construction d'un indicateur agrégé optimal, un travail qui doit se poursuivre et être approfondi. Comme le rappelle Shakil (2021), il existe plus de 140 indicateurs ESG différents, développés par des acteurs poursuivant des objectifs variés. Il est crucial de développer une mesure unique et standardisée ECG, par exemple à l'échelle de l'UE, qui faciliterait la comparaison entre les entreprises et stimulerait une dynamique d'amélioration continue.

Table 7 – Régressions de l'appartenance au groupe délocalisateur sur les caractéristiques des entreprises

		Logit	Probit	Linéaire
	Nombre d'employés	1.947e-06***	1.996e-06***	1.799e-06***
		(2.3e-07)	(2.46e-07)	(2.86e-07)
Régions	Afrique / Moyen-Orient	-0.2700***	-0.2577***	-0.2854***
		(0.019)	(0.024)	(0.016)
	Asie / Pacifique (sans Chine)	-0.0147**	-0.0149***	-0.0163**
		(0.006)	(0.005)	(0.007)
	Chine	-0.0983***	-0.0951***	-0.0920***
		(0.002)	(0.004)	(0.002)
	Europe	0.0797**	0.0812**	0.0778**
		(0.039)	(0.041)	(0.034)
	Amérique latine et Caraïbes	0.0599	0.0714	0.0630
		(0.079)	(0.069)	(0.082)
Années	2019	-0.0390***	-0.0446***	-0.0318**
		(0.015)	(0.017)	(0.016)
	2020	-0.0050	-0.0072	-0.0013
		(0.012)	(0.011)	(0.009)
Secteurs	Services de communication	0.1722***	0.1678**	0.1931***
		(0.065)	(0.067)	(0.068)
	Biens non-essentiels	0.1571***	0.1560***	0.1589***
		(0.048)	(0.050)	(0.043)
	Biens de consommation courante	0.0627***	0.0552***	0.0678**
		(0.022)	(0.021)	(0.032)
	Énergie	-0.1199	-0.1296	-0.1467
		(0.096)	(0.095)	(0.106)
	Secteur financier	0.2063***	0.2020***	0.2234***
		(0.067)	(0.069)	(0.059)
	Santé	-0.0763	-0.0863	-0.1110
		(0.138)	(0.147)	(0.169)
	Biens industriels	-0.0586***	-0.0656***	-0.0750***
		(0.018)	(0.017)	(0.025)
	Commerce, transport	-0.3576***	-0.3575***	-0.3735***
		(0.064)	(0.065)	(0.061)
	Technologie de l'information	0.0160	0.0096	0.0038
		(0.131)	(0.140)	(0.159)
	Matériaux	-0.0657*	-0.0727**	-0.0919**
		(0.035)	(0.036)	(0.043)
	Services publics	-0.4157***	-0.4096***	-0.3888***
		(0.054)	(0.049)	(0.047)

Régressions de la variable binaire d'appartenance au groupe 1 (scope 1 faible, scope 3 élevé) : effets marginaux moyens pour les régressions Logit et Probit, et coefficients pour la régression linéaire.
Erreurs standards (avec regroupement par entreprises) entre parenthèses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Les catégories de référence utilisées sont : États-Unis et Canada pour la région, Immobilier pour le secteur, 2018 pour l'année. Nous nous intéressons aux effets marginaux moyens pour les régressions Logit et Probit, et non pas aux coefficients des régressions car ces derniers ne fournissent pas les corrélations partielles.

Table 8 – Régressions de l'appartenance au groupe des entreprises exemplaires (ayant émissions scope 1 et scope 3 faibles) sur les autres variables environnementales

