

Un outil pour l'étude des dépenses de santé et des « restes à charge » des ménages : le modèle *Omar*

Rémi Lardellier, Renaud Legal, Denis Raynaud et Guillaume Vidal*

Les « restes à charge » des ménages correspondent à la part de leurs dépenses de santé qui n'est couverte ni par l'assurance obligatoire de base ni par la couverture complémentaire. Leur connaissance est indispensable au pilotage du système d'assurance maladie. Au niveau macroéconomique, ils sont suivis grâce aux Comptes de la santé. Mais on s'attend à ce que leur poids soit très variable d'un individu à l'autre. Or il n'y a pas de source individuelle qui permette leur observation directe à niveau fin.

Cet article présente une démarche qui vise à combler cette lacune, celle du modèle *Omar* (Outil de Microsimulation pour l'Analyse des Restes à charge). Ce modèle reconstitue le partage de la dépense individuelle entre les trois financeurs : la Sécurité sociale, l'organisme complémentaire et l'individu. Il le fait en s'appuyant sur deux sources. La première est la source *Epas-SPS* qui apparie l'enquête *Santé et Protection Sociale* de l'Irdes et des données administratives de la Caisse nationale d'assurance maladie. La seconde est une enquête de la Drees auprès des organismes de couverture complémentaire. Cette dernière permet d'identifier les principales caractéristiques des contrats qu'ils offrent à leurs affiliés. Les informations de ces deux sources sont combinées et complétées par diverses techniques d'imputation.

Ce modèle a commencé à être utilisé pour des travaux d'études appliqués, tels que l'analyse de la redistribution opérée par le système de soins selon l'âge et selon le niveau de vie. L'objectif du présent article est plutôt méthodologique : il consiste à détailler les principales composantes du modèle, et notamment les procédures d'imputation retenues pour compléter les données qui ne sont pas directement observables dans les deux sources. Des tests de robustesse permettent d'apprécier la qualité des résultats obtenus.

* Rémi Lardellier, Renaud Legal et Denis Raynaud appartiennent au bureau Dépenses de santé et Relations avec l'Assurance Maladie (BDSRAM) de la Direction de la Recherche des Études, de l'Évaluation et des Statistiques (Drees) du Ministère de la Santé et des Sports. Lors de la rédaction de cet article, Guillaume Vidal appartenait au même bureau. Ce travail a bénéficié de présentations au séminaire *microsimula* de l'École d'Économie de Paris et au séminaire 3S de la Drees. Ces travaux font l'objet d'une collaboration avec l'Irdes et la DG-Trésor. Les auteurs remercient particulièrement Laurent Davezies, Valérie Albouy, Bérengère Mesqui, Claire Montialoux, Grégoire de Lagasnerie, Pierre-Yves Geoffard et Thierry Debrand pour leurs conseils et remarques méthodologiques, ainsi que Anne-Marie Brocas et Lucile Olier pour leurs relectures attentives de l'article. Les auteurs restent évidemment seuls responsables des limites ou inexactitudes de leur travail.

La part des dépenses de santé qui reste à la charge des ménages après intervention des couvertures de base et complémentaire constitue un enjeu important pour les politiques de santé. Sa connaissance permet de mesurer l'accessibilité financière des soins ou d'apprécier la solidarité entre malades et bien-portants induite par le système d'assurance maladie.

Grâce aux Comptes de la santé, la structure de financement et la dynamique des dépenses de santé sont bien connues au niveau macroéconomique. Ainsi, en 2009, la Sécurité sociale a financé 75,5 % des dépenses de santé, les organismes complémentaires 13,8 %, les ménages 9,4 % et l'État 1,3 % au travers essentiellement de la Couverture maladie universelle complémentaire, la CMU-C (source : Comptes de la santé).

En revanche, il n'existe pas à ce jour de sources d'information permettant de décliner l'exercice au niveau microéconomique. Les données de l'Assurance maladie renseignent uniquement sur la dépense individuelle présentée au remboursement et la part remboursée par l'assurance maladie obligatoire. Ces données ne permettent pas d'estimer la part de cette dépense restant à la charge de l'individu et de son ménage après intervention des organismes complémentaires, qui couvrent 94 % de la population. Les données des organismes complémentaires sont, quant à elles, souvent incomplètes et ne permettent pas d'avoir une vue d'ensemble, représentative pour l'ensemble des ménages, des remboursements qu'elles prennent en charge. Quant aux cas-types parfois mis en avant pour analyser la structure du financement des dépenses de santé et le reste à charge pour différents types de ménages, ils ont le défaut majeur de ne pas tenir compte de la fréquence réelle des cas décrits dans la population.

Pourtant, les structures de financement de la dépense de santé observées au niveau agrégé n'ont aucune raison d'être les mêmes pour tous les ménages, en raison notamment de l'hétérogénéité de la population face au risque maladie d'une part, et face à sa prise en charge par la Sécurité sociale et les organismes complémentaires d'autre part. Le reste à charge des ménages est probablement distribué de façon assez inégale au sein de la population. C'est pour évaluer cette hétérogénéité qu'a été entreprise la construction du modèle *Omar* (Outil de Microsimulation pour l'Analyse des Restes à charge). Il a été développé par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) du ministère en charge

de la Santé, en collaboration avec la Direction générale du Trésor du ministère en charge de l'Économie.

Ce modèle s'appuie en premier lieu sur l'échantillon 2006 de l'enquête *Santé et Protection Sociale (SPS)* conduite par l'Irdes, apparié avec les données de remboursement de l'*Échantillon Permanent d'Assurés Sociaux* des caisses d'assurance maladie (*Epas*). Cette source *Epas-SPS* doit être complétée sur trois points. D'une part, les données de revenu qu'elle fournit sont incomplètes, soit que les revenus n'aient pas été déclarés, soit qu'ils ne l'aient été qu'en tranches. Ensuite, l'*Epas* ne renseigne les dépenses de santé et leur prise en charge par l'assurance obligatoire que pour environ une moitié des effectifs de l'enquête *SPS*. Enfin, en matière de couverture complémentaire, l'enquête se borne à dire si l'individu est couvert ou non, dans quel type d'organisme, et par un contrat collectif ou individuel. On n'a pas davantage d'informations sur le niveau de garanties offertes.

Compléter des données de revenu qui ne sont connues qu'en tranche est un problème d'imputation assez classique. On le résout en estimant une équation de revenu et en appliquant la méthode des résidus simulés (cf. annexe 1). L'article se concentre sur les deux autres problèmes d'imputation. Pour les dépenses de santé, on recourt à une méthode de *hot deck* aléatoire stratifié consistant à dupliquer les dépenses observées pour d'autres individus de l'échantillon. Pour reconstituer les garanties offertes par la couverture complémentaire, on s'appuie sur une deuxième source, l'enquête sur les contrats d'assurance complémentaire les plus souscrits (*ECPS*) conduite par la Drees en 2008, dits encore « contrats modaux ». La méthode d'imputation consiste à attribuer un de ces types de contrats à chacun des individus de la base. Ceci permet de simuler le montant remboursé par l'assurance complémentaire au niveau individuel et donc d'avoir une vision complète de la structure de financement au niveau microéconomique.

La construction de ce modèle a ainsi impliqué un grand nombre de choix méthodologiques : le but de cet article est d'exposer le détail de ces choix. On reviendra tout d'abord sur l'enjeu de la démarche. Les données et leurs limites seront ensuite décrites. Puis on présentera les deux grandes étapes d'imputation mentionnées précédemment. Un exemple de résultats sera enfin fourni, le montant des remboursements complémentaires par décile de niveau de vie,

qui donnera l'occasion de tester à quel point les résultats sont sensibles au caractère aléatoire de la méthode d'imputation.

Mesurer les restes à charge : quels enjeux ?

La mesure des restes à charge est importante à plusieurs égards. Ils sont un des indicateurs du degré de solidarité du système de santé et de sa capacité à garantir l'accès aux soins. Ils constituent aussi un enjeu de santé publique puisqu'ils influent sur les comportements effectifs de recours aux soins. Or ces restes à charge sont *a priori* variables d'un individu à l'autre, compte tenu de l'hétérogénéité de la couverture complémentaire.

Une caractéristique importante du système de santé

En France, la participation à l'assurance maladie publique est obligatoire, d'où l'appellation « régime obligatoire ». La philosophie d'ensemble de ce régime est souvent résumée par l'expression : « contribution de chacun selon ses moyens, couverture de chacun selon ses besoins ». Les cotisations maladie sont en effet assises sur les revenus, sans lien *a priori* avec le niveau de risque individuel¹ et le niveau des remboursements. En revanche, les niveaux de remboursement des soins ne dépendent pas de la situation financière du patient, mais du degré de morbidité supposé et de son coût de prise en charge (taux de remboursement de 80 % des frais hospitaliers, de 70 % des honoraires de médecins, etc.), du service médical rendu (médicaments à 35 % ou 65 %, 100 %, etc.) ainsi que, le cas échéant, de la situation particulière du patient au regard des mécanismes d'exonération au moment des soins : affections de longue durée (ALD), hospitalisation de plus de 30 jours, maternité, etc.. Plus concrètement, on peut dire que le système d'assurance obligatoire français poursuit deux objectifs distincts : un objectif de solidarité entre bien portants et malades (centré sur le coût de la maladie indépendamment des revenus) et un objectif plus global d'accessibilité financière des soins. En outre, dans une perspective dynamique, chaque individu pouvant au cours de sa vie connaître des problèmes de santé, le système d'assurance publique et obligatoire a aussi pour objectif de procurer une assurance de long terme contre le risque maladie (Geoffard 2000).

Le reste à charge après remboursement par les seuls régimes de base, est un des éléments qui permet d'apprécier le degré de solidarité entre bien portants et malades induite par le système d'Assurance maladie obligatoire (AMO). Dans un système visant un niveau élevé de solidarité, les personnes les plus malades ne devraient pas avoir de dépenses sensiblement plus élevées que les autres personnes.

Le reste à charge qui intègre non seulement les remboursements de l'AMO mais aussi les remboursements des organismes complémentaires (Assurance maladie complémentaire, AMC) permet d'apprécier le second objectif, celui d'accessibilité financière des soins. On peut souhaiter que ce reste à charge soit faible pour les individus aux faibles revenus ou que, corollairement, le taux d'effort² correspondant soit raisonnable.

Les données de l'Assurance maladie permettent d'apprécier le reste à charge après AMO seulement. En revanche, elles n'offrent que peu d'informations sur les individus, hormis leur âge, leur sexe, le fait qu'ils bénéficient ou non de la CMU-C ou du dispositif ALD. De ce point de vue, l'appariement *Epas-SPS* qui sera mobilisé dans cet article permet une analyse plus riche, en mobilisant les informations sociodémographiques (catégories socioprofessionnelles, niveau d'études, état de santé déclaré, etc.) de l'enquête SPS. Par contre, cet appariement ne permet pas d'opérer le partage du reste à charge, après AMO, entre les ménages et les organismes complémentaires : la vision du reste à charge après AMO et AMC n'existe donc pas à ce jour. Or, du fait de son poids croissant, l'intervention des organismes complémentaires ne peut rester absente des études portant sur notre système de santé.

...qui influe sur l'accès aux soins...

Évaluer le reste à charge après intervention de la couverture complémentaire importe également d'un point de vue de santé publique, car les études empiriques, ont démontré le rôle déterminant du taux global de remboursement dans l'accès aux soins, à état de santé donné. Tel est le cas des études américaines appuyées sur l'expérience menée par la RAND Corporation dans les années 1970 (Newhouse, 1993) ou les

1. Contrairement au principe de tarification actuarielle, basé sur l'espérance mathématique du coût individuel, conditionnellement aux caractéristiques de la personne observées par l'assureur.

2. Défini comme le ratio du reste à charge sur le revenu disponible.

études sur l'extension du programme *Medicaid* (Currie, 2000). En allégeant la contrainte financière que la consommation de soins exerce sur le budget des ménages, l'assurance de base et l'assurance complémentaire permettent de réduire les disparités de consommation de soins entre les ménages modestes et les autres et d'éviter notamment une sous-consommation de soins au regard de l'état de santé.

En France, de nombreux travaux empiriques ont mis en évidence cet effet positif de la couverture complémentaire sur le recours aux soins (Caussat et Glaude, 1993, Breuil-Genier *et al.*, 1999, Dourgnon *et al.*, 2001, Raynaud, 2002 et 2005, Boisguérin, 2004). Les dépenses de spécialistes, comme de soins ambulatoires, des personnes disposant d'une assurance complémentaire santé sont plus élevées toutes choses égales par ailleurs, et notamment à état de santé identique (Caussat et Glaude, 1993, Breuil-Genier *et al.*, 1999). La couverture complémentaire joue un rôle important dans l'accès des ménages modestes à des prestations peu remboursées par la Sécurité sociale, comme les dépenses d'optique et en soins dentaires (Lengagne et Perronin, 2005).

C'est pourquoi les pouvoirs publics se sont efforcés de favoriser l'accès à la couverture universelle, notamment des plus modestes. Les études existantes confirment en effet que l'absence de couverture complémentaire dépend essentiellement du niveau des revenus et du milieu social (Kambia-Chopin *et al.*, 2008). Globalement, le taux de couverture par une complémentaire santé augmente avec le niveau de vie (cf. graphique I). La mise en place de la CMU-C, sous condition de ressources, instaurée par la loi du 27 juillet 1999, a représenté à cet égard un progrès décisif (cf. encadré 1). Plus récemment, l'Aide à l'acquisition d'une complémentaire santé, mise en place le 1^{er} janvier 2005, vise à faciliter l'accès des ménages dont les revenus dépassent légèrement le seuil d'éligibilité à la CMU-C à une complémentaire santé, en prenant en charge une partie des primes d'assurance.

...mais le niveau de couverture complémentaire est très hétérogène

Si la CMU-C et l'ACS permettent d'atténuer les inégalités en offrant une couverture complémentaire aux plus démunis, la proportion de personnes sans couverture complémentaire demeure toutefois bien plus élevée dans les premiers déciles de niveau de vie (20 % dans le premier décile contre 3 % dans le dernier).

D'autre part, le fait de disposer d'une couverture complémentaire ne présage pas du taux de prise en charge. Le marché de la complémentaire santé se répartit entre trois types d'organismes : les mutuelles, les sociétés d'assurances et les institutions de prévoyance. Les bénéficiaires d'une couverture complémentaire santé sont assurés dans 58 % des cas par un contrat individuel³ et dans 42 % des cas par un contrat collectif, acquis par l'intermédiaire de l'entreprise (Garnero et Rattier, 2011, et cf. encadré 2). Or, le niveau de garanties des personnes assurées par les organismes complémentaires santé varie fortement selon le type de contrat adopté : les contrats collectifs sont en moyenne de meilleure qualité que les contrats individuels⁴ (Couffinhal, Perronin, 2004 ; Francesconi *et al.*, 2006). Le niveau varie également selon le type d'organisme et l'âge du souscripteur.

