

LE MODELE PROMETHEUS DE MICROSIMULATION DE LA FACTURE ENERGETIQUE DES MENAGES

Alexandre GODZINSKI (*), Tristan LOISEL (*), Élise MARTIN (*), Clémentine MINS (*),
Mathilde NIAY (*), Quentin NICOD (*), Lucie NEIRAC (*),
Caroline PINTON (*), Bruno QUILLE (*), Laouen RYO LE ROUX (*)

(*) CGDD, Service de l'économie verte et solidaire

alexandre.godzinski@developpement-durable.gouv.fr

tristan.loisel@developpement-durable.gouv.fr

clementine.mins@developpement-durable.gouv.fr

lucie.neirac@developpement-durable.gouv.fr

laouen.ryo-le-roux@developpement-durable.gouv.fr

Mots-clés : Facture énergétique, ménages, logement, transport, fiscalité énergétique, transition énergétique

Domaine concerné : Microsimulation

Résumé

Le modèle de microsimulation Prometheus (PROgramme de Microsimulation des Énergies du Transport et de l'Habitat pour ÉvalUations Sociales) estime la consommation énergétique des ménages résidant en France métropolitaine pour leur logement et pour leurs déplacements en voiture. La microsimulation permet d'obtenir ces estimations à un niveau fin de granularité, permettant des analyses selon diverses dimensions socio-économiques, par exemple pour chaque dixième de niveau de vie, par type de commune (au sens du zonage en aire d'attraction des villes), par énergie principale de chauffage ou selon les équipements en véhicules. Des croisements entre dimensions sont également possibles.

Sur son champ d'étude (facture énergétique complète des ménages et grandeurs associées), le modèle est le plus précis existant en France. Il est mobilisé notamment :

- pour réaliser des estimations descriptives de la facture énergétique des ménages, de leur taux d'effort énergétique, de leur consommation d'énergie et de leurs émissions directes de CO₂ ;
- pour estimer chaque année le taux de précarité énergétique (nombre de ménages en situation de précarité énergétique au sens du taux d'effort énergétique) ;
- pour évaluer l'impact de dispositifs d'aide tels que le chèque énergie ;
- pour simuler l'impact sur la facture des ménages de variations ex post ou ex ante des prix de l'énergie ou de la fiscalité énergétique. Ce fut le cas lors des évaluations du bouclier tarifaire en 2022.

Ce document de travail porte sur la version 3 du modèle, où l'Enquête nationale logement 2013 (ENL 2013) de l'Insee est appariée avec l'Enquête mobilité des personnes 2019 (EMP 2019) du SDES (encadré). Dans la première version du modèle développée en 2016, seule l'ENL 2006 était exploitée et le champ du modèle se limitait donc à la facture énergétique pour le logement. La version 2, mobilisait l'ENL 2013

et l'Enquête nationale transport et déplacement (ENTD 2008), conduisant donc à une estimation de la facture énergétique complète des ménages. Cette version 2 n'a pas fait l'objet d'un document de travail. Cette troisième version substitue l'EMP 2019 à l'ENTD 2008, et l'ensemble du modèle est désormais codé avec le langage de programmation R, le langage SAS, utilisé pour les versions 1 et 2 ayant été écarté.

Ce document de travail poursuit deux objectifs. D'une part, il développe les choix méthodologiques sur lesquels le modèle s'appuie et vise à servir de référence à laquelle renvoyer pour l'ensemble des travaux conduits désormais avec le modèle. D'autre part, il présente un cahier de variantes, permettant d'illustrer les résultats canoniques du modèle pour différents chocs sur les prix de l'énergie, sur la fiscalité, et sur les mesures d'accompagnement. Ces résultats sont présentés pour l'année 2019, dernière année disponible au cours de laquelle l'énergie des ménages n'a pas fait l'objet d'évolution notable.

Le cahier de variantes présente notamment l'impact, sur la facture énergétique des ménages et le taux d'effort énergétique, de hausses des prix de l'énergie (carburants, gaz, électricité), de hausses de la fiscalité énergétique (composante carbone, droit d'accise sur l'électricité, TVA sur l'énergie) et des versements de chèques énergie (chèque énergie classique et augmenté d'un chèque additionnel). Les résultats mettent notamment en évidence le caractère régressif du renchérissement des énergies : si la facture en euros augmente davantage pour les ménages aisés que pour les ménages modestes, du fait des volumes consommés, cette hausse pèse davantage sur le revenu des ménages modestes ; cela vaut notamment dans le cas d'une hausse des carburants. Le versement d'aides ciblées telles que le chèque énergie contribue à alléger significativement la facture énergétique des ménages les plus modestes.

Abstract

The Prometheus microsimulation model estimates the energy consumption of households living in metropolitan France, accounting for both housing and car travel. It enables a quantitative evaluation of how changes in energy prices, energy taxation, and policy support instruments (such as the energy voucher) affect households in terms of energy expenditures, energy burden, and energy poverty. These effects can be evaluated both at the aggregate level and across a range of socio-economic dimensions, including household income level, main heating energy source, car ownership, and city type.

The third version of the model has recently been finalized and now integrates data from the 2019 Personal Mobility Survey, thereby updating the representation of household fuel expenditures. Given that the previous and only documentation of the model dated from 2016, this new version provides the opportunity to release an updated working paper. The aim of this publication is to present a comprehensive and detailed description of the model's structure, functioning, and to report its principal benchmark results.



MINISTÈRES
AMÉNAGEMENT
DU TERRITOIRE
TRANSITION
ÉCOLOGIQUE

*Liberté
Égalité
Fraternité*

DOCUMENT DE TRAVAIL

Le modèle Prometheus de microsimulation de la facture énergétique des ménages

Méthodologie et résultats

Mars 2025

Résumé

Les ménages font face à des dépenses énergétiques pour se loger et pour se déplacer. Ces factures d'énergie sont très hétérogènes et dépendent de multiples déterminants socio-économiques. Le modèle de microsimulation de la facture énergétique, intitulé « PROgramme de Microsimulation des Énergies du Transport et de l'Habitat pour ÉvalUation Sociales » ou « Prometheus », est développé au Commissariat général au développement durable (CGDD) et a bénéficié de nombreux enrichissements au cours du temps. Il s'appuie sur des enquêtes et données de l'Insee et du ministère de la Transition écologique. Le champ est celui des ménages ordinaires vivant en France métropolitaine. Les dépenses d'énergie du logement sont celles des résidences principales des ménages, les dépenses de carburants sont celles des véhicules des ménages. Ce document de travail présente la version actuelle du modèle, une photographie de la situation des ménages en 2019, ainsi qu'un cahier de variantes. Cette connaissance fine des niveaux de facture énergétique et des effets de chocs permet d'éclairer la conception de politiques publiques de transition énergétique socialement acceptables.

Mots clés : *facture énergétique, ménages, logement, transport, prix de l'énergie, fiscalité énergétique, transition énergétique.*

Remerciements

Nous remercions les membres de l'équipe précédente de Prometheus qui ont développé la *version 2* du modèle Prometheus, Mathilde Clément, Camille Parent, Charles Peroumal, Annaïck Rolland, Augustin Vicard, ainsi que ceux ayant contribué au développement de la version actuelle, Hana Bouhalli, Kendrick Herzberg, Vincent Marcus. Nous remercions également Olivier Simon et Mathieu Fouquet pour leur relecture attentive.

Citation pour ce document : Godzinski, A. *et al.* (2025), « Le modèle Prometheus de microsimulation de la facture énergétique des ménages », CGDD, document de travail, mars 2025.

Auteurs

Alexandre GODZINSKI (SEVS)

Élise MARTIN (SEVS)

Mathilde NIAY* (SEVS)

Quentin NICOD* (SEVS)

Caroline PINTON (SEVS)

Bruno QUILLE* (SEVS)

* *En poste au SEVS lors de la rédaction du document.*

Sommaire

I. REVUE DE LITTÉRATURE.....	8
II. PRÉSENTATION DU MODÈLE PROMETHEUS	20
III. BASES DE DONNÉES D'ENQUÊTES MOBILISÉES	25
IV. TRAITEMENT DES BASES	31
V. PSEUDO-APPARIEMENT.....	43
VI. VIEILLISSEMENT ANNUEL.....	56
VII. MISE À JOUR DE LA FISCALITÉ	63
VIII. SYNTHÈSE ET UTILISATION PRATIQUE DU MODÈLE	69
IX. CAHIER DE VARIANTES.....	71
ANNEXES	99

Synthèse

Le modèle Prometheus est l'outil de microsimulation du CGDD de la facture énergétique des ménages. Il est développé, maintenu et exploité en interne au CGDD, dans le bureau de la transition énergétique et solidaire (SEVS/SDEE2). Il permet de chiffrer l'impact sur les ménages (facture énergétique, taux d'effort énergétique, précarité énergétique...) d'évolution des prix de l'énergie, de la fiscalité des énergies ou encore de dispositifs d'accompagnement (chèque énergie). Ces impacts peuvent être quantifiés en moyenne sur l'ensemble des ménages, mais aussi selon diverses dimensions socio-économiques (niveau de vie du ménage, énergie de chauffage, commune d'appartenance, etc.). La troisième version du modèle vient d'être finalisée et intègre notamment l'Enquête mobilité des personnes 2019 du SDES pour mettre à jour la description des dépenses de carburant des ménages. Le précédent et unique document de présentation du modèle datant de 2016, cette troisième version est l'occasion de publier un nouveau document de travail. Son objectif est de détailler avec précision l'ensemble du fonctionnement du modèle et à en donner les principaux résultats canoniques.

Le modèle de microsimulation Prometheus (PROgramme de Microsimulation des Énergies du Transport et de l'Habitat pour Évaluations Sociales) estime la consommation énergétique des ménages résidant en France métropolitaine pour leur logement et pour leurs déplacements en voiture. La microsimulation permet d'obtenir ces estimations à un niveau fin de granularité, permettant des analyses selon diverses dimensions socio-économiques, par exemple pour chaque dixième de niveau de vie, par type de commune (au sens du zonage en aire d'attraction des villes), par énergie principale de chauffage ou selon les équipements en véhicules. Des croisements entre dimensions sont également possibles.

Sur son champ d'étude (facture énergétique complète des ménages et grandeurs associées), le modèle est le plus précis existant en France. Il est mobilisé notamment :

- pour réaliser des estimations descriptives de la facture énergétique des ménages, de leur taux d'effort énergétique, de leur consommation d'énergie et de leurs émissions directes de CO₂¹ ;
- pour estimer chaque année le taux de précarité énergétique (nombre de ménage en situation de précarité énergétique au sens du taux d'effort énergétique)² ;
- pour évaluer l'impact de dispositifs d'aide tels que le chèque énergie ;
- pour simuler l'impact sur la facture des ménages de variations ex post ou ex ante des prix de l'énergie ou de la fiscalité énergétique. Ce fut le cas lors des évaluations du bouclier tarifaire en 2022³.

¹ Dans la *partie III.C* du Rapport sur l'impact environnemental du budget de l'État annexé chaque année au PLF1 (rapport « budget vert de l'État »), la fiscalité énergétique payée par les ménages est détaillée par usage (logement et transport) et entre ses différentes composantes (composante carbone, hors composante carbone, fiscalité sur l'électricité) en utilisant le modèle Prometheus.

² [La précarité énergétique diminue en 2022 grâce aux mesures de soutien aux ménages et aux baisses des consommations](#), Théma Essentiel, mai 2024

³ [L'impact distributif des mesures de soutien aux ménages face à la hausse des prix de l'énergie en 2022](#), Théma Essentiel, septembre 2023.

Ce document de travail porte sur la *version 3* du modèle, où l'Enquête nationale logement 2013 (ENL 2013) de l'Insee est appariée avec l'Enquête mobilité des personnes 2019 (EMP 2019) du SDES (encadré). Dans la première version du modèle développée en 2016, seule l'ENL 2006 était exploitée et le champ du modèle se limitait donc à la facture énergétique pour le logement. La *version 2*, mobilisait l'ENL 2013 et l'Enquête nationale transport et déplacement (ENTD 2008), conduisant donc à une estimation de la facture énergétique complète des ménages. Cette *version 2* n'a pas fait l'objet d'un document de travail. Cette troisième version substitue l'EMP 2019 à l'ENTD 2008, et l'ensemble du modèle est désormais codé avec le langage de programmation R, le langage SAS, utilisé pour les *versions 1 et 2* ayant été écarté.

Ce document de travail poursuit deux objectifs. D'une part, il développe les choix méthodologiques sur lesquelles le modèle s'appuie et vise à servir de référence à laquelle renvoyer pour l'ensemble des travaux conduits désormais avec le modèle. D'autre part, il présente un cahier de variantes, permettant d'illustrer les résultats canoniques du modèle pour différents chocs sur les prix de l'énergie, sur la fiscalité, et sur les mesures d'accompagnement. Ces résultats sont présentés pour l'année 2019, dernière année disponible au cours de laquelle l'énergie des ménages n'a pas fait l'objet d'évolution notable.

Le cahier de variantes présente notamment l'impact, sur la facture énergétique des ménages et le taux d'effort énergétique, de hausses des prix de l'énergie (carburants, gaz, électricité), de hausses de la fiscalité énergétique (composante carbone, droit d'accise sur l'électricité, TVA sur l'énergie) et des versements de chèques énergie (chèque énergie classique et augmenté d'un chèque additionnel). Les résultats mettent notamment en évidence le caractère régressif du renchérissement des énergies : si la facture en euros augmente davantage pour les ménages aisés que pour les ménages modestes, du fait des volumes consommés, cette hausse pèse davantage sur le revenu des ménages modestes ; cela vaut notamment dans le cas d'une hausse des carburants. Le versement d'aides ciblées telles que le chèque énergie contribue à alléger significativement la facture énergétique des ménages les plus modestes.

Introduction

La transition énergétique représente un changement structurel de l'économie, qui implique notamment une évolution du mix énergétique et du prix des énergies. Cette évolution affecte la facture énergétique des agents économiques, notamment celle des ménages. Or, la situation de ceux-ci au regard de leur facture énergétique est très hétérogène et dépend de nombreux déterminants. La connaissance fine des niveaux de facture énergétique et de leurs évolutions lors de la mise en place de différentes politiques de transition énergétique est donc nécessaire pour concevoir des politiques publiques pertinentes énergétiquement et acceptables socialement. Le modèle de microsimulation du Commissariat général au développement durable (CGDD) intitulé « Prometheus » pour « PROgramme de Microsimulation des Énergies du Transport et de l'Habitat pour ÉvalUation Sociales », développé depuis 2016, a pour objectif de contribuer à cette connaissance. Il estime finement les niveaux de facture énergétique payés au niveau du ménage individuel, c'est-à-dire en tenant compte de l'hétérogénéité de ces derniers, comme les différences de revenus, de zone d'habitation, d'énergie principale de chauffage, de taille, etc. Il permet également d'analyser l'effet, sur la facture énergétique des ménages, de variations de prix hors taxe et de fiscalité des énergies, ou encore l'effet de mesures de compensation, comme le chèque énergie. Par rapport à la version initiale de 2016 du modèle, (*Thao Khamsing et al.* (2016)), les principales nouveautés de cette nouvelle version sont la prise en compte de la facture énergétique du transport, en plus de la facture énergétique du logement, l'augmentation du nombre de mesures simulables, et le passage du code du modèle dans le langage R.

Ce document présente la construction du modèle, son fonctionnement ainsi qu'un échantillon des résultats qu'il permet d'obtenir.

Prometheus (*PROgramme de Microsimulation des Énergies du Transport et de l'Habitat pour ÉvalUation Sociales*) est un modèle de microsimulation des dépenses de consommations d'énergies résidentielles et de carburants des ménages qui s'inscrit dans la tradition des modèles de microsimulation statiques portant sur les ménages. La littérature distingue usuellement deux grandes familles de modèles de microsimulation : les modèles statiques où les changements de comportements individuels ne sont pas modélisés, et les modèles dynamiques qui les modélisent. Les modèles de microsimulation peuvent également être distingués selon qu'ils sont généralistes ou spécialisés dans une thématique donnée, ainsi, la microsimulation énergétique peut être réalisée par des modèles spécialisés sur l'énergie, tel que Prometheus, ou par des modèles généralistes dont on utilise les aspects énergétiques (*chapitre I*).

Dans Prometheus, les prix et les taxes qui s'appliquent aux consommations énergétiques sont intégrés au modèle, permettant de simuler la facture liée à l'énergie consommée directement par le ménage suivant ses caractéristiques sociodémographiques, avec une hétérogénéité verticale (dixième de niveau de vie, catégorie socio-professionnelle) et/ou horizontale (composition du ménage, zone d'habitation...). Le modèle permet aussi d'analyser les résultats selon les caractéristiques du logement des ménages (énergie de chauffage, type de logement, surface...) et selon son usage des voitures (nombre de kilomètres parcourus, nombre de véhicules possédés, type de carburant...). Ces différentes dimensions peuvent être combinées afin de produire une analyse à un niveau microéconomique (*chapitre II*).

S'agissant des sources de données le constituant, Prometheus s'appuie sur deux enquêtes de la statistique publique qui fournissent un échantillon de ménages représentatifs : l'édition

2013 de « l'Enquête nationale logement » de l'Insee (ENL 2013), pour obtenir, au niveau de chaque ménage de l'échantillon, les consommations énergétiques de son logement ; « l'Enquête mobilité des personnes » du SDES⁴ (EMP 2019), pour obtenir les consommations en carburant. Ces enquêtes sont en effet représentatives de l'ensemble des ménages vivant en logement ordinaire en France métropolitaine (*chapitre III*). Les données de ces deux enquêtes sont redressées pour prendre notamment en compte les données manquantes, puis sont traitées pour obtenir les variables d'intérêt du modèle (factures déclarées, consommations énergétiques pour les ménages en chauffage collectif, kilométrages parcourus, etc.). Du fait de sa plus grande ancienneté, l'ENL 2013 fait l'objet de redressements plus importants que l'EMP 2019 (*chapitre IV*). Prometheus a l'originalité et la spécificité de lier ces deux enquêtes afin de permettre l'analyse conjointe des factures d'énergie pour un même ménage, tant sur le logement que sur le transport. Un pseudo-appariement a rendu possible ce rapprochement de deux enquêtes différentes (*chapitre V*).

La base de données des ménages pseudo-appariée comprenant l'ensemble des ménages et des variables conservées est actualisée chaque année à l'aide de données agrégées afin de refléter l'évolution observée des comportements des ménages relativement aux énergies (caractéristiques des logements et des véhicules, volumes de consommations énergétiques, etc.). C'est ce qu'on appelle le « vieillissement » de la base de données (*chapitre VI*). Le modèle permet en outre de représenter au mieux les réformes les plus récentes de la législation fiscale (taxes) et les dernières évolutions des prix et des tarifs de l'énergie, avec une prise en compte possible de la réaction des ménages face à ces évolutions (élasticités-prix) – (*chapitre VII*). Le vieillissement des données répond ainsi à la difficulté de disposer de données individuelles à la fois récentes et détaillées, en l'absence actuellement de telles bases de données d'enquêtes statistiques ou administratives.

Le modèle s'inscrit dans la lignée des modèles de microsimulation, notamment dans celle du modèle Ines de l'Insee et de la Drees. L'organisation des données utilisées (données d'enquêtes initiales, et données externes actualisées chaque année dites *d'inputs*) et des programmes (sous R principalement) est relativement similaire (*chapitre VIII*). Prometheus permet ainsi d'estimer et d'analyser les factures d'énergie annuelles en temps réel autant que possible (*nowcasting*), d'affiner les résultats disponibles en croisant de nombreuses dimensions d'analyse (avec les variables socio-économiques), mais aussi de quantifier les variations des factures ou des taux d'effort (en part du revenu) à la suite d'un choc externe (par exemple une hausse de prix, de fiscalité, etc.). Il permet enfin de prendre en compte de l'effet de dispositifs de compensations (comme le chèque énergie), de réaliser des estimations de coûts budgétaires, etc. (*chapitre IX*). Actuellement basé sur deux enquêtes de la statistique publique, le modèle évolue au rythme des différentes éditions de ces enquêtes. En outre, à long terme, l'émergence et la mise à disposition de nouvelles données et de bases administratives poseront également de nouveaux défis.

⁴ Service des données et études statistiques, du ministère de la Transition écologique, au sein du CGDD.

I. REVUE DE LITTÉRATURE

A. Objectifs et typologie des modèles de microsimulation

L'idée de la microsimulation remonte à *Orcutt* (1957), qui notait que les modèles économiques usuels estiment seulement les agrégats et n'arrivent pas à prédire les distributions des individus, ménages, entreprises selon une ou plusieurs dimensions. La microsimulation permet ainsi d'aller au-delà de cette limite et d'entrer dans le détail d'un résultat macroéconomique. Elle permet en particulier, au niveau des agents individuels d'interpréter les variations de revenus comme celles de bien-être, suite à une variation de prix ou de taxes (*Bourguignon et Spadaro* (2006)).

Une définition simple est ainsi donnée par *Klevmarken* (2022) : « la microsimulation est une technique qui utilise la capacité des ordinateurs modernes pour faire agir et interagir les micro-unités de telle manière qu'il soit possible de les agréger au niveau d'intérêt. Un modèle de microsimulation peut être considéré comme un ensemble de règles qui fonctionnent sur un échantillon de micro-unités telles que les individus, les ménages et les entreprises. Chaque micro-unité est définie et caractérisée par un ensemble de propriétés (variables) et au fur et à mesure que le modèle est simulé, ces propriétés sont mises à jour pour chaque micro-unité ».

La microsimulation permet ainsi d'appréhender plusieurs problématiques de politiques publiques (*Blanchet et al.* (2016)). D'une part, c'est un outil de quantification *ex ante* comme *ex post* de l'impact distributif d'une ou plusieurs mesures de politiques publiques, en permettant une évaluation allant au-delà d'une réponse binaire (du type « oui, la mesure a un effet significatif ») ou moyenne (en regardant les distributions) ou pour calculer comment des indices d'inégalité (Gini par exemple) sont modifiés ; dans le but de mieux la calibrer, par exemple de prévoir d'éventuelles compensations des agents les plus vulnérables ou des transferts monétaires entre agents, d'anticiper son coût budgétaire, ou pour évaluer qui a été le plus ciblé⁵. D'autre part, la microsimulation permet de réaliser une estimation de données en temps réel ou avancée (*nowcasting*), ou même des prévisions (*forecasting*) avec les limites inhérentes à la modélisation. Elle permet surtout de modéliser des scénarios en modifiant des données en entrée ou des relations comptables au sein du modèle et aller plus finement dans l'analyse que les approches par des cas-types ou des résultats agrégés qui raisonnent en moyenne (*Legendre* (2004)), lesquels peuvent néanmoins contenir des biais d'agrégation d'après *Rausch et Schwarz* (2016), qui montrent que l'évaluation de l'incidence d'une taxe carbone est biaisée si l'on ne tient pas compte de l'hétérogénéité des ménages. Deux grandes catégories de modèles de microsimulation peuvent être distingués⁶, les modèles « statiques » (aussi appelés « comptables » tel *Prometheus*) et les modèles « dynamiques ».

La première catégorie, constituée des modèles statiques, permet, à partir de données individuelles (issues d'un échantillon représentatif) auxquelles le modèle applique d'autres données comme paramètres, d'obtenir un état du monde différent de l'état initial. Le cadre le plus simple est l'application d'une évolution de paramètres aux données individuelles, comme une évolution du système socio-fiscal, à date inchangée. À partir de là, il est possible

⁵ Comme le rappellent *Bourguignon et Landais* (2022), ne pas utiliser des réponses comportementales (i.e. les élasticités) ne fausse pas l'analyse en matière d'impacts redistributifs. En revanche, il est préférable de les utiliser concernant les estimations de coût budgétaire.

⁶ Pour plus de détail sur cette classification voir *Dekkers* (2015), *Blanchet et al.* (2015) et *Blanchet* (2020).

d'aller un peu plus loin et de modéliser des évolutions temporelles simples. Il s'agit alors d'appliquer un système de prix, de fiscalité ou de revenu en année N+1 à des individus en année N, pour en déduire les grandeurs d'intérêt en année N+1, en supposant que les individus conservent en année N+1 leurs autres caractéristiques de l'année N. Éventuellement, un modèle statique peut intégrer un changement de comportement (en implémentant des élasticités-prix par exemple) dans le système d'hypothèses permettant d'aboutir au nouvel état du monde.

Par ailleurs, il est possible, en statique, de faire « vieillir » la base initiale sur des dimensions spécifiques (pour les concepts présentés ici voir *Jingjing et al. (2014)* et *Dekkers (2015)*). Il existe alors trois méthodes :

- mise à jour par groupe. Les données sont mises à niveau (*uprating*), par exemple en leur appliquant des évolutions externes ou en les indexant ;
- mise à jour en fonction des caractéristiques individuelles. Dans certains cas, cela peut se faire via l'actualisation (*updating*) des données de bases ou de la législation ;
- repondération : le vieillissement se fait grâce à la repondération des individus. Cela permet d'aligner les données microéconomiques sur une information externe (*reweighting*, voir *Immervoll et al. (2005)*), i.e. de faire de la « pseudo-dynamisation » (*static-ageing*). Dans ce cas, le modèle permet de prendre en compte les changements plutôt que de les simuler (*O'Donoghue et Loughrey (2014)*). En revanche, à terme, certains individus ayant un poids faible dans la base initiale, risquent d'avoir un poids total trop élevé au regard du nombre d'observations, ce qui pourrait distordre le poids d'autres caractéristiques. Dans ce cas, un modèle dynamique serait plus adapté.

La seconde catégorie, constituée des modèles dynamiques, modélise les comportements individuels au sein de la population initiale. Chaque individu peut avoir sa propre dynamique, qui dépend des changements de l'état du monde. Ce second type de modèle fait intervenir des transitions aléatoires entre états et des probabilités associées (sans intervention sur les poids), lesquelles peuvent être plus ou moins contrôlées à l'aide de calages sur des données externes. En outre, il existe deux sous-catégories de modèles dynamiques : ceux « de période » qui projettent l'ensemble d'une population au cours du temps et ceux « de cohorte » qui projettent une génération sur son cycle de vie. Pour cette seconde grande catégorie de modèles, une approche moins fine sur les individus, dite « cellulaire » ou « matricielle », avec une granularité intermédiaire entre échelle micro et macroscopique peut parfois être plus indiquée (voir *Blanchet et al. (2015)* pour une discussion).

En théorie, il est possible de rendre équivalentes les sorties et propriétés d'un modèle statique à celles d'un modèle dynamique, via l'application dans le modèle statique d'un bon système de poids et de repondération successive. Cela peut être le cas s'il n'y a pas besoin de simulation à trop long terme, et que la structure de la population représentée ne se déforme pas trop fortement en réponse, notamment à des chocs exogènes. Par ailleurs, il faut également que les caractéristiques que l'on veut faire évoluer soient présentes chez un nombre minimal d'individus dans la base initiale. *Dekkers (2015)* va jusqu'à réfuter la distinction entre modèles statiques et dynamiques, en ramenant le choix entre chaque catégorie à des considérations purement techniques (voir aussi *O'Donoghue (2001)* et *De Blander et al. (2013)*)⁷.

Plusieurs modèles de microsimulation existent en France, laquelle a vu l'émergence progressive, depuis le milieu des années 1960, puis la consolidation, depuis les années 2000,

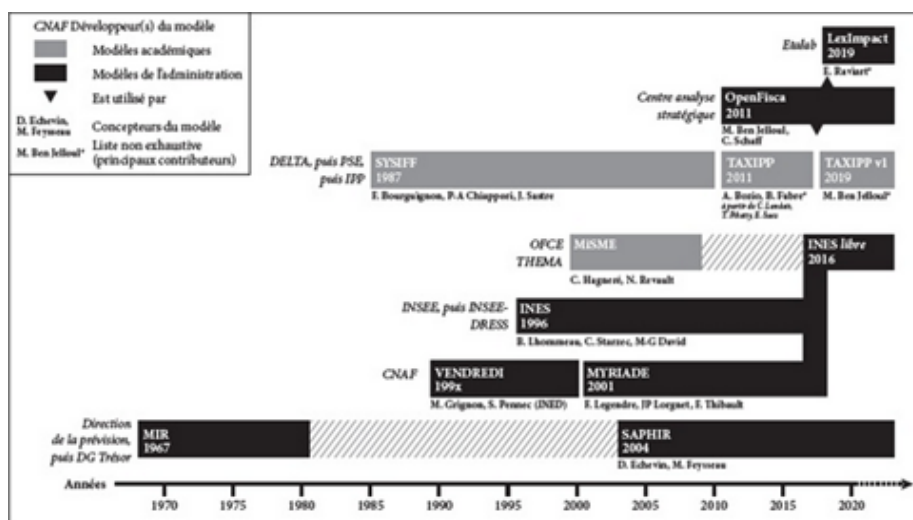
⁷ Il est aussi possible d'étudier des intervalles de confiances des résultats de la microsimulation (voir *McClelland et al. (2020)* pour une courte revue de papiers et un exemple), mais c'est assez peu fait dans la littérature et les sorties des modèles.

de modèles dédiés à l'analyse des politiques sociales et fiscales (*Legendre (2019)*). Deux modèles de microsimulation sont habituellement cités pour illustrer en France la modélisation statique et dynamique :

- le modèle statique Ines (initialement pour *Insee études sociales*, puis pour *Ines-Drees*⁸, ces deux institutions développant conjointement le modèle) co-développé par l'Insee, la Drees et la Cnav. Ce modèle est en particulier utilisé pour établir le bilan annuel des effets redistributifs des mesures socio-fiscales ou encore produire l'estimateur avancé du taux de pauvreté⁹ en France ;
- le modèle dynamique Destinie (*démographique économique et social de trajectoires individuelles simulés*) développé par l'Insee. Il permet de simuler l'évolution des individus au cours de leur cycle de vie, en modélisant notamment leur passage à la retraite, prend également en compte des changements démographiques, et se place ainsi sur le long terme pour évaluer la soutenabilité du régime de retraites¹⁰.

Concernant la France, la *figure 1*, extraite de *Bessis et Cotton (2021)* permet de synthétiser les principaux modèles statiques existants¹¹. Pour les modèles étrangers, voir par exemple *Bourguignon et Landais (2022a)*.

Figure 1 : historique des principaux modèles de microsimulation statiques en France



Source : Bessis et Cotton (2021)

Le modèle Prometheus, qui s'intéresse à la facture énergétique des ménages, appartient aussi à cette catégorie. Il en est à sa troisième version (*encadré 1*).

Au-delà des modèles de microsimulation socio-fiscaux classiques, comme noté par *Blanchet (2020)*, plusieurs déclinaisons peuvent être considérées : des extensions vers un super-modèle englobant plusieurs dimensions au risque d'être parfois moins précis, sauf à développer des sous-modules qui constituent autant de surcouches du modèle initial ; ou au

⁸ Direction de la recherche, de l'évaluation, des études et des statistiques.

⁹ Voir : www.insee.fr/fr/information/2021951 ou ici : <https://Drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/le-modele-de-microsimulation-ines>

¹⁰ Voir la méthodologie et des publications ici : www.insee.fr/fr/information/3606338#titre-bloc-7 ; pour une revue plus large des différents modèles sur les retraites, voir *Cheloudko et Martin (2020)*.

¹¹ Ce papier permet également de faire l'historique des avancées de la connaissance sur la microsimulation et les volontés de monopoles ou de partenariats entre administrations pour chaque modèle. Pour des recensions plus complètes, bien qu'un peu anciennes, voir *Blanchet et al. (2015)* sur les modèles statiques et dynamiques en France, *Li et O'Donoghue (2013)* sur les modèles dynamiques au niveau mondial et *O'Donoghue et Loughrey (2014)* sur les modèles statiques en particulier de *nowcasting*.

contraire des déclinaisons vers des modèles portant sur une problématique donnée, avec des « thèmes envisageables à la commande, par exemple la simulation de l'exposition à la précarité énergétique, déjà traitée au CGDD dans un cadre statique, celui du modèle Prometheus ».

Le modèle Prometheus répond en effet au besoin identifié par *Legendre (2019)* pour la microsimulation d'étudier l'impact des politiques de transition écologique. Il cite Prometheus comme réponse à ce besoin, car il permet « d'identifier les conséquences anti-redistributives des politiques de taxation indirecte ». Plus largement, pour calibrer les mesures d'une transition juste et éclairer le débat en la matière, les modèles thématiques et précis comme Prometheus (voir par exemple pour une évaluation du « bouclier tarifaire » par Prometheus, *Nicod et al. (2023)*) semblent plus appropriés que les modèles de microsimulation socio-fiscaux classiques. Ainsi, la plupart des modèles de microsimulation socio-fiscaux simulent des prélèvements directs et des dispositifs de redistributions aux ménages. Peu de modèles en revanche portent sur la taxation de la consommation et les prélèvements indirects. Ces derniers, comme Prometheus, ont plusieurs spécificités (*Capéau et al. (2014)*) : les taxes portent sur des consommations, donc à partir de la consommation des individus, on peut calculer le niveau de taxation, à l'inverse, lorsqu'il s'agit de compiler des prélèvements payés ou des transferts reçus, les individus doivent être identifiés via des critères d'éligibilité plus difficilement identifiables. Par ailleurs, les enquêtes et les données administratives sur les quantités consommées étant plus rares et moins fréquemment collectées que celles fiscales, les données sous-jacentes à ces modèles sont plus souvent mises à jour au niveau des groupes (pour coller à des données macroéconomiques) qu'au niveau individuel.

Il est à noter que les modèles de microsimulation (et c'est le cas de Prometheus) peuvent utiliser parfois des élasticités-prix. Sans que cela ne permette de parler de dynamique, ces élasticités modélisent néanmoins une adaptation de court terme des agents. En particulier, des élasticités-prix à la consommation permettent de modéliser comment les ménages adaptent à court terme leur consommation d'un bien, en réponse à une hausse du prix de ce même bien.

Enfin, une autre approche plus complexe est celle qui consiste à coupler les effets de bouclage macroéconomique et l'analyse microéconomique issue d'un modèle de microsimulation. Par exemple concernant l'empreinte environnementale engendrée par la consommation de biens, ce type d'approche permet de prendre en compte la déformation de la structure de l'économie sur les résultats (*Boccanfuso et al. (2011)* et *Yusuf et Resosudarmo (2015)*). Au travers d'une approche comptable, toute une littérature se base sur l'analyse input/output à la Leontief (*O'Donoghue (1997)*, *Gay et Propos (1993)* et *Casler et Rafiqui (1993)*). Une autre démarche possible est de lier modèle de microsimulation et modèle macroéconomique (par exemple un « modèle d'équilibre général calculable » (ECG) ; *Bourguignon et al. (2008)* pour une présentation et des applications). Comme le rappelle *Klevmarken (2022)*, ce « vœu pieux » était déjà présent chez Orcutt qui, à la première conférence internationale de microsimulation en 1977, avait exposé un modèle de microsimulation intégrant un modèle macroéconomique auxiliaire. Ce type de modèle « micro-macro » est plus ambitieux et encore peu répandu, mais il permet de pallier les insuffisances de chaque outil, à la fois du côté microéconomique en permettant des effets de bouclage, et du côté macroéconomique en allant dans l'hétérogénéité des agents économiques. Pour ce faire, plusieurs approches sont possibles : sans rétroaction, c'est-à-dire de l'échelle macroéconomique vers l'échelle microéconomique (avec réponse ou non de court terme des individus) ; ou au contraire de l'échelle microéconomique vers l'échelle macroéconomique, mais cette fois-ci avec rétroaction à l'aide d'une approche itérative convergeant vers un état stable. Il est aussi envisageable d'avoir un modèle qui intègre à la fois les deux approches ce qui nécessite néanmoins des bases de données cohérentes et des capacités de calcul importantes. Concernant cette typologie et des exemples de modèles voir *Cockburn et al. (2014)*. En

particulier, pour modéliser l'impact d'une taxe environnementale sur la structure de l'économie comme sur les ménages pris avec leur hétérogénéité, il est en théorie plus approprié de passer par une analyse avec un modèle ECG couplé avec un modèle micro ou qui présente de l'hétérogénéité des ménages.

Encadré 1 : les documents de travail des modèles de microsimulation

La précédente version de document de travail du modèle Prometheus date de 2016 (Thao Khamsing et al. (2016)). Le modèle, qui en était à sa première version, s'intéressait alors uniquement à la composante logement de la facture énergétique, grâce à l'utilisation de l'ENL 2006.

Le présent document de travail a ainsi pour but de présenter de manière détaillée la construction et le fonctionnement du modèle Prometheus version 3, en particulier concernant les aspects de calage et de vieillissement, qui sont souvent peu détaillés dans les documentations méthodologiques des modèles de microsimulation, comme le relève la note du Conseil d'analyse économique (Bourguignon et Landais (2022b)).

Ce document vise aussi à présenter un cahier de variantes de manière similaire à ce qui est souvent fait dans les documents de présentations des modèles macroéconomiques appliqués, et plus rarement fait dans les documents présentant les modèles de microsimulation (pour quelques exemples des cahiers de variantes d'un modèle de microsimulation, voir, pour le modèle Ines, Fontaine et Sicsic (2016) et Cornuet et Sicsic (2020)).

B. La microsimulation énergétique

Dans ce chapitre, nous faisons une recension des modèles de microsimulation portant en partie ou entièrement sur le sujet des dépenses énergétiques des ménages. Plus que sur les conclusions elles-mêmes de ces modèles, nous nous focalisons sur l'orientation, la méthode et l'originalité de l'approche de chacun d'entre eux.

Les politiques environnementales, qui constituent la thématique d'évaluation de Prometheus, ont des effets très hétérogènes sur les ménages et constituent donc un sujet d'étude particulièrement intéressant pour les analyses distributives. En particulier, la microsimulation permet d'étudier de nombreuses thématiques environnementales spécifiques (analyse spatiale, agriculture, ressources, transport et congestion, coût d'abattement, etc.¹²). Serret et Johnstone (2006) font un panorama de différentes études utilisant la microsimulation et portant sur ces thématiques.

Ainsi, dans une des premières études sur le sujet mobilisant un modèle de microsimulation comprenant un modèle de demande pour évaluer l'impact sur les prix, y compris des biens non énergétiques, suite à l'introduction d'une taxe carbone¹³, Symons et al. (1994) montrent que relativement aux revenus des ménages, les taxes environnementales touchent davantage ceux situés en bas de la distribution, c'est-à-dire les plus modestes. Des transferts peuvent alors être introduits pour réduire ces effets régressifs qui s'expliquent par le fait que les taxes

¹² Pour une recension, voir O'Donoghue (2014), chapitre 15 « *Environmental model* ».

¹³ Voir O'Donoghue (2014) chapitre 15 également, pour une petite revue de littérature sur la modélisation de l'impact distributif des taxes environnementales entre 1997 à 2012.

environnementales s'appliquent à des quantités de biens qui ont un impact négatif sur l'environnement, et non à des valeurs nominales (contrairement à la TVA) qui concernent des biens de consommation souvent essentiels comme le carburant et les énergies de chauffage.

L'évaluation des politiques environnementales via la microsimulation permet également d'étudier l'hétérogénéité de la quantité de gaz à effet de serre émise par les ménages. Ainsi, *Symons et al.* (1994), modélisent également des réactions comportementales des ménages pour quantifier l'effet d'une taxe carbone sur les émissions de CO₂.

Comme évoqué dans le chapitre précédent, l'introduction de mécanismes comportementaux dans les modèles devient essentielle dès lors qu'il s'agit d'étudier l'impact, sur la consommation ou l'environnement, d'une politique publique donnée. Par exemple, concernant l'effet d'une taxe carbone, *Symons et al.* (1994) modélisent explicitement les effets comportementaux de la mesure, et donc ses effets sur les émissions de CO₂, tandis que *Casler et Rafiqi* (1993) adoptent une approche input/output pour en quantifier l'effet purement distributif et monétaire. Dans le cas des modèles de microsimulation intégrant des effets comportementaux, leur modélisation peut se faire soit avec un modèle structurel de demande (par exemple avec une fonction d'utilité), ou plus simplement avec des élasticités-prix.

L'utilisation de modèles de microsimulation spécialisés sur l'énergie

Un des premiers modèles de microsimulation portant en partie sur l'énergie est celui exposé dans *Ruiz et Trannoy* (2008). Cette approche montre et estime le caractère régressif de la taxation indirecte des biens, à l'aide d'un modèle de microsimulation appelé « indirect taxes simulation » (ITS¹⁴). Il permet en effet de modéliser la hausse des prix TTC induite par la hausse d'une taxation des biens, via un taux de répercussion, c'est-à-dire l'effet d'une taxe sur le prix TTC, (pour 71 biens), d'évaluer ainsi les nouvelles dépenses de consommation des ménages en tenant compte de leurs changements de comportement (modélisés par huit élasticités-prix non différenciées entre les ménages des produits estimées dans cette même approche) et d'en déduire les recettes fiscales associées (avec extrapolation au niveau macroéconomique des résultats microéconomiques). Pour estimer les élasticités, le modèle met en place la méthode de « prix personnalisé » (pour éviter des biais d'estimation et garantir la variabilité des prix), en utilisant l'enquête budget de familles 2001 (BDF 2001) et la méthode *QUAIDS*¹⁵. Bien que non limité à l'énergie, ce modèle est une première tentative de modélisation en propre des dépenses énergétiques de logement et de transport.

Également à l'aide d'un modèle de microsimulation, *Berry* (2017) étudie l'impact distributif d'une taxe carbone et son effet sur la précarité énergétique (*encadré 2* pour plus de détail sur la notion dont *Prometheus* permet la mesure). Le modèle permet également de mettre en évidence quelles compensations introduire pour corriger les effets régressifs de la taxe. Ce modèle, statique, utilise l'enquête de performance de l'habitat, équipements, besoins et usages de l'énergie (Phébus 2013), produite par le SDES et qui porte sur l'année 2012, et son échantillon de 5 405 ménages, contenant des informations sur les logements et les consommations énergétiques logement et transport (mais en excluant les 20 % de ménages en logement collectif). La méthode utilisée modélise le système de taxes énergétiques de 2012, ainsi qu'un scénario contrefactuel avec introduction d'une taxe carbone, mais estime aussi des élasticités-prix des consommations d'énergie à partir de BDF 2016, différenciées par type d'énergie et par dixième de revenu des ménages.

¹⁴ Développé sous SAS et Excel, mais sans plus de détail et sans qu'il n'ait donné lieu à beaucoup d'études par la suite.

¹⁵ *Quadratic Almost Ideal Demand System*, voir *Banks, Blundell et Lewbel* (1997).

Encadré 2 : la précarité énergétique

La notion de précarité énergétique¹ a été définie par la loi du 12 juillet 2010 dite « loi Grenelle II », qui a suivi le Grenelle de l'environnement. Les ménages en précarité énergétique y sont définis comme les ménages qui « éprouvent dans leur logement des difficultés particulières à disposer de la fourniture d'énergie nécessaire à la satisfaction de leurs besoins élémentaires en raison de l'inadaptation de leurs ressources ou de leurs conditions d'habitat ».

La notion de précarité énergétique peut être mesurée de différentes manières. L'enquête nationale logement de l'Insee, contient une question sur le « froid ressenti » et offre donc une mesure subjective de la précarité énergétique. Celle-ci peut également être mesurée objectivement : c'est généralement le cas en définissant un seuil de taux d'effort (ratio entre les dépenses énergétiques liées au logement et le revenu) en dessous duquel les ménages sont considérés comme précaire. Devalière (2009) étudie la précarité énergétique à l'aide d'un tel indicateur, à partir de l'ENL 2006, en fixant ce seuil à 10 % du revenu disponible par unité de consommation (niveau de vie). Devalière et al. (2011) mettent en regard cet indicateur et celui du froid ressenti. Le seuil de 10 % avait été retenu au regard des premiers travaux menés en Angleterre en 1988 à partir de l'enquête Family Expenditure Survey²). Depuis 2018, dans le cadre d'un partenariat avec l'ONPE (Observatoire national de la précarité énergétique), le CGDD produit une estimation annuelle de ce même indicateur à partir du modèle Prometheus³, en retenant le seuil de 8 % pour les ménages des trois premiers dixièmes (voir Devalière et al. (2018) et la dernière publication en date Martin (2024)). Notons que ces mesures objectives, réalisées à partir des consommations observées ne capture pas l'effet d'une auto-restriction des dépenses énergétiques des ménages pour raisons budgétaires ou autres, comme un effort de sobriété.

Cette notion de précarité est proche du concept de vulnérabilité énergétique étudié dans Cochez et al. (2015), lequel tient compte en plus des transports et, en particulier, des déplacements contraints⁴, avec une définition un peu différente (les consommations considérées peuvent être celles observées comme dans la définition de la précarité énergétique, ou théoriques comme pour la vulnérabilité telle que mesurée par l'Insee en 2015). Dans Cochez et al. (2015), un ménage est dit dans une situation de vulnérabilité énergétique si son taux d'effort énergétique est supérieur au double de la médiane des taux d'effort observés en France métropolitaine l'année considérée, soit 8 % en 2015 pour les dépenses énergétiques logement et 4,5 % pour les déplacements contraints. Sont exclus par ailleurs les ménages ayant un revenu par unité de consommation supérieur au double du revenu par unité de consommation médian. L'étude de Cochez et al. (2015) mobilise les données du recensement de la population (RP), de l'enquête revenus fiscaux et sociaux (ERFS), de l'enquête nationale transport et déplacements (ENTD) 2008, les revenus disponibles localisés (RDL) de 2008, et utilise un modèle de l'Agence nationale de l'habitat (Marchal, 2008) et le modèle Copert de l'Agence européenne pour l'environnement pour modéliser respectivement les dépenses énergétiques logement et transport. En 2023, des travaux sur la vulnérabilité énergétique des ménages ont été lancés par l'Insee pour la réalisation de nouvelles études dont les résultats ont vocation à pouvoir couvrir une échelle spatiale fine⁵. Ils visent à apparier des données administratives, soient celles du RP, du Fichiers démographiques sur les logements et les individus (Fidéli) et du Répertoire statistique des véhicules routiers (RSVERO)⁶.

Encadré 2 : la précarité énergétique (suite et fin)

Au niveau national et dans un cadre européen, une autre grande enquête permet d'appréhender les notions de précarité et de vulnérabilité énergétique. Depuis 2020, l'enquête statistiques sur les ressources et les conditions de vie (SRCV⁷) de l'Insee a en effet été modifiée pour permettre le calcul de l'indice européen de privation matérielle et sociale. Cet indicateur interroge le fait de « ne pas pouvoir maintenir le logement à bonne température pour des raisons financières ». SRCV est une enquête annuelle effectuée en début d'année civile et qui porte sur l'année précédente, sur près de 23 000 ménages représentatifs et dont les résultats, notamment en matière de privation matérielle et sociale, ont vocation à pouvoir se décliner selon les cinquièmes de niveau de vie⁸.

¹ Pour une discussion en particulier sur ce seuil de 10 % et sa mesure au Royaume-Uni, voir Hills (2011 et 2012). Pour une discussion de la notion de précarité énergétique, à la fois sur la sémantique, sa catégorisation pour l'action publique et sa place dans le champ institutionnel, se référer à la thèse d'Adèle Sébert, « *Qualification et prises en charge de la précarité énergétique : une analyse économique institutionnaliste* (2022).

² Ces travaux ont été conduits à la demande du gouvernement suite au Warm Homes and Energy Conservation Act 2000 et la publication en 2001 de la Fuel Poverty Strategy for England.

³ On parle « d'estimation », car seule l'ENL permet de mesurer finement, en approchant via l'échantillonnage, la réalité de la précarité énergétique à date donnée de l'enquête.

⁴ Pour se rendre sur son lieu de travail et/ou d'étude, pour des achats, la santé ou raison administrative.

⁵ Plus d'information : www.insee.fr/fr/information/6965693.

⁶ Nous mentionnons également une thèse en cours qui a un but similaire et intitulé *Géographie de la vulnérabilité énergétique : une approche par les trajectoires résidentielles et mobilitaires des ménages. L'exemple de la Nouvelle-Aquitaine* (Julien Haine).

⁷ Voir méthodologie : www.insee.fr/fr/metadonnees/source/serie/s1220.

⁸ Voir publication 2022 et données : www.insee.fr/fr/statistiques/7651550#documentation.

Un autre travail s'inscrivant dans la lignée du précédent est celui de Douenne (2020). Il développe un modèle de microsimulation dans le but de quantifier l'impact distributif d'une taxe carbone en France. Il montre que la taxe est régressive, mais que cet effet peut être contrebalancé par des transferts. Il montre aussi que l'hétérogénéité horizontale est plus difficilement appréhendable que la verticale, notamment dans une optique de ciblage. Cette approche est focalisée sur l'évaluation de la politique fiscale énergétique entre 2016 et 2018, années de montée en puissance de la composante carbone introduite en 2014 dans la fiscalité énergétique. Pour obtenir ses résultats, Douenne a utilisé le modèle de microsimulation TAXIIP de l'IPP (Institut des politiques publiques) qu'il a complété avec un appariement de l'Enquête nationale des transports et déplacements (ENTD) de 2008 et de BDF 2011. Cette dernière est utilisée en particulier, car le modèle estime une demande des ménages, ce qui peut être fait grâce à la mesure de l'ensemble des dépenses des ménages contenus dans la base. Puis, il réalise un *matching* statistique (méthode *hot-deck* de la « distance des plus proches voisins ») entre BDF et l'ENTD afin de redresser les consommations de carburant dont l'hétérogénéité entre ménage est surestimée dans BDF en raison de la faible période d'enquête (les dépenses sont relevées sur sept jours, puis annualisées). Au contraire de l'enquête BDF, l'ENTD 2008 fournit des distances annuelles. In fine, ces traitements permettent d'obtenir 10 342 observations¹⁶. Enfin, pour simuler l'année de base 2016, ce travail utilise également les comptes nationaux pour redresser les dépenses énergétiques et les revenus. Pour mieux quantifier l'impact de la taxe carbone, il prend ensuite en compte les réponses comportementales des ménages à l'aide des élasticités (de prix non compensés et

¹⁶ Pour le détail du *matching* voir la *Online appendix*.

de revenus¹⁷) estimées à partir de la méthode *QUAIDS* sur BDF 2001/2006/2011. Ces élasticités sont estimées au niveau de l'ensemble des ménages, puis au niveau décile, croisé avec la taille de l'unité urbaine, en distinguant consommation de transport, de logement et de biens non durables.

Les trois précédentes approches n'utilisent toutefois pas « l'enquête nationale logement » (ENL), *Chaton et Gouraud (2020)*, mais élaborent un modèle de microsimulation portant spécifiquement sur l'aspect logement. Ce modèle vise à comprendre, estimer et prédire l'évolution de la précarité énergétique, avec trois modèles linéaires qui permettent d'estimer pour chaque ménage : 1) le revenu disponible 2) la probabilité de faire une rénovation thermique 3) les dépenses énergétiques. Le modèle est constitué de trois blocs correspondants aux trois estimations linéaires et est calibré à l'aide de l'ENL 2006 (qui contient des données sur les ménages, le logement en lui-même, et en lien avec la rénovation). Il peut ainsi estimer la précarité énergétique et étudier les effets de la variation du taux de chômage, des prix de l'énergie sur celle-ci, ainsi qu'estimer le nombre de rénovations thermiques. Le premier modèle, celui qui estime le revenu disponible, tient compte notamment du taux de chômage. Ainsi, faire varier le taux de chômage correspond à changer l'état des ménages du statut de *chômeur* à *non chômeur*, et donc peut conduire à un changement de statut de précarité, lequel dépend du revenu dans le troisième modèle. Ce dernier permet également d'estimer la facture en tenant compte des variations des prix de l'énergie ou de l'effet rebond lié à une rénovation. La modélisation de la rénovation du second modèle est faite via un modèle logit (peu satisfaisant néanmoins d'après l'étude) qui prend également en compte la variation du chômage et des revenus. Les trois simulations faites portent sur 2012-2014 (pour répliquer les dates d'enquête de l'ENL 2013, soit entre juin 2012 et juin 2014), 2017 et 2018.

Un autre angle d'approche de la microsimulation énergétique peut être celui focalisé sur les logements et avec une dimension géographique. *Zhang et al. (2018)* construisent un modèle de consommation d'énergie (électricité et gaz) désagrégé pour les États-Unis, à partir notamment de la *Residential Energy Consumption Survey (RECS)*¹⁸. En utilisant la méthode de *statistical matching* qui prend les variables communes de deux bases les plus corrélées avec les consommations, ils appariant les données de consommations de celle-ci avec celles de la *Public Use Microdata Sample (PUMS)*¹⁹ qui contient des informations sur les caractéristiques des ménages et des logements. Ensuite, un modèle explicatif des consommations est construit afin d'imputer les valeurs manquantes aux ménages qui n'ont pas été appariés. Enfin les données sont appariées avec l'*American Community Survey (ACS)*²⁰ afin d'avoir des informations sur les distributions géographiques et spatiales : sur les ménages selon une échelle de quartiers par exemple. Toutes ces enquêtes portent sur l'année 2009 et le modèle est testé et validé pour la métropole d'Atlanta en confrontant les résultats à d'autres sources de données.

De manière similaire, *Chai et al. (2021)* élaborent un modèle de microsimulation qui estime : 1) la précarité énergétique 2) la capacité des ménages à s'adapter à la hausse des prix. En outre, il permet de réaliser une analyse géographique et de construire un indice de pauvreté énergétique localisé. L'étude se concentre sur la consommation de l'électricité par les ménages, à l'aide d'un échantillon de 1 595 ménages australiens (*ABS Household Energy Consumption Survey 2012*²¹). Ainsi, le modèle utilise l'estimation de la probabilité d'être en

¹⁷ Pour une définition voir l'encadré 2 dans *Calvet et Marical (2011)*.

¹⁸ www.eia.gov/consumption/residential/.

¹⁹ www.census.gov/programs-surveys/acs/microdata.html.

²⁰ www.census.gov/programs-surveys/acs.

²¹ www.abs.gov.au/ausstats/abs@.nsf/Lookup/4670.0main+features12012.

précarité énergétique conditionnellement à des variables observées au sein de l'État du Queensland en Australie, et s'en sert pour estimer à un niveau géographique plus fin cette probabilité (avec les données du recensement 2016). Par ailleurs, le modèle estime des élasticités à l'aide de la modélisation d'une demande d'énergie ménage, y compris par tranche de revenu. Ces élasticités sont injectées dans le modèle de microsimulation pour identifier les ménages susceptibles de basculer en précarité. La microsimulation permet ainsi de modéliser, suite à une hausse de prix, les effets sur les consommations des ménages et donc leurs factures, et *in fine* sur la variation de la précarité énergétique par tranches de revenus, ainsi qu'au niveau régional.

L'utilisation de modèles de microsimulations généralistes

INES

Ines, le modèle de microsimulation phare de l'Insee (développé en partenariat avec la Drees et la Cnaf depuis 1995), est un modèle de microsimulation statique²². Il se base sur l'enquête « revenus fiscaux et sociaux » (ERFS, environ 50 000 ménages) laquelle est disponible pour l'année N en année N+2, et permet avec le vieillissement des données d'estimer la situation socio-fiscale des ménages en temps réel (*nowcasting*) ou de réaliser des simulations avec des variations de fiscalité notamment²³. En traitant ainsi l'ensemble des revenus, cotisations, prestations et impôts directs, il permet de modéliser pour chaque ménage le *revenu disponible*²⁴. Outre l'enquête ERFS, ce modèle utilise également l'enquête BDF, afin d'imputer les données de consommation permettant ainsi de simuler notamment la TVA acquittée et la TICPE (voir André et al. (2016) pour la méthodologie), passant ainsi au concept de *revenu disponible corrigé*. Par ailleurs, il peut aussi modéliser les transferts en nature, dont le chèque énergie, définissant ainsi un concept de *revenu disponible élargi*.