		Logit	Probit	Linéaire
	Nombre d'employés	3.948e-07***	3.798e-07***	5.163e-07**
		(5.81e-08)	(6.72e-08)	(2.1e-07)
	Production de déchets par employé	4.445e-06***	4.389e-06***	2.722e-07
		(6.13e-07)	(5.83e-07)	(2.02e-07)
	Recyclage des déchets par employé	-0.0001***	-0.0001***	-7.462e-07***
		(1.62e-05)	(2.21e-05)	(6.24e-08)
	Consommation d'eau par employé	-3.604e-06***	-3.493e-06***	-1.428e-07***
		(8.09e-07)	(6.31e-07)	(5.22e-08)
	Prélèvement d'eau par employé	-1.432e-06***	-1.397e-06***	-3.482e-08
		(2.2e-07)	(1.62e-07)	(2.27e-08)
	Consommation d'énergie par employé	-3.548e-06***	-4.022e-06***	5.636e-09**
		(6.85e-07)	(7.3e-07)	(2.22e-09)
Régions	Afrique / Moyen-Orient	0.2386***	0.2404***	0.4377***
		(0.028)	(0.027)	(0.035)
	Asie / Pacifique (sans Chine)	0.1158***	0.1181***	0.1978***
		(0.024)	(0.025)	(0.029)
	Chine	-0.0084	-0.0023	-0.0329
		(0.016)	(0.010)	(0.034)
	Europe	-0.0305	-0.0299	-0.0207
		(0.043)	(0.045)	(0.048)
	Amérique latine et Caraïbes	0.2770***	0.2743***	0.4832***
		(0.029)	(0.029)	(0.029)
Années	2019	0.0055***	0.0044***	0.0011
		(0.000)	(0.001)	(0.001)
	2020	0.1464***	0.1428***	0.1108***
		(0.002)	(0.005)	(0.001)
Secteurs	Services de communication	0.0807***	0.0819***	0.0883*
		(0.016)	(0.020)	(0.052)
	Biens non-essentiels	-0.0289	-0.0168	-0.0273
		(0.025)	(0.030)	(0.027)
	Biens de consommation courante	-0.0127***	-0.0100***	-0.0151***
		(0.003)	(0.003)	(0.004)
	Énergie	-0.1473***	-0.1476***	-0.0945***
		(0.003)	(0.005)	(0.031)
	Secteur financier	0.2344***	0.2425***	0.3469***
		(0.018)	(0.015)	(0.048)
	Santé	0.0675***	0.0716***	0.0635***
		(0.014)	(0.014)	(0.016)
	Biens industriels	-0.0121	-0.0125	-0.0130*
		(0.012)	(0.012)	(0.008)
	Commerce, transport	0.0273***	0.0337***	0.0275*
		(0.004)	(0.006)	(0.016)
	Technologie de l'information	0.0964***	0.1046***	0.1236***
		(0.020)	(0.022)	(0.015)
	Matériaux	-0.0492***	-0.0373***	-0.0426***
		(0.002)	(0.002)	(0.016)
	Services publics	-0.0629***	-0.0513***	-0.0548*
		(0.004)	(0.005)	(0.030)

Régressions de la variable binaire d'appartenance au groupe des entreprises exemplaires sur les émissions scope 1 et scope 3 : effets marginaux moyens pour les régressions Logit et Probit, et coefficients pour la régression linéaire.
Erreurs standards (avec regroupement par entreprises) entre parenthèses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Les catégories de référence utilisées sont : États-Unis et Canada pour la région, Immobilier pour le secteur, 2018 pour l'année. Nous nous intéressons aux effets marginaux moyens pour les régressions Logit et Probit, et non pas aux coefficients des régressions car ces derniers ne fournissent pas les corrélations partielles.

Table 9 – Régressions de la non-divulgation des écarts de salaire sur les autres proxies des préoccupations sociales