En définitive, ce sont à la fois le taux d'adhésion à une couverture complémentaire et la nature de cette couverture complémentaire qui varient fortement selon le milieu social (Bocognano *et al.*, 2000). Avec une proportion croissante de contrats collectifs, et symétriquement une proportion décroissante d'individus non-couverts, au fur et à mesure que le niveau de vie s'élève, les résultats de l'appariement *Epas-SPS* illustrent ce résultat (cf. graphique I). Cette hétérogénéité des couvertures complémentaires rend indispensable une observation microéconomique des garanties offertes par ces contrats. Il serait illusoire de vouloir approcher les remboursements complémentaires par une simple imputation globale de remboursements observés à l'échelle macroéconomique. Or, les systèmes d'informations disponibles à ce jour ne permettent pas de connaître les garanties et les remboursements de l'AMC à l'échelle individuelle. En attendant la mise en place de nouveaux systèmes d'information, l'imputation des données manquantes par microsimulation constitue l'une des solutions.

3. Les contrats de mutuelles proposés aux fonctionnaires, bien qu'obtenus par l'intermédiaire de l'employeur, sont classés en individuel.

4. Même si des disparités existent au sein des contrats collectifs entre contrats à adhésion obligatoire et contrats à adhésion facultative. Ainsi une étude sur les données de l'enquête PSCE de 2003 a très clairement montré que les niveaux de garanties des contrats collectifs à adhésion obligatoire étaient en moyenne plus élevés que les contrats collectifs à adhésion facultative. Cette différence tient à plusieurs raisons. Premièrement, les contrats obligatoires ne sont pas soumis à l'anti-sélection. Les assureurs peuvent donc proposer des niveaux de garanties élevés sans risquer de voir certains salariés refuser de souscrire le contrat. Ensuite, les contrats obligatoires bénéficient d'avantages financiers spécifiques (réduction d'assiette des charges sociales pour les employeurs, déductions fiscales pour les salariés), qui leur permettent à budget donné, de proposer des niveaux de garanties supérieurs (Francesconi *et al.*, 2006).

Un modèle s'appuyant sur deux sources

Le modèle *Omar* procède aux imputations des données manquantes à partir de deux sources. La première est une base de données individuelles qui renseigne principalement sur les dépenses de santé et les caractéristiques sociodémographiques des individus. En matière de couverture complémentaire, elle n'informe que sur sa détention. L'autre est collectée auprès des organismes de couverture complémentaire : elle renseigne sur les caractéristiques des contrats qu'elles proposent.

L'appariement *Epas-SPS*

À ce jour en France, seules deux sources offrent à la fois des informations fines sur les caractéristiques sociodémographiques des individus et des données précises sur leur consommation médicale : l'enquête Insee *Handicap-Santé* appariée aux données de l'Assurance maladie (2003 et 2008) et l'appariement *Epas-SPS* (bisannuel). Le modèle *Omar* s'appuie sur l'appariement *Epas-SPS* 2006.

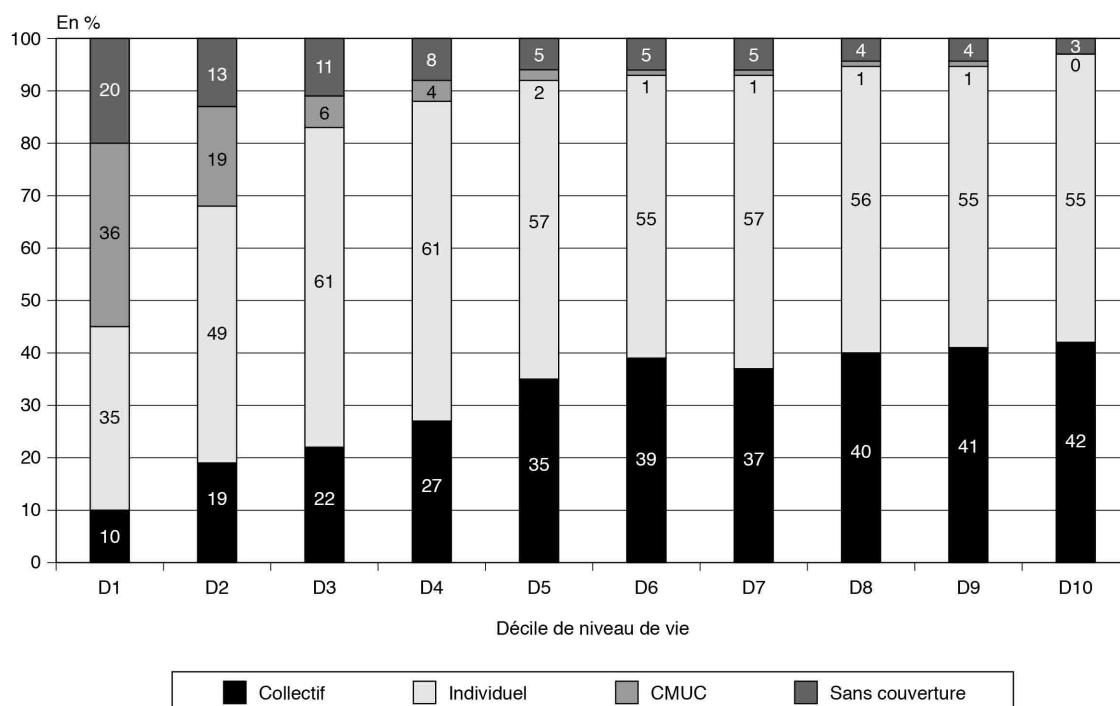
L'enquête *Santé et Protection Sociale* est une enquête réalisée par l'Irdes qui interroge tous les

deux ans 8 000 ménages ordinaires environ, soit 22 000 individus résidant en France métropolitaine. Elle permet de recueillir des données sur les caractéristiques sociodémographiques (âge, sexe, revenus, catégorie socioprofessionnelle, composition du ménage, etc.), sur l'affiliation à une couverture complémentaire ainsi que sur l'état de santé. La base de sondage servant à définir les individus interrogés est par construction incluse dans l'*Échantillon Permanent des Assurés Sociaux*, de manière à appairer les données de l'enquête avec les données de remboursement de l'Assurance maladie.

L'*Epas* est un échantillon au 1/600^{ième} d'assurés sociaux du régime général. En tant que fichier administratif de liquidation, l'*Epas* permet de connaître les dépenses médicales présentées au remboursement. Ce champ ne recouvre donc pas l'ensemble des dépenses⁵, contrairement au champ des Comptes de la santé. Il s'agit d'une particularité importante de l'étude présentée ici. Un traitement des données brutes a permis de

5. Ainsi, les dotations non individualisables reçues par les hôpitaux et les dépenses de santé non présentées au remboursement (automédication notamment) ne sont pris en compte dans cette étude.

Graphique I
Couverture maladie complémentaire et niveau de vie



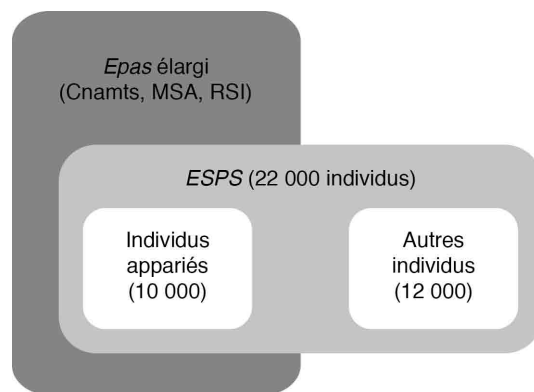
Lecture : 10 % des individus appartenant à un ménage du premier décile de niveau de vie sont couverts par un contrat collectif contre 42 % des individus appartenant à un ménage du dernier décile.
Champ : ménages ordinaires, France métropolitaine.
Source : appariement *Epas-SPS* 2006, traitement de la Drees.

reconstituer par type de soins et pour l'année 2006 les agrégats suivants : le montant total des dépenses de santé présentées au remboursement, des remboursements par la Sécurité sociale, des tickets modérateurs, des dépassements et des participations forfaitaires ventilés par grands postes de soins (omnipraticien, spécialiste, pharmacie, biologie, infirmière, masseur-kinésithérapeute, optique, dentiste et hôpital).

Enfin, la base de sondage de l'enquête *SPS* est en réalité un *Epas* élargi puisqu'à l'*Epas* de la Caisse nationale de l'assurance maladie des travailleurs salariés (Cnamts) s'ajoutent deux bases de données similaires relevant de la Mutualité sociale agricole (MSA) et du Régime social des indépendants (RSI). Ainsi, l'appariement *Epas-SPS* permet de couvrir les trois grands régimes de l'Assurance maladie. En revanche, la vision « assurés et ayants droit » offerte par l'*Epas* ne coïncidant pas avec la vision « ménages » de l'enquête *SPS*, l'appariement

ne permet de récupérer les dépenses de santé que pour une petite moitié des effectifs de l'enquête *SPS* (cf. graphique II) : en 2006, ce sont 53 % des individus interrogés dans l'enquête *SPS* qui n'étaient pas présents dans l'*Epas* et

Graphique II
Articulation de l'enquête SPS de l'Irdes et des données de l'Assurance maladie



Source : auteurs, appariement *Epas-ESPS* 2006.

Encadré 1

LA CMU-C ET L'AIDE À L'ACQUISITION D'UNE COMPLÉMENTAIRE SANTÉ

La Couverture Maladie Universelle Complémentaire (CMU-C), instaurée par la loi du 27 juillet 1999, permet d'accorder une couverture complémentaire gratuite aux Français et aux étrangers en situation régulière. Ce dispositif est un dispositif sous conditions de ressources : le plafond de ressources varie selon la composition du foyer. Il est de 627 euros mensuels pour une personne seule en métropole (698 euros dans les Dom). Il est majoré de 50 % pour la deuxième personne, de 30 % pour les troisième et quatrième personnes, et de 40 % par personne à partir de la cinquième personne. Lorsque ces conditions sont respectées, la CMU-C permet la prise en charge avec dispense d'avance de frais du ticket modérateur, du forfait journalier et des frais supplémentaires concernant les prothèses dentaires, l'orthopédie dento-faciale et certains dispositifs médicaux à usage individuel (lunettes, audio-prothèses, etc.). Par ailleurs, pour les patients concernés par ce dispositif, les professionnels de santé ont l'obligation de respecter les tarifs reconnus par la Sécurité sociale. Financé intégralement par les organismes complémentaires depuis 2009, le dispositif comptait 4,1 millions de bénéficiaires au 31 décembre 2009 pour les trois grands régimes d'assurance maladie : le régime général, le régime des travailleurs indépendants et le régime agricole, en métropole et dans les Dom.

L'enquête *Santé et Protection Sociale* permet de connaître les caractéristiques sociodémographiques des bénéficiaires de la CMU-C. Il s'agit d'une population jeune, dont un tiers appartient à une famille monoparentale, issue d'un milieu social modeste. Les bénéficiaires de la CMU-C déclarent un moins bon

état de santé que le reste de la population du même âge. Le taux de personnes souffrant d'une affection de longue durée (ALD) est d'ailleurs plus élevé parmi les bénéficiaires de la CMU-C que parmi les personnes ayant une couverture privée, à âge équivalent (Boisguérin, 2004, 2007, 2009).

Par rapport à l'absence de couverture, la CMU-C permet de diminuer le renoncement aux soins pour raisons financières (Boisguérin *et al.*, 2010). « Toutes choses égales par ailleurs », le bénéfice de la CMU-C permet de diviser par trois le risque de renoncement aux soins par rapport à une personne sans assurance complémentaire (Raynaud, 2005).

Mise en place dès le 1^{er} janvier 2005, l'Aide à l'acquisition d'une Complémentaire Santé (ACS) s'adresse aux ménages dont les revenus dépassent légèrement le seuil d'éligibilité à la CMU-C. Toute personne dont les ressources sont comprises entre le plafond de la CMU-C et 20 % de ce plafond peut prétendre à l'ACS. Ce dispositif prend la forme d'un bon d'achat qui vient réduire le montant que l'assuré doit payer au titre des primes d'assurance. Il est destiné à gommer les effets de seuil induits par la CMU-C. Toutefois, malgré des niveaux de subvention non négligeables (proches de 50 % du coût du contrat moyen acheté par les bénéficiaires de l'ACS en 2006), le dispositif ne semble pas susciter une demande à la hauteur de celle qui était attendue : en novembre 2009, ce sont seulement 501 681 attestations qui avaient été utilisées (Fonds de financement de la CMU, 2010) pour une population de personnes éligibles estimée à environ 2 millions en 2005 (Grignon et Kambia-Chopin, 2010).

dont les dépenses ne sont donc pas connues. Il s'agit d'un premier ensemble de données que nous sommes amenés à imputer.

L'enquête de la Drees sur les contrats complémentaires les plus souscrits

L'*Epas-SPS* se bornant par ailleurs à indiquer si l'individu est affilié ou non à une couverture complémentaire⁶, il faut s'appuyer sur une autre source pour évaluer comment cette couverture vient compléter les remboursements de l'AMO. Cette deuxième source est l'enquête sur les contrats les plus souscrits (*ECPS*), menée par la Drees.

Cette enquête est menée chaque année auprès des trois types d'organismes de couverture complémentaire (OC) : les sociétés d'assurances, les instituts de prévoyance et les mutuelles. Les questions portent sur les contrats couvrant le plus grand nombre de personnes, qu'ils fassent encore partie de l'offre commerciale ou non, d'où l'appellation de contrats modaux. Pour les contrats individuels, les questions portent sur les trois contrats couvrant le plus grand nombre de personnes au sein de l'ensemble des contrats individuels. Ils sont dénommés 1^{er} contrat

modal, 2^{ème} contrat modal et 3^{ème} contrat modal. Pour les contrats collectifs, les questions portent sur les deux contrats couvrant le plus grand nombre de personnes au sein de l'ensemble des contrats collectifs, que l'adhésion soit obligatoire ou facultative, que le contrat soit « sur-mesure » ou standard. Ils sont dénommés 1^{er} et 2^{ème} contrat modal.

Les contrats couvrant les bénéficiaires de la CMU complémentaire sont hors du champ de l'enquête. Les organismes concernés relèvent de la France métropolitaine et des départements d'Outre-mer (Dom). L'enquête permet d'une part de rassembler quelques informations sur l'activité globale de l'organisme et d'autre part de décrire de façon détaillée les caractéristiques et les garanties proposées par les cinq contrats modaux.

Le nombre de contrats ainsi collectés est cependant important, de l'ordre de 900 environ. Pour les besoins des imputations qui suivent, un regroupement intermédiaire en grandes catégories est aussi utilisé. Il a été construit au moyen

6. L'enquête SPS fournit aussi des éléments sur la qualité ressentie de la couverture par les assurés. Le caractère subjectif de cette information n'a pas permis de l'intégrer pour l'instant dans le modèle Omar.