Il existe ainsi une part de modélisation des dépenses énergétiques dans Ines. Cornuet (2022) modélise par exemple les effets des variations des prix de l'énergie sur le revenu des ménages. Cependant, celle-ci nécessite de nombreuses hypothèses et est moins directe que Prometheus. En effet, elle requière, d'une part l'imputation des consommations des ménages de BDF vers ceux de ERFS, et d'autre part la modélisation des prix via l'IPC (indice des prix à la consommation) des postes concernés.

TAXIPP

Le modèle TAXIPP²⁵ construit par l'Institut des politiques publiques (IPP) utilise les données du « fichier démographique sur les logements et les individus » (Fidéli) de l'Insee et du « fichier échantillonné de l'impôt sur le revenu » (Felin, environ 500 000 foyers fiscaux) de la DGFIP, mais aussi les fichiers de « déclarations annuelles de données sociales » (DADS) de l'Insee qui recensent toutes les informations des déclarations sociales des travailleurs salariés pour le recouvrement des cotisations sociales, les fichiers de la « base non-salariés » (BNS) de l'Insee, similaires aux DADS et les données ERFS. De manière similaire à Ines le modèle (également statique) permet de décrire et d'évaluer des modifications du système socio-fiscal. Tout comme Ines, le modèle est doté d'un module complémentaire tenant compte de la TVA, des taxes sur le tabac et l'alcool, ainsi que de la fiscalité indirecte et en particulier de la fiscalité énergétique via la TICPE (voir sur ce dernier aspect Ben Jelloul et Douenne (2016)). Pour cela, l'enquête BDF est utilisée. Par ailleurs, le modèle peut prendre en compte des élasticités, notamment sur la consommation énergétique, provenant de papier de Douenne (2020).

²² La documentation initiale est fournie par David et al. (1999), et plus récemment dans Fredon et Sicsic (2020).

²³ En particulier chaque année dans « France, portrait social », ouvrage de l'INSEE, Ines permet la publication d'un bilan redistributif et budgétaire de l'année précédente.

²⁴ Lequel est proche du concept de revenu disponible brut en comptabilité nationale : www.insee.fr/fr/metadonnees/definition/c1633.

²⁵ www.ipp.eu/methodes/taxipp-outils/.

Le modèle est mobilisé à la fois pour faire de la recherche académique et des évaluations afin d'alimenter le débat public. Par exemple, à l'automne le projet de loi de finances (PLF) en cours d'examen fait l'objet d'un exercice d'évaluation par TAXIPP qui estime l'effet distributif du budget sur les ménages. TAXIPP a la particularité de reposer largement sur des données administratives plutôt que d'enquêtes.

L'utilisation combinée de modèles de microsimulations et macroéconomiques

Pour conclure et comme évoqué précédemment, il est également possible de compléter la microsimulation avec une approche macroéconomique. Ainsi, Vandyck (2013) utilise un modèle d'équilibre général calculable propre à la Belgique, basé sur le modèle GEM-E3 (*General Equilibrium Model for Economy-Energy-Environnement*) développé par la Commission européenne²⁶ et deux modèles de microsimulation. La relation est *top-down*, c'est-à-dire que les résultats du modèle macroéconomique sont intégrés en tant qu'input du premier modèle de microsimulation. Ce dernier est basé sur les données de l'*European Union Statistics on Income and Living Conditions* (EU-SILC²⁷) d'Eurostat sur 14 000 individus belges contenant des informations sur leur niveau d'éducation, la situation sur le marché du travail, l'âge, la région de résidence etc. Il utilise aussi l'enquête « budget des ménages²⁸ » pour construire un indice de prix à la consommation individuel (*IPCI*). La méthodologie consiste à harmoniser les sources macro et microéconomiques, assurer un vieillissement de ces données via les poids des individus (modèle pseudo-dynamique) qui reflète dans la partie micro les réactions de l'emploi suite à des réformes, et intégrer au niveau individuel les variations du niveau de vie induit par le modèle macro. Un autre modèle de microsimulation est utilisé pour obtenir les revenus est Euromod²⁹ et l'impact des variations de prix ou de taxes est également obtenu à partir des *IPCI*. Le papier modélise une hausse de la taxe d'accise sur le pétrole et deux scénarios de redistributions par transferts monétaires (+ 5 %) ou réduction des cotisations employeurs (- 2 points de pourcentage). Les effets macro concernant le PIB, l'emploi, la consommation etc. sont tout d'abord quantifiés dans les deux scénarios. La microsimulation permet ensuite de quantifier l'impact distributif sur les ménages des deux scénarios.

Ravné et al. (2022) combinent également microsimulation et modélisation macroéconomique à trois horizons temporels (2025, 2030 et 2035) en tenant compte : 1) d'élasticités de long terme différenciées en 14 catégories de biens et services et 40 catégories de ménages (par déciles, croisés avec quatre catégories de ménages types) ; 2) du changement technologique induit par l'adoption de véhicules électriques et la mise en place de rénovations thermiques. Le modèle quantifie l'impact distributif de différentes politiques sur les entreprises et les ménages : subventions, transferts, taxe carbone. Pour cette dernière le papier liste trois canaux principaux de transmission de la hausse de taxe : 1) l'impact direct sur le prix de l'énergie et les biens à fort contenu énergétique ; 2) l'impact indirect sur les coûts de production qui entraîne une modification des prix relatifs pour les consommateurs ; 3) la réaction des autres marchés. Au total, la consommation et le revenu des ménages, ainsi que les choix et prix de production des entreprises sont modifiés,

²⁶ <https://web.jrc.ec.europa.eu/policy-model-inventory/explore/models/model-gem-e3/>.

²⁷ En français SRCV : *Statistiques (de l'Union européenne) sur le revenu et les conditions de vie*, voir : <https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions>.

²⁸ <https://statbel.fgov.be/fr/themes/menages/budget-des-menages#documents>.

²⁹ C'est un modèle de microsimulation impôts-prestations pour l'Union européenne qui permet de calculer, de manière comparable, les effets sur les revenus des ménages et les incitations au travail pour la population de chaque pays et pour l'UE en tant que pays (voir *présentation dans O'Donoghue (2014), chapitre 8*). Il est développé par la Commission européenne, Eurostat et de nombreux pays : <https://euromod-web.jrc.ec.europa.eu/overview>.

et ce en modélisant les changements radicaux de long terme tels que l'électrification du parc automobile et la rénovation de logements. Plus précisément :

- sur le plan microéconomique, le modèle de microsimulation utilisé est le modèle Matisse, qui s'appuie sur BDF 2011 et son échantillon de 10 000 ménages pour la France métropolitaine ³⁰ . Le modèle Matisse fonctionne en trois étapes, en effet il : 1) utilise les revenus issus du modèle macroéconomique ; 2) adapte les choix de consommation en fonction des prix macroéconomiques et des revenus des ménages (utilisation des élasticités) ; 3) tient compte de l'électrification des voitures et des rénovations de logement résidentiel ;
- sur le plan macroéconomique (pour revue des différents types de modèles environnementaux, voir *Gantois et al. (2022)*), le modèle considéré est une adaptation de IMACLIM (*Impact Assessment of CLIMate*), modèle du Centre international de recherche sur l'environnement et le développement (CIRED), qui est semblable et compatible avec *ThreeME* ; le modèle s'appelle ainsi IMACLIM-3ME. La calibration est issue du modèle *ThreeME*³¹ (*Multi-sector Macroeconomic Model for the Evaluation of Environmental and Energy Policy*, voir *Callonnec et Cancé (2022)*), qui est un modèle de l'Ademe et de l'OFCE, et correspond à des scénarios de la Stratégie nationale bas carbone (SNBC).

Enfin, il y a un recalage des données microéconomiques pour correspondre à certaines grandeurs macroéconomiques, puis récupération et injection dans le modèle macroéconomique de certaines variables pour sa mise à jour, etc. avec des itérations successives jusqu'à convergence du modèle.

³⁰ Il y a également d'autres enquêtes, mais la liste n'est pas explicite, le papier renvoyant à *De Lauretis (2017)* et *Douenne (2017)* ; ces enquêtes semblent néanmoins être notamment l'ENL, l'ENTD, Phébus, l'enquête emploi du temps (EDT), et une enquête de fournisseur d'électricité.

³¹ www.threeme.org/.

II. PRÉSENTATION DU MODÈLE PROMETHEUS

A. Un modèle de microsimulation statique avec vieillissement

Prometheus est un modèle de microsimulation, sans prise en compte d'effets de bouclage macroéconomique. Par exemple, la modélisation d'une hausse de la fiscalité énergétique n'implique pas un changement de revenu total disponible par un report d'une consommation sur d'autres produits.

Par conséquent, les hausses de prix et taxes ou changement de législation socio-fiscale n'induisent pas de changements de comportement des ménages, par exemple au niveau des rénovations de logement, du nombre de voitures possédées, ou des changements d'équipements domestiques. Certaines distorsions des distributions au sein de la population peuvent être par conséquent mal captées. Néanmoins, il est possible d'utiliser des élasticités-prix pour modéliser une variation des consommations induite par une variation de prix. Le seul aspect dynamique présent est l'évolution de la répartition du parc d'énergie principale de logement et la conversion des voitures d'un type de carburant à l'autre, afin de s'ajuster sur la répartition actuelle du parc de véhicules. Ainsi Prometheus est un modèle statique avec de la pseudo-dynamique³².

Le modèle peut être utilisé de plusieurs manières : tout d'abord, grâce à la mise à niveau du modèle lors du vieillissement des données, il offre la possibilité de faire de l'estimation en temps réel en l'absence d'enquêtes sur l'année en cours. Il permet, d'une part, de simuler des variantes ou contrefactuels en statique pure, par l'application de deux systèmes de prix par exemple. D'autre part, il facilite la projection à court terme avec l'application d'un état de consommation, de prix, de taxes et de revenus prévu, et à moyen terme en tenant compte de la déformation du parc de chauffage des logements à travers la repondération des ménages.

B. L'architecture du modèle

Le modèle Prometheus peut être décomposé en deux grandes parties distinctes : la première est le traitement et le pseudo-appariement des bases de données d'enquête sur lesquelles se fonde le modèle (ENL 2013 et EMP 2019). La seconde est ce qu'on peut appeler la microsimulation en tant que telle, c'est-à-dire le vieillissement des données d'enquête pour les caler sur les grandeurs agrégées les plus récentes, ainsi que l'actualisation des prix de l'énergie et du système fiscal, ce qui permet d'obtenir les factures énergétiques de l'ensemble des ménages de l'échantillon.

³² Un modèle dynamique serait, en revanche, un modèle où les ménages possèdent une utilité individuelle et arbitrent leur changement d'énergie de chauffage en matière de logement, ou encore leur type de carburant et le nombre de voiture possédées en matière de transport, selon des paramètres et des préférences, idéalement avec des fonctions d'offre respectivement du nombre de logements et de véhicules. Cela correspondrait davantage à un ou des modèles de parcs de logements et de véhicules.

Le traitement, puis le pseudo-appariement de l'ENL 2013 et de l'EMP 2019

Au cours de la 1^e phase de construction du modèle, les données d'enquête, l'ENL 2013 et l'EMP 2019, sont traitées, puis pseudo-appariées pour former la base initiale des ménages, laquelle sera vieillie par la suite.

Tout d'abord, dans l'EMP 2019 comme dans l'ENL 2013, de nombreux pré-traitements sur les variables des bases sont effectués : suppression des variables hors sujet pour le modèle, regroupements de modalités de variables, etc. Plus précisément, la construction de la base pseudo-appariée se fait en trois étapes :

1. la 1^e étape consiste à traiter l'ENL 2013 pour obtenir, au niveau de chaque ménage, la consommation d'énergie de logement. Or, seules les factures énergétiques sont renseignées dans l'ENL 2013. Tout d'abord, le type de chauffage est attribué à chaque ménage en logement individuel ou collectif, de même qu'un mode (et une énergie) principal(e) de chauffage. Les factures sont donc retraitées et les valeurs manquantes imputées. Puis, grâce aux prix et tarifs de l'énergie, des consommations d'énergie en volume sont obtenues pour chaque ménage ;
2. la 2^e étape consiste à traiter l'EMP 2019. La variable d'intérêt principale est la consommation de carburant des ménages. Les données portant sur les véhicules des ménages interrogés sont ainsi rattachées au niveau du ménage. Par véhicule, les kilomètres totaux parcourus ainsi que les consommations unitaires sont disponibles. Il est donc possible d'obtenir pour chaque ménage ses consommations selon le carburant utilisé et bien entendu le nombre total de véhicules ;
3. enfin, le pseudo-appariement des deux bases de données d'enquête est réalisé. Les deux bases ne sont pas appariables directement, car les ménages interrogés ne sont pas les mêmes : elles ne comportent donc pas d'identifiant ménage commun. Les bases sont donc pseudo-appariées au sens où à chaque ménage de l'ENL 2013 est associé un ménage de l'EMP 2019. En sortie de ce traitement statistique, la base de données ménages obtenue s'apparente à une nouvelle base, à la fois redressée en partie et contenant pour chaque ménage des informations sur ses caractéristiques socio-démographiques, sa résidence principale et ses voitures.

La microsimulation : vieillissement et application du système de prix

La partie microsimulation proprement dite est composée de deux blocs principaux exécutés à chaque utilisation du modèle, de deux blocs complémentaires utilisés de manière optionnelles et de deux blocs de formatage des données qui permettent leur affichage. Les deux blocs principaux sont les suivants :

- un bloc « vieillissement » utilise des données annuelles agrégées de l'Insee et du SDES comme données de calage pour actualiser la base initiale (pseudo-appariement de l'ENL 2013 et de l'EMP 2019). Côté logement sont ainsi actualisés le parc de logements et leurs consommations unitaires ainsi que le parc de moyens de chauffage. Côté transport, le bloc « vieillissement » permet d'actualiser le parc automobile (la répartition des carburants en son sein) et les pratiques de mobilité des ménages (kilomètres parcourus, part des véhicules utilitaires légers utilisés par les ménages, consommations unitaires). Enfin, les revenus des ménages sont aussi actualisés. Au total, les consommations énergétiques de chaque ménage sont vieillies dès que l'année est disponible ;
- un bloc « prix et fiscalité » des énergies permet de calculer les factures énergétiques à partir des consommations issues du bloc précédent et de l'ensemble des prix et taxes renseignés pour chaque année disponible.

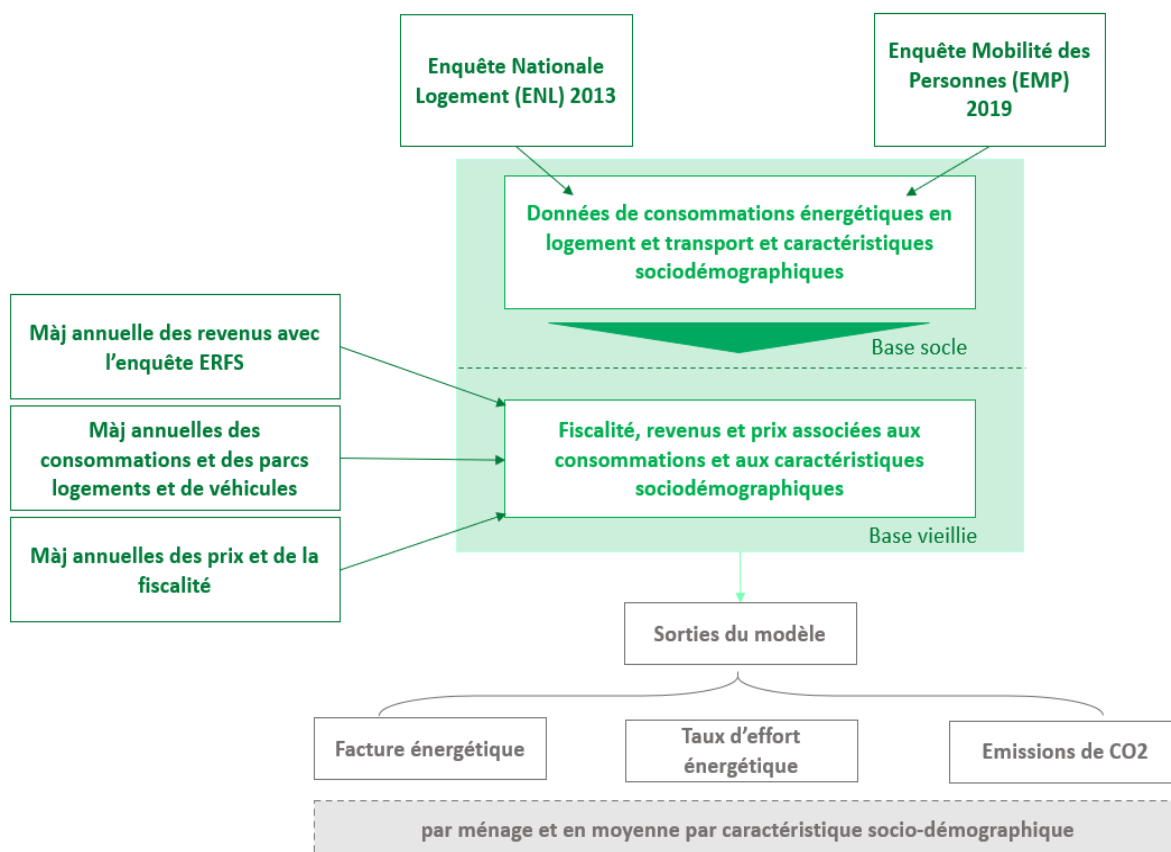
Les deux blocs complémentaires utilisés selon le problème étudié sont les suivants :

- un bloc « élasticité » qui intègre au modèle des élasticités-prix. Ces dernières permettent de modéliser une réaction sur les consommations d'énergie des ménages en matière de logement et de transport à la suite d'un choc de prix de l'énergie ;
- un bloc « compensation » qui permet de modéliser le versement du chèque énergie ou d'autres transferts aux ménages selon leurs caractéristiques (par exemple énergie principale de chauffage, dixième de niveau de vie), mais aussi de distinguer les ménages précaires et d'obtenir le taux de précarité énergétique.

Enfin, deux blocs de formatage existent et sont les suivants :

- un bloc « sortie » qui permet le formatage et la mise au propre des variables en sortie (nom parlant, format en caractère ou nombre, certains regroupements, etc.) ;
- un code qui fait le lien pour intégrer les sorties du modèle avec une application Rshiny de datavisualisation développée pour un usage interne³³, laquelle permet des calculs de grandeurs statistiques sur des sous-catégories de population.

Figure 2 : structure simplifiée du modèle Prometheus



Source : CGDD, modèle Prometheus

³³ Cette application très utile a été développée par Kendrick Herzberg en 2023.

Le calcul des factures énergétiques des ménages

Prometheus est un modèle qui se focalise sur les factures énergétiques des ménages (*encadré 3* pour une courte présentation d'une approche plus large). En matière de logement, les factures sont principalement celles portant sur le chauffage, et selon une gamme détaillée d'énergies (électricité, gaz, fioul, bois, etc.), et celle portant sur l'utilisation de l'électricité pour les usages domestiques (éclairage, cuisson des aliments...). En matière de transport, Prometheus modélise les factures de carburants (essence, diesel, ...) des véhicules utilisés par les ménages.

Le modèle permet de décliner ces factures selon des variables propres aux ménages : socio-démographiques ou caractéristiques de leur logement et de leurs véhicules. Ces variables sont issues des bases initiales sur lesquelles se fondent le modèle (l'ENL 2013 et l'EMP 2019). Les revenus sont également pris en compte, ils correspondent à l'ensemble des revenus et prestations perçus par les ménages avant la redistribution opérée par les prélèvements directs (non disponibles dans les données utilisées). Le champ est celui des ménages ordinaires vivant en France métropolitaine. Les consommations d'énergie du logement sont celles des résidences principales des ménages, excluant dès lors les résidences secondaires ou certains types de logement, notamment collectifs (prisons, résidences collectives pour personnes âgées, résidences étudiantes etc.). Les consommations de carburants sont celles des ménages et des entrepreneurs individuels. Les consommations réalisées via des transports autres que les véhicules particuliers (transport en commun, avion, etc.) sont exclues.

Ainsi, à partir des données externes comme les consommations en énergie, les caractéristiques des parcs de logements et de véhicules, et en tenant compte de la législation fiscale (TVA, taxes) ainsi que des prix en vigueur pour une année (prix moyen, détaillé par exemple en parts variables ou abonnement, etc.), Prometheus calcule au niveau de chaque ménage :

- les consommations énergétiques directes en unités physiques ;
- les émissions de CO₂ afférentes, le cas échéant, à ces consommations ;
- les factures TTC, ainsi que la décomposition entre le prix HT, la TVA et les différentes taxes spécifiques à l'énergie (accises sur les énergies, dont composante carbone dans le cas des énergies fossiles, CTA dans le cas de l'électricité, etc.).

Le modèle permet ainsi de calculer ces trois types de grandeur pour l'ensemble des ménages ainsi que selon différentes dimensions d'hétérogénéité verticale (dixième de niveau de vie) et horizontale (zone d'habitation, caractéristique du logement, distance au transport en commun, nombre de véhicules, âge de la personne de référence du ménage, etc.). Il est également possible de calculer des « taux d'efforts », c'est-à-dire de rapporter au revenu du ménage la facture énergétique, ou sa seule composante fiscale.

Encadré 3 : la notion de coût résidentiel

Les dépenses énergétiques, c'est-à-dire de chauffage, de transport, etc. peuvent être vues comme des dépenses de première nécessité, au même titre que la nourriture. Mais, elles peuvent aussi être considérées comme des coûts variables, en ce qu'elles dépendent de variations de prix qui s'ajoutent à des coûts relativement fixes portant sur un périmètre similaire : le loyer, les impôts locaux, les frais d'assurance et d'entretien de l'automobile, etc. Cet ensemble large de coûts liés au logement et aux déplacements constitue le « coût résidentiel », notion qui provient des travaux d'économie urbaine menés aux États-Unis dans les années 1990¹. Par exemple, un ménage périurbain peut avoir des dépenses en carburant élevées liées à une mobilité contrainte forte, mais celles-ci peuvent être compensées par un loyer ou un prix d'achat du logement plus faible (Nicolas et al. (2012)). Le coût résidentiel constituerait ainsi l'échelle d'analyse pertinente pour évaluer leur degré de précarité ou de vulnérabilité énergétique.

Toutefois, les limites de cette approche tiennent concrètement à la disponibilité des données permettant de mesurer toutes ces dimensions, a fortiori si l'on veut mener une analyse géographique territorialisée et locale, mais elles sont aussi plus conceptuellement liées aux limites à poser à l'analyse. En effet, étendre excessivement le champ des dépenses considérées peut finalement aboutir à retomber sur des problématiques plus générales liées au niveau de vie et à la pauvreté.

¹ Pour la France, voir les travaux de Maresca (2013). Nous signalons également une thèse en préparation au laboratoire de recherche LAET associé à l'ENTPE, intitulée *Coût résidentiel, vulnérabilité des ménages et politiques publiques territoriales* (Maria Besselièvre).

III. BASES DE DONNÉES D'ENQUÊTES MOBILISÉES

A. L'enquête nationale logement de 2013

L'une des sources de données constitutives de Prometheus est « l'enquête nationale logement » (ENL) de l'Insee. Elle sert de base au modèle grâce à son large échantillon et à l'ensemble des données permettant de caractériser et de distinguer pour les ménages les factures énergétiques, les équipements du logement et le parc de logement.

La toute première version de Prometheus s'appuyait sur l'ENL de 2006, tandis que la version actuelle s'appuie sur l'ENL 2013³⁴. C'est une enquête au statut d'intérêt général et de qualité statistique à caractère obligatoire³⁵, réalisée par l'Insee à intervalles réguliers (tous les quatre ou sept ans depuis 1955). Elle constitue la source nationale statistique majeure pour mesurer l'évolution des conditions de logement des ménages en France entière, et détailler ainsi leurs consommations énergétiques et équipements associés pour leur résidence principale. Le millésime 2013 a donné lieu, au moment de sa mise à disposition, à la publication *Les conditions de logement fin 2013 – Premiers résultats de l'enquête logement*, Insee, avril 2015³⁶.

Champ de l'enquête et échantillonnage

Le champ de l'enquête porte sur les logements « ordinaires » occupés à titre de résidence principale, y compris les logements-foyers pour personnes âgées, et excluent des logements collectifs tels que les foyers, internats ou casernes, ainsi que les habitations mobiles. L'unité statistique sondée est celle du logement, tandis que l'unité interrogée est le ménage.

Les réponses ont été collectées en face à face avec un enquêteur, soit en collecte assistée par informatique (Capi) entre juin 2013 et juin 2014³⁷. Les fiches d'adresses ont été tirées dans deux fichiers différents : dans la base de sondage principale qui est l'enquête annuelle de recensement (EAR) de 2011, et dans un fichier recensant les permis de construire pour suivre les logements neufs (un appariement entre deux fichiers Sitadel10 et des données de la DGI).

L'édition 2013 de l'ENL de l'Insee pour la France métropolitaine contient 27 000 logements répondants. D'autres échantillons ont également été tirés pour les DOM (9 000), Mayotte (3 000) et des extensions de l'échantillon pour permettre une représentativité de certaines régions ou zones urbaines sensibles (15 000)³⁸.

³⁴ Au moment de la rédaction de ce document de travail, l'ENL réalisée exceptionnellement par le SDES portant sur les années 2019 à 2021 n'est pas complètement disponible, en particulier les modules énergies utilisés dans Prometheus, ainsi que les revenus consolidés afférents. La crise sanitaire de 2020 a en particulier retardé le calendrier de production de l'enquête. Par ailleurs, l'Insee est chargé de la réalisation de l'ENL 2023-2024, avec, prévisionnellement, une phase d'enquête en 2023/2024, une publication de référence fin 2025, et une mise à disposition des données ensuite : www.cnis.fr/enquetes/enquete-logement-2023-2024-2023a075lo/.

³⁵ www.cnis.fr/enquetes/logement-2013-enquete/.

³⁶ www.insee.fr/fr/statistiques/1287961?sommaire=1912749.

³⁷ Pour le questionnaire, voir : www.insee.fr/fr/metadonnees/source/fichier/DOC_ENQ_LOGEMENT_2013_QUEST.pdf.

³⁸ Pour plus de détail, voir : www.insee.fr/fr/metadonnees/source/operation/s1251/presentation.

Points d'attention méthodologiques

De nombreuses données ont été redressées de manière classique pour corriger l'effet de la non-réponse, réduire les biais de l'enquête et de la base de sondage, ainsi que l'aléa du plan de sondage. Les variables sont principalement les mêmes que pour l'ENL 2006.

Une variable de revenu total du ménage a été construite pour la première fois depuis la création de l'ENL. Elle combine les informations déclarées par le répondant et les sources fiscales de 2013, via un appariement et un calage sur l'enquête revenus fiscaux et sociaux (ERFS). Cette variable (MRTOTA3) concerne tous les ménages à l'exception de ceux dont la personne de référence est étudiante et ceux à revenus négatifs³⁹.

Le taux de réponse de l'enquête est de 70,2 % pour la France métropolitaine (périmètre de Prometheus). Les résultats de l'ENL 2013 utilisés par Prometheus couvrent environ 27 000 ménages représentatifs des 29 millions de ménages, sans représentativité infra nationale. Le questionnaire de l'ENL 2013 permet donc d'intégrer à Prometheus de nombreux aspects ménages et logement : les caractéristiques des ménages (la composition du ménage, les revenus perçus, etc.), les caractéristiques physiques et la qualité du parc de logement (chauffage, taille, confort sanitaire, sécurité, bruit, qualité de l'équipement, etc.), les dépenses associées, ainsi que certains aspects juridiques d'occupation (aides perçues par l'État, législation de loyer, propriété, etc.).

B. L'enquête mobilité des personnes de 2019

L'autre base de données constitutive du modèle est « l'enquête mobilité des personnes » (EMP). Elle sert de complément à l'ENL 2013, afin de caractériser et de distinguer les factures de carburant, le parc de véhicules, et certaines informations sur les pratiques de mobilité des ménages.

La 1^{re} version de Prometheus n'intégrait pas la microsimulation de la facture de carburant, se limitant à celle des énergies du logement. En revanche, la deuxième version du modèle⁴⁰ a intégré « l'enquête nationale des transports et déplacements » (ENTD) de 2008. S'inscrivant dans le cadre des enquêtes nationales de transport depuis les années 1960⁴¹, l'ENTD a été conduite par le SDES en 2007/2008⁴².

L'enquête « mobilité des personnes », qui succède l'ENTD 2008⁴³, a quant à elle été conduite par l'Insee et exploitée par le SDES, entre 2018 et 2019. Reconnue d'intérêt général et de qualité statistique à caractère obligatoire⁴⁴, sa production et sa mise à disposition en libre accès⁴⁵, s'est accompagnée de la publication *Comment les Français se déplacent-ils en 2019 ? Résultats de l'enquête mobilité des personnes*, SDES, 2020⁴⁶.

³⁹ Sur ces différentes précisions méthodologiques et pour le dictionnaire des variables, voir *enquête logement 2013 - Dictionnaire des codes* : https://data.progedo.fr/documents/adisp/lil-1022/dico_codes_enl13_1710.pdf.

⁴⁰ Comme déjà mentionnée, cette version n'a pas donné lieu à un document de travail, néanmoins une présentation de la mise à jour avait été faite dans le cadre des « Journées de méthodologie statistique » de l'Insee en 2018 : www.jms-insee.fr/2018/S06_4_PPT_PEROUMAL_JMS2018.pdf.

⁴¹ Voir : www.insee.fr/fr/metadonnees/source/serie/s1277.

⁴² Pour plus de détail sur cette enquête, voir : www.statistiques.developpement-durable.gouv.fr/enquete-nationale-transports-et-deplacements-entd-2008.

⁴³ Voir l'avis du CNIS à ce sujet : www.cnis.fr/wp-content/uploads/2017/03/AC_2017_sdes_EMP_2018.pdf.

⁴⁴ www.cnis.fr/enquetes/enquete-mobilite-des-personnes/.

⁴⁵ Les données et questionnaire sont notamment disponibles : www.statistiques.developpement-durable.gouv.fr/resultats-detailles-de-lenquete-mobilite-des-personnes-de-2019.

⁴⁶ www.statistiques.developpement-durable.gouv.fr/comment-les-francais-se-deplacent-ils-en-2019-resultats-de-lenquete-mobilite-des-personnes.

Par rapport à l'ENTD 2008, de nombreuses informations supplémentaires ont été ajoutées pour enrichir l'enquête : les aires d'attraction des villes, les types d'unités urbaines, l'accès aux transports en commun, le nombre de kilomètres parcourus en vélo (il y a ainsi une table de données spécifiques sur les vélos, même si l'information sur les kilomètres parcourus n'est disponible que pour la semaine précédant l'enquête), etc.

Champ de l'enquête et échantillonnage

Le champ de l'enquête porte sur les ménages ordinaires pour ce qui relève de l'équipement en véhicules et des moyens de transports utilisés, et sur les individus pour ce qui concerne les déplacements. Elle inclut les individus âgés de six ans ou plus résidant en France métropolitaine. L'unité statistique sondée est le logement, et les unités enquêtées sont le ménage ainsi que le ou les individus le composant.

Le mode de collecte est en face à face et l'enquête s'est déroulée entre avril 2018 et avril 2019, en six vagues de deux mois. Les fichiers de tirage des échantillons sont constitués du recensement de la population (Octopusse⁴⁷) et de fichiers fiscaux, afin que l'enquête soit représentative également des quartiers prioritaires de la ville (QPV). Deux échantillons sont ainsi tirés, un échantillon principal, constitué de 20 000 unités, et un échantillon supplémentaire constitué de 1 150 unités⁴⁸. L'échantillon supplémentaire est utilisé comme échantillon de contrôle. Le taux de réponse de l'enquête est de 75,6 % pour la France métropolitaine. Les résultats de l'EMP 2019 couvrent ainsi près de 14 000 ménages représentatifs des 29 millions de ménages.

La méthodologie de production des deux enquêtes (ENTD 2008 et EMP 2019) est très similaire, même si quelques points diffèrent et méritent des précisions s'agissant de l'EMP 2019.

Points d'attention méthodologiques

ENJEUX LIÉS AUX VARIABLES RELEVANT DES INDIVIDUS KISH

Le questionnaire de l'EMP 2019 comporte quatre parties :

- la 1^{ère} partie concerne le ménage dans son ensemble (caractéristiques du ménage, lieu d'habitation, distance aux transports en commun, etc.) et recense les véhicules et les informations en lien avec ces derniers (caractéristiques, kilomètres parcourus par an, kilométrage total au compteur, etc.) ;
- les trois dernières parties du questionnaire concernent uniquement l'individu Kish (voir ci-dessous) sélectionné : elles recensent ses caractéristiques et ses déplacements et voyages (abonnement, mobilité locale et mobilité longue distance).

Pour Prometheus, les informations rattachées aux véhicules et aux ménages sont primordiales, en particulier concernant le nombre de kilomètres parcourus par an par les voitures de chaque ménage.

Ainsi, lorsque l'EMP 2019 utilise la méthode d'interrogation individuelle sur un individu « Kish » cela ne permet pas d'avoir l'information exhaustive sur un ménage. En effet, ce mode d'interrogation permet, avec le système de pondération, d'obtenir des statistiques agrégées nationales, sur un jour ou une semaine type. Toutefois, les réponses des individus Kish ne sont pas représentatives de la réponse du ménage où ils résident. Il n'est donc pas possible d'assimiler leur réponse au niveau du ménage.

⁴⁷ www.insee.fr/fr/information/5211274.

⁴⁸ Pour le détail méthodologique, voir : www.statistiques.developpement-durable.gouv.fr/enquete-sur-la-mobilite-des-personnes-2018-2019.

Il est seulement possible d'en faire des variables de comptage en tant que bornes inférieures. En effet, si le l'individu Kish du ménage possède une carte d'abonnement au transport en commun par exemple, alors au moins un membre du ménage en possède une, cependant s'il n'en possède pas, a contrario cela ne signifie pas qu'aucun membre du ménage n'en possède une.

Ainsi de nombreuses informations sont disponibles au niveau individuel sans qu'on ne puisse les utiliser dans Prometheus : déplacement sur les sept jours précédents, possession d'une carte d'abonnement transport, covoiturage, mobilité longue distance etc. L'information peut aussi être parcellaire.

Par ailleurs, le nombre de jours en étude ou garderie ne sont plus exhaustifs, ni au niveau individuel, et encore moins au niveau ménages (au contraire de ce qui était possible dans l'ENTD 2008) : une variable de mobilité contrainte en nombre de kilomètres n'est alors plus constructible comme elle l'était à partir de l'ENTD 2008.

REDRESSEMENT DES VALEURS MANQUANTES

En plus des méthodes traditionnelles de redressement, certaines données administratives ont pu être utilisées pour renforcer les traitements post-enquête, en particulier sur la non-réponse⁴⁹. Le répertoire statistique des véhicules routier (RSVERO⁵⁰), diffusé par le SDES, a ainsi pu se substituer au besoin de questionner le ménage concernant son ou ses véhicules, si le numéro d'immatriculation déclaré était disponible dans ce répertoire.

Les nombreux traitements sur la non-réponse et les biais, expliquent aussi, notamment sur les variables d'intérêt pour Prometheus, que les données apparaissent comme mieux remplies et plus robustes que dans l'ENTD 2008.

AUTRES POINTS : NON-REPRÉSENTATIVITÉ LOCALE ET PRISE EN COMPTE DES EFFETS DU COVID-19

Enfin, l'EMP 2019 n'est pas représentative des régions. Un dernier point d'attention important est que l'enquête a été réalisée avant l'épidémie Covid-19 de 2020 : les résultats portent donc sur la mobilité avant la crise sanitaire. Les éléments comme le télétravail et le covoiturage sont marginaux et peu exploitables par exemple.

C. Comparaison avec d'autres enquêtes

Les deux enquêtes présentées et utilisées par le modèle sont des enquêtes labellisées, c'est-à-dire validées par le Comité du label du CNIS. Elles font ainsi partie des enquêtes phares du Service statistique publique (SSP). Même si les factures énergétiques produites par Prometheus ne constituent pas des données observées, mais simulées, les bases initiales qui sous-tendent le modèle représentent des données réelles observées, permettant ainsi de quantifier et modéliser au plus près de la réalité de nombreuses dimensions relatives au logement et au transport, et en particulier celles impactant les factures énergétiques.

Comparaison avec l'enquête budget de famille

Pour construire Prometheus, le choix de l'ENL 2013 et de l'EMP 2019 a été préféré à celui de l'enquête « budget de famille » (BDF), qui aurait pu constituer une source alternative (*voir le précédent document de travail de Prometheus*). C'est aussi une enquête phare de l'Insee, sur laquelle repose en partie par exemple le modèle de microsimulation Ines.

⁴⁹ En revanche, les kilométrages des deux roues, également disponibles, ne sont pas consolidés.

⁵⁰ RSVERO recense le nombre de véhicule du parc automobile français (véhicules routiers, véhicules automoteurs) en utilisant l'immatriculation des véhicules et des informations du ministère de l'intérieur, voir : www.statistiques.developpement-durable.gouv.fr/rsvero-repertoire-statistique-des-vehicules-routiers.

L'enquête BDF est réalisée tous les cinq ans et permet de reconstituer l'intégralité de la comptabilité des ménages, c'est-à-dire à la fois leurs dépenses et leurs ressources, pour les ménages résidant en France (Métropole et DOM)⁵¹. La dernière édition en date porte sur l'année 2017⁵². BDF a l'avantage de couvrir l'ensemble des dépenses d'un ménage, en cohérence avec la comptabilité nationale, et utilise pour cela la nomenclature *European Classification of Individual Consumption by Purpose*, dite ECOICOP⁵³. En outre, elle permet de mieux couvrir certaines dépenses prises dans leur ensemble : par exemple, BDF permet de mieux quantifier l'ensemble des dépenses de transports (carte transport, dépenses avions, pièces détachées de véhicules, etc.) et de relier les montants associés à ceux de la comptabilité nationale. En particulier :

- contrairement à BDF, et c'est une limite, les données utilisées par Prometheus ne permettent pas de replacer les dépenses énergétiques au sein de l'ensemble des dépenses des ménages. En revanche, les données de Prometheus sont plus précises sur deux aspects très spécifiques que sont le logement et le transport, ce qui rend le modèle très robuste et quasi exhaustif dans le domaine des factures énergétiques. Ainsi, BDF ne distingue pas systématiquement les factures d'électricité de celles de gaz, et ne les détaille pas selon les usages ;
- sur le champ métropolitain en particulier, l'échantillon de logements est plus élevé pour l'ENL 2013 : 42 500 ménages enquêtés, contre 20 700 pour BDF 2017⁵⁴. L'EMP 2019 dispose aussi d'une taille d'échantillon conséquente (20 000 ménages).

Enfin, le taux de réponse est plus élevé côté ENL 2013 et EMP 2019 (plus de 70 %) que pour BDF (environ 50 %). Il est à noter que le prochain millésime de BDF sera celui de 2026, pour une disponibilité en 2028⁵⁵.

Comparaison avec l'enquête performance de l'habitat, équipements, besoins et usages de l'énergie (Phébus)

Au moment de la construction de la première version de Prometheus, l'enquête Phébus du SDES n'était pas disponible, mais constituait une base de données prometteuse pour le développement du modèle. Il s'agit d'une enquête d'intérêt général et de qualité statistique à caractère obligatoire⁵⁶, qui s'inspire d'autres enquêtes européennes.

Le seul millésime est celui de 2013 (portant sur 2012). Cette enquête permet d'étudier les performances énergétiques des résidences principales des ménages, selon les caractéristiques de ces derniers, y compris de leurs équipements domestiques et automobiles. Elle permet également de quantifier les consommations et factures domestiques des ménages (*Consommations énergétiques des ménages en 2012*, CGDD, 2015⁵⁷).

Cependant, l'échantillon est faible : environ 8 000 logements⁵⁸, et nombre de variables est plus limitée que dans l'ENL. Comme le note *Douenne* (2018) la très petite taille de l'échantillon et la quantité limitée d'information de cette enquête ne permet pas une exploration fine des déterminants des effets distributifs horizontaux. Enfin, cette enquête n'a pas été reconduite

⁵¹ Plus de détail ici : www.insee.fr/fr/metadonnees/source/serie/s1194.

⁵² www.insee.fr/fr/metadonnees/source/operation/s1341/presentation.

⁵³ www.insee.fr/fr/information/2402696.

⁵⁴ 15 800 pour Bdf 2011.

⁵⁵ www.cnis.fr/enquetes/enquete-budget-de-famille-2026-en-cours-de-labellisation/.

⁵⁶ www.cnis.fr/enquetes/performance-de-lhabitat-equipements-besoins-et-usages-de-lenergie-phebus-enquete-sur-la/.

⁵⁷ www.statistiques.developpement-durable.gouv.fr/consommations-energetiques-des-menages-en-2012.

⁵⁸ Voir méthodologie : www.statistiques.developpement-durable.gouv.fr/enquete-performance-de-lhabitat-equipements-besoins-et-usages-de-lenergie-phebus.

et il n'est pas prévu qu'elle le soit. Par conséquent elle n'a pas été utilisée depuis la création du modèle et dans ses développements ultérieurs.

Une approche fine et cohérente

L'ENL 2013 et l'EMP 2019 ont chacune des liens étroits avec les données agrégées de consommations annuelles d'énergie produites par le SDES. Ces liens facilitent à la fois le travail de vieillissement des données (*voir ci-dessous*) et la cohérence d'ensemble du modèle :

- les consommations agrégées concernant les logements produites par le SDES utilisent dans leur construction des informations des enquêtes ENL (2013 et celles à venir) ; par exemple les nomenclature d'énergie sont assez proches ;
- les comptes du transport, qui servent au vieillissement des données du modèle du côté du transport (données de kilométrages, consommations unitaires, etc.), sont produits au SDES par les mêmes producteurs que l'EMP 2019, ces deux productions s'appuyant en particulier sur le répertoire administratif RSVERO concernant les consommations des ménages.

Au total, l'ENL 2013 et de l'EMP 2019 ont l'avantage d'être des données précises, fines et détaillées, apportant tout un ensemble de variables sociodémographiques annexes qui permettent de caractériser au mieux les dépenses énergétiques. Le vieillissement de la base initiale formée du pseudo-appariement de l'ENL 2013 et de l'EMP 2019 utilise des données qui sont relativement bien homogènes et cohérentes avec ces enquêtes, notamment en termes de nomenclature des énergies. Ainsi, le pseudo-appariement entre ces deux enquêtes permet à la fois d'obtenir une solide base représentative des ménages français et d'ouvrir de nombreuses perspectives dans l'optique de la microsimulation.

IV. TRAITEMENT DES BASES

Ce chapitre présente les traitements qui ont été réalisés sur l'ENL 2013 et l'EMP 2019 avant de les pseudo-appariées, afin que les enquêtes puissent être utilisées et exploitées dans Prometheus.

A. Traitements de l'enquête nationale logement 2013

Dans ce chapitre, l'ENL 2013 est présentée, ainsi que les différentes étapes qui ont été réalisées d'une part afin de redresser de nombreuses valeurs manquantes concernant les variables utiles pour Prometheus, d'autre part pour permettre d'obtenir des consommations et in fine des factures énergétiques selon la nomenclature voulue. Les traitements sur l'ENL 2013 sont relativement importants. Ces traitements sont classiques, toutefois certains choix qui ont prévalu à l'époque peuvent être discutés. Néanmoins, ce document de travail et cette *version 3* du modèle ne vise pas à modifier le traitement de l'ENL 2013, mais à les expliciter, en l'attente d'une future actualisation de la partie logement de la base ménages utilisée par Prometheus.

Champ et objectifs

L'ENL 2013 servant de base au modèle Prometheus, le champ de ce dernier correspond à celui couvert par cette enquête, c'est-à-dire les ménages vivant en logements dits « ordinaires ». Sont donc exclus les ménages vivant en collectivité (foyers, prisons, hôpitaux...), ainsi que les personnes vivant dans des habitations mobiles (mariniers...) et les sans-domiciles. Par ailleurs, seuls les ménages vivant en France métropolitaine sont pris en compte par le modèle (comme évoqué supra, un échantillonnage sur les DOM est présent, mais l'enquête n'est pas représentative par région et l'ENTD 2008 tout comme l'EMP 2019 ne portent que sur la France métropolitaine).

L'ENL 2013 fournit les factures énergétiques déclarées par les ménages, mais pas les volumes de consommation. Le traitement principal apportée à l'ENL 2013 consiste donc à calculer les volumes de consommation à partir des factures d'énergie⁵⁹. Seules les factures, et donc les consommations, des résidences principales sont considérées. Toutes les consommations liées aux résidences secondaires ou à un usage professionnel sont donc exclues.

Concrètement, le programme de traitement de l'ENL 2013 s'appelle *Bloc_Logement* et il est composé de trois sous-blocs principaux et d'un petit bloc qui traite les revenus :

- préalable : import des tables, suppression des ménages non métropolitains ;
- le 1^{er} sous-bloc attribue à chaque ménage de l'enquête une énergie principale de chauffage ;
- le 2^e sous-bloc traite et redresse les factures d'énergie déclarées par les ménages ;
- le 3^e sous-bloc calcule les consommations énergétiques des ménages à partir des factures, en attribuant au préalable un tarif ou un prix de l'énergie à chaque ménage ;
- le bloc complémentaire recalcule la variable de revenu sur le niveau de vie.

⁵⁹ L'ensemble de ces traitements a été effectué en 2016 et 2017.

Préalable : import des tables

Le préalable aux traitements est la récupération des variables utiles issues des trois tables indiquées dans le *tableau 1* afin de les fusionner dans une nouvelle table après les avoir triées par la variable identifiante ménage. Ne sont conservés que les ménages métropolitains⁶⁰.

Tableau 1 : tables en entrée du bloc logement

Nom de la table	MENLOG_DIFF	MENLOG_COMP	IDLOGZC
Définition de la table	Chaque ligne de la table correspond à un ménage, avec une unité de vie principale.	Variabes complémentaires à celles de la table MENLOG_DIFF, avec le même champ.	Variabes de la zone climatique « officielle » du ministère. Elle est aussi utilisée par le Centre d'études et de recherches économiques sur l'énergie.

Premier sous-bloc : traitement des énergies

L'attribution d'une énergie principale de chauffage à chaque ménage, laquelle n'est pas disponible en tant que telle dans l'enquête, se déroule en trois temps, à partir de : 1) du type de chauffage (individuel ou collectif) du logement ; 2) du mode de chauffage principal du ménage (chaudière, pompe à chaleur, etc.) ; 3) de l'énergie principale de chauffage associée à ces types et modes de chauffage (fioul, gaz, électricité etc.). Des traitements similaires et simplifiés sont effectués concernant la cuisson et l'eau chaude (*encadré 4*).

LE TYPE DE CHAUFFAGE : INDIVIDUEL OU COLLECTIF

Il est essentiel de distinguer les logements en chauffage collectif de ceux chauffés individuellement. En effet, lorsque le chauffage est collectif, les factures énergétiques déclarées par le ménage dans l'enquête n'intègrent pas les dépenses associées aux consommations de chauffage : celles-ci font partie des charges collectives payées par le ménage. Il est souvent difficile, d'une part, d'identifier les ménages en chauffage collectif (ils ne se déclarent pas nécessairement comme tels), mais aussi, d'autre part, de connaître précisément l'énergie de chauffage utilisée par ce chauffage collectif (souvent mal connue des ménages), ainsi que les dépenses associées pour le ménage (qu'il faut être capable d'isoler des autres charges collectives). Il faut donc opérer des redressements spécifiques pour les ménages en chauffage collectif.

Prometheus considère ainsi que le chauffage du ménage est :

- un chauffage individuel si le ménage déclare dans l'enquête⁶¹ disposer du mode de chauffage « chaudière individuelle », « chauffage tout électrique », « appareils indépendants » ou « autre moyen de chauffage » ;
- un chauffage collectif si le ménage déclare disposer du mode de chauffage « chauffage ou chaufferie collective » ou bien « chauffage mixte »⁶², ou si le ménage déclare des charges collectives de chauffage non nulles dans l'enquête⁶³, et ce même s'il n'a pas déclaré au préalable être en chauffage collectif ou mixte⁶⁴.

⁶⁰ Variable DOM de l'enquête, avec comme valeur 2 pour les ménages métropolitains.

⁶¹ Variables KMOD_1 à KMOD_7 de l'enquête pour les modes de chauffage.

⁶² On considère en effet les ménages en chauffage mixte comme étant en chauffage collectif : la proportion des ménages en chauffage mixte déclarant par la suite que le chauffage est inclus dans leurs charges collectives (8 ménages sur 10) est équivalente à la proportion des ménages en chauffage collectif dans le même cas.

⁶³ Variables LCHAUF ou FCHAUF de l'enquête pour respectivement les ménages locataires et propriétaires.

⁶⁴ On considère alors qu'il a mal répondu à la question sur le mode de chauffage (probablement par méconnaissance), mais que le fait qu'il déclare des charges collectives de chauffage prévaut.

Tableau 2 : répartition du type de chauffage entre individuel et collectif en 2013

Type de chauffage	Individuel	Collectif
Part	82 %	18 %

Source : CGDD, modèle Prometheus à partir de l'ENL 2013

LE MODE DE CHAUFFAGE

L'enquête interroge chaque ménage sur les modes de chauffage dont il dispose dans son logement (chaudière individuelle, appareils indépendants, etc.) et les émetteurs de chaleur associés à ces modes⁶⁵ (radiateurs, poêle, etc.).

L'enquête interroge également sur le type de chaudière ou installation centrale individuelle⁶⁶ (chaudière, pompe à chaleur, solaire thermique ou autres), ce qui correspond à une clarification du mode de chauffage « chaudière individuelle ».

Puis, deux variables complémentaires précisent : i) le type d'appareils indépendants pour se chauffer s'ils en utilisent⁶⁷ (radiateurs, cheminées) ii) les autres moyens de chauffage qu'ils utilisent⁶⁸ (radiateurs, cheminée...).

In fine la variable de mode de chauffage présente 17 modalités (*annexe 1*). À noter qu'il y a également une modalité « aucun moyen de chauffage ».

L'ÉNERGIE PRINCIPALE DE CHAUFFAGE

À partir de la variable mode de chauffage de l'ENL 2013 (17 modalités), on construit une variable énergie principale de chauffage (9 modalités) qui associe au ménage l'énergie utilisée⁶⁹ et non plus le mode de chauffage. Par exemple, les granulés et bûches sont regroupés au titre de l'énergie principale bois (les traitements et regroupements sont précisés en *annexe 2*). À ce stade chaque ménage se voit attribué une énergie principale de chauffage. Un traitement particulier est réalisé pour les ménages en chauffage collectif, car de nombreux déclarent mal ou ne déclarent pas d'énergie de chauffage.

Définition de l'énergie principale de chauffage dans le cas du chauffage collectif : afin d'être cohérent avec les données macroéconomiques du Ceren, l'énergie du type chauffage collectif est redressée dans l'ENL 2013 pour Prometheus. Seuls les ménages qui ont un type d'immeuble collectif⁷⁰ et qui utilisent un chauffage individuel collectif sont compris dans le redressement. Pour ces ménages, les logements dont on ne connaît pas l'énergie principale de chauffage ainsi que ceux qui déclarent l'électricité, sont basculés en chauffage principal au gaz ou chauffage urbain. La redistribution se fait par une randomisation suivant une loi normale, qui permet de s'aligner sur la distribution fournie par le Ceren entre le chauffage principal gaz et urbain (*tableau 3* pour la répartition retenue et les comparaisons avec l'ENL 2013 et les données du Ceren).

⁶⁵ Variables KEMETTER_1 à KEMETTER_11 de l'enquête pour les émetteurs de chaleur associés aux appareils indépendants de chauffage et variables KEMETTERB_1 à KEMETTERB_11 pour les autres moyens de chauffage utilisés occasionnellement.

⁶⁶ Variables KMODPRIM_1 à KMODPRIM_3 de l'enquête pour le type d'installation centrale.

⁶⁷ Variable KEMETTER.

⁶⁸ Variable KEMETTERB.

⁶⁹ Variables KCB_X, X désignant le combustible.

⁷⁰ Variables INL4 de l'enquête pour le type d'immeuble (immeuble collectif ou immeuble individuel).

Tableau 3 : répartition de l'énergie de chauffage collectif pour les ménages

En %

Part des énergies dans le chauffage collectif	Énergie principale de chauffage			
	Gaz	Chauffage urbain	Électricité	Aucun
ENL 2013	44,5	9,4	8,0	19,7
Prometheus	60,7	20,9	0,0	0,0
Ceren	63,5	23,4	0,0	0,0

Source : CGDD, modèle Prometheus

Enfin, après que chaque ménage s'est vu attribué une énergie principale de chauffage, la nomenclature retenue et les parts des énergies qui en découlent sont présentées dans le tableau 4.

Tableau 4 : part des ménages par énergie principale en 2013

Énergie principale de chauffage	Part des ménages (en %)
Gaz	35,3
Électricité	35,0
Fioul	12,8
Aucun	4,6
Bois	4,3
Autre	3,5
Chauffage urbain	1,8
GPL	1,6
Multi énergie	1,2
Total	100

Source : CGDD, modèle Prometheus à partir de l'ENL 2013

Encadré 4 : traitement de l'énergie principale de cuisson et de l'eau chaude sanitaire

Des traitements similaires à ceux présentés sur l'énergie principale de chauffage ont été effectués pour affecter une énergie principale de cuisson et d'eau chaude sanitaire². Toutefois les catégories « multi énergie » et « aucun » n'ont pas été construites. Surtout, dans la suite du traitement de l'ENL 2013 pour le passage des factures aux consommations, c'est l'information concernant l'énergie principale de chauffage qui est exploitée.

¹ KCUI_1 à KCUI_6 de l'enquête pour l'énergie principale de cuisson et KCUI_N pour le nombre d'énergies utilisées pour la cuisson.

² KCEC_1 à KCEC_9 de l'enquête pour le combustible utilisé par les appareils indépendants fournissant de l'eau chaude et KCEC_N pour le nombre de combustible utilisé.

Redressement des factures énergétiques du logement

Dans le *chapitre VIII énergie* du questionnaire de l'ENL 2013, les ménages sont interrogés sur leurs montants annuels de dépenses (non professionnelles) pour chacune des six énergies suivantes : électricité, gaz, fioul, GPL, bois et charbon⁷¹.

Les factures annuelles par énergie n'ont cependant pas fait l'objet de redressements spécifiques de la part de l'Insee dans les fichiers de diffusion de l'enquête. Seule la facture énergétique totale du logement (définie comme la somme des factures déclarées), a été redressée par l'Insee. Or, environ 20 % des ménages sont au moins concernés par une facture aberrante et le modèle Prometheus a besoin de disposer pour chaque ménage de factures détaillées par énergie.

