
Observer la rénovation énergétique par des enquêtes statistiques : l'exemple de Tremi 2020

Mathieu BAUDRY (), Thomas MERLY-ALPA (**), Jean-Philippe RATHLE (***)*

() MESRI-SIES, Département des études statistiques de l'enseignement supérieur*

*(**) Ined, Service des enquêtes et sondages*

*(***) SDES, Sous-direction des statistiques de l'énergie*

`jean-philippe.rathle@developpement-durable.gouv.fr`

Mots-clés. (6 maximum) : Rénovation énergétique, non-réponse, collecte, confinement, enquête, données administratives.

Domaines. redressement des données, collecte.

Résumé (entre 350 et 900 mots environ)

La politique énergétique a pris de l'ampleur depuis plus de cinquante ans, et en particulier depuis les chocs pétroliers des années 1970. En 2015, la France s'est engagée dans une Stratégie nationale bas carbone (SNBC), destinée à diminuer la consommation d'énergies fossiles. Cette politique implique une amélioration des performances énergétiques des logements, en particulier *via* la baisse de leurs émissions de gaz à effet de serre.

La mesure de ses effets, dont l'observation de la rénovation énergétique des logements, s'effectue à la fois au travers de sources administratives et d'enquêtes. À ce jour, cela repose principalement sur les données fiscales concernant les aides à la rénovation perçues par les ménages (notamment le CITE – crédit d'impôt pour la transition énergétique - et les CEE – certificats d'économies d'énergie), ainsi que sur les enquêtes Tremi (Travaux de rénovation énergétique dans les maisons individuelles) et Treco (Travaux de rénovation énergétique dans le collectif).

L'objet de cette publication est de présenter le contexte dans lequel l'enquête Tremi 2020 a été menée, la façon dont son terrain s'est déroulé, les traitements post-collecte effectués puis la diffusion de ses résultats.

Cette enquête vise à collecter des informations sur les travaux de rénovation permettant d'améliorer la performance énergétique des maisons individuelles, les ménages les ayant réalisés et leur avis sur les travaux. Par un questionnement sur les leviers et freins à la rénovation, ainsi que la connaissance des différentes aides, l'enquête contribue à évaluer l'impact des politiques d'encouragement à la rénovation énergétique des maisons.

L'enquête s'est déroulée du 9 janvier au 14 avril 2020, principalement en ligne et *via* un questionnaire papier, pour les ménages n'ayant pas accès au questionnaire par Internet. 100 000 ménages en maisons individuelles ont été enquêtés, pour un taux de réponse de 53 %. Les mesures restrictives dues à la crise sanitaire liée à l'épidémie de Covid-19 ont sans doute eu un impact

positif, le taux de réponse ayant augmenté de 5 points de pourcentage la semaine suivant l'annonce du confinement de mars 2020.

Le tirage de son échantillon a été réalisé par l'Insee dans le fichier Fidéli issu des sources fiscales, et stratifié sur un certain nombre de variables essentielles pour la diffusion des résultats (régions, date de construction, notamment). Cependant, la non-réponse totale n'est pas uniformément répartie sur la population interrogée et les résultats bruts de l'enquête n'étaient donc pas représentatifs de la population.

En particulier, on peut s'interroger sur le risque que la non-réponse soit, dans notre situation, très liée à la thématique de l'enquête ; ce que l'on appelle non-réponse non ignorable (NMAR). En effet, on peut supposer que des ménages ayant effectué des travaux de rénovation, qui plus est à une échéance proche de leur interrogation, sont plus enclins à répondre à un questionnaire portant sur ce thème.

Afin de corriger ce problème, une repondération en deux temps a été effectuée.

Tout d'abord, une modélisation du taux de réponse a été effectuée, en utilisant une régression logistique puis une repondération par groupe de réponse homogène (GRH) a été mobilisée. L'indicatrice de réponse est modélisée de la façon suivante :

$$i_{\text{rep}} \sim \text{typologie_ménage} + \text{tranche_année_construction} + i_{\text{locataire}} + i_{\text{CITE ou CEE}}$$

où les indicatrices CITE et CEE correspondent à la perception par le ménage entre 2017 et 2019 d'aides à la rénovation ; ces informations ont été reconstruites sur l'ensemble de l'échantillon en mobilisant des millésimes ultérieures des données fiscales (pour le CITE) et par un appariement statistique à l'adresse (pour le CEE). L'intégration de ces variables conduit à une évolution sensible de l'estimation du nombre total de gestes de rénovation (- 2,7 %), et du montant total des travaux (- 4,3 %).

Ensuite, un calage sur marges a été opéré, afin de finaliser la repondération de l'échantillon et coller à des totaux annexes. Il a été réalisé sur quelques variables structurelles, pour lesquelles on dispose des totaux calculés à partir du fichier Fidéli, et que l'on connaît pour l'ensemble des répondants. En complément, d'autres variables spécifiques à notre thématique ont été intégrées : après avoir testé un grand nombre de variables, des indicatrices sur le bénéfice d'un crédit d'impôt à la rénovation (CITE) ou d'un certificat d'économies d'énergie (CEE) ont été retenues, ainsi que l'année de construction du logement ou le statut d'occupation du ménage.

Les résultats de l'enquête ont été publiés en mai 2021 *via* l'Observatoire national de la rénovation énergétique (ONRE), une instance regroupant différents acteurs tels que le Ministère de la transition énergétique et l'Ademe. Un éclairage a été effectué sur les gains énergétiques associés aux travaux de rénovation, sur les travaux pour lesquels les ménages n'ont pas bénéficié d'aides ainsi que sur les aspects qualitatifs, qui ne sont pas renseignés par les statistiques sur les aides perçues par les ménages.

Cette enquête sera par la suite complétée par l'enquête Treco, qui étudiera pour sa part les travaux de rénovation énergétique effectués dans les immeubles et les copropriétés.

Abstract en Anglais (Texte de 5 à 10 lignes)

France, as many other countries, is engaged to reduce energy consumption and warming emissions. Upgrading the energy efficiency of existing buildings is one of the key components of French strategy to achieve these goals. Measuring the dynamics of renovation is crucial; as such, French officials launched a national observatory on these subjects. This observatory relies mainly on administrative data to measure the rate of upgrading each year, but these data cannot cover the whole scope. For instance, some people might repair their homes without asking for financial help from the government. To improve these indicators, we conduct an household surveys, Tremi, on the scope of individual houses. This communication presents some challenges faced while designing and conducting this survey, such as NMAR non-response treatment.

Introduction

Le plan de rénovation énergétique des bâtiments (PREB), promulgué le 26 avril 2018 réaffirme les ambitions de la France en matière de rénovation énergétique. Il fixe pour objectifs, entre autres, de :

- Rénover l'ensemble du parc de bâtiments au niveau bâtiment basse consommation d'ici 2050 ;
- Éliminer les 7 à 8 millions de passoires thermiques¹ du parc privé d'ici 2025 (logements classés en étiquettes F et G du Diagnostic de performance énergétique), et rénover 500 000 logements par an ;
- Lutter contre la précarité énergétique en concentrant ses efforts pour éradiquer les passoires thermiques habitées par des ménages propriétaires aux faibles revenus en visant la rénovation de 150 000 logements par an dès 2018.

Suivre de tels objectifs nécessite d'être en mesure de qualifier le parc de logement et d'évaluer finement les dynamiques de rénovation. À cette fin, le Commissariat général au développement durable (CGDD) a été missionné pour la mise en place d'un Observatoire de la rénovation énergétique, visant à produire un tableau de bord d'indicateurs assurant le suivi des objectifs de la PREB et plus généralement les dynamiques de rénovation des logements, individuels et collectifs, et des locaux à usage tertiaire.

L'alimentation de ce tableau de bord repose en grande majorité sur les sources administratives liées aux aides à la rénovation :

- Le CITE (Crédit d'impôt pour la transition énergétique), progressivement remplacé par « Ma prime renov » ;
- Les CEE (Certificats d'économie d'énergie) ;
- Les aides dédiées de l'Anah telles que « Habiter mieux ».