Encadré 2

LA PROTECTION SOCIALE COMPLÉMENTAIRE D'ENTREPRISE

Conduite pour la première fois en 2003 par l'Irdes auprès des établissements et reconduite en 2009 grâce à un cofinancement de la Cnamts et de la Drees, l'enquête *PSCE (Protection Sociale Complémentaire d'Entreprise)* a permis de mettre en évidence une grande hétérogénéité de l'offre de contrats collectifs selon la taille des entreprises et les secteurs d'activité. Plus l'entreprise comporte de salariés et plus elle est susceptible de proposer un contrat collectif à ses employés. Les salariés des petites entreprises du secteur des services et encore plus de l'industrie restent majoritairement exclus de la protection maladie collective. Enfin, ce sont les cadres et les professions intermédiaires qui ont les probabilités les plus élevées d'avoir une complémentaire maladie d'entreprise (Couffinhal *et al.*, 2004, Francesconi *et al.*, 2006).

Les résultats de l'édition 2009 de l'enquête *PSCE* montrent que tous les établissements de plus de 500 salariés proposent une couverture complémentaire d'entreprise contre seulement 34 % des très petites entreprises (TPE de moins de 10 salariés), confirmant ainsi le lien entre taille de l'entreprise et offre de complémentaire. En 2009, les cadres ont toujours plus de chance d'avoir une couverture d'entreprise. En 2009, le principal changement par rapport

à 2003 tient à la recomposition récente de l'offre de complémentaire santé des entreprises aux salariés sous l'effet de la loi Fillon du 21 août 2003 (Guillaume et Rochereau, 2010).

Avant celle-ci, les employeurs et les salariés bénéficiaient d'exonérations fiscales et sociales sur leurs cotisations respectives quel que soit le type de contrat de complémentaire santé collective souscrit. La loi Fillon du 21 août 2003 a modifié les conditions d'exonérations fiscale et sociale. Au-delà d'une période transitoire prenant fin le 31 décembre 2008, les contrats doivent désormais respecter un certain nombre de contraintes pour pouvoir bénéficier de ces exonérations. Le régime doit notamment avoir un caractère collectif. Les contrats doivent être « responsables ». L'adhésion doit avoir un caractère obligatoire (pour une présentation plus complète de ces contraintes, voir Guillaume et Rochereau, 2010).

Afin de conserver les exonérations fiscales et sociales, bon nombre d'entreprises ont refondu leur offre de contrats, si bien qu'en 2009, le tiers des contrats datent de moins de deux ans. Sous l'effet de la loi Fillon, en 2009 le ou les contrats sont exclusivement obligatoires dans plus de trois établissements sur quatre.

d'une Classification Ascendante Hiérarchique (CAH) réalisée à la suite d'une Analyse en Composantes Principales (ACP)⁷ (Arnould et Vidal, 2008). Cette typologie aboutit à un regroupement des contrats en quatre classes : les contrats A qui offrent les meilleures garanties jusqu'aux contrats D offrant les moins bonnes.

La classe d'appartenance - le niveau de couverture - d'un contrat est très dépendante du type de contrat (collectif/individuel) et du type d'organisme complémentaire qui le propose (mutuelle, institut de prévoyance, assureur privé). Par ailleurs, la structure par âge des assurés varie fortement d'une classe à une autre, signe d'un lien fort entre âge et niveau de couverture (cf. tableau 1). Cependant, la structure par âge au sein des contrats n'est demandée dans l'enquête que depuis 2008. C'est pour tenir compte de cette information fortement reliée à la qualité de la couverture que nous avons choisi de retenir l'édition 2008 de l'enquête. Le décalage temporel qui en résulte par rapport à l'appariement *Epas-SPS* n'est pas gênant, la répartition des contrats (A, B, C, D) en fonction du type d'organisme et du type de contrat ayant très peu évolué entre 2006 et 2008.

Une représentativité globalement satisfaisante

Ces sources présentent quelques limites. Les dépenses enregistrées dans l'*Epas* sont uniquement celles présentées au remboursement. Ainsi par exemple, les dépenses d'automédication n'apparaîtront pas dans *Omar*, alors qu'elles sont comptabilisées dans les Comptes de la santé. D'une manière générale, l'écart entre les résultats présentés dans les Comptes de la santé et ceux disponibles dans *Omar*, devra être apprécié suivant la différence de champ existant entre ces deux bases de données : les Comptes de la santé couvrent les Dom alors qu'*Omar* est restreint à la France métropolitaine.

Par ailleurs, l'enquête *SPS* n'interroge pas les ménages en institutions. Par conséquent, ces ménages aux dépenses élevées ne rentrent pas dans le champ de l'outil de microsimulation. Enfin, la taille de l'échantillon (environ 10 000 dépenses observées seulement) est faible pour apprécier correctement les dépenses

7. Ces techniques d'analyses de données permettent de maximiser la variance inter-classes et de minimiser la variance intra-classe.

Tableau 1
Répartition des contrats complémentaires les plus souscrits en fonction de la classe de qualité du contrat (A, B, C et D) du type d'organisme, du type de contrat et de l'âge de l'assuré

Type d'organisme	Mutuelle		Institut de prévoyance		Société d'assurance	
	Collectif	Individuel	Collectif	Individuel	Collectif	Individuel
Assurés de moins de 25 ans						
A	14	2	19	9	42	3
B	10	36	42	47	37	22
C	14	29	38	43	20	54
D	62	33	0	0	0	21
Assurés de 25 à 59 ans						
A	39	2	25	11	48	5
B	31	57	42	44	47	22
C	26	33	33	44	5	57
D	4	8	0	1	0	16
Assurés de 60 ans et plus						
A	7	3	25	12	28	5
B	11	53	43	29	69	23
C	80	37	33	58	2	52
D	2	6	0	1	0	20

Lecture : 14 % des assurés de moins de 25 ans couverts par un contrat collectif de mutuelle ont un contrat de qualité A (qualité supérieure).

Champ : contrats modaux.

Source : enquête Drees sur les contrats les plus souscrits, 2008.

hospitalières, alors même que les dépenses élevées sont très concentrées⁸.

Une autre limite est le fait que l'enquête Drees sur les contrats ne concerne que les contrats les plus souscrits. Par conséquent, les contrats de niches (offrant une très forte couverture ou à l'inverse proposant une couverture très faible), peu souscrits, sont notamment sous représentés, voire absents de l'enquête, et donc absents également du modèle de microsimulation.

Plus généralement, les contrats recueillis par l'enquête Drees ne couvrent que 63,2 % des personnes couvertes en individuel et 22,2 % des personnes couvertes en collectif⁹ (Garnero et Rattier, 2011). La faible représentativité en termes de personnes couvertes des contrats collectifs de l'enquête peut conduire à s'interroger sur la représentativité de ces contrats en termes de niveau de couverture. D'après une comparaison des résultats de la dernière enquête Drees sur les contrats modaux avec les résultats de la dernière édition de l'enquête *Protection Sociale Complémentaire d'Entreprise (PSCE 2009)* réalisée par l'Irdes, il semblerait que la qualité des contrats collectifs de l'enquête Drees sur les contrats modaux, et donc d'Omar, soit légèrement sous-estimée (Garnero et Rattier, 2011).

Ces problèmes de représentativité des contrats modaux sont susceptibles de biaiser nos résultats de restes à charge après Assurance maladie et couverture maladie complémentaire. Toutefois, l'imputation des contrats issus de l'enquête Drees conduit à retrouver au niveau agrégé des chiffres très proches de ceux des Comptes de la santé (cf. annexe 4). Cela ne garantit certes pas la fiabilité du partage AMO/AMC offert par l'outil de microsimulation au niveau microéconomique. Mais, *a contrario*, des chiffres agrégés très éloignés des Comptes auraient invalidé l'outil. Tel n'est pas le cas.

Reconstituer les dépenses de santé lorsqu'elles ne sont pas connues

En laissant de côté la question de l'imputation des revenus (cf. annexe 1), la première étape dans l'élaboration d'Omar va consister à imputer des dépenses de santé aux individus présents dans l'enquête SPS mais non-appariés à l'Epas.

Les dépenses de santé offrent deux types de spécificités dont cette imputation doit tenir compte. Certaines sont purement arithmétiques : par exemple la dépense totale portée au remboursement doit être égale à la somme du remboursement de la Sécurité sociale, du ticket modérateur, des dépassements et des participations forfaitaires. D'autres sont d'ordre statistique : par exemple, les assurés ayant des dépenses importantes en omnipraticien auront généralement des dépenses élevées en pharmacie (cf. tableau 2), si bien qu'une forte corrélation doit exister entre ces deux postes au niveau des dépenses observées, comme des dépenses imputées.

Le choix du *hot deck* stratifié

Deux méthodes peuvent être envisagées pour imputer les dépenses de santé des individus non appariés avec l'Epas. La première est une méthode d'imputation par équations et résidus simulés, la seconde est une méthode de *hot deck*.

Avant de s'intéresser aux avantages et aux limites de chacune des méthodes, il convient de rappeler les objectifs de cette imputation. Le but final de ce travail est de pouvoir reconstituer, pour chacun des ménages présents dans l'enquête SPS, l'ensemble de ses dépenses de santé, de ses remboursements AMO et de ses remboursements AMC. Le point essentiel à retenir ici est la reconstitution des remboursements AMC. En effet, afin de pouvoir simuler ces remboursements il faut pouvoir appliquer les garanties d'un contrat complémentaire aux dépenses de santé des individus. Or, les niveaux de remboursements ne sont pas équivalents pour tous les postes et tous ne sont pas spécifiés sous la même forme par les organismes d'AMC (taux ou forfait). Si l'on veut simuler ces garanties, il faut donc que les dépenses de santé des individus soient présentées à un niveau désagrégé.

8. Cette dernière limite explique aussi en partie pourquoi la structure de financement des dépenses hospitalières livrée par Omar est sensiblement différente de celles fournies par les Comptes de la santé. Une autre explication tient au fait que l'Epas fournit une dépense hospitalière individualisée (celle utilisée pour établir la facture du patient au moment de sa sortie), tandis que les comptes représentent la ventilation des coûts (qui ne sont pas tous individualisables).

9. Il apparaît donc que la représentativité de l'enquête Drees sur les contrats modaux est moins bonne sur les contrats collectifs que sur les contrats individuels. En effet, les contrats collectifs sont souvent bâtis sur mesure pour les entreprises si bien que l'offre de contrats collectifs est plus protéiforme que l'offre de contrats individuels (resserrée autour de quelques contrats seulement). Dès lors, interroger sur les contrats les plus souscrits pénalise davantage la représentativité des contrats collectifs modaux recueillis dans l'enquête que la représentativité des contrats individuels modaux recueillis par la même enquête.

Si l'on décide d'imputer les dépenses de santé à partir d'équations et de résidus simulés, plusieurs étapes sont nécessaires. Dans un premier temps, il faut que la modélisation mise en œuvre détermine à la fois la consommation ou la non-consommation puis, le cas échéant, le montant de cette consommation engagée. Surtout, il faut que cette modélisation soit effectuée pour chacun des postes de soins pour lesquels les organismes d'AMC spécifient un niveau de remboursement. Dès lors ce n'est plus *une* équation qu'il faut envisager mais plus d'une dizaine d'équations. Enfin, on a indiqué que les postes de soins étaient, hors optique et dentaire, fortement corrélés entre eux (cf. tableau 2) : l'imputation nécessite donc autre chose qu'une succession d'équations indépendantes, mais un système d'équations toutes corrélées les unes avec les autres. À supposer qu'un tel système puisse être construit, il faut ensuite songer à ajouter un résidu simulé à chaque valeur imputée. Dans *Omar*, cette méthode par équations et résidus simulés a été appliquée à l'imputation des données manquantes pour les revenus (cf. annexe 1), mais, après les premières tentatives, elle s'est avérée trop complexe à mettre en œuvre pour l'imputation des dépenses de santé.

Les techniques de *hot deck* permettent de traiter le même problème de manière plus simple, en répliquant des données existantes plutôt qu'en inventant de nouvelles. Pour imputer des données, il faut que la population soit segmentée entre répondants et non-répondants : la méthode consiste alors à associer à chaque non-répondant un répondant, et à dupliquer les données des

répondants pour remplacer la non-réponse. On peut voir le *hot deck* comme un pendant non-paramétrique de l'imputation par équations et résidus simulés. Ici, le résidu est empirique.

L'avantage des techniques *hot deck* est majeur : là où l'imputation par équations impose un lourd système d'équations pour déterminer les dépenses de santé individuelles, le *hot deck* permet, par l'association d'un donneur et d'un receveur de dupliquer immédiatement chacun des postes de soins. Ces valeurs imputées étant des dépenses constatées, les deux contraintes évoquées plus haut sont immédiatement satisfaites, à savoir l'identité comptable entre les dépenses et la somme des remboursements de l'AMO et du ticket modérateur, et la corrélation entre dépenses sur les différents postes.

Parmi les méthodes de *hot deck* existantes, celle que nous avons retenue est la variante dite du *hot deck* aléatoire stratifié. Il est qualifié d'aléatoire, car le donneur est sélectionné par un tirage aléatoire et deux jeux d'imputation n'associeront pas le même donneur à un receveur précis (contrairement au *hot deck* métrique par exemple). Il est qualifié de stratifié, car le tirage aléatoire du donneur se fait parmi une population de donneurs ayant les mêmes caractéristiques que le receveur.

On peut ainsi résumer les étapes à mettre en œuvre pour imputer les dépenses de santé des individus non appariés à l'*Epas* :

- 1) cerner les facteurs explicatifs de la consommation de soins ;

Tableau 2
Corrélation entre les dépenses totales des différents segments de soins

Dépenses	Coefficient de corrélation linéaire								
	Biologie	Dentistes	Hôpital	Infirmières	Kinésithérapie	Omnipraticiens	Optique	Pharmacie	Spécialistes
Biologie	1,00	0,06	0,23	0,17	0,15	0,39	0,08	0,41	0,36
Dentistes	0,06	1,00	0,00	0,01	0,02	0,03	0,05	0,05	0,02
Hôpital	0,23	0,00	1,00	0,12	0,11	0,20	0,01	0,24	0,07
Infirmiers	0,17	0,01	0,12	1,00	0,17	0,23	0,00	0,22	0,13
Kinésithérapie	0,15	0,02	0,11	0,17	1,00	0,28	0,05	0,16	0,06
Omnipraticiens	0,39	0,03	0,20	0,23	0,28	1,00	0,07	0,38	0,15
Optique	0,08	0,05	0,01	0,00	0,05	0,07	1,00	0,05	0,03
Pharmacie	0,41	0,05	0,24	0,22	0,16	0,38	0,05	1,00	0,24
Spécialistes	0,36	0,02	0,07	0,13	0,06	0,15	0,03	0,24	1,00

Lecture : 1^{ère} ligne sixième colonne, le coefficient de corrélation linéaire entre les dépenses en omnipraticien et la biologie est de 0,39.
Champ : ménages ordinaires, France métropolitaine, dépenses observées.
Source : appariement Epas-SPS 2006, Irdes, Cnamts, RSI, MSA, - traitement Drees.