Le modèle redresse donc ces six factures annuelles, et ajoute une 7^e facture pour l'énergie de chauffage urbain. Cette dernière correspond aux charges collectives de chauffage déclarées dans les *chapitres XI et XII* du questionnaire de l'enquête sur les ménages locataires ou propriétaires (qui n'ont pas fait non plus l'objet de redressement dans les données de l'enquête) ajoutée à celle qui correspond à l'énergie associée au chauffage collectif du ménage.

Ce second bloc consiste donc à redresser les factures individuelles par énergie : un ménage peut utiliser plusieurs énergies (gaz pour se chauffer, électricité pour s'éclairer...) et dispose potentiellement de plusieurs factures d'énergie. Il s'agit également d'imputer les factures manquantes ou aberrantes par énergie. L'objectif étant qu'in fine, chaque ménage dispose de factures cohérentes pour ce qui concerne à la fois son énergie principale de chauffage et l'électricité (quand ce n'est pas son énergie principale de chauffage).

IDENTIFICATION DES FACTURES ABERRANTES

Pour chaque énergie et chaque ménage, Prometheus calcule une facture unitaire comme étant la facture non redressée rapportée à la surface du logement⁷², ce qui correspond à la dépense annuelle au mètre carré. Les valeurs aberrantes ou trop atypiques de ces factures unitaires sont identifiées : d'une part, les valeurs extrêmes même peu nombreuses, peuvent biaiser fortement les résultats lors de l'utilisation du modèle ; d'autre part, si elles sont trop basses ou avec des valeurs nulles elles sont considérées comme aberrantes et doivent être redressées :

- factures « trop élevées » : une méthode de *winsorisation* est appliquée pour les 0,5 % de factures unitaires trop élevées qui sont alors corrigées⁷³ ;
- factures « trop faibles » : les factures nulles et les 5 % de factures unitaires les plus faibles⁷⁴ font l'objet d'une imputation de valeur (*voir infra*). En effet, dans le premier cas, elles sont considérées comme aberrantes, car il est postulé que tous les ménages consomment a minima de l'électricité dite « spécifique », c'est-à-dire l'électricité consommée pour l'éclairage et l'alimentation des appareils électroménagers dans le logement. Dans le second cas, elles sont jugées trop faibles au regard de l'utilisation intensive faite généralement de l'énergie principale de chauffage.

⁷¹ Variables COML11, COML12, COML2, COML3, COML41, COML42 (pour les logements collectifs, normalement les montants sont égaux à 0). La variable totale redressée est COML, la somme des composantes n'est pas égale à ce total.

⁷² Variable hsh1 de l'enquête, qui correspond à la surface habitable du logement.

⁷³ Cette méthode consiste à tronquer les valeurs extrêmes d'une distribution en ramenant à un seuil donné (ici le percentile 99,5, calculé pour chaque énergie) toutes les valeurs situées au-delà. Autrement dit, pour chaque énergie, les 0,5 % de factures unitaires les plus élevées sont abaissées à ces seuils.

⁷⁴ Toutes les factures en dessous de ce seuil, c'est-à-dire les 5 % des factures unitaires non nulles les plus faibles pour chaque énergie.

IMPUTATION PAR HOT DECK

Pour une énergie donnée, les ménages dont la facture de cette énergie est aberrante sont regroupés dans un groupe dit de « receveurs ». Les autres ménages sont regroupés dans un groupe de « donneurs ». L'imputation par « hot deck » consiste à utiliser une méthode économétrique dans le but d'imputer à chaque ménage « receveur », la facture réelle d'un ménage « donneur » (annexe 3 pour le détail de la méthode, la synthèse des variables retenues et les résultats des régression).

Calcul des consommations d'énergie de logement

Bien que l'ENL 2013 interroge les ménages sur leurs factures d'énergie, il ne les interroge ni sur les volumes consommés (par exemple le nombre de KWh d'électricité), ni sur les prix payés pour ces énergies. Or, pour évaluer l'impact distributifs de politiques de transition énergétique, il est nécessaire de pouvoir disposer et s'appuyer sur une photographie cohérente des consommations. En effet, le modèle de microsimulation fait vieillir les consommations afin d'obtenir les factures correspondantes. Par ailleurs, les consommations sont nécessaires pour calculer les émissions de CO₂ ou de polluants atmosphériques⁷⁵.

La construction des volumes d'énergies consommés dans Prometheus s'effectue par une décomposition prix-volume des factures énergétiques redressées. La méthode de décomposition est relativement simple et souvent directe pour le fioul, le bois ou le chauffage urbain notamment (voir *infra*). Elle est en revanche plus complexe pour les énergies de réseaux (gaz et électricité) qui sont des énergies en grande partie tarifées et où les ménages payent d'un côté une part fixe (l'abonnement) et de l'autre une part variable, fonction de leur consommation⁷⁶.

Remarque : la plupart des prix utilisés pour ce redressement (réalisé une fois sur l'ENL 2013) proviennent de la base PEGASE « Pétrole, Électricité, Gaz et Autres Statistiques de l'Énergie » du SDES⁷⁷.

VOLUMES DE CONSOMMATIONS DES ÉNERGIES À PRIX « SIMPLES » (FIOUL, BOIS, CHAUFFAGE URBAIN)

Les consommations des énergies à tarification « simples » sont reconstituées de la façon suivante, à partir de la facture :

- fioul et GPL : pour le fioul, la facture de chaque ménage est divisée par son prix unitaire. Pour le GPL, la facture est divisée par le prix unitaire en citerne ou en bouteille selon que le GPL est ou non l'énergie principale de chauffage.

⁷⁵ En outre, la plupart des taxes énergétiques sont des accises, c'est-à-dire qu'elles dépendent des volumes d'énergie consommés (hormis la TVA). La prise en compte de la réaction des ménages face aux variations des prix de l'énergie, du fait de mesures fiscales ou autres et au travers d'élasticités-prix, repose sur la donnée de leur volume de consommation d'énergie et non sur la facture.

⁷⁶ Les grilles tarifaires des énergies de réseau correspondent à des prix moyens plus élevés pour les petits consommateurs, de sorte que les volumes ne sont pas linéairement proportionnels aux factures. Un calcul fondé sur les factures individuelles avec un prix identique pour tous les ménages ne permettrait pas de reconstituer la distribution des consommations individuelles réelles.

⁷⁷ La base PEGASE a depuis été remplacée (y compris avec des changements méthodologiques et conceptuels) par les *Données mensuelles de l'énergie* (www.statistiques.developpement-durable.gouv.fr/donnees-mensuelles-de-lenergie) et l'application de diffusion DiDo) – (www.statistiques.developpement-durable.gouv.fr/catalogue?page=dataset&datasetId=631b03afb61e5c6479370169) et les séries longues annuelles publiées lors de la publication annuelle du « Bilan énergétique de la France » (www.statistiques.developpement-durable.gouv.fr/bilan-energetique-de-la-france-en-2021-synthese).

- bois : si le combustible de chauffage est le bois ou les granulés, la facture est divisée par le prix des granulés en sacs PCI⁷⁸, autrement (cas où le combustible est la bûche), la facture est divisée par le prix du bois-bûche en vrac PCI.
- chauffage urbain : on utilise un tarif de longue utilisation, en retranchant la partie abonnement et en divisant par un prix pondéré à 85 % de tarif hiver et 15 % de tarif été.

Les prix utilisés sont disponibles en *annexe 4*. Par ailleurs, les séries de prix dont elles émanent sont détaillées au *chapitre VII* de ce document de travail, car celles-ci sont généralement les mêmes utilisées pour actualiser et faire vieillir le modèle, et donc obtenir en annuel les factures à partir des consommations actualisées.

VOLUMES DE CONSOMMATIONS DE GAZ

En pratique, les ménages peuvent choisir entre plusieurs fournisseurs de gaz et, une fois le fournisseur sélectionné, ils peuvent aussi choisir parmi différents tarifs (chacun combinant un abonnement fixe indépendant de la consommation, et un tarif variable indexé sur la consommation). Il existe par ailleurs un tarif réglementé de vente (TRV), les autres tarifs étant proposés à la discrétion des fournisseurs.

Or, les tarifs choisis par les ménages ne sont pas collectés dans l'ENL 2013 : ils ont donc été reconstitués à partir de la distribution des factures. Ainsi, en faisant l'hypothèse que les ménages suivent les recommandations des fournisseurs d'énergie (qui permettent de minimiser la facture à consommation donnée), et qu'ils sont de surcroît au TRV, la répartition des tarifs est inférée à partir des factures. Les tarifs du gaz utilisés dans Prometheus sont de ce fait les tarifs réglementés de vente (*encadré 4* pour les limites induites).

Tous les ménages ayant une facture de gaz positive se voient attribuer un TRV (dans le cas contraire la consommation est mise à nulle). Deux cas sont distingués :

- les ménages avec un chauffage au gaz individuel ou consommant du gaz pour la cuisson ou l'eau chaude : la tarification du gaz de réseau comprend quatre tarifs selon son usage dans le logement. Chaque option tarifaire est composée d'un abonnement (part fixe) et d'un prix proportionnel à la consommation (part variable ou prix unitaire)⁷⁹. Pour déterminer la tarification applicable à ces ménages, on se cale sur la répartition effective des ménages telle qu'elle ressort des données réelles (données fournies par GDF Suez, voir *annexe 5*). Les ménages se voient attribuer un tarif croissant avec leurs factures de gaz ;
- les ménages avec un chauffage au gaz collectif : un tarif commun (B2S) est imputé aux ménages dont l'énergie de chauffage principale est le gaz et dont le type de chauffage principal est collectif.

Par ailleurs la facture de gaz doit être au moins égale pour chaque ménage au coût de son abonnement (correspondant au tarif attribué)⁸⁰. Pour chaque ménage, le montant de l'abonnement est déduit de sa facture, et le résultat divisé par le prix variable pour obtenir sa consommation en gaz.

⁷⁸ Pouvoir calorifique inférieur.

⁷⁹ Voir *annexe 5* pour la tarification et les montants.

⁸⁰ 1 % des ménages utilisant le gaz sont redressés.

CONSOMMATIONS D'ÉLECTRICITÉ

Tout comme pour le gaz, l'ENL 2013 ne fournit pas d'indications sur les tarifications des ménages concernant l'électricité, ceux-ci sont également considérés comme au tarif réglementé de vente (encadré 5).

Il existe deux grands types de contrats au sein du TRV de l'électricité : le contrat à tarif unique (option base) et le contrat à double tarification (heures pleines HP/heures creuses HC). Ces deux types de contrats présentent des parts fixes (abonnement) et des parts variables (prix unitaire en €/kWh), qui diffèrent selon la puissance des équipements installés. Par ailleurs, dans le cas du contrat à double tarification, les tarifs des parts variables diffèrent également selon la période de consommation (HP ou HC). Les ménages du modèle Prometheus se voient affecter le TRV de l'électricité parmi ces deux types de contrats⁸¹, dès lors que leur facture d'électricité est positive (dans le cas contraire la consommation est considérée comme nulle).

Il reste alors à déterminer au sein de chaque contrat la puissance à laquelle chaque ménage souscrit : de manière analogue que pour le gaz, il est supposé que les ménages suivent les recommandations d'Engie pour minimiser la facture à consommation donnée. Ces dernières prennent en compte plusieurs critères : équipement en appareils électriques, nombre d'habitants dans le logement, taille du logement. Les informations de l'ENL 2013 sont utilisées pour attribuer à chaque type de ménage une puissance qui leur correspond (en particulier à l'aide du type et de la surface du logement). Face à la complexité de ces grilles tarifaires, quelques hypothèses ont été formulées par souci de simplification (voir annexe 6 pour les hypothèses et les choix opérés).

Par ailleurs, la facture des ménages utilisée pour le calcul des consommations doit être au moins égale à l'abonnement souscrit⁸², et le redressement se fait sur le même principe que pour le gaz si elle est trop faible. Le volume de consommation d'électricité se déduit alors de la facture d'électricité moins la partie abonnement, divisé par le tarif variable considéré (annexe 6 pour le détail de la tarification).

Encadré 5 : les prix du gaz et de l'électricité dans Prometheus

Il est supposé dans Prometheus que les prix du gaz et de l'électricité sont ceux donnés par les tarifs réglementés de vente. Or, aujourd'hui, des offres de marché sans lien avec les tarifs réglementés sont proposés aux ménages pour le gaz et pour l'électricité, et le tarif réglementé pour le gaz n'existe plus depuis le 1^{er} juillet 2023¹. Néanmoins, en 2022, concernant le gaz, un peu plus de 50 % des ménages disposent encore d'un contrat au tarif réglementé ou avec un prix indexé sur celui-ci. La limite du modèle est moins forte en ce qui concerne l'électricité que le gaz, puisqu'en 2022, près de 80 % des ménages ont un contrat au tarif réglementé de vente ou avec un prix indexé sur celui-ci.

Une piste future, avec une nouvelle version de l'ENL, serait d'utiliser des prix moyens par type de consommateur. Des travaux en ce sens ont déjà été entrepris.

¹ La CRE publie depuis un « prix repère de vente de gaz naturel ».

⁸¹ Annexe 6 pour le détail. En revanche, par souci de simplification, les tarifs « Tempo » et « Effacement des jours de pointe » (EJP) d'EDF ne sont pas attribués. Ce dernier tarif n'est d'ailleurs plus disponible à la souscription depuis 1998. Ces deux types d'abonnement représentaient environ 3 % des sites résidentiels en 2013.

⁸² Ce redressement concerne 0,6 % des ménages utilisant de l'électricité.

B. Les revenus

La variable de revenu de l'ENL 2013 dont l'utilisation est préconisée par l'Insee, est la variable de revenu total du ménage⁸³ qui a été calculée à partir du revenu déclaré par les répondants et d'un appariement avec des sources sociales et fiscales, puis recalée sur le même concept de revenu que celui des données de ERFS, à savoir le revenu disponible après impôts et redistribution (pour plus de détail voir annexe 7). Pour calculer le niveau de vie des ménages, dans Prometheus, cette variable de revenu total, à laquelle est ajoutée le montant mensuel annualisé des aides au logement⁸⁴, et en tenant compte des unités de consommation, est calée sur la distribution de niveau de vie de ERFS (au niveau vingtiles, P5, P10 etc. jusqu'à P95)⁸⁵.

Par ailleurs, comme le revenu fiscal de référence (RFR) n'est pas fourni par l'ENL 2013, il est construit par Prometheus, en additionnant 90 % du montant total des revenus annuels d'activité et de remplacement⁸⁶ (pour tenir compte de l'abattement de 10 %), au montant total des revenus annuels fonciers⁸⁷ et tient compte aussi des UC⁸⁸.

À l'issue de ces traitements de l'ENL 2013, la base logement est prête à recevoir les informations de l'EMP 2019 par pseudo-appariement.

C. Traitements de l'enquête mobilité des personnes

Champ et objectifs

L'EMP 2019 est pseudo-appariée avec l'ENL 2013 afin de lui associer des informations sur les déplacements motorisés des ménages, et notamment sur leurs consommations en carburant, afin de calculer pour chaque ménage sa facture d'énergie du transport, et in fine sa facture énergétique totale⁸⁹.

L'EMP 2019 est traitée dans le but de créer deux tables. Une table véhicule qui recense, pour chaque ménage, l'ensemble des véhicules qu'il détient avec leurs caractéristiques (distance annuelle parcourue, type d'énergie de carburant, consommations unitaires...); et une table transport, dont l'unité d'observation est le ménage, construite à partir de la table véhicule, qui évalue la distance annuelle totale parcourue par les ménages (en nombre de kilomètres) et ce, par énergie de carburant utilisée. Les traitements apportés visent à redresser les valeurs manquantes. Le programme associé s'appelle *Bloc_Transport*.

⁸³ Variable MRTOTA3.

⁸⁴ Variable AMR, et ce, afin de respecter la définition du niveau de vie dans ERFS. On a alors la variable : $Rev_tot = MRTOTA3 + 12 * AMR$

⁸⁵ Rev_tot_uc permet ainsi d'obtenir la variable de niveau de vie NIVVIE.

⁸⁶ Variable ZRMACT3.

⁸⁷ Variable ZRFONC3. Toutefois, certains revenus de l'épargne sont imposables, d'autres non, il serait donc aussi possible d'en tenir compte. Pour l'instant, la principale utilisation du RFR est pour le calcul de l'éligibilité au chèque énergie. Or, l'expertise de l'époque a montré que le RFR calculé sans les revenus financiers (RMEPAR), avait une distribution plus proche de celle du RFR fournie par la DGFIP à la DGEC en 2014 (probablement, car les plus modestes, i.e. ceux éligibles au chèque énergie n'ont que des revenus de l'épargne non imposables (livret A etc.). Néanmoins, le modèle a construit deux concepts de RFR : un sans revenus de l'épargne, l'autre avec.

⁸⁸ La fin de ce programme contient quelques calculs portant sur des indicateurs de précarité énergétique en 2013, lesquels seront recalculés et actualisés dans la partie de microsimulation en tant que telle.

⁸⁹ Contrairement à l'ENL 2013, dans l'EMP 2019 les factures ne sont pas disponibles.

Import des tables et agrégation des données au niveau ménage

L'EMP 2019 comporte plusieurs tables :

- deux tables dont l'unité d'observation est le ménage : le tronc commun des ménages (TCM de l'Insee) qui comporte les caractéristiques des ménages⁹⁰, notamment des caractéristiques communes à l'ENL 2013 et l'EMP 2019 qui vont servir au pseudo-appariement ; et une table synthétique (q_ménage) qui présente, par ménage, le nombre de véhicules qu'il détient ;
- plusieurs tables dont les unités d'observation sont les véhicules : les tables voiture (q_voitvul), vélo (q_velo), deux-roues (q_moto) etc., qui détaillent l'équipement des ménages pour chaque type de véhicule qu'ils détiennent⁹¹. Deux autres tables portent sur les quads et les tricycles à moteurs des ménages. Celles-ci ne sont pas conservées, car ces véhicules sont très peu nombreux ;
- deux tables dont les unités d'observation sont les individus Kish⁹² interrogés et qui présentent, entre autres, des variables complémentaires sur sa possession de cartes d'abonnement transport et autoroute, sa fréquence de la pratique du télétravail et du covoiturage, sa satisfaction quant aux transports en commun.

Une fois toutes les variables d'intérêt sélectionnées et les valeurs manquantes corrigées (voir ci-dessous), la table transport est construite afin de comporter une ligne par ménage :

1. les variables de l'EMP 2019 qui sont déjà données au niveau ménage sont conservées ;
2. les variables portant sur les individus Kish sont conservées pour être utilisées comme des bornes inférieures de comptage ;
3. enfin, à partir des tables véhicules qui présentent une ligne par véhicule possédé par les ménages, des valeurs agrégées sont créées au niveau ménage.

C'est cette table qui sera ensuite pseudo-appariée avec la table de l'ENL 2013 afin de former la base socle de Prometheus.

À partir de la table voiture, les variables suivantes sont créées pour la table ménage :

- le nombre de voitures dans le ménage, décomposé par type de carburant ;
- la distance annuelle parcourue par les voitures du ménage, décomposée par type de carburant.

À partir de la table vélo, les variables suivantes sont créées pour la table ménage :

- le nombre de vélo du ménage, décomposé par usage principal (sport et promenade, trajets domicile-travail ou domicile-études et pour faire des achats ou démarches personnelles) et le nombre de vélo décomposé selon la catégorie (à assistance électrique, de course, de ville ou non) ;
- la distance totale parcourue par les vélos du ménage la semaine précédant l'interrogation⁹³ ;
- le nombre de vélos utilisé au cours des trois derniers mois dans le ménage (par fréquence⁹⁴).

⁹⁰ www.insee.fr/fr/information/6666398.

⁹¹ Dans ces tables, une ligne correspond à un véhicule et chaque identifiant ménage est répété autant de fois que ce ménage possède de véhicules.

⁹² Voir la partie présentation de l'enquête.

⁹³ Pour les vélos, la table ne présente pas la somme des kilomètres parcourus annuellement, comme c'est le cas pour les voitures.

⁹⁴ En regroupant les occurrences de la variable (KVELOFREQ).

À partir des tables deux roues et quads et tricycle à moteur, on crée pour la table ménage, les variables suivantes :

- le nombre de véhicule quads ou tricycle à moteur dans le ménage ;
- le nombre de deux roues dans le ménage.

Traitement des tables véhicules

Dans un premier temps, chaque table véhicule est traitée afin de permettre par la suite, l'agrégation des informations qu'elle contient au niveau ménage.

TRAITEMENT DE LA TABLE VOITURE

La table voiture de l'EMP 2019 est celle sur laquelle le plus de traitement ont été effectués. Cette table contient toutes les voitures des ménages interrogés dans l'EMP 2019.

SÉLECTION DES VARIABLES D'INTÉRÊT DANS LA TABLE VOITURE

De la table voiture sont extraites les variables d'intérêt pour le modèle : énergie de motorisation, consommations unitaires des véhicules, kilométrages annuels parcourus. Sont également conservées des variables caractérisant les véhicules, telles que les variables vignette Crit'Air, l'année d'acquisition, l'année de mise en service, cylindrée, la puissance fiscale ou encore la possibilité de disposer d'une prise électrique pour alimenter la voiture.

TRAITEMENTS DES VALEURS MANQUANTES ASSOCIÉES AUX VOITURES

Les traitements de la table voiture de l'EMP 2019 consistent à redresser les valeurs manquantes des consommations unitaires, des énergies et des kilométrages : 2,3 % des voitures de la base comportent une valeur manquante sur l'une au moins de ces trois variables.

Après avoir isolé les véhicules concernés, il a été convenu d'écarter de la base tous les véhicules qui n'ont pas roulé au cours des 12 derniers mois⁹⁵. Cela a permis d'écarter 11,8 % des 2,3 % de valeurs manquantes.

Les voitures sans consommations unitaires représentent alors 0,05 % des voitures de l'EMP 2019 qui ont roulé au cours de l'année précédant l'enquête. Celles sans kilométrages en représentent 0,9 %, et celles sans énergie en représente 1,3 %. Les redressements finaux de ces trois variables sont alors effectués :

- kilométrage annuel : la variable de kilométrage choisie est KVKM1ANV, car elle été redressée par le SDES à l'aide des données administratives de la base RSVERO⁹⁶. Les kilométrages sont redressés en imputant aux véhicules la moyenne des kilomètres parcourus par les véhicules dont les ménages sont du même dixième, ce choix est fait, car l'EMP 2019 montre que la distance totale parcourue en voiture croît avec les revenus. Ainsi, celle parcourue par les 40 % des ménages les plus riches est deux fois et demie supérieure à celle des 10 % des plus modestes (17 000 km par an, contre 6 700 km). Ensuite on « winsorise » avec un intervalle à 1 % (voir note de bas de page n° 87) ;
- consommations unitaires : les consommations unitaires manquantes concernent uniquement les véhicules électriques (et l'intégralité de ces véhicules). Celles-ci sont

⁹⁵ Variable et modalité : KVROUL = 2. En effet, tous les véhicules qui n'ont pas roulé au cours des 12 derniers mois présentent une valeur manquante pour ces trois variables.

⁹⁶ Le répertoire statistique des véhicules routiers recense les véhicules routiers (voitures particulières, autobus et autocars, camionnettes, camions, tracteurs routiers, remorques, semi-remorques et véhicules automoteurs spécialisés) immatriculés sur le territoire français (départements d'outre-mer compris) à partir des informations transmises par le ministère de l'Intérieur (Agence nationale des titres sécurisés) issues du système d'immatriculation des véhicules (SIV).

recodées à 17 kWh au 100 km⁹⁷ ce qui correspond à une moyenne des consommations unitaires moyennes trouvées dans la littérature ;

- énergies : au sein des véhicules dont l'énergie est manquante, celle-ci est attribuée par tirage aléatoire en conservant les proportions de chaque énergie présente dans l'enquête.

TRAITEMENT DES AUTRES TABLES DE VÉHICULES

Des traitements similaires à ceux appliqués à la table voiture ont été également réalisés sur les autres tables véhicules (vélo, deux-roues etc.).

Dans la table vélo, les variables d'intérêt conservées sont la catégorie de vélo (de ville, à assistance électrique, de course, etc.), la fréquence d'utilisation au cours de la semaine précédant l'interrogation, le mode d'utilisation (sport et promenade, trajets domicile-travail ou domicile-études et pour faire des achats ou démarches personnelles, etc.) et le kilométrage parcouru au cours de la dernière semaine précédant l'interrogation⁹⁸.

On corrige les variables caractérisant la fréquence d'utilisation, la catégorie de vélo, et l'usage du vélo⁹⁹. L'usage et la catégorie ont été corrigés conjointement en supprimant les vélos pour lesquels l'usage n'était pas renseigné sauf lorsqu'ils étaient d'une catégorie parmi les trois les plus utilisées : VTT, vélos de ville ou vélo à assistance électrique, et alors on leur a affecté l'usage majoritaire de leur catégorie de vélo. Ensuite, la fréquence d'utilisation a été corrigée en remplaçant les valeurs manquantes par le mode de la fréquence d'utilisation de la catégorie du vélo.

La variable portant sur la distance parcourue à vélo n'est pas corrigée, car son utilisation future dans Prometheus n'est pas envisagée. En effet, il paraît difficile de ramener cette variable qui concerne un vélo du ménage une semaine donné à un kilométrage annuel réalisé par le ménage. La variable est corrigée à titre indicatif.

D. Traitement des tables individus et ménages

En plus des variables issues des tables véhicules, quelques variables issues des autres tables de l'EMP 2019, sont conservées.

Les variables du tronc commun des ménages qui sont sélectionnées sont en particulier celles qui sont communes à l'ENL 2013 et qui permettront donc ensuite le pseudo-appariement. Les tables comportant des informations sur les individus Kish ne sont pas conservées, exceptées pour les variables portant sur le covoiturage, les péages et l'utilisation des transports en communs. Ces variables pourront être utilisées comme des bornes inférieures de comptage (voir III. B.). De fait, ces variables deviennent des bornes inférieures en attribuant la modalité « non » aux valeurs manquantes pour le covoiturage, la possession d'un abonnement au transport en commun ou aux péages d'autoroute des individus « Kish ».

Remarque : la variable portant sur le télétravail des individus Kish n'a pas été conservé en raison de son taux élevé de non-réponse et compte-tenu du fait que l'enquête a été conduite en 2019, soit avant le confinement du printemps 2020¹⁰⁰.

⁹⁷ Voir *Les coûts d'abattement Partie 2 – Transports, Le remplacement des véhicules conventionnels par des véhicules à faibles émissions*, France Stratégie, Commission sur les coûts d'abattement, Juin 2021. Par ailleurs, on peut trouver une estimation comprise entre 16 kWh/100km et 18 kWh/100km respectivement dans *Panorama de l'offre de véhicules : comparer l'impact carbone systémique des technologies de mobilité*, le Shift Project, Février 2020 et dans le rapport européen *Individual mobility: From conventional to electric cars* JRC Science & Policy report, 2015.

⁹⁸ Respectivement pour chacune de ces variables, le nom de la variable correspondante est KVELOCAT, KVELOFREQ, KVELOUTIL, KVKM1SEM.

⁹⁹ Respectivement KVELOCAT, KVELOFREQ et KVELOUTIL.

¹⁰⁰ 4 % des salariés pratiquaient le télétravail de manière régulière en 2019, contre 27 % en 2021 d'après Erb, Inan et al. (2022) : *Télétravail durant la crise sanitaire*, Dares Analyses, n° 9.

V. PSEUDO-APPARIEMENT¹⁰¹

Ce chapitre présente le pseudo-appariement réalisé entre l'EMP 2019 et l'ENL 2013. Sa qualité a une grande importance, car la base obtenue sous-tend le modèle de microsimulation Prometheus ainsi que les simulations qui en résultent. De plus, une spécificité de Prometheus est de ne pas actualiser chaque année ses bases de données, lesquelles ne sont pas disponibles à une telle fréquence, ce qui rend le pseudo-appariement particulièrement important. En outre, bien que peu d'exemples existent dans la littérature portant sur le pseudo-appariement conjointement à la microsimulation, ce chapitre présente en détail : la réflexion préalable, l'exécution de l'appariement lui-même et les vérifications effectuées ensuite.

A. Brève revue de littérature et choix de la méthode

Historiquement, les premiers appariements statistiques ont eu lieu au Canada et aux États-Unis dans les années 1970. Jusqu'alors contraint par des restrictions légales, le premier « *merge file* » a été réalisé en 1966 par l'*US office of Economic Opportunity* (voir Okner (1966)). Par la suite, les appariements ont été beaucoup utilisés pour des travaux empiriques portant sur de nombreux sujets (sociologiques, économiques, biomédicales, etc.). De nombreux travaux se basaient principalement sur l'étude du score de propension pour pouvoir apparier deux populations différentes, mais possédant des caractéristiques observables similaires. Beaucoup de méthodes se sont développées depuis.

Quelques éléments théoriques sur l'appariement statistique

FORMALISATION DU PSEUDO-APPARIEMENT STATISTIQUE

Supposons deux enquêtes A et B indépendantes de taille respective $n(a)$ et $n(b)$ échantillonnées sur la même population et identiquement distribuées¹⁰². Soit (X, Y, Z) un vecteur de variables aléatoires de densité $f(x, y, z)$, avec $x \in X, y \in Y, z \in Z$. A ne dispose pas de l'information Z ni B de l'information Y .

L'union $A \cup B$ est donc un échantillon de la population étudiée de taille $n(a) + n(b)$, issue de $f(x, y, z)$ où Z est manquant dans A et Y dans B . Par conséquent, le problème d'appariement statistique se réfère à un problème d'imputation de valeurs manquantes partielles. L'objectif consiste alors à étudier la distribution jointe de (X, Y, Z) , en dépit du fait que nous n'observons pas cette information de façon conjointe, mais seulement à travers A ou B . Le problème se résume ainsi : il faut choisir une base de donneurs (ici A), une autre de receveurs (B) et estimer l'information sur Y , noté Y^* , à imputer dans la base de receveurs (*tableau 5*).

¹⁰¹ Ce chapitre s'appuie largement sur les travaux réalisés par Mathilde Niay en 2022, à l'époque agente du bureau, dans le cadre de sa formation continue et de son mémoire de master Ensaï. Elle avait alors mené à bien un premier pseudo-appariement : la revue de littérature, la sélection de variables, l'écriture de l'algorithme, son exécution, et les vérifications afférentes. Ce travail avait été salué par Éric Lesage et Stéphane Legleye lors de la soutenance du mémoire. Il a été actualisé et prolongé à la marge : le code a été simplifié et des vérifications supplémentaires ont été effectuées notamment au regard des quelques commentaires fait pendant la soutenance.

¹⁰² A se réfère dans notre cas à l'EMP 2019 et B à l'ENL 2013.

L'hypothèse principale de l'appariement statistique est l'hypothèse d'indépendance conditionnelle (CIA¹⁰³) qui traduit l'indépendance de Y et Z sachant X. Elle signifie que les relations entre Y et Z sont expliquées uniquement à travers le jeu de variables X. Cette hypothèse permet de rendre le modèle de l'union de A et B identifiable et donc estimable, car l'ensemble des informations est disponible dans les échantillons. En effet, sous cette hypothèse la structure de la fonction de densité $f(X, Y, Z)$ est la suivante (pour plus de détails voir D'Orazio et al. (2006), chapitre 2) : $f(x, y, z) = f[Y|X](y|x) * f[Z|X](z|x) * f[X](x)$.

Tableau 5 : structure de l'appariement

Échantillon A (donneurs)	Échantillon B (receveurs)	Base de données synthétiques (appariées)
X,Y	X,Z	X,Y*,Z

DIFFÉRENTES MÉTHODES DE PSEUDO-APPARIEMENT

Il existe des méthodes paramétriques et non paramétriques (tableau 6) pour mettre en œuvre sur données microéconomique l'appariement, que nous présentons rapidement (pour une présentation théorique voir D'Orazio et al. (2006), chapitre 3) :

Tableau 6 : synthèse des principales méthodes d'appariement

Type	Méthodes	Avantages	Inconvénients
Paramétriques	<i>Conditional mean matching.</i> <i>Drawn from the predicted distribution.</i>	Parcimonie du modèle.	Demande une spécification d'un modèle, valeurs imputées « artificielles ».
Non paramétriques	<i>Random hot deck.</i> <i>Distance hot deck.</i> <i>Rank hot deck.</i>	Pas de spécification d'un modèle, les valeurs imputées sont « observées », adaptées au cas complexes.	Demande un travail de sélection de variables communes.

- les méthodes paramétriques sont basées sur l'estimation des valeurs d'une variable à imputer Y en utilisant des régressions multiples. La méthode la plus utilisée est le *conditional mean matching*, soit l'imputation d'une variable Y issue de A dans B, en estimant des régressions sur A sachant X, appliquées sur l'échantillon B. La deuxième méthode, parmi les plus utilisées, est le *draw from the predicted distribution*, qui est similaire à celle décrite précédemment, la seule différence étant l'ajout d'un résidu (aléatoire) pour chaque valeur prédite par les régressions. Ces méthodes requièrent donc la spécification d'un modèle économétrique ;
- les méthodes non paramétriques sont des méthodes, qui, à la différence des paramétriques, viennent imputer une variable Y observée et non estimée. Le *random hot deck* (*hot deck* aléatoire stratifié) consiste en la création de strates homogènes dans la population des donneurs, répliquées dans celle des receveurs. Un donneur est par la suite choisi aléatoirement pour le receveur de la même strate et la distribution jointe estimée y est répliquée. Le *distance hot deck* est similaire, à la différence qu'une fonction de distance choisie est implémentée : le donneur ayant la distance la plus petite est choisie pour être apparié au receveur. Enfin, le *rank hot deck* permet d'exploiter une variable X ordinaire pour imputer des valeurs.

¹⁰³ *Conditional Independence Assumption.*

Considérations pratiques¹⁰⁴

TRAITEMENTS STATISTIQUES SUR LES BASES

En amont de la réalisation d'un appariement statistique, il est essentiel d'étudier et de préparer les données à appairer. En effet, les potentielles sources de disparités entre les bases sont diverses : elles peuvent concerner la collecte des données, la méthodologie de conception d'enquête, les informations manquantes ou mal interprétées, etc. Ainsi il est important de savoir si le champ des ménages interrogés est similaire, si les variables communes sont homogènes dans leur contenu statistique (distribution par exemple) et d'essayer de quantifier le degré de similarité ou de disparité entre ces variables. Différents types d'harmonisation ressortent de la littérature : ajustement des valeurs manquantes, harmonisation des modalités et définition des variables, harmonisation de la période de référence, etc.

En aval, il est important de réaliser certaines vérifications sur le jeu de données apparié. Comme un nouvel ensemble de données est obtenu à partir de deux échantillons, l'appariement ne préserve pas les valeurs individuelles des individus à l'identique de la base donneuse. Cependant, le but n'est pas tant de préserver les niveaux individuels de valeur que de préserver la distribution des données et les relations multivariées entre les variables d'intérêt (*Rubin (1986)*). Différentes méthodes de vérifications existent : comparaisons visuelles des distributions, comparaison des moyennes, examens de corrélations ou exhibition de tests statistiques.

LA VALIDITÉ DE L'HYPOTHÈSE D'INDÉPENDANCE CONDITIONNELLE

Dans la pratique, l'hypothèse d'indépendance conditionnelle n'est pas vérifiable statistiquement. Néanmoins, pour s'assurer de sa probable validité, il est possible d'utiliser un ensemble de variables auxiliaires ayant un très fort pouvoir prédictif sur les variables d'intérêts à imputer, ces variables auxiliaires faisant ainsi l'intermédiaire de la relation entre Y et Z. Dans tous les cas, et même dans celui d'un appariement non paramétrique, la base de données obtenue sera considérée comme synthétique ou fictive. Bien qu'étant issues de données réellement observées, les données imputées et celles déjà présentes dans l'échantillon receveur ne sont pas celles d'un unique individu dans la pratique.

Pour identifier ces variables jouant le rôle de prédicteurs, des méthodes classiques sont préconisées : la sélection automatique de variable via des régressions et des analyses factorielles type arbres de régressions.

Choix méthodologiques nécessaires à la réalisation d'un appariement statistique

CHOIX DE LA BASE RECEVEUSE

Pour la réalisation du pseudo-appariement, l'ENL 2013 est la base receveuse et l'EMP 2019 est la base donneuse.

D'après l'ouvrage théorique de *D'Orazio et al. (2006) (op. cit.)*, il peut y avoir discussion quant au choix de la base des receveurs et de la base des donneurs. En pratique, plus souvent, la base receveuse retenue est celle avec le plus d'observations comme par exemple l'appariement OMAR-Ines (*encadré 6*), et c'est également le cas de Prometheus dans sa précédente version.

Il s'agit en effet de prendre en compte la manière dont l'information à imputer est distribuée dans la population. Par exemple, dans le modèle OMAR-Ines, la variable d'intérêt est la dépense de santé. Or, ces dépenses sont réparties de façon très hétérogènes au sein de la population, peu d'individus ayant de très fortes, ou de très faibles, dépenses de santé. Si ces

¹⁰⁴ Cette partie se base notamment sur la méthodologie issue d'Eurostat : *Leulescu et Agafitei (2013)*.

individus très particuliers se retrouvent dans une mauvaise strate, cela va déformer la distribution de la variable à imputer, en ajoutant des valeurs aberrantes et détériorer la qualité de l'appariement. En choisissant d'imputer la plus petite base à la plus grande, il y a moins de risques que ces observations soient mises dans la mauvaise strate.

Dans Prometheus, une problématique similaire se pose, dans la mesure où le besoin consiste à imputer des dépenses de carburant, celles-ci étant également très disparates selon les ménages. Par le passé, même si l'ENL 2013 présentait un nombre de ménages plus élevé que l'ENTD 2008 (27 137 individus, contre 20 178 pour l'ENTD), elle avait été choisie comme base receveuse. Elle était jugée plus robuste, disposait d'un grand nombre de variables d'intérêts pour le modèle et de données plus précises concernant le revenu des ménages, ce qui est essentiel pour mener des analyses distributives.

Dans la nouvelle version de Prometheus, l'ENL 2013 a été conservée comme base receveuse. La base donneuse est l'EMP 2019. En effet, utiliser la même méthode que dans la version précédente du modèle permet une bonne comparabilité entre les deux versions et, par ailleurs, l'ENL 2013 reste la base de microsimulation originelle du modèle, celui-ci ayant été d'abord développé pour l'analyse redistributive des politiques énergétiques dans le domaine du logement. Ce choix oblige néanmoins à prendre certaines précautions et notamment à contraindre les classes de donneurs à un effectif minimal.

Encadré 6 : OMAR, un modèle avec pseudo-appariement

Le modèle de microsimulation avec l'outil de microsimulation pour l'analyse des restes à charges (OMAR) de la Drees permet d'analyser les « restes à charges » pour les soins des ménages et d'étudier l'efficacité de l'assurance maladie concernant la redistribution du système de soin. Dans une première version du modèle (Lardellier et al. (2011)), la méthode retenue est un appariement hot deck aléatoire stratifié pour reconstituer les dépenses de santé en associant une base répertoriant les dépenses médicales remboursées (EPAS) et une autre recensant l'affiliation à une complémentaire (SPS). Après avoir identifié les variables explicatives de la consommation de soin, un choix est effectué parmi celles-ci pour définir les strates d'imputation : pour un receveur appartenant à une classe donnée, un donneur est tiré (avec remise) parmi les donneurs appartenant à la même strate, son profil est appliqué au ménage receveur.

Plus récemment, OMAR a été apparié avec Ines afin de disposer d'une vision plus globale du système de redistribution (Fouquet (2021)). Dans ce travail, un appariement par hot deck aléatoire a été préféré à des méthodes comme les forêts aléatoires ou les arbres de régressions (qui ont néanmoins servis pour sélectionner les variables dans la première étape du hot deck). En outre l'option choisie a été de transmettre aux individus d'Ines les dépenses de santé et complémentaire des individus d'OMAR.

CHOIX DE LA MÉTHODE D'APPARIEMENT

Dans notre cas, la méthode retenue est celle du *hot deck* aléatoire stratifiée (*Random Hot deck*). Cette méthode est la plus classique (de Waal (2015)) et était déjà retenue pour la précédente version de Prometheus, ce qui permet de travailler à méthode constante et d'offrir une meilleure comparabilité entre les versions du modèle¹⁰⁵.

¹⁰⁵ D'autres méthodes comme les scores de propensions (Rosenbaum et Rubin (1983)) ou des méthodes d'apprentissage type forêts aléatoires pourraient aussi être utilisées ; cependant le nombre de variables communes existant entre les deux enquêtes ne semble pas suffisant pour les mettre en œuvre et certaines de ces méthodes peuvent par ailleurs apparaître

Conformément à la méthode retenue du *hot deck* aléatoire stratifié, les individus des deux enquêtes sont classés au préalable dans des groupes d'individus homogènes que l'on appelle des strates. Dans ces strates, une sélection de variables catégorielles communes (voir ci-dessous) aux deux échantillons *A* et *B* ont la même valeur pour tous les individus. Ensuite, pour chaque individu de *A*, est sélectionné aléatoirement un individu donneur de *B* du même groupe, dont la valeur d'intérêt observée *Z* est imputée au receveur *A*. Ainsi, cette procédure équivaut à reproduire la distribution conditionnelle de *Z* selon le jeu de variables catégorielles communes *X*. Concrètement, les valeurs des variables d'intérêt appartenant à un individu de la base EMP 2019 sont imputées à un individu de la base ENL 2013. Nous parlons ici de pseudo-appariement et non d'appariement, car les échantillons sont disjoints, même si la population reflétée est la même (ménages de France métropolitaine).

B. Mise en œuvre pratique du pseudo-appariement

Le but de l'appariement est d'imputer dans l'ENL 2013 les kilométrages annuels et le nombre de véhicules, afin de pouvoir calculer les consommations et factures de carburants des ménages, à partir de l'EMP 2019. Pour cela, il s'agit de rapprocher des ménages similaires du point de vue des kilométrages annuels parcourus en voiture.

Présentation globale

La procédure d'appariement statistique par *hot deck* aléatoire stratifiée nécessite d'abord de créer des strates de profil similaires. Dans ce contexte, une strate représente un ensemble de ménages dont les variables de stratification ont toutes la même modalité. Ces strates vont permettre, pour chaque ménage de l'ENL 2013 (table receveuse) d'identifier un profil de ménage similaire à celui de l'EMP 2019 (table donneuse) du point de vue de ses autres caractéristiques observables.

De fait, la procédure d'appariement statistique par *hot deck* aléatoire stratifiée comprend deux étapes. Dans un premier temps les populations de ménages des deux enquêtes sont stratifiées en utilisant les caractéristiques communes explicatives de la principale variable d'intérêt à imputer. Dans un second temps, un appariement *hot deck* est réalisé au sein de chaque strate.

Sélection des variables de stratification

Dans la phase de stratification, des variables communes aux deux tables sont sélectionnées ou construites, en lien avec la modélisation économétrique de la variable de kilométrage annuel. Puis, les ménages de la base donneuse (EMP 2019) et de la base receveuse (ENL 2013) qui présentent les mêmes modalités pour toutes les variables de stratification sélectionnées se voient affecter le même numéro de strates. Ensuite, pour les ménages auxquels aucune strate n'a été affectée, c'est-à-dire lorsque la strate à laquelle ils auraient dû être affecté comporte moins de cinq individus, on itère la création de strates en supprimant une variable de stratification de l'ensemble considéré au départ, en commençant par la variable jugée la moins prépondérante dans l'explication de la variable d'intérêt à savoir le kilométrage annuel.

comme « boîte noire ». De plus, dans le cas de l'appariement OMAR-Ines, la méthode des forêts avait été comparée avec le *Random Hot Deck* et le gain résultant était apparu marginal (Fouquet (2021)).

MÉTHODE DE SÉLECTION DES VARIABLES ET HOMOGENÉISATION DES MODALITÉS DES VARIABLES DE STRATIFICATION

Les variables choisies pour former les strates – nombre de véhicules, âge, sexe et situation face à l’emploi de la personne de référence, catégorie de commune de résidence, dixième de niveau de vie, nombre de personne du ménage – ont été sélectionnées en respectant les étapes préconisées par la littérature.

Afin de réaliser une première analyse de la base, des arbres de décision *Classification and Regression Tree* (CART) ont été utilisés. Ils ont permis de faire émerger de l’EMP 2019 un premier regroupement de variables ayant un fort pouvoir prédictif du kilométrage annuel (nombre de véhicule, type de ménage, etc., voir annexe 8 pour plus de détail).

Ensuite, les variables issues des arbres de régressions ont fait l’objet de regroupements sur lesquels des régressions et des procédures de sélection automatiques de variable en fonction des AIC et BIC dites *stepwise* ont été ensuite lancées, permettant de déterminer les variables à sélectionner et les modalités à définir pour ces variables. Autrement dit, la procédure *stepwise* a été exécutée en faisant varier les variables de stratification envisagées et leurs modalités catégorielles, afin de sélectionner non seulement les variables de stratification, mais aussi leurs modalités (annexe 9 pour les résultats de la régression retenue, notamment contenant la catégorie âge, le dixième de niveau de vie, etc.).

Plus précisément, plusieurs versions des modalités des variables qualitatives ont été testées. En effet, ces variables communes destinées à permettre la stratification des ménages de la base doivent être catégorielles et identiques dans les deux bases de données. Aussi, doit-on s’assurer, pour toutes les variables communes qui servent à la stratification, que les modalités sont identiques et, pour toutes celles qui font l’objet d’un regroupement (tel que l’âge de la personne de référence), que le regroupement codé est le plus pertinent possible. L’arbitrage entre ces différentes versions a été rendu selon les critères suivants :

- lorsqu’une modalité n’était pas significative dans la régression, sa fusion avec une autre modalité plus significative lui était préférée : par exemple, à la classe d’âge des « 26 - 40 ans » non significative, une classe comprenant les « moins de 50 ans » a été préférée ;
- un principe de cohérence de la définition des variables vis-à-vis des travaux de Prometheus a été appliqué : par exemple les modalités de la catégorie de la commune de résidence ont été réalisées selon des précédents travaux conduits avec Prometheus ;
- l’effectif pondéré entre les deux bases par modalité devait être proche et les classes construites non disproportionnées.

L’annexe 10 présente pour chaque variable les distributions des effectifs pondérés en fonction des modalités finalement choisies.

HIÉRARCHISATION DES VARIABLES D’APPARIEMENT

Les variables de stratification ont été ensuite hiérarchisées selon la significativité de chacune par rapport à la variable d’intérêt, pour convenir à l’hypothèse d’indépendance conditionnelle. Cette hiérarchie est en effet importante, car dans le processus de *hot deck*, la variable la moins significative sera supprimée à chaque tour. À partir de la régression *stepwise* précédente obtenue pour le choix des variables, il est possible d’ordonner les variables en fonction de la significativité de leurs modalités. Ce procédé est largement utilisable dans ce cas présent, étant donné que le nombre de variables que l’on doit imputer est faible et que

celles-ci contiennent une information similaire (kilométrage annuel, nombre de véhicules et nombre de véhicules distingué selon la motorisation)¹⁰⁶.

Dans le cas présent, la significativité n'est pas extrêmement discriminante, sauf pour la variable du nombre de véhicules, étant clairement la plus significative d'un point de vue de l'explication du kilométrage annuel. Il a ainsi été décidé d'ordonner les variables selon un arbitrage entre la significativité, l'utilisation que l'on souhaite faire du modèle et le pseudo-appariement de la version précédente de Prometheus. En effet, d'après le niveau des significativités entre des variables, le choix d'un ordre différent ne modifierait qu'à la marge les résultats de l'appariement. Les variables retenues et leur ordre sont les suivants :

Tableau 7 : hiérarchie des variables d'appariement

Hiérarchie	Variables
1	Le nombre de véhicules
2	L'âge de la personne de référence
3	La catégorie de la commune
4	Le dixième de niveau de vie
5	Le sexe
6	La situation face à l'emploi
7	Le nombre de personne dans le ménage

Ces sept variables ont été retenues en considérant qu'elles suffiraient à appairier l'ensemble des ménages. Empiriquement, seules six itérations ont été nécessaires. De plus, tous les ménages ont été pseudo-appariés avec au moins trois variables communes.

MISE EN ŒUVRE ALGORITHMIQUE DE LA STRATIFICATION

Les strates d'imputation ont été créées séquentiellement via un algorithme. Dans un premier temps, toutes les strates pouvant être créées avec les sept variables de stratification selon la condition d'un effectif minimal de cinq ménages dans chaque strate de donneurs ont été formées. Cette condition a été posée afin de ne pas entraîner une surreprésentation dans la base appariée d'un individu spécifique de l'EMP 2019 : il est ainsi vérifié qu'il y a au moins cinq ménages de la base donneuse (EMP 2019) présents dans chaque strate. Aux deux tables réduites, sont ainsi ajoutée une variable « strate » avec son numéro de strate associée. Les observations n'ayant pu être classées dans des strates (parce qu'inférieures à cinq donneurs) sont stockées dans deux vecteurs, l'un pour la table transport, l'autre pour la table logement.

L'opération de création de strate est ensuite répétée sur les deux tables contenant les observations n'ayant pas de strate, en retirant à chaque fois la dernière variable de stratification. Cette étape est répétée itérativement, jusqu'à ce que toutes les observations soient rangées dans des strates d'imputations¹⁰⁷. Dans notre cas, six itérations ont été nécessaires pour un total de 15 632 strates créées. À chaque itération, il a été possible de vérifier que la procédure se déroule correctement en comparant le nombre de strates créées avec un calcul simple du nombre théorique de strates que chaque étape est censée produire.

¹⁰⁶ En effet, dans le cas où la quantité d'information à imputer est grande, cette technique est alors impossible à utiliser (une régression ne suffirait pas à expliquer la quantité disparate d'information à imputer). Dit autrement, comme les variables à imputer sont corrélées, il y a peu de risque d'avoir des hiérarchies différentes entre les variables explicatives.

¹⁰⁷ Initialement le package « Stat Match » disponible en open source (D'Orazio (2017)) avait été utilisé pour réaliser ce nouvel pseudo-appariement. Cependant, il ne permettait d'utiliser au maximum que quatre variables, car par la suite il se bloquait avec l'ajout de variables supplémentaires en raison de strates vides, puisqu'il n'y a pas d'option permettant de mettre une condition sur le nombre de donneur ou ignorer les classes pour lesquelles l'effectif des donneurs est nul. Un code spécifique a donc été créé pour lever cet obstacle.

Mise en œuvre du *hot deck* aléatoire au sein de chaque strate

Une fois les tables stratifiées, un appariement par *hot deck* aléatoire a été réalisé au sein de chaque strate. Deux versions de tirage ont été réalisées, l'une sans prendre en compte le poids des donneurs, l'autre avec pondération :

- le hot-deck aléatoire sans pondération a été construit de la manière suivante : pour chaque ménage de l'ENL 2013, un ménage de l'EMP 2019 au sein de la même strate est attribué selon un tirage aléatoire uniforme avec remise ;
- pour la version avec pondération, le tirage avec remise a été réalisé non plus de façon uniforme, mais en le pondérant par le poids relatif des individus donneurs. Par exemple, une personne ayant deux fois plus de poids dans la table transport a deux fois plus de chance d'être tirée comme donneur. Pour cela, un poids individuel par strate a été créé en rapportant la pondération individuelle de chaque ménage sur la somme de l'ensemble de la pondération des ménages de la strate.

Au total, pour chaque ménage de l'ENL 2013 un ménage de l'EMP 2019 a été tiré.

Jointure des tables

La dernière étape du pseudo-appariement consiste à joindre les variables d'intérêts de la table transport à la table logement. Grâce à l'identifiant des ménages de la table transport imputé dans l'étape précédente, il est possible de faire une jointure entre les deux tables et de conserver uniquement les variables d'intérêt issue de l'EMP 2019 : kilométrage annuel total et par carburant de transport des ménages, nombre de véhicules total (décomposé par type de carburant, essence, diesel etc.) et quelques variables supplémentaires (distance aux transports en commun, etc.).

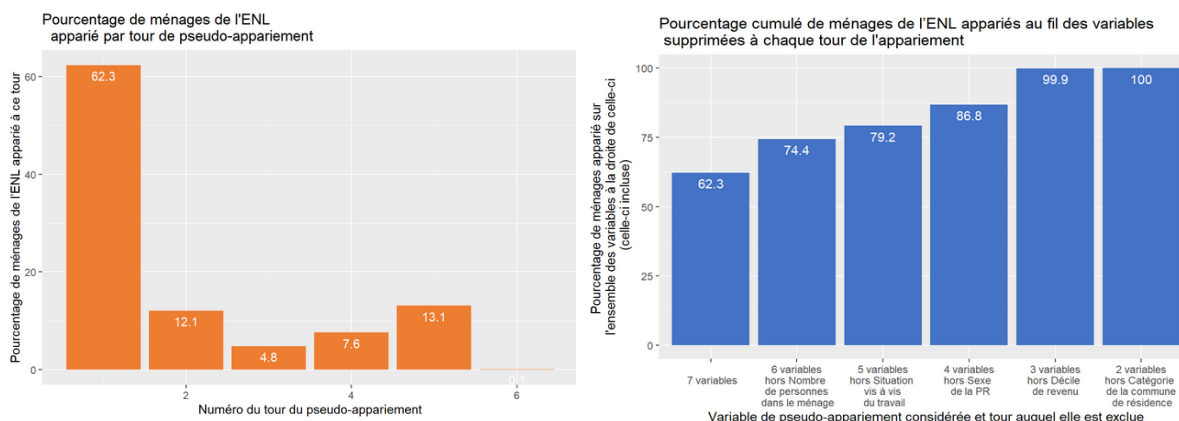
C. Analyse de la qualité de l'appariement

À la fin du *hot deck* aléatoire, tous les ménages de l'ENL 2013 ont été appariés avec 75 % de ménages distincts de l'EMP 2019. En effet, un même donneur a pu être tiré plusieurs fois, puisque le *hot deck* a été développé dans la version de référence selon un tirage aléatoire pondéré avec remise dans les différentes strates constituant la population des donneurs. Ce nombre de donneurs distincts utilisé a son importance : il s'agit de ne pas utiliser un même donneur à de trop nombreuses reprises, afin d'une part de ne pas déformer la distribution, et d'autre part de limiter le risque de choisir un donneur « dans la mauvaise case », qui pourrait conduire à une imputation de données aberrantes.

Statistiques sur les strates d'imputation

Un premier tour de l'algorithme d'appariement par *hot deck* aléatoire avec les sept variables retenues a permis l'appariement de 62,3 % de la population des receveurs de l'ENL 2013 avec un donneur de l'EMP 2019 (*figure 3* à gauche). Lorsqu'il a été retiré une variable d'appariement (nombre de personnes dans le ménage), 12,1 % supplémentaires de la population des receveurs ont pu être appariés. Dit autrement, avec six variables d'appariement, 74,4 % de la population de l'ENL 2013 est appariée avec un donneur de l'EMP 2019 (*figure 3* à droite), ce qui est très satisfaisant. En moyenne, les ménages appariés l'ont été 2,6 fois (l'écart-type est de 2,0). Au total, 35,1 % des ménages de l'EMP 2019 qui ont été appariés l'ont été une seule fois, 51,9 % l'ont été entre deux et quatre fois et 13 % cinq fois ou plus. À cette étape, les identifiants des ménages de la table transport tirés ont été ajoutés à la table logement.