		Logit	**Probit**	**Linéaire**	**Logit Winsorisé**
	Heures de formation par employé	3.189e-05***	2.965e-05***	3.003e-06***	3.228e-05***
		(1.21e-05)	(1.07e-05)	(7.26e-07)	(1.16e-05)
	Membres indépendants du	-0.0533***	-0.0557***	-0.0541***	-0.0629***
	conseil d'administration (%)	(0.006)	(0.008)	(0.012)	(0.019)
	Frais de justice payés par employé	0.0616	0.0560	0.0103*	0.0598
		(0.038)	(0.041)	(0.004)	(0.037)
	Nombre d'employés	-1.898e-08	-3.643e-08	-5.114e-08	-1.803e-08
		(4.01e-08)	(4.43e-08)	(4.51e-08)	(3.91e-08)
Régions	Afrique / Moyen-Orient	-0.0521	-0.0464	-0.0333	-0.0534
		(0.035)	(0.035)	(0.034)	(0.034)
	Asie / Pacifique (sans Chine)	0.1056***	0.0891***	0.0368*	0.1034***
		(0.030)	(0.027)	(0.016)	(0.033)
	Chine	0.1274***	0.1060***	0.0368**	0.1254***
		(0.036)	(0.029)	(0.015)	(0.040)
	Europe	-0.1060***	-0.0987***	-0.1009***	-0.1069***
		(0.021)	(0.022)	(0.031)	(0.022)
	Amérique latine et Caraïbes	-0.0261	-0.0209	-0.0174	-0.0272
		(0.049)	(0.046)	(0.035)	(0.052)
Années	2019	-0.0277***	-0.0280***	-0.0223***	-0.0280***
		(0.001)	(0.001)	(0.000)	(0.001)
	2020	-0.0757***	-0.0752***	-0.0746***	-0.0760***
		(0.001)	(0.002)	(0.004)	(0.001)
Secteurs	Services de communication	0.0063	0.0048	-0.0015	0.0063
		(0.048)	(0.051)	(0.057)	(0.047)
	Biens non-essentiels	0.0317	0.0287	0.0317	0.0309
		(0.043)	(0.045)	(0.044)	(0.043)
	Biens de consommation courante	-0.0008	-0.0014	-0.0043	-0.0016
		(0.017)	(0.018)	(0.017)	(0.017)
	Énergie	0.0858	0.0836	0.0652	0.0845
		(0.055)	(0.053)	(0.035)	(0.053)
	Secteur financier	0.0923***	0.0849***	0.0768***	0.0922***
		(0.021)	(0.021)	(0.017)	(0.020)
	Santé	0.0441	0.0431	0.0461	0.0439
		(0.046)	(0.045)	(0.043)	(0.045)
	Biens industriels	0.0400	0.0395	0.0379	0.0396
		(0.046)	(0.048)	(0.044)	(0.045)
	Commerce, transport	0.0278	0.0259	0.0243	0.0274
		(0.015)	(0.016)	(0.015)	(0.014)
	Technologie de l'information	0.0272	0.0259	0.0299	0.0272
		(0.015)	(0.016)	(0.016)	(0.014)
	Matériaux	0.0357***	0.0316	0.0326**	0.0378**
		(0.013)	(0.016)	(0.013)	(0.012)
	Services publics	-0.0266	-0.0239	-0.0396	-0.0251
		(0.026)	(0.028)	(0.023)	(0.023)

Régressions de la variable binaire de non-divulgation des écarts de remunéraion femmes-hommes : effets marginaux moyens pour les régressions Logit et Probit, et coefficients pour la régression linéaire. Régression Logit (winsorisé) présénte les effets marginaux avec winsorisation des variables heures de formation par employé, Membres indépendant du conseil d'administration et Frais de justice payés par employé au 5ème et 95ème centile.
Erreurs standards (avec regroupement par entreprises) entre parenthèses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Les catégories de référence utilisées sont : États-Unis et Canada pour la région, Immobilier pour le secteur, 2018 pour l'année. Nous nous intéressons aux effets marginaux moyens pour les régressions Logit et Probit, et non pas aux coefficients des régressions car ces derniers ne fournissent pas les corrélations partielles.

Table 10 – Régressions de la non-divulgation des écarts de rémunération sur le niveau des émissions dans les différents scopes