- 2) choisir, parmi ces facteurs, des variables, de manière à définir des strates d'imputation ;
- 3) ventiler les individus au sein de ces strates ;
- 4) pour un *receveur* appartenant à une classe donnée, tirer au sort (avec remise) un *donneur* parmi les donneurs appartenant à la même strate ;
- 5) appliquer au receveur le profil de dépense du donneur tiré au sort

Les paragraphes suivants visent à détailler quelques points importants de cette méthode.

Choisir des strates d'imputation

La mise en œuvre du *hot deck* suppose que les donneurs aient des caractéristiques proches des receveurs, autrement dit de raisonner à l'échelle de strates d'imputations. Ces strates sont construites au moyen de variables auxiliaires qui expliquent bien la dépense. Le détour par des équations de dépense reste donc nécessaire, même s'il est plus léger que lorsque l'on combine équations et résidus simulés.

Pour choisir ces équations, il faut tenir compte de ce que les dépenses d'optique et de dentaire ne sont pas corrélées avec les autres dépenses de soins. C'est pourquoi ces deux types de dépenses ont été imputés séparément. De manière à bâtir les strates d'imputation pour la mise en œuvre du *hot deck* stratifié, trois modèles de consommation de soins ont donc été estimés au préalable. Le premier pour modéliser la dépense en optique, le deuxième pour modéliser la dépense en dentaire, et le dernier pour le reste de la dépense de santé.

Du fait de l'existence d'un nombre important d'individus n'ayant aucune dépense de santé dans l'année, la distribution des dépenses de santé présente un point d'accumulation en 0. D'un point de vue économétrique, cette particularité oblige à modéliser la dépense en deux temps. D'abord, on modélise le fait d'avoir ou non une dépense de santé, à l'aide d'un modèle Probit : c'est l'équation de participation. Ensuite, on modélise le montant de la dépense (ou son logarithme, compte tenu de la distribution log-normale des dépenses de santé), conditionnellement au fait d'avoir recouru : c'est l'équation de dépense.

$$\begin{cases} P_i^* = X_{2i}'\beta_2 + u_{i2} & (\text{équation de participation}) \\ \log(dep_i) = (X_{1i}'\beta_1 + u_{i1}) P_i & (\text{équation de dépense}) \end{cases}$$

Où P_i^* est la probabilité d'engager des dépenses de santé, dep_i le montant des dépenses de santé engagées, X_{1i} et X_{2i} les vecteurs de variables explicatives (âge, sexe, état de santé déclaré niveau de vie etc.) et u_{1i} et u_{2i} des résidus aléatoires indépendants suivant une loi normale $(0, \sigma^2)$.

Deux types de méthode économétriques peuvent être utilisées dans ce cas de figure : le *two-part model*, utilisé notamment aux États-Unis dans les travaux de la RAND (Newhouse, 1993) ou le modèle de sélection (*Sample selection model*, Heckman, 1979). Ces deux méthodes ne sont pas équivalentes. La première s'intéresse à l'espérance de la dépense conditionnellement à une consommation positive, et les équations de participation et de dépense peuvent alors être estimées de manière indépendante. La deuxième s'intéresse à l'espérance non conditionnelle de la dépense, si bien que dès lors que la décision de consommer et le montant de dépense ne sont pas indépendants, il est nécessaire dans l'estimation de tenir compte de la corrélation des résidus des deux équations, en introduisant dans la deuxième équation l'inverse du ratio de Mills estimé à partir de la première équation.

Les travaux cherchant à mesurer les déterminants de la dépense de santé privilégient généralement l'estimation par un modèle de sélection car l'intérêt se porte sur l'espérance non conditionnelle de dépense (voir ainsi Pichetti *et al.*, 2009). En revanche, les travaux de microsimulation privilégient plutôt l'estimation par un *two-part model* (voir ainsi Dormont *et al.*, 2006) car l'objectif principal n'est pas d'interpréter les coefficients estimés de la deuxième équation. De nombreux travaux théoriques et empiriques débattent du meilleur choix possible (voir ainsi Leung et Yu, 1996, Beeuwkes et Zaslavsky, 2003). En particulier, des travaux de simulation par la méthode de Monte-Carlo ont montré que l'estimation par *two-part model* pouvait conduire à de meilleurs résultats que les modèles de sélection même quand le vrai modèle générateur des données est un modèle de sélection (Manning, Duan, Rogers 1987). Leung et Yu ont montré que ce cas de figure pouvait arriver s'il y a trop de colinéarité entre l'inverse du ratio de Mills et les variables explicatives de la deuxième équation.

Dans la présente étude, compte tenu de notre objectif de microsimuler des dépenses de soins plutôt que de mesurer les facteurs explicatifs de la dépense pour pouvoir les interpréter, l'estimation de ces deux équations a donc été

faite dans le cadre d'un *two-part model*, qui présente de meilleures propriétés empiriques (cf. annexe 2, tableaux A à F).

Dans le cadre d'une imputation par *hot deck* stratifié, il n'est pas possible de retenir la totalité des facteurs explicatifs de la dépense de santé dans la construction des classes d'imputation. En effet, un nombre trop important de variables conduirait à des sous-populations de taille très faibles avec très peu de donneurs. C'est pourquoi, dans une logique de parcimonie, nous avons sélectionné les déterminants les plus pertinents, quitte à ne pas retenir toutes les variables significatives au seuil de 5 %.

Le problème des donneurs en nombre insuffisant ou des donneurs inexistant

Malgré le principe de parcimonie précédent, il se peut qu'au sein de certaines strates le nombre de donneurs soit encore trop faible (voire nul). C'est pourquoi il est nécessaire de s'imposer deux contraintes techniques supplémentaires, dans la mise en œuvre du *hot deck* :

- un nombre minimal de donneurs par classe (C_1) ;
- le fait que le nombre de donneurs doit être supérieur au nombre de receveurs¹⁰ (C_2)

Ces deux contraintes permettent notamment d'éviter qu'un individu atypique soit utilisé un trop grand nombre de fois.

Soumis à ces contraintes et à l'absence de donneurs dans certaines classes, le *hot deck* s'effectue de manière itérative. Dans un premier temps, les classes d'imputation sont formées à partir de toutes les variables auxiliaires. L'imputation par donneurs est réalisée pour tous les receveurs des classes pour lesquelles les contraintes C_1 et C_2 sont respectées. Pour les classes pour lesquelles l'une de ces contraintes ne serait pas vérifiée, on supprime alors une des variables auxiliaires ayant servi à la définition des strates et on reforme des strates sur la base des variables auxiliaires restantes. Le choix de retirer une variable se fonde sur son niveau de significativité aussi bien dans l'équation de participation que dans l'équation de montant. L'imputation par donneurs est alors à nouveau réalisée pour tous les receveurs des classes pour lesquelles les contraintes C_1 et C_2 sont respectées. Ce processus itératif se poursuit jusqu'à ce que tous les individus soient imputés. Au fil des différentes

étapes, les variables auxiliaires sont retirées les unes après les autres (cf. tableau 3).

Après un premier essai d'imputation par *hot deck*, l'ordre des variables peut éventuellement être modifié à la marge afin d'imputer un plus grand nombre d'individus au cours des premières étapes. Si on prend l'exemple des dépenses hors optique et dentaire, c'est plus de la moitié de la population qui est appariée avant le quatrième tour et le R^2 des modèles associés à ces stratifications est généralement compris entre 0,41 et 0,36. Moins de 3 % de la population est traitée lors des deux dernières étapes où le R^2 des modélisations associées est de 0,28 (cf. tableau 3)

Calage des dépenses sur les données hospitalières de l'Epas

Du fait des sources utilisées, cette méthode d'imputation conduit à une sous-estimation des dépenses hospitalières. En effet, le taux de réponse à l'enquête *SPS* est plus faible chez les personnes en mauvaise santé. Par ailleurs, seules les personnes en « ménage ordinaire » sont interrogées dans l'enquête *SPS*. Les personnes vivant en institution, qui ont potentiellement des dépenses hospitalières importantes, ont été exclues du champ de l'enquête. Les données issues de *SPS* ont donc tendance à sous-estimer les dépenses hospitalières par rapport à l'*Epas*. Ce biais peut être redressé en calant les données *SPS* sur la « queue de distribution » des dépenses hospitalières de l'*Epas*, pour obtenir une proportion de personnes ayant des dépenses hospitalières importantes équivalente à celle de l'*Epas*. Ce calage doit être fait à la fois au niveau des individus et des ménages.

Les variables de calage retenues sont celles utilisées dans *SPS* pour préserver la représentativité initiale de l'échantillon : âge, sexe, régime d'affiliation et taille du ménage (Allonier *et al.*, 2008) auxquelles sont ajoutées une variable de dépenses hospitalières et la proportion d'ALD par niveau de dépenses à l'hôpital (cf. annexe 3). Le calage est réalisé à partir de la macro Calmar mise au point par l'Insee (Sautory, 1993)¹¹.

10. Cette dernière contrainte étant obligatoire dans le cadre d'un *hot deck* avec tirage sans remise mais elle peut être levée si l'on procède à un tirage des donneurs avec remise

11. Parmi les quatre méthodes proposées, c'est la méthode linéaire tronquée qui a été retenue ici. Elle présente l'avantage de ne pas aboutir à des poids négatifs et permet, en faisant varier les bornes maximale et minimale des rapports de poids, d'arbitrer entre une faible dispersion, une faible étendue ou l'allure générale de la distribution.

Un test de validité

Pour s'assurer que la méthode mise en place ne conduit pas à imputer des valeurs aberrantes, un essai d'imputation a été effectué sur des individus pour lesquels les dépenses de santé sont connues (*i.e.* présentes dans l'*Epas*). On peut par exemple comparer les moyennes obtenues après *hot deck* avec les « vraies » moyennes pour les dépenses en omnipraticien (cf. tableau 4).

Plus précisément, un échantillon représentant un tiers de la population appariée entre l'*Epas* et l'*ESPS* a été sélectionné aléatoirement. Pour ces 3 135 individus, les dépenses de santé sont observées et n'ont pas besoin d'être imputées par *hot deck*. Pour les besoins de l'exercice, on a supposé que leur dépenses de santé n'étaient pas connues et la procédure d'imputation leur a été appliquée comme aux non-répondants. Les données de ces faux non-répondants ont été imputées par *hot deck* stratifié sur la base des valeurs observées pour les deux tiers restants de l'appariement. La non-réponse étant fictive, il est possible de confronter valeurs observées et valeurs imputées. Les résultats de cet exercice sont satisfaisants. Cela était prévisible au niveau agrégé : le taux de

consommateurs (part des individus ayant engagé des dépenses d'omnipraticien dans l'ensemble de la population) ne varie que de 1,2 points entre les deux sources et les montants moyens ne présentent que des écarts très faibles ; dépenses, remboursements AMO, tickets modérateurs, dépassements et participations forfaitaires ne diffèrent jamais de plus d'un euro, en moyenne, suivant la source. Mais une bonne concordance s'observe aussi à un niveau plus fin. Les diagrammes en boîte des séries observée et simulée ne font pas apparaître de divergences majeures entre l'*Epas* et les données simulées pour ce qui concerne les valeurs centrales (moyennes, 1^{er} quartile et 3^{ème} quartile). Des différences ne commencent à apparaître qu'aux extrémités de la distribution. Le 99^{ème} percentile est ainsi supérieur de 60 euros pour les données simulées par rapport aux données observées (cf. graphique III-A).

Les différences ne deviennent très importantes que lorsque l'on passe au niveau totalement individuel. Au niveau individuel, la valeur imputée a en effet très peu de chances d'être équivalente à celle observée. Ce sont ainsi 74 % des écarts entre dépenses observées en omnipraticiens et dépenses simulées qui sont supérieurs, en valeur

Tableau 3
Déroulement de l'imputation par *hot deck* des dépenses de santé hors optique et dentaire

	1 ^{er} tour	2 ^{ème} tour	3 ^{ème} tour	4 ^{ème} tour	5 ^{ème} tour	6 ^{ème} tour	7 ^{ème} tour	8 ^{ème} tour	9 ^{ème} tour
Individus auxquels sont affectés des dépenses de santé									
Nombre	3 227	1 717	1 679	2 242	1 755	571	208	299	4
%	27,58	14,67	14,35	19,16	15,00	4,88	1,78	2,56	0,03
Variables de stratification									
Exonération	x	x	x	x	x	x	x	x	x
Consultation d'un omnipraticien	x	x	x	x	x	x	x	x	x
Régime	x	x	x	x	x	x	x	x	
Sexe et âge	x	x	x	x	x	x	x		
État de santé perçu	x	x	x	x	x	x			
Nature de la couverture complémentaire	x	x	x	x	x				
Nb de personnes dans le ménage	x	x	x	x					
Niveau de revenus	x	x	x						
Consultation d'un spécialiste	x	x							
Consommation de médicaments	x								
R ² du modèle associé (équation de montant)	0,4074	0,3997	0,3765	0,3764	0,3744	0,3687	0,3588	0,288	0,2824

Lecture : avec la totalité des variables de stratification (dix) et sous la condition que le ratio receveurs/donneurs n'excède pas 1,4, il est possible d'imputer 3 227 individus au premier tour de la procédure *hot deck*.

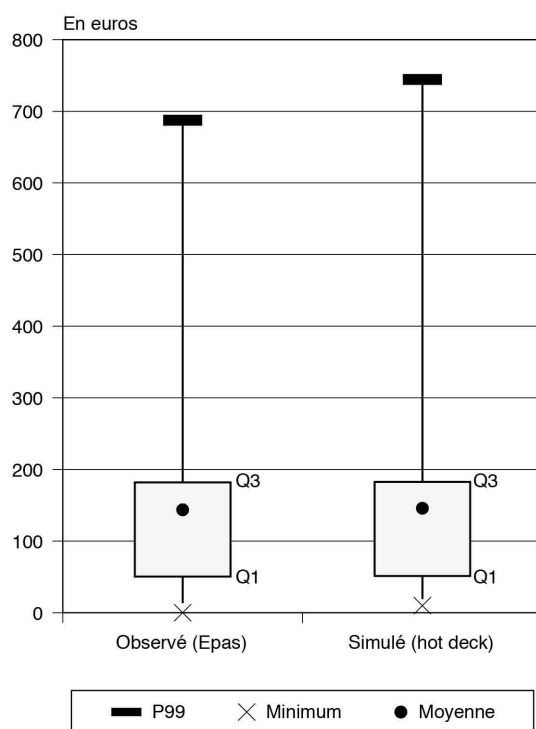
Champ : individus non appariés avec l'*Epas*.