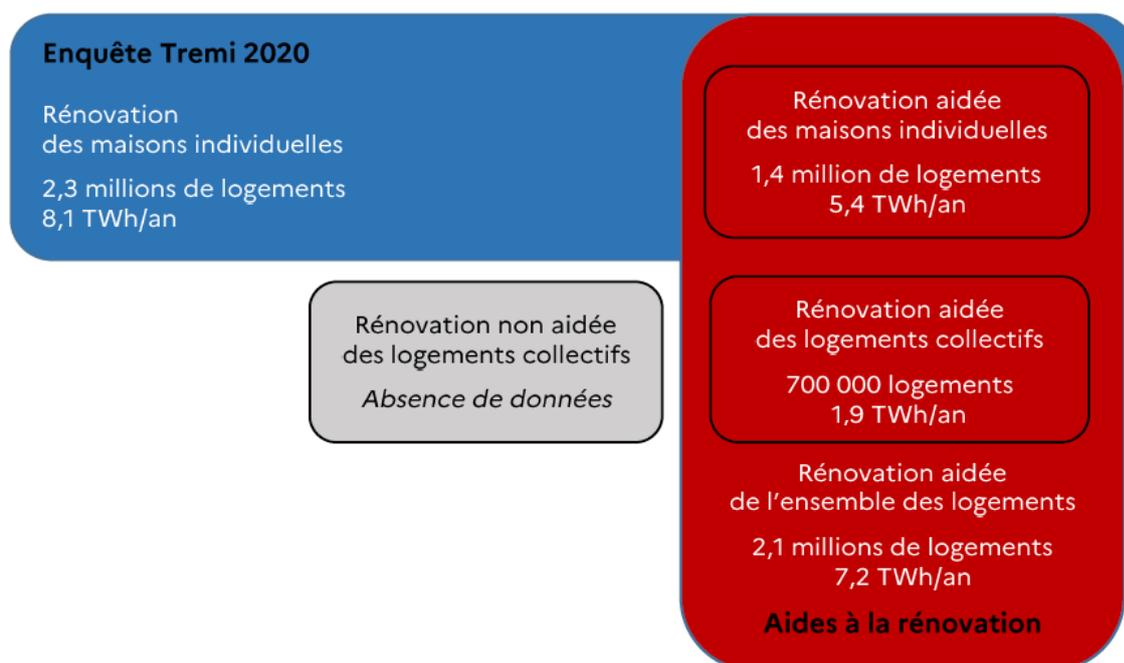
Cependant, ces aides ne permettent pas de connaître l'ensemble des rénovations réalisées dans les logements français. En effet, environ une rénovation sur trois serait réalisée sans bénéficier d'aucun dispositif d'aide. La *figure 1* permet de visualiser l'ensemble des données sur la rénovation.

Il est donc nécessaire de compléter ces informations par le biais d'enquêtes statistiques à échantillonnage aléatoire. En particulier, l'enquête Tremi (Travaux de rénovation énergétique dans les maisons individuelles) permet d'obtenir des informations sur :

- Les motivations des ménages (confort, économies d'énergie) réalisant des rénovations ;
- À l'inverse, les freins à la rénovation (complexité, coût, etc.) ;

1. Cette estimation en vigueur en 2018 a depuis été revue, avec une estimation de 5,1 millions de passoires thermiques.

FIGURE 1 – La rénovation énergétique des logements en 2019



Source : SDES

- La connaissance et le recours éventuel à des services d'appui à la rénovation (réseau Faire, sites Internet dédiés, etc.) ;
- Le coût des rénovations et la part couverte par les aides.

L'enquête Tremi 2020 a été menée conjointement par le SDES et l'Ademe², en s'appuyant pour la maîtrise d'œuvre sur un groupement de prestataires externes recruté par appel d'offres. Le service des données et études statistiques s'est particulièrement occupé des aspects méthodologiques de l'enquête, de sa préparation à sa diffusion.

Après une présentation de la collecte de l'enquête, nous détaillerons les traitements post-collecte qui ont été effectués, en particulier en ce qui concerne le traitement de la non-réponse, ainsi que la façon dont ont été traitées les opérations de diffusion.

1 Échantillonnage et collecte

1.1 Méthodologie

1.1.1 Champ et base de sondage

Les unités statistiques étaient les maisons individuelles habitées en tant que résidences principales par les ménages occupants. Les logements/ménages ont été tirés dans le millésime 2018 de Fidéli (comprenant les revenus déclarés pour l'année 2017). L'unité d'observation était le ménage pour toutes les questions relatives à la connaissance des aides, aux avis sur les professionnels, etc.,

2. Jusqu'en 2017, seule l'Ademe était maître d'ouvrage. Le SOeS avait participé au comité de pilotage de la campagne 2017.

et le logement qu'il occupe à titre de résidence principale pour toutes les questions relatives aux équipements présents, aux travaux effectués, etc.

Les logements en immeuble collectif ont été exclus pour des raisons d'hétérogénéité des champs : les dynamiques de rénovation sont très différentes entre maisons individuelles et appartements. La réalisation d'une enquête complémentaire à Tremi pour les logements collectifs (appelée Treco) est envisagée pour la campagne 2023 de l'enquête [9].

Le champ géographique de Tremi 2020 était la France, les départements et régions d'outre-mer (Drom) ayant été interrogés pour la première fois, à titre expérimental. Les Drom ont pu être interrogés sans surcoût car l'enquête fut auto-administrée et par Internet. Le questionnaire a été adapté pour mieux correspondre aux spécificités de ces régions grâce aux retours des directions régionales ultramarines de l'Ademe. Cependant, compte-tenu des spécificités de ces territoires et de l'aspect relativement nouveau de cette interrogation, la diffusion de résultats sur chacun des Drom individuellement n'a pas été effectuée.

1.1.2 Échantillonnage

L'échantillon était composé de 100 000 ménages, établi en une seule fois (pas d'échantillons de réserve). Le tirage de cet échantillon a été confié à l'Insee. Une disjonction de l'échantillon Tremi avec les autres enquêtes de la statistique publique, et en particulier l'ENL³ 2019-2020, fut effectuée, afin de limiter la charge d'interrogation pour les ménages (limiter le risque d'être interrogé plusieurs fois pour des enquêtes différentes sur une même période).

Il a été stratifié explicitement selon la région, par allocation manuelle. Pour les régions, l'un des objectifs de l'enquête était de réaliser une diffusion régionale des résultats sur les principaux paramètres d'intérêt. Or, une répartition proportionnelle conduisait à des nombres de répondants estimés trop bas dans certaines régions métropolitaines. La répartition de l'échantillon a donc été ajustée manuellement pour prendre en compte ce problème, tout en limitant au maximum la déformation des taux de sondage afin de conserver une dispersion des poids finaux la plus faible possible.

Le tirage était ensuite un tirage systématique à probabilités égales dans chacune des régions, trié selon les variables suivantes :

- **Le statut d'occupation** qui était regroupé en deux modalités : propriétaire (modalités « occupation par le propriétaire » et « occupation à titre gratuit ») et locataire (autres modalités) ;
- **L'année de construction du logement**, qui a été utilisée en construisant des tranches dont les bornes étaient déterminées par les réglementations thermiques françaises successives : 1948, 1974, 1988, 2000, 2006 et 2012 ;
- **Le revenu du ménage**, défini comme le revenu disponible divisé par le nombre d'unité de consommation au sein du ménage, en tranches construites à partir des quartiles de la distribution ;
- **Le code commune** Insee.

Les deux premières variables (statut d'occupation et année de construction) sont fortement corrélées aux variables d'intérêt de l'enquête, en particulier le fait d'avoir effectué des travaux de rénovation énergétique. Les deux dernières ont été utilisées pour disperser le plus possible l'échantillon.

1.2 Questionnaire

Le questionnaire de l'enquête Tremi était calibré pour être rempli en une trentaine de minutes⁴, auto-administré en ligne⁵. Pour répondre de manière la plus précise possible, l'enquêté était encouragé à aller consulter les factures relatives à ses travaux. Néanmoins, des ordres de grandeur sur les coûts des travaux ont été acceptés. Les ménages n'ayant pas réalisé de travaux entrant dans le champ de l'enquête devaient quand même répondre à un questionnaire allégé pour en préciser justement les freins, et répondre à quelques questions relatives à la notoriété des aides et dispositifs. Ce questionnaire dit « court » durait environ 10 minutes.

Les membres du comité de pilotage ont décidé d'administrer le questionnaire à propos du logement que le ménage occupe à la date du remplissage du questionnaire (donc il y a une censure des travaux réalisés dans « l'ancienne maison » dans le cas où le ménage a déménagé pendant la période d'enquête). Afin de palier ce biais, il est prévu dans les prochaines éditions d'interroger, en plus des ménages occupants, les propriétaires bailleurs.