Figure 3 : statistiques sur l'appariement des données



Source : CGDD, modèle Prometheus

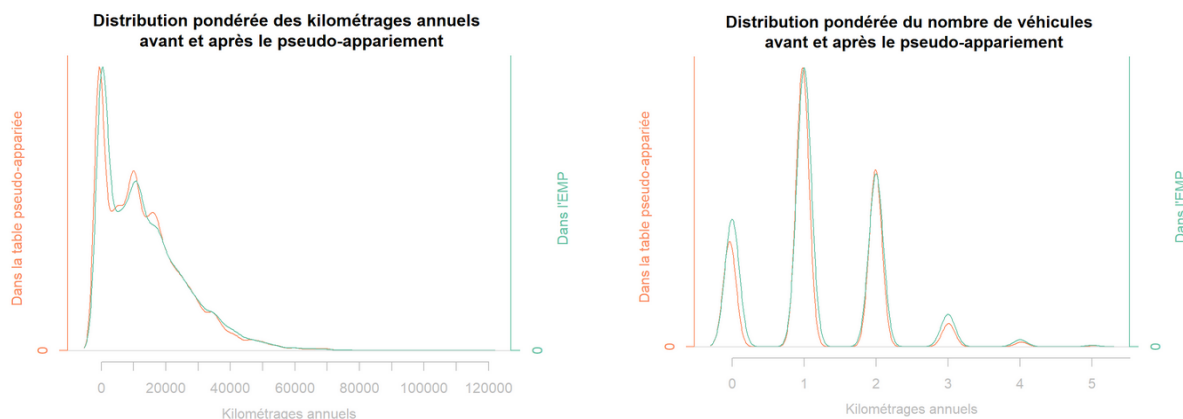
Contrôles sur les distributions des variables d'intérêt imputées

Il s'agit ensuite de vérifier que les variables d'intérêts imputées dans l'ENL 2013 gardent la même distribution que dans l'EMP 2019. Un des avantages de l'utilisation d'une méthode *hot deck* est qu'elle permet, en théorie, de préserver cette distribution. Il s'agit dans ce cas de s'en assurer afin de vérifier que l'appariement s'est bien déroulé.

VÉRIFICATIONS VISUELLES

Les graphiques de la *figure 4* représentent la distribution des variables d'intérêts à imputer permettant in fine le calcul des consommations et factures, à savoir le kilométrage annuel (*figure 4* à gauche) et le nombre de véhicules (*figure 4* à droite ; d'autres variables secondaires utiles pour nos analyses comme la possession d'une carte de transport, la distance au transport etc., mais qui restent secondaires, sont également imputées par la suite, voir ci-dessous). Les courbes de densité de probabilité des variables sont représentées avant et après appariement, soit les distributions des variables dans la base initiale issue de l'enquête mobilité des personnes (EMP 2019 avec ses pondérations) et de ces mêmes variables dans la base pseudo-appariée (en utilisant les pondérations issues de l'ENL 2013). Visuellement, les résultats semblent très concluants, les distributions représentées étant très proches.

Figure 4 : graphiques de vérification des distributions



Source : CGDD, modèle Prometheus

Source : CGDD, modèle Prometheus

VÉRIFICATIONS NUMÉRIQUES

Afin de conforter ces résultats visuels, l'effectif par dixième de niveau de vie a été calculé pour chaque variable d'intérêt imputée. En effet, Prometheus visant à réaliser des analyses distributives, il est nécessaire que les résultats par dixième soient les plus robustes possibles.

Le *tableau 8* ci-dessous présente la moyenne du nombre de véhicules selon le dixième de revenu des ménages avant et après appariement. La moyenne reste quasi identique pour l'ensemble des dixièmes des ménages, a fortiori arrondie au nombre entier de véhicule (par exemple, pour le D1 l'arrondi de 0,68 ou 0,76 est 1 véhicule).

Tableau 8 : statistiques distributives sur le nombre de véhicules possédés par les ménages dans l'EMP 2019 et à l'issue du pseudo-appariement

Variables de comparaisons	EMP	Table pseudo-appariée	Écart-relatif (en %)
Somme (en millions)	35,74	33,53	6
Moyenne	1,22	1,23	- 1
Q1	0,88	0,85	3
Médiane	1,29	1,15	11
Q3	1,53	1,36	11
D1	0,68	0,76	- 12
D2	0,87	0,95	- 9
D3	1,02	1,07	- 5
D4	1,18	1,14	3
D5	1,24	1,24	0
D6	1,36	1,35	1
D7	1,48	1,36	9
D8	1,43	1,43	- 1
D9	1,54	1,49	3
D10	1,44	1,54	- 7

Source : CGDD, modèle Prometheus

L'*annexe 11* présente un tableau sur les kilométrages moyens ventilés par dixième. Les résultats sont similaires.

STABILITÉ ET ROBUSTESSE DES RÉSULTATS

Afin de tester la stabilité des résultats, 50 pseudo-appariements différents ont été simulés, selon le même principe qu'une technique de *bootstrap* : la réplication multiple d'appariements permet de calculer des mesures de dispersions pour les variables de mobilité imputées. Le *tableau 9* suivant décrit ces indicateurs pour les variables de kilométrage et nombre de véhicules. Les résultats montrent que malgré le caractère aléatoire de l'appariement, le kilométrage annuel total et le nombre de véhicules total (et différencié par motorisation) sont stables.

Tableau 9 : indicateurs de dispersion calculés sur 50 pseudo-appariements

Variabiles	Moyenne	Écart-type	Coefficient de variation*
Kilométrage annuel	12 430,14	129,066	1,04
Proportion de ménages appariées au quatrième tour de hot-deck	93,89	5,987	6,38
Proportion de donneurs distincts	75,44	0,227	0,30
Nombre de véhicules	1,11	0,004	0,36
Nombre de véhicules diesel	0,41	0,006	1,46
Nombre de véhicules essence	0,68	0,009	1,32
Nombre de véhicules GPL et autres	0,02	0,001	5,00

*Coefficient de variation : écart-type/moyenne * 1000

Source : CGDD, modèle Prometheus

Préservation des relations de dépendances entre les variables

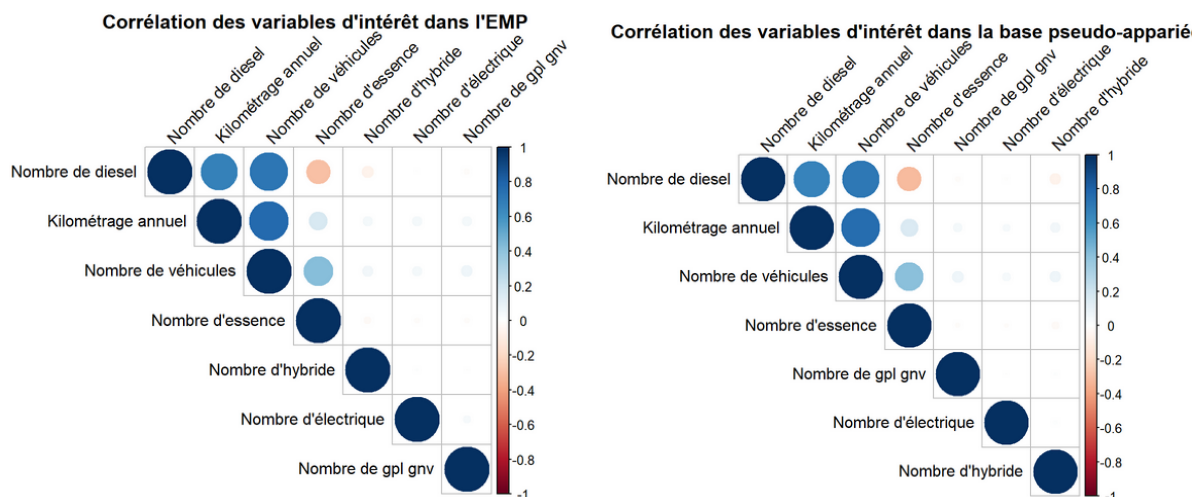
ENTRE LES VARIABLES D'INTÉRÊT IMPUTÉES

Le jeu de données appariées doit également préserver les corrélations (positives ou négatives) présentes dans le jeu de données des donneurs. On s'intéresse en premier lieu aux relations entre les variables d'intérêt que l'on a cherché à imputer, soit le kilométrage et le nombre de véhicules (distingué également selon la motorisation). Ces variables étant quantitatives, l'étude des relations entre ces variables avant et après appariement peut se faire à partir des matrices de corrélations (*figure 5*) :

- la première matrice de corrélations, calculée à partir de la table transport (*figure 5, gauche*), montre une relation de corrélation positive entre le kilométrage annuel et le nombre total de véhicules, le kilométrage annuel et le nombre de véhicules diesel (les détenteurs de véhicules diesel parcourent en effet de plus grandes distances que la moyenne), le kilométrage annuel et le nombre de véhicules essence (bien que moins prononcée). Elle montre également une relation négative entre le nombre de véhicules essence et le nombre de véhicules diesel ainsi qu'entre le nombre de véhicules essence et le nombre de véhicules GPL ou autre ;
- la seconde matrice de corrélations (*figure 5, droite*) a été calculée sur la base finale appariée. Les relations décrites précédemment sont conservées. La procédure d'appariement par *hot deck* aléatoire stratifié ne semble donc pas avoir distordu les relations entre les variables d'intérêts¹⁰⁸.

¹⁰⁸ Les valeurs des corrélations sont fournies en *annexe 12*.

Figure 5 : lien entre les variables d'intérêt présentes dans l'EMP 2019



Note : plus le cercle est gros et bleu, plus le lien entre les variables en colonne et en ligne est fort.

Source : CGDD, modèle Prometheus

ENTRE LE KILOMÉTRAGE ANNUEL LES VARIABLES AUXILIAIRES DE STRATIFICATION

Dans un deuxième temps, on vérifie les relations entre la principale variable d'intérêt imputée (le kilométrage annuel) et les variables auxiliaires, celles ayant servies à la formation des strates d'imputations. Les relations entre ces variables auxiliaires et la principale variable d'intérêt imputée doivent avoir été conservées pour que l'appariement soit considéré comme de bonne qualité.

Pour toutes ces variables, un test du khi-deux (qui vérifie le lien statistique entre deux variables) a été réalisé et l'hypothèse du test H0 d'indépendance des variables a été rejetée au seuil de 5 %, avant et après l'imputation. Par conséquent, les variables semblent avoir préservé leur relation de dépendance. Afin de quantifier cette relation, le V de Cramer (mesure de la taille d'effet pour le test d'indépendance du khi-carré, il mesure le degré d'association de deux champs catégoriels) a été calculé pour chacune de ces variables et confirme la similarité des relations avant et après pseudo-appariement (tableau 10)¹⁰⁹.

¹⁰⁹ Les tests réalisés ici ne tiennent pas compte des pondérations. La fonction *svychisq* du package « survey » ne tourne pas en raison du trop grand nombre de valeurs prises par la variable de kilométrage ; il n'a pas été trouvé de fonction permettant de calculer le V de Cramer en tenant compte des pondérations différentes selon le plan de sondage.

Tableau 10 : V de Cramer pour les variables auxiliaires utilisées pour le pseudo-appariement dans la base initiale et dans la base pseudo-appariée

Variables	Base initiale EMP 2019	Base pseudo-appariée	Différence entre les V de Cramer
Âge	0,23	0,20	0,03
Dixième de niveau de vie	0,18	0,18	- 0,00
Catégorie de la commune de résidence	0,12	0,10	0,20
Nombre de véhicules	0,67	0,69	0,01
Nombre de personne du ménage	0,31	0,23	0,08
Sexe de la PR	0,25	0,21	0,04
Situation vis-à-vis du travail	0,38	0,31	0,06

Source : CGDD, modèle Prometheus

VI. VIEILLISSEMENT ANNUEL

La base de données des ménages issue du pseudo-appariement entre l'ENL 2013 et l'EMP 2019 est actualisée chaque année à l'aide de données agrégées afin de refléter l'évolution des comportements des ménages vis-à-vis des énergies (caractéristiques des logements et des véhicules, consommations unitaires, etc.). Les données agrégées utilisées sont relatives au logement, au transport et aux revenus des ménages. Au regard du chapitre précédent, les variables servant au pseudo-appariement étant des variables dont les valeurs n'évoluent quasiment pas au cours du temps et par souci d'ergonomie du modèle, il est plus optimal de pseudo-apparier les enquêtes en une seule base, puis de faire vieillir cette base en une seule fois, plutôt que de faire vieillir indépendamment chaque base et de relancer un pseudo-appariement à chaque fois.

A. Logement

Afin de calculer les factures énergétiques des ménages en matière de logement, les volumes de consommations d'énergie du logement sont vieillis, ensuite, il suffit de compiler les prix et la fiscalité pour en déduire des factures :

- le vieillissement des consommations est conduit en deux étapes auxquelles est éventuellement ajoutée une étape de modélisation de l'effet de la météo. Dans un premier temps, le parc de logement¹¹⁰ et de chauffage est actualisé en faisant évoluer la répartition des logements selon l'énergie de chauffage, puis en recalant les données individuelles en fonction du nombre de logements individuels et collectifs ;
- dans un second temps, la consommation énergétique unitaire de chaque logement est actualisée grâce à la consommation énergétique agrégée ;
- par ailleurs les données sur la météo sont actualisées afin de prendre éventuellement en compte une correction des variables climatiques dans le modèle qui neutralise les effets de la consommation liés à un hiver plus ou moins rigoureux par rapport à la tendance.

Vieillissement du parc de logement

Pour vieillir le parc de logement, les pondérations de l'ENL 2013 sont modifiées en tenant compte des évolutions des types de logement (maison ou appartement), des modes de chauffage (chauffage central individuel ou chauffage collectif) et des énergies principales de chauffage (électricité, gaz, fioul, etc.). Les données utilisées sont les suivantes :

- le « recensement de la population » (RP) de l'Insee, pour faire évoluer la répartition des logements par type de logement (maison ou appartement), type de chauffage (individuel ou collectif) et énergie principale de chauffage. Ces données sont disponibles en année N-3 (par exemple, pour l'actualisation des données 2022, la dernière année disponible est 2019) : pour les années postérieures manquantes, une moyenne glissante des trois derniers taux de croissance est appliquée ;
- les « Comptes satellites du logement » (CSL) du SDES, pour faire évoluer le nombre de logements individuels et collectifs.

¹¹⁰ Le « parc de logement » désigne le nombre de logements collectifs et individuels ainsi que la répartition par type d'habitation et de chauffage, et d'énergie principale de chauffage. On ne fait pas évoluer la typologie des logements (nombre de pièces, surface etc.), car la variable d'intérêt est la consommation annuelle totale des ménages.

ÉVOLUTION DES PARTS DES ÉNERGIES DE CHAUFFAGE DANS LES LOGEMENTS

Les données du RP de l'Insee¹¹¹ permettent d'obtenir la structure croisée des ménages selon le type de logement, le mode de chauffage et l'énergie principale de chauffage. Les pondérations des ménages dans la base pseudo-appariée du modèle sont alors modifiées de façon à répliquer la structure croisée issue du RP de l'Insee.

Plus précisément, pour une catégorie « type de logement * mode de chauffage* énergie principale de chauffage » donnée (*annexe 13*)¹¹², les ménages concernés voient leurs pondérations respectives évoluer de la même façon que le poids de cette catégorie dans le RP (par exemple, si le poids de cette catégorie évolue de 5 % dans le RP, les pondérations des ménages concernés sont augmentées de 5 %). Les autres caractéristiques des ménages (structure par âge, par sexe, par catégorie socio-professionnelle, etc.) ou de leur logement (date de construction, surface, etc.) sont supposées inchangées.

ÉVOLUTION ET CALAGE DU NOMBRE DE LOGEMENTS INDIVIDUELS ET COLLECTIFS

Pour actualiser le nombre total de logement, et le détail entre logement individuels et collectifs, les données utilisées sont celles du SDES¹¹³ disponibles assez rapidement à l'occasion de la publication des Comptes satellites du logement (CSL) : l'agrégation des données sur les résidences principales en France métropolitaine permet d'obtenir les nombres de logements individuels, collectifs et totaux qui sont intégrés au modèle.

Après le traitement explicité précédemment pour l'actualisation de la structure du parc de logement, les nombres de logements individuels et collectifs sont calés sur leur totaux issus des CSL, tout en conservant cette structure. Ce calage est obtenu en rapportant le nombre respectif en logements individuels et collectifs cible à celui de la base : les pondérations sont alors multipliées par ce taux pour correspondre aux nombres cibles de logements.

Vieillessement des consommations énergétiques unitaires des logements

DONNÉES AGRÉGÉES UTILISÉES

Le vieillissement des consommations énergétiques unitaires des logements est effectuée à partir du « Bilan énergétique de la France » publié par le SDES en fin d'année¹¹⁴, et qui porte sur l'année N-1¹¹⁵. Ce bilan énergétique fournit en effet une description de la consommation d'énergie en France, par type d'énergie et par secteur, notamment dans le résidentiel, avec également un détail du nombre de logements par type d'énergie. Une mise en cohérence préalable des nomenclatures d'énergies est toutefois nécessaire, entre Prometheus et le bilan énergétique de la France du SDES. Ce travail est toutefois rendu léger, du fait de la forte proximité entre les deux nomenclatures (*annexe 14*).

VIIEILLISSEMENT DES CONSOMMATIONS DES MÉNAGES

Pour chaque énergie, on divise la consommation totale résidentielle issue des données du SDES par le nombre de logements où l'énergie de chauffage principale est la même. Cela permet de calculer les consommations unitaires pour chaque énergie. En revanche, pour la

¹¹¹ Plus précisément les fichiers « Logements » : www.insee.fr/fr/statistiques/4171558.

¹¹² Seules les parts supérieures à 1 % sont actualisées dans le modèle. Les parts non renseignées dans le modèle sont supposées constantes. Ce choix a été fait pour éviter qu'il y ait trop de volatilité, certains croisements gardent donc le même poids qu'en 2013.

¹¹³ Disponibles en ligne : www.statistiques.developpement-durable.gouv.fr/rapport-du-compte-du-logement-2021-0.

¹¹⁴ Jusqu'au données sur 2020, les données récupérées provenaient du Ceren, lesquelles servaient au SDES pour la construction de son « Bilan énergétique de la France » (résultats définitifs et détaillés). Depuis l'actualisation de l'année 2021, les données sont fournies directement par le SDES et cohérentes avec les séries longues du « Bilan énergétique de la France ».

¹¹⁵ Fin 2023, voir début 2024, l'année 2022 est disponible.

consommation d'électricité spécifique, on divise la consommation par le nombre total de logement.

Les consommations unitaires ainsi obtenues sont donc disponibles pour 2013 et pour toutes les années postérieures. Le vieillissement de la consommation d'énergie des ménages, dans Prometheus, se fait alors en faisant évoluer, pour chaque ménage, la consommation des différentes énergies par leurs consommations unitaires (voir annexe 14 pour le traitement de l'électricité spécifique).

Ainsi, ce vieillissement correspond à des hypothèses simplificatrices au niveau ménage : ceux-ci ne changent pas d'énergie, ne rénovent pas leur logement et ont un comportement d'évolution de consommation qui correspond à la moyenne. Néanmoins le modèle reflète les évolutions de consommation unitaire et les changements de part des énergies dans les logements, ce qui permet en moyenne de faire des analyses pertinentes faute d'actualisation plus fréquentes des données d'enquêtes et dans le cadre d'un modèle statique (même pseudo-dynamique). Ces problèmes ne se poseraient dans un cadre dynamique.

Correction des conditions météorologiques

Le modèle Prometheus permet aussi la prise en compte des conditions météorologiques d'une année donnée, en simulant si besoin des consommations énergétiques corrigées des variations climatiques (CVC) et donc des factures qui le sont également (les données utilisées dans le Bilan du SDES sont prises en brutes). Cela permet d'obtenir une autre chronique sur les consommations et les factures énergétiques en neutralisant l'effet de la température en hiver. Par exemple, dans le cadre de l'estimation annuelle qui est faite par Prometheus de la précarité énergétique des ménages, il est nécessaire de comparer l'évolution du taux de précarité CVC par rapport au taux brut. Un recul de ce dernier peut être lié à un hiver plus chaud et non pas à une amélioration des conditions de logement des précaires.

CONCEPTS UTILISÉS

La correction des conditions météorologiques vise à neutraliser l'effet de la température en hiver afin d'aller plus loin qu'une analyse de 1^{er} rang sur une hausse de la consommation d'énergie si l'hiver a été très rigoureux par exemple. Il repose sur l'indice de rigueur (IR) publié chaque année par le SDES¹¹⁶. Il se fonde lui-même sur le concept de degrés-jours unifiés (DJU)¹¹⁷ : pour chaque jour de l'année, une valeur T est calculée comme la moyenne des extrêmes de températures observées en France métropolitaine en degrés Celsius, comparée à une température de référence de 17 ° (fixée par le SDES). Le nombre de DJU d'une journée vaut 17-T si T est inférieur à 17, et 0 sinon, et le DJU annuel est la somme de ces valeurs journalières sur la période de chauffe qui va de janvier à mai et d'octobre à décembre. Cette valeur est alors comparée à la moyenne des degrés-jours sur la période de référence 1986-2015, dite DJU0. Pour une année donnée : $IR = DJU/DJU0$. Lorsque $IR > 1$, l'année est plus froide que la moyenne.

TRAITEMENT DANS LE MODÈLE

En janvier de chaque année, le SDES publie pour chaque énergie un coefficient relatif aux ménages (p) traduisant à quel point la consommation de cette énergie par les ménages est sensible aux variations météorologiques.

¹¹⁶ www.statistiques.developpement-durable.gouv.fr/indice-de-rigueur-degres-jours-unifies-aux-niveaux-national-regional-et-departemental.

¹¹⁷ www.statistiques.developpement-durable.gouv.fr/sites/default/files/2020-01/methodologie-degres-jours-unifies_0.pdf.

Pour chaque ménage et après vieillissement de ses consommations énergétiques, sa consommation brute (C, c'est-à-dire non corrigée des conditions météorologiques) est ajustée en consommation CVC à l'aide de la formule suivante¹¹⁸ :

$$CVC = \frac{C}{1 + p \times (IR - 1)}$$

Ainsi, pour une année plus froide que la moyenne, les ménages consomment plus de chauffage et la consommation observée est donc considérée plus élevée qu'une année moyenne, toutes choses égales par ailleurs. La consommation CVC est donc inférieure à sa valeur observée. À l'inverse, à une année plus chaude que la moyenne correspond une consommation CVC supérieure à sa valeur observée.

La notion d'indice de rigueur ne permet de tenir compte que de la rigueur des hivers et non pas des épisodes caniculaires des étés qui pourraient éventuellement pousser à la hausse les dépenses énergétiques pour se refroidir. Pour évaluer ce phénomène, une utilisation des DJU de rafraîchissement pourrait être développée¹¹⁹.

B. Transport

Afin d'actualiser les factures énergétiques des ménages côté transport, la mise à jour des données relatives au transport comporte plusieurs étapes.

La première étape consiste à vieillir le parc de véhicules, en faisant en sorte qu'au niveau agrégé, la répartition des kilométrages parcourus selon les types de carburant soit calée sur celle du bilan annuel des transports publié par le SDES. Ce calage conduit à ce que les véhicules de certains ménages soient transformés en véhicules diesel ou essence, c'est l'étape qu'on nomme « diésélisation » ou « essentialisation ». Ensuite, les kilométrages parcourus par les véhicules des ménages sont recalés de manière à ce que le kilométrage moyen réalisé par les ménages égalise celui publié en agrégé par le SDES. Enfin, les consommations unitaires des véhicules des ménages sont également recalées de la même manière afin d'égaliser en moyenne celles observées par le SDES au niveau agrégé.

Données sources

Chaque année, dans son « bilan annuel des transports »¹²⁰, le SDES publie de façon agrégée, par type de carburant (concernant l'électrique voir encadré 7), le kilométrage annuel moyen des véhicules, le nombre de véhicules sous pavillon français, et la consommation unitaire moyenne des véhicules. En parallèle, le SDES publie les « données sur le parc automobile français »¹²¹ dans lesquelles sont détaillés, là aussi de façon agrégée, le nombre de véhicules utilitaires légers (VUL) utilisés par des particuliers, par carburant, et avec leur kilométrage annuel moyen et leur consommation unitaire.

Ces deux sources de données permettent de calculer, pour une année donnée et par carburant, le kilométrage annuel moyen parcouru par les voitures et les véhicules utilitaires légers utilisés par les particuliers, ainsi que leur consommation unitaire moyenne. S'en déduisent, au sein de l'ensemble des kilomètres parcourus par les ménages, la proportion de kilomètres parcourus par des véhicules diesel. Cette proportion sert de cible à l'étape de diésélisation.

¹¹⁸ Dans le cas de l'électricité, du gaz et du fioul, seulement pour les ménages qui ont une de ces énergies en tant qu'énergie principale de chauffage. Par ailleurs toutes les consommations de gpl, bois et urbain sont aussi ajustées.

¹¹⁹ Voir par exemple une utilisation possible avec une valeur de référence de 25° C par l'ONPE ([tableau de bord 2022](#) p.17).

¹²⁰ www.statistiques.developpement-durable.gouv.fr/bilan-annuel-des-transports-en-2022.

¹²¹ www.statistiques.developpement-durable.gouv.fr/donnees-sur-le-parc-automobile-francais-au-1er-janvier-2023?rubrique=58&dossier=1348.

La diésélisation ou essencialisation

Comme présenté dans la partie précédente (VI A.), concernant logement, le vieillissement du parc est réalisé en ajustant les pondérations. La même méthode n'est pas utilisée dans le domaine du transport, car nous ne disposons pas de données croisant les aspects logement et transport qui permettraient de réaliser un tel vieillissement. Or, la structure du parc en termes de répartition des véhicules par motorisation évolue. En particulier les ménages ont été incités à s'équiper davantage en véhicule diesel puis en véhicule essence à partir du milieu des années 2010. Aussi, la part de véhicule diesel dans le parc a crû jusqu'à 2015 jusqu'à atteindre 64 %¹²² avant de diminuer au profit des véhicules essences.

Ainsi pour faire vieillir le parc de véhicules, on transforme directement la motorisation de certains véhicules. Ceux-ci sont sélectionnés en estimant la probabilité pour chaque ménage en fonction de leur type de motorisation, de changer de motorisation de véhicule.

CALCUL DE LA PROBABILITÉ DE TRANSFORMATION DE LA MOTORISATION DU VÉHICULE

Dans la base pseudo-appariée, il existe six types de motorisation de véhicules pour un ménage (tableau 11). Pour chaque type de motorisation, on régresse sur la base pseudo-appariée la probabilité de changer de motorisation, d'être dans la catégorie 0 ou dans la catégorie 1, où la différence entre chacune des catégories tient à la motorisation d'un des véhicules. De fait, la probabilité pour un ménage d'avoir un véhicule diesel « essentielisé » est définie comme le complémentaire de la probabilité d'avoir un véhicule essence « diésélisé ».

Concrètement, on réalise six régressions logistiques sur la base pseudo-appariée, une par type de composition de la motorisation des ménages (nombre de véhicules par type de carburant), et dans la limite de trois véhicules possédés au total par ménage (pour le détail des estimations voir annexe 15).

Tableau 11 : probabilité logistique modélisée

Numéro du cas	Catégorie	
	0	1
1	1 ES	1 GO
2	2 ES	1GO / 1 ES
3	1 ES / 1 GO	2 GO
4	3 ES	2 ES / 1 GO
5	2 ES / 1 GO	1 ES / 2 GO
6	1 ES / 2GO	3GO

Puis, pour chaque ménage, si la probabilité de transformer un de ses véhicules est supérieure à celle donnée par une loi uniforme allant de 0 à 1, alors on considère qu'il est « diésélisable ». En comparant la probabilité obtenue à celle d'une loi uniforme, on réintroduit de l'aléatoire dans le fait de réellement choisir de transformer le véhicule.

CHOIX DES VÉHICULES SUBSTITUÉS

Puis, on sélectionne aléatoirement les ménages qui seront réellement diésélisés parmi les diésélisables, de façon à ce que la part des kilomètres parcourus par des véhicules diesel dans les kilomètres parcourus par l'ensemble des véhicules des ménages, calculée à partir des données publiées annuellement par le SDES, soit atteinte. On ne remplace qu'un seul véhicule par ménage et si notre cible n'est pas atteinte on tire « au hasard » des véhicules supplémentaires.

¹²² D'après le « Bilan de la circulation 2021 » publié par le SDES.

Lorsqu'on « diésélise », le véhicule essence parcourant le plus grand nombre de kilomètres des ménages qui ont le plus de chance d'avoir un véhicule diesel est remplacé par un véhicule de ce carburant. Inversement, lorsqu'on « essentialise », on remplace le véhicule diesel parcourant le moins de kilomètres des ménages qui ont le plus de chance d'avoir un véhicule essence. Par exemple, pour un ménage donné, à la place de posséder un diesel âgé de 10 ans, on va lui imputer un véhicule essence âgé de 10 ans. Cela revient à faire l'hypothèse que le même ménage aurait acheté un véhicule essence, à la place du véhicule diesel il y a 10 ans.

AJUSTEMENT DES CARACTÉRISTIQUES DU VÉHICULE TRANSFORMÉ

Aussi, les ménages qui font l'objet d'une transition ont un de leur véhicule qui change de motorisation, son kilométrage annuel est conservé, de même que l'ensemble de ces caractéristiques, exceptée sa consommation unitaire. En effet, à tout véhicule diésélisé on associe la consommation unitaire du véhicule diesel de l'EMP 2019 le plus proche en âge à travers un processus de *matching* sans remise. Si l'âge est manquant, l'âge initial du véhicule, (donc lorsqu'il était un véhicule essence) est conservée.

DISCUSSION SUR LES HYPOTHÈSES SOUS-JACENTE À CETTE MÉTHODE

Comme pour la partie vieillissement du parc logement, la méthodologie ne vise pas à faire émerger des causalités ou à définir un modèle structurel en portant sur le parc de véhicules en tant que tel. Elle vise à refléter dans le modèle les changements à moyen terme observés dans les données servant à caler le modèle, plutôt qu'à les expliquer. En revanche, un modèle dynamique portant sur un parc de véhicules serait davantage approprié pour cela, avec des utilités et des arbitrages à chaque date, des coûts de transactions, et avec la modélisation des mises au rebut et d'un marché de l'occasion (pour un exemple voir *Schiraldi* (2011) ou *Li et al.* (2022)). Cependant, cette approche serait alors antinomique à l'idée de calage et à la modélisation statique.

Les calages

LES KILOMÉTRAGES PARCOURUS PAR LES VÉHICULES

À l'issue de l'étape de diésélisation, la part des kilométrages parcourus par les véhicules diesels (voitures et VUL particuliers) dans l'ensemble des kilométrages est calée sur les données du SDES.

On modifie ensuite le nombre total des kilomètres parcourus par l'ensemble des véhicules afin qu'il corresponde aux données du SDES. La cale qui va être utilisée pour les kilomètres est égale au ratio entre le kilométrage moyen parcouru par les ménages observés dans la base de Prometheus en année N-1 et le kilométrage moyen parcouru par les ménages obtenu via les données du SDES. On multiplie ainsi le kilométrage de tous les véhicules présents dans la base par cette cale ce qui permet de caler le kilométrage total des véhicules sur celui donné en entrée du modèle.

LES CONSOMMATIONS UNITAIRES DES VÉHICULES

De la même manière que les kilométrages, les consommations unitaires sont recalées pour tous les véhicules du parc présents dans la base.

D'abord on construit deux cales, une pour les véhicules diesel et une autre pour les véhicules essence. Celles-ci sont égales au ratio de la consommation unitaire moyenne des véhicules du carburant donné dans la base de données et la consommation unitaire moyenne des véhicules français du carburant donné pour l'année d'actualisation présentée dans les données du SDES. Ensuite, on multiplie les consommations unitaires des véhicules par la cale correspondant à leur carburant.

Encadré 7 : vers une électrification et une hybridation du parc ?

Actuellement, dans les données de l'EMP 2019, le nombre d'observations concernant les véhicules électriques et hybrides est inférieur à 30, soit le seuil de robustesse de représentativité statistique. Ce qui est insuffisant pour modéliser le passage à l'électrique ou hybride au sein de la base des ménages à partir de la méthodologie actuellement utilisée pour le passage au diesel ou à l'essence.

Des travaux sont en cours pour pallier ce problème, notamment en mobilisant la base appariée RSVERO-Fidéli² produite par le SDES (Maamari (2024) : « Acquérir une voiture électrique : pas seulement une question de revenu », Datalab Essentiel, SDES³).

¹ Dans la précédente enquête socle du modèle, l'ENTD 2008, ces véhicules étaient absents.

² www.statistiques.developpement-durable.gouv.fr/media/7230/download?inline.

³ www.statistiques.developpement-durable.gouv.fr/media/7182/download?inline.

C. Revenus

Les revenus sont mis à jour avec la même méthodologie qui a prévalu initialement dans la construction et traitements des variables (voir ci-dessus). Chaque année l'Insee publie les données de l'Enquête revenus fiscaux et sociaux (ERFS) qui fournissent des informations sur les revenus des ménages à l'année N-2. Ces données de distribution des niveaux de vie annuels des ménages sont utilisées pour la mise à jour des revenus dans Prometheus¹²³ : elles fournissent les montants des bornes supérieures de chaque décile (et P95), lesquels sont exprimés en euros constants. À l'aide des données de l'IPC, les montants sont traduits en euros courants.

Il existe quatre variables de revenus dans le modèle¹²⁴. Pour chacune de ces variables, ses déciles sont calculés. Ces derniers sont actualisés en les faisant évoluer comme les déciles de la dernière année ERFS disponible. Ensuite, le modèle cale la distribution de chaque variable sur les niveaux des déciles obtenus précédemment à l'aide d'ERFS¹²⁵.

Par ailleurs, annuellement, Prometheus est utilisé pour estimer le taux de précarité énergétique des ménages. Pour cela, il est nécessaire de disposer des revenus de l'année N-1. Comme mentionné auparavant, les niveaux d'ERFS ne sont alors pas encore disponibles. Par conséquent les indicateurs avancés de pauvreté du modèle Ines sont utilisés lorsqu'ils sont disponibles pour estimer en particulier l'évolution des trois premiers déciles de ERFS.

¹²³ Différents des « niveaux de vie moyen par décile ». Champ : France métropolitaine, individus vivant dans un ménage (en logement ordinaire) dont le revenu déclaré est positif ou nul et dont la personne de référence n'est pas étudiante.

¹²⁴ Il s'agit des variables nivvie, rfrUC, MRTOTA3UC et RevenuTEE. Cette dernière variable correspond à la restriction de la variable Rev_tot (voir III. 1) sur le champ des ménages hors étudiants et revenu négatif (pour mémoire, i.e. : champrev = 1).

¹²⁵ Par ailleurs le modèle permet également si l'on ne dispose pas (encore) des données d'ERFS, d'utiliser en entrée une seule valeur comme le revenu disponible brut par UC de la comptabilité nationale. Même si cela peut être fait « à la main » dans les inputs, en appliquant un taux de croissance commun aux valeurs des derniers déciles renseignées de ERFS.

VII. MISE À JOUR DE LA FISCALITÉ

Prometheus permet de calculer des factures énergétiques moyennes par catégorie de ménages. Pour cela, les prix moyens annuels de toutes les énergies du logement et du transport sont mis à jour chaque année, intégrant notamment les évolutions du système socio-fiscal en matière énergétique. Cela permet ainsi d'obtenir des factures énergétiques hors taxes (HT) et toutes taxes comprises (TTC), ainsi que le détail de la fiscalité.

Après une présentation de la prise en compte dans fiscalité dans Prometheus, ce chapitre présente, par énergie, le détail de la mise à jour des prix et du calcul des factures TTC ou HT¹²⁶.

A. La fiscalité dans Prometheus

Prometheus permet de quantifier, au niveau de chaque ménage, le montant payé des différentes composantes de la fiscalité énergétique.

Typologie

LA FISCALITÉ PRISE EN COMPTE DANS PROMETHEUS

La plupart des taxes qui entre en compte dans la facture énergétique des ménages sont intégrées dans Prometheus. On distingue la taxe sur la valeur ajoutée (TVA) et la fiscalité spécifique qui s'applique sur l'énergie. La TVA s'applique sur toute la facture énergétique des ménages, y compris sur les taxes énergétiques ; elle peut être à taux plein (20 %) ou à taux réduit (10 %, pour le bois, et 5,5 %, pour les contributions tarifaires d'acheminement). S'agissant de la fiscalité spécifique qui s'applique sur l'énergie, deux catégories de taxes peuvent être distinguées selon le type d'énergie concernée (combustibles et carburants d'un côté, électricité de l'autre)¹²⁷ :

- la fiscalité qui porte sur les combustibles et carburants s'applique sur, le fioul, le gaz, l'essence et le diesel. Sur le gaz naturel s'applique la fraction perçue sur le gaz (qui a remplacé la taxe intérieure de consommation sur le gaz naturel (TICGN)), dont le montant est proportionnel au volume de consommation, ainsi que la contribution tarifaire d'acheminement (CTA), de nature forfaitaire ;
- sur le carburant (essence et diesel), le fioul et le GPL, s'applique la *fraction sur les produits énergétiques autres que les gaz naturels et les charbons* (qui a remplacé la taxe intérieure de consommation sur les produits énergétiques (TICPE)), dont le montant est proportionnel au volume de consommation.

Au sein de ces fractions, peut être distinguée la « composante carbone » intégrée dans la fiscalité des énergies fossiles en 2014, et qui a augmenté continûment jusqu'en 2018 pour se situer à 44,6 €/tCO₂.

¹²⁶ Côté logement, les prix concernant les énergies sont renseignés dans les fichiers d'input du modèle toutes taxes comprises (TTC). Ensuite, la mise à jour des taxes permet de calculer des prix hors taxes (HT). Prometheus compile alors les factures HT et TTC des différentes consommations d'énergie, elles-mêmes actualisées, de chaque ménage. Côté transport, les prix renseignés initialement sont des prix HT, et la mise à jour de la fiscalité permet de calculer les prix et factures TTC.

¹²⁷ C'est cette typologie qui est utilisée dans le [rapport annuel sur l'impact environnemental du budget de l'Etat. Partie III.C.](#)

S'agissant de la fiscalité qui porte sur l'électricité, elle se décompose entre la fraction perçue sur l'électricité qui s'applique proportionnellement au volume d'électricité consommée (et qui a remplacé en 2022 formellement la taxe intérieure sur la consommation finale d'électricité (TICFE)) et au sein de laquelle ont été fondu respectivement en 2022 puis 2023, les deux taxes locales sur la consommation finale d'électricité (TLCFE), d'abord la taxe départementale en 2022, puis la taxe communale en 2023 ; et la contribution tarifaire d'acheminement (CTA), forfaitaire.

LA FISCALITÉ NON-PRISE EN COMPTE DANS PROMETHEUS

Trois autres taxes énergétiques ne sont pas prises en compte dans la simulation des factures énergétiques par le modèle Prometheus :

- la fraction perçue sur les charbons (anciennement taxe intérieure de consommation sur le charbon (TICC)), car les montants en jeu sont négligeables devant ceux de la fraction sur les produits énergétiques autres que les gaz naturels et les charbons et de la fraction gaz ;
- la fraction perçue en outre-mer sur les produits énergétiques, autres que les gaz naturels et les charbons (anciennement taxe spéciale de consommation (TSC)), en outre-mer, n'est pas prise en compte, car le modèle Prometheus ne porte que sur la France métropolitaine ;
- la taxe incitative relative à l'utilisation d'énergie renouvelable dans le transport (TIRUERT, anciennement taxe incitative relative à l'incorporation des biocarburants, TRIB) qui est une taxe additionnelle à la fraction perçue en métropole sur les produits énergétique autres que les gaz naturels et les charbons, car son rendement budgétaire est négligeable¹²⁸.

Sources

Avant 2022, les niveaux d'impositions des accises énergétiques se trouvaient renseignés dans le code des douanes. Depuis 2022, ils sont fixés dans le Code d'imposition sur les biens et services (CBIS)¹²⁹, y compris ceux des différentes exemptions et taux réduits. Pour les taxes non présentées dans ce code d'imposition (comme les CTA), il faut consulter les arrêtés en vigueur ; certaines taxes fixées dans le CBIS font l'objet d'une minoration ou majoration régionale. C'est le cas de la fraction sur le gaz et de la taxation des carburants.

Par ailleurs, il est parfois nécessaire de convertir l'unité de la base d'imposition (unité d'expression des taux d'imposition) en MWh (unité utilisée dans Prometheus). Les coefficients de conversion se trouvent dans *l'Arrêté du 13 décembre 2022 constatant divers tarifs et seuils de régime d'impositions relatifs à certaines impositions sur les biens et services*¹³⁰.

B. Électricité

Actualisation des prix

Les prix résidentiels de l'électricité auxquels font face les ménages sont actualisés grâce au tarif réglementé de vente TTC publiés par la Commission de régulation de l'énergie (CRE)¹³¹.

¹²⁸ Cette taxe fixe un objectif d'utilisation d'énergie renouvelable dans le transport au-delà duquel le montant dû au titre de cette taxe est nul pour le redevable.

¹²⁹ Partie législative > Livre III : énergies, alcools et tabacs > Titre I : régimes d'accise > Chapitre II : énergie > Section 3 : montant de l'accise > Sous-section 2.

¹³⁰ www.legifrance.gouv.fr/loda/id/LEGIARTI000048907823/2024-01-01/.

¹³¹ www.cre.fr/Pages-annexes/Open-Data. Dans la rubrique « Marché de détail » plus précisément les fichiers disponibles dans « historique des tarifs réglementés de vente d'électricité pour les consommateurs résidentiels ». Auparavant, comme déjà mentionné, les prix de la base Pégase du SDES étaient utilisés. Ainsi pour les prix de l'électricité, on suppose que tous

Les tarifs correspondent à la nomenclature évoquée dans le *chapitre IV. A.*, c'est-à-dire avec les différentes options « base » et « heures pleines/heures creuses », à la fois pour les parts variables et les abonnements, ainsi que pour les puissances¹³². Les tarifs ne sont actualisés que deux fois par an (février et août), donc un calcul au prorata est réalisé pour obtenir les tarifs moyens annuels. De plus, les données dites « TURPE » (tarifs d'utilisation des réseaux publics d'électricité) qui sont utiles (*voir infra*) sont récupérées.

Calculs des factures

Dans Prometheus, pour un ménage et pour chacune des catégories issues du croisement entre option et puissance :

- l'abonnement HT est égal à l'abonnement renseigné ;
- le montant CTA est égal au TURPE multiplié par le taux de CTA ;
- le prix HT correspond au tarif du ménage.

Le calcul de la facture par le modèle se fait alors de la façon suivante :

- facture HT = abonnement + prix HT* consommation ;
- fraction élec = consommation * taux de fraction élec/1 000 ;
- la part TVA est calculée comme l'application de la TVA à taux plein à ces trois composantes hors abonnement. Celle à taux réduite (5,5 %) s'applique à l'abonnement et au montant de CTA et à la partie abonnement.

$$TVA = \text{taux plein de TVA} * (\text{Facture HT} + \text{Fraction élec}) + \text{taux réduit TVA} * (\text{CTA} + \text{abonnement HT})$$

Enfin, la facture d'électricité TTC agrège ces différentes composantes :

$$\text{Facture électricité TTC} = \text{Facture électricité HT} + \text{Fraction élec} + \text{TVA} + \text{CTA}$$

C. Gaz naturel

Actualisation des prix

Les prix résidentiels du gaz sont actualisés également grâce au TRV TTC publié par la CRE tous les mois¹³³. Les tarifs correspondent à la nomenclature évoquée dans le *chapitre IV.1*, c'est-à-dire avec les différentes options (base, B0, B1, B2I) à la fois pour les parts variables et les abonnements¹³⁴. Un traitement est fait pour annualiser les tarifs moyens. Les données sur le CTA sont également récupérées sur le site de la CRE. Toutes ces données permettent de mettre à jour les fichiers d'inputs du modèle.

les ménages sont au tarif réglementé de vente (TRV) ou qu'il dispose un contrat indexé sur ce TRV, c'est en effet le cas de presque 80 % des ménages, *voir Chapitre IV* pour les limites concernant les offres à prix de marchés.

¹³² Par ailleurs, les évolutions de ces données sont « rattachées » aux évolutions des données historiques de Pégase. En effet, sur la CRE les données ne commencent qu'en 2013.

¹³³ www.cre.fr/Pages-annexes/Open-Data. Dans la rubrique « Marché de détail » plus précisément les fichiers disponibles dans « historique des tarifs réglementés de vente de gaz naturel pour les consommateurs résidentiels ». Auparavant, comme déjà mentionné, les prix de la base Pégase du SDES étaient utilisés. Aussi, de la même manière que pour l'électricité, pour les prix de l'électricité, on suppose que tous les ménages sont au tarif réglementé de vente ou avec un contrat indexé sur ce TRV, *voir chapitre IV* pour les limites concernant les offres à prix de marchés.

¹³⁴ Ici aussi, les évolutions de ces données sont « rattachées » à celles historiques de Pégase. En effet, sur la CRE les données ne commencent qu'en 2013.

Calculs des factures

Dans Prometheus, pour un ménage et pour chacune des catégories d'options :

- l'abonnement HT est égal à l'abonnement renseigné ;
- le montant de CTA est égal à la somme des tarifs de distribution et transport, eux-mêmes multipliés par leur taux de CTA respectif ;
- le prix HT correspond au tarif du ménage.

Le calcul de la facture par le modèle se fait alors de la façon suivante :

- facture HT = abonnement + prix HT * consommation ;
- CC^{135} gaz = consommation * CC * facteur d'émission du gaz PCS ;
- fraction gaz = CC gaz + consommation * taux de Fraction gaz hors composante carbone¹³⁶ ;
- la part TVA est calculée comme l'application de la TVA à taux plein à la facture HT hors abonnement et au montant de la fraction gaz, à laquelle s'ajoute celle à taux réduit (5,5 %) s'appliquant à l'abonnement et au montant de CTA :

$$TVA = \text{taux plein de TVA} * (\text{Facture HT} + \text{Fraction gaz}) + \text{taux réduit TVA} * (\text{CTA} + \text{abonnement HT})$$

Enfin, la facture gaz TTC agrège ces différentes composantes.

$$\text{Facture gaz TTC} = \text{Facture électricité HT} + \text{Fraction gaz} + \text{TVA} + \text{CTA}$$

D. Autres énergies de logement

Fioul

Les prix du fioul sont récupérés dans les séries longues du Bilan du SDES¹³⁷ (données annuelles) ou dans le tableau disponible sur la base Dido du SDES¹³⁸ (données mensuelles).

Dans Prometheus, pour un ménage qui possède une consommation de fioul, le calcul de la facture se fait de la façon suivante :

- facture HT = prix HT * consommation ;
- fraction fioul : somme de la composante carbone fioul et du taux de fraction sur les produits énergétiques autres que les gaz naturels et les charbons hors composante carbone, qui s'applique à la consommation¹³⁹ ;
- la part TVA est calculée comme l'application de la TVA à taux plein à la facture HT ajoutée du montant de la fraction fioul.

La facture TTC agrège ces différentes composantes.

¹³⁵ CC pour composante carbone.

¹³⁶ Ce taux est le taux de fraction gaz hors composante carbone = taux de fraction gaz - CC * facteur d'émission gaz PCS.

¹³⁷ « Prix fioul domestique à la consommation livraison 2-5 kl (TTC) », €/l.

¹³⁸ www.statistiques.developpement-durable.gouv.fr/donnees-mensuelles-de-lenergie -> « Accéder aux données conjoncturelles de l'énergie »/« PX_PETRO_FOD_100KHW_C1 » (« Prix au détail de 10 kWh PCI de FOD au tarif C1 »).

¹³⁹ Où : Taux de fraction sur les produits énergétiques autres que les gaz naturels et les charbons hors composante carbone = taux de fraction sur les produits énergétiques autres que les gaz naturels et les charbons - CC * facteur d'émission du fioul.

Bois

Les données de prix des granulés et bois-bûches sont disponibles dans les séries longues du « Bilan annuel du SDES »¹⁴⁰. La série de prix des granulés est également disponible sur Dido en mensuel¹⁴¹.

Dans Prometheus, pour un ménage qui possède une consommation de bois, le calcul de la facture se fait de la façon suivante :

- facture HT = prix HT * consommation ; le prix est différencié selon que le combustible est le granulé ou la bûche ;
- la part TVA est calculée comme l'application de la TVA à taux réduit (10 %) à la facture HT.

La facture bois TTC est la somme de ces deux composantes.

Chauffage et réseau de chaleur urbain

Pour le chauffage urbain, l'input de Prometheus fait intervenir une tarification qui varie selon la saison (entre l'hiver et l'été), et la durée d'utilisation (longue, moyenne) et qui comprend une part variable et une part abonnement. Cela correspond à la grille tarifaire de la Compagnie parisienne de chaleur urbaine (CPCU)¹⁴².

Dans Prometheus, pour un ménage qui possède une consommation de chauffage urbain, l'hypothèse est faite qu'il est concerné par la tarification destinée aux longues utilisations :

- l'abonnement HT est égal à l'abonnement destiné à la longue utilisation ;
- le prix est un prix pondéré à 85 % du tarif hiver et 15 % du tarif été.

Le calcul de la facture se fait de la façon suivante :

- facture HT = abonnement + prix HT pondéré* consommation ;
- la part TVA est calculée comme l'application de la TVA à taux plein à la facture HT hors abonnement, tandis que la TVA à taux réduit (5,5 %) s'applique à l'abonnement.

La facture chauffage urbain TTC est la somme ces deux composantes.

¹⁴⁰ « Prix granulés en sacs PCI au 1^{er} trimestre : palette départ fournisseur (TTC) » et « Prix bois-bûche en vrac PCI au 1^{er} trimestre 50 cm H1 : hors livraison (TTC) » et les données sont divisées par 1 000 pour passer de €/MWh à un prix unitaire.

¹⁴¹ « PX_BOIS_GRANDSACS_100KWH » : comme la série annuelle correspond au 1^{er} trimestre, c'est utile dans le cas d'une mise à jour avancée de la donnée.

¹⁴² www.cpcu.fr/documentation/. Ce traitement est cohérent avec le traitement initial détaillé en IV. Auparavant, les données des prix de la chaleur provenaient de la base de données Pégase avec la même nomenclature. Le SDES remplissait cette base de données avec en partie les données de la CPCU, sachant que la moitié des ménages utilisant ce mode de chauffage vie dans l'unité urbaine de Paris, le maintien de cette source parisienne reste une bonne approximation. Il est également possible d'utiliser la série publiée dans le bilan de l'énergie « Prix de la chaleur à la consommation pour les secteurs autres que l'industrie (TTC) », mais d'une part cela implique d'appliquer un prix unique à tous les ménages possédant cette énergie, et d'autre part ce prix concerne certaines consommations du secteur tertiaire et ne concerne donc pas que le secteur résidentiel (i.e. des ménages).

GPL logement

Les prix unitaires en bouteille (butane)¹⁴³ et en citerne (propane)¹⁴⁴, sont fournis par la base Dido (même tableau que le fioul et même fonctionnement), ou encore pour les prix en bouteille, dans les données annuelles disponibles dans les séries longues du Bilan du SDES.

Dans Prometheus, le calcul de la facture en GPL domestique se fait de la façon suivante pour chaque ménage :

- facture HT = prix HT* consommation ; le prix est différencié selon que l'énergie principale de chauffage est le GPL ou non¹⁴⁵ ;
- fraction GPL : somme de la composante carbone GPL et du taux de fraction sur les produits énergétiques autres que les gaz naturels et les charbons hors composante carbone¹⁴⁶, qui s'applique à la consommation ;
- la part TVA est calculée par l'application du taux plein de TVA à la facture HT.

La facture TTC agrège ces différentes composantes.

E. Les carburants

Les prix HT des carburants (essence et diesel) sont obtenus sur le site du ministère Territoires Écologie Logement et mis à jour en annuel dans le modèle¹⁴⁷. Pour chaque carburant, le calcul de la facture par le modèle est réalisé de la façon suivante :

- facture HT = prix HT * consommation ;
- fraction carburant = somme de la composante carbone carburant et du taux de la fraction sur les produits énergétiques autres que les gaz naturels et les charbons hors CC¹⁴⁸, qui s'applique à la consommation ;
- la part TVA est calculée comme l'application de la TVA à taux plein à la facture HT et au montant de la fraction carburant.

Enfin, la facture de carburant TTC agrège ces différentes composantes.

¹⁴³ « PX_PETRO_BUTANE_BOUTEILLE13KG », qui correspond au « prix au détail d'une bouteille de butane de 13 kg » et disponible sous le même nom dans les séries longues en annuel.

¹⁴⁴ La série n'est pas disponible dans la base Dido du SDES, cependant elle se trouve dans les séries longues du Bilan du SDES : « prix du gaz propane liquéfié PCI à la consommation (TTC) ».

¹⁴⁵ Il est fait l'hypothèse que si c'est le cas, le ménage utilise une citerne, et dans le cas contraire, plutôt des bouteilles.

¹⁴⁶ Ce taux de fraction sur les produits énergétiques autres que les gaz naturels et les charbons hors composante carbone est égal au taux de la fraction sur les produits énergétiques autres que les gaz naturels et les charbons auquel on soustrait le facteur d'émission du GPL multiplié par le taux de la hausse de composante carbone.

¹⁴⁷ www.ecologie.gouv.fr/prix-des-produits-petroliers.

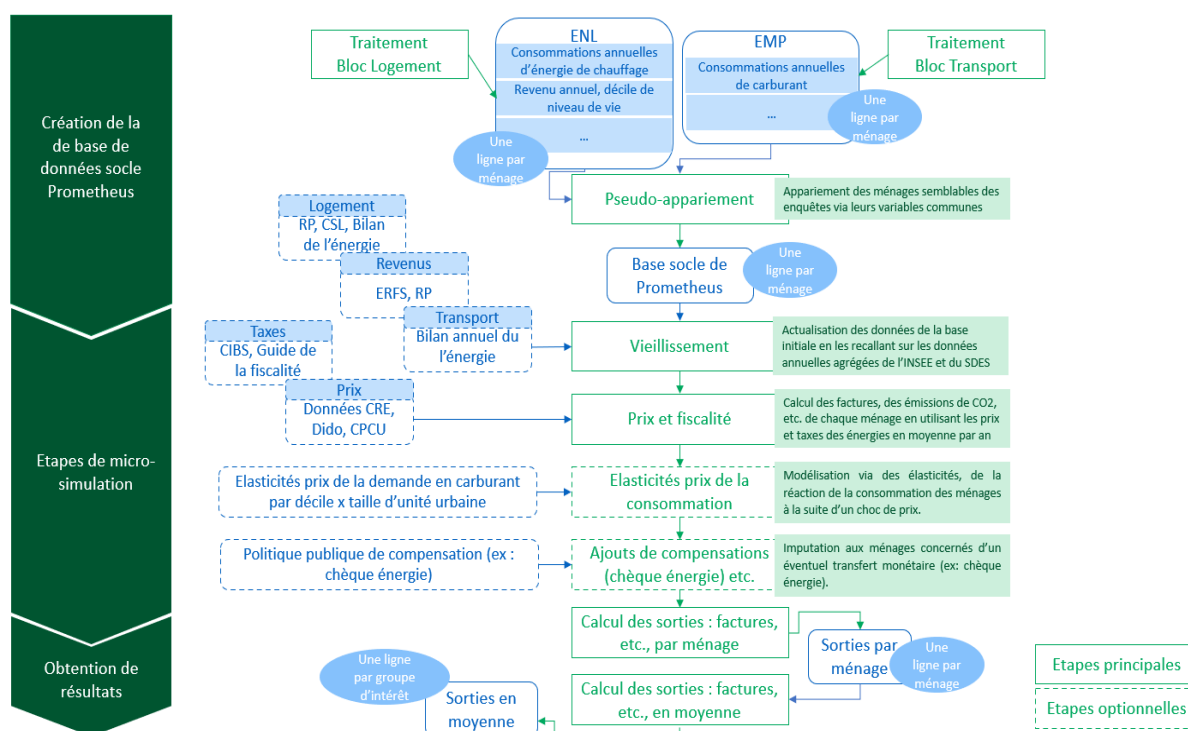
¹⁴⁸ Ce taux est le taux de fraction sur les produits énergétiques autres que les gaz naturels et les charbons hors composante carbone = taux de fraction sur les produits énergétiques autres que les gaz naturels et les charbons hors composante carbone - CC * facteur d'émission carburant.