Source : Omar 2006.

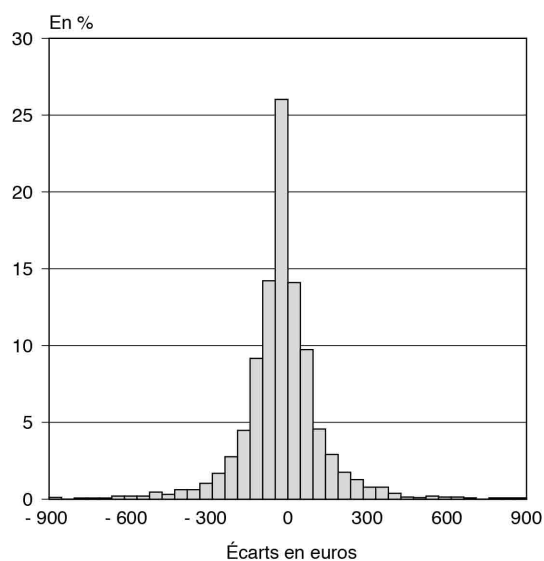
absolue, à 25 euros (cf. graphique III-B). Cette variabilité est à rapprocher des faibles valeurs de R^2 des régressions du *two-part model* présentées précédemment. Observer l'âge, l'état de santé perçu et diverses variables sociodémographiques ne permet évidemment pas de déduire précisément la dépense de santé d'un individu. Seules

des données très fines de diagnostic pourraient améliorer significativement la variance expliquée. Mais cette limite ne peut être opposée à la méthode : son but n'est pas de reconstruire l'existant individu par individu, mais de construire une base qui reproduit bien les caractéristiques des distributions observées dans la réalité.

Graphique III
A-Distribution des dépenses de santé en omnipraticiens, observées dans l'Epas et simulées par hot deck



B-Répartition des écarts entre dépenses observées et dépenses simulées (poste omnipraticiens)



Lecture : 74 % des écarts entre dépenses observées en omnipraticiens et dépenses simulées sont supérieurs, en valeur absolue, à 25 €. Champ : dépenses en omnipraticiens d'un sous-échantillon aléatoire de l'appariement Epas-ESPS. Source : Omar 2006.

Tableau 4
Comparaison des montants moyens observés dans l'Epas à ceux imputés par hot deck pour le poste « Omnipraticiens » sur non-réponse fictive

Poste de soins « Omnipraticiens »	Observé dans l'Epas	Imputé par hot deck stratifié
Taux de consommateurs (%)	83,4	82,2
Pour les consommateurs, montant (€) moyen des...		
... dépenses	145,3	144,8
... remboursements AMO	103,8	103,3
... tickets modérateurs	31,9	31,5
... dépassements	6,0	6,4
... participations forfaitaires	3,6	3,6

Lecture : pour un tiers des individus présents dans l'appariement Epas-ESPS, le taux de consommateurs en Omnipraticiens est 83,4 % et les dépenses moyennes de ces consommateurs s'élèvent à 145 €. Lorsque ces individus sont traités par hot deck le taux est plus faible, à 82,2 % et les dépenses moyennes sont quasiment identiques. Champ : échantillon de 3 135 individus sélectionné par tirage aléatoire dans l'appariement Epas-ESPS. Source : appariement Epas-SPS 2006, traitements de la Drees.

Attribuer un type de contrat complémentaire à chaque ménage qui en bénéficie

Une fois reconstituées l'ensemble des dépenses, le cœur du modèle consiste à simuler pour chaque individu les remboursements de son assurance complémentaire, lorsqu'il en a une. Ils seront calculés sur la base des dépenses constatées lesquelles, compte tenu des sources utilisées, correspondent aux seules dépenses présentées au remboursement¹². Mais le niveau de garantie offert par l'assurance complémentaire n'est connu directement que pour les bénéficiaires de la CMU-C. Pour les autres individus, l'enquête *SPS* ne renseigne que sur le type de contrat (individuel ou collectif) et la nature de l'organisme assureur (assurance, mutuelle, institut de prévoyance). C'est ici qu'intervient l'enquête de la Drees de 2008 sur les contrats modaux. On va attribuer à chaque individu de *SPS* ayant une couverture privée un contrat issu de l'enquête sur les contrats modaux. Plus exactement, cette attribution se fait dans chaque ménage pour le ou les souscripteurs du ou des contrats du ménage. Il est en effet important que les individus ayant déclaré dans *ESPS* être rattachés à un même contrat soient effectivement couverts par les mêmes garanties. *Omar* procède donc en deux temps, traitant d'abord le cas des ouvrants droit, et en déduisant ensuite les caractéristiques de l'AMC des ayants droit.

L'association d'un contrat modal à chaque ouvrant droit présent dans *Omar* se fait elle-même en deux étapes. La première consiste à déterminer la classe de qualité du contrat complémentaire et se fonde sur les caractéristiques individuelles connues de l'ouvrant droit : âge, type d'organisme auprès duquel le contrat est souscrit et caractère individuel ou collectif de ce contrat. Pour une combinaison X_i donnée de ces trois caractéristiques, on connaît les probabilités ($P_A(X_i)$, $P_B(X_i)$, $P_C(X_i)$, $P_D(X_i)$) de bénéficier d'une des quatre classes de couvertures identifiées plus haut (A, B, C, ou D) (cf. tableau 1). Par ce tirage aléatoire du niveau de garantie, la distribution globale de la qualité des contrats est respectée dans *Omar*.

Une fois déterminée la qualité du contrat, l'attribution d'un contrat modal à proprement parler peut être effectuée. Cette allocation se fait par tirage aléatoire d'un des contrats de l'enquête parmi ceux ayant la qualité retenue. Le choix se fait à nouveau en respectant les caractéristiques

associées à l'ouvrant droit, à savoir le fait que le contrat soit individuel ou collectif et le type d'organisme de couverture complémentaire. Ce tirage n'est pas à probabilités égales. En effet, les contrats de l'enquête Drees couvrent un nombre variable d'individus. Afin d'en tenir compte, la probabilité pour un contrat d'être tiré est proportionnelle au nombre de personnes ayant souscrit ce contrat. Ainsi, plus un contrat couvre un nombre important d'ouvrants droit dans l'enquête Drees, plus il a de chances d'être sélectionné.

À la fin de cette étape, un contrat de l'enquête Drees a ainsi été attribué à chaque ouvrant droit déclaré dans l'enquête *SPS* et par voie de conséquences, à ses ayants droit. Pour chacun de ces individus, on connaît donc les garanties pour un certain nombre de soins (spécialiste, optique, dentaire, etc.).

Un calcul des remboursements qui dépend du type de dépense à couvrir

La dernière étape de la simulation consiste à appliquer ces garanties aux dépenses de l'individu, dépenses constatées ou imputées, pour déterminer le montant simulé des remboursements au titre de la couverture complémentaire (ROC). La procédure diffère selon le type de dépense considérée.

Le cas des actes médicaux

Pour les actes médicaux¹³, les garanties des assurances complémentaires sont exprimées en pourcentage de la base de remboursement, si bien que le montant maximal remboursé par l'organisme complémentaire est obtenu par la formule suivante :

$$ROC^{max} = (RSS + TM) * \text{garantieOC} - RSS$$

où *RSS* désigne le montant remboursé par la Sécurité sociale, *TM* le montant de ticket modérateur et *garantieOC* le taux de remboursement de l'organisme complémentaire, exprimé en pourcentage de la base de remboursement.

Cependant, il est possible que les dépenses de l'assuré soient inférieures à ce remboursement

12. Mais incluant le cas échéant les dépassements d'honoraires.
13. On entend ici par « actes médicaux » les honoraires d'omnipraticiens, de spécialistes, de sages-femmes, les actes d'infirmiers, de kinésithérapeutes et d'autres auxiliaires médicaux ainsi que les examens de biologie et les transports.

maximum (par exemple, en l'absence de dépassements d'honoraires). De ce fait, le montant effectivement remboursé par l'organisme complémentaire est en réalité donné par la formule suivante :

$$ROC = \min(ROC^{max}, dépenses - RSS)$$

Soins et prothèses dentaires

En matière de soins dentaires, les données de l'*Epas* (2006) ne permettent pas de distinguer les dépenses imputables aux prothèses de celles dues aux soins conservateurs. Ces dépenses sont en effet regroupées dans une variable unique « dépense en dentaire ». C'est pourquoi le premier travail est de distinguer les individus qui ont eu une pose de prothèse dentaire des autres. Pour ce faire, nous considérons qu'un individu dont la base de remboursement annuelle est inférieure à celle d'une couronne (107,5 euros) a eu uniquement des soins conservateurs¹⁴.

Les individus dont la base de remboursement est supérieure à 107,5 euros et qui présentent des dépassements en dentaire se sont fait poser une prothèse. Pour ces individus, il faut alors distinguer dans la dépense agrégée en dentaire la part liée aux prothèses de celle liée aux soins conservateurs. Pour ce faire, nous faisons l'hypothèse qu'un individu ne peut s'être fait poser plus de trois prothèses dans l'année, les autres dépenses étant de fait liées à des soins conservateurs. Le nombre de prothèse est alors estimé de la manière suivante :

$$\text{nombre de prothèses dentaires} = \min \left(\left[\frac{\text{bases de remboursements en dentaire}}{107,5} \right], 3 \right)$$

La simulation du remboursement en dentaire repose sur la comparaison entre le maximum pris en charge par le contrat de l'individu et le reste à charge observé après remboursement de l'AMO dans la base de données. Le montant maximal pris en charge par la couverture complémentaire correspond au forfait « Prothèses » du contrat, appliqué aux prothèses décelées le cas échéant (trois au maximum), et aux tickets modérateurs des soins conservateurs. Si ce montant calculé est supérieur au reste à charge en dentaire après remboursements de l'AMO, alors le remboursement AMC est égal à ce seul reste à charge. Si le montant calculé est inférieur, alors il constitue le remboursement AMC retenu *in fine*.

Notons que, de manière à améliorer la qualité de la simulation du remboursement par l'orga-

nisme complémentaire sur le dentaire, depuis l'édition 2008 de l'outil *Omar*, la Drees accède à deux variables « Prothèses dentaires » d'une part et « Autres soins » d'autre part, et non une seule comme cela était le cas auparavant, ce qui permet de s'affranchir des hypothèses évoquées ci-dessus.

Le remboursement des dépenses d'optique

En matière d'optique, les données de l'*Epas* consistent également en une seule variable de dépense en optique agrégée, sans distinguer les différents types de prestations. Or, l'enquête Drees distingue les garanties couvrant les verres simples et les garanties s'appliquant aux verres complexes. Considérant qu'il est rare d'acheter plusieurs paires dans l'année, nous avons supposé que le montant de la dépense observé relevait d'un seul achat de lunettes. Ensuite, la confrontation de la base de remboursement observée avec celle de verres complexes standards nous a permis d'inférer si l'individu avait acheté des verres simples ou des verres complexes. La dernière opération a consisté à appliquer la garantie adéquate du contrat attribué à l'individu lors de la précédente étape.

Dépenses en pharmacie ou en hospitalisation

Les autres dépenses sont également agrégées par poste dans les données de l'*Epas*. De fait, cela ne nous permet pas de ventiler la dépense de pharmacie entre ce qui relève de médicaments pris en charge à 65 % ou à 35 %. Concernant les dépenses hospitalières, nous ne sommes pas non plus capables de distinguer ce qui relève du ticket modérateur à proprement parler de ce qui relève du forfait journalier dans le montant non remboursé par la Sécurité sociale : dans nos données, la variable « Ticket modérateur » englobe les deux montants.

Cependant, les contrats complémentaires ne prenant pas en charge les médicaments remboursés à 35 % par l'assurance maladie obligatoire ou le ticket modérateur à l'hôpital sont minoritaires. Aussi, sur ces deux postes de dépenses, faute d'avoir des informations détaillées, nous avons

14. Pour ces individus consommant uniquement des soins conservateurs, nous adoptons comme conventions que leur contrat rembourse la totalité du ticket modérateur mais non les éventuels dépassements. Cette convention est indispensable dans la mesure où l'enquête sur les contrats les plus souscrits ne contient pas cette information.

considéré que les contrats de couverture complémentaire remboursaient systématiquement le montant du ticket modérateur.

Là encore, à compter de l'édition 2008 de l'outil *Omar*, la Drees dispose d'informations plus détaillées sur les dépenses pharmaceutiques (notamment le taux de remboursement à 15 %, 35 % ou 65 %), ce qui permet d'appliquer directement les garanties du contrat d'assurance complémentaire.

Des résultats finaux peu affectés par le caractère aléatoire des imputations

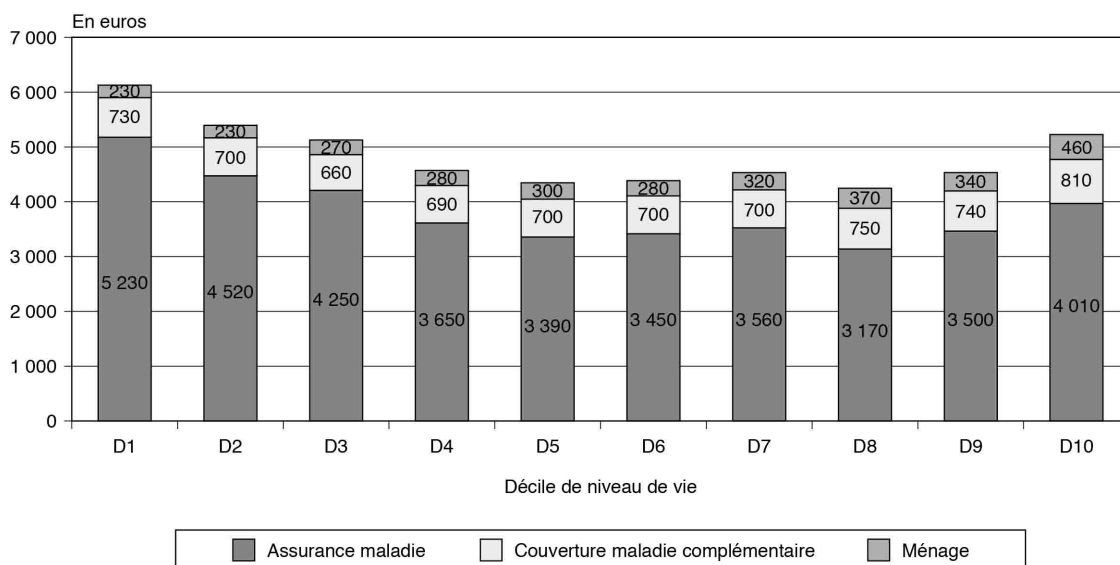
A titre d'illustration des résultats du modèle, on peut examiner la structure du financement des dépenses par déciles de revenu du ménage (cf. graphique IV). On constate que les dépenses présentent un profil en « J » inversé, avec des restes à charge légèrement progressifs, passant de 230 euros par an pour les ménages du premier décile à 460 euros par an pour les ménages du dernier décile. Le profil de la prise en charge par l'assurance complémentaire suit un profil en « J » non inversé : il part de 730 euros par an, atteint un minimum de

660 euros pour le troisième décile, et remonte à environ 810 euros pour le décile le plus élevé.