Une séance du comité des utilisateurs a été dédiée aux variables du questionnaire. Sa trame a été conservée par rapport à l'édition précédente de l'enquête :

- un bloc « Informations sur le foyer » : commune, âge, PCS⁶, etc.
- un bloc « Informations sur le logement » : année de construction, système de chauffage et d'eau chaude sanitaire (individuel ou collectif)
- un bloc « Travaux » : types de gestes effectués ou raisons d'une absence de travaux. En cas de travaux, le questionnaire « long », destiné aux répondants ayant fait au moins un travaux de rénovation énergétique qui s'est terminé en 2019, continuait sur les sous-blocs suivants :
 - un sous-bloc « Détails techniques des travaux » : filtré par geste, quel type d'isolant a été posé, etc.
 - un sous-bloc « Réalisation des travaux » : qui a réalisé les travaux, pour quel montant, etc.
 - un sous-bloc « Caractéristiques techniques du logement » : énergie avant travaux, etc.
 - un sous-bloc « Informations sur le logement bis » : forme de la maison (avec des dessins pour expliquer les formes des bâtiments), altitude, etc.
 - un sous-bloc « Avis post-travaux » : opinion, ressenti, etc.
- un bloc « Aides et prêt » : connaissance et utilisation, avec les logos des dispositifs d'aides mis en avant pour faciliter les réponses
- un bloc « Accompagnement » : organismes et structures d'appui.

Des filtres de cohérence ont été mis en place sur certaines variables, directement dans le questionnaire. Cela concernait notamment la surface du logement, bornée par une valeur minimale et une maximale, l'année d'emménagement, qui ne pouvait pas être inférieure à l'année de construction du logement, etc. Une question-filtre a également été ajoutée au questionnaire, sur le type de logement occupé. Si le ménage déclarait ne pas habiter en maison individuelle, le questionnaire prenait immédiatement fin.

4. Au final, la durée moyenne de complétude a été de 24 minutes pour les plus longs.

5. Un mode exclusivement *web* étant susceptible d'introduire un biais de couverture, nous avons autorisé les enquêtés qui le souhaitaient à demander un questionnaire papier.

6. Professions et catégories socioprofessionnelles

1.3 Terrain

1.3.1 D roul 

Apr s avoir re u un avis d'opportunit  de la part du Cnis en d cembre 2018, et un avis d'examen favorable avec obligation de r ponse de la part du Comit  du label en novembre 2019, la collecte de Tremi 2020 s'est d roul e, une fois la phase de test termin e, du 9 janvier au 14 avril 2020. Elle a  t  confi e   une groupement de prestataires, dont CSA Research qui pilotait la collecte.

Selon le protocole, le r pondant n' tait pas forc ment la personne de r f rence du m nage : il  tait possible d'enqu ter toute personne ayant une connaissance suffisante des travaux effectu s dans le logement. Les adresses *mails* et num ros de t l phone disponibles dans les fiches-adresses ont  t  utilis s pour les phases de relances. Le rythme de relance avait pour objectif de maximiser les chances d'atteindre l'objectif de 40 % de r pondants.

Une lettre-avis a ainsi  t  envoy e   100 000 m nages,  chantillonn s dans le mill sime 2018 de Fideli, les encourageant   r pondre   l'enqu te en ligne. Au total, 51 485 questionnaires *online* ont  t  compl t s dont :

- 10 265 questionnaires « longs »
- 41 220 questionnaires « courts »

De plus, 1 529 questionnaires papier ont  t  re us, trait s et saisis. Ils  taient destin s   toute personne en ayant fait la demande *via* la *hotline* ou la *webmail* d di es   l'enqu te,   l'exception des habitants des Drom.

Plusieurs relances  taient pr vues tout au long de la collecte (voir *Figure 2*) :

- Des relances par *mail*, pour les foyers pour lesquels cette information  tait disponibles, ont  t  effectu es trois fois durant la collecte ;
- Deux relances postales ont  t  r alis es apr s respectivement un tiers et deux tiers de la p riode de collecte ;
- En fin de collecte, une p riode de relance t l phonique (pour les foyers pour lesquels l'information  tait disponible) a  t  r alis e.

1.3.2 Confinement

La *Figure 2* pr sente l' volution du taux de collecte, qui s'est d roul e de janvier   mi avril 2020. Entre le 13 janvier et le 10 f vrier, il progresse et atteint 33 %, puis jusqu'au 19 mars, il se stabilise. Enfin, d s le 20 mars il augmente de fa on significative pour atteindre 53 % en fin de p riode de collecte.

Ce pic inhabituel du taux de r ponse pourrait s'expliquer par le confinement g n ralis  d cr t  par le gouvernement suite   l' mergence du coronavirus⁷, qui a bloqu  une grande partie des m nages chez eux et a pu faire augmenter le taux de r ponse de l'enqu te, ind pendamment de l'effet des relances.

L'Insee estime en effet   34 % la part des personnes en emploi qui ont t l travaill  durant la p riode de confinement [1]. L'institut met en avant le fait que ce taux est tr s li    la cat gorie sociale et aux revenus. Il monte   58 % pour les cadres et professions interm diaires, tandis que le t l travail a concern  21 % des personnes les plus modestes (1^{er} quintile de niveau de vie), et

7. Le 16 mars 2020, le pr sident de la R publique puis le ministre de l'int rieur ont annonc  pour une dur e de quinze jours - ensuite prolong e - des mesures de confinement pour endiguer l' pid mie de COVID-19 provoqu e par le coronavirus.



53 % des plus aisés (dernier quintile). Nous ne disposons toutefois pas dans nos fichiers d'analyse d'information sur les dates de réponse de chaque questionnaire. Ainsi, n'avons pas pu discriminer les répondants durant cette période selon leur catégorie socioprofessionnelle ou leurs revenus.

De façon plus précise, plusieurs phases dans la collecte peuvent être identifiées :

- du 13 au 27 janvier : deux premières semaines avec un fort taux de réponse ;
- du 28 janvier au 20 février : trois semaines faisant suite aux premières relances, avec des remontées continues de questionnaires, à un rythme moins soutenu cependant ;
- du 21 février au 19 mars : un mois de stagnation ;
- du 20 mars au 14 avril : remontée du taux de réponse, sûrement due à la période de confinement obligatoire sur l'ensemble du territoire.

2 Traitements post-collecte

L'échantillon final contenait 53 014 questionnaires exploitables. La correction de la non-réponse de l'enquête a suivi un processus classique de redressements : imputations, pondération, calage.

2.1 Correction de la non-réponse partielle

Afin de corriger la non-réponse partielle de certaines variables, les valeurs manquantes ont été imputées à partir de l'information provenant de sources externes.

2.1.1 Sexe, surface et année de construction du logement : utilisation du fichier de redressement

Le sexe de la personne de référence, le statut d'occupation, l'année de construction et la surface du logement sont des variables présentes dans les bases fiscales et donc disponibles dans le fichier de redressement directement. Renseignées en clair dans ces sources, nous avons toutefois considéré plus prudent de la renseigner en tranche dans nos fichier d'analyses.

Pour la surface cependant, afin de faciliter les calculs « surfaciques » prévus⁸, une variable en clair a été imputée, en prenant l'information du fichier de redressement si celle-ci est cohérente avec la tranche renseignée par le ménage, soit en prenant le milieu de la tranche renseignée par le ménage s'il y a incohérence avec l'information du fichier de redressement⁹ (*Table 1*).

2.1.2 Coût des travaux : par imputation

Pour les variables de coût des travaux, nous avons considéré comme non-répondants tout ménage qui n'a pas répondu ou qui a déclaré un coût jugé incohérent ; notamment, les coûts trop faibles (en dessous d'un seuil repris de celui établi pour le traitement de Tremi 2017 : 50 €) sont considérés comme incohérents, car le ménage doit déclarer le coût des travaux avant le bénéfice d'aides. Cela exclut notamment les rénovations dites à « un euro », qui correspondent à des aides bonifiées pour les ménages les plus précaires.

Les variables de coût des travaux étaient de bonnes candidates à une correction par imputation. Leur taux de non-réponse se situait en effet entre 10 % et 20 %, et le coût est très certainement corrélé à des variables présentes dans le questionnaire (type d'équipement installé, surface, etc.).

8. Ces calculs sont utiles en particulier dans la perspective des calculs de gains énergétiques, cf. 3.2.

9. Cette imputation par la moyenne, effectuée également pour l'année de construction, était toutefois mineure, car elle n'a concerné qu'au plus 10 observations.

TABLE 1 – Bilan des corrections par variable

La surface vient de	Variable en clair	Variable en tranche
une réponse du répondant	43 027	9 316
un recodage de la réponse en tranche du ménage	4 271	
un recodage de la réponse en clair du ménage		43 027
l'attribution de l'information du fichier de redressement	5 711	666
une imputation par la moyenne	5	5
L'année de construction du logement vient de		
un recodage de la réponse en clair du ménage	49 134	
l'attribution de l'information du fichier de redressement	3 871	
une imputation par la moyenne	9	
Le sexe de la personne de référence vient de		
une réponse du répondant	52 916	
l'attribution de l'information du fichier de redressement	98	
Le statut d'occupation du logement vient de		
une réponse du répondant		
l'attribution de l'information du fichier de redressement	136	

Lecture : 5 711 corrections de la non-réponse partielle concernant la surface du logement provient du fichier de redressement lorsque la variable est en clair, 666 lorsqu'elle est continue.