		Logit	**Probit**	**Linéaire**
	Scope 1 par employé	3.303e-06	4.674e-06	2.639e-07***
		(3.34e-06)	(3.64e-06)	(2.17e-08)
	Scope 2 par employé	1.441e-05	1.263e-05	6.1e-08
		(1.24e-05)	(8.44e-06)	(4.64e-07)
	Scope 3 par employé	1.482e-07***	1.569e-07***	2.453e-07***
		(2.22e-08)	(6.14e-08)	(8.18e-08)
	Nombre d'employés	-4.478e-08***	-5.306e-08***	-8.732e-08***
		(8.39e-09)	(1.27e-08)	(1.23e-08)
Régions	Afrique / Moyen-Orient	-0.0228**	-0.0186*	-0.0141
		(0.009)	(0.011)	(0.011)
	Asie / Pacifique (sans Chine)	0.0178***	0.0143***	0.0058***
		(0.004)	(0.002)	(0.001)
	Chine	0.0689***	0.0495***	0.0125***
		(0.009)	(0.007)	(0.003)
	Europe	-0.0670***	-0.0620***	-0.0947***
		(0.010)	(0.009)	(0.019)
	Amérique latine et Caraïbes	-0.0189***	-0.0148***	-0.0094***
		(0.002)	(0.002)	(0.002)
Secteurs	Services de communication	-0.0057	-0.0052*	-0.0119***
		(0.003)	(0.003)	(0.004)
	Biens non-essentiels	0.0097*	0.0090*	0.0097*
		(0.004)	(0.004)	(0.004)
	Biens de consommation courante	0.0079	0.0089	0.0073
		(0.008)	(0.008)	(0.008)
	Énergie	-0.0009	-0.0024	3.187e-06
		(0.006)	(0.006)	(0.006)
	Secteur financier	0.0144*	0.0145*	0.0144*
		(0.006)	(0.006)	(0.006)
	Santé	0.0411***	0.0388***	0.0260***
		(0.012)	(0.009)	(0.008)
	Biens industriels	0.0219***	0.0229***	0.0219***
		(0.007)	(0.006)	(0.008)
	Transport commercial	-0.0009	-0.0011	-0.0034
		(0.004)	(0.004)	(0.005)
	Technologie de l'information	0.0083***	0.0083***	0.0083***
		(0.001)	(0.001)	(0.002)
	Matériaux	0.0019	0.0003	0.0024
		(0.004)	(0.004)	(0.005)
	Services publics	-0.0239***	-0.0271***	-0.0440***
		(0.005)	(0.005)	(0.017)

Régressions de la variable binaire de divulgation des données d'inégalités salariales, effets marginaux moyens pour les régressions Logit et Probit, et coefficients pour la régression linéaire.
Erreurs standards (avec regroupement par entreprises) entre parenthèses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Les catégories de référence utilisées sont : États-Unis et Canada pour la région, Immobilier pour le secteur, 2018 pour l'année. Nous nous intéressons aux effets marginaux moyens pour les régressions Logit et Probit, et non pas aux coefficients des régressions car ces derniers ne fournissent pas les corrélations partielles.

ANNEXES

Détails concernant le modèle Tobit

Premier modèle : Données censurées

L'écart de rémunération théorique est :

$$Y_i^* = X_i'\beta_0 + \epsilon_i$$

et la la variable observée est $Y_i = \min(s, Y_i^*)$ avec s un seuil à partir duquel les entreprises ne révèlent plus l'écart de salaire.

On suppose que X_i regroupe les caractéristiques de l'entreprise et donc nos variables d'intérêt, et que $\epsilon|X$ suit une loi $N(0,1)$.

$$\begin{aligned} E[Y|X, Y<s] &= E[Y^*|X, Y^*<s] \\ &= E[X'\beta_0 + \epsilon|X, Y^*<s] \\ &= X'\beta_0 + E[\epsilon|X, X'\beta_0 + \epsilon < s] \\ &= X'\beta_0 + \frac{E[\epsilon 1_{\epsilon \leq s - X'\beta_0}|X]}{P(\epsilon \leq s - X'\beta_0|X)} \end{aligned}$$

En faisant rentrer s dans la constante de β_0 :

$$X'\beta_0 + \frac{E[\epsilon 1_{\epsilon \leq -X'\beta_0}|X]}{P(\epsilon \leq -X'\beta_0|X)}$$

En simplifiant :

$$E[Y|X, Y<s] = X'\beta_0 - \frac{\phi(-X'\beta_0)}{\Phi(-X'\beta_0)}$$

Dans ce premier modèle, Y^* est déterminé par les caractéristiques de l'entreprise, alors que s est un seuil exogène qui ne dépend pas de ceux-ci. Ainsi, le même paramètre β_0 détermine à la fois la décision de révéler les écarts de rémunération ($Y = 0$ par rapport à $Y > 0$) , et le montant de ceux-ci.

Dans ce modèle, les paramètres peuvent être estimés par la méthode du maximum des moments, ce qui permet de calculer les effets marginaux de nos variables d'intérêt, notamment le niveau des émissions même s'il n'est pas possible d'identifier séparément s et la constante du modèle Y^*.