VIII. SYNTHÈSE ET UTILISATION PRATIQUE DU MODÈLE

A. Architecture du modèle

Le modèle Prometheus peut être synthétisé dans son organisation, ses traitements et ses principales sources comme dans la *figure 6* (les différents sigles ont tous été expliqués *ci-dessus*) :

Figure 6 : structure détaillée du modèle Prometheus



Source : CGDD, modèle Prometheus

B. Actualisation des fichiers d'inputs

Le code *R* en lui-même est relativement complexe dans son architecture et dans certains traitements, notamment dans les parties « vieillissement annuel ». Néanmoins, en pratique, le code n'a pas vocation à être modifié à chaque lancement de simulation (ponctuellement bien sûr le code peut être amélioré ou modifié). Il suffit de renseigner les dates voulues de chargement des paramètres dans le programme de lancement et de l'exécuter : il va alors appeler l'ensemble des autres programmes nécessaires à la bonne exécution du modèle.

En revanche, le travail de mise à jour en cours d'année sur le modèle consiste avant tout à actualiser des fichiers de paramètres (ce qui nécessite pas mal de traitements en amont à partir des données sources). Concrètement, dans chaque fichier *Excel* il y a un onglet par type de données, et une colonne par année pour l'ensemble des valeurs à renseigner pour chaque variable.

Le fichier de calage et vieillissement

Le fichier compile l'ensemble des données annuelles concernant la structure et les consommations des ménages, côté logement et côté transport, nécessaire pour le vieillissement du modèle. Comme montré précédemment, ce fichier permet d'actualiser dans Prometheus les consommations d'énergie par type de logement, de chauffage et d'énergie, le nombre de logements, mais aussi les kilométrages moyens selon les carburants et les consommations unitaires par types de véhicules. Ce fichier contient aussi les données concernant les revenus et la météo.

Le fichier de prix et fiscalité

Le fichier compile l'ensemble des données annuelles concernant les prix et la fiscalité selon toutes les énergies présentes dans Prometheus, côté logement et côté transport. Ainsi le fichier contient l'ensemble des prix ou tarifs de gaz, électricité, GPL bois, fioul, chauffage urbain, des différents carburants, ainsi que toutes les taxes spécifiques aux énergies ou la fiscalité plus générale comme la TVA ou la composante carbone.

C. Actualisation du programme de lancement

Les dates de paramètres

Avant chaque lancement, après avoir précisé les années que l'on veut simuler, il faut veiller à bien renseigner les dates de chargement de chaque type de paramètres. Ainsi, quand toutes les données sont à jour pour l'année N, pour obtenir l'ensemble des factures à jour, il faut indiquer l'année N pour chaque type de paramètre (taxes, prix, revenus, consommations, etc.). Si l'on veut par exemple faire un scénario avec les prix de l'année précédente, toutes choses égales par ailleurs, alors il faut indiquer N-1 pour l'année des prix.

Les paramètres globaux

Pour chaque lancement, il est aussi possible de configurer si l'on veut appliquer la correction météo, un mécanisme compensatoire, les élasticités, mais aussi les fichiers que l'on veut en sortie. En effet, en plus de la base dite « brute », il est aussi possible d'avoir un fichier de synthèse avec les principaux résultats en moyenne pour les ménages ou encore un fichier avec toutes les données qui sera au bon format pour être lu par une application interne *RShiny*.

IX. CAHIER DE VARIANTES¹⁴⁹

Le modèle Prometheus permet d'estimer la facture énergétique des ménages, ainsi que les écarts de cette facture entre deux situations. D'autres grandeurs peuvent également être estimées, comme le taux d'effort énergétique (facture énergétique rapportée au revenu disponible du ménage). Par ailleurs, la facture énergétique peut être décomposée entre ce qui relève du logement et du transport, entre les différentes énergies consommées, ou encore entre la partie hors taxes et celle relative à la fiscalité.

Les grandeurs peuvent être estimées selon diverses catégories de ménages, notamment sociodémographiques. Deux dimensions d'hétérogénéité sont retenues comme fils conducteurs dans ce chapitre. D'une part une dimension verticale, à savoir le niveau de vie, et d'autre part une dimension horizontale, à savoir la catégorie de commune dans la nomenclature d'aire d'attraction des villes de l'Insee¹⁵⁰. En annexe, deux autres dimensions d'hétérogénéité horizontales sont également considérées : l'énergie principale de chauffage du logement et les équipements en véhicules.

Ce chapitre commence par une description de la facture énergétique de 2019, telle que simulée par Prometheus et prise comme année de référence¹⁵¹. Sont ensuite considérées trois types de « chocs » : des chocs sur le prix TTC des principales énergies (carburants routiers, gaz, électricité) ; des chocs fiscaux (composante carbone, accise sur l'électricité, TVA) ; des évolutions dans les dispositifs d'accompagnement des ménages (chèque énergie classique, chèque énergie augmenté de 100 euros). Par simplicité, les effets des chocs sont estimés toutes choses égales par ailleurs, sans ajustement des volumes d'énergie consommées. Ils peuvent donc s'interpréter comme l'effet instantané du « choc » sur la facture énergétique des ménages (ou sur les autres grandeurs documentées).

Les grandeurs présentées sont annuelles. Les montants sont exprimés en euros courants de l'année de référence (2019). Les consommations d'énergie du logement, et donc les factures et les taux d'effort qui en découlent, sont corrigés des conditions météorologiques (consommations d'énergie à conditions météorologiques « normales »). Les factures énergétiques et les taux d'effort énergétiques calculés sont bruts du chèque énergie, sauf mention contraire. Pour une explication de la construction des différentes catégories d'énergie ou de taxes, (voir les chapitres IV. A. sur le traitement de l'ENL 2013 et VII. sur la fiscalité du présent document).

¹⁴⁹ Ce chapitre est issu des travaux de *Quentin Nicod* en 2024, à l'époque agent du bureau, dans le cadre de sa formation continue et de son mémoire de master Ensaï.

¹⁵⁰ Le zonage en aires d'attraction des villes a succédé en 2020 au zonage en aires urbaines de 2010. Selon l'Insee, « l'aire d'attraction d'une ville est un ensemble de communes, d'un seul tenant et sans enclave, qui définit l'étendue de l'influence d'un pôle de population et d'emploi sur les communes environnantes, cette influence étant mesurée par l'intensité des déplacements domicile-travail ». Une aire d'attraction est constituée d'un pôle et d'une couronne. Ces pôles sont déterminés principalement à partir de critères de densité et de population totale. Les communes qui envoient au moins 15 % de leurs actifs travailler dans le pôle constituent la couronne de l'aire. Au sein des pôles, la commune la plus peuplée est appelée commune-centre. (Voir le site de l'Insee www.insee.fr/fr/information/4803954 et la Note méthodologique sur la constitution des aires d'attraction des villes 2020 de l'Insee).

¹⁵¹ Cette année est choisie comme référence, car elle est antérieure à la crise du Covid de 2020, ainsi qu'à la crise énergétique qui a débuté fin 2021 (pour une analyse redistributive de l'effet des hausses du prix des énergies et des mesures de soutien aux ménages, voir notamment *Nicod et al. (2023)*). Par ailleurs, les données d'entrée utilisées pour calibrer le modèle et relatives à l'année 2023 ne sont pas encore disponibles au moment de la rédaction de ce document.

A. La facture énergétique des ménages en 2019 : estimation et analyse descriptive

Cette section présente une photographie de la situation en 2019, prise comme année de référence. L'analyse est d'abord réalisée pour un ménage moyen. Elle est ensuite précisée selon deux dimensions centrales d'hétérogénéité : le niveau de vie et la catégorie de commune dans la nomenclature d'aire d'attraction des villes de l'Insee. Les analyses de la facture énergétique sont déclinées en niveau et en part du revenu via le taux d'effort énergétique. Différentes décompositions de la facture énergétique et du taux d'effort sont proposées : selon la composante logement et la composante transport de la facture¹⁵², selon les types d'énergie, selon la composante hors taxes et la partie fiscalité, avec une ventilation par grandes catégories de taxe, et en faisant apparaître la composante carbone.

Des statistiques complémentaires sont disponibles dans les *annexes 15 à 17*. Les *annexes 15 et 16* fournissent des statistiques respectivement univariées et bivariées sur les caractéristiques des ménages. L'*annexe 17* présente quant à elle des extensions des résultats sur la facture énergétique et sur le taux d'effort en croisant deux dimensions : niveau de vie et catégories de commune, niveau de vie et énergie principale de chauffage, niveau de vie et équipement en véhicules.

Ces résultats permettent de dresser un panorama précis de la situation de l'année de référence, qui sert de base aux variantes de hausses de fiscalité et de mise en place de mesures d'accompagnement présentées par la suite.

Niveaux moyens

La facture énergétique moyenne d'un ménage en 2019 est de 2 796 € en 2019, soit 6,62 % de son revenu.

DÉCOMPOSITION ENTRE LES COMPOSANTES LOGEMENT ET TRANSPORT

Le *tableau 12* montre que les composantes relatives au logement et au transport sont quasi équivalentes et représentent donc chacune environ la moitié de cette facture.

DÉCOMPOSITION PAR ÉNERGIE

Les dépenses moyennes par type d'énergie présentées dans le *tableau 13* où elles sont classées par montant décroissant, sont quant à elles très hétérogènes.

Le gazole et l'électricité représentent de loin les deux premiers postes de dépenses par énergie au sein de la facture énergétique moyenne des ménages. La première place du gazole s'explique par la prépondérance toujours forte des véhicules diesels dans le parc automobile. En 2019, 74 % des kilomètres parcourus par les ménages étaient réalisés dans une voiture diesel. La seconde place de l'électricité reflète son rôle central dans l'alimentation des logements en énergie et un prix en 2019 plus élevé que celui du gaz.

Les deux postes suivants de dépenses par énergie sont l'essence et le gaz. Les autres postes, à commencer par le fioul, reflètent principalement les énergies principales de chauffage utilisées.

DÉCOMPOSITION PAR DIMENSIONS FISCALES

La facture moyenne peut aussi être décomposée entre une partie « hors toutes taxes » et une partie « fiscalité ». Cette décomposition est présentée dans le *tableau 14*. La partie hors toutes taxes représente 1 520 €, soit un peu plus de la moitié de la facture énergétique, tandis que l'ensemble des taxes représente 1 276 €.

¹⁵² Les énergies qui composent la facture transport sont le diesel, l'essence et le GNV (voir chapitres III. B. et VI. B. pour du détail sur le traitement du transport et des carburant ainsi que la place de l'électrique).

Tableau 12 : décomposition de la facture énergétique entre les composantes logement et transport en 2019

Composante	Factures (en €)	Taux d'effort énergétique (en %)
Logement	1 402	3,32
Transport	1 394	3,30
Total	2 796	6,62

Tableau 13 : décomposition de la facture énergétique par énergie¹⁵³ en 2019

Décomposition	Factures (en €)	Taux d'effort énergétique (en %)
Gazole	1 048	2,48
Électricité	814	1,93
Essence	343	0,81
Gaz	335	0,79
Fioul	152	0,36
GPL	40	0,09
Bois	30	0,07
Urbain	20	0,05
GNV	3	0,01
Autres	11	0,03
Total	2 796	6,62

Tableau 14 : décomposition de la facture énergétique et du taux d'effort énergétique entre le hors toutes taxes et la fiscalité en 2019

Décomposition	Factures (en €)	Taux d'effort énergétique (en %)
Facture énergétique hors toutes taxes	1 520	3,60
Fiscalité totale	1 276	3,02
TVA	441	1,04
Fiscalité sur les combustibles et carburants	668	1,58
Composante carbone	166	0,39
Hors composante carbone	502	1,19
Fiscalité sur l'électricité	167	0,40
Facture énergétique totale	2 796	6,62

Source : CGDD, modèle Prometheus

¹⁵³ Pour des précisions sur cette nomenclature, voir chapitre IV. sur le traitement des bases ENL 2013 et EMP 2019.

Selon le niveau de vie

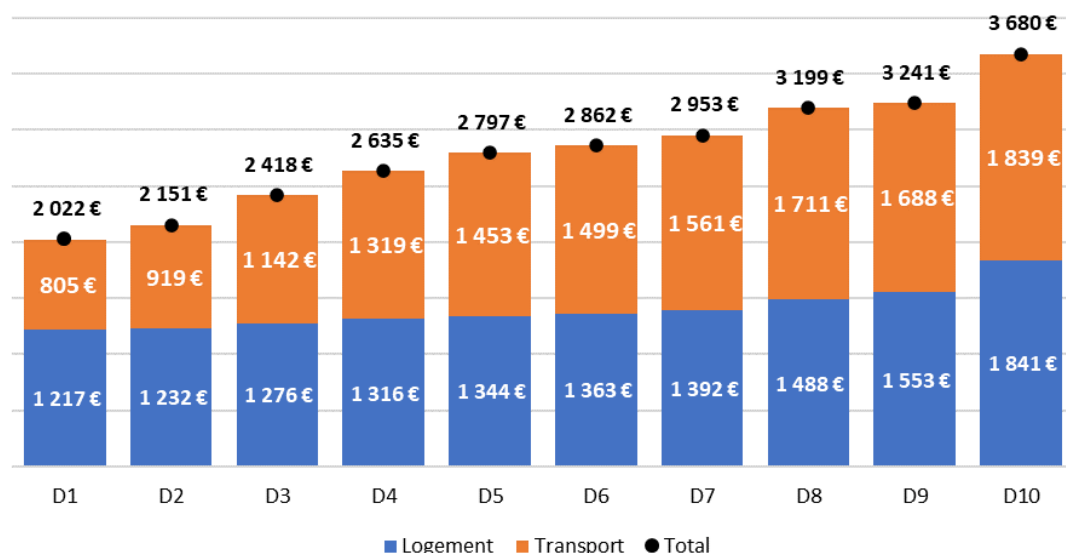
Une analyse similaire est présentée selon le dixième de niveau de vie (voir chapitres IV. B. et VI. C. pour le détail sur cette variable et son vieillissement). Dans ce chapitre, les factures énergétiques moyennes selon les différentes décompositions sont présentées dans un premier temps, suivies, dans un second temps, des taux d'effort énergétique moyens.

FACTURE ÉNERGÉTIQUE

La facture énergétique est croissante avec le niveau de vie comme le montre la *figure 7*.

La facture énergétique totale augmente avec le niveau de vie du ménage, traduisant l'augmentation des quantités d'énergie consommée. Les composantes logement et transport sont plutôt équivalentes au sein de chaque dixième de niveau de vie. Pour les trois premiers dixièmes, qui concernent les ménages les plus modestes, la facture logement en moyenne est un peu supérieure à la facture transport. Tandis que pour les sept dixièmes suivants, qui concernent des ménages plus aisés, c'est la facture transport qui est un peu supérieure. Cette inversion de l'ordre des composantes s'explique par des consommations d'énergie qui augmente plus vite avec le niveau de vie en matière de transport que pour logement.

Figure 7 : décomposition de la facture énergétique par dixièmes de niveau de vie et entre les composantes logement et transport en 2019



Source : CGDD, modèle Prometheus

(a) DÉCOMPOSITION PAR ÉNERGIE

En ventilant les factures énergétiques par dixième de niveau de vie et par énergie, de nouvelles tendances, dont témoignent le *tableau 15*, se dégagent.

Les dépenses dans les cinq principaux types d'énergie (gazole, électricité, essence, gaz et fioul) sont toutes croissantes avec le niveau de vie, et traduisent encore le profil croissant des quantités consommées de ces énergies. Pour les deux premiers dixièmes de niveau de vie, l'électricité est la forme d'énergie qui représente le premier poste de dépense, suivi du gazole, du gaz et de l'essence. Pour les huit dixièmes suivants, le gazole représente le premier poste de dépense, suivi de l'électricité. Les deux postes suivants sont, pour les D3 à D5, le gaz puis l'essence, et, à partir du D6, l'essence puis le gaz. Ces différences d'ordre des postes par dixième sont principalement le reflet, comme pour la décomposition entre logement et transport, des distances parcourues en moyenne croissantes avec le niveau de vie.

Tableau 15 : décomposition de la facture énergétique par dixièmes de niveau de vie et par énergies en 2019

En euros

Dixième de niveau de vie	Gazole	Electricité	Essence	Gaz	Fioul	GPL	Bois	Urbain	GNV	Autres	Total
D1	645	713	159	296	106	34	25	27	-	17	2 022
D2	689	690	230	303	128	45	26	26	-	14	2 151
D3	875	724	266	307	144	41	27	21	1	12	2 418
D4	1037	768	281	316	137	39	30	17	1	9	2 635
D5	1167	786	284	305	148	50	29	18	2	8	2 797
D6	1157	816	339	305	146	38	32	14	4	11	2 862
D7	1176	832	381	309	153	39	35	15	4	9	2 953
D8	1246	891	463	329	167	44	36	16	1	6	3 199
D9	1236	901	447	387	167	36	36	20	5	6	3 241
D10	1251	1024	578	496	227	36	28	25	10	7	3 680

Source : CGDD, modèle Prometheus

(b) DÉCOMPOSITION PAR DIMENSIONS FISCALES

Pour chaque dixième de revenu, la ventilation entre facture énergétique hors toutes taxes et les différentes catégories de taxes est présentée dans le *tableau 16*.

La partie hors toutes taxes de la facture énergétique ainsi que les montants de taxes par catégories sont croissants avec le niveau de vie, traduisant de nouveau la croissance des quantités d'énergie consommée avec le niveau de vie. La fiscalité est légèrement croissante avec le niveau de vie. Par ailleurs, chez les ménages du D1, elle représente 57 % de la facture, contre 54 % pour ceux du D10. Cela vient du fait que la part de la composante transport dans la facture énergétique augmente avec le niveau de vie et que les carburants sont plus taxés que les énergies du logement.

Tableau 16 : décomposition de la facture énergétique par dixièmes de niveau de vie et entre le hors toutes taxes et la fiscalité en 2019

Dixième de niveau de vie	Facture énergétique (en euros)							Total
	Facture hors toutes taxes	Fiscalité					Total	
		TVA	Fiscalité sur l'électricité	Fiscalité sur les combustibles et carburant				
				Composante Carbone	Hors Composante Carbone			
D1	1 161	314	146	110	291	401	861	2 022
D2	1 219	335	141	121	336	457	932	2 151
D3	1 337	379	148	142	413	555	1 081	2 418
D4	1 434	415	157	156	474	630	1 201	2 635
D5	1 507	441	161	167	520	687	1 290	2 797
D6	1 535	452	168	170	537	707	1 327	2 862
D7	1 579	466	171	176	561	737	1 374	2 953
D8	1 701	507	184	191	617	808	1 498	3 199
D9	1 737	513	186	197	608	805	1 504	3 241
D10	1 987	584	212	228	669	897	1 693	3 680

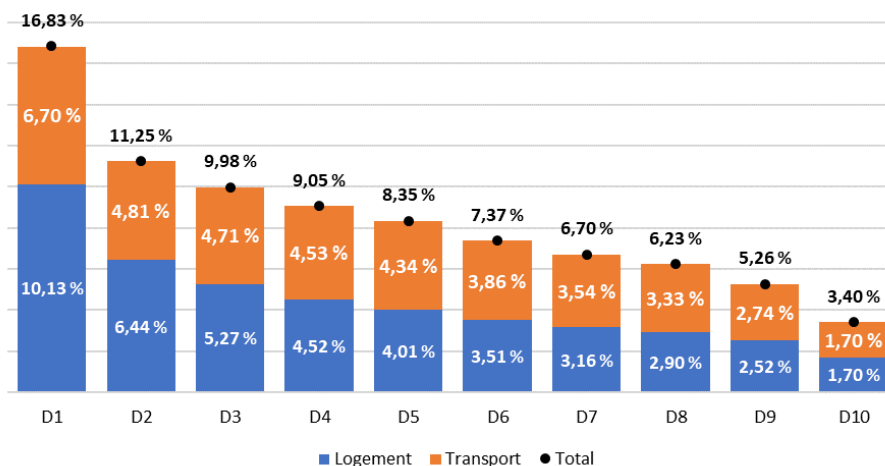
Source : CGDD, modèle Prometheus

TAUX D'EFFORT ÉNERGÉTIQUE

Le taux d'effort énergétique est fortement décroissant avec le niveau de vie comme le montre la figure 8.

En effet, même si les niveaux de consommation sont croissants en le niveau de vie, l'énergie représente une plus grande part de la consommation pour les ménages modestes. Pour les ménages du D1, la facture énergétique représente un sixième de leur revenu, tandis qu'elle en représente un trentième pour les ménages du D10.

Figure 8 : décomposition du taux d'effort énergétique par dixièmes de niveau de vie et entre les composantes logement et transport en 2019



Source : CGDD, modèle Prometheus

(c) DÉCOMPOSITION PAR ÉNERGIE

Cette tendance se retrouve en décomposant pour chaque dixième de niveau de vie le taux d'effort énergétique par énergies comme cela est fait dans le *tableau 17*. Les taux d'effort pour les cinq principaux types d'énergie (gazole, électricité, essence, gaz et fioul) sont décroissants en le niveau de vie.

Tableau 17 : décomposition du taux d'effort énergétique par dixièmes de niveau de vie et par énergie en 2019

En euros

Dixième de niveau de vie	Gazole	Electricité	Essence	Gaz	Fioul	GPL	Bois	Urbain	GNV	Autres	Total
D1	5,37	5,93	1,32	2,46	0,88	0,28	0,21	0,22	0,00	0,16	16,83
D2	3,60	3,61	1,20	1,58	0,67	0,24	0,14	0,14	0,00	0,07	11,25
D3	3,61	2,99	1,10	1,27	0,59	0,17	0,11	0,09	0,00	0,05	9,98
D4	3,56	2,64	0,97	1,09	0,47	0,13	0,10	0,06	0,00	0,03	9,05
D5	3,48	2,35	0,85	0,91	0,44	0,15	0,09	0,05	0,01	0,02	8,35
D6	2,98	2,10	0,87	0,79	0,38	0,10	0,08	0,04	0,01	0,02	7,37
D7	2,67	1,89	0,86	0,70	0,35	0,09	0,08	0,03	0,01	0,02	6,70
D8	2,43	1,74	0,90	0,64	0,33	0,09	0,07	0,03	0,00	0,00	6,23
D9	2,01	1,46	0,73	0,63	0,27	0,06	0,06	0,03	0,01	0,00	5,26
D10	1,16	0,95	0,53	0,46	0,21	0,03	0,03	0,02	0,01	0,00	3,40

Source : CGDD, modèle Prometheus

(d) DÉCOMPOSITION PAR DIMENSIONS FISCALES

Pour chaque dixième de revenu, la ventilation entre taux d'effort lié à la facture énergétique hors toutes taxes et taux d'effort liés aux différentes catégories de taxes est présentée dans le *tableau 18*.

Tableau 18 : décomposition du taux d'effort énergétique par dixièmes de niveau de vie et entre le hors toutes taxes et la fiscalité en 2019

Dixième de niveau de vie	Facture énergétique (en %)							
	Facture hors toutes taxes	Fiscalité					Total	Total
		TVA	Fiscalité sur l'électricité	Fiscalité sur les combustibles et carburant				
				Composante Carbone	Hors Composante Carbone	Total		
D1	9,65	2,61	1,22	0,92	2,43	3,35	7,18	16,83
D2	6,38	1,75	0,74	0,63	1,75	2,38	4,87	11,25
D3	5,53	1,56	0,61	0,58	1,70	2,28	4,45	9,98
D4	4,92	1,42	0,54	0,54	1,63	2,17	4,13	9,05
D5	4,50	1,32	0,48	0,50	1,55	2,05	3,85	8,35
D6	3,95	1,17	0,43	0,44	1,38	1,82	3,42	7,37
D7	3,58	1,06	0,39	0,40	1,27	1,67	3,12	6,70
D8	3,31	0,99	0,36	0,37	1,20	1,57	2,92	6,23
D9	2,82	0,83	0,30	0,32	0,99	1,31	2,44	5,26
D10	1,83	0,54	0,20	0,00	0,62	0,62	1,36	3,19

Source : CGDD, modèle Prometheus

Ces taux d'effort sont également nettement décroissants avec le niveau de vie et mettent en évidence le caractère régressif¹⁵⁴ des catégories de taxes présentées (fiscalité sur les combustibles et carburants, fiscalité sur l'électricité, TVA sur la facture énergétique). Cela s'explique de la même manière que la décroissance du taux d'effort, l'énergie représentant une plus grande part de la consommation pour les ménages modestes.

Selon la catégorie de commune¹⁵⁵

DÉCOMPOSITION DES FACTURES ÉNERGÉTIQUES

Les factures énergétiques sont plus faibles pour les habitants des pôles que pour les habitants des communes de couronne et des communes situées hors d'une aire d'attraction de ville, comme le montre la *figure 9*.

Les habitants des communes-centres de pôle possèdent les factures énergétiques les plus faibles. Viennent ensuite les habitants résidant en dehors des communes-centres de pôle. Enfin, les habitants des communes de couronnes et ceux de communes situées hors d'une aire d'attraction de ville ont les factures énergétiques les plus élevées, avec des niveaux assez proches entre ces deux catégories. Ces faits stylisés sont présents en regardant la facture énergétique totale, mais aussi en regardant isolément la facture énergétique logement et la facture énergétique transport.

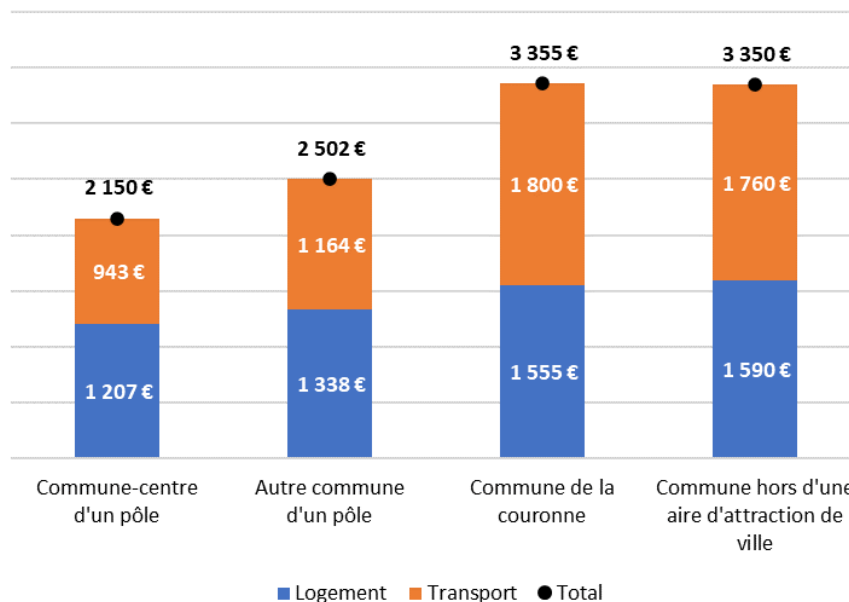
La disparité en matière de logement s'explique par des logements de plus grande surface à mesure que l'on s'éloigne des centres (*voir figure 12 de l'annexe 16*), nécessitant davantage

¹⁵⁴ Pour une analyse comparée de la régressivité des taxes indirectes, voir par exemple André et al. (2016), page 40.

¹⁵⁵ Pour du détail sur la définition des catégories de communes, voir la note de bas de page numéro 148.

d'énergie pour le chauffage. En ce qui concerne le transport, l'explication réside dans de plus grandes distances à parcourir et la moindre disponibilité de transport en commun, qui se matérialisent par une détention de véhicule accrue (voir figure 13 de l'annexe 16). Entre la facture énergétique logement et la facture énergétique transport, la gradation est un peu plus marquée du côté transport.

Figure 9 : décomposition de la facture énergétique par catégories de commune et entre les composantes logement et transport en 2019



Source : CGDD, modèle Prometheus

(a) DÉCOMPOSITION PAR ÉNERGIE

Il existe également une hétérogénéité des factures selon les différentes énergies selon la catégorie de communes des ménages. Cette décomposition est présentée dans le tableau 19.

Le gazole et l'électricité représentent les deux premiers postes de dépenses dans toutes les catégories de commune considérées, avec des niveaux croissants à mesure que l'on s'éloigne des centres. L'explication réside comme précédemment dans de plus grandes distances à parcourir en véhicule particulier et dans de plus grands logements.

Les autres énergies présentent des profils différents. La facture d'essence tend aussi à être croissante lorsqu'on s'éloigne des centres, mais cette gradation est bien moins marquée que pour le gazole. Cela traduit qu'en dehors des pôles, les ménages sont davantage équipés en véhicules diesel. Les factures de gaz les plus élevées sont retrouvées dans les pôles, dans la mesure où il s'agit de l'énergie principale de chauffage de plus de la moitié des ménages des pôles (voir figure 6 de l'annexe), en lien avec la possibilité d'être raccordé au gaz de réseau.

À l'inverse, les factures de fioul les plus élevées sont retrouvées en dehors des pôles. Les niveaux hors des aires d'attraction des villes sont supérieurs à ceux de la couronne, traduisant l'utilisation du fioul comme énergie principale de chauffage dans une plus forte proportion hors des aires d'attraction des villes qu'en couronne (voir figure 6 de l'annexe).

Par ailleurs, les dépenses de chauffage urbains sont principalement localisées dans les pôles, tandis que les dépenses de bois de chauffage le sont principalement hors des pôles.

Tableau 19 : décomposition de la facture énergétique par catégories de commune et par énergies en 2019

En euros

Catégorie de commune	Gazole	Electricité	Essence	Gaz	Fioul	GPL	Bois	Urbain	GNV	Autres	Total
Commune-centre d'un pôle	660	640	283 €	426	72	14	10	0 €	10	35	2 150
Autre commune d'un pôle	810	724	353	465	76	13	11	1	12	37	2 502
Commune d'une couronne	1 410	970	384	242	218	64	50	5	10	2	3 355
Commune hors d'une aire d'attraction de ville	1 427	977	329	47 €	380	106	70	5	8	1	3 350

Source : CGDD, modèle Prometheus

(b) DÉCOMPOSITION PAR DIMENSIONS FISCALES

Les niveaux de factures correspondant aux différentes taxes, présentés dans le *tableau 20*, évoluent selon la catégorie de communes des ménages.

Les niveaux des différentes composantes fiscales de la facture énergétique augmentent quand on s'éloigne des communes-centres. En outre, la part des différentes taxes dans la facture totale est légèrement croissante quand on s'éloigne des communes-centres. Cela vient du fait que la part de la composante transport dans la facture énergétique augmente quand on s'éloigne des communes-centres et que les carburants sont plus taxés que les énergies du logement.

Tableau 20 : décomposition de la facture énergétique par catégories de commune et entre le hors toutes taxes et la fiscalité en 2019

Catégorie de commune	Facture énergétique (en euros)							
	Facture hors toutes taxes	Fiscalité						Total
		TVA	Fiscalité sur l'électricité	Fiscalité sur les combustibles et carburant			Total	
				Composante Carbone	Hors Composante Carbone	Total		
Commune-centre d'un pôle	1 212	334	130	346	474	820	1 284	2 150
Autre commune d'un pôle	1 385	392	148	425	577	1 002	1 542	2 502
Commune d'une couronne	1 782	533	200	643	840	1 483	2 216	3 355
Commune hors d'une aire d'attraction de ville	1 789	534	202	630	825	1 455	2 191	3 350

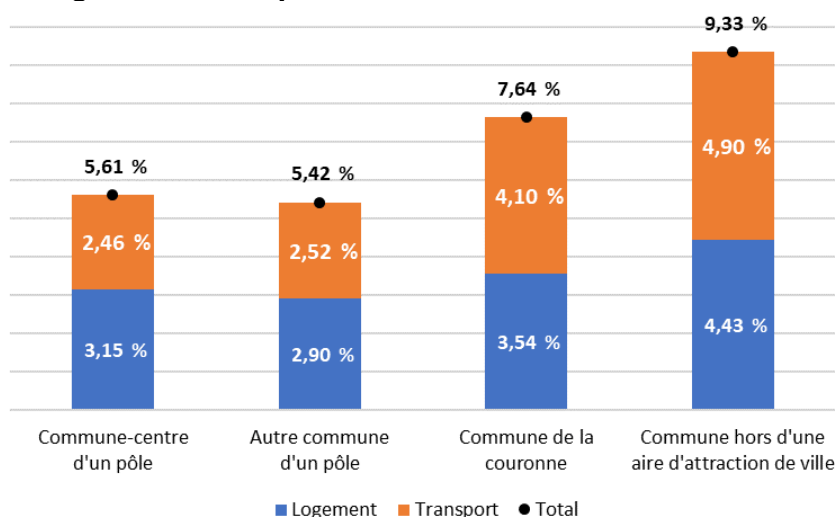
Source : CGDD, modèle Prometheus

DÉCOMPOSITION DES TAUX D'EFFORT ÉNERGÉTIQUES

Les taux d'effort énergétiques des ménages sont croissants avec l'éloignement aux communes-centres comme le montre la *figure 10*.

Les différences de taux d'effort énergétique entre les zones géographiques viennent des composantes logement et transport. Ces deux dernières composent chacune environ la moitié du taux d'effort total. Dans les pôles, la partie logement est légèrement supérieure tandis que hors des pôles c'est la partie transport qui représente la plus grosse part du taux d'effort.

Figure 10 : décomposition du taux d'effort énergétique par catégories de commune et entre les composantes logement et transport en 2019



Source : CGDD, modèle Prometheus

(c) DÉCOMPOSITION PAR ÉNERGIE

L'analyse de l'hétérogénéité des taux d'effort selon les différentes énergies et la catégorie de communes des ménages présentés dans le *tableau 21* amène aux mêmes conclusions que celle en niveaux de factures.

En part du revenu, les carburants ainsi que l'électricité représentent les plus hauts postes de dépense. Puis viennent le gaz (plutôt dans les pôles), le fioul (plutôt hors de ceux-ci) et enfin les énergies moins répandues comme le bois, le GPL et le chauffage urbain.

Tableau 21 : décomposition du taux d'effort énergétique par catégories de commune et par énergie en 2019

En %

Catégorie de commune	Gazole	Électricité	Essence	Gaz	Fioul	GPL	Bois	Urbain	GNV	Autres	Total
Commune-centre d'un pôle	1,72	1,67	0,74	1,11	0,19	0,04	0,03	0,09	0,00	0,02	5,61
Autre commune d'un pôle	1,76	1,57	0,77	1,01	0,17	0,03	0,02	0,08	0,00	0,01	5,42
Commune d'une couronne	3,21	2,21	0,87	0,55	0,50	0,15	0,11	0,00	0,01	0,03	7,64
Commune hors d'une aire d'attraction de ville	3,98	2,72	0,92	0,13	1,06	0,29	0,20	0,00	0,01	0,02	9,33

Source : CGDD, modèle Prometheus

(d) DÉCOMPOSITION PAR DIMENSIONS FISCALES

Le détail par composante de taxation est présenté dans le *tableau 22*.

La fiscalité combustibles et carburants est le premier poste de dépense fiscal. Puis vient la TVA, ensuite la composante carbone et enfin la fiscalité sur l'électricité. Au total, environ 45 % du taux d'effort énergétique d'un ménage quelle que soit sa commune de résidence, vient des différentes taxes sur l'énergie.

Tableau 22 : décomposition du taux d'effort énergétique par catégories de commune et entre le hors toutes taxes et la fiscalité en 2019

Catégorie de commune	Facture énergétique (en %)							
	Facture hors toutes taxes	Fiscalité						Total
		TVA	Fiscalité sur l'électricité	Fiscalité sur les combustibles et carburant			Total	
				Composante Carbone	Hors Composante Carbone	Total		
Commune-centre d'un pôle	3,16	0,87	0,34	0,34	0,90	1,24	2,45	5,61
Autre commune d'un pôle	3,00	0,85	0,32	0,33	0,92	1,25	2,42	5,42
Commune d'une couronne	4,06	1,21	0,46	0,45	1,46	1,91	3,58	7,64
Commune hors d'une aire d'attraction de ville	4,98	1,49	0,57	0,54	1,75	2,29	4,35	9,33

Source : CGDD, modèle Prometheus

B. Hausse du prix TTC des énergies

Dans cette partie, l'impact sur les factures et taux d'effort énergétiques de hausses du prix TTC de 10 % des trois principales énergies (carburants routiers, gaz et électricité) est présenté selon trois scénarios distincts de hausse de prix. Il s'agit d'une hausse de la part variable des prix de l'énergie uniquement (donc hors abonnements). La situation de référence est celle de 2019, précédemment décrite.

Hausse des prix TTC des carburants de 10 %

La première hausse concerne les prix des carburants routiers. Sont touchés tous les ménages qui utilisent un véhicule thermique, soit 82 % des ménages. Cette hausse correspond à une hausse de 10 % de l'ensemble des prix TTC des carburants à la pompe (diesel, SP98, SP95, SP95E10). Cela augmente la facture transport de l'ensemble des ménages de 10 % de façon homogène, les quantités d'énergie consommées étant supposées inchangées.

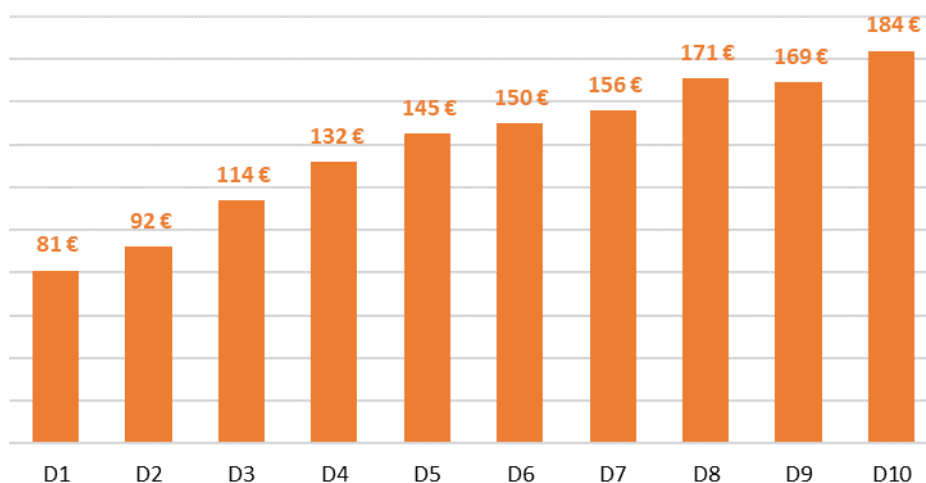
En niveau, la facture de transport augmente de 139 € en moyenne, soit une hausse du taux d'effort énergétique moyen de 0,33 point de pourcentage. Pour les ménages qui disposent d'au moins un véhicule, la facture transport augmente en moyenne de 170 €.

Cette hausse, exprimée en euros, est croissante avec le niveau de vie, allant de 81 € pour le premier dixième à plus de 184 € pour le dernier (*figure 11a*). Mais elle est décroissante,

lorsqu'elle est exprimée en part de revenus, avec le niveau de vie, représentant 0,67 % du revenu pour le premier dixième, contre seulement 0,17 % de celui du dernier (figure 11b).

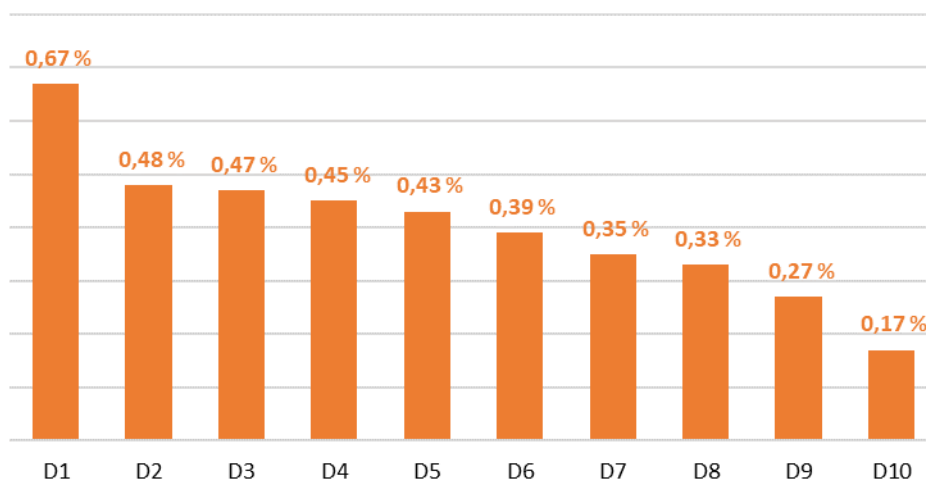
Cette hausse est inégale en fonction des zones géographiques. Elle va de 94 € en commune-centre, 116 € dans les autres communes du pôle, et jusqu'à environ 180 € hors des pôles (figure 12a). Cette augmentation ne représente que 0,25 point de pourcentage de hausse du taux d'effort chez les ménages des communes-centres, contre 0,49 point de pourcentage pour les ménages résidant hors d'une aire d'attraction de ville (figure 12b), soit une part plus de deux fois supérieure.

Figure 11a : écart de facture énergétique par dixième de niveau de vie



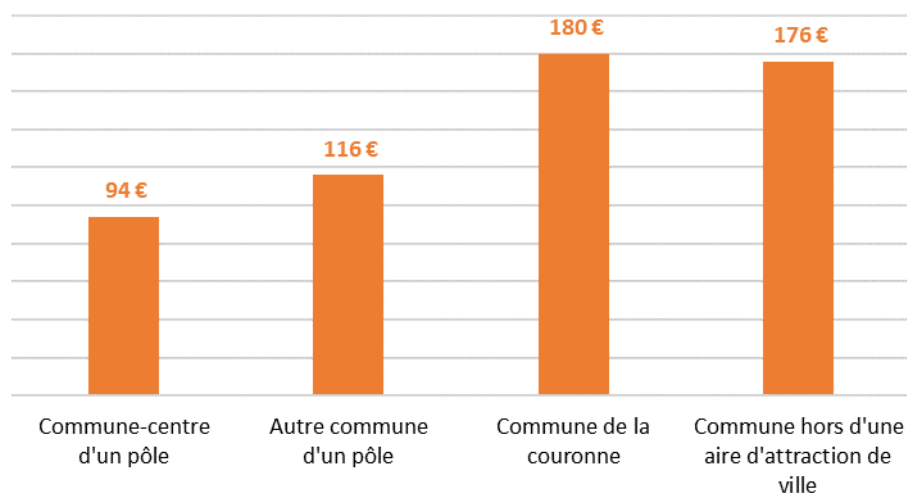
Source : CGDD, modèle Prometheus

Figure 11b : écart de taux d'effort énergétique par dixièmes de niveau de vie



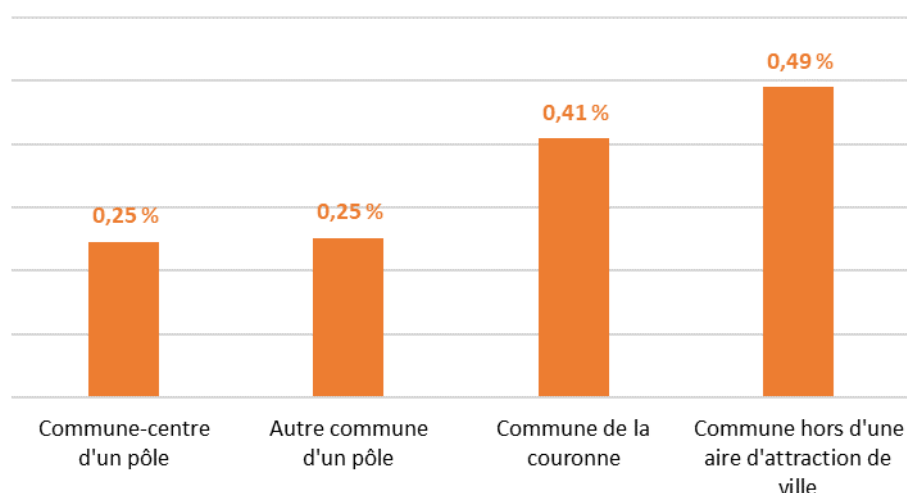
Source : CGDD, modèle Prometheus

Figure 12a : écart de facture énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Figure 12b : écart de taux d'effort énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Hausse des prix TTC du gaz de 10 %

La deuxième hausse considérée concerne le prix du gaz, et plus précisément la part variable des contrats de gaz (donc hors abonnement)¹⁵⁶. Cette hausse ne concerne pas tous les ménages. En effet, 39 % des ménages en France métropolitaine utilisent cette énergie comme énergie principale de chauffage (voir annexe 16) et ce sont plutôt les ménages urbains qui utilisent cette source d'énergie pour se chauffer (voir partie B).

En moyenne, une hausse des prix de gaz de 10 % entraîne une hausse de la facture totale moyenne de 27 €, soit une hausse du taux d'effort énergétique de 0,06 point de pourcentage. Pour les ménages dont l'énergie principale est le gaz, cette hausse de prix du gaz entraîne une augmentation de la facture énergétique totale de 68 €.

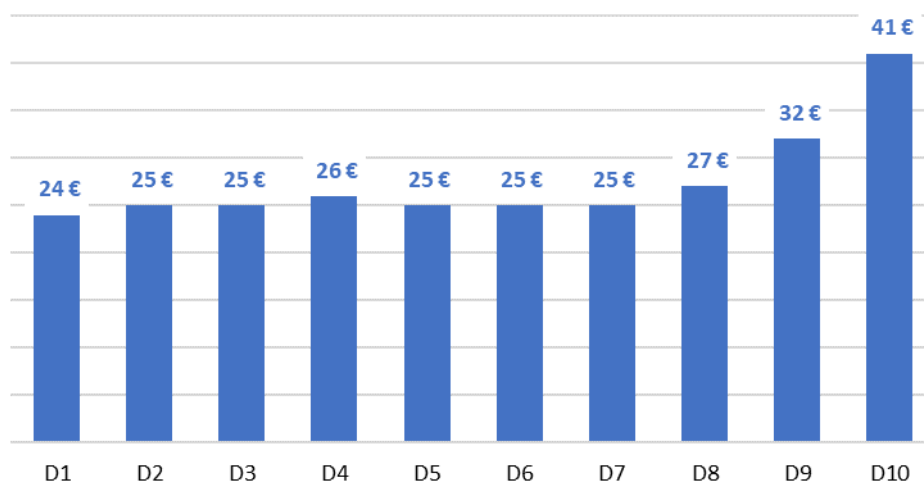
Cette hausse est quasi équivalente, en euros, pour les huit premiers dixièmes de niveau de vie, à près de 25 €, puis est plus conséquente pour les deux derniers dixièmes, allant jusqu'à

¹⁵⁶ Le gaz considéré dans ce contexte est le gaz de réseau. Les bouteilles de propane et de butane ne sont pas considérées dans cette catégorie de même que le GPL utilisé à usage de carburant.

presque 41 € pour le dernier (figure 13a). Exprimée en part du revenu, cette hausse est décroissante avec le niveau de vie, allant de 0,20 % du revenu à 0,04 % (figure 13b).

Cette hausse concerne principalement les habitants des pôles, où le gaz est plus répandu (graphique 14a) et où elle conduit en moyenne à une hausse de la facture énergétique entre 35 € (communes-centres) et 39 € (autres communes des pôles) et à une élévation du taux d'effort énergétique d'un peu moins de 0,1 point de pourcentage. En dehors des pôles, les habitants des couronnes sont moins concernés, en moyenne leur facture augmente de 19 €, et ceux hors attraction des villes ne sont presque pas touchés.

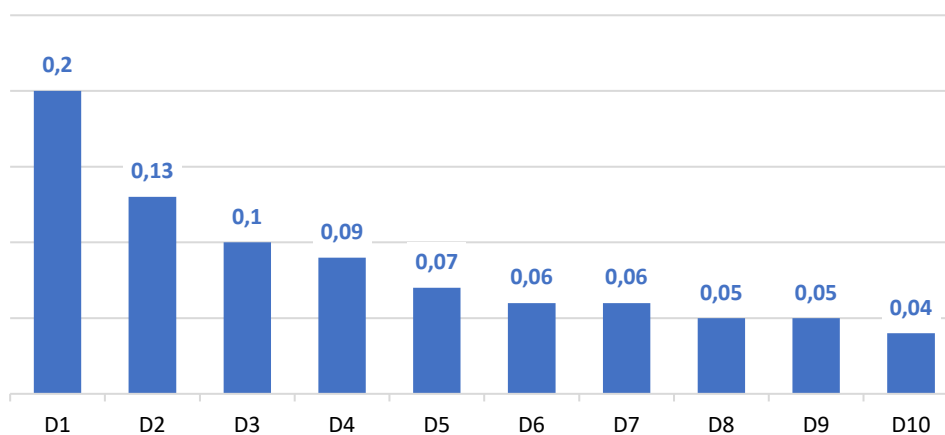
Figure 13a : écart de facture par dixièmes de niveau de vie



Source : CGDD, modèle Prometheus

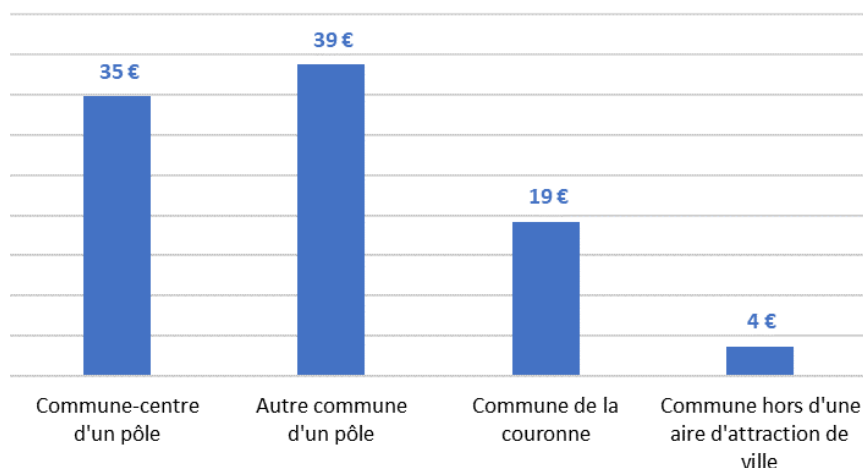
Figure 13b : écart de taux d'effort énergétique par dixièmes de niveau de vie

En %



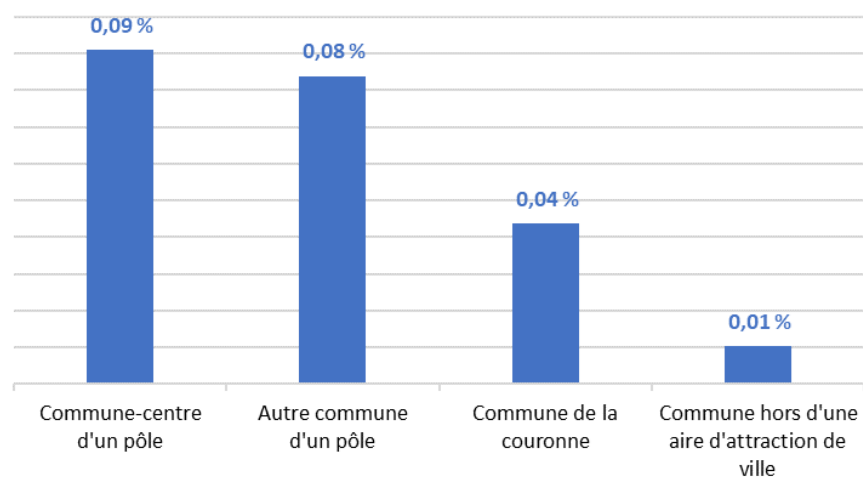
Source : CGDD, modèle Prometheus

Figure 14a : écart de facture énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Figure 14b : écart de taux d'effort énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Hausse des prix TTC de l'électricité de 10 %

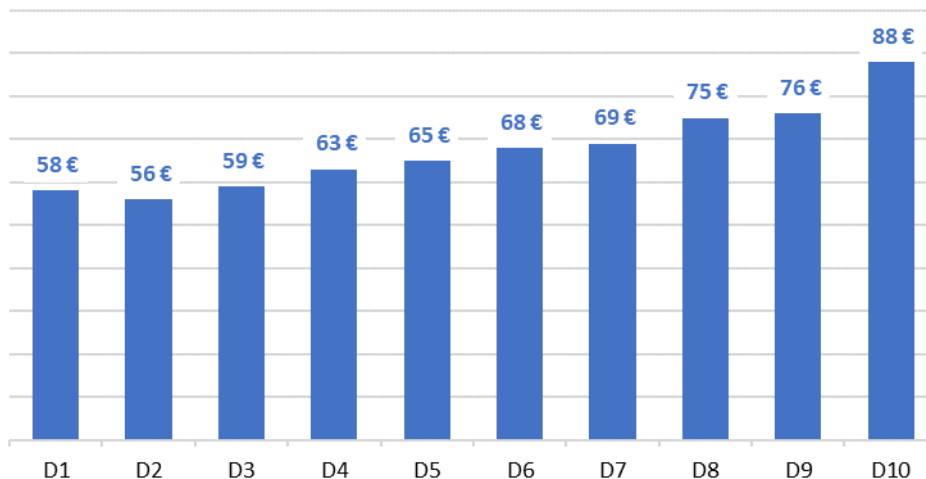
La troisième hausse considérée concerne le prix de l'électricité, là aussi sur la part variable des contrats (donc hors abonnement). Cette hausse touche l'ensemble des ménages, car, même si ce n'est l'énergie principale de chauffage que de 35 % des ménages (voir annexe 16), soit la deuxième énergie de chauffage la plus répandue, tous utilisent de l'électricité pour s'éclairer ou pour leur électroménager.

Une hausse du prix de l'électricité entraîne une hausse de la facture énergétique de 68 € en moyenne, soit une hausse du taux d'effort énergétique de 0,16 point de pourcentage du revenu. Pour les ménages qui utilisent l'électricité comme énergie principale de chauffage, la hausse de facture totale correspondante à cette hausse de prix est de 97 € en moyenne tandis qu'elle est de 52 € en moyenne pour les ménages qui n'utilisent pas l'électricité comme énergie principale de chauffage.