Source : SDES

La première étape du traitement de la non-réponse a consisté à trouver ces variables expliquant le coût des travaux (pris au logarithme). Par geste, un modèle statistique de sélection de variables (régression pénalisée, *via le package glmnet*) a été mis en place, et un maximum de trois ou quatre variables ont été retenues comme étant les plus explicatives.

La deuxième étape du traitement a consisté à croiser les variables choisies pour former des groupes de ménages, puis d'imputer les valeurs manquantes avec la moyenne du groupe auquel appartient le ménage qui n'a pas rempli le coût des travaux.

La *Table 2* présente l'ensemble des corrections effectués sur chacun des gestes du questionnaire. L'*Annexe 1* donne la correspondance entre les codes des gestes et leurs intitulés.

Par exemple, pour les travaux « Isolation d'un pan ou de la totalité d'une toiture sans rénovation de la toiture », les variables sélectionnées par le modèle statistique ont été : « part de surface isolée », « épaisseur de l'isolant », « qui a réalisé les travaux » et « nombre d'étages ». Un ménage qui n'a pas répondu au coût de ses travaux s'est donc vu imputer la valeur moyenne de l'ensemble des ménages ayant répondu et qui partagent les mêmes modalités pour ces quatre variables.

Si une des variables explicatives n'était pas renseignée pour un ménage pour lequel le coût est manquant ou si un des croisements donne un groupe « vide », nous avons formé des groupes plus larges en retirant une à une les variables explicatives (en partant de la moins explicative) jusqu'à ce que l'appartenance à un groupe « non vide » puisse être établi pour chaque ménage.

Dans l'exemple précédent, si un ménage n'a pas répondu au coût de ses travaux et n'a pas renseigné l'épaisseur de l'isolant, la moyenne des coûts a été calculée sur l'ensemble des ménages ayant la même modalité à la variable « part de la surface isolée ».

2.1.3 Traitement des incohérences

Lors de l'étude de la non-réponse partielle, et notamment en cherchant à reproduire les filtres pour connaître les ménages concernés par les questions, certaines incohérences sont apparues.

Certains ménages, qui n'auraient pas dû se voir poser certaines questions, ont quand même une modalité enregistrée dans le questionnaire pour ce champ.

Quelques exemples d'incohérences :

- La question « Depuis la construction de votre logement, la toiture a-t-elle fait l'objet d'une isolation ? », qui est censée être filtrée selon « Si pas de travaux sur la toiture déclaré », a en fait été posée à tous les questionnaires longs (aucune non-réponse sur ce champ) ;
- La question « Savez-vous que des organismes publics peuvent vous conseiller gratuitement et/ou vous accompagner dans vos travaux de rénovation énergétique ? » n'aurait pas dû être posée aux personnes ayant déclaré avoir contacté un de ces organismes publics quelques questions plus tôt. Or, il se trouve que de ces ménages ont parfois répondu à cette question de manière contradictoire ;
- D'autres questions ont un nombre de répondants supérieur au nombre de ménages censés être concernés par la question.

Les réponses ne correspondant pas à une situation dans laquelle le ménage aurait dû se voir poser la question ont été effacées, afin de faciliter l'étude de la non-réponse.

TABLE 2 – Bilan des corrections par geste

	Effectif concerné	Nombre de réponses imputées	Taux de non-réponse	Moyenne avant imputation	Moyenne après imputation	Différence entre les moyennes
Q11	882	170	19,3%	15 269	14 728	-3,5%
Q12	503	61	12,1%	10 831	10 892	0,6%
Q13	672	187	27,8%	4 464	4 367	-2,2%
Q14	2 608	791	30,3%	2 613	2 625	0,5%
Q15	119	29	24,4%	10 256	10 193	-0,6%
Q16	98	16	16,3%	7 136	7 649	7,2%
Q21	497	99	19,9%	13 400	13 416	0,1%
Q22	917	155	16,9%	6 835	6 796	-0,6%
Q23	1 003	241	24,0%	6 020	6 257	3,9%
Q24	856	135	15,8%	2 958	2 932	-0,9%
Q31	753	233	30,9%	3 480	3 450	-0,9%
Q32	395	71	18,0%	3 428	3 346	-2,4%
Q41	3 408	488	14,3%	7 797	7 736	-0,8%
Q42	1 465	304	20,8%	4 857	4 963	2,2%
Q43	1 699	297	17,5%	3 753	3 794	1,1%
Q51	3 035	354	11,7%	7 309	7 205	-1,4%
Q52	896	196	21,9%	2 556	2 595	1,5%
Q53	2 084	337	16,2%	3 956	3 896	-1,5%
Q54	196	71	36,2%	2 314	2 245	-3,0%
Q55	1 236	271	21,9%	1 185	1 190	0,4%
Q56	1 054	115	10,9%	3 918	3 900	-0,5%
Q57	123	21	17,1%	2 367	2 243	-5,2%
Q58	285	24	8,4%	16 102	15 844	-1,6%

Lecture : Pour le geste Q11 (Rénovation d'un pan ou de la totalité d'une toiture avec isolation de la toiture), le coûts moyens des travaux était de 15 269 € avant imputation, et de 14 728 € après.

Source : SDES

Seules les variables pertinentes ont été traitées. Ainsi, les questions obligatoires n'ont pas été étudiées car, dans ce cas, une réponse vide correspond forcément à un ménage non concerné¹⁰. De même, pour les questions à choix multiple, qui ont été codées sur plusieurs colonnes indicées, seule la première a été étudiée (la deuxième colonne aura forcément moins de valeurs remplies que la première, calculer la non-réponse dessus n'aurait aucun intérêt).

Afin de distinguer, en cas de besoin, les réponses imputées de celles collectées directement auprès des ménages, des indicatrices ont été créées dans le fichier d'analyse et de diffusion, indiquant l'origine de la modalité.

2.2 Correction de la non-réponse totale

Le mode de collecte de l'enquête Tremi, auto-administré, associé au thème de l'enquête, très spécifique, induit un risque lié au caractère non-ignorable de la non-réponse (appelé aussi « *Non Missing at Random* », ou NMAR) par Little et Rubin [5] ; c'est-à-dire, lorsque le fait d'avoir réalisé ou non des travaux, toutes choses égales par ailleurs, a un impact sur la motivation à répondre à l'enquête d'un ménage. Ce risque est vrai pour la grande majorité des enquêtes ([6], [7]), mais pourrait être particulièrement important ici. Pour pallier ce problème, un protocole de traitement des non-répondants a été mis en place, afin de limiter ou de corriger les effets de sélection (ménages n'ayant pas fait de travaux et ne répondant pas de ce fait à l'enquête, car se sentant non concernés).

2.2.1 Enrichissement et sélection des variables

Ainsi, la phase de correction de la non-réponse a mobilisé le plus d'informations possibles afin d'enrichir l'échantillon de Tremi. Nous avons ainsi utilisé d'une part les variables issues du fichier d'échantillonnage : surface, année de construction, type de ménage, revenus disponibles, etc., et d'autre part les variables qui ont pu être appariées grâce aux coordonnées géographiques : variables communales, zonages, etc. En plus des nombreuses variables sociodémographiques et liées aux revenus issues des sources fiscales, des variables plus spécifiques ont été mobilisées à partir de sources plus directement liées aux variables d'intérêt de l'enquête.

Les différentes sources de données mobilisables à cette fin sont les informations relatives aux aides à la rénovation : le Crédit d'impôt pour la transition énergétique (CITE) et les Certificats d'économie d'énergie (CEE).

1. Le CITE vise à encourager les ménages à réaliser des travaux de rénovation énergétique. Cette prime est associée à la déclaration de gestes précis dans le cadre des déclarations de revenus des ménages ; elle ne concerne que la réalisation de travaux dans leur résidence principale. Elle correspond à de nombreux gestes tels que le changement d'une chaudière, l'installation d'une pompe à chaleur, l'isolation de parois extérieures ou intérieures, etc.

Le SDES, *via* sa sous-direction des statistiques du logement et de la construction, a mené des études sur les ménages bénéficiaires du CITE [10]. Il en concluait qu'environ 1,4 million de foyers ont bénéficié du crédit d'impôt en 2017, et, pour 75 % d'entre eux, cela ne concernait qu'un seul geste. Même si le nombre de bénéficiaires est en croissance entre 2015 et 2017 (de l'ordre de 150 000 ménages concernés supplémentaires chaque année), une évolution de la liste des dépenses éligibles à cette aide a conduit à une baisse du nombre de bénéficiaires en 2018 et 2019 (moins d'un million de foyers chaque année).