Second modèle : Modèle de sélection

Dans ce modèle, nous observons $Y_i = D_i Y_i^*$ avec Y^* l'écart de rémunération entre hommes et femmes théoriques et D_i une variable binaire qui vaut 1 si l'entreprise divulgue l'écart de rémunération femmes-hommes, tel que

$$\begin{aligned} Y_i^* &= X_i'\alpha_0 + \epsilon_i \\ D_i &= 1_{Z_i'\gamma_0 + \nu_i < s} \end{aligned}$$

on a également $\nu|Z$ suivant une loi $N(0,1)$, $E[\epsilon|\nu] = \delta_0\nu$, et (ϵ, ν) indépendant de (X, Z).

En suivant les mêmes étapes que précédemment, on trouve :

$$\begin{aligned} E[Y|X,Z,D=1] &= X'\alpha_0 + \delta_0 \frac{E[\nu 1_{\nu \leq -Z'\gamma_0}|X]}{P(\nu \leq -Z'\gamma_0|X)} \\ &= X'\alpha_0 - \delta_0 \lambda(-Z'\gamma_0) \end{aligned}$$

avec $\lambda(-Z'\gamma_0)$ le ratio inverse de Mills.

Dans ce second modèle, le fait de divulguer n'est plus lié à un seuil exogène, mais bien déterminé par un processus de sélection séparé qui ne dépend plus de α_0 mais de γ_0

Ce modèle peut être estimé via une procédure Heckit en deux étapes :

1. Régression Probit de D_i sur Z_i qui permet d'estimer γ_0 et donc d'obtenir une estimation de $\lambda(Z_i'\hat{\gamma_0})$
2. Régression linéaire de Y_i sur $(X_i, \lambda(Z_i'\hat{\gamma_0}))$ pour i tel que $D_i = 1$ pour obtenir une estimation de δ_0 et α_0

Identification du modèle pertinent

La situation optimale serait d'avoir un instrument pour réaliser la procédure Heckit. Malheureusement, aucune variable dans notre jeu de données n'explique la divulgation des écarts de rémunération sans expliquer également le montant de ceux-ci. Dans le cas $Z_i = X_i$, le modèle est identifiable car $\lambda(\cdot)$ est un opérateur non-linéaire mais l'estimation est peu robuste car l'hypothèse est particulièrement forte.

Sous les hypothèses $Z_i = X_i$ et $\epsilon|X$ suit une loi $N(0,1)$, il est cependant possible de comparer la pertinence de ces deux modèles. En effet, dans ce cas, le modèle de données censurées est un cas particulier de notre modèle de sélection dans lequel le même paramètre (β_0) détermine à la fois la décision de révéler les écarts de rémunération et le montant de ceux-ci. Ainsi, il est possible de construire un test du rapport de vraisemblance qui compare la log-vraisemblance du modèle contraint (modèle de données censurées) et la log-vraisemblance du modèle non-contraint (modèle de sélection) pour séparer H_0 et H_1 où :

H_0 : le modèle est un modèle de données censurées.

H_1 : le modèle est un modèle de sélection.

On rejette H_0 si :

$$2\left(\ln\left(L(Y|X;\text{Selection})\right) - \ln\left(L(Y|X;\text{Censure})\right)\right) > q_{95\%}^{\chi^2(d)}$$

où $L(Y|X;\text{Selection})$ est la vraisemblance du modèle de sélection, et $L(Y|X;\text{Censure})$ est la vraisemblance du modèle de censure. $q_{95\%}^{\chi^2(d)}$ est le quantile à 95% de la distribution chi-carré avec d degrés de liberté.

production.