Cette hausse est croissante, en euros, avec le niveau de vie, allant de 58 € pour le premier dixième à 88 € pour le dernier (figure 15a), mais décroissante, en part de revenus, avec le niveau de vie, représentant 0,49 % du revenu pour le premier dixième, contre seulement 0,08 % de celui du dernier (figure 15b).

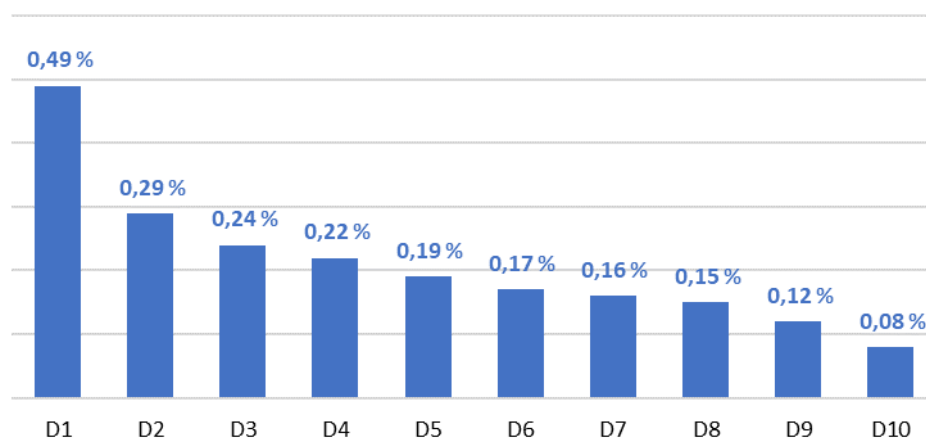
Cette hausse est également croissante, en niveau, en s'éloignant des communes-centres. Allant de 51 € pour les habitants de celles-ci, à enfin 82 € en dehors des pôles (figure 16a). Elle est également croissante, en part de revenu, en s'éloignant des communes-centres, allant de 0,13 point de pourcentage dans les pôles à 0,23 point de pourcentage dans les communes situées hors d'une aire d'attraction de ville (figure 16b).

Figure 15a : écart de facture par dixièmes de niveau de vie



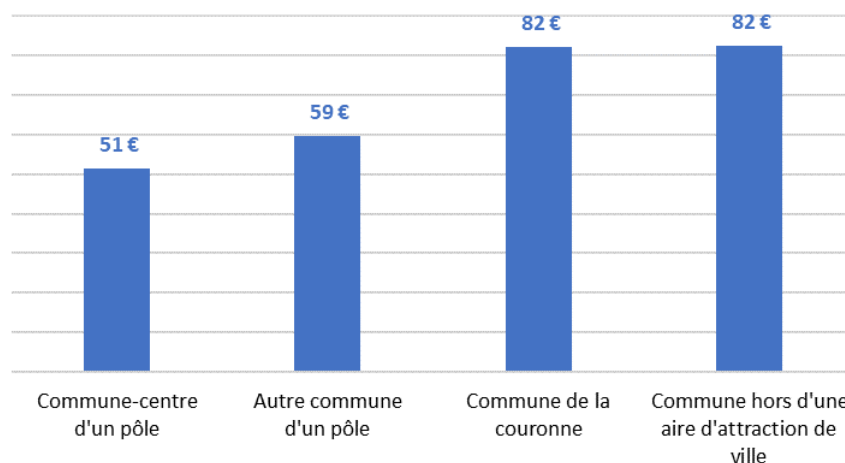
Source : CGDD, modèle Prometheus

Figure 15b : écart de taux d'effort énergétique par dixièmes de niveau de vie



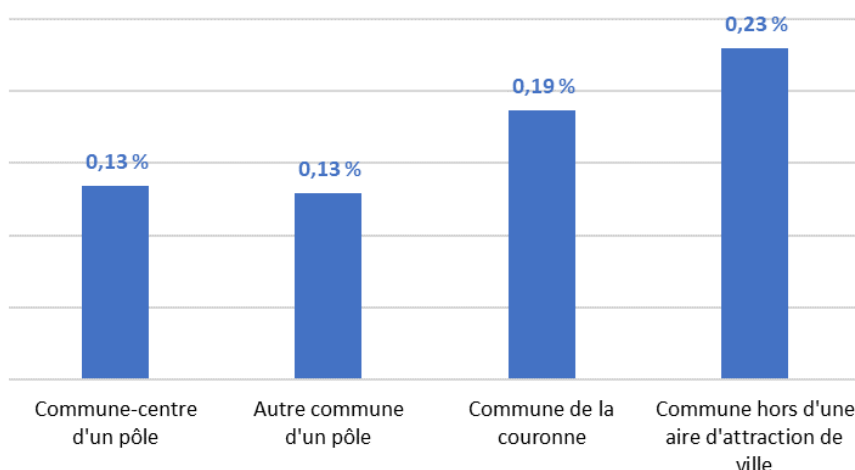
Source : CGDD, modèle Prometheus

Figure 16a : écart de facture par catégories de commune



Source : CGDD, modèle Prometheus

Figure 16b : écart de taux d'effort énergétique par catégories de commune



Source : CGDD, modèle Prometheus

C. Hausse de fiscalité énergétique

Dans cette troisième partie, trois scénarios de hausses de la fiscalité énergétique sont présentés : composante carbone, droit d'accise sur l'électricité, et TVA sur les prix de l'énergie. Pour rappel, la fiscalité représente en moyenne près de la moitié de la facture énergétique des ménages.

Hausse de la composante carbone de 10 €/tCO₂

La « composante carbone », intégrée en 2014 dans la fiscalité des énergies fossiles, est une partie de l'accise sur les énergies fossiles (gaz, fioul, gazole, essence), proportionnelle à leurs émissions de CO₂ (voir chapitre VII). Sa valeur est fixée depuis 2018 à 44,6 €/tCO₂. Une augmentation de 10 €/tCO₂ de cette taxe entraîne une hausse de la facture énergétique pour la plupart des ménages. Les ménages non concernés sont ceux qui ne possèdent pas de voiture à moteur thermique et qui n'utilisent dans leur logement ni gaz, ni fioul.

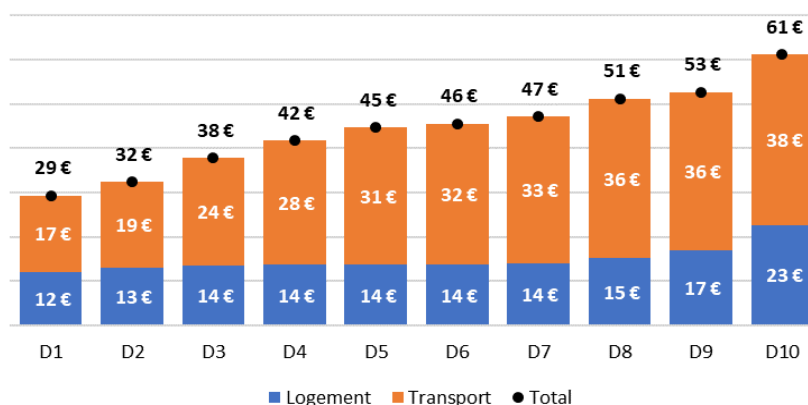
Une hausse de 10 €/tCO₂ de la composante carbone entraîne en moyenne une hausse de la facture énergétique totale de 44 €, soit 0,11 % du revenu en moyenne. Cette hausse est portée

à hauteur de 15 € par la composante logement de la facture et de 29 € par la composante transport. Cette hausse augmente les recettes de l'État de 1,3 Md€¹⁵⁷.

Cette hausse est croissante, en euros, avec le niveau de vie, allant de 29 € pour le premier dixième à 61 € pour le dernier (figure 17a), mais décroissante, en part de revenus, avec le niveau de vie, représentant 0,24 % du revenu des ménages du premier dixième, contre seulement 0,06 % de celui du dernier (figure 17b). Pour tous les dixièmes, la hausse de facture est portée plus particulièrement par la composante transport. Pour le premier dixième, cette composante représente une hausse de 17 €, contre 12 € pour la composante logement. Pour le dernier dixième, cette composante représente une hausse de 38 €, contre 23 € pour la composante logement (figure 17a).

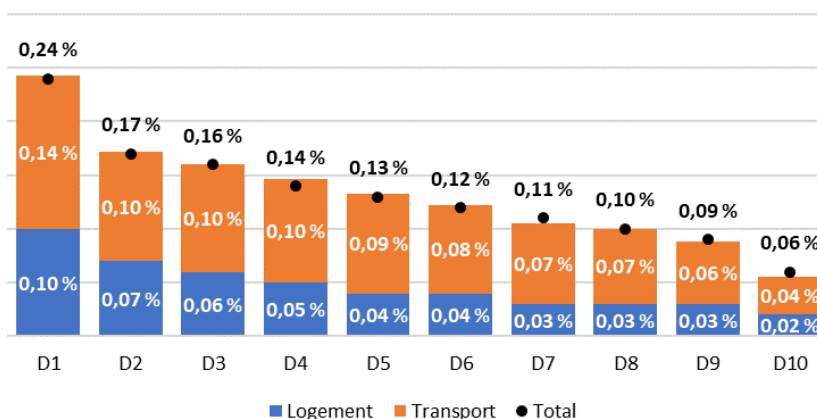
Cette hausse est croissante, en euros, en s'éloignant des communes-centres. Elle s'étale de 35 € en moyenne pour les habitants de celles-ci, à 41 € pour les autres communes du pôle, à enfin plus de 52 € en dehors des pôles (figure 18a). Elle est également croissante, en part de revenu, en s'éloignant des communes-centres, allant de 0,09 point de pourcentage dans les pôles à 0,14 point de pourcentage pour les communes hors d'une aire d'attraction de ville (figure 18b). À nouveau, cette hausse est surtout portée par la composante transport de la facture. En particulier pour les habitants hors des pôles pour qui la composante transport représente une hausse de 38 €, contre 14 € pour la composante logement.

Figure 17a : écart de facture énergétique par dixièmes de niveau de vie



Source : CGDD, modèle Prometheus

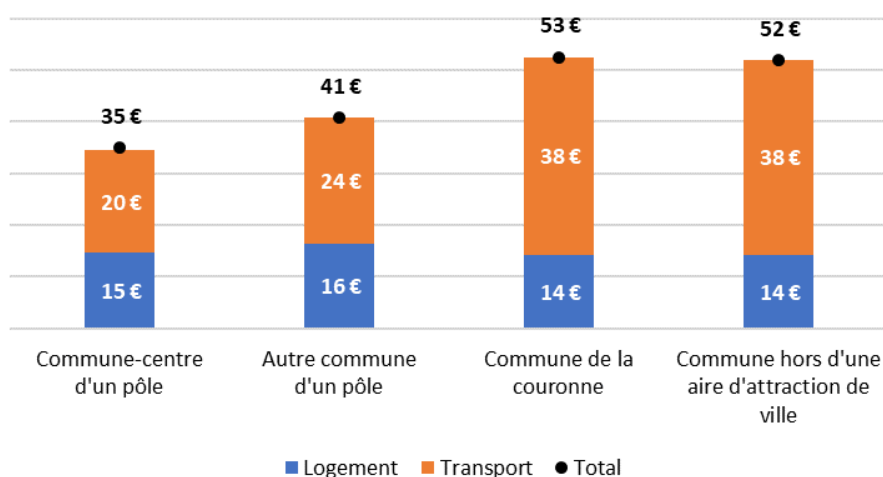
Figure 17b : écart de taux d'effort énergétique par dixièmes de niveau de vie



Source : CGDD, modèle Prometheus

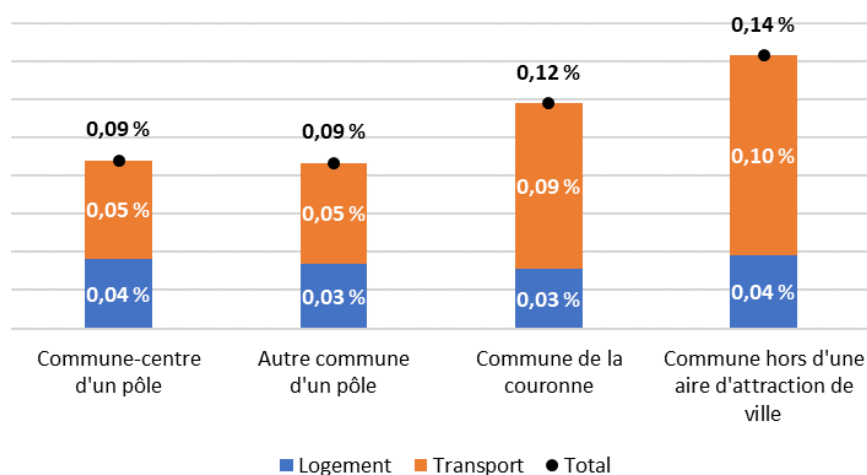
¹⁵⁷ Le champ est celui des ménages considérés dans le modèle Prometheus, c'est-à-dire celui des ménages résidant en France métropolitaine. Par ailleurs, s'agissant des consommations d'énergie du logement, seules celles relatives à la résidence principale sont prises en compte.

Figure 18a : écart de facture énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Figure 18b : écart de taux d'effort énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Hausse de l'accise sur l'électricité de 10 €/MWh

La fraction électricité est la partie de l'accise sur les énergies qui concerne l'électricité. En 2019, cette fraction n'existait pas encore sous ce nom. Trois taxes d'accise sur l'électricité existaient alors : une taxe nationale (la TICFE), dont le taux s'élevait à 22,5 €/MWh, et deux taxes locales (les TLCFE), dont le taux moyen, pour ces deux taxes prises ensemble, est estimé à 9,37 €/MWh.

Une hausse de la fiscalité d'accise sur l'électricité touche tous les ménages, car tous utilisent cette énergie dans leur quotidien de plus son impact est très similaire à celui de la hausse de 10 % du prix TTC présenté dans le chapitre IX.B.

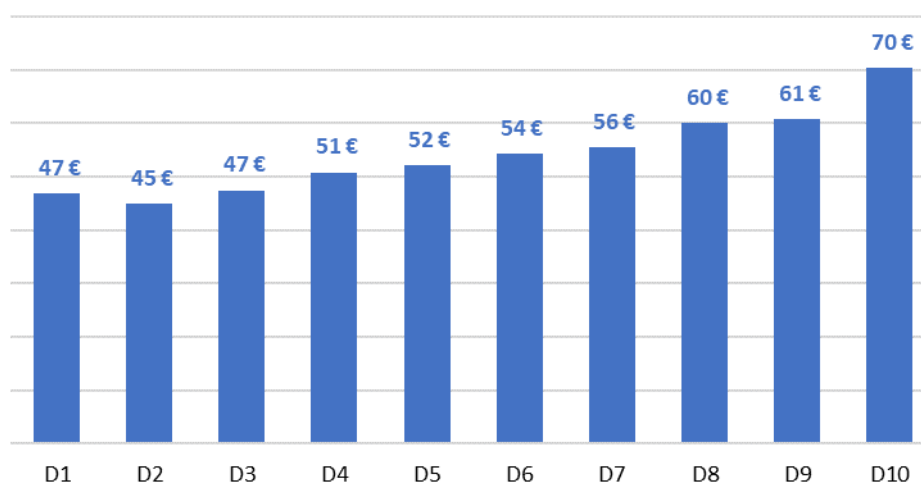
Une hausse de la fraction électricité de 10 €/MWh entraîne en moyenne une hausse de la facture énergétique totale de 54 €, soit 0,13 % du revenu en moyenne. Cette hausse de la fiscalité entraîne une hausse des recettes de l'État de 1,6 Md€¹⁵⁸.

¹⁵⁸ Le champ est celui des ménages considérés dans le modèle Prometheus, c'est-à-dire celui des ménages résidant en France métropolitaine. Par ailleurs, s'agissant des consommations d'énergie du logement, seules celles relatives à la résidence principale sont prises en compte.

Cette hausse est croissante, en niveau, avec le niveau de vie, allant de 47 € pour le premier dixième à 70 € pour le dernier (figure 19a), mais décroissante, en part de revenus, avec le niveau de vie, représentant 0,39 % du revenu des ménages du premier dixième, contre seulement 0,07 % de celui des ménages du dernier (figure 19b).

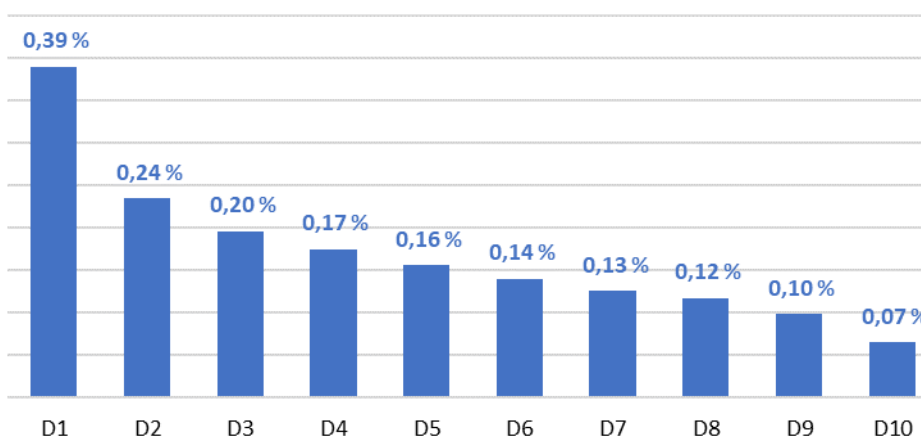
Cette hausse est croissante, en niveau, en s'éloignant des communes-centres. Allant de 41 € pour les habitants de celles-ci, à 48 € pour les autres communes du pôle, à enfin plus de 66 € en dehors des pôles (figure 20a). Elle est également croissante, en part de revenu, en s'éloignant des communes-centres, allant de 0,10 point de pourcentage dans les pôles à 0,18 point de pourcentage pour les communes situées hors d'une aire d'attraction de ville (figure 20b).

Figure 19a : écart de facture énergétique par dixièmes de niveau de vie



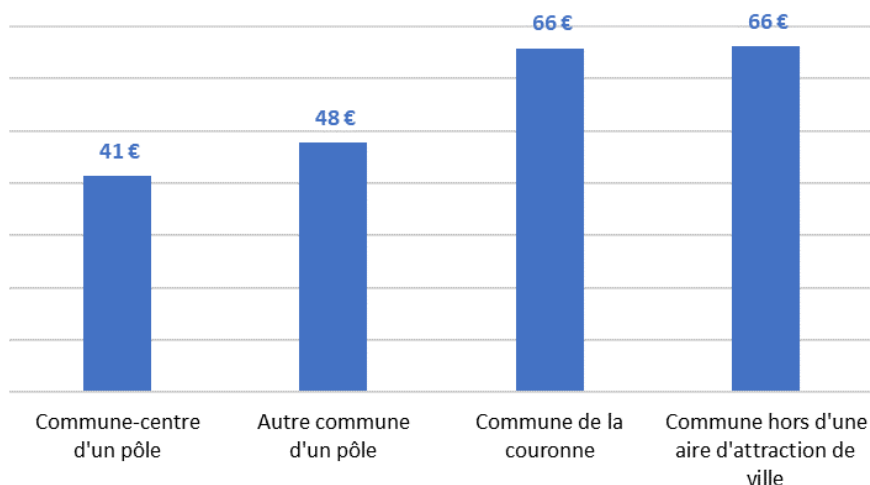
Source : CGDD, modèle Prometheus

Figure 19b : écart de taux d'effort énergétique par dixièmes de niveau de vie



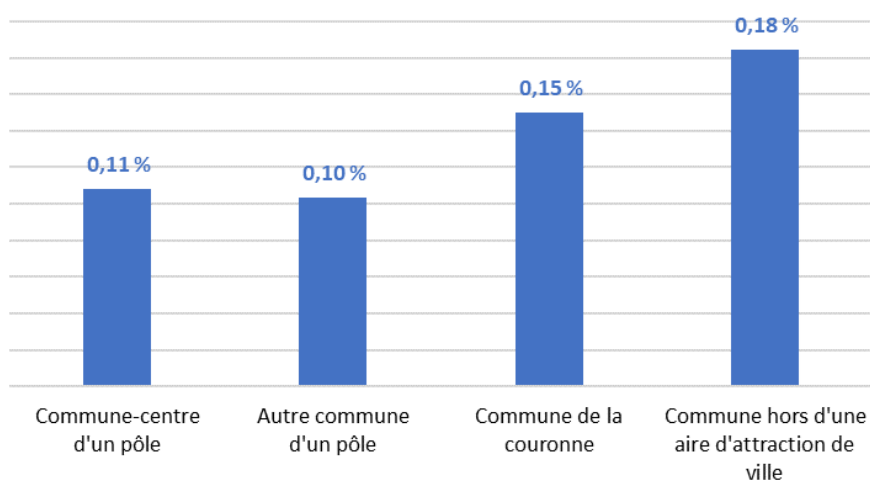
Source : CGDD, modèle Prometheus

Figure 20a : écart de facture énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Figure 20b : écart de taux d'effort énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Hausse de la TVA sur les prix de l'énergie d'un point de pourcentage

La taxe sur la valeur ajoutée (TVA) s'applique sur l'ensemble des prix de l'énergie, qu'il s'agisse des carburants ou des énergies du logement. Son taux plein s'élève à 20 % du prix hors TVA. Pour rappel, il existe des taux réduits de 10 %, s'appliquant sur le prix du bois, et de 5,5 % s'appliquant des abonnements (*voir chapitre VII*). Dans cette variante, seul le taux plein de TVA voit sa valeur modifiée, passant de 20 % à 21 %.

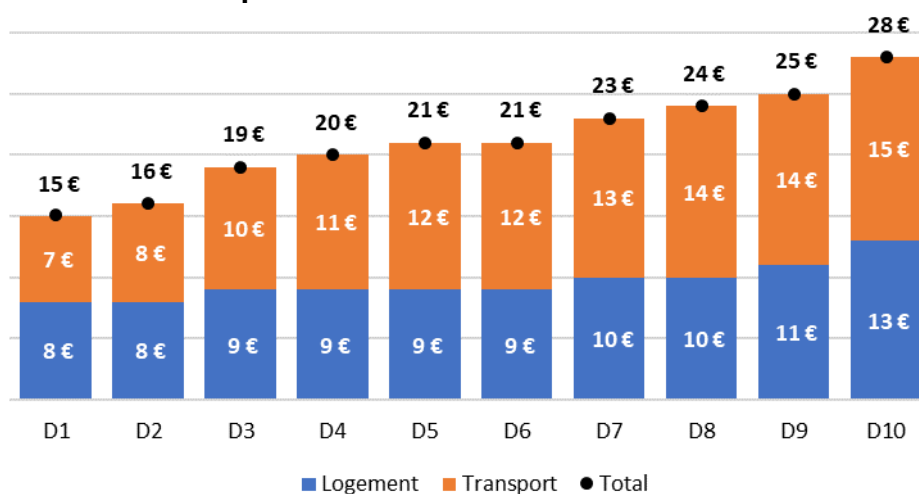
La hausse de la TVA d'un point de pourcentage entraîne en moyenne une hausse de la facture énergétique totale de 21 €, soit 0,05 % du revenu en moyenne. Cette hausse est portée à hauteur de 10 € par la composante logement de la facture énergétique et de 11 € par la composante transport. Cette hausse de la fiscalité entraîne une hausse des recettes de l'État de 0,6 Md€. ¹⁵⁹

¹⁵⁹ Le champ est celui des ménages considérés dans le modèle Prometheus, c'est-à-dire celui des ménages résidant en France métropolitaine. Par ailleurs, s'agissant des consommations d'énergie du logement, seules celles relatives à la résidence principale sont prises en compte.

Cette hausse est croissante, en euros, avec le niveau de vie, allant de 15 € pour le premier dixième à 28 € pour le dernier (figure 21a), mais décroissante, en part du revenu, avec le niveau de vie, représentant 0,12 % du revenu pour le premier dixième, contre 0,03 % de celui du dernier (figure 21b). Pour tous les dixièmes de niveau de vie, la hausse est portée légèrement davantage par la composante transport et l'écart entre les composantes logement et transport croit avec le niveau de vie.

Cette hausse est croissante, en euros, en s'éloignant des communes-centres. Allant de 16 € pour les habitants de celles-ci, à 19 € pour les autres communes du pôle, à enfin plus de 26 € en dehors des pôles (figure 22a). Elle est également croissante, en part de revenu, en s'éloignant des communes-centres, allant de 0,04 point de pourcentage dans les pôles à 0,07 point de pourcentage pour les communes hors d'une aire d'attraction de ville (figure 22b). Ces hausses sont réparties assez équitablement entre composantes logement et transport de la facture au sein des pôles, tandis que hors des pôles, c'est la partie transport qui augmente le plus avec cette hausse de la fiscalité.

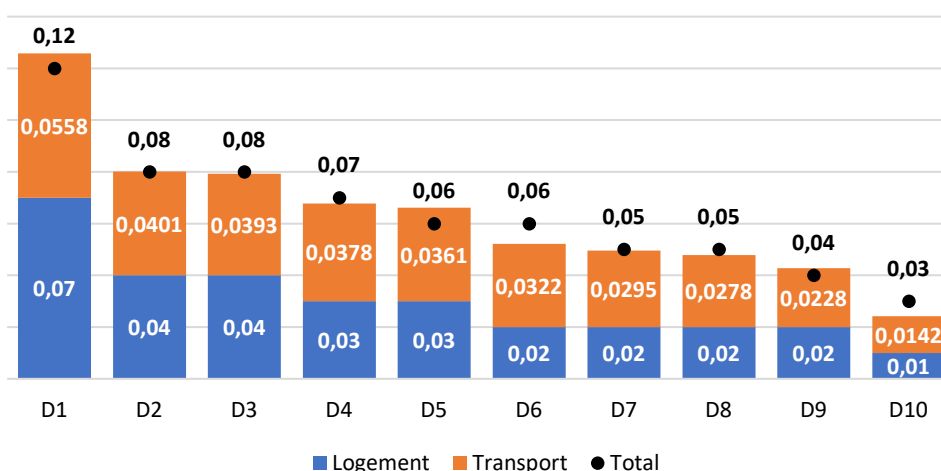
Figure 21a : écart de facture par dixièmes de niveau de vie



Source : CGDD, modèle Prometheus

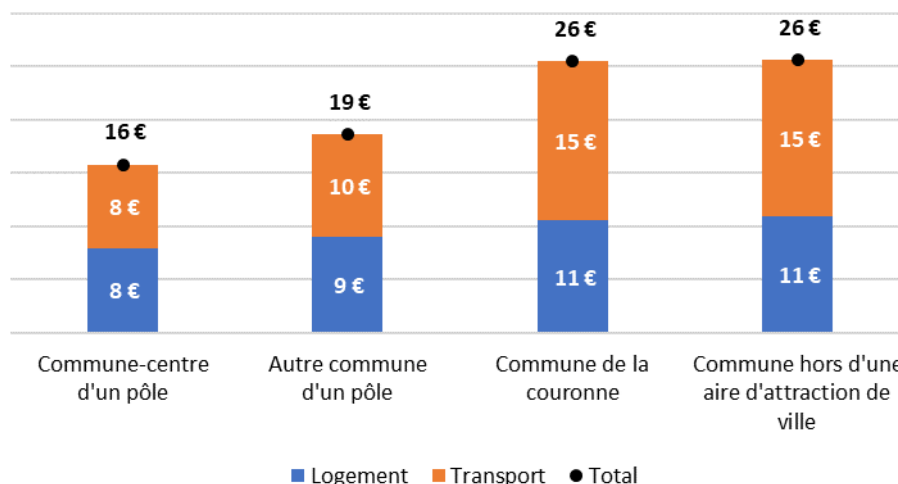
Figure 21b : écart de taux d'effort énergétique par dixièmes de niveau de vie

En %



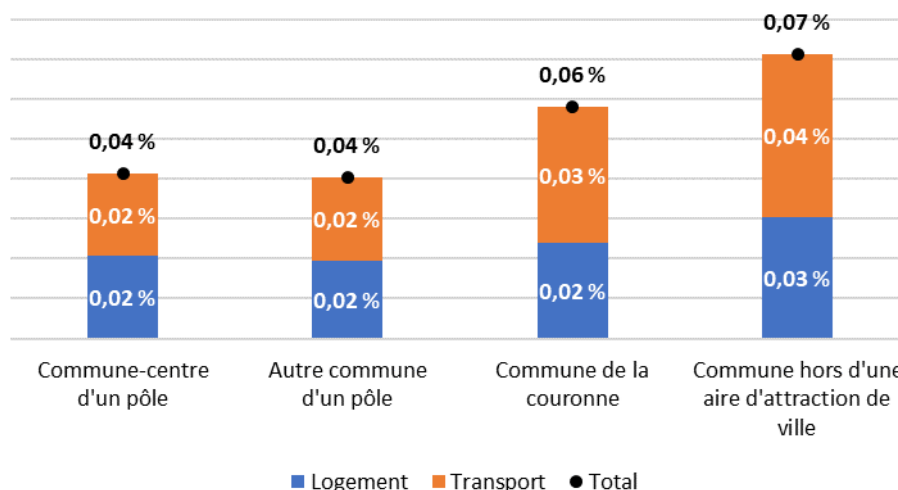
Source : CGDD, modèle Prometheus

Figure 22a : écart de facture énergétique par catégories de commune



Source : CGDD, modèle Prometheus

Figure 22b : écart de taux d'effort énergétique par catégories de commune



Source : CGDD, modèle Prometheus

D. Aides au paiement de la facture énergétique

Cette dernière partie présente l'impact, sur la facture énergétique des ménages, de dispositifs d'aides tels que le chèque énergie. Ces dispositifs permettent de réduire la facture énergétique des ménages. La situation de référence est, comme précédemment, celle de 2019.

Le chèque énergie est un dispositif permettant d'aider les ménages modestes à payer des dépenses d'énergie liées au logement¹⁶⁰, et ainsi de réduire la précarité énergétique¹⁶¹. Il est versé chaque année depuis 2018, et a connu plusieurs évolutions depuis lors.

¹⁶⁰ Le ménage peut également choisir d'utiliser le chèque énergie pour des travaux de rénovation énergétique, mais cet usage est rare.

¹⁶¹ Pour une analyse des effets du chèque énergie sur la précarité énergétique en 2019, voir Parent et al. (2021).

Deux versions du chèque énergie sont considérées : le chèque énergie dans sa version de 2019 (dit dans la suite « chèque énergie classique ») et une version étendue, dans laquelle chaque ménage éligible au chèque énergie en 2019 reçoit, en plus du chèque énergie classique, un chèque énergie additionnel de 100 €. Ce versement additionnel correspond à ce qui a pu être mis en place fin 2021 et fin 2022 dans le contexte du fort renchérissement des énergies.

Les factures énergétiques et les taux d'effort énergétiques présentés ci-après sont calculés nets du chèque énergie. Le taux de recours et d'utilisation du chèque énergie est supposé par simplicité égal à 100 %¹⁶².

Chèque énergie classique

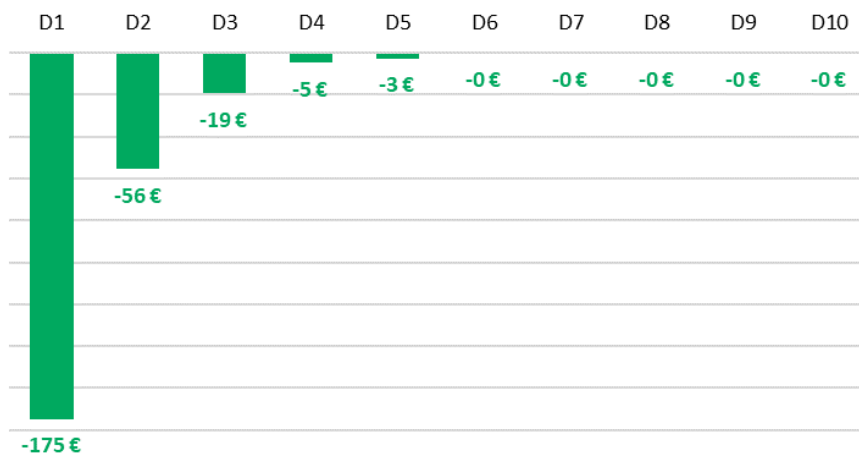
Le chèque énergie simulé dans ce cas est celui de l'année de référence 2019¹⁶³. Son montant varie de 48 € à 277 €, en fonction du revenu fiscal de référence et de la composition du ménage, définie en unité de consommation (UC). Le montant moyen pour un ménage éligible¹⁶⁴ est de 147 €.

En moyenne, le chèque énergie réduit la facture énergétique des ménages de 29 €, soit une baisse de 0,07 % en part du revenu. Son coût budgétaire est de 0,8 Md€¹⁶⁵.

Cette baisse se concentre sur les ménages des premiers dixièmes de niveau de vie. Les trois premiers dixièmes voient leur facture logement baisser respectivement de 175 €, 56 €, et 19 € (figure 23a). Cela représente des taux d'effort énergétique respectifs de 1,45, 0,29, 0,08 point de pourcentage (figure 23b).

Une analyse sous l'angle géographique montre que les habitants des communes-centres bénéficient davantage du chèque, ils touchent en moyenne 35 €, tandis que ceux vivant en dehors des communes-centres touchent en moyenne entre 20 € et 25 € (figure 24a).

Figure 23a : écart de facture énergétique nette du chèque énergie classique par dixièmes de niveau de vie



Source : CGDD, modèle Prometheus

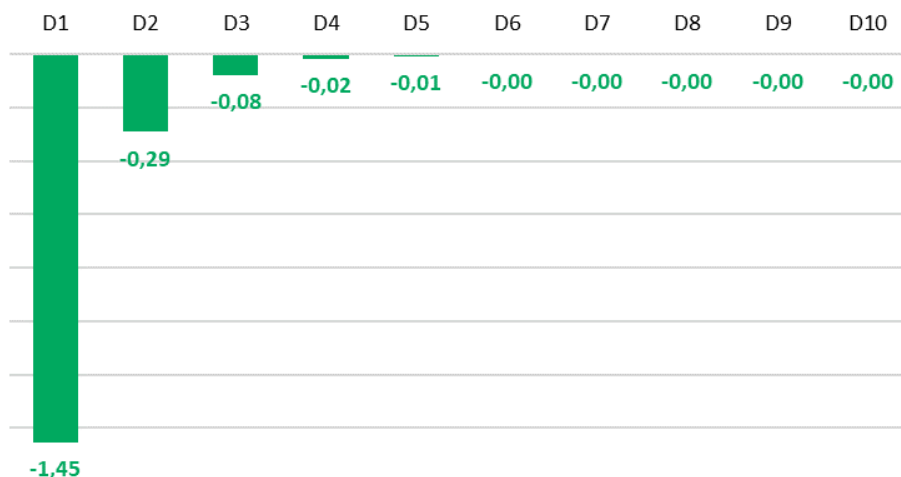
¹⁶² Le taux d'utilisation réel est usuellement compris entre 80 % et 90 %.

¹⁶³ Pour une description du barème du chèque énergie classique, voir également Parent et al. (2021).

¹⁶⁴ Ces ménages éligibles sont estimés à 5,2 millions dans le modèle, qui porte sur la France métropolitaine. Des écarts de périmètre géographique (le chèque énergie concerne aussi les départements d'outre-mer) et des raisons méthodologiques expliquent que ces estimations soient légèrement différentes du nombre de chèques réellement émis, qui est de 5,7 millions pour la France en 2019.

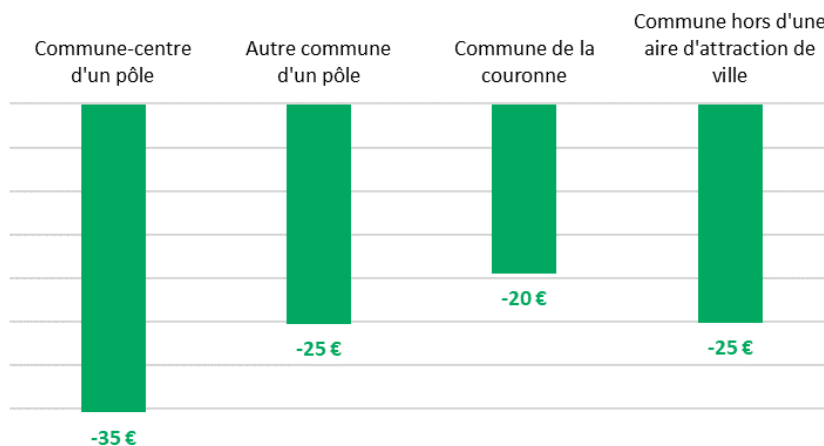
¹⁶⁵ Le champ est celui des ménages considérés dans le modèle Prometheus, c'est-à-dire celui des ménages résidant en France métropolitaine. Par ailleurs, s'agissant des consommations d'énergie du logement, seules celles relatives à la résidence principale sont prises en compte.

Figure 23b : écart de taux d'effort énergétique net du chèque énergie classique par dixièmes de niveau de vie



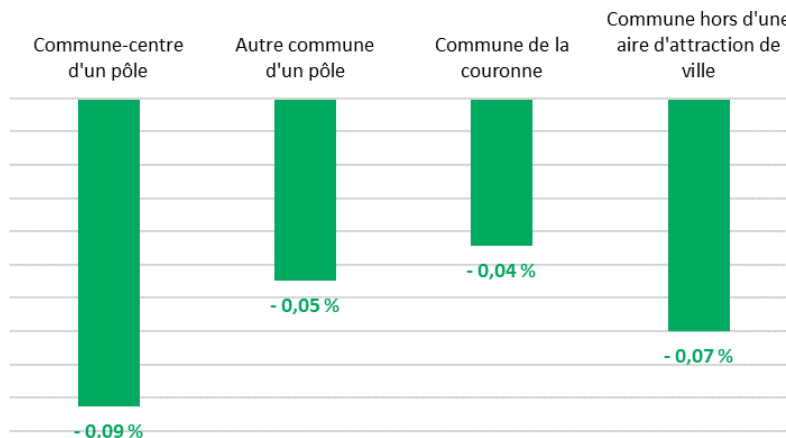
Source : CGDD, modèle Prometheus

Figure 24a : écart de facture énergétique nette du chèque énergie classique par catégories de commune



Source : CGDD, modèle Prometheus

Figure 24b : écart de taux d'effort énergétique net du chèque énergie classique par catégorie de commune



Source : CGDD, modèle Prometheus

Chèque énergie classique augmenté d'un chèque additionnel de 100 €

Une variante du chèque énergie classique de 2019, inspirée du chèque énergie exceptionnel versé fin 2021, est présentée dans ce contexte. Chaque ménage éligible au chèque énergie classique reçoit le chèque énergie classique et un chèque énergie exceptionnel de 100 euros. L'aide analysée est la somme de ces deux chèques, appelée pour simplifier « chèque énergie augmenté ».

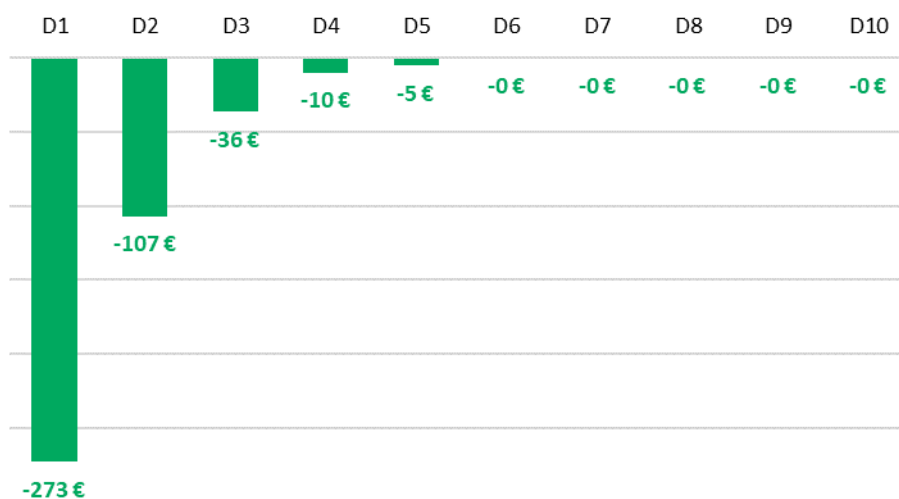
Ce dispositif permet donc de réduire davantage la facture énergétique et le taux d'effort énergétique que le chèque énergie classique. Le montant moyen de ce chèque augmenté pour un ménage éligible au dispositif est de 247 €.

En moyenne, le chèque énergie réhaussé réduit la facture énergétique des ménages de 49 €, soit 0,12 % en part de revenu. Son coût budgétaire est de 1,3 Md d'euros¹⁶⁶.

Comme pour le chèque énergie classique, cette baisse se concentre sur les ménages des premiers dixièmes de niveau de vie. Le premier dixième voit sa facture logement baisser de presque 273 €, puis 107 € pour le deuxième dixième, puis 36 € pour le troisième (*figure 25a*). Cela représente une baisse du taux d'effort énergétique allant de 2,27 points de pourcentage pour les ménages du premier dixième à 0,15 point de pourcentage pour ceux du troisième dixième (*figure 25b*).

De nouveau comme pour le chèque énergie classique, une analyse sous l'angle géographique montre que les habitants des communes-centres et des communes situées hors d'une aire d'attraction de ville bénéficient plus du chèque, ils perçoivent en moyenne environ 59 €, tandis que ceux vivant en dehors des communes-centres perçoivent entre 33 € et 45 € en moyenne (*figure 26a*).

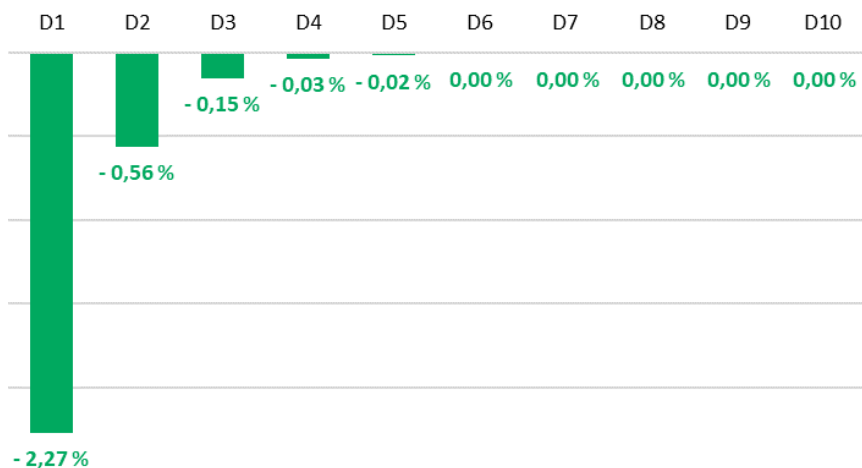
Figure 25a : écart de facture énergétique nette du chèque énergie réhaussé par dixièmes de niveau de vie



Source : CGDD, modèle Prometheus

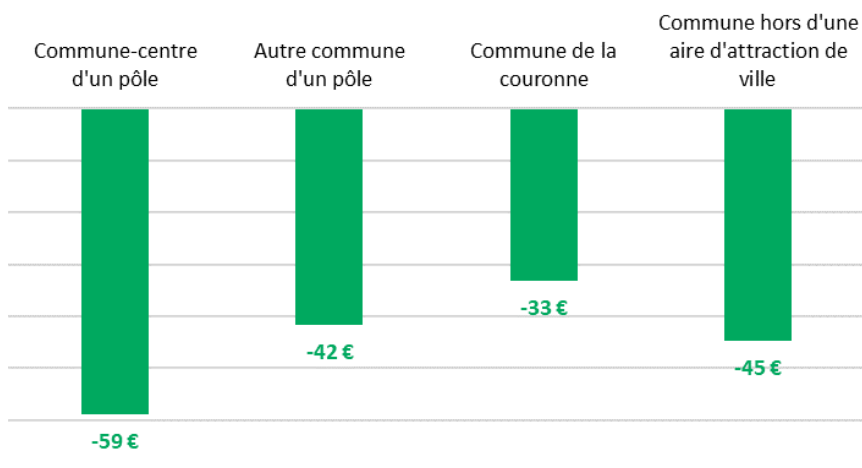
¹⁶⁶ Le champ est celui des ménages considérés dans le modèle Prometheus, c'est-à-dire celui des ménages résidant en France métropolitaine. Par ailleurs, s'agissant des consommations d'énergie du logement, seules celles relatives à la résidence principale sont prises en compte.

Figure 25b : écart de taux d'effort énergétique nette du chèque énergie réhaussé par dixièmes de niveau de vie



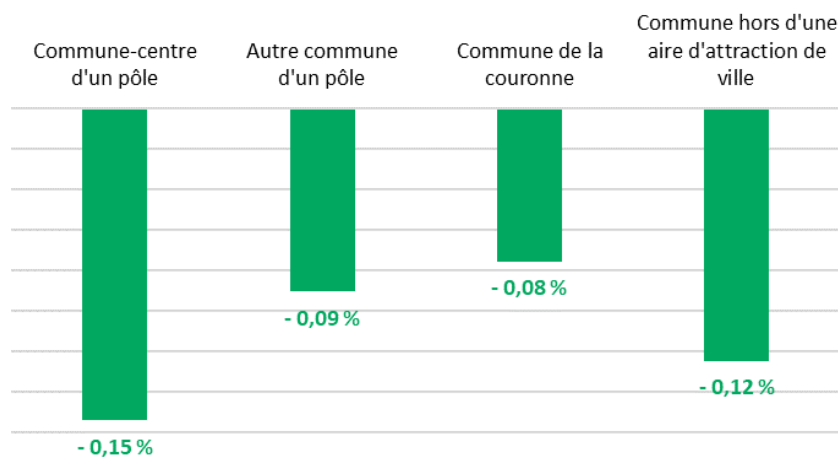
Source : CGDD, modèle Prometheus

Figure 26a : écart de facture énergétique nette du chèque énergie réhaussé par catégories de commune



Source : CGDD, modèle Prometheus

Figure 26b : écart de taux d'effort énergétique net du chèque énergie réhaussé par catégories de commune



Source : CGDD, modèle Prometheus

ANNEXES

ANNEXE 1 - DÉCOMPOSITION DE LA VARIABLE MODE DE CHAUFFAGE EN 17 MODALITÉS

KMOD_P		KMODPRIM		KEMETTER		KEMETTERB		i
KMOD_P	Mode de chauffage principal	KMODPRIM	Type d'installation centrale	KEMETTER	émetteurs de chaleur associés aux appareils indépendants	KEMETTERB	autre moyen de chauffage utilisé occasionnellement	
1	Chaudière individuelle ou installation centrale individuelle	_1=1	Installation centrale = chaudière					1
		_2=1	Installation centrale = pompe à chaleur					2
		_3=1	Installation centrale = solaire thermique					3
		_4=1	Installation centrale = autre installation centrale individuelle					4
2	chauffage ou chaufferie collective							5
3	chauffage mixte							6
4	chauffage tout électrique individuel							8
5	appareil(s) indépendant(s)			_1 ou _2 =1	radiateurs ou convecteurs fixes / mobiles			9
				_3 ou _4 =1	cuisinières chauffantes /poeles			10
				5 =1	cheminee a foyer ouvert			11
				6 =1	cheminee radiante			12
				7 =1	cheminee a foyer ferme			13
				8 =1	panneaux rayonnants non lumineux			14
				9 =1	appareils à accumulation			15
				10 =1	appareils de chauffage soufflant			16
				11=1	AUTRES (panneaux lumineux ou à infrarouge, etc.)		17	
6	Autre moyen de chauffage					_1=1 ou _2=1	radiateurs ou convecteurs fixes / mobiles	9
						_3=1 ou _4=1	cuisinières chauffantes /poeles	10
						5 =1	cheminee a foyer ouvert	11
						6 =1	cheminee radiante	12
						7 =1	cheminee a foyer ferme	13
						8 =1	panneaux rayonnants non lumineux	14
						9 =1	appareils à accumulation	15
						10 =1	appareils de chauffage soufflant	16
						11=1	AUTRES (panneaux lumineux ou à infrarouge, etc.)	17
7	Aucun							18

Source : CGDD, modèle Prometheus

ANNEXE 2 : ATTRIBUTION DE L'ÉNERGIE PRINCIPALE DE CHAUFFAGE

KCB_	Combustibles	Mode de chauffage associé	Observation
A	Fioul		
B	Gaz de réseau		
C	Butane, propane en bouteille		
D	Butane, propane en citerne		
E	Charbon		
F	Granulés, bois		
G	Bûches		
H	Autre végétal		
I	Electricité	i = 8,14,15,16,17	
J	Pétrole lampant		
K	Chauffage urbain		
L	Autre combustible		
M	Multi énergie		Catégorie créée
S	Solaire thermique	i = 2 et 3	Catégorie créée
Z	Pas d'énergie	i = 18	Catégorie créée

Note : la modalité i se réfère à l'Annexe 1.

Source : CGDD, modèle Prometheus

S'il y a plusieurs combustibles différents, il est considéré que l'énergie de chauffage principale est :

- le « gaz » si plusieurs combustibles différents compris dans B, C et D. Il y a une priorisation du gaz par rapport au GPL lorsque les deux sont utilisés ;
- le « bois » si plusieurs combustibles différents compris dans F et G ;
- le « multi énergie », soit M en dehors de ces deux cas particuliers.

Les regroupements de combustibles sont les suivants :

- butane et propane en bouteille (C) ou citerne (D) au sein du « GPL » ;
- les granulés, bois (F) et les bûches (G) au sein du « Bois » ;
- le charbon (E), autre végétal (H), pétrole lampant (J), autre combustible (L) et le solaire thermique (S) dans « Autre ».

ANNEXE 3 : VARIABLES ET RÉSULTATS DES RÉGRESSIONS DE L'IMPUTATION HOT DECK

Étapes de la méthode hot deck :

1. Pour chaque énergie de chauffage, une régression linéaire est estimée sur le groupe de ménages « donneurs » pour expliquer la facture en fonction d'un certain nombre de variables¹⁶⁷. Ces variables reprennent certaines caractéristiques du chauffage, du logement, et du ménage. Les variables retenues diffèrent selon les trois modèles propres à l'électricité, l'électricité spécifique et pour les autres énergies (sont présentées en Annexe 3) ;
2. Chaque régression est utilisée pour calculer pour chaque ménage (receveurs et donneurs de cette énergie de chauffage) une « facture estimée » ;
3. Pour chaque ménage receveur on identifie parmi les ménages donneurs, celui dont la facture estimée se rapproche le plus de la facture estimée de ce ménage receveur ;
4. On impute au ménage receveur la facture réelle¹⁶⁸ déclarée par le ménage donneur (facture imputée définitive).

L'écart moyen entre la facture totale redressée de l'ENL 2013 et la somme des factures d'énergies différentes redressées pour chaque ménage par Prometheus est de 131 €. Dans Prometheus on redresse les factures d'énergie de 10 631 ménages interrogés dans l'enquête.

Variables des régressions pour les redressements des factures par énergie de chauffage

Variables	Facture					
	Électricité	Électricité spécifique	Gaz	Fioul	Bois	Chauffage urbain
Constante						
Type de logement : individuel (maison) ou collectif (appartement)						
Surface du logement selon le type de chauffage : individuel ou collectif						
Nombre de pièces d'habitation						
Date de construction du logement : <1949, 1949-1974, 1975-1981, 1982-1989, 1990-1998 ou >1998						
Climatisation : oui ou non						
Isolation des murs : Moyenne ou mauvaise isolation ou bonne isolation						
Double vitrage : Absence ou présence						
Montant de l'assurance habitation : < 200€, 200€ à < 400€ ou >= 400€						
Zone climatique : Ouest et Sud ouest, Sud Est ou Reste de la France						
Redressement des revenus : Etudiants et revenus négatifs ou revenus redressés valides						
Logarithme du revenu par UC						
Présence d'une personne « au foyer » dans le ménage : oui ou non						
Âge de la personne de référence : <35 ans, 35 - 50 ans, 50 - 65 ans ou >65 ans						

Note de lecture : en vert la variable est présente, en rouge absente.

Source : CGDD, modèle Prometheus

¹⁶⁷ Ces variables sont pour beaucoup également utilisées par l'Insee dans le redressement de la facture énergétique totale opéré dans les données de l'enquête.

¹⁶⁸ Ce choix permet de ne pas déformer la distribution des factures (contrairement à une estimation de facture uniquement à partir des résultats de la régression linéaire).

Résultats des régressions pour les redressement des factures par énergie de chauffage

Variables	Électricité	Gaz	Électricité spécifique	Fioul	Bois	Chauffage urbain
	Estimation des paramètres	Estimation des paramètres	Estimation des paramètres	Estimation des paramètres	Estimation des paramètres	Estimation des paramètres
Constante	-74.3	-52.2	251.1 ***	-574.9 *	228.0	-448.7 *
Surface du logement selon le type de chauffage						
Surface (m²) en chauffage individuel	6.1 ***	4.7 ***	2.5 ***	5.8 ***	1.4 **	4.3 **
Surface (m²) en chauffage collectif	7.4 *	5.5 ***	2.7 ***	9.8 ***	6.5 ***	3.7 ***
Nombre de pièces d'habitation	52.3 ***	60.9 ***	23.0 ***	64.3 ***	33.4 *	99.9 ***
Zone climatique						
2 : Ouest et Sud Ouest	-143.5 ***	-174.8 ***		ref.	ref.	ref.
3 : Sud Est	-163.4 ***	-140.6 ***		-329.8 ***	-120.0 ***	-141.1 ***
1 : Reste de la France	ref.	ref.		-376.5 ***	-162.4 **	-151.8 *
Montant de l'assurance habitation						
123	ref.					
45	29.1					
678	173.8 ***					
mis	-23.6					
Climatisation						
Non	ref.		ref.			
Oui	-8.2		365.6 ***			
Date de construction du logement						
Avant 1949	ref.	ref.	ref.	ref.	ref.	ref.
Entre 1949 et 1974	31.1	-38.1 **	-29.3 ***	5.7	0.7	-17.0
Entre 1975 et 1981	19.9	-93.2 ***	-9.4	-124.7 *	-31.1	-6.8
Entre 1982 et 1989	-0.5	-117.2 ***	58.1 ***	-121.9	2.7	58.0
Entre 1990 et 1998	3.5	-121.4 ***	31.6 *	-180.7 **	-4.5	-26.7
Après 1999	-43.0 *	-195.0 ***	72.0 ***	-345.7 ***	-141.4 *	-22.3
Type de logement						
Individuel (maison)	ref.	ref.	ref.	ref.	ref.	ref.
Collectif (appartement)	-122.1 ***	-165.7 ***	-101.2 ***	-709.8 ***	-76.3	-9.7
Redressement des revenus						
Personne de référence est étudiante, ménage déclarant des revenus négatifs	ref.	ref.		ref.	ref.	ref.
Revenus redressés valides	704.3 ***	833.9 ***		1878.1 ***	471.6	732.7 ***
Logarithme du revenu par UC	64.4 ***	73.6 ***		171.9 ***	8.3	69.2 ***
Isolation des murs						
Moyenne ou mauvaise isolation	ref.	ref.		ref.	ref.	ref.
Bonne isolation des murs	-47.9 **	-56.3 ***		-82.2 **	-47.0	-5.8
Double vitrage						
Absence de double vitrage	ref.	ref.		ref.	ref.	ref.
Présence de double vitrage	21.5	-115.6 ***		-247.9 ***	-19.3	-8.2
Présence d'une personne « au foyer » dans le ménage						
Non	ref.	ref.	ref.	ref.	ref.	ref.
Oui	45.6 *	69.3 ***	9.9	189.5 ***	63.2	25.9
Âge de la personne de référence						
Moins de 35 ans	ref.		ref.			
Entre 35 et 50 ans	21.2		18.8			
Entre 50 et 65 ans	-49.4 *		-23.9 *			
De plus de 65 ans	-66.8 *		-127.0 ***			
Montant de l'assurance habitation						
123			ref.			
45			20.1 **			
678			94.9 ***			
mis			-5.3			
Redressement des revenus						
Personne de référence est étudiante, ménage déclarant des revenus négatifs			ref.			
Revenus redressés valides			92.6			
Logarithme du revenu par UC			2.8			

Source : CGDD, modèle Prometheus

ANNEXE 4 : DESCRIPTION ET PRIX PÉGASE UTILISÉS POUR CALCULER LES CONSOMMATIONS DE FIOUL GPL, BOIS ET DE CHAUFFAGE URBAIN À PARTIR DE L'ENL EN 2013

Énergie	Description	Unité	Prix TTC
Fioul	Prix unitaire	Tarif au kWh PCI	0,094
GPL	Prix unitaire en citerne	Tarif au kWh PCI	0,141
	Prix unitaire en bouteille	Tarif au kWh PCI	0,201
BOIS	Prix bois-bûche en vrac PCI au 1er trimestre 50 cm H1 : hors livraison en €/kWh	Tarif au kWh PCI	0,054
	Prix granulés en sacs PCI au 1er trimestre : palette départ fournisseur	Tarif au kWh PCI	0,061
Urbain	Prime fixe annuelle au tarif T100 LU	€/an	30,534
	Prime fixe annuelle au tarif T110 MU	€/an	23,81
	kwh vapeur au tarif T100 LU hiver	Tarif au kWh	0,067
	kwh vapeur au tarif T100 LU été	Tarif au kWh	0,042
	kwh vapeur au tarif T110 MU hiver	Tarif au kWh	0,076
	kwh vapeur au tarif T110 MU été	Tarif au kWh	0,053

Source : CGDD, modèle Prometheus

ANNEXE 5 : INFORMATIONS UTILISÉES POUR CALCULER LES CONSOMMATIONS DE GAZ

Tarification au gaz des ménages retenue pour le calcul des consommations

Tarif	Prix unitaire TTC (en €/kWh PCI)	Abonnement TTC (en €/an)
Base	0,104	52,96
B0	0,088	66,63
B1	0,059	209,10
B2I	0,059	209,10
B2S	0,066	15,78

Note de lecture : à partir de l'ENL 2013, pour un ménage qui a une facture de gaz et qui est à la tarification de « Base », sa consommation est obtenue à l'aide d'un prix unitaire de 0,104 €/kWh PCI (proportionnel à sa consommation) et un abonnement (fixe) d'une valeur de 52,96 €/an.