10. Même si dans les faits, certains questionnaires papiers contredisent cette règle

Les variables associées au CITE étant fiscales, elles sont directement disponibles dans les fichiers mobilisés par l’Insee pour l’échantillonnage, et n’ont pas nécessité d’appariement spécifique. En effet, les variables associées au CITE sont présentes dans les fichiers POTE auxquels a accès la division Sondages de l’Insee. Elles ont pu être fournies dès la livraison du fichier redressement pour l’année 2017 ; un enrichissement complémentaire associé aux mêmes logements a été demandé pour les crédits d’impôt au titre des années 2018 et 2019, et ont pu être utilisés pour les redressements de l’enquête. Le taux d’enrichissement a été ainsi de 100 %, et une attention particulière a été portée aux ménages ayant déménagé, ménages que la division Sondages nous a indiqué *via* une variable indicatrice de déménagement.

Dans l’enquête Tremi 2017, la question du recours au CITE/CIDD (nom du crédit d’impôt avant 2015) indiquait un recours à ces aides de 11 % à 14 %, suivant que la rénovation était jugée performante ou non. Comme indiqué dans la brochure, ce taux était jugé « anormalement bas en regard des chiffres officiels ». L’hypothèse était que l’intitulé de la question était difficilement compréhensible par les ménages ; il a été adapté et amélioré pour l’édition 2020¹¹. Dans tous les cas, il est raisonnable de penser que le taux de recours au CITE est une variable de redressement de bonne qualité, car fortement corrélée à la réalisation de travaux.

2. Le dispositif des CEE repose quant à lui sur une obligation de réalisation d’économies d’énergie imposée par les pouvoirs publics aux vendeurs d’énergie (électricité, gaz, GPL¹², chaleur et froid, fioul domestique et carburants pour automobiles). Ceux-ci sont ainsi incités à promouvoir activement l’efficacité énergétique auprès de leurs clients et des autres consommateurs d’énergie : ménages, collectivités territoriales ou professionnels.

Dans l’enquête Tremi 2017, la question du recours à des CEE (à noter que cette dénomination est très mal connue des ménages, la question étant formulée sous l’angle « aides des fournisseurs d’énergie ») indiquait un recours à ces aides de 8 %, et jusqu’à 17 % pour des travaux de rénovation ayant permis deux sauts de classe DPE. Cette variable est donc fortement corrélée à la réalisation de travaux, en particulier de travaux de rénovation énergétique performants.

Les données sur les CEE sont centralisées par la Direction générale de l’énergie et du climat, au sein du Ministère de la transition écologique, à partir des bases de données des vendeurs d’énergie, qui les délivrent. Ces bases de données sont actualisées en temps réel ; les données disponibles en 2020 couvraient l’ensemble de la période 2017 – 2019 à laquelle s’intéresse l’enquête Tremi 2020. Ces données ont été mise à disposition auprès du SDES dans le cadre des travaux de l’Observatoire national de la rénovation énergétique.

Un appariement a pu être réalisé à l’adresse (opération facilitée car le champ de l’enquête est limité aux maisons individuelles), en utilisant des informations sur la position géographique des ménages dans Fidéli (coordonnées, etc.) ; cette opération a permis un enrichissement de l’échantillon par cette variable pour la plus grande partie de l’échantillon.

11. Une attention particulière a, de plus, été apportée à la modalité « autre », les répondants ne connaissant pas toujours la dénomination des aides concernées

12. Gaz de pétrole liquéfié

Afin d'être sélectionnée, une variable devait être corrélée au fait de répondre ou non et au fait d'avoir réalisé des travaux de rénovation ou non. La *figure 3* explicite le taux de réponse et le taux de travaux selon les modalités d'une partie des variables étudiées (d'autres variables ont également été testées, mais n'ont finalement pas été retenues).

Les variables écrites en vert sont celles que l'on a, *a priori*, retenues : le taux de réponse et le taux de travaux varient significativement selon leurs modalités. On y retrouve l'année de construction en tranches, le statut d'occupation, et une typologie des ménages croisant le nombre d'adulte le composant et un certain niveau de revenu.

Les variables en rouge sont celles que l'on a rejetées : le taux de réponse ou le taux de travaux ne varient pas significativement entre leurs modalités. La surface en tranches par exemple, pour laquelle le taux de réponse varie bien mais pas le taux de travaux, a été rejetée de la ré pondération par GRH¹³.

Les variables en bleu sont celles qui remplissaient les conditions et auraient pu être sélectionnées individuellement, mais qu'on a préféré regrouper afin d'éviter un effet de sur-apprentissage (*i.e.* les données entrées sont trop nombreuses et contraignent trop le modèle, qui s'ajuste donc mieux aux données présentes mais qui perdrait en efficacité avec des données supplémentaires, ce que l'on veut éviter). C'est le cas des variables indicatrices sur le bénéfice d'un CITE ou d'un CEE, qui ont été regroupées en une seule indicatrice résumant l'information de 5 indicatrices. Cette indicatrice globale indique donc si le ménage a bénéficié ou non d'une aide.

Les informations obtenues grâce à ces variables ont été utilisées à différentes étapes du traitement de la non-réponse totale dans l'enquête Tremi 2020 : d'une part lors de la ré pondération par création de groupes de réponses homogènes (GRH), d'autre part lors du calage sur marges des données.

2.2.2 Groupes de réponses homogènes

Le principe des GRH est de construire des groupes de personnes « homogènes », *i.e.* ayant des comportements de réponse et des caractéristiques similaires, au sein desquels la corrélation entre comportement de réponse et variables d'intérêt pourra être négligée [4]. Ces groupes sont construits par la méthode dite des scores, qui consiste à modéliser des probabilités de réponse pour chaque individu et à regrouper ceux qui partagent les mêmes probabilités. La pondération initiale, qui ne prenait pas en compte l'information sur le recours aux aides, a ainsi été corrigée au sein de chaque groupe en la multipliant par l'inverse du taux de réponse au sein du groupe, avant de procéder à un calage classique.

Nous avons alors utilisé une régression logistique, modélisant l'indicatrice de réponse selon l'ensemble des variables sélectionnées précédemment (*cf.* 2.2).

$$i_{\text{rep}} \sim \text{typologie_ménage} + \text{tranche_année_construction} + i_{\text{locataire}} + i_{\text{CITE ou CEE}}$$

Le modèle a été testé et optimisé, pour vérifier que les variables explicatives amélioreraient bien significativement son efficacité, et que la sélection effectuée plus haut était bien cohérente. L'efficacité des modèles a été contrôlée *via* le taux de bonnes prédictions, calculé sur un échantillon de test, différent de celui sur lequel le modèle est construit, afin d'éviter tout problème

13. Groupes de réponses homogènes, *cf.* 2.3

FIGURE 3 – Taux de réponse et de travaux

		Taux de réponse	Taux de travaux	Eff.
Surface (tranches)	(0,29]	23%	38%	454
	(29,39]	29%	37%	984
	(39,49]	33%	39%	2 128
	(49,69]	40%	41%	10 706
	(69,99]	50%	43%	37 562
	(99,149]	59%	42%	36 303
	(149,999]	62%	40%	11 457
Année construction (tranches)	<1949	49%	43%	30 273
	1949-1974	51%	46%	17 976
	1975-1981	56%	49%	10 686
	1982-1989	58%	50%	10 739
	1990-2000	57%	45%	10 843
	2001-2011	56%	31%	13 731
	>2012	56%	17%	5 387
Sexe	Femme	43%	41%	26 682
	Homme	56%	42%	73 318
Statut d'occupation	Propriétaire	57%	43%	82 299
	Locataire	33%	34%	17 701
Typologie ménage <i>XA = X adulte</i> <i>REV1 = revenu faible</i> <i>REV3 = revenu élevé</i>	1A-REV1	31%	36%	9 605
	1A-REV2	41%	40%	12 615
	1A-REV3	51%	39%	10 318
	2A-REV1	48%	43%	22 684
	2A-REV2	62%	42%	23 706
	2A-REV3	67%	43%	20 283
Indicatrice CITE 2018	0	52%	39%	95 666
	1	67%	90%	4 334
Indicatrice au moins un CITE en 2017 ou 2018	0	51%	35%	89 492
	1	67%	85%	10 508
Indicatrice CEE 2018	0	53%	41%	98 424
	1	58%	78%	1 576
Indicatrice au moins un CEE entre 2016 et 2018	0	53%	40%	95 533
	1	59%	70%	4 467
Indicatrice au moins un CITE ou un CEE	0	51%	34%	86 374
	1	64%	79%	13 626
Indicatrice de déménagement entre 2017 et 2018	0	58%	42%	89 492
	1	14%	41%	10 508
Zone de tension immobilière	A	58%	45%	8 947
	Abis	61%	41%	1 039
	B1	50%	43%	20 642
	B2	54%	42%	21 068
	C	53%	41%	47 314
Type de commune	Banlieue	55%	43%	32 274
	Centre ville	51%	43%	20 835
	Isolé	51%	41%	12 389
	Rural	53%	40%	33 506
		53%	42%	100 000

Source : SDES

de sur-apprentissage. Une procédure de sélection automatisée a également été mise en place (*stepwise*) pour confirmer nos comparaisons et le modèle ci-dessus.