Mais un reproche couramment adressé aux méthodes de microsimulation ou d'imputation aléatoire est le fait que les données obtenues sont affectées d'une variabilité stochastique. On peut donc se demander si les résultats obtenus sont robustes, ou s'ils sont tributaires des tirages aléatoires effectués pour imputer les données manquantes. Ces profils semblent certes assez réguliers pour ne pas traduire qu'un bruit aléatoire, mais il est intéressant d'essayer de le vérifier. Pour ce faire, il suffit de générer un grand nombre de distributions générées par le même modèle et d'étudier leur variance. On va plus spécialement examiner le montant des remboursements versés par l'AMC. Celui-ci est en effet directement dépendant des spécificités du contrat imputé.

Une procédure *bootstrap* a été mise en place permettant de produire 150 jeux de données microsimulées. Entre chaque itération, les dépenses de santé associées aux individus restent inchangées : seules les étapes d'attribution d'un niveau de garantie et, par la suite, d'un contrat modal sont renouvelées. Dès lors, la seule source de variation des résultats relève de l'imputation de la couverture maladie complémentaire.

Graphique IV
Structure de financement des dépenses de santé des ménages selon le décile de niveau de vie



Lecture : les dépenses de santé d'un ménage du 1^{er} décile de niveau de vie s'élèvent en moyenne à 6 190 € ; sur ce montant, 5 230 € sont pris en charge, en moyenne, par l'Assurance maladie, 730 € par la couverture maladie complémentaire et 230 € restent à la charge du ménage.

Champ : ménages ordinaires ; dépenses présentées au remboursement de l'Assurance maladie.

Source : Omar 2006.

Avec les 150 versions de données générées par *Omar*, il est possible de calculer, à l'aide d'une procédure *bootstrap*, des moyennes et des intervalles de confiance à 95 %. Les remboursements d'AMC simulés et bootstrapés sont faiblement dispersés (cf. graphique V). L'écart entre la borne supérieure et la borne inférieure de l'intervalle de confiance est, pour l'ensemble des déciles de niveau de vie, comprise entre 12 euros (1^{er} décile de niveau de vie) et 27 euros (9^{ème} décile de niveau de vie). Rapportée au remboursement moyen, cette incertitude s'élève au maximum à 3,6 % de la moyenne *bootstrap* estimée¹⁵.

Même si les montants calculés via *Omar* ne doivent pas être étudiés à la dizaine près, ces intervalles de confiance montrent que l'on peut attribuer une certaine fiabilité à des phénomènes tels que les remboursements d'assurance maladie complémentaire plus élevés des ménages appartenant aux trois déciles de niveaux de vie les plus favorisés.

Avec cette confirmation de la faible volatilité des montants simulés, c'est finalement l'objectif

de départ qui est atteint. Il est désormais possible de simuler la structure de financement des dépenses de santé à un niveau plus fin que celui disponible dans les Comptes de la santé.

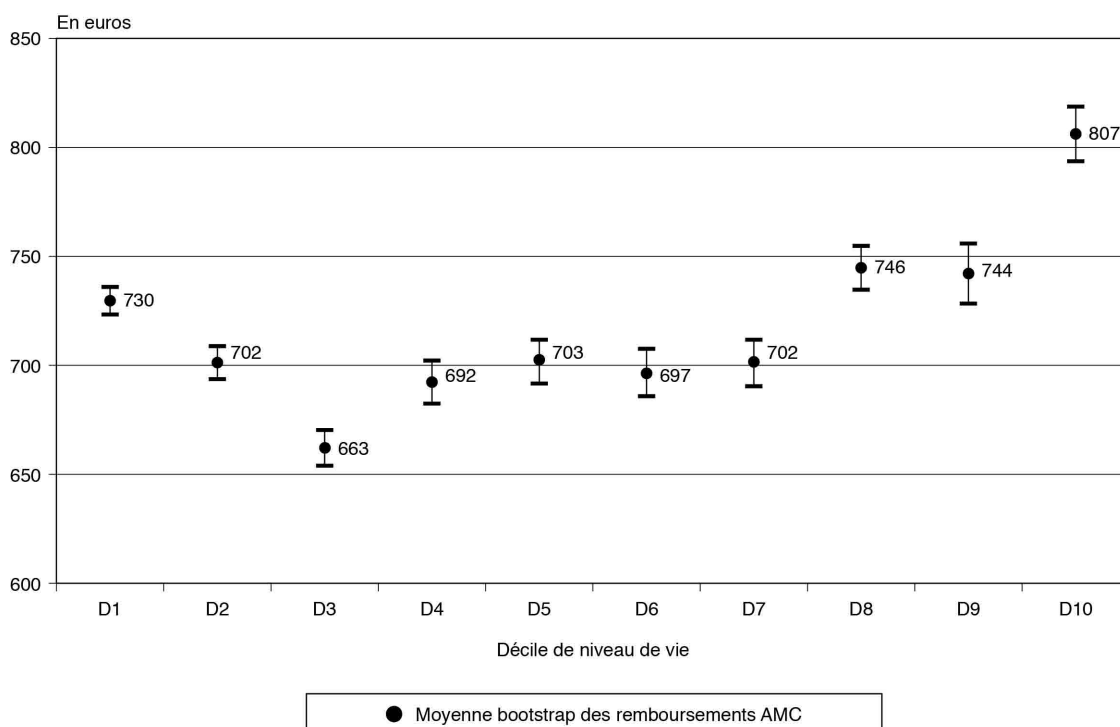
* *
*

Les résultats présentés dans cette étude montrent les potentialités offertes par *Omar* pour mieux évaluer le poids des dépenses de santé dans le budget des ménages.

À l'avenir, la Drees prévoit d'actualiser et d'améliorer cet outil tous les deux ans, suivant la mise à disposition des données de l'appariement *Epas-ESPS*. De plus, une réflexion est en cours afin d'améliorer l'imputation des contrats d'assurance complémentaire santé : il s'agirait de tenir compte de l'appréciation subjective de l'assuré sur les niveaux de garanties offerts par

15. Ce maximum est atteint pour le remboursement complémentaire des ménages appartenant au 9^e décile de niveau de vie (moyenne bootstrap de 744 euros par ans).

Graphique V
Remboursements de l'Assurance maladie complémentaire (niveau ménage) : moyennes et intervalles de confiance à 95 % calculés à l'aide d'une procédure *bootstrap* selon le décile de niveau de vie



Lecture : moyennes et intervalles de confiance à 95 % des remboursements de l'Assurance maladie complémentaire (AMC), au niveau des ménages, calculés à l'aide de la procédure *bootstrap*, selon le décile de niveau de vie.
Champ : ménages ordinaires, France métropolitaine ; dépenses présentées au remboursement.
Source : Omar 2006, Drees.

le contrat qu'il détient. Par ailleurs, conjointement au renouvellement de l'enquête sur les contrats les plus souscrits, il est prévu d'étudier les changements résultant d'une modification de la classification des contrats suivant leur niveau de garantie. En effet, des travaux méthodologiques sur la typologie des contrats d'assurance maladie complémentaire vont être reconduits par la Drees à partir des données 2010 de l'enquête sur les contrats. Ces travaux permettront de tester la sensibilité des résultats de l'actualisation du modèle *Omar* aux différentes typologies essayées.

Enfin, la Drees a intégré ce modèle de microsimulation dans le modèle *Ines*¹⁶. Cette intégration permet de conduire des analyses redistributives globales puisqu'*Ines* contient des informations sur les cotisations individuelles versées à la Sécurité sociale. Ainsi, l'intégration de ces deux modèles permet de mettre en regard, au niveau individuel, le volet prestations et le volet cotisations et primes à la fois pour les assurances publique et privée. Dès lors, la réunion de ces deux outils permet de compléter et d'affiner les travaux macroéconomiques et microéconomiques existant sur l'analyse de la redistribution opérée par l'Assurance maladie (Caussat *et al.*, 2005) et d'estimer l'impact de mesures portant sur les règles de remboursement par les différents organismes de couverture obligatoire ou facultative. Deux études, réalisées grâce à l'outil *Ines-Omar* et portant sur les transferts entre tranches d'âge et sur les transferts entre déciles de niveaux de vie ont été

publiées dans les ouvrages des Comptes de la santé 2010 et 2011 (Duval *et al.*, 2011, Duval et Lardellier, 2012).

Finalement, le fait qu'*Omar* soit un outil encore naissant et amené à évoluer dans les prochaines années empêche pour le moment d'en faire un outil pour mener des analyses longitudinales. L'intérêt d'*Omar* réside d'une part, dans sa capacité à fournir des premiers chiffrages de l'intervention des organismes d'AMC et, d'autre part, à rendre possible des comparaisons transversales, entre déciles de niveau de vie pour une année donnée par exemple. Ces travaux de microsimulation ne sont qu'un substitut imparfait à l'absence d'observation de données réelles qui seraient directement issues d'un appariement des données de l'assurance maladie obligatoire et des assurances complémentaires. Des expérimentations en ce sens sont en cours, comme le projet Monaco mis en œuvre par l'Irdes avec la collaboration de la Cnamts et des organismes complémentaires volontaires sous l'égide de l'Institut des données de santé. Il est souhaitable que ce type de rapprochement de données se développe, comme le recommande le Conseil national de l'information statistique (Cnis). □

16. Il s'agit d'un outil d'analyse et d'évaluation des effets redistributifs des changements introduits dans le système de prélèvement et de transferts. Le projet de réalisation de ce modèle a vu le jour en 1995, sous la conduite de l'Insee. Depuis 1995, le modèle a beaucoup évolué, notamment en 1998, date à laquelle il a atteint un stade opérationnel de large diffusion et a pu être utilisé dans plusieurs analyses et simulations couvrant l'ensemble du système redistributif français. Actuellement, le modèle *Ines* est géré par la Drees et l'Insee.

BIBLIOGRAPHIE

Allonier C., Dourgnon P. et Rochereau T (2008), « Enquête sur la Santé et la Protection Sociale », *Série résultats*, Irdes.

Arnould M-L et Vidal G. (2008), « Typologie des contrats les plus souscrits auprès des complémentaires santé en 2006 », *Études et résultats*, n° 663, Drees.

Beeuwkes Buntin M. et Zaslavsky A.M. (2003), « Too much about two-part models and transformation ? Comparing methods of modelling Medicare expenditures » *Journal of Health Economics*, n° 23, pp. 525-542.

Bocognano A., Couffinal A., Dusmenil S. et Grignon M. (2000), « La complémentaire maladie en France : qui bénéficie de quels remboursements », *Questions d'économie de la santé*, n° 32, Credes.

Boisguérin B. (2004), « état de santé et recours aux soins des bénéficiaires de la CMU-un impact qui se consolide entre 2000 et 2003 », *Études et résultats*, n° 294, Drees.

Boisguérin B. (2007), « Les allocataires des minima sociaux : CMU, état de santé et recours aux soins », *Études et résultats*, n° 603, Drees.

- Boisguérin B. (2009)**, « Quelles caractéristiques et quel recours aux soins pour les bénéficiaires de la CMUC ? », *Études et résultats*, n° 675, Drees.
- Boisguérin B., Després C., Dourgnon P., Fantin R. et Legal R. (2010)**, « étudier l'accès aux soins des assurés CMU-C, une approche par le renoncement aux soins », dans *Enquête sur la santé et la protection sociale 2010*, Irdes.
- Breuil-Genier P., Grandfils N. et Raynaud D. (1999)**, « Revenus, assurance et santé : le problème de l'accès aux soins des plus démunis », *Les cahiers du Gratice*, n° 15, pp. 243-76.
- Caussat L. et Glaude M. (1993)**, « Dépenses médicales et couverture sociale », *Économie et Statistique*, n° 265, pp. 31-43.
- Caussat L., Le Minez S. et Raynaud D. (2005)**, « L'assurance-maladie contribue-t-elle à redistribuer les revenus ? », *Dossiers Solidarité et Santé*, n° 1.
- Couffinhal A. et Perronnin M. (2004)**, « Accès à la couverture complémentaire maladie en France : une comparaison des niveaux de remboursement », *Questions d'économie de la santé*, n° 47, Credes.
- Couffinhal A., Grandfils N., Grignon M. et Rochereau T. (2004)**, « La complémentaire maladie d'entreprise. Premiers résultats nationaux d'une enquête menée fin 2003 auprès de 1700 établissements », *Questions d'économie de la santé*, n° 83, Irdes.
- Currie J. (2000)**, « Child Health in Developed Countries », dans A. Cutler et J.P. Newhouse (eds.), *Handbook of Health Economics*, pp. 1054-1090.
- Dormont B., Hubert H. et Grignon M. (2006)**, « Health expenditures growth : reassessing the threat of ageing », *Health economics*, vol. 15, n° 9.
- Dourgnon P., Grignon M. et Jusot F. (2001)**, « L'assurance maladie réduit-elle les inégalités sociales de santé ? Une revue de littérature », *Questions d'économie de la santé*, n° 43, Irdes.
- Duval J., Lardellier R. et Legal R. (2011)**, « La redistribution opérée par l'assurance maladie obligatoire et par les assurances complémentaires selon l'âge », dans *Les comptes nationaux de la santé 2010*, Document de travail Drees,-série statistiques n° 161.
- Duval J. et Lardellier R. (2012)**, « La redistribution verticale opérée par l'assurance maladie », dans *Les comptes nationaux de la santé 2011*, Document de travail Drees,-série statistiques n° 172.
- Fonds de Financement de la Couverture Maladie Universelle (2010)**, *Références*, n° 39.
- Francesconi C., Perronnin M. et Rochereau T. (2006)**, « La complémentaire maladie d'entreprise : niveaux de garanties des contrats selon les salariés et le secteur d'activité », *Questions d'économie de la santé*, n° 112, Irdes.
- Franc, C. et Perronnin M. (2007)**, « Aide à l'acquisition d'une assurance maladie complémentaire : une première évaluation du dispositif ACS », *Questions d'économie de la santé*, n° 121, Irdes.
- Francesconi C., Perronnin M. et Rochereau T. (2006)**, « Complémentaire maladie d'entreprise : contrats obligatoires ou facultatifs, lutte contre l'antiselection et conséquences pour les salariés », *Questions d'économie de la santé*, n° 115, Irdes.
- Garnero M. et Rattier M.O. (2011)**, « Les contrats les plus souscrits auprès des complémentaires santé en 2008 », *Études et résultats*, n° 52, Drees.
- Geoffard P.Y. (2000)**, « Assurance maladie : la gestion du risque long », *Revue d'économie politique*, vol. 110, n° 4, pp. 457-482.
- Grignon M. et Kambia-Chopin B. (2010)**, « Quelle subvention optimale pour l'achat d'une complémentaire santé ? », *Questions d'économie de la santé*, n° 153, Irdes.
- Guillaume S. et Rochereau T. (2010)**, « La complémentaire santé collective : une offre aux salariés très inégale selon les entreprises », *Questions d'économie de la santé*, n° 154, Irdes.
- Heckman J.J. (1979)**, « Sample selection bias as a specification error », *Études et résultats*, n° 663, Drees *Econometrica*, vol. 47, pp. 153-161.
- Kambia-Chopin B., Perronnin M., Pierre A. et Rochereau T. (2008)**, « La complémentaire santé en France en 2006 : un accès qui reste inégalitaire », *Questions d'économie de la santé*, n° 132, Irdes.
- Lengagne P. et Perronnin M. (2005)**, « Impact des niveaux de garanties des complémentaires santé sur les consommations de soins peu remboursées par l'Assurance maladie : le cas des lunettes et des

prothèses dentaires », *Questions d'économie de la santé*, n° 100, Irdes.