Tableaux complémentaires

Table A.1 – Régressions de la non-divulgation des écarts de salaire sur les caractéristiques de l'entreprise

		Logit	**Probit**	**Linéaire**	**Trimmed Logit**
	Revenus	-1.396e-07***	-1.428e-07***	-3.447e-07***	-1.364e-07***
		(6.26e-09)	(5.37e-09)	(7.96e-09)	(7.63e-09)
	Nombre d'employés	2.756e-09***	-4.273e-09	-1.703e-08	2.033e-09**
		(9.05e-10)	(2.91e-09)	(2.05e-08)	(8.95e-10)
Régions	Afrique / Moyen-Orient	-0.0049*	-0.0044	-0.0034*	-0.0050*
		(0.003)	(0.003)	(0.001)	(0.003)
	Asie / Pacifique (sans Chine)	0.0055**	0.0041**	0.0021	0.0061**
		(0.002)	(0.002)	(0.001)	(0.002)
	Chine	0.0375***	0.0261***	0.0046**	0.0397***
		(0.003)	(0.002)	(0.002)	(0.003)
	Europe	-0.0293***	-0.0270***	-0.0563**	-0.0305***
		(0.002)	(0.003)	(0.024)	(0.002)
	Amérique latine et Caraïbes	-0.0048***	-0.0038***	-0.0022*	-0.0046***
		(0.001)	(0.001)	(0.001)	(0.001)
Années	2019	-0.0045***	-0.0046***	-0.0030***	-0.0047***
		(1.07e-05)	(0.000)	(8.55e-05)	(1.06e-05)
	2020	-0.0148***	-0.0142***	-0.0146***	-0.0154***
		(1.24e-05)	(0.000)	(0.000)	(1.27e-05)
Secteurs	Services de communication	-0.0032	-0.0028	-0.0049	0.0036
		(0.002)	(0.002)	(0.004)	(0.002)
	Biens non-essentiels	8.214e-05	0.0003	0.0002	0.0072***
		(0.001)	(0.001)	(0.001)	(0.001)
	Biens de consommation courante	0.0012***	0.0019***	0.0014**	0.0081***
		(0.000)	(0.000)	(0.001)	(0.000)
	Énergie	0.0016	0.0002	0.0016	0.0069***
		(0.002)	(0.002)	(0.002)	(0.002)
	Secteur financier	0.0048***	0.0046***	0.0053*	0.0118***
		(0.001)	(0.001)	(0.002)	(0.000)
	Santé	0.0079***	0.0079***	0.0066***	0.0133***
		(0.001)	(0.001)	(0.002)	(0.002)
	Biens industriels	0.0059*	0.0065**	0.0056***	0.0129**
		(0.004)	(0.003)	(0.001)	(0.004)
	Commerce, transport	0.0008	0.0011	0.0012	0.0076***
		(0.001)	(0.001)	(0.001)	(0.001)
	Technologie de l'information	-0.0002	-0.0006	0.0004	0.0067***
		(0.002)	(0.003)	(0.002)	(0.002)
	Matériaux	0.0002	-2.814e-05	0.0004	0.0073***
		(0.002)	(0.002)	(0.002)	(0.002)
	Services publics	-0.0091***	-0.0097***	-0.0173***	-0.0038*
		(0.002)	(0.002)	(0.001)	(0.002)

Régressions de la variable binaire de non-divulgation des écarts salariaux femmes-hommes : effets marginaux moyens pour les régressions Logit et Probit, et coefficients pour la régression linéaire. Trimmed Logit présente les effets marginaux sur les entreprises ayant un nombre d'employés supérieur au 5% centile. Erreurs standards (avec regroupement par entreprises) entre parenthèses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Les catégories de référence utilisées sont : États-Unis et Canada pour la région, Immobilier pour le secteur, 2018 pour l'année. Nous nous intéressons aux effets marginaux moyens pour les régressions Logit et Probit, et non pas aux coefficients des régressions car ces derniers ne fournissent pas les corrélations partielles.

Table A.2 – Pourcentage des observations par sous-secteur

Secteur	**% des Observations**
Produits pharmaceutiques	5.92
Matériel et équipement technologique	5.68
Produits chimiques	5.64
Banques	5.16
Machinerie	4.92
Matériaux de construction et minage	4.74
Alimentation et tabac	4.63
Logiciels et services	4.51
Construction et aérospatial	4.08
Services financiers	3.99
Immobilier	3.95
Commerces de proximité	3.86
Équipements de santé	3.79
Énergie	3.48
Services publics	3.39
Commerces professionnels	3.34
Transports	3.21
Appareil ménagers	3.21
Médias et divertissement	3.06
Équipements électriques	2.84
Services consommation	2.83
Semiconducteurs	2.66
Conglomérat de distributeurs	1.82
Assurance	1.77
Services de télécomunnication	1.37
Automobile	1.31
Papiers et emballages	1.19