Source : CGDD, modèle Prometheus

Détail de la répartition de la tarification au gaz des ménages (pour le chauffage individuel et autres usages)

Répartition des factures de gaz	Tarif imputé
< P18	Base
[P18 ; P32 [B0
[P32 ; P95 [B1
> P95	B2I

Note de lecture : les ménages dont la facture de gaz est inférieure au 18^e centile, se voient imputer le tarif « Base ».

Source : CGDD, modèle Prometheus

ANNEXE 6 : HYPOTHÈSES, CRITÈRES ET INFORMATIONS UTILISÉES POUR CALCULER LES CONSOMMATIONS D'ÉLECTRICITÉ

Hypothèses utilisées pour l'imputation des tarifs d'électricité

- La partie variable du tarif du contrat à double tarification est la même respectivement en HP et HC, quelle que soit sa puissance. Un tarif moyen a donc été calculé, en appliquant à chaque ménage un tarif pondéré par la répartition moyenne de la consommation en heures pleines et en heures creuses¹⁶⁹. Les ménages auxquels le tarif HP/HC a été attribué, ont donc une facture dont la part variable est calculée sur la base d'une répartition de leurs consommations électriques de 57 % en heures pleines et de 43 % en heures creuses¹⁷⁰. En revanche la partie abonnement est différente selon la puissance souscrite ;
- l'affectation type de contrat (base ou tarif HC/HP) à un ménage repose sur l'hypothèse que les ménages suivent la recommandation d'EDF, qui est de choisir le tarif de base si le ménage n'utilise ni chauffage, ni chauffe-eau électrique, et dans le cas contraire, de choisir le tarif HP/HC ;
- une catégorie tarifaire a été créée pour les « très gros consommateurs », qui correspond aux puissances supérieures à 12 kVA. En effet, ces consommateurs sont peu nombreux (environ 1,5 %) ;
- pour les « très gros consommateurs », Engie ne formule pas de recommandation sur l'option tarifaire, et l'hypothèse est faite qu'ils souscrivent tous à l'option HP/HC qui paraît plus avantageuse. La partie abonnement est une moyenne des abonnements des puissances supérieures à 12 kVA.

Critères utilisés pour l'imputation de la puissance d'électricité et du contrat à chaque ménage

Option 3 kVA :

(logement = collectif) ET (surface = < 50 m²) ET (énergie de chauffage ≠ électricité) ET (énergie principale d'eau chaude ≠ électricité)

→ Contrat = Base

Option 6 kVA :

(logement = collectif) ET (surface = < 50 m²) ET (énergie de chauffage ≠ électricité) ET (énergie d'eau chaude = électricité)

OU

(logement = collectif) ET (surface = > 50 m²) ET (énergie de chauffage ≠ électricité)

OU

¹⁶⁹ $\text{prix_TTC_HPHC} = 0,57 * \text{prix_TTC_HP} + 0,43 * \text{prix_TTC_HC}$

¹⁷⁰ Cette clé de répartition est issue des données fournies à l'époque par la direction générale de l'Énergie et du Climat.

(logement = individuel) ET (surface = < 100 m²) ET (énergie de chauffage ≠ électricité)

OU

(logement = collectif) ET (surface = < 80 m²) ET (énergie de chauffage = électricité)

→ Contrat = Base : si (énergie de chauffage ≠ électricité) ET (énergie d'eau chaude ≠ électricité), sinon Contrat = HCHP

Option 9 kVA :

(logement = individuel) ET (surface = < 100 m²) ET (énergie de chauffage = électricité)

OU

(logement = individuel) ET (surface = > 100 m²) ET (énergie de chauffage ≠ électricité)

OU

(logement = collectif) ET (surface = > 80 m²) ET (énergie de chauffage = électricité)

→ Contrat = HCHP

Option 12 kVA :

(logement = individuel) ET (surface = > 100 m²) ET (énergie de chauffage = électricité)

→ Contrat = HCHP

Option >12 kVA (de 15 à 36 kVA) :

(logement = individuel) ET (surface = > 160 m²) ET (énergie de chauffage = électricité)

→ Contrat = HCHP

Tarification de l'électricité des ménages retenue pour le calcul des consommations

Tarif	Prix unitaire TTC (en €/kWh)	Abonnement TTC (en €/an)
Base 3 kVA	0,131	62,85
Base 6 kVA	0,132	84,33
HP/HC 6 kVA	0,126	97,37
HP/HC 9 kVA	0,126	122,32
HP/HC 12 kVA	0,126	205,38
>12 kVA	0,126	507,58

Note de lecture : à partir de l'ENL 2013, pour un ménage qui a une facture d'électricité et qui est à la tarification de « Base » et à la puissance 3 kVA, sa consommation est obtenue à l'aide d'un prix unitaire de 0,131 €/kWh (proportionnel à sa consommation) et un abonnement (fixe) d'une valeur de 62,85 €/an.

Source : CGDD, modèle Prometheus

ANNEXE 7 : LES REVENUS DANS L'ENL 2013

Dans l'ENL 2013, la variable de base initiale est MRTOTA3 qui est la variable de revenu total qui ne tient pas compte des impôts directs et indirects ; elle est la somme des :

- revenus d'activité ou de remplacement (revenus salariaux et non salariaux, chômage, allocation de solidarité spécifique, pensions et retraites) (ZRMACT3) ;
- revenus des prestations (prestations familiales, aide sociale, allocation RSA) (ZRAIDE3) ;
- revenus fonciers (ZRFONC3) ;
- revenus de l'épargne (RMEPAR3) ;
- revenus autres (pensions alimentaires, indemnités de licenciement etc.) – (RMAUT3).

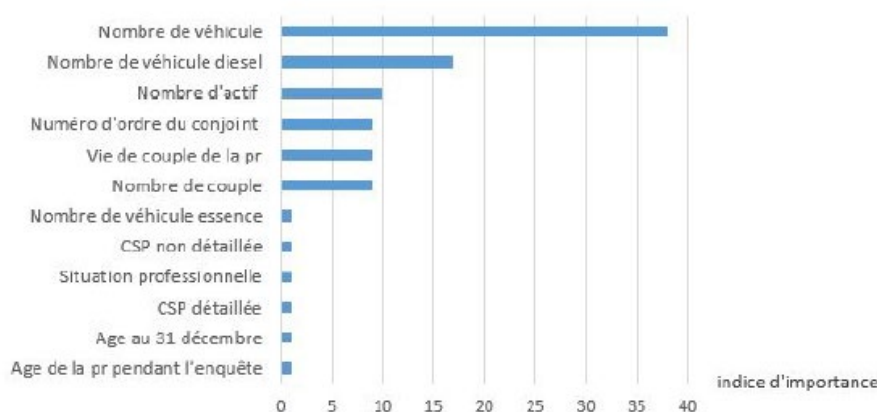
L'Insee avait calé l'enquête et cette variable sur les centiles de revenus de ERFS 2013, sur le champ des ménages hors étudiants et hors ménages ayant déclarés des revenus négatifs. Cette variable a vocation à rendre compte des revenus par grandes catégories de ménages, en niveau et en évolution, notamment pour le calcul de taux d'effort comparables dans le temps. Pour plus de détail, se référer à *l'Enquête logement 2013 – Dictionnaire des codes* déjà citée, *paragraphe 8.1* (Préconisations et précautions d'usage – Le revenu des ménages).

ANNEXE 8 : PROCÉDURE DE SÉLECTION AVEC ARBRE DE RÉGRESSION

La procédure CART a été lancée désignant douze variables explicatives du kilométrage annuel.

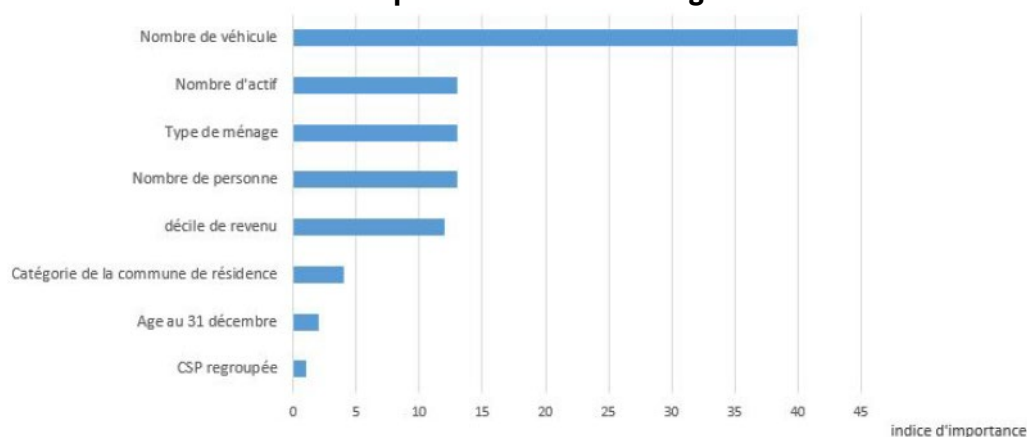
Nombreuses sont les variables qui apportent une information similaire : par exemple, le nombre de véhicules et le nombre de véhicules diesel ou essence ; ou encore les deux variables d'âge de la personne de référence, ou sa situation de couple. Il s'agit donc de répéter cette procédure en lançant plusieurs arbres de régression itérativement, en retirant à chaque étape certaines variables choisies par l'algorithme pour lesquelles l'information est redondante. L'analyse exploratoire par arbre de régression a ainsi été répétée à trois reprises jusqu'à obtenir une version satisfaisante.

Variables classées selon l'indice d'importance en termes de prédiction du kilométrage annuel du premier arbre de classification



Source : CGDD, modèle Prometheus

Variables classées selon l'indice d'importance de l'arbre élagué



Source : CGDD, modèle Prometheus

ANNEXE 9 : RÉSULTAT DE LA RÉGRESSION

```
lm(formula = knage_total ~ agepr + nb_veh + decile + nperso +
    situa + Catcom + TAA, data = base_menage_recodage, weights = base_menage$pond_menc)
```

Weighted Residuals:

Min	1Q	Median	3Q	Max
-1657440	-202414	-24935	151193	4650459

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2891.2	467.7	6.182	6.50e-10 ***
agepr26-40	17.0	364.8	0.047	0.962821
agepr41-60	-489.3	354.7	-1.379	0.167831
agepr60+	-2103.6	442.5	-4.754	2.02e-06 ***
nb_veh1	8713.2	190.0	45.864	< 2e-16 ***
nb_veh2	19062.8	245.3	77.698	< 2e-16 ***
nb_veh3 ou plus	30582.5	341.4	89.577	< 2e-16 ***
decile10	3648.6	362.0	10.080	< 2e-16 ***
decile2	-56.3	298.2	-0.189	0.850262
decile3	115.6	302.1	0.383	0.701961
decile4	555.2	323.0	1.719	0.085661 .
decile5	1121.8	326.4	3.437	0.000591 ***
decile6	1288.5	333.3	3.866	0.000111 ***
decile7	2051.4	337.7	6.074	1.28e-09 ***
decile8	2116.4	350.8	6.033	1.65e-09 ***
decile9	3777.5	353.9	10.674	< 2e-16 ***
nperso2	-581.4	185.8	-3.129	0.001756 **
nperso3	431.6	251.3	1.717	0.085938 .
nperso4	1639.0	272.9	6.006	1.95e-09 ***
nperso5 ou plus	1757.2	323.9	5.425	5.90e-08 ***
situaAutres	-2435.8	436.5	-5.580	2.45e-08 ***
situaInactifs	-2280.2	306.2	-7.446	1.02e-13 ***
situaSt/rech	-1037.7	321.2	-3.231	0.001236 **
Catcomcommune-centre	-322.8	192.0	-1.681	0.092813 .
Catcomcommune de la couronne	2375.2	187.0	12.701	< 2e-16 ***
Catcomcommune hors attraction	2563.8	347.3	7.382	1.65e-13 ***
TAA+700 000 + paris	-2013.3	223.0	-9.028	< 2e-16 ***
TAA200 000-700 000 hab	-986.5	227.1	-4.344	1.41e-05 ***
TAA50 000-200 000 hab	-1167.7	242.2	-4.821	1.44e-06 ***
TAACom hors attraction	NA	NA	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 353200 on 13796 degrees of freedom
Multiple R-squared: 0.6554, Adjusted R-squared: 0.6547
F-statistic: 937 on 28 and 13796 DF, p-value: < 2.2e-16

Source : CGDD, modèle Prometheus

ANNEXE 10 : HOMOGÉNÉISATION DES VARIABLES

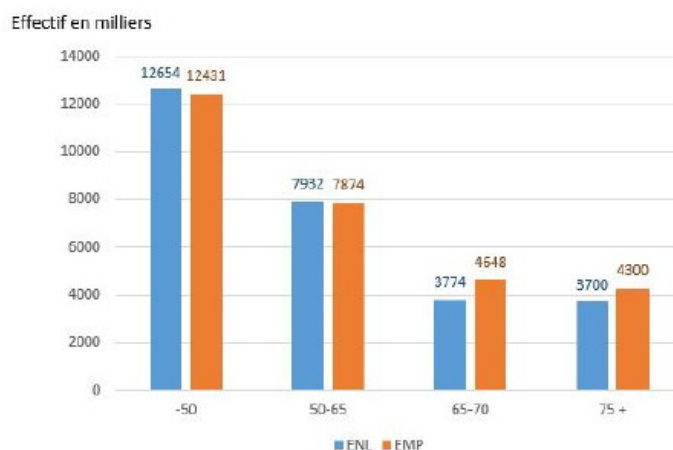
Les figures suivantes représentent l'effectif pondéré pour les variables de stratification. Ils correspondent à la variable choisie après recodage des modalités.

Effectif pondéré calculé pour les modalités de la variable du nombre de véhicule



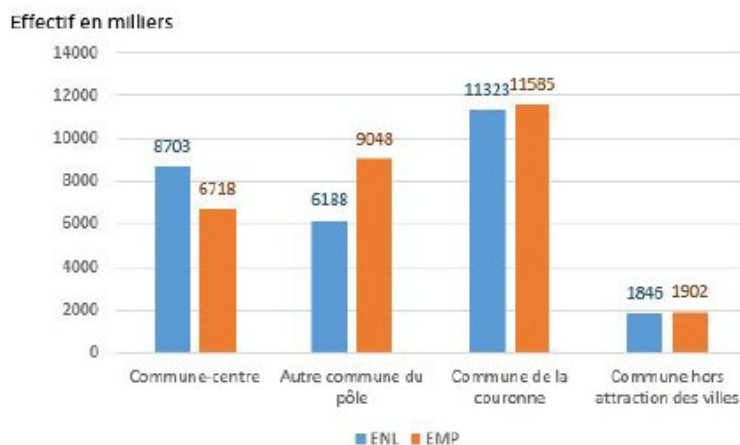
Source : CGDD, modèle Prometheus

Effectif pondéré calculé pour les modalités de la variable de la tranche d'âge de la PR



Source : CGDD, modèle Prometheus

Effectif pondéré calculé pour les modalités de la variable de la catégorie de la commune



Source : CGDD, modèle Prometheus

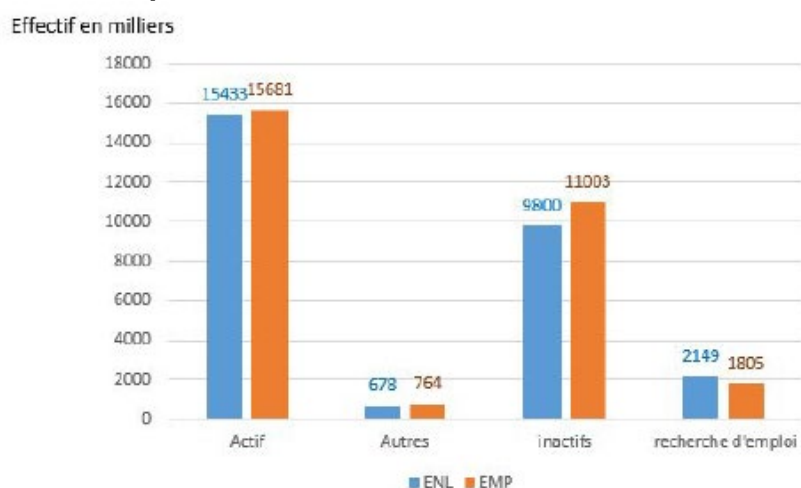
Effectif pondéré calculé pour les modalités de la variable du dixième de niveau de vie



Note : dans les données de l'EMP, apparaît un déséquilibre entre les dixièmes qui a été corrigé.

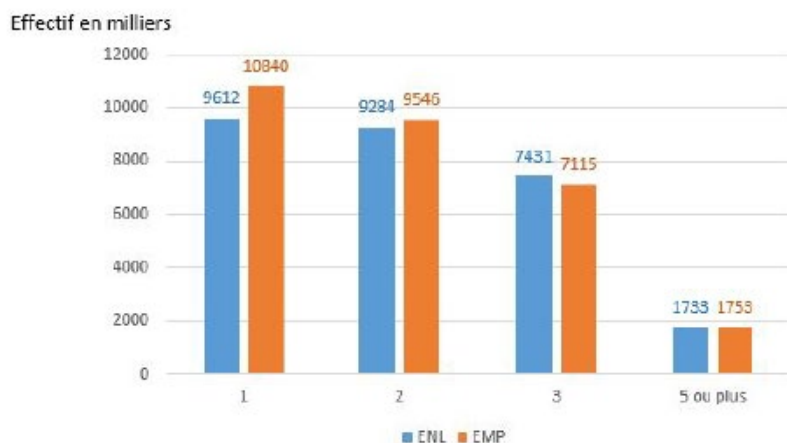
Source : CGDD, modèle Prometheus

Effectif pondéré calculé pour les modalités de la variable de situation vis-à-vis du travail



Source : CGDD, modèle Prometheus

Effectif pondéré calculé pour les modalités de la variable du nombre de personne dans le ménage



Source : CGDD, modèle Prometheus

ANNEXE 11 : VALEURS DES CORRÉLATIONS

Statistiques distributives sur le kilométrage annuel réalisé par les ménages dans l'EMP 2019 et à l'issue du pseudo-appariement

Variables de comparaisons	EMP	Table pseudo-appariée	Écart-relatif (en %)
Somme (en millions)	395 350	375 676	5
Moyenne	13 515	13 803	- 2
Q1	9 037	8 603	5
Médiane	13 977	13 004	7
Q3	17 291	15 501	10
D1	6 578	7 240	- 10
D2	8 745	9 395	- 7
D3	10 837	11 591	- 7
D4	12 765	12 581	1
D5	13 986	14 248	- 2
D6	14 715	15 071	- 2
D7	16 937	15 793	78
D8	16 556	16 962	- 4
D9	17 457	16 962	2
D10	16 727	17 117	- 8

Source : CGDD, modèle Prometheus

ANNEXE 12 : VALEURS DES CORRÉLATIONS

Matrice de corrélation entre variables initialement dans l'EMP

	Kilométrage annuel	Nombre de véhicules	Nombre de diesel	Nombre d'essence	Nombre d'électrique	Nombre d'hybride	Nombre de GPL
Kilométrage annuel	1	0,77	0,68	0,17	0,05	0,05	0,05
Nombre de véhicules	0,77	1	0,72	0,43	0,05	0,06	0,06
Nombre de diesel	0,68	0,72	1	- 0,30	0,00	- 0,07	- 0,02
Nombre d'essence	0,17	0,43	- 0,30	1	- 0,02	- 0,03	- 0,02
Nombre d'électrique	0,05	0,05	0,00	- 0,02	1	0,01	0,03
Nombre d'hybride	0,05	0,06	- 0,07	- 0,03	0,01	1	0,00
Nombre de GPL	0,05	0,06	- 0,02	- 0,02	0,03	0,00	1

Matrice de corrélation entre variables dans la base pseudo-appariée

	Kilométrage annuel	Nombre de véhicules	Nombre de diesel	Nombre d'essence	Nombre d'électrique	Nombre d'hybride	Nombre de GPL
Kilométrage annuel	1	0,76	0,67	0,16	0,03	0,05	0,04
Nombre de véhicules	0,76	1	0,71	0,42	0,04	0,06	0,07
Nombre de diesel	0,67	0,71	1	- 0,31	- 0,01	- 0,06	- 0,02
Nombre d'essence	0,16	0,42	- 0,31	1	- 0,01	- 0,03	- 0,01
Nombre de GPL/autres	0,03	0,04	- 0,01	- 0,01	1	0,01	0,00
Nombre d'électrique	0,05	0,06	- 0,06	- 0,03	0,01	1	0,01
Nombre de GPL	0,04	0,07	- 0,02	- 0,01	0,00	0,01	1

Source : CGDD, modèle Prometheus

ANNEXE 13 : NOMENCLATURE RP (INSEE) UTILISÉE DANS PROMETHEUS

Croisement conservés et actualisés dans Prometheus

Type de logement	Mode de chauffage	Énergie
Maison	Individuel	Gaz
		Fioul
		Électricité
		GPL
		Autre
Appartement	Collectif	Urbain
		Gaz
		Fioul
		Électricité
		Autre
	Individuel	Gaz
		Électricité

Note : les ménages en appartement, se chauffant individuellement au bois évoluent dans Prometheus comme la catégorie « Maison * Individuel * Autre ».

Source : CGDD, modèle Prometheus

ANNEXE 14 : PRÉCISIONS SUR LES CALCULS DES CONSOMMATIONS

Nomenclature SDES des consommations énergétiques réelles dans le résidentiel, en TWh (Métropole)

Consommation finale de chaleur (vendue)
Consommation finale de charbon
Consommation finale d'électricité (et détail spécifique)
Consommation finale d'énergies renouvelables thermiques et déchets
<i>Consommation de biomasse solide dans le résidentiel</i>
<i>Consommation de chaleur renouvelable issue des pompes de chaleur dans le résidentiel</i>
<i>Consommation d'énergie solaire thermique dans le résidentiel</i>
Consommation finale de gaz naturel*
Consommation finale de produits raffinés

*Disponible en PCS également

Quelques traitements sont nécessaires à l'heure actuelle :

- l'usage spécifique de l'électricité n'est pas disponible avant 2011, une rétopolation avec les taux de croissance de la consommation d'électricité est effectuée ;
- la catégorie « produits raffinés » comprend à la fois l'énergie fioul et GPL, par conséquent on applique une clef de répartition correspondant à la moyenne sur 2014-2021 (0,8 en faveur du fioul) ;
- la catégorie « biomasses solide » correspond au bois.

La consommation d'électricité spécifique selon le type de logement et l'énergie de chauffage principale

Part en %

Type de logement	Énergie principale de chauffage	Part de la consommation électrique totale
Individuel	Électricité	40
	Bois	60
	Pompe à chaleur/ solaire thermique	50
	Autres énergies	100
Collectif	Électricité	40
	Pompe à chaleur/ solaire thermique	50
	Autres énergies	100

Note de lecture : en moyenne, 40 % de l'électricité consommée dans un logement individuel dont l'énergie principale de chauffage est l'électricité, est de l'électricité spécifique. L'électricité consommée dans un logement individuel dont l'énergie principale de chauffage n'est ni l'électricité, ni le bois, ni la pompe à chaleur/solaire thermique, est entièrement de l'électricité spécifique.

Source : CGDD, modèle Prometheus

ANNEXE 15 : ESTIMATIONS ET DÉTAILS DES MODÈLES LOGISTIQUES DE TRANSITION

Sélection des variables

Variables considérées dans la sélection automatique

Intitulé	Signification
km_voit_per_men_sum	Nombre annuel de Km du ménage
sexe_pr	Sexe de la personne de référence
age_pr	Âge de la personne de référence
nperso	Nombre de personnes dans le ménage
mpao	Nombre d'actifs occupés dans le ménage
mne1	Nombre d'enfant à charge
decile_nivvie	Dixième de niveau de vie
logement	Type de logement
Catcom	Catégorie de commune
tdaav	Taille de l'aire d'attractivité de la commune
bcartpage	Présence d'une carte de péage

Quelques recodages sont faits sur les formats, et les regroupements sont les suivants :

- on distingue commune « du pôle » et « hors pôle » ;
- on regroupe les tailles d'aires d'attractivité des villes comme ce qui sera in fine utilisé dans le modèle en sortie ;
- des indicatrices sont faites pour couvrir tous les cas (on s'arrête à un maximum de trois voitures comme dans la V2 et du fait du faible nombre d'observations au-delà).

À noter :

- la présence d'une carte de péage est utilisée, car les précédentes régressions le faisaient (des variables telles que les *distances au transport* ont été testées sans succès) ;

- La variable d'âge est la même que celle utilisée pour le pseudo-appariement (donc avec des regroupements qui éliminaient les modalités non significatives dans l'explication des Km parcourus) ;
- la variable de diplôme présente auparavant a aussi été testée sans succès.

Synthèses des modèles

Probabilité modélisée	Variable de la regression
P(GO=1/VUL=1)	<i>km_voit_per_men_sum + sexe_pr + age_pr + nperso + decile_nivvie + Catcom</i>
P(GO=2/VUL=2)	<i>km_voit_per_men_sum + age_pr + mpao + decile_nivvie + Catcom</i>
P(GO=1, ES=1/VUL=2)	<i>km_voit_per_men_sum + age_pr + nperso + decile_nivvie + catcom + bcartpage</i>
P(GO=1, ES=2/VUL=3)	<i>mne1 + Catcom</i>
P(GO=2, ES=1/VUL=3)	<i>km_voit_per_men_sum + age_pr + Catcom</i>
P(GO=3 /VUL=3)*	<i>age_pr + logement + tdaav + bcartpage</i>

La sélection automatique *stepwise* est utilisée, parfois des variables dont les modalités étaient peu significatives ont été enlevées, puis la sélection relancée, et les modèles présentaient alors des AIC plus forts. L'ensemble des tests usuels ont été faits et validés (basés sur https://lrouviere.github.io/doc_cours/poly_logistique.pdf).

Les modèles

Modèle 1

```
Call:
glm(formula = ivoitvul_1go ~ km_voit_per_men_sum + sexe_pr +
    age_pr + nperso + mpao + dec_nivvie + catcom, family = "binomial",
    data = data1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6231 -1.1129  0.6397  0.9709  1.9516

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.789e-01  8.860e-02 -2.019  0.043495 *
km_voit_per_men_sum  9.332e-05  3.557e-06  26.237 < 2e-16 ***
sexe_prH         4.471e-01  4.177e-02  10.705 < 2e-16 ***
age_pr50-65     -8.114e-02  5.113e-02 -1.587  0.112547
age_pr65-70    -2.792e-01  7.244e-02 -3.854  0.000116 ***
age_pr75 +     -7.201e-01  7.580e-02 -9.500 < 2e-16 ***
nperso2         1.634e-01  4.890e-02  3.342  0.000831 ***
nperso3         2.493e-01  6.387e-02  3.904  9.47e-05 ***
nperso5 ou plus  5.037e-01  9.542e-02  5.273  1.30e-07 ***
mpao           -8.290e-02  3.648e-02 -2.273  0.023041 *
dec_nivvie10   -7.248e-01  9.665e-02 -7.499  6.42e-14 ***
dec_nivvie2    -3.755e-01  8.474e-02 -4.431  9.39e-06 ***
dec_nivvie3    -4.194e-01  8.832e-02 -4.748  2.05e-06 ***
dec_nivvie4    -3.510e-01  8.879e-02 -3.954  7.69e-05 ***
dec_nivvie5    -2.934e-01  8.993e-02 -3.263  0.001104 **
dec_nivvie6    -1.745e-01  9.155e-02 -1.906  0.056599 .
dec_nivvie7    -5.778e-01  9.293e-02 -6.218  5.03e-10 ***
dec_nivvie8    -5.483e-01  9.365e-02 -5.855  4.76e-09 ***
dec_nivvie9    -5.187e-01  9.524e-02 -5.446  5.14e-08 ***
catcomp0le    -2.748e-01  4.219e-02 -6.512  7.41e-11 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16619  on 12268  degrees of freedom
Residual deviance: 14937  on 12249  degrees of freedom
AIC: 14977

Number of Fisher Scoring iterations: 4
```

Modèle 2

```
Call:
glm(formula = ivoitvul_2go ~ km_voit_per_men_sum + age_pr + mpao +
    dec_nivvie + catcom, family = "binomial", data = data2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3108 -1.2059  0.7948  1.0578  1.5932

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.863e-02  1.473e-01  0.330  0.74132
km_voit_per_men_sum  3.342e-05  2.705e-06  12.354 < 2e-16 ***
age_pr50-65     8.609e-02  6.416e-02  1.342  0.17965
age_pr65-70    -1.509e-01  1.045e-01 -1.444  0.14868
age_pr75 +     -7.098e-02  1.402e-01 -0.506  0.61275
mpao          -8.737e-02  3.922e-02 -2.227  0.02592 *
dec_nivvie10   -7.922e-01  1.438e-01 -5.508  3.63e-08 ***
dec_nivvie2    2.102e-01  1.603e-01  1.311  0.18984
dec_nivvie3    -1.795e-01  1.533e-01 -1.171  0.24157
dec_nivvie4    -2.422e-02  1.494e-01 -0.162  0.87119
dec_nivvie5    -1.914e-01  1.461e-01 -1.310  0.19011
dec_nivvie6    -2.498e-01  1.441e-01 -1.734  0.08296 .
dec_nivvie7    -4.663e-01  1.424e-01 -3.275  0.00106 **
dec_nivvie8    -6.062e-01  1.414e-01 -4.285  1.82e-05 ***
dec_nivvie9    -7.750e-01  1.418e-01 -5.466  4.59e-08 ***
catcomp0le    -2.665e-01  5.374e-02 -4.959  7.08e-07 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8842.1  on 6457  degrees of freedom
Residual deviance: 8484.3  on 6442  degrees of freedom
AIC: 8516.3

Number of Fisher Scoring iterations: 4
```

Modèle 3

Call:
glm(formula = ivoitvul_1goies ~ km_voit_per_men_sum + age_pr + nperso + dec_nivvie + catcom + bcartpeage, family = "binomial", data = data3)

Deviance Residuals:
Min 1Q Median 3Q Max
-2.5445 -1.0167 0.5961 0.7944 1.5477

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.891e-01 4.873e-01 -1.004 0.315518
km_voit_per_men_sum 6.344e-05 5.257e-06 12.069 < 2e-16 ***
age_pr50-65 1.213e-01 1.002e-01 1.210 0.226261
age_pr65-70 4.816e-01 1.398e-01 3.444 0.000573 ***
age_pr75 + -3.027e-02 1.665e-01 -0.182 0.855709
nperso2 -4.204e-01 1.508e-01 -2.788 0.005300 **
nperso3 -9.759e-02 1.587e-01 -0.615 0.538574
nperso5 ou plus 1.647e-02 2.018e-01 0.082 0.934949
dec_nivvie10 -8.214e-01 2.229e-01 -3.684 0.000229 ***
dec_nivvie2 -2.927e-01 2.611e-01 -1.121 0.262314
dec_nivvie3 -1.040e-01 2.546e-01 -0.409 0.682806
dec_nivvie4 2.211e-01 2.625e-01 0.842 0.399675
dec_nivvie5 1.385e-01 2.526e-01 0.548 0.583574
dec_nivvie6 -1.425e-01 2.404e-01 -0.593 0.553383
dec_nivvie7 -1.093e-01 2.337e-01 -0.468 0.639978
dec_nivvie8 -5.640e-01 2.278e-01 -2.476 0.013299 *
dec_nivvie9 -2.409e-01 2.270e-01 -1.062 0.288432
CatcomPôle -3.879e-01 8.012e-02 -4.841 1.29e-06 ***
bcartpeage 4.667e-01 2.064e-01 2.261 0.023740 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4303.5 on 3776 degrees of freedom
Residual deviance: 3963.2 on 3758 degrees of freedom
AIC: 4001.2

Number of Fisher Scoring iterations: 4

Modèle 5

Call:
glm(formula = ivoitvul_2goies ~ km_voit_per_men_sum + age_pr + Catcom, family = "binomial", data = data5)

Deviance Residuals:
Min 1Q Median 3Q Max
-2.3055 -1.0707 0.6442 0.9299 1.7764

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.079e-01 3.342e-01 -1.520 0.12856
km_voit_per_men_sum 4.526e-05 8.800e-06 5.143 2.71e-07 ***
age_pr50-65 -2.287e-02 2.186e-01 -0.105 0.91666
age_pr65-70 -8.156e-01 2.978e-01 -2.739 0.00616 **
age_pr75 + -1.357e+00 5.570e-01 -2.436 0.01486 *
CatcomPôle -4.911e-01 1.989e-01 -2.469 0.01356 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 719.82 on 538 degrees of freedom
Residual deviance: 634.78 on 533 degrees of freedom
AIC: 646.78

Number of Fisher Scoring iterations: 4

Modèle 6

Call:
glm(formula = ivoitvul_3go ~ age_pr + logement + Tdaav + bcartpeage, family = "binomial", data = data6)

Deviance Residuals:
Min 1Q Median 3Q Max
-1.7398 -1.1255 0.8474 1.0679 1.9888

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.163209 1.563217 -2.663 0.00774 **
age_pr50-65 0.016101 0.166816 0.097 0.92311
age_pr65-70 -0.686146 0.324409 -2.115 0.03442 *
age_pr75 + -0.766358 0.751280 -1.020 0.30770
logementMaison -0.592288 0.251801 -2.352 0.01866 *
TdaavAire de 200 000 à moins de 700 000 habitant -0.471312 0.249871 -1.886 0.05927 .
TdaavAire de 50 000 à moins de 200 000 habitants 0.006385 0.259419 0.025 0.98036
TdaavAire de moins de 50 000 habitants 0.242710 0.306379 0.792 0.42825
TdaavAire de Paris -0.386478 0.268242 -1.441 0.14965
TdaavCommune hors attraction des villes 0.409348 0.326390 1.254 0.20978
bcartpeage 2.501292 0.765863 3.266 0.00109 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 964.28 on 696 degrees of freedom
Residual deviance: 926.20 on 686 degrees of freedom
AIC: 948.2

Number of Fisher Scoring iterations: 4

Source : CGDD, modèle Prometheus

Modèle 4

Call:
glm(formula = ivoitvul_1go2es ~ mnel + Catcom, family = "binomial", data = data4)

Deviance Residuals:
Min 1Q Median 3Q Max
-2.1706 0.4464 0.5347 0.6942 1.2777

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) 2.2562 0.3092 7.296 2.97e-13 ***
mnel -0.3832 0.1541 -2.487 0.01287 *
CatcomPôle -0.9560 0.3456 -2.766 0.00567 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

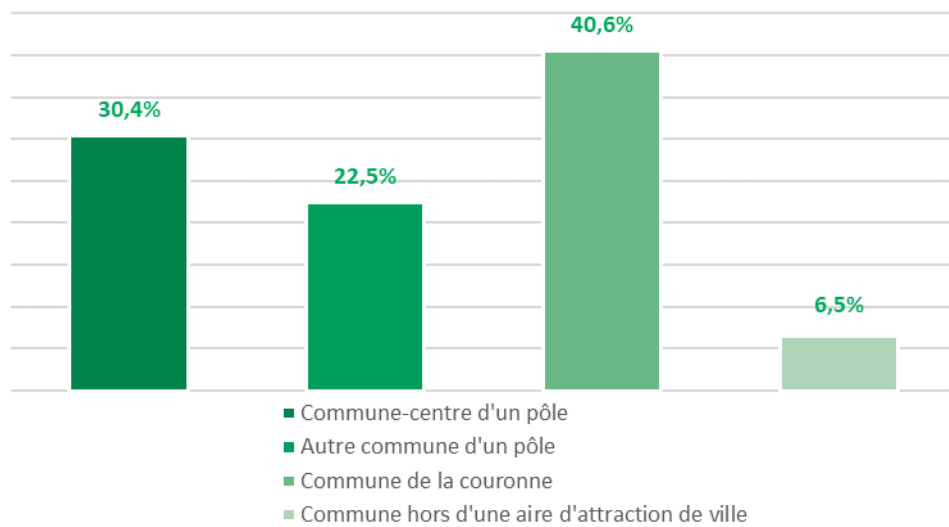
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 247.49 on 256 degrees of freedom
Residual deviance: 235.48 on 254 degrees of freedom
AIC: 241.48

Number of Fisher Scoring iterations: 4

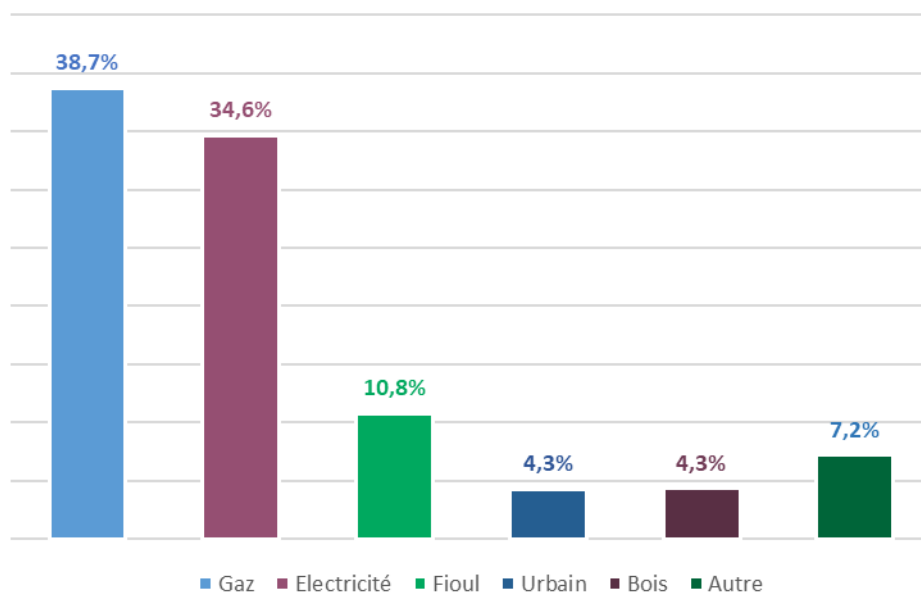
ANNEXE 16 : STATISTIQUES UNIVARIÉES

Figure 1 : répartition des ménages selon la catégorie de commune dans laquelle ils habitent



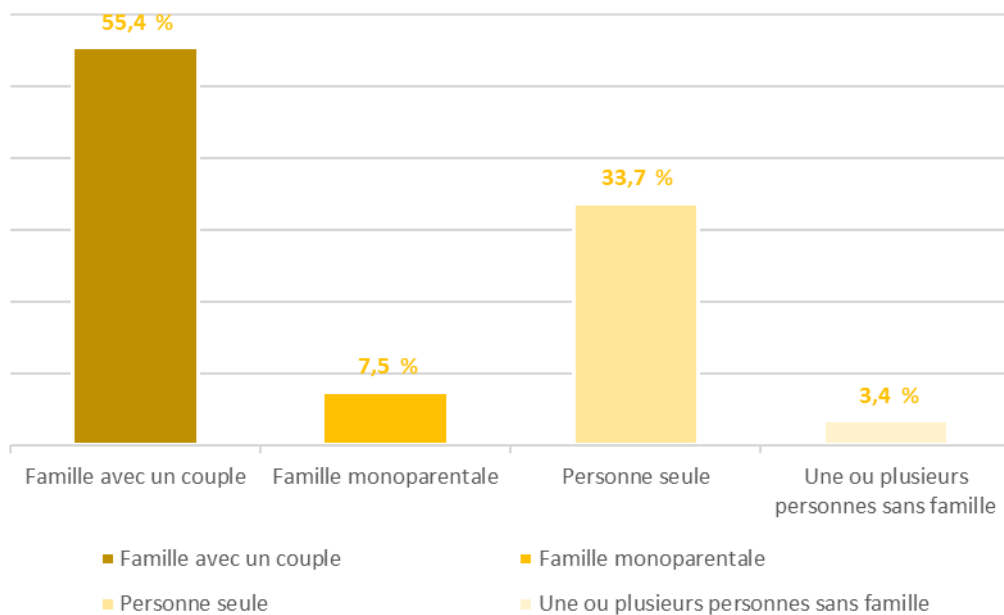
Source : CGDD, modèle Prometheus

Figure 2 : répartition des ménages selon leur énergie principale de chauffage



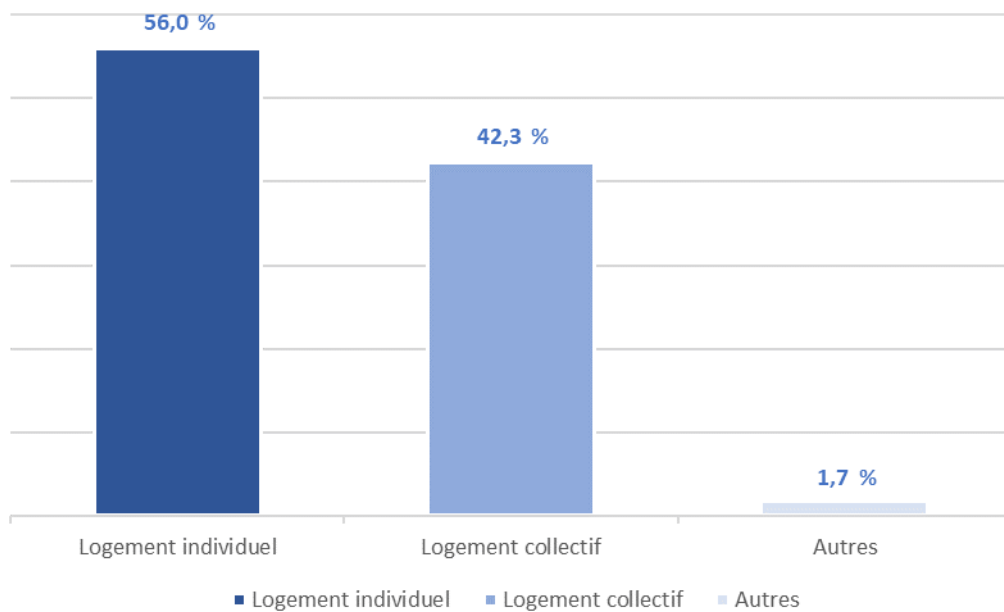
Source : CGDD, modèle Prometheus

Figure 3 : répartition des ménages selon leur composition



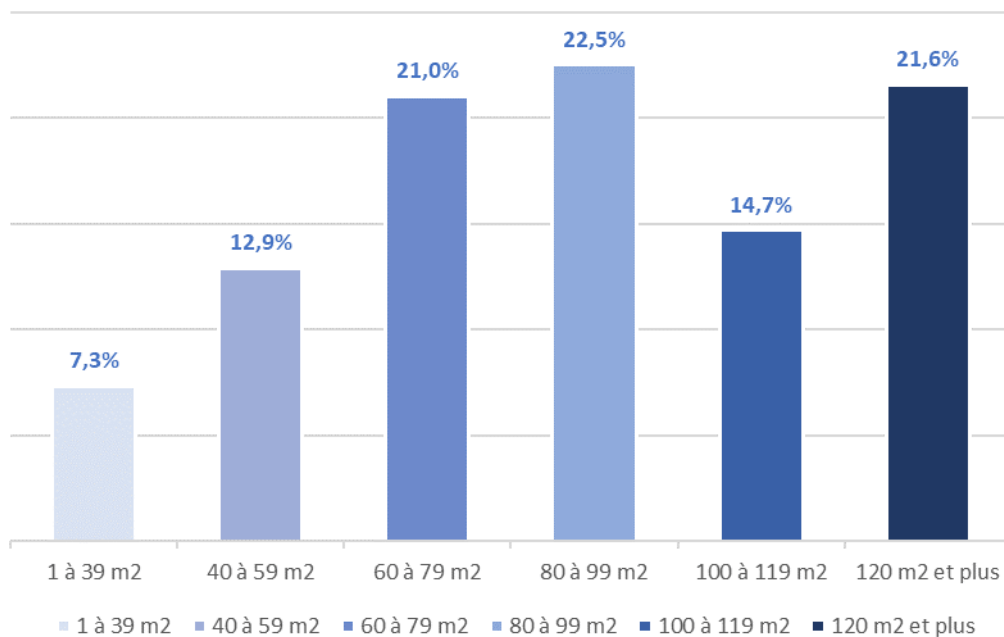
Source : CGDD, modèle Prometheus

Figure 4 : répartition des ménages selon leur type de logement



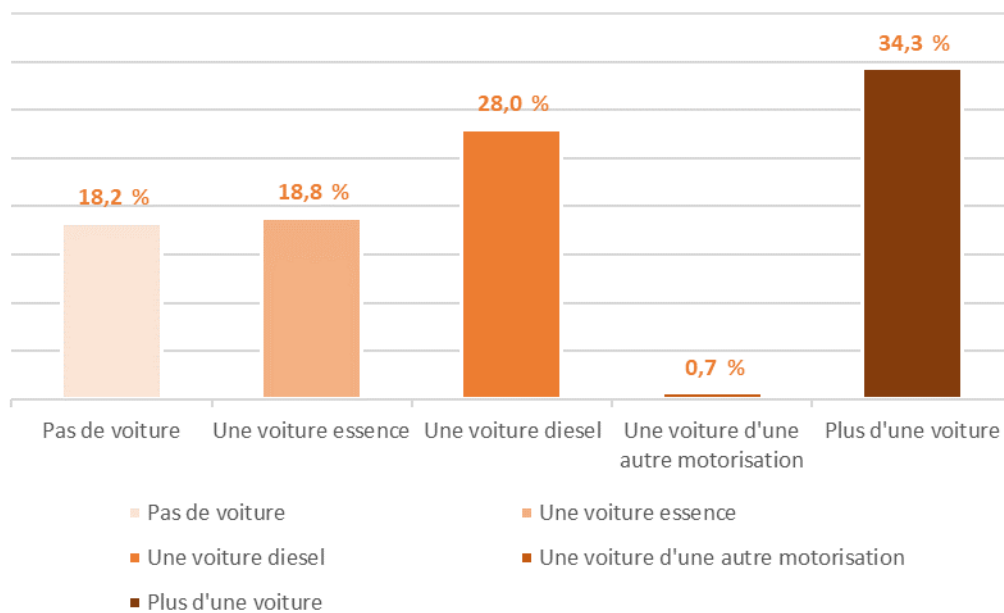
Source : CGDD, modèle Prometheus

Figure 5 : répartition des ménages selon la surface habitable de leur logement



Source : CGDD, modèle Prometheus

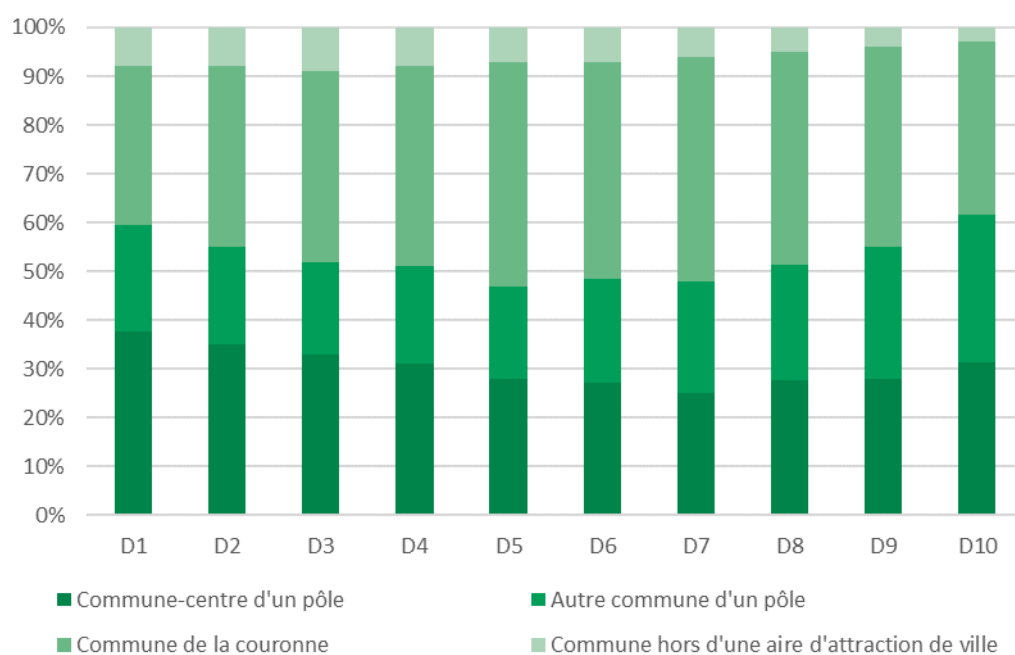
Figure 6 : répartition des ménages selon leur équipement en véhicule



Source : CGDD, modèle Prometheus

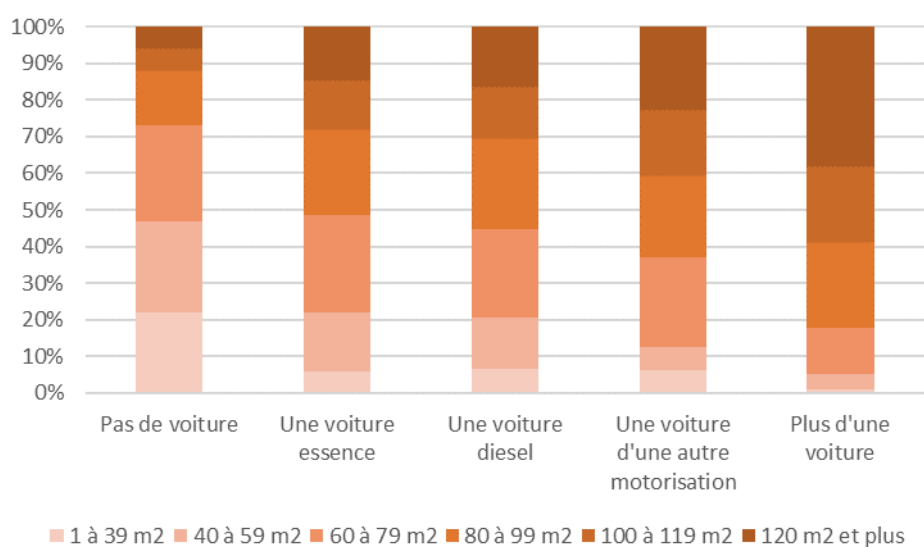
ANNEXE 17 : STATISTIQUES BIVARIÉES

Figure 1 : répartition des ménages par catégories de communes et selon leur dixième de niveau de vie



Source : CGDD, modèle Prometheus

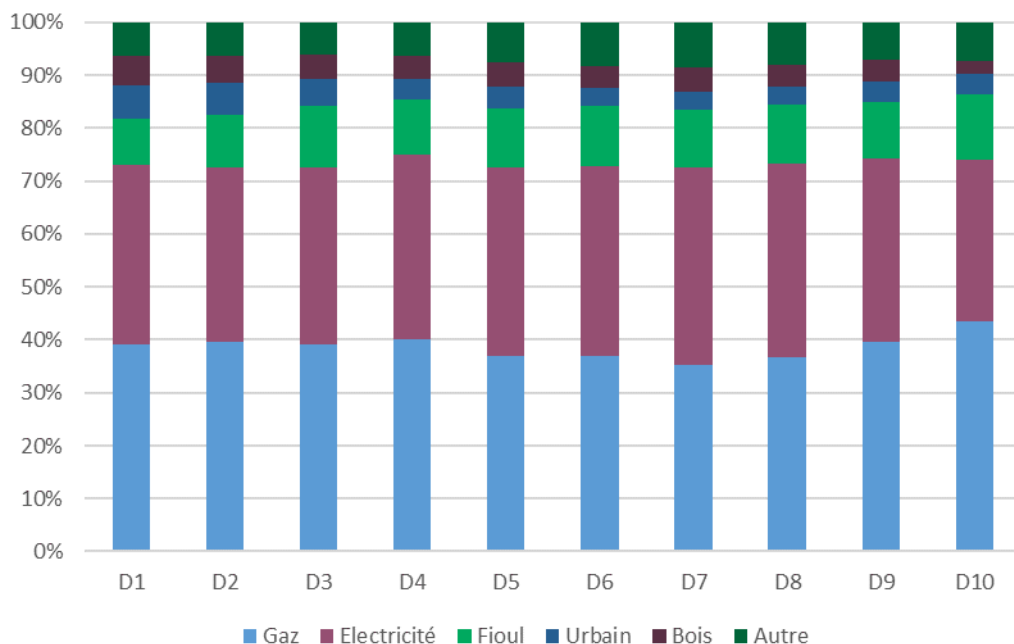
Figure 2 : répartition des ménages par surface du logement et selon leur équipement en véhicule