Leung, S.F. et Yu S. (1996), « One the choice between sample selection and two-part models », *Journal of Econometrics*, vol. 72, n° 1-2, pp. 197-229.

Lollivier S. (1997), « Modèles univariés et modèles de durée sur données individuelles », Série des documents de travail, Méthodologie Statistique, Insee.

Manning, W., Duan, N. et Rogers W. (1987), « Monte carlo evidence on the choice between sample selection and two-part models », *Journal of econometrics*, vol. 35, pp. 59-82.

Newhouse J.-P. (1993), *Free for All ? Lessons from the RAND experiment*, Harvard University Press.

Pichetti, S., Raynaud D. et Vidal G. (2009), « Les déterminants individuels des dépenses de santé » dans *Traité d'économie et de gestion de la santé*, presses de Sciences Po, ouvrage coordonné par P.L. Bras, G. de Pourville, D. Tabuteau.

Raynaud D. (2002), « Les déterminants individuels des dépenses de santé », *Études et résultats*, n° 182, Drees.

Raynaud D. (2005), « Les déterminants individuels des dépenses de santé : l'influence de la catégorie sociale et de l'assurance maladie complémentaire », *Études et résultats*, n° 378, Drees.

Sautory, O. (1993), *La macro calmar : redressement d'un échantillon par calage sur marges*, document de travail n° F9310, Insee.

L'IMPUTATION DES REVENUS

L'étape d'imputation des revenus vise à reconstituer les revenus des ménages qui ont été déclarés en tranches ou ceux qui sont totalement inconnus. Avec ou sans pondération, ces deux groupes représentent respectivement un dixième et un quart de l'échantillon total (cf. tableau A).

Tableau A
Proportion et nombre de ménages ayant répondu à la question sur le revenu en montant ou en tranche ou n'ayant pas répondu à la question sur les revenus

Type de réponse	Données avec pondération		Données sans pondération	
	Proportion (%)	Nombre de ménages	Proportion (%)	Nombre de ménages
Revenu en montant	67	5 430	66	5 335
Revenu par tranche	9	731	10	802
Revenu inconnu	24	1 968	24	1 992

Champ : ménages ordinaires, France métropolitaine.
Source : appariement EPAS-SPS 2006, Irdes, Cnamts, RSI, MSA, traitement Drees.

Des sources de revenus très hétérogènes

Suspectant des sources de revenus assez hétérogènes selon le statut vis-à-vis de l'emploi des membres du ménage, nous avons distingué *a priori* quatre types de ménages : les ménages dont les adultes sont uniquement des actifs occupés, ceux dont les adultes sont uniquement des inactifs ou des chômeurs, ceux dont les adultes sont uniquement des retraités et les autres (cf. tableau B). Sont considérés comme *actifs occupés* les individus qui occupent un travail au moment de l'enquête, qui sont en congé maternité ou qui sont en congé longue maladie. Sont considérées comme *retraités* les personnes qui se déclarent à la retraite ou retirées des affaires ou celles qui ont plus de 65 ans et qui n'exercent plus d'activité professionnelle. Ce dernier critère permet de classer dans cette catégorie les personnes qui ont déclaré une PCS mais qui n'exercent plus leur métier et les veuves et les veufs de plus de 65 ans qui perçoivent une pension de réversion. Sont rassemblés dans la catégorie *inactifs ou chômeurs* les adultes qui ne sont ni classés comme actifs occupés ni comme retraités. Cette classe comprend donc les individus scolarisés de plus de 16 ans, les enfants, les demandeurs d'emploi, les hommes et les femmes au foyer de moins de 65 ans, les inactifs avec pension d'invalidité, etc.

Tableau B
Part des différentes sources de revenus selon le type de ménage

En %

	Actifs occupés	Retraités	Inactifs ou chômeurs	Autres
Part dans l'ensemble des ménages (%)	45.8	19.8	8.5	25.9
Revenus d'activité	91.6	1.3	5.7	67.1
Pensions de retraite	0.4	90.5	4.0	14
Revenus de transfert	6.1	4.3	84.3	16.2
Autres	1.9	3.9	6.0	2.7

Champ : ménages ordinaires, France métropolitaine.
Source : appariement EPAS-SPS 2006, Irdes, Cnamts, RSI, MSA, traitement Drees.

L'estimation de quatre équations de revenu

Compte tenu de la distribution log-normale des revenus, c'est le logarithme du revenu qui est modélisé, et non le revenu lui-même. Par ailleurs, il est possible que l'équation de revenu souffre d'un biais de sélection. C'est le cas si certaines variables inobservables jouent à la fois sur la décision de répondre à la question du revenu et sur le montant du revenu déclaré. Pour tester l'existence d'un tel phénomène et le corriger le cas échéant, il convient donc d'estimer l'équation d'imputation de revenu au moyen d'un modèle Tobit généralisé en faisant l'hypothèse d'une distribution jointe des résidus (en l'occurrence, via une loi normale bivariée) :

$$\begin{cases} P_i^* = X_{2i}'\beta_2 + u_{i2} & (\text{équation de réponse}) \\ \log(R_i) = (X_{1i}'\beta_1 + u_{i1}) P_i & (\text{équation de revenu}) \end{cases}$$

$$\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} \rightarrow N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$$

Où P_i^* est la probabilité de déclarer ses revenus, R_i le niveau des revenus, X_1 et X_2 les vecteurs de variables explicatives (âge, sexe, PCS etc.) et u_1 et u_2 des résidus aléatoires suivant une loi normale $(0, \sigma^2)$. L'existence d'un biais de sélection se matérialise par un coefficient de corrélation des résidus ρ significativement non nul.

L'estimation des quatre modèles Tobit généralisés est obtenue par la procédure proc qlim du logiciel SAS (méthode du maximum de vraisemblance). Les résultats montrent qu'il n'existe pas de biais de sélection (ρ non significativement différent de 0) (cf. tableau C). En fin de compte, il est donc possible d'estimer directement l'équation de revenu pour obtenir une estimation non biaisée de β_1 .

Tableau C
Coefficients de corrélation des résidus
des quatre modèles Tobit généralisés de revenu

Catégorie de ménages	ρ estimé	p-value
Actifs	- 0,093	0,4066
Retraités	- 0,0303	0,9467
Inactifs	0,0065	0,98
Mixtes	- 0,0798	0,7356

Source : auteurs.

Environ 10 % des individus ont déclaré leur revenu en tranches. En notant E_1 (resp. E_2) le sous-échantillon des individus qui ont fourni leur revenu en euros (resp. en tranches), et R1 et R2 les bornes respectivement inférieure et supérieure des tranches déclarées, nous estimons β_1 et σ_1 par maximisation de la fonction de vraisemblance suivante au moyen de la procédure proc lifereg sous SAS :

$$L = \prod_{i \in E_1} \varphi \left(\frac{\log(R_i) - X_i \beta_1}{\sigma_1} \right) \prod_{i \in E_2} \left(\Phi \left(\frac{\log(R_{2i}) - X_i \beta_1}{\sigma_1} \right) - \Phi \left(\frac{\log(R_{1i}) - X_i \beta_1}{\sigma_1} \right) \right)$$

De cette manière, nous prenons en compte l'information des individus qui ont déclaré leurs revenus en euros mais aussi en tranche (Lollivier, 1997).

Imputation des revenus manquants par la méthode des résidus simulés

Les revenus des non-répondants (non-réponse totale et non-réponse partielle, i.e. en tranche) sont alors imputés

par la méthode des résidus simulés. Pour cela, on simule le logarithme du revenu d'un individu de caractéristiques observables X_i qui n'aurait pas déclaré son revenu à l'aide de l'espérance conditionnelle du revenu à laquelle on ajoute une perturbation :

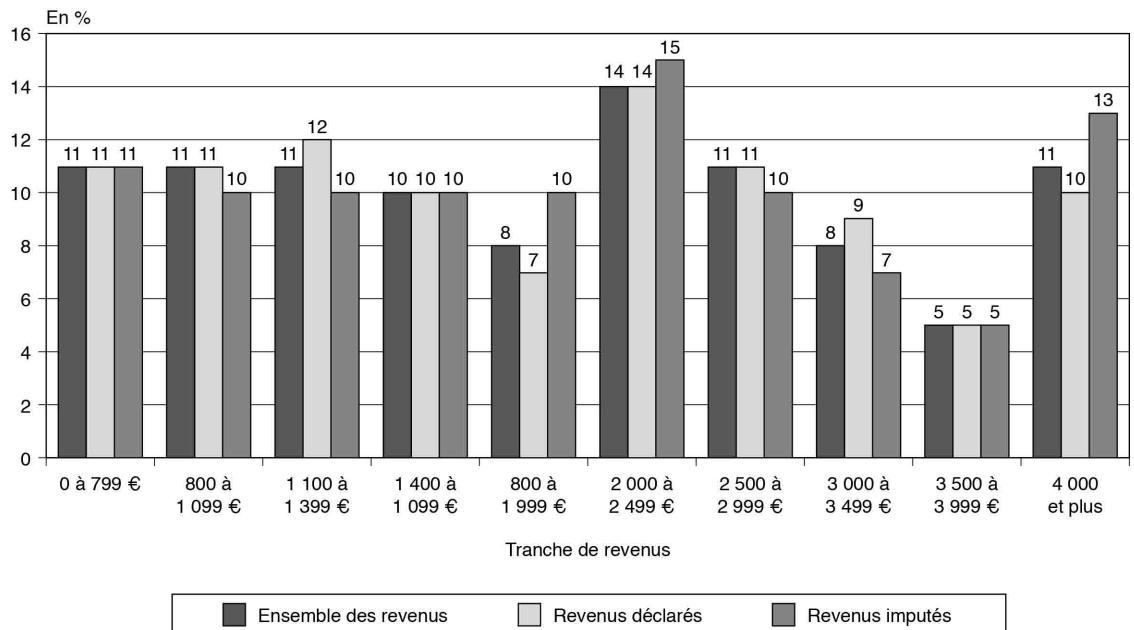
$$(\log(R_i))^{sim} = E(\log(R_i) | X_i) + u_i^{sim} = X_i \hat{\beta}_1 + u_i^{sim}$$

Si bien que $R_i^{sim} = \exp(X_i \hat{\beta}_1 + u_i^{sim})$.

Le résidu simulé u_i^{sim} est tiré dans une loi normale de variance $\hat{\sigma}_1^2$. Pour les individus qui ont déclaré leur revenu en tranches, tant que le revenu simulé n'est pas dans la tranche, un nouveau résidu est tiré au sort. Pour les autres, on contraint le revenu simulé à être supérieur au minimum des revenus déclarés et inférieur au maximum des revenus déclarés (pour la catégorie de ménages correspondants). Ce dernier choix suppose implicitement que les revenus minimum et maximum par catégorie de ménages sont bien observés. Cette hypothèse peut paraître forte, mais elle permet d'éviter d'introduire aléatoirement des revenus qui seraient atypiques, et par ailleurs, elle n'a pas d'impact sur les résultats car les revenus sont ensuite traités en tranche (déciles).

Par construction, les distributions des revenus avant et après imputations sont proches l'une de l'autre (cf. graphique).

Graphique
Distribution des revenus avant et après imputation



Champ : ensemble des ménages.
 Source : Omar 2006.

RÉSULTATS DES ESTIMATIONS DES TROIS MODÈLES TOBIT DE DÉPENSE

Tableau A
Modélisation de la présence ou non d'une consommation hors optique et dentaire

Variable	Coefficient estimé	Écart-type
Constante	1,73 ***	0,11
Nature de la couverture complémentaire		
Aucune	- 0,48 ***	0,12
Privée	Réf.	Réf.
CMU	0,25 *	0,13
Régime		
MSA	0,33 **	0,16
RSI	- 0,19	0,14
Cnamts et autres	Réf.	Réf.
Exonération		
Non et NSP	Réf.	Réf.
Oui	0,99 ***	0,17
Consultation d'un omnipraticien		
0	- 1,09 ***	0,1
1 ou 2	Réf.	Réf.
3 ou 4	0,37 ***	0,13
> 5	0,54 ***	0,15
NSP	- 0,64 ***	0,15
Consultation d'un spécialiste		
0	Réf.	Réf.
> 1	0,64 ***	0,1
NSP	0,31 **	0,12
Consommation de médicaments		
Oui	0,29 **	0,11
Non	Réf.	Réf.
Inconnu	0,14	0,14
État de santé perçu		
Dégradé	0,08	0,14
Bon à très bon	Réf.	Réf.
Inconnu	0,02	0,18
Sexe et âge		
0 à 19 ans	0,90 ***	0,11
40 à 59 ans	0,33 ***	0,1
60 à 69 ans	0,48 ***	0,17
> 70 ans	0,79 ***	0,2
Homme de 20 à 39 ans	Réf.	Réf.
Femme de 20 à 39 ans	0,55 ***	0,13
Niveau de revenus		
Élevé	- 0,01 *	0,09
Moyen	Réf.	Réf.
Faible	- 0,28 ***	0,09
Nombre de personnes dans le ménage		
1 à 3	Réf.	Réf.
> 4	- 0,13 *	0,08

Lecture : * : significatif au seuil de 10 % ; ** : Significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

Le pourcentage de paires concordantes est de 73,3.

Champ : individus vivant au sein d'un ménage ordinaire de France métropolitaine.

Source : appariement Epas-ESPS.