Table A.3 – Description des variables avant winsorisation et trimming

Variable	Min	Max	Mean	Mediane	5eme centile	95eme centile
Nombre d'employés	1.00	2.3e6	13142.4	3049.0	192.0	53000.0
Nombre d'heures de formation	0.00	1.5e8	983245.7	144885.5	15.0	4.11e6
Frais de justice payés	0.00	1.4e4	34.6	0.0	0.0	84.5
Membres indépendants du conseil d'administration (%)	0.00	1.00	0.6	0.6	0.1	1.00

Table A.4 – Régression par variable instrumentale de la variation de scope 3 sur la variation de scope 1

	Variable explicative	**Coefficient**	**Variable dépendante**
Première étape	Élasticité prix de la consommation d'électricité	0.005	Variation des émissions scope 1
		(0.004)	
	Services de communication	-0.1349***	
		(0.019)	
	Biens non essentiels	0.0940***	
		(0.013)	
	Biens de consommation courante	0.0422***	
		(0.012)	
	Énergie	-0.0780	
		(0.019)	
	Santé	-0.0055	
		(0.016)	
	Biens industriels	0.0898***	
		(0.012)	
	Commerce, Transport	0.0789***	
		(0.014)	
	Technologie de l'information	0.0282**	
		(0.012)	
	Matériaux	0.0369***	
		(0.012)	
	Immobilier	-0.0654**	
		(0.035)	
	Services publics	-0.0953***	
		(0.014)	
Seconde étape	Variation des émissions scope 1 prédites	-0.407	Variation des émissions scope 3
		(0.900)	
	Services de communication	0.4260***	
		(0.122)	
	Biens non essentiels	0.3030***	
		(0.087)	
	Biens de consommation courante	0.1428***	
		(0.042)	
	Énergie	0.0472	
		(0.076)	
	Santé	-0.1227**	
		(0.054)	
	Biens industriels	0.2285***	
		(0.083)	
	Commerce, Transport	0.2730***	
		(0.083)	
	Technologie de l'information	0.2458***	
		(0.078)	
	Matériaux	0.1369***	
		(0.012)	
	Immobilier	-0.0396	
		(0.074)	
	Services publics	-0.3366***	
		(0.088)	

Régression par variable instrumentale de la variation des emissions scope 3 sur la variation des émissions scope 1 contrôlée par secteurs : coefficients pour la régression linéaire. Le secteur de référence utilisé est : Secteur financier.
Erreurs standards entre parenthèses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

RÉFÉRENCES

Austin, Chelsea Rae, Donna D Bobek et Ling L Harris (2021). "Does information about gender pay matter to investors ? An experimental investigation". In : *Accounting, Organizations and Society* 90, p. 101193.

Blundell, Jack (2020). "Wage Responses to Gender Pay Gap Reporting Requirements". In : *SSRN Electronic Journal*.

Burkhardt, Kirsten, Pascal Nguyen et Evelyne Poincelot (2020). "Agents of change : Women in top management and corporate environmental performance". In : *Corporate Social Responsibility and Environmental Management* 27(4), p. 1591-1604.

Cragg, John G (1971). "Some statistical models for limited dependent variables with application to the demand for durable goods". In : *Econometrica : journal of the Econometric Society*, p. 829-844.

Flammer, Caroline (2013). "Corporate social responsibility and shareholder reaction : The environmental awareness of investors". In : *Academy of Management journal* 56(3), p. 758-781.

ILO (2018). *Global Wage Report 2018/19 : What lies behind gender pay gaps*. International Labour Office.

OECD (2021). *Pay Transparency Tools to Close the Gender Wage Gap*, p. 134.

Shakil, Mohammad Hassan (2021). "Environmental, social and governance performance and financial risk : Moderating role of ESG controversies and board gender diversity". In : *Resources Policy* 72, p. 102144.

Wang, Xia et Dipak K. Dey (2010). "Generalized extreme value regression for binary response data : An application to B2B electronic payments system adoption". In : *The Annals of Applied Statistics* 4(4), p. 2000-2023.