Source : CGDD, modèle Prometheus

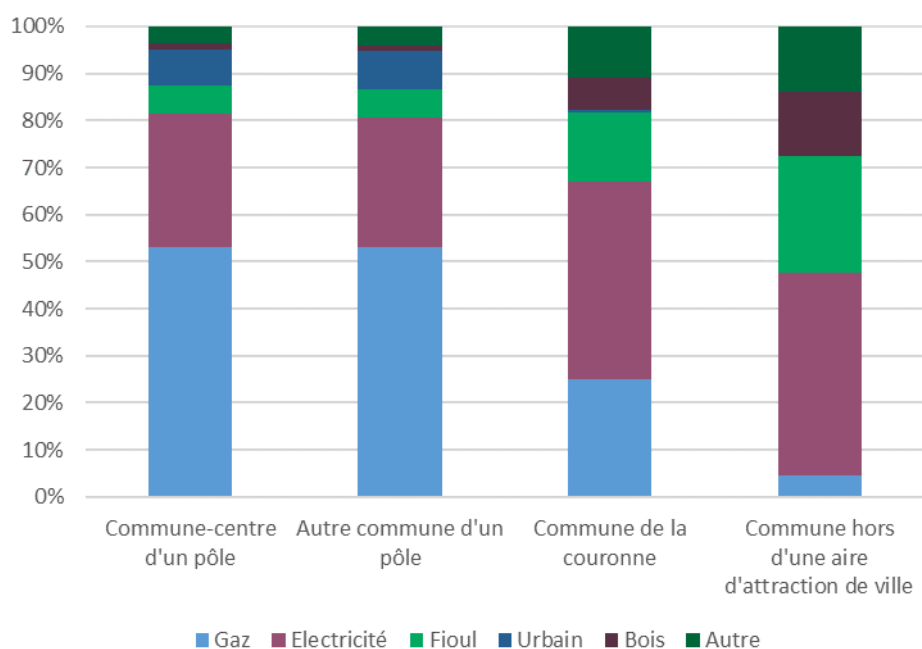
Figure 3 : répartition des ménages par énergie principale de chauffage selon leur dixième de niveau de vie et leur catégorie de commune

3-A selon leur dixième de niveau de vie



Source : CGDD, modèle Prometheus

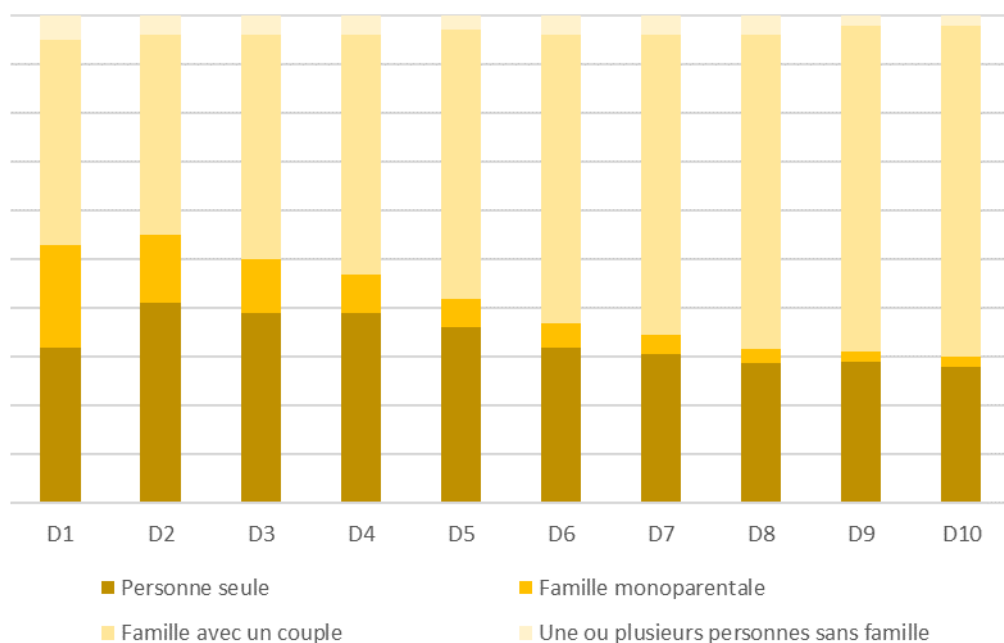
3-A selon leur catégorie de commune



Source : CGDD, modèle Prometheus

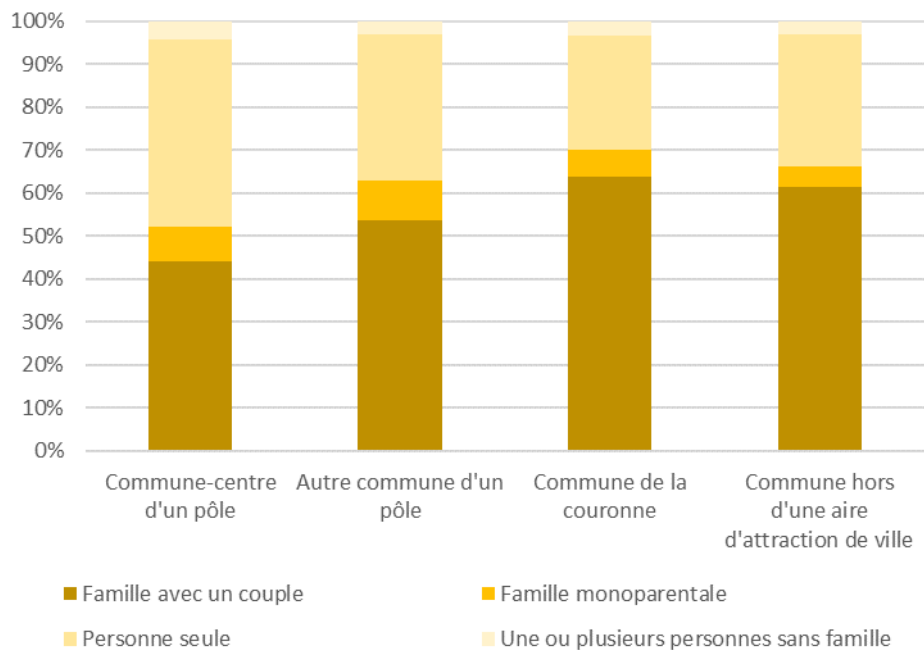
Figure 4 : répartition des ménages par composition familiale selon leur dixième de niveau de vie et leur catégorie de commune

4-A selon leur dixième de niveau de vie



Source : CGDD, modèle Prometheus

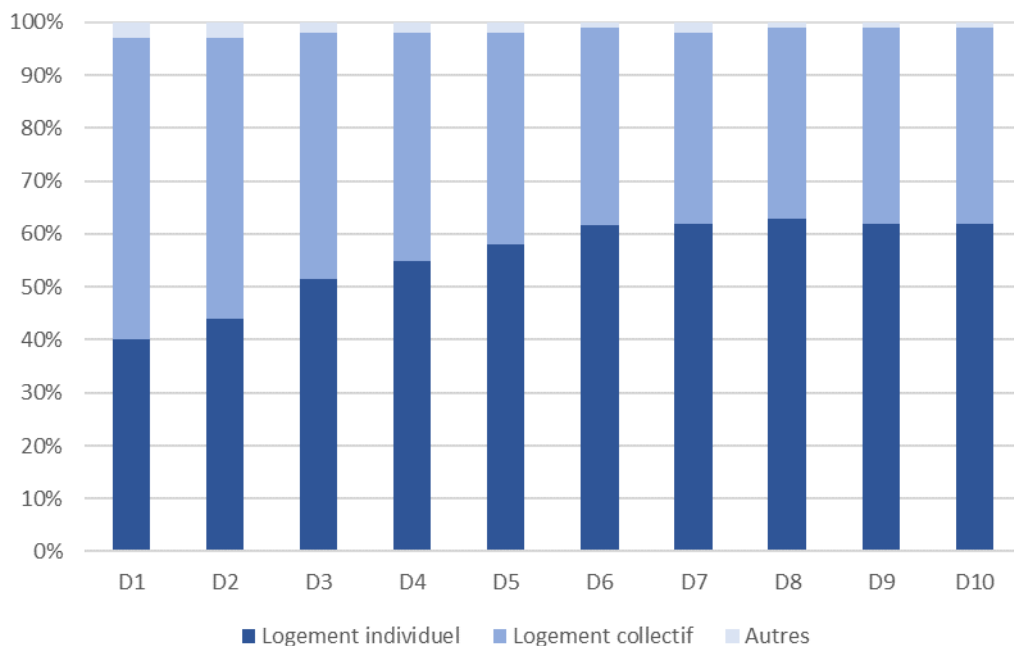
4-B selon leur catégorie de commune



Source : CGDD, modèle Prometheus

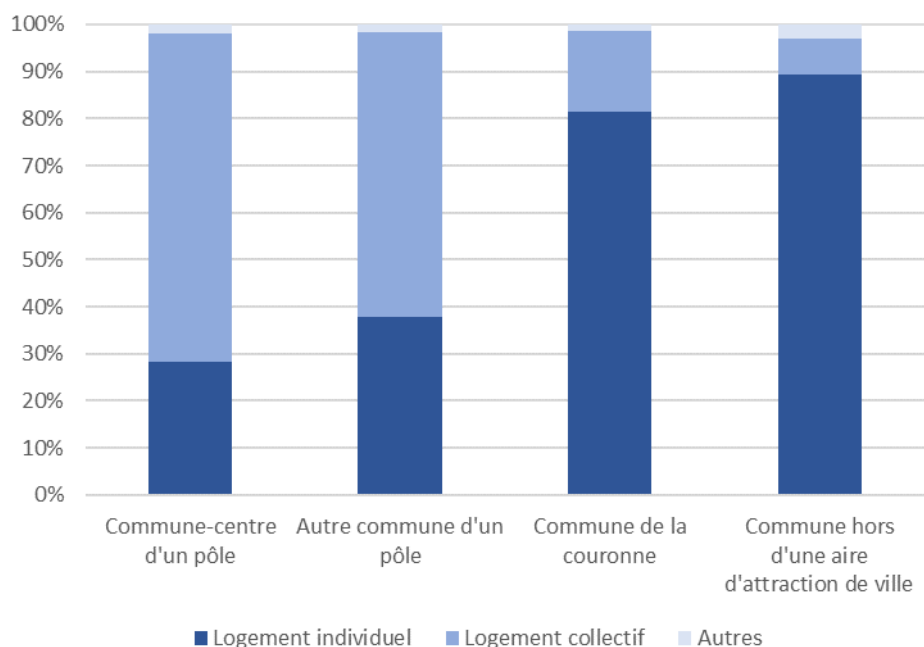
Figure 5 : répartition des ménages par type de logement selon leur dixième de niveau de vie et leur catégorie de commune

5-A selon leur dixième de niveau de vie



Source : CGDD, modèle Prometheus

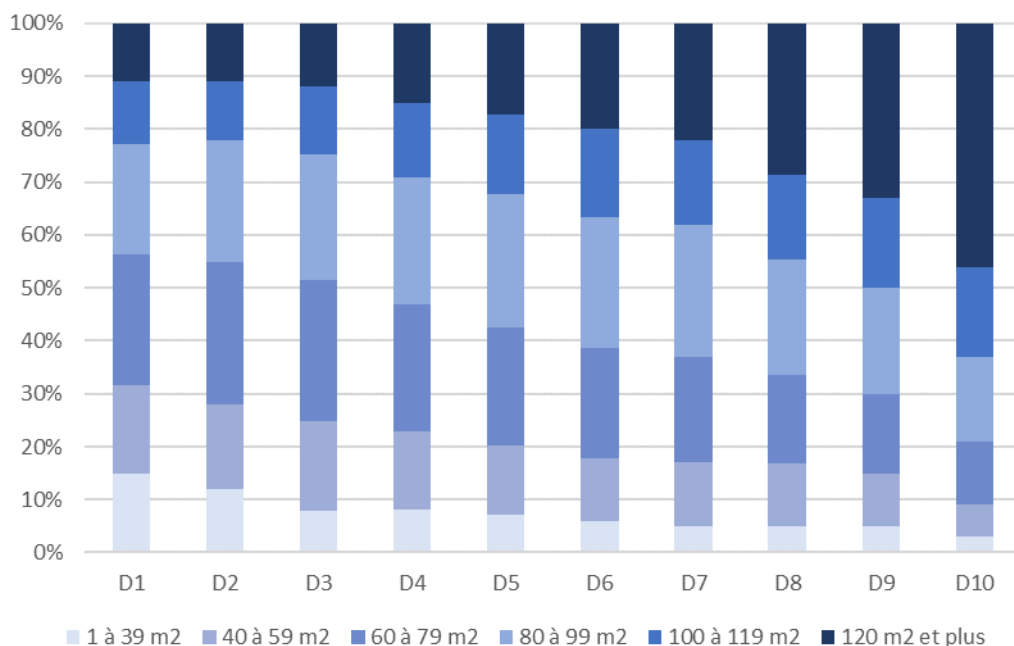
5-B selon leur catégorie de commune



Source : CGDD, modèle Prometheus

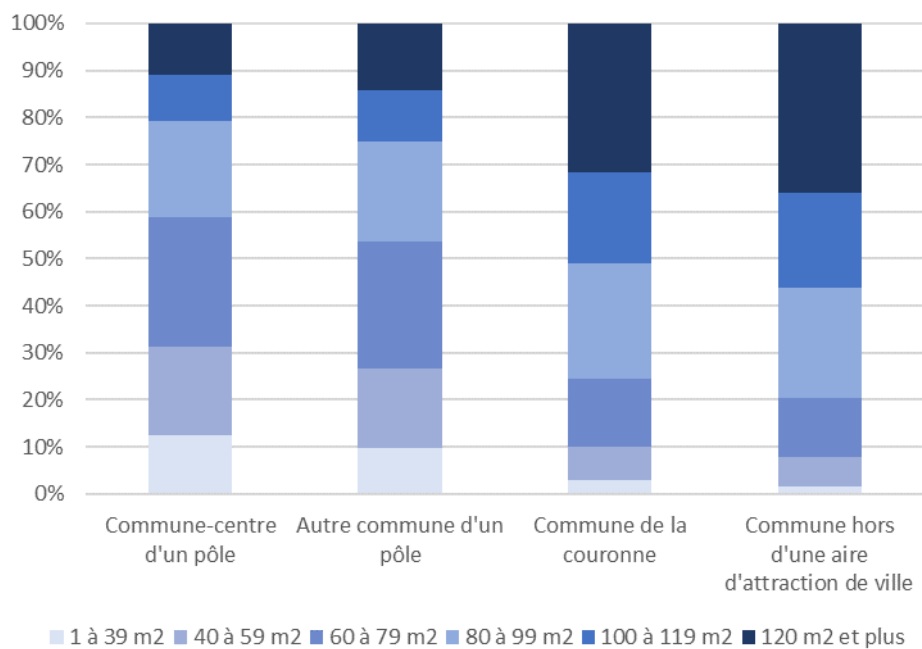
Figure 6 : répartition des ménages par surface du logement selon leur dixième de niveau de vie et leur catégorie de commune

6-A selon leur dixième de niveau de vie



Source : CGDD, modèle Prometheus

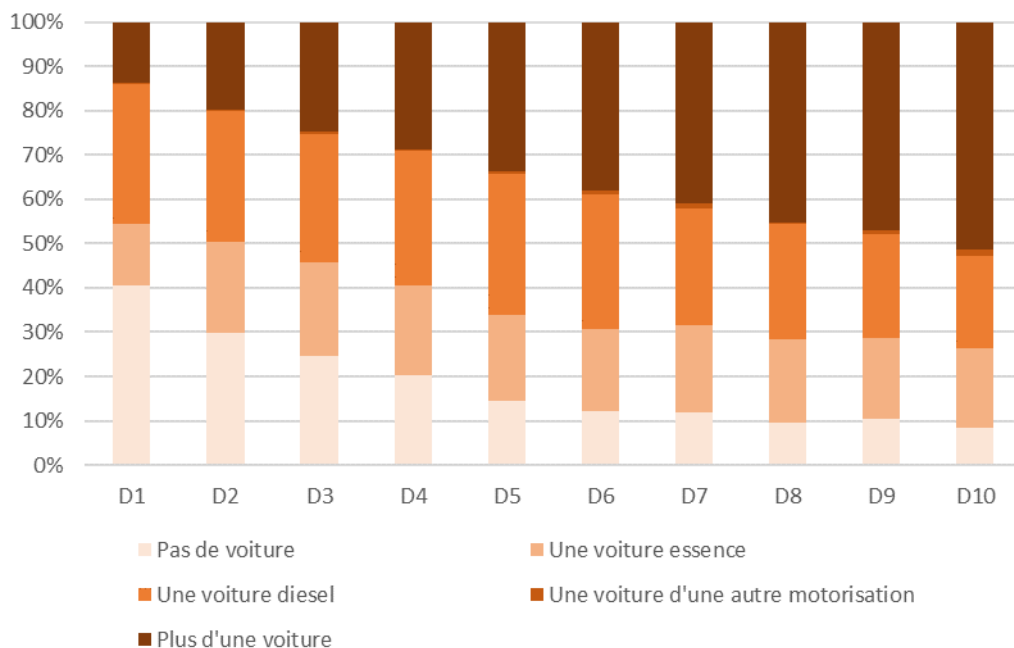
6-B selon leur catégorie de commune



Source : CGDD, modèle Prometheus

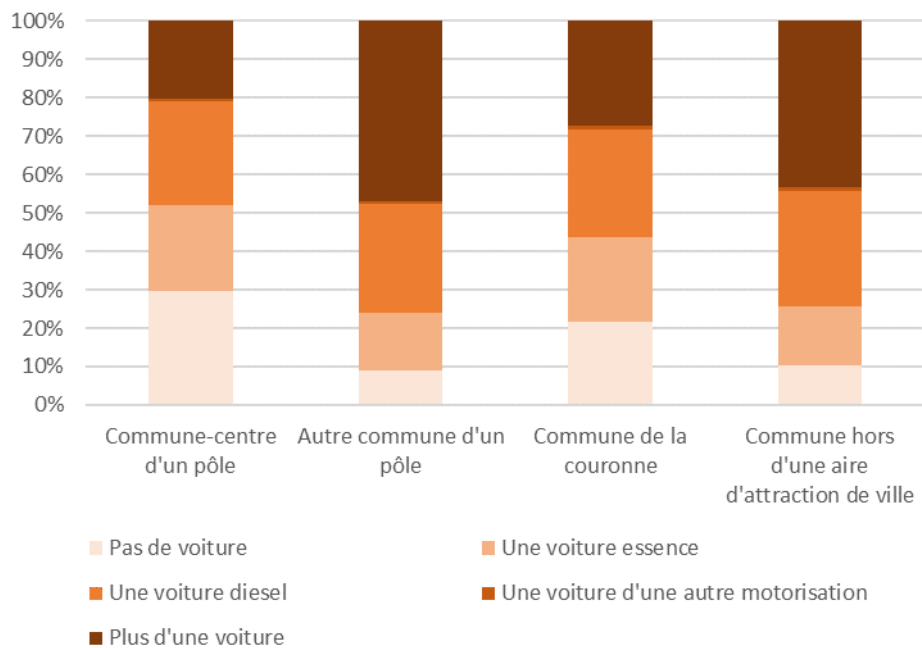
Figure 7 : répartition des ménages par équipement en véhicule selon leur dixième de niveau de vie et leur catégorie de commune

7-A selon leur dixième de niveau de vie



Source : CGDD, modèle Prometheus

7-B selon leur catégorie de commune



Source : CGDD, modèle Prometheus

ANNEXE 18 : TABLEAUX CROISÉS

Tableau 1 : facture énergétique selon la catégorie de commune et le dixième de niveau de vie

Catégorie de commune	Dixième de niveau de vie	Facture totale (en €)	Facture logement (en €)	Facture transport (en €)
Commune-centre d'un pôle	D1	1 669	1 086	583
	D2	1 753	1 074	679
	D3	1 872	1 102	771
	D4	2 078	1 159	919
	D5	2 231	1 201	1 031
	D6	2 141	1 177	964
	D7	2 207	1 174	1 033
	D8	2 415	1 256	1 159
	D9	2 430	1 306	1 125
	D10	2 942	1 587	1 355
	Ensemble	2 149	1 207	943
Autre commune d'un pôle	D1	1 809	1 195	614
	D2	2 034	1 166	868
	D3	2 083	1 222	861
	D4	2 309	1 217	1 092
	D5	2 369	1 220	1 149
	D6	2 398	1 252	1 146
	D7	2 504	1 322	1 182
	D8	2 814	1 406	1 408
	D9	2 859	1 454	1 405
	D10	3 289	1 693	1 596
	Ensemble	2 502	1 337	1 164
Commune d'une couronne	D1	2 466	1 357	1 109
	D2	2 530	1 393	1 137
	D3	2 919	1 408	1 511
	D4	3 068	1 443	1 625
	D5	3 185	1 449	1 736
	D6	3 378	1 486	1 892
	D7	3 492	1 507	1 986
	D8	3 808	1 645	2 163
	D9	3 997	1 752	2 245
	D10	4 570	2 146	2 424
	Ensemble	3 355	1 555	1 800
Commune hors d'une aire d'attraction de ville	D1	2 448	1 326	1 122
	D2	2 454	1 347	1 107
	D3	2 936	1 450	1 486
	D4	3 439	1 526	1 913
	D5	3 615	1 564	2 051
	D6	3 809	1 638	2 171
	D7	3 565	1 652	1 913
	D8	4 188	1 843	2 345
	D9	3 810	1 939	1 871
	D10	4 702	2 357	2 345
	Ensemble	3 349	1 590	1 759

Source : CGDD, modèle Prometheus

Table 2 : taux d'effort énergétique selon la catégorie de commune et le dixième de niveau de vie

Catégorie de commune	Dixième de niveau de vie	TEE totale (en %)	TEE logement (en %)	TEE transport (en %)
Commune-centre d'un pôle	D1	15,28	9,95	5,34
	D2	9,89	6,06	3,83
	D3	8,64	5,08	3,56
	D4	7,79	4,35	3,44
	D5	7,29	3,92	3,37
	D6	6,07	3,34	2,73
	D7	5,48	2,92	2,57
	D8	5,04	2,62	2,42
	D9	4,32	2,32	2,00
	D10	2,75	1,48	1,26
	Ensemble	5,61	3,15	2,46
Autre commune d'un pôle	D1	14,34	9,47	4,87
	D2	9,92	5,69	4,23
	D3	8,08	4,74	3,34
	D4	7,82	4,12	3,70
	D5	7,02	3,61	3,41
	D6	6,29	3,28	3,00
	D7	5,72	3,02	2,70
	D8	5,55	2,77	2,78
	D9	4,57	2,32	2,25
	D10	3,06	1,57	1,48
	Ensemble	5,42	2,90	2,52
Commune d'une couronne	D1	19,69	10,83	8,85
	D2	12,75	7,02	5,73
	D3	11,45	5,52	5,92
	D4	9,96	4,69	5,28
	D5	9,16	4,17	4,99
	D6	8,21	3,61	4,60
	D7	7,55	3,26	4,29
	D8	7,06	3,05	4,01
	D9	6,19	2,72	3,48
	D10	4,13	1,94	2,19
	Ensemble	7,63	3,54	4,09
Commune hors d'une aire d'attraction de ville	D1	18,12	9,82	8,31
	D2	13,36	7,33	6,03
	D3	11,88	5,87	6,01
	D4	11,98	5,31	6,66
	D5	10,01	4,33	5,68
	D6	9,55	4,11	5,44
	D7	8,00	3,71	4,29
	D8	8,33	3,67	4,67
	D9	6,03	3,07	2,96
	D10	4,81	2,41	2,40
	Ensemble	9,33	4,43	4,90

Source : CGDD, modèle Prometheus

Tableau 3 : facture énergétique selon l'énergie principale de chauffage et le dixième de niveau de vie

Énergie principale de chauffage	Dixième de niveau de vie	Facture totale (en €)	Facture logement (en €)	Facture transport (en €)
Gaz	D1	2 016	1 264	752
	D2	2 087	1 238	849
	D3	2 270	1 255	1 015
	D4	2 347	1 261	1 086
	D5	2 580	1 310	1 270
	D6	2 592	1 320	1 271
	D7	2 697	1 369	1 327
	D8	2 947	1 439	1 507
	D9	3 058	1 547	1 511
	D10	3 534	1 797	1 737
	Ensemble	2 619	1 385	1 235
Electricité	D1	1 854	1 054	800
	D2	2 025	1 066	959
	D3	2 376	1 119	1 257
	D4	2 613	1 187	1 426
	D5	2 726	1 178	1 548
	D6	2 787	1 213	1 574
	D7	2 833	1 195	1 637
	D8	3 069	1 295	1 774
	D9	3 054	1 311	1 743
	D10	3 346	1 532	1 814
	Ensemble	2 671	1 213	1 458
Fioul	D1	2 826	1 891	935
	D2	2 895	1 894	1 000
	D3	3 042	1 879	1 163
	D4	3 491	2 023	1 467
	D5	3 670	2 048	1 622
	D6	3 819	2 013	1 806
	D7	3 801	2 053	1 748
	D8	4 130	2 297	1 833
	D9	4 249	2 379	1 870
	D10	4 929	2 776	2 153
	Ensemble	3 721	2 137	1 584
Urbain	D1	1 564	1 006	557
	D2	1 688	1 032	656
	D3	1 662	1 012	650
	D4	1 820	960	860
	D5	1 836	978	858
	D6	1 616	966	649
	D7	1 818	1 044	774
	D8	1 889	1 021	867
	D9	1 991	1 084	906
	D10	2 129	1 338	791
	Ensemble	1 778	1 042	737
Bois	D1	2 186	1 081	1 105
	D2	2 599	1 232	1 367
	D3	2 890	1 124	1 766
	D4	3 504	1 452	2 052
	D5	3 409	1 337	2 073
	D6	3 500	1 313	2 187
	D7	3 496	1 429	2 067
	D8	3 939	1 415	2 524
	D9	3 732	1 503	2 229
	D10	4 376	1 756	2 620
	Ensemble	3 272	1 337	1 936
Autre	D1	2 149	1 201	948
	D2	2 147	1 219	928
	D3	2 667	1 451	1 216

	D4	3 110	1 354	1 756
	D5	3 001	1 438	1 563
	D6	3 267	1 485	1 782
	D7	3 567	1 587	1 979
	D8	3 800	1 699	2 100
	D9	4 040	1 819	2 221
	D10	4 477	2 157	2 319
	Ensemble	3 268	1 553	1 715

Source : CGDD, modèle Prometheus

Tableau 4 : taux d'effort énergétique selon l'énergie principale de chauffage et le dixième de niveau de vie

Énergie principale de chauffage	Dixième de niveau de vie	Facture totale (en %)	Facture logement (en %)	Facture transport (en %)
Gaz	D1	16,07	10,08	5,99
	D2	10,80	6,41	4,40
	D3	9,44	5,22	4,22
	D4	8,22	4,42	3,80
	D5	7,98	4,05	3,93
	D6	6,80	3,47	3,34
	D7	6,19	3,15	3,05
	D8	5,74	2,81	2,94
	D9	4,94	2,50	2,44
	D10	3,16	1,61	1,56
	Ensemble	6,08	3,22	2,87
Electricité	D1	16,97	9,64	7,32
	D2	11,05	5,81	5,23
	D3	9,95	4,68	5,26
	D4	9,08	4,12	4,95
	D5	8,22	3,55	4,67
	D6	7,20	3,13	4,06
	D7	6,61	2,79	3,82
	D8	6,11	2,58	3,53
	D9	5,03	2,16	2,87
	D10	3,37	1,54	1,83
	Ensemble	6,63	3,01	3,62
Fioul	D1	22,11	14,79	7,32
	D2	14,72	9,63	5,08
	D3	12,48	7,71	4,77
	D4	11,62	6,73	4,88
	D5	10,25	5,72	4,53
	D6	9,53	5,02	4,51
	D7	8,41	4,54	3,87
	D8	7,88	4,38	3,50
	D9	6,97	3,90	3,07
	D10	4,41	2,48	1,93
	Ensemble	8,30	4,77	3,53
Urbain	D1	12,55	8,07	4,47
	D2	8,70	5,32	3,38
	D3	7,28	4,43	2,85
	D4	7,04	3,71	3,33
	D5	5,73	3,06	2,68
	D6	4,63	2,77	1,86
	D7	4,45	2,55	1,89
	D8	4,27	2,31	1,96
	D9	3,69	2,01	1,68
	D10	2,02	1,27	0,75
	Ensemble	4,92	2,88	2,04
Bois	D1	16,65	8,23	8,41
	D2	11,78	5,59	6,20
	D3	10,85	4,22	6,63
	D4	10,65	4,41	6,23

	D5	8,99	3,52	5,46
	D6	8,89	3,34	5,56
	D7	7,29	2,98	4,31
	D8	7,21	2,59	4,62
	D9	5,74	2,31	3,43
	D10	4,36	1,75	2,61
	Ensemble	8,11	3,31	4,80
Autre	D1	17,75	9,92	7,83
	D2	11,53	6,55	4,99
	D3	10,13	5,51	4,62
	D4	9,66	4,21	5,45
	D5	8,44	4,05	4,40
	D6	7,82	3,55	4,27
	D7	7,19	3,20	3,99
	D8	6,80	3,04	3,76
	D9	5,97	2,69	3,28
	D10	3,62	1,75	1,88
	Ensemble	6,87	3,27	3,61

Source : CGDD, modèle Prometheus

Tableau 5 : facture énergétique selon l'équipement en véhicules et le dixième de niveau de vie

Type de possession de véhicules	Dixième de niveau de vie	Facture totale (en €)	Facture logement (en €)	Facture transport (en €)
Pas de voiture	D1	1 083	1 083	0
	D2	1 038	1 038	0
	D3	1 085	1 085	0
	D4	1 113	1 113	0
	D5	1 098	1 098	0
	D6	1 044	1 044	0
	D7	1 055	1 055	0
	D8	1 103	1 103	0
	D9	1 088	1 088	0
	D10	1 227	1 227	0
Ensemble	1 084	1 084	0	
Une voiture essence	D1	1 969	1 248	721
	D2	1 984	1 267	716
	D3	2 036	1 268	768
	D4	2 002	1 254	748
	D5	2 018	1 217	801
	D6	2 036	1 227	809
	D7	2 101	1 304	797
	D8	2 297	1 419	878
	D9	2 290	1 450	840
	D10	2 502	1 599	903
Ensemble	2 122	1 324	798	
Une voiture diesel	D1	2 370	1 262	1 108
	D2	2 326	1 200	1 125
	D3	2 522	1 299	1 223
	D4	2 603	1 301	1 302
	D5	2 495	1 314	1 181
	D6	2 445	1 289	1 157
	D7	2 705	1 311	1 394
	D8	2 652	1 355	1 298
	D9	2 805	1 475	1 331
	D10	3 027	1 614	1 413
Ensemble	2 572	1 330	1 243	
Plus d'une voiture	D1	4 046	1 481	2 564
	D2	3 761	1 531	2 230
	D3	3 954	1 444	2 509
	D4	4 202	1 522	2 680
	D5	4 280	1 561	2 719

	D6	4 201	1 588	2 613
	D7	4 107	1 581	2 526
	D8	4 337	1 671	2 666
	D9	4 340	1 741	2 599
	D10	4 809	2 122	2 687
	Ensemble	4 276	1 673	2 603

Note : les ménages possédant une unique voiture d'une motorisation autre que diesel ou essence (électrique, hybride rechargeable ou non rechargeable, GPL, hydrogène...) ne sont pas inclus dans le tableau, car ils sont trop peu nombreux dans l'EMP 2019 pour permettre de réaliser des statistiques robustes.

Source : CGDD, modèle Prometheus

Tableau 6 : taux d'effort énergétique selon l'équipement en véhicules et le dixième de niveau de vie

Type de possession de véhicules	Dixième de niveau de vie	Facture totale (en %)	Facture logement (en %)	Facture transport (en %)
Pas de voiture	D1	12,87	12,87	0,00
	D2	6,71	6,71	0,00
	D3	5,58	5,58	0,00
	D4	4,92	4,92	0,00
	D5	4,09	4,09	0,00
	D6	3,41	3,41	0,00
	D7	2,92	2,92	0,00
	D8	2,70	2,70	0,00
	D9	2,21	2,21	0,00
	D10	1,47	1,47	0,00
	Ensemble	4,35	4,35	0,00
Une voiture essence	D1	15,37	9,74	5,63
	D2	11,28	7,21	4,08
	D3	9,08	5,66	3,42
	D4	7,81	4,89	2,92
	D5	7,09	4,27	2,82
	D6	6,09	3,67	2,42
	D7	5,50	3,42	2,09
	D8	5,17	3,19	1,98
	D9	4,17	2,64	1,53
	D10	2,62	1,67	0,94
	Ensemble	5,71	3,56	2,15
Une voiture diesel	D1	17,38	9,25	8,13
	D2	11,42	5,90	5,53
	D3	10,21	5,26	4,95
	D4	9,13	4,56	4,57
	D5	7,73	4,07	3,66
	D6	6,64	3,50	3,14
	D7	6,73	3,26	3,47
	D8	5,57	2,84	2,72
	D9	4,80	2,52	2,27
	D10	3,18	1,69	1,48
	Ensemble	6,90	3,57	3,33
Plus d'une voiture	D1	22,29	8,16	14,13
	D2	15,40	6,27	9,13
	D3	13,25	4,84	8,41
	D4	11,42	4,14	7,28
	D5	10,57	3,85	6,71
	D6	9,22	3,49	5,74
	D7	7,94	3,06	4,89
	D8	7,40	2,85	4,55
	D9	6,32	2,54	3,79
	D10	3,96	1,75	2,21
	Ensemble	7,39	2,89	4,50

Note : les ménages possédant une unique voiture d'une motorisation autre que diesel ou essence (électrique, hybride rechargeable ou non rechargeable, GPL, hydrogène...) ne sont pas inclus dans le tableau, car ils sont trop peu nombreux dans l'EMP 2019 pour permettre de réaliser des statistiques robustes.

Source : CGDD, modèle Prometheus

BIBLIOGRAPHIE

- André M., Biotteau A-L. et Duval J. (2016) : « [Module de taxation indirecte du modèle Ines](#) », Document de travail, *Insee*
- Ben Jelloul M. et Douenne T. (2016) : « [La TCIFE dans le module de fiscalité indirecte de TAXIPP](#) », Guide méthodologique, *IPP*
- Bessis F. et Cotton P. (2021) : « [La réforme, le chiffrage, son modèle et ses données - Les évolutions du monopole de l'expertise économique au prisme d'un instrument de microsimulation de la législation socio-fiscale](#) », *Politix n° 134*
- Blanchet D., Buffeteau S., Crenner E. et Le Minez S. (2011) : « [Le modèle de microsimulation Destinie 2 : principales caractéristiques et premiers résultats](#) », Économie et Statistique, *Insee*
- Blanchet D., Hagneré C., Legendre F. et Thibault F. (2015) : « [Microsimulations statique et dynamique appliquées aux politiques fiscales et sociales : modèles et méthodes - Introduction](#) », Économie et Statistique, *Insee*
- Blanchet D., Hagneré C. Legendre F. et Thibault F. (2016) : « [Évaluation des politiques publiques, ex post et ex ante : l'apport de la microsimulation – Introduction](#) », *Revue économique*
- Blanchet D. (2020) : « [Des modèles de microsimulation dans un institut de statistique – Pourquoi, comment, jusqu'où ?](#) », Courrier des statistiques N4, *Insee*
- Boccanfuso D., Estache A. et Savard L. (2018) : « [Intra-Country Distributional Impact of Policies to Fight Climate Change: A Survey](#) », *The Journal of Development Studies*
- Bourguignon F. et Spadaro A. (2006) : « [Microsimulation as a tool for evaluating redistribution policies](#) », *The Journal of Economic Inequality*
- Bourguignon F., Bussolo M. et Pereira da Silva L. A. (2008) : « [The impact on macro economics policies on poverty and income distribution – Macro-micro evaluation techniques and tools](#) », *The World Bank*
- Bourguignon F. et Landais C. (2022a) : « [La micro-simulation de la redistribution dans les économies avancées](#) », Focus n° 87, *Conseil d'analyse économique*.
- Bourguignon F. et Landais C. (2022b) : « [Micro-simuler l'impact des politiques publiques sur les ménages : pourquoi, comment et lesquelles ?](#) », Note n° 74, *Conseil d'analyse économique*.
- Calvet L. et Marical F. (2011) : « [Consommation de carburant : effets des prix à court et à long terme par type de population](#) », Économie et Statistique, *Insee*
- Callonec G. et Cancé R. (2022) : « [Évaluation macroéconomique de la Stratégie nationale bas-carbone \(SNBC2\) avec le modèle ThreeME](#) », Document de travail, *CGDD*
- Capéau B., Decoster A. et Phillips D. (2014) : [Handbook of Microsimulation Modelling – Chapter 8 "Consumption and Indirect Tax Models"](#), O'Donoghue C. (éd.), *Emerald Group Publishing Limited*
- Casler S.D. et Rafiqui A. (1993) : « [Evaluating fuel tax equity: direct and indirect distributional effects](#) », *National Tax Journal*
- Chai A., Bonner S., Tennet S., Ratnasiri S. et Wagner L. (2021) : « [The impact of rising energy prices on energy poverty in Queensland: a microsimulation exercise](#) », *Economic Analysis and Policy*
- Chaton C. et Gouraud A. (2019) : « [Simulation of fuel poverty in France](#) », *Energy Policy*
- Cheloudko P. et Martin H. (2020) : « [Une décennie de modélisation du système de retraite : la genèse du modèle de microsimulation TRAJECTOIRE](#) », Courrier des statistiques N4, *Insee*

- Cochez N., Durieux É. et Lévy D. (2015) : « Vulnérabilité énergétique-Loin des pôles urbains, chauffage, carburant pèsent fortement dans le budget », *Insee*
- Cockburn J., Savard L. et Tiberti L., (2014): *Handbook of Microsimulation Modelling – Chapter 9 “Macro-Micro Models”*, O’Donoghue C. (éd.), Emerald Group Publishing Limited
- Cournuet F. et Sicsic M. (2020) : « Effets d’une variation de transferts socio-fiscaux sur le niveau de vie et les inégalités », *Insee*
- Cornuet F. (2022) : « Entre janvier 2021 et juin 2022, la hausse des prix de l’énergie a entraîné une perte de pouvoir d’achat, malgré la mise en œuvre des mesures exceptionnelles », *Insee*
- David M-G., Lhommeau B. et Starzec C. (1999) : « Le Modèle de Microsimulation INES », Documents de travail, *Insee*
- De Blander R., Schockaert I., Decoster A. et Deboosere P. (2013): “The impact of demographic change on policy indicators and reforms”, *Flemosi Discussion Paper DP25, 1-44*
- Dekkers G. (2015): “The simulation properties of microsimulation models with static and dynamic ageing – a brief guide into choosing one type of model over the other”, *International Journal of Microsimulation*
- De Lauretis S. (2017) : « Modélisation des impacts énergie/carbone de changements de modes de vie. Une prospective macro-micro fondée sur les emplois du temps », *Université Paris-Saclay* (PhD Thesis)
- Devalière I. (2009) : « De l’inconfort thermique à la précarité énergétique, profils et pratiques des ménages pauvres », *Informations sociales*
- Devalière I., Briant P. et Arnault S. (2011) : « La précarité énergétique : avoir froid ou dépenser trop pour se chauffer », *Insee*
- Devalière I., Clément M. et Dussud F-X. (2018) : « Suivi annuel de la précarité énergétique : un nouvel enjeu pour l’ONPE », *Ademe-CGDD*
- Douenne T. (2017) : « Documentation sur l’appariement des enquêtes Budget de Familles, Enquête Logement et Enquête Nationale Transports et Déplacements »
- Douenne T. (2020): “The vertical and horizontal distributive effects of energy taxes: a case study of a French policy”, *The Energy Journal*
- Fredon S. et Sicsic M. (2020) : « INES, le modèle qui simule l’impact des politiques sociales et fiscales », *Courrier des statistiques N4, Insee*
- Fontaine M. et Sicsic M. (2016) : « L’effet d’une variation du montant de certains transferts du système socio-fiscal sur le niveau de vie : résultats sur 2016 à partir du modèle de microsimulation Ines (Cahier de variantes) », Document de travail, *Insee*
- Fouquet M. (2021) : « Le modèle de micro-simulation Ines-Omar 2017 – Note méthodologique détaillée », *Les dossiers de la Drees, Drees*
- Gantois T., Girard P-L et Le Gall C. (2022) : « Évaluation de l’impact macroéconomique de la transition écologique : revue des modèles macro-environnementaux, usages et limites », Documents de travail, *DG Trésor*
- Gay P. W. et Propos J. L. (1993): “Carbon dioxide production by the UK economy: An input-output assessment”, *Applied Energy*
- Hills J. (2011): “Fuel poverty – The problem and its measurement. Interim report of the fuel poverty review”, *Centre for Analysis of Social Exclusion*
- Hills J. (2012): “Getting the measure of fuel poverty. Finale report of the fuel poverty review”, *Centre for Analysis of Social Exclusion*

- Immervoll H., Lindström K., Mustonen E., Rihela M. et Vitamäki H. (2005): "Static ageing techniques: Accounting for population changes in tax-benefit microsimulation models", *EUROMOD Working Paper*
- Khamsing W. T., Ceci-Renaud N. et Guillot L. (2016) : « Simuler l'impact social de la fiscalité énergétique : le modèle Prometheus », Étude & documents, *CGDD*
- Klevmarken A. (2022): "Should We Invest in Microsimulation Models?", *International Journal of Microsimulation*
- Klevmarken A. (2022): "Microsimulation. A Tool for Economic Analysis", *International Journal of Microsimulation*
- Jingjing L., O'Donoghue C. et Loughrey J. (2014): *Handbook of Microsimulation Modelling – Chapter 3 "Static Models"*, O'Donoghue C. (éd.), *Emerald Group Publishing Limited*
- Landais C., Piketty T. et Saez E. (2011) : *Pour une révolution fiscale*, édition *Le Seuil*
- Lardellier R., Legal R., Raynaud D. et Vidal G. (2011) : « Un outil pour l'étude des dépenses de santé et des « restes à charge » des ménages : le modèle OMAR », *Économie et statistiques, Insee*
- Leulescu, A. et Agafitei M. (2013): "Statistical matching: a model based approach for data integration", *Eurostat-Methodologies and Working papers*
- Lengendre F. (2004) : « Micro-simulation et évaluation des politiques économiques et sociale : un panorama des développements récents en France », *Revue d'économie politique*
- Legendre F. (2019) : « L'émergence et la consolidation des méthodes de microsimulation en France », *Économie et Statistique, Insee*
- Li J. et O'Donoghue C. (2013): "A survey of dynamic microsimulation models: uses, model structure and methodology", *International Journal of Microsimulation*
- Li S., Liu Y. et Wei C. (2022): "The cost of greening stimulus: A dynamic analysis of vehicle scrappage programs", *International Economic Review*
- Marchal J. (2008) : « Modélisation des performances énergétiques du parc de logement. Etat énergétique du parc en 2008 », *Agence nationale de l'habitat*
- Maresca B. (2013) : « La précarité énergétique pose la question du coût du logement en France », *Consommation et modes de vie, CREDOC*
- Martin E. (2024) : « La précarité énergétique diminue en 2022 grâce aux mesures de soutien aux ménages et aux baisses de consommation », *Théma Essentiel, CGDD*
- McClelland R., Khitatrakun S. et Lu C. (2020): "Estimating Confidence Intervals in a Tax Microsimulation Model", *International Journal of Microsimulation*
- Newhouse J-P. (1993): "Free for All? Lessons from the RAND experiment", *Harvard University Press*
- Nicod. Q, Pinton C. et Quille B. (2023) : « L'impact distributif des mesures de soutien aux ménages face à la hausse des prix de l'énergie en 2022 », *Théma Essentiel, CGDD*
- Nicolas J-P., Vanco F. et Verry D. (2012) : « Mobilité quotidienne et vulnérabilité des ménages », *Revue d'économie régionale et urbaine*
- O'Donoghue C. (1997): "Carbon dioxide, energy taxes and household income", *ESRI working paper*
- O'Donoghue C. (2001): "Dynamic Microsimulation: A Methodological Survey", *Brazilian Electronic Journal of Economics*
- O'Donoghue C. et Loughrey J. (2014): "Nowcastings in Microsimulation Models: A Methodological Survey", *Journal of Artificial Societies and Social Simulation*

- O'Donoghue, C. (2014): [Handbook of microsimulation modelling](#), Emerald Group Publishing
- Okner B. (1972): [Annals of Economic and Social Measurement – Chapter “Constructing a New Data Base from Existing Microdata”](#), NBER
- D’Orazio M., Di Zio M., et Scanu M. (2006): [“Statistical matching: Theory and practice”](#), John Wiley & Sons
- D’Orazio M. (2017): [“Statistical matching and imputation of survey data with statmatch”](#), Italian National Institute of Statistics
- Orcutt G. H. (1957): [“A new type of socio-economic system”](#), Review of Economics and Statistics
- Paass, G. (1986): Microanalytic Simulation Models to Support Social and Financial Policy - [“Statistical match: evaluation of existing procedures and improvements by using additional information”](#) (p.401-420)
- Rausch S. et Schwarz G. (2016): [“Household heterogeneity, aggregation, and the distributional impacts of environmental taxes”](#), Journal of Public Economics
- Ravigné E., Ghersi F. et Nadaud F. (2022): [“Is a fait energy transition possible ? Evidence from the French low-carbon strategy”](#), Ecological Economics
- Rosenbaum P. R. et Donald R. (1983): [“The central role of the propensity score in observational studies for causal effects”](#), Biometrika
- Rubin D.B. (1986): [“Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations”](#), Journal of Business and Economic Statistics
- Ruiz N. et Trannoy A. (2008) : « [Le caractère régressif des taxes indirectes : les enseignements d’un modèle de microsimulation](#) », Économie et statistiques, Insee
- Schiraldi P. (2011): [“Automobile replacement: a dynamic structural approach”](#), the RAND journal of economics
- Serret Y. et Johnstone N. (2006): [“The Distributional Effects of Environmental Policy”](#), OCDE
- Symons E., Propos J. Et Gay P. (1994): [“Carbon taxes, consumer demand and carbon dioxide emissions: A simulation analysis for the UK”](#), Fiscal Studies
- Yusuf A. et Resosudarmo B. (2015): [“On the Distributional Effect of Carbon Tax in Developing Countries: The Case of Indonesia”](#), Environmental Economics and Policy Studies
- Vandyck T. (2013): [“Efficiency and equity aspects of energy taxation”](#), EUROMOD Working Paper
- de Waal T. (2015): [“Statistical matching: Experimental results and future research questions”](#), Statistics Netherlands
- Zhang W., Robinson C., Guhathakurta S., Garikapti V. M., Dilkina B., Brown M. A. et Pendyala R. M. (2018): [“Estimating Residential Energy Consumption in Metropolitan Areas: A Microsimulation Approach”](#), Energy

SIGLES

Ademe	Agence de l'environnement et de la maîtrise de l'énergie
BDF	Budget de famille
Ceren	Centre d'étude et de recherches économiques sur l'énergie
Cired	Centre international de recherche sur l'environnement et le développement
Cnaf	Caisse nationale des allocations familiales
Cnav	Caisse nationale d'assurance vieillesse
CNIS	Conseil national de l'information statistique
CO₂	Dioxyde de carbone
CRE	Commission de régulation de l'énergie
CSL	Comptes satellite du logement
DGEC	Direction générale de l'énergie et du climat
DGFIP	Direction générale de Finances publiques
DJU	Degrés jours unifiés
Drees	Direction de la recherche, des études, de l'évaluation et des statistiques
EMP	Enquête mobilité des personnes
ENL	Enquête nationale Logement
ENTD	Enquête nationale transports et déplacements
ERFS	Enquête revenus fiscaux et sociaux
Fidéli	Fichier démographique sur les logements et les individus
GPL	Gaz de pétrole liquéfié
Ines	Insee-Drees
IPC	Indice des prix des prix à la consommation
IPP	Institut des politiques publiques
OFCE	Observatoire français des conjonctures économiques
OMAR	Outil de microsimulation pour l'analyse des restes à charge
ONPE	Observatoire national de la précarité énergétique
Phébus	Enquête performance de l'habitat, équipements, besoins et usages de l'énergie
RP	Recensement de la population
RSVERO	Répertoire statistique des véhicules routiers
SNBC	Stratégie nationale bas-carbone
SRCV	Statistiques sur les ressources et les conditions de vie
SSP	Service statistique public
TRV	Tarifs réglementés de vente
VUL	Véhicule utilitaire léger

TABLE DES MATIÈRES

I. REVUE DE LITTÉRATURE.....	8
A. Objectifs et typologie des modèles de microsimulation	8
B. La microsimulation énergétique.....	12
L'utilisation de modèles de microsimulation spécialisés sur l'énergie.....	13
L'utilisation de modèles de microsimulations généralistes.....	17
L'utilisation combinée de modèles de microsimulations et macroéconomiques	18
II. PRÉSENTATION DU MODÈLE PROMETHEUS	20
A. Un modèle de microsimulation statique avec vieillissement	20
B. L'architecture du modèle	20
Le traitement, puis le pseudo-appariement de l'ENL 2013 et de l'EMP 2019.....	21
La microsimulation : vieillissement et application du système de prix.....	21
Le calcul des factures énergétiques des ménages.....	23
III. BASES DE DONNÉES D'ENQUÊTES MOBILISÉES	25
A. L'enquête nationale logement de 2013	25
Champ de l'enquête et échantillonnage	25
Points d'attention méthodologiques.....	26
B. L'enquête mobilité des personnes de 2019	26
Champ de l'enquête et échantillonnage	27
Points d'attention méthodologiques.....	27
C. Comparaison avec d'autres enquêtes.....	28
Comparaison avec l'enquête budget de famille	28
Comparaison avec l'enquête performance de l'habitat, équipements, besoins et usages de l'énergie (Phébus)	29
Une approche fine et cohérente.....	30
IV. TRAITEMENT DES BASES.....	31
A. Traitements de l'enquête nationale logement 2013	31
Champ et objectifs.....	31
Préalable : import des tables.....	32
Premier sous-bloc : traitement des énergies	32
Redressement des factures énergétiques du logement.....	35
Calcul des consommations d'énergie de logement.....	36
B. Les revenus.....	39
C. Traitements de l'enquête mobilité des personnes.....	39
Champ et objectifs.....	39
Import des tables et agrégation des données au niveau ménage.....	40
Traitement des tables véhicules.....	41
D. Traitement des tables individus et ménages	42

V. PSEUDO-APPARIEMENT	43
A. Brève revue de littérature et choix de la méthode	43
Quelques éléments théoriques sur l'appariement statistique	43
Considérations pratiques	45
Choix méthodologiques nécessaires à la réalisation d'un appariement statistique	45
B. Mise en œuvre pratique du pseudo-appariement.....	47
Présentation globale.....	47
Sélection des variables de stratification	47
Mise en œuvre du <i>hot deck</i> aléatoire au sein de chaque strate	50
Jointure des tables	50
C. Analyse de la qualité de l'appariement	50
Statistiques sur les strates d'imputation.....	50
Contrôles sur les distributions des variables d'intérêt imputées.....	51
Source : CGDD, modèle Prometheus	51
Préservation des relations de dépendances entre les variables	53
VI. VIEILLISSEMENT ANNUEL	56
A. Logement	56
Vieillessement du parc de logement	56
Vieillessement des consommations énergétiques unitaires des logements	57
Correction des conditions météorologiques	58
B. Transport.....	59
Données sources.....	59
La diésélisation ou esencialisation.....	60
Les calages.....	61
C. Revenus	62
VII. MISE À JOUR DE LA FISCALITÉ	63
A. La fiscalité dans Prometheus.....	63
Typologie.....	63
Sources	64
B. Électricité	64
Actualisation des prix.....	64
Calculs des factures.....	65
C. Gaz naturel	65
Actualisation des prix.....	65
Calculs des factures.....	66
D. Autres énergies de logement	66
Fioul.....	66
Bois	67
Chauffage et réseau de chaleur urbain	67
GPL logement	68
E. Les carburants.....	68

VIII. SYNTHÈSE ET UTILISATION PRATIQUE DU MODÈLE	69
A. Architecture du modèle.....	69
B. Actualisation des fichiers d'inputs.....	69
Le fichier de calage et vieillissement	70
Le fichier de prix et fiscalité.....	70
C. Actualisation du programme de lancement.....	70
Les dates de paramètres.....	70
Les paramètres globaux.....	70
IX. CAHIER DE VARIANTES.....	71
A. La facture énergétique des ménages en 2019 : estimation et analyse descriptive.....	72
Niveaux moyens.....	72
Selon le niveau de vie.....	74
Selon la catégorie de commune.....	78
B. Hausse du prix TTC des énergies	82
Hausse des prix TTC des carburants de 10 %.....	82
Hausse des prix TTC du gaz de 10 %.....	84
Hausse des prix TTC de l'électricité de 10 %.....	86
C. Hausse de fiscalité énergétique	88
Hausse de la composante carbone de 10 €/tCO ₂	88
Hausse de l'accise sur l'électricité de 10 €/MWh.....	90
Hausse de la TVA sur les prix de l'énergie d'un point de pourcentage.....	92
D. Aides au paiement de la facture énergétique	94
Chèque énergie classique.....	95
Chèque énergie classique augmenté d'un chèque additionnel de 100 €.....	97
ANNEXES	99
Annexe 1 : décomposition de la variable mode de chauffage en 17 modalités	99
Annexe 2 : attribution de l'Énergie principale de chauffage.....	100
Annexe 3 : variables et résultats des régressions de l'imputation hot deck	101
Annexe 4 : description et prix PÉGASE utilisés pour calculer les consommations de fioul GPL, bois et de chauffage urbain à partir de l'ENL en 2013.....	103
Annexe 5 : informations utilisées pour calculer les consommations de gaz.....	104
Annexe 6 : hypothèses, critères et informations utilisées pour calculer les consommations d'électricité.....	105
Hypothèses utilisées pour l'imputation des tarifs d'électricité	105
Critères utilisés pour l'imputation de la puissance d'électricité et du contrat à chaque ménage	105
Annexe 7 : les revenus dans l'ENL 2013.....	107
Annexe 8 : procédure de sélection avec arbre de régression	108
Annexe 9 : résultat de la régression.....	109
Annexe 10 : homogénéisation des variables	110

Annexe 11 : valeurs des corrélations	112
Annexe 12 : valeurs des corrélations	113
Annexe 13 : nomenclature RP (Insee) utilisée dans Prometheus	113
Annexe 14 : précisions sur les calculs des consommations.....	115
Annexe 15 : estimations et détails des modèles logistiques de transition.....	116
Sélection des variables.....	116
Les modèles	117
Annexe 16 : statistiques univariées	119
Annexe 17 : statistiques bivariées	122
Annexe 18 : tableaux croisés	128
Bibliographie.....	134
Sigles	138
Table des matières	139



**MINISTÈRES
AMÉNAGEMENT
DU TERRITOIRE
TRANSITION
ÉCOLOGIQUE**

*Liberté
Égalité
Fraternité*

**Commissariat général
au développement durable**

Service de l'économie verte et solidaire (SEVS)

Sous-direction de l'économie et de l'évaluation

Tour Séquoia – 92055 La Défense cedex

Courriel : diffusion.cgdd@developpement-durable.gouv.fr

www.ecologie.gouv.fr