En sortie du modèle, nous avons pris les prédictions calculées, qui sont des valeurs continues comprises entre 0 et 1, puis nous avons entré ces valeurs dans un algorithme qui a généré les groupes maximisant l’homogénéité intra-groupes. Cet algorithme a été développé par des agents de la division Sondages de l’Insee, à partir d’un article de Haziza et Beaumont « *On the construction of imputation classes in surveys* » [2].

En sortie de l’algorithme, un numéro de groupe a été affecté à chaque observation. Il ne restait plus qu’à modifier la pondération initiale au sein de chaque groupe, en la multipliant par le rapport entre le nombre total de ménages dans le groupe et le nombre de répondants dans le groupe. Cette nouvelle pondération a alors pu être utilisée en entrée d’un calage sur marges classique.

2.2.3 Calage sur marges

Une étape de calage sur marges mobilisant ces variables a ensuite été réalisée. Le montant régional total de CITE observé a été introduit comme variable de calage pour la plupart des régions de France continentale. Cette variable a servi de *proxy* pour la réalisation des travaux mais également pour leurs caractéristiques (coût total des travaux, qualité des matériaux, etc., qui sont des déterminants de la dépense éligible au CITE).

Le calage a été réalisé avec le *package* « *icarus* » [8], qui adapte la macro SAS « CALMAR » développée par l’Insee pour le logiciel statistique R. Dans le cas de Tremi 2020, ce redressement a été fait par région, de sorte qu’implicitement, le nombre de maisons individuelles par région correspond aux marges régionales. Les variables de calage choisies sont au nombre de 7 (5 qualitatives et 2 quantitatives) :

- Le statut d’occupation du logement (Locataire / Propriétaire)
- L’année de construction du logement (Avant 1948 / De 1948 à 1974 / De 1975 à 2000 / Après 2000)
- La tranche d’âge de la personne de référence (Moins de 50 ans / De 50 à 64 ans / 65 ans et plus)
- La PCS de la personne de référence (PCS sup. / PCS inf.)
- La tranche de revenu du ménage (Moins de 19 000 € / De 19 000 à 31 700 € / Plus de 31 700 €)
- Le revenu total (variable quantitative)
- Le montant total de CITE déclaré¹⁴ (variable quantitative).

Toutefois, pour certaines régions, la convergence de l’algorithme de calage a nécessité d’exclure certaines de ces variables. En ce qui concerne le revenu, l’information par tranche a été privilégiée. Dans les cas où il n’est pas apparu possible de l’intégrer au calage, la prise en compte alternative de la variable de revenu total a été testée (et s’est avérée concluante dans un cas, pour la Martinique).

14. Les valeurs des travaux déclarés supérieures à 100 000 € ont été corrigées et mises à 100 000 €.

FIGURE 4 – Informations utilisées pour la calage par région dans l'enquête Tremi 2020

	Nombre de maisons individuelles de la région	Statut d'occupation du logement	Tranche d'année de construction	Tranche d'âge de la personne de référence	PCS de la personne de référence	Tranche de revenu du ménage	Revenu total des ménages de la région	Montant total tronqué de CITE déclaré
Guadeloupe								
Martinique								
Guyane								
La Réunion								
Mayotte								
Île-de-France								
Centre-Val de Loire								
Bourgogne-Franche-Comte								
Normandie								
Hauts-de-France								
Grand Est								
Pays de la Loire								
Bretagne								
Nouvelle-Aquitaine								
Occitanie								
Auvergne-Rhône-Alpes								
Provence-Alpes-Côte-d'Azur								
Corse								

Lecture : Dans l'enquête Tremi 2020, pour la région Auvergne-Rhône-Alpes, la variable « revenu total des ménages » n'a pas été utilisée pour le calage.

Source : SDES

Les informations sont issues des résultats de l'Enquête Emploi en continu de l'Insee ou des sources fiscales (Fidéli). Dans les deux cas, les totaux ont été recalculés pour être cohérents avec les derniers chiffres du Compte satellite du logement (CSL), statistique officielle concernant le logement.¹⁵

La repondération préalable avec la méthode des GRH a amélioré la pondération initiale en entrée du calage et a permis d'effectuer un calage de meilleure qualité, notamment dans les Drom, en y ajoutant des variables (*Figure 4*).

2.3 Pondérations

Une fois obtenu l'accès aux données relatives aux CITE 2019, le SDES a pu estimer les poids définitifs utilisés pour produire les résultats de l'enquête. La distribution des poids a été étudiée et les résultats ainsi obtenus ont été comparés à ceux obtenus sans repondération.

La dispersion des poids est standard compte-tenu des taux de réponse observés et des méthodes employées (*Figure 5*). En métropole, les poids définitifs s'étalent de 116 à 1 248 , avec 80 % des poids compris entre 198 et 475. Dans les Drom, les poids sont compris entre 69 et 4 156, avec 80 % des poids compris entre 184 et 657. Les estimations demeurent fragiles, surtout à Mayotte. Les exploitations sur ce champ géographique ont nécessité de ne publier les résultats que sur la

15. Le CSL synthétise un grand nombre de sources de données afin de dresser un bilan comptable complet et structuré des dépenses de logement de l'ensemble des acteurs économiques. Il est réalisé tous les ans au SDES par la Sous-direction des statistiques du logement et de la construction.

FIGURE 5 – Distribution des poids dans l'enquête Tremi 2020

Région	Min	D1	D2	D3	D4	D5	D6	D7	D8	D9	Max	Rapport max/min	Rapport D9/D1
DROM	69	184	220	254	310	345	388	428	522	657	4 156	60,2	3,6
Métropole	116	198	222	241	261	281	303	333	383	475	1 248	10,8	2,4
France	69	198	222	242	261	282	305	337	386	481	4 156	60,2	2,4

Lecture : Dans l'équête Tremi 2020, les poids de calage varient entre 116 et 1 248 en Métropole.

Source : SDES

France métropolitaine et de regrouper l'ensemble des Drom lors de la diffusion individuelle des données.

La variable « long » désigne, dans la *Table 3*, les ménages ayant répondu au questionnaire complet, *i.e.* ayant achevé au moins un geste de rénovation en 2019. Ce taux de travaux diminue quelque peu (de 19,9 % à 19,1 %) lorsqu'on tient compte des aides à la rénovation dans le calage, car les ménages ayant bénéficié d'aides ont plus répondu que les autres. L'intégration de ces variables conduit de plus à une évolution sensible de l'estimation du nombre total de gestes de rénovation (- 2,7 %), et du montant total des travaux (- 4,3 %)

Pour chacun des postes de travaux, la pondération définitive conduit à un moindre nombre de rénovations, mais avec un écart variable selon les postes de travaux (*Table 4*). La plus grosse différence porte sur les travaux concernant les ouvertures (portes et fenêtres), mais elle reste modérée (- 1 point de pourcentage, calculé comme l'écart entre le nombre de travaux ramené à l'effectif total¹⁶, avec et sans pondération).

La prise en compte des variables d'aides dans la repondération fait également évoluer la structure des montants des travaux, dans une certaine mesure (*Table 5*). Ce sont les travaux sur la toiture dont la structure par rapport aux montants totaux des travaux est le plus affectée par l'amélioration de la méthode de repondération (+ 1,2 point de pourcentage).

Le nombre de bénéficiaires du CITE entre 2017 et 2019 d'après les déclarations des enquêtés s'élève à 1,74 million avec la pondération définitive. Sa part dans la population a diminué par rapport à l'échantillon non pondéré (- 1,9 % environ), baisse due au fait que les bénéficiaires d'une aide répondent en moyenne mieux que les autres et qu'ils ont donc été sous-pondérés par rapport au reste de la population dans le calage définitif prenant en compte l'information sur les bénéficiaire d'au moins une aide.

3 Diffusion

3.1 Gains énergétiques

Afin de compléter les données de l'enquête Tremi 2020, le SDES souhaitait disposer des consommations des ménages, avant et après leurs travaux.