Tableau B
Modélisation du montant de la consommation hors optique et dentaire

Variable	Coefficient estimé	Écart-type
Constante	7,15***	0,1
Nature de la couverture complémentaire		
Aucune	- 0,57***	0,06
Privée	- 0,28***	0,04
CMU	Réf.	Réf.
Régime		
MSA	- 0,57***	0,07
RSI	Réf.	Réf.
Cnamts et autres	- 0,40***	0,06
Exonération		
Non et NSP	- 0,91***	0,04
Oui	Réf.	Réf.
Consultation d'un omnipraticien		
0	- 0,45***	0,07
1 ou 2	- 0,22***	0,06
3 ou 4	0,22***	0,06
> 5	0,56***	0,06
NSP	Réf.	Réf.
Consultation d'un spécialiste		
0	- 0,44***	0,04
> 1	0,09**	0,04
NSP	Réf.	Réf.
Consommation de médicaments		
Oui	0,17***	0,05
Non	- 0,22***	0,05
Inconnu	Réf.	Réf.
État de santé perçu		
Dégradé	0,35***	0,07
Bon à très bon	0,01	0,07
Inconnu	Réf.	Réf.
Sexe et âge		
0 à 19 ans	0,11**	0,05
40 à 59 ans	0,54***	0,05
60 à 69 ans	0,86***	0,06
> 70 ans	1,10***	0,06
Homme de 20 à 39 ans	Réf.	Réf.
Femme de 20 à 39 ans	0,55***	0,05
Niveau de revenus		
Élevé	Réf.	Réf.
Moyen	- 0,01	0,03
Faible	- 0,01	0,04
Nombre de personnes dans le ménage		
1 à 3	0,15***	0,03
> 4	Réf.	Réf.

Lecture : * : significatif au seuil de 10 % ; ** : Significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

Le R² du modèle vaut 0,4074.

Champ : individus vivant au sein d'un ménage ordinaire de France métropolitaine.

Source : appariement Epas-ESPS.

Tableau C
Modélisation de la présence ou non d'une consommation en optique

Variable	Coefficient estimé	Écart-type
Constante	- 0,20***	0,13
Nature de la couverture complémentaire		
Aucune	- 0,54***	0,13
Privée	Réf.	Réf.
CMU	- 0,14	0,11
Spécialiste		
Ophtalmologue	1,12***	0,07
Autres et non	Réf.	Réf.
NSP	0,58***	0,11
Lunettes		
Oui	Réf.	Réf.
Non	- 1,70***	0,12
Inconnu	- 1,58***	0,16
Indicateur synthétique de problèmes de vue		
Loin et près	0,45***	0,1
Loin ou près	0,36***	0,09
Aucun	Réf.	Réf.
Inconnu	0,96***	0,14
Maladie vue		
Non	- 0,32***	0,11
Oui	Réf.	Réf.
NSP	0,45**	0,18
Classes d'âge		
0 à 15 ans	0,14	0,1
16 à 44 ans	Réf.	Réf.
Homme de 45 à 64 ans	0,21**	0,09
Femme de 45 à 64 ans	0,29***	0,09
> 65	- 0,28***	0,1
Nombre de personnes dans le ménage		
1 à 3	Réf.	Réf.
> 4	0,16**	0,07
Niveau de revenus		
Élevé	0,03	0,07
Moyen	Réf.	Réf.
Faible	- 0,18 **	0,07
Régime		
MSA	0,01	0,12
Cnamts et autres	Réf.	Réf.
RSI	- 0,02	0,12
Niveau d'études		
1	Réf.	Réf.
2	- 0,08	0,07

Lecture : * : significatif au seuil de 10 % ; ** : Significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.
 Le pourcentage de paires concordantes est de 76,1 %.
 Champ : individus vivant au sein d'un ménage ordinaire de France métropolitaine.
 Source : appariement Epas-ESPS.

Tableau D
Modélisation du montant de la consommation en optique

Variable	Coefficient estimé	Écart-type
Constante	5,21***	0,12
Nature de la couverture complémentaire		
Aucune	0,59***	0,10
Privée	1,07***	0,07
CMU	Réf.	Réf.
Spécialiste		
Autres et non	- 0,01	0,08
NSP	Réf.	Réf.
Ophtalmologue	0	0,07
Lunettes		
Oui	0,25**	0,12
Inconnu	Réf.	Réf.
Non	- 0,01	0,12
Indicateur synthétique de problèmes de vue		
Loin et près	0,11	0,09
Loin ou près	- 0,06	0,09
Inconnu	Réf.	Réf.
Aucun	- 0,05	0,09
Maladie vue		
Oui	- 0,33**	0,13
Non	- 0,24**	0,12
NSP	Réf.	Réf.
Classes d'âge		
0 à 15 ans	- 0,36***	0,07
16 à 44 ans	- 0,25***	0,06
Homme de 45 à 64 ans	0,05	0,06
Femme de 45 à 64 ans	0,16**	0,06
> 65	Réf.	Réf.
Niveau de revenus		
Élevé	Réf.	Réf.
Faible	- 0,22***	0,05
Moyen	- 0,11**	0,04
Nombre de personnes dans le ménage		
1 à 3	- 0,03	0,04
> 4	Réf.	Réf.
Régime		
Cnamts et autres	- 0,34***	0,07
MSA	- 0,41***	0,10
RSI	Réf.	Réf.
Niveau d'études		
1	0,15***	0,04
2	Réf.	Réf.

Lecture : * : significatif au seuil de 10 % ; ** : Significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.
 Le R² du modèle vaut 0,2984.

Champ : individus vivant au sein d'un ménage ordinaire de France métropolitaine.

Source : appariement Epas-ESPS.

Tableau E
Modélisation de la présence ou non d'une consommation en dentaire

Variable	Coefficient estimé	Écart-type
Constante	- 0,97 ***	0,09
Nature de la couverture complémentaire		
Aucune	- 0,41 ***	0,09
Privée	Réf.	Réf.
CMU	0,05	0,08
Classes d'âge		
0 à 5 ans	- 1,75 ***	0,17
6 à 15 ans	0,67 ***	0,10
Femmes de 16 à 29 ans	0,49 ***	0,11
Hommes de 16 à 29 ans	Réf.	Réf.
Femmes de 30 à 49 ans	0,50 ***	0,10
Hommes de 30 à 49 ans	0,19 **	0,10
Femmes de 50 à 64 ans	0,52 ***	0,11
Hommes de 50 à 64 ans	0,13	0,11
Femmes de plus de 65 ans	0,09	0,11
Niveau de revenus		
Élevé	0,24 ***	0,05
Moyen	Réf.	Réf.
Faible	- 0,1576 ***	0,05
Régime		
MSA	0,11	0,09
Cnamts et autres	Réf.	Réf.
RSI	0,07	0,09
Nombre de personnes dans le ménage		
1 à 3	Réf.	Réf.
> 4	0,05	0,05
État de la dentition		
Inconnu	0,28 ***	0,09
Mauvais	- 0,29 ***	0,07
Bon	Réf.	Réf.
Très bon	- 0,11	0,07
Problèmes de dents		
Oui	0,74 ***	0,06
Non	Réf.	Réf.
Inconnu	- 0,16 **	0,08
Prothèse		
Dentier	- 0,12	0,09
Fixe	0,57 ***	0,06
Inconnu	0,01	0,08
Non	Réf.	Réf.
Redressement	1,64 ***	0,21

Lecture : * : significatif au seuil de 10 % ; ** : Significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

Le pourcentage de paires concordantes est de 67,4 %.

Champ : individus vivant au sein d'un ménage ordinaire de France métropolitaine.

Source : appariement Epas-ESPS.

Tableau F
Modélisation du montant de la consommation en dentaire

Modalité	Coefficient estimé	Écart-type
Constante	5,61***	0,16
Nature de la couverture complémentaire		
Aucune	- 0,21*	0,11
Privée	- 0,14*	0,08
CMU	Réf.	Réf.
Classes d'âge		
0 à 5 ans	- 1,31***	0,21
6 à 15 ans	- 0,38***	0,10
Femmes de 16 à 29 ans	- 0,35***	0,10
Hommes de 16 à 29 ans	- 0,18	0,11
Femmes de 30 à 49 ans	- 0,10	0,08
Hommes de 30 à 49 ans	- 0,06	0,09
Femmes de 50 à 64 ans	- 0,19**	0,09
Hommes de 50 à 64 ans	0,01	0,09
Femmes de plus de 65 ans	Réf.	Réf.
Niveau de revenus		
Élevé	Réf.	Réf.
Moyen	0,04	0,05
Faible	0,00	0,06
Régime		
MSA	- 0,42***	0,12
Cnamts et autres	- 0,45***	0,09
RSI	Réf.	Réf.
Nombre de personnes dans le ménage		
1 à 3	- 0,13**	0,05
> 4	Réf.	Réf.
État de la dentition		
Inconnu	Réf.	Réf.
Mauvais	0,2*	0,12
Bon	- 0,11	0,10
Très bon	- 0,39***	0,11
Problèmes de dents		
Oui	0,06	0,16
Non	- 0,14*	0,08
Inconnu	Réf.	Réf.
Prothèse		
Dentier	0,3**	0,12
Fixe	0,41***	0,10
Inconnu	Réf.	Réf.
Non	- 0,03	0,08
Redressement	1,7***	0,15

Lecture : * : significatif au seuil de 10 % ; ** : Significatif au seuil de 5 % ; *** : significatif au seuil de 1 %.

Le R² du modèle vaut 0,1006.

Champ : individus vivant au sein d'un ménage ordinaire de France métropolitaine.

Source : appariement Epas-ESPS.

RÉSUMÉS DES DÉROULEMENTS DES *HOT DECK* STRATIFIÉS

Tableau A
Imputation des dépenses en optique

	1 ^{er} tour	2 ^{ème} tour	3 ^{ème} tour	4 ^{ème} tour	5 ^{ème} tour	6 ^{ème} tour	7 ^{ème} tour	8 ^{ème} tour	9 ^{ème} tour
Individus auxquels sont imputées des dépenses en optique									
Nombre	5 843	2 309	436	1 140	613	974	296	87	4
%	49,93	19,73	3,73	9,74	5,24	8,32	2,53	0,74	0,03
Variables de stratification									
Spécialiste	x	x	x	x	x	x	x	x	x
Nature de la couverture complémentaire	x	x	x	x	x	x	x	x	x
Classes d'âge	x	x	x	x	x	x	x	x	
Régime	x	x	x	x	x	x	x		
Indicateur synthétique de problèmes de vue	x	x	x	x	x	x			
Niveau de revenus	x	x	x	x	x				
Lunettes	x	x	x	x					
Niveau d'études	x	x	x						
Maladie vue	x	x							
Nb de personnes dans le ménage	x								
R ² du modèle associé (équation de montant)	0,2984	0,2981	0,2957	0,2911	0,2874	0,2742	0,2663	0,2553	0,1973

Lecture : lors du premier tour d'imputation, toutes les variables de stratifications sont utilisées et 49,93 % des individus se voient imputer des dépenses en optique.

Champ : individus de l'enquête SPS, non appariés avec l'Epas.

Source : Omar 2006.

Tableau B
Imputation des dépenses en dentaire

	1 ^{er} tour	2 ^{ème} tour	3 ^{ème} tour	4 ^{ème} tour	5 ^{ème} tour	6 ^{ème} tour	7 ^{ème} tour	8 ^{ème} tour
Individus auxquels sont imputées des dépenses en dentaire								
Nombre	5 228	2 048	2 354	871	399	138	81	583
%	44,68	17,50	20,12	7,44	3,41	1,18	0,69	4,98
Variables de stratification								
Prothèse	x	x	x	x	x	x	x	x
Classes d'âge	x	x	x	x	x	x	x	
Régime	x	x	x	x	x	x		
Problèmes de dents	x	x	x	x	x			
Nature de la couverture complémentaire	x	x	x	x				
État de la dentition	x	x	x					
Niveau de revenus	x	x						
Nb de personnes dans le ménage	x							
R ² du modèle associé (équation de montant)	0,1006	0,0992	0,0991	0,0878	0,0865	0,0849	0,0795	0,0593

Lecture : lors du premier tour d'imputation, toutes les variables de stratifications sont utilisées et 44,68 % des individus se voient imputer des dépenses en soins dentaires.

Champ : individus de l'enquête SPS, non appariés avec l'Epas.

Source : Omar 2006.

COMPARAISON DES RÉSULTATS D'OMAR AVEC LES COMPTES DE LA SANTÉ

Omar présente une structure de financement des dépenses de santé relativement proche de celle décrite par les Comptes de la santé. Les écarts qui peuvent apparaître tiennent surtout à des différences de champ.

Omar aboutit à un taux de remboursement de la Sécurité sociale un peu plus important que celui des Comptes de la santé (79 % contre 77 %) car il ne porte que sur les dépenses présentées au remboursement. Cette différence de champ est surtout perceptible pour la pharmacie, puisque l'automédication est prise en considération dans les Comptes de la santé alors qu'elle est exclue du champ de l'outil de microsimulation.

Les dépenses hospitalières sont sous-estimées dans *Omar*. Cet écart tient d'abord au champ retenu par

l'enquête SPS qui n'interroge que les ménages ordinaires. Or, les personnes en institution ont tendance à avoir des dépenses hospitalières plus importantes. De plus, les personnes les plus malades ont un taux de réponse plus faible à l'enquête SPS. Enfin, certaines dépenses hospitalières ne sont pas individualisables. Ce type de dépense est inclus dans les Comptes de la santé, mais pas dans *Omar*.

Enfin, une part des écarts s'explique par les aléas statistiques dus à la méthodologie qui repose sur des imputations successives.

Les similitudes entre *Omar* et les Comptes de la santé n'équivalent pas à une validation d'*Omar*. Mais *a contrario*, des chiffres agrégés très éloignés des comptes auraient invalidé l'outil. Ce n'est pas le cas.

Tableau
Comparaison des taux de remboursement entre *Omar* et les Comptes de la santé (2006)

En %

	<i>Omar</i>				Comptes de la santé			
	AMO	CMU-C	AMC	Ménage	AMO	CMU-C	AMC	Ménage
Médecins	73	2	17	9	71	2	18	9
Dentistes	37	3	33	26	35	2	35	28
Actes de biologie	72	2	22	4	74	2	21	3
Auxiliaires médicaux	84	1	13	1	81	1	9	9
Transports	89	1	7	3	94	1	3	2
Pharmacie	77	2	21	1	68	1	19	12
Autres biens médicaux	38	2	27	15	43	1	30	26
Total ambulatoire	68	2	21	10	65	1	21	13
Hôpital	96	0	2	2	92	1	4	3
Total	79	1	13	6	77	1	13	9

Champ : ménages ordinaires, France métropolitaine.

Source : *Omar* et Comptes de la santé.