16. Le nombre total de ménage comprend les ménages qui n'ont pas fait de travaux

TABLE 3 – Taux de questionnaires longs par région, selon la pondération appliquée

Région	Calage définitif			Sans repondération		
	Long	Court	% long	Long	Court	% long
11	262 957	1 140 362	18,7%	891	3 722	19,3%
24	163 343	673 481	19,5%	684	2 657	20,5%
27	151 426	676 196	18,3%	654	2 864	18,6%
28	189 449	822 474	18,7%	677	2 765	19,7%
32	339 784	1 488 505	18,6%	950	3 914	19,5%
44	279 745	1 139 564	19,7%	917	3 536	20,6%
52	230 975	970 352	19,2%	836	3 282	20,3%
53	185 104	904 423	17,0%	661	3 062	17,8%
75	401 366	1 583 360	20,2%	1 108	4 142	21,1%
76	357 626	1 353 726	20,9%	1 115	3 967	21,9%
84	312 495	1 505 239	17,2%	1 000	4 460	18,3%
93	198 751	768 414	20,5%	776	2 834	21,5%
94	9 989	60 291	14,2%	32	186	14,7%
01	28 873	98 561	22,7%	58	198	22,7%
02	21 670	84 886	20,3%	49	174	22,0%
03	9 022	39 798	18,5%	50	163	23,5%
04	31 527	192 581	14,1%	92	519	15,1%
06	15 892	38 042	29,5%	5	14	26,3%
	3 189 995	13 540 255	19,1%	10 555	42 459	19,9%

Lecture : Après calage, dans la région 11 (Île-de-France), 262 957 ménages ont répondu au questionnaire « long » de l'enquête Tremi, et 1 140 362 ménages ont répondu au questionnaire « court » de cette enquête.

Source : Tremi 2020, SDES

TABLE 4 – Nombre de travaux par poste, avant et après calage

	Murs	Toits	Sols	Systèmes	Ouvertures	Ventilation	Total
Pond. def. (millions)	1,6	2,6	0,6	0,3	0,3	1,2	16,2
Sans pond. (milliers)	5,1	8,3	1,9	8,8	9,4	3,9	51,7
Écart (point de %)	-0,1	-0,2	0,0	-15,3	-16,5	-0,4	

Lecture : 51 700 ménages ont déclaré des travaux dans l'enquête Tremi, soit, après pondération, 16,2 millions de ménages.

Source : Enquête Tremi 2020, SDES

TABLE 5 – Montant des dépenses par postes

	en M€						Total
	Ouvertures	Systèmes	Toitures	Murs	Ventilation	Sols	
Pond. def.	11 650	9 508	9 609	6 580	1 632	1 202	40 181
Sans pond.	40	33	31	22	6	4	135
Écart (point de %)	-0,8	-0,6	1,2	0,3	-0,3	0,1	

Lecture : La somme des montants de travaux déclarés par les ménages dans l'enquête Tremi se monte à 135 millions d'euros, soit, après pondération, 40,2 milliards d'euros.

Source : *Enquête Tremi 2020, SDES*

3.1.1 Calculs des économies d'énergie

Les données de l'enquête Tremi 2017 avaient déjà permis d'évaluer, pour chaque logement ayant fait des travaux, l'étiquette DPE avant et l'étiquette DPE après travaux, en calculant des consommations énergétiques basées sur une méthode type 3CL 2021 (calcul de la consommation conventionnelle des logements)¹⁷. Lors de l'enquête Tremi 2020, un cabinet de consultants a réalisé ces calculs énergétiques conformément à la nouvelle méthode de calcul du DPE¹⁸, en intégrant les principaux changements concernant l'évaluation.

Certaines adaptations ont dû être menées pour prendre en compte les données disponibles dans l'enquête Tremi ; en effet, réglementairement, réaliser un tel diagnostic demande plus d'une centaine de données en entrée, établies par un expert visitant le logement ou le bâtiment. Ici, les données sont issues des données contextuelles de l'enquête (localisation des logements, date de construction, etc.) et des réponses aux questionnaires, ce qui peut induire des biais déclaratifs, les ménages n'ayant pas nécessairement de bonnes connaissances des performances thermiques de leur logement ou des systèmes de chauffage et d'eau chaude sanitaire. Plusieurs hypothèses, nécessaires mais forcément simplificatrices, ont donc été faites.

On peut notamment citer :

- Le fait que la détermination des surfaces au sens thermique repose sur la surface habitable déclarée (parfois en tranches) et sur les questions relatives à la géométrie du bâtiment, ce qui ne permet pas une analyse aussi précise qu'à partir de mètres sur plans ;
- Le fait que l'état initial des murs, plafonds et planchers bas n'est pas décrit dans l'enquête, et repose donc sur des hypothèses liées aux caractéristiques du logement et en particulier à sa période de construction ;
- Le fait que l'identification des systèmes de chauffage (avant et après travaux) peut être source d'erreurs pour des ménages non spécialistes du sujet, ce qui peut conduire à des écarts très forts sur l'estimation, ce paramètre étant central dans l'évaluation de la performance énergétique du logement ;
- Le fait que lorsque le ménage ne répond pas à une question technique (par exemple sur le type d'isolant utilisé), une valeur par défaut est renseignée, ce qui peut altérer la précision de l'estimation des gains.

17. La méthode conventionnelle dite 3CL s'appuie sur une analyse des caractéristiques du bâti (isolation, surfaces vitrées, matériaux de mise en œuvre, modes de chauffage et de production d'eau chaude sanitaire (ECS), etc.). Le mode de calcul conventionnel considère donc un usage standard du logement, ce qui correspond à une occupation et un comportement standard moyen.

18. En vigueur depuis juillet 2021

Les estimations de consommation ont été établies en énergie primaire et finale, avant toute rénovation, après l'ensemble des rénovations déclarées entre 2017 et 2019 et après chaque geste de rénovation pris isolément, ce qui permet d'analyser les gains associés à un geste en particulier, y compris pour les ménages ayant réalisé des bouquets de travaux. En effet, une difficulté dans l'évaluation des gains énergétiques associés aux gestes de rénovation tient au fait que l'effet d'une combinaison de gestes peut être différente de la somme des effets qu'auraient eu les gestes s'ils avaient été effectués isolément. L'estimation des gains geste par geste est en conséquence réalisée en deux étapes. Tout d'abord, on simule l'impact de chaque geste en ignorant les gestes effectués simultanément. Un redressement de ces gains est ensuite réalisé en comparant, pour chaque ménage répondant à l'enquête Tremi et faisant partie du champ de l'étude :

- les gains associés à l'ensemble des travaux réalisés entre 2017 et 2019 ;
- la somme des gains associés à chaque geste réalisé entre 2017 et 2019.

Le ratio de ces deux grandeurs est alors utilisé comme facteur correctif des gains associés aux gestes réalisés par le ménage en 2019. Dans certains cas (économies non estimables, incohérences trop fortes, etc.), aucune correction n'est appliquée et le chiffre est pris tel quel. Par ailleurs, afin de limiter la variabilité induite par ce traitement, les facteurs correctifs sont bornés à une multiplication ou une division des économies d'énergie d'un facteur 2 au maximum. À cette fin, les consommations de chacun des gestes ont donc été pondérées par le coefficient :

$$\frac{Conso_{menage}}{somme(Conso_{gestes})}$$

Ainsi la somme des gains de chaque geste pris isolément n'est pas égale aux gains de la somme des gestes. Cette correction conduit à diminuer d'environ 11 % à 21 % les gains énergétiques totaux, selon qu'ils sont estimés en énergie finale ou primaire.

3.1.2 Économies conventionnelles ou Économies réelles

L'estimation des gains énergétiques calculés par cette méthode est dite conventionnelle : elle repose sur un calcul de consommation basé sur des hypothèses (températures moyennes de chauffage, périodes de chauffe, durée d'occupation, etc.) qui peuvent conduire à un écart avec les économies réelles d'énergie.

Ces économies d'énergie conventionnelles peuvent être différentes des économies d'énergie réelles pour diverses raisons : la qualité technique réelle des rénovations peut être inférieure à celle anticipée ; l'amélioration du confort induite par les rénovations peut entraîner un changement de comportement, notamment en terme de choix de température ambiante du logement des ménages après rénovation. Le ménage choisit alors de « reconvertir » une partie des gains énergétiques et donc économiques en gains de confort.

C'est ce que l'on appelle « l'effet rebond », qui peut induire une augmentation des consommations réelles une fois certains gestes de rénovation réalisés. Plus largement, la consommation conventionnelle, avant comme après travaux, demeure un calcul théorique qui peut s'éloigner de la consommation réelle, celle-ci dépendant des usages du ménage occupant.

C'est pourquoi des travaux ultérieurs seront conduits dans le cadre de l'Observatoire national de la rénovation énergétique pour mettre en regard les rénovations énergétiques et l'évolution des données réelles de consommation d'énergie.

3.2 Gestion du secret statistique

Dans la perspective de la diffusion des données individuelles de l'enquête Tremi 2020, le traitement du secret statistique associé a constitué une étape cruciale. Avec l'aide de la division Sondages de l'Insee, des règles de diffusion ont pu être établies, tant pour les fichiers destinés au grand public (*open data*) que pour le réseau des chercheurs (*via* le CASD¹⁹).

La problématique à laquelle nous faisons face était la réidentification des personnes. En effet, nous ne disposions dans nos bases d'aucune variable directement identifiable (ni nom, ni prénom, ni adresse, ni téléphone, ni *mail* ...). Celles-ci avaient été transmises à notre prestataire de collecte, en appliquant les règles établies par la Statistique publique, en particulier en établissant une demande d'habilitation, destinée aux sous-traitants de l'enquête, pour la réception de fiches-contacts issues de Fideli.

Ainsi, comme préconisé par le guide du secret statistique²⁰, nous avons retiré des bases accessibles au grand public :

- la commune de résidence ;
- le code postal ;
- le département de résidence.

Restaient quelques variables indirectement identifiables comme l'âge, le sexe ou la PCS de la personne de référence (PR). Afin de limiter tout risque de réidentification des personnes à partir de ces caractéristiques :

- pour l'âge des répondants, notre base de données comportait une variable en tranche ainsi qu'une variable quantitative. Nous n'avons sélectionné dans les livrables destinés à l'*open data* que la variable reconstituée, ce qui ne semblait pas trop gênant pour les besoins de l'enquête et limitait largement les risques de réidentification ;
- nous avons choisi de restreindre la maille géographique la plus fine à la région. Le plan de sondage de notre enquête ayant été conçu afin que les données diffusées soient représentatives au niveau régional, cette solution nous a donc semblé cohérente.

Concernant les revenus, notre questionnaire proposait 10 tranches aux répondants. Compte tenu du caractère sensible de cette variable, qui n'est pas une variable d'intérêt principale de notre enquête, nous avons fait le choix de construire trois tranches de revenus, qui nous semblaient apporter suffisamment d'information aux utilisateurs « grands publics ».

De plus, nous avons considéré préférable de fusionner, dans ces fichiers de diffusion, les codes régions des Drom, et d'inclure les données de la Corse à celles de la région Provence - Alpes - Côte d'Azur. Cela permettait d'éviter que les utilisateurs ne puissent effectuer des calculs sur le nombre de rénovations sur ces territoires, où les données étaient trop fragiles pour être diffusées.

En procédant ainsi, nous avons fortement réduit le risque de violation du secret statistique. La fonction `createSdcObj` du *package* `sdcMicro` nous a permis d'estimer le risque d'identification des individus statistiques. Le fait d'utiliser des âges en tranches au lieu de la variable discrète a permis de le réduire de 18 %. Puis, restreindre les tranches de revenus a permis une diminution de ce risque de de 1,5 % supplémentaire. Enfin, le fait de regrouper les Drom ensemble, et la Corse avec la région Paca, a baissé ce risque de réidentification de 0,3 % de plus. Nous sommes ainsi passés de 20 % des données potentiellement identifiantes à 0,2 %.

19. Centre d'accès sécurisé aux données

20. accessible ici : <https://www.insee.fr/fr/information/1300624>

Toutefois, nous avons fourni une base de données enrichie à destination de la communauté des chercheurs, au travers du CASD. Voici ses caractéristiques :

- sa maille géographique la plus fine est la commune ;
- elle possède les 10 tranches de revenus de notre questionnaire ;
- la variable discrète de revenus y a été conservée.

Conclusion

Les résultats de l'enquête Tremi 2020 ont été publiés sous l'égide de l'Observatoire national de la rénovation énergétique. Cette instance, créée au sein du Commissariat général au développement durable (CGDD), produit un tableau de bord sur les dynamiques de rénovation.

Si l'alimentation du tableau de bord de l'observatoire repose fortement sur des sources administratives (dispositifs fiscaux et incitatifs, recueils de diagnostics techniques, etc.), certaines informations, par exemple les gestes de rénovation réalisés en dehors de tout cadre incitatif, nécessitent la mise en place d'enquêtes statistiques dédiées, telles que Tremi, sur le champ des maisons individuelles, et Treco, sur les copropriétés.

L'enquête Treco, encore en projet, vise à compléter les indicateurs sur le champ du logement collectif, qui représente environ 40 % du parc résidentiel, est moins bien connu et plus difficile à appréhender en raison des interactions entre parties communes et parties privatives ainsi que de la multiplication des parties prenantes (occupants, copropriétaires, syndicats, etc.). Le choix du protocole et du plan d'échantillonnage de l'enquête forment des défis méthodologiques.

Bibliographie

- [1] Albouy V. et Legleye S., « Conditions de vie pendant le confinement : des écarts selon le niveau de vie et la catégorie socioprofessionnelle », Insee Focus n°197, juin 2020.
- [2] Beaumont J.F. et Haziza D., « On the construction of imputation classes in surveys », *International statistical review*, 2007.
- [3] Caumont R., Merly-Alpa T., Rathle J.P. « La rénovation énergétique des maisons individuelles : résultats de l'enquête Tremi 2020 », *in* « La rénovation énergétique des logements : bilan des travaux et des aides entre 2016 et 2019 », Document de travail, Observatoire de la rénovation énergétique, pp. 61-116, mai 2021.
- [4] Deroyon T., « La correction de la non-réponse par repondération », Document de travail, Insee, 2017.
- [5] Little A., Roderick J. et Rubin B., « *Statistical Analysis with Missing Data* », Wiley Series in Probability and Statistics, 2002.
- [6] Matei A. et Ranalli G., « Traitement de la non-réponse non ignorable dans les enquêtes : une approche de modélisation par variables latentes », *Techniques d'enquête*, Statcan, juin 2015.
- [7] Osier G., « Traitement de la non-réponse non ignorable par calage généralisé : une simulation à partir de l'enquête budget des ménages au Luxembourg », *JMS* 2012.
- [8] Rebecq, A., « Icarus : un package R pour le calage sur marges et ses variantes », 9^e colloque francophone sur les sondages, Gatineau (Canada), 2016.
- [9] Merly-Alpa T., « Mesurer la rénovation énergétique des logements, de Tremi à Treco », 11^e colloque francophone sur les sondages, Bruxelles, 2021.

[10] Le Jeannic T. et Kraszewski M., « Les rénovations énergétiques aidées du secteur résidentiel entre 2016 et 2019 », *in* « La rénovation énergétique des logements : bilan des travaux et des aides entre 2016 et 2019 », Document de travail, Observatoire de la rénovation énergétique, pp. 11-60, mai 2021.

A Annexe 1 : correspondance entre les gestes et leurs codes

TABLE 6 – Q1 : Concernant votre résidence principale actuelle, parmi les travaux suivants, quels sont ceux qui ont débutés et se sont achevés entre 2017 et 2019 ?

11	Rénovation d'un pan ou de la totalité d'une toiture avec isolation de la toiture
12	Rénovation d'un pan ou de la totalité d'une toiture sans isolation de la toiture
13	Isolation d'un pan ou de la totalité d'une toiture sans rénovation de la toiture
14	Isolation du plancher des combles
15	Réfection de la toiture-terrasse avec isolation
16	Réfection de la toiture-terrasse sans isolation
21	Rénovation extérieure des murs donnant sur l'extérieur avec isolation (isolant + ravalement / bardage / crépi / peinture ...)
22	Rénovation extérieure des murs donnant sur l'extérieur sans isolation (ravalement, bardage, crépi, peinture, ...)
23	Rénovation intérieure des murs donnant sur l'extérieur avec isolation (isolant + peinture / papier peint...)
24	Rénovation intérieure des murs donnant sur l'extérieur sans isolation (peinture, papier peint...)
31	Rénovation du plancher bas avec isolation
32	Rénovation du plancher bas sans isolation
41	Pose ou remplacement de fenêtres, portes fenêtres ou baies vitrées
42	Pose ou remplacement sur les ouvertures de volets, ventelles, jalousies, ...
43	Remplacement ou pose de portes donnant sur l'extérieur
51	Remplacement ou installation d'un système de production de chauffage principal (énergies renouvelables comprises)
52	Amélioration du dispositif de régulation du chauffage (robinets thermostatiques, programmateur, systèmes de gestion du chauffage électrique)
53	Remplacement ou installation d'un système de production d'Eau Chaude Sanitaire (énergies renouvelables comprises)
54	Calorifugeage (isolation thermique) de tout ou partie d'une installation de production ou de distribution de chaleur ou d'eau chaude sanitaire
55	Mise en place ou remplacement d'un équipement de ventilation
56	Mise en place ou remplacement d'un équipement de climatisation
57	Mise en place ou remplacement d'un équipement permettant le rafraîchissement, hors climatisation (brasseur d'air, brise soleil, ...)
58	Installation de panneaux photovoltaïques (production d'électricité)