

LES IMPUTATIONS ÉCONOMÉTRIQUES : UN MAL NÉCESSAIRE ?

L'exemple de l'enquête Revenus fiscaux 1992

Didier Contencin et Daniel Verger

L'INSEE réalise en collaboration avec la DGI, depuis 1956 et tous les cinq ans environ, une enquête sur les revenus fiscaux des ménages. Pour l'édition 1992, 33 000 ménages ont été tirés aléatoirement, pour les deux tiers dans le recensement de 1990 et pour le tiers restant parmi les répondants de l'enquête Actifs financiers 1991, déjà présents dans leur logement au 31/12/1990, date de référence de l'enquête Revenus. Ce couplage partiel permet un enrichissement mutuel des deux opérations. Les services fiscaux repèrent tous les foyers fiscaux correspondant à ces ménages et récupèrent dans leurs fichiers et leurs dossiers les déclarations de revenus correspondantes, les montants d'impôt sur le revenu et de taxe d'habitation dus.

Cette enquête donne une information de bonne qualité sur le revenu soumis à l'impôt, d'excellente qualité même pour les revenus des contribuables imposables. Ses limites sont toutefois réelles : les revenus déclarés par les contribuables non imposables et qui ne font une déclaration que pour obtenir un certificat de non imposition sont moins précis ; à l'autre extrémité de l'échelle sociale certains revenus du capital, soumis à prélèvement libératoire, ne sont pas portés sur la déclaration de revenus et échappent donc au recensement. De plus, pour les indépendants, les conventions comptables et l'existence de forfaits rendent le revenu fiscal peu comparable avec celui des salariés. La limite qui va nous concerner ici est autre : elle tient à ce que la statistique obtenue, à savoir le revenu fiscal, avant ou après impôt, n'est pas la statistique qui intéresse au premier chef les partenaires sociaux, le monde politique ou les chercheurs analysant l'inégalité de ressources.

L'ensemble des prestations versées aux ménages, qui sont par nature non imposables, échappe totalement à la mesure, puisqu'elles ne figurent pas sur les documents fiscaux. Sauf à produire une statistique purement fiscale, d'intérêt assez confidentiel, et à laisser aux enquêtes auprès des ménages, et à elles seules, le soin d'éclairer la répartition des revenus disponibles, ce qui ne va pas sans difficultés (les ménages étant parfois réticents

à déclarer leur revenu à un enquêteur ou étant aussi victimes d'oublis), il faut compléter l'opération fiscale afin de reconstruire ce revenu disponible, en s'efforçant de bien rendre compte de la dispersion réelle des sommes perçues, de façon à permettre une analyse fiable de l'inégalité et de son évolution.

La tâche pourrait sembler aisée à première vue, chaque prestation étant régie par des règles d'attribution précises et par des barèmes officiels sans ambiguïté. Même si quelques zones d'ombre subsistent, en particulier dans le cas du RMI, par exemple pour définir les revenus pris en compte pour le calcul de cette indemnité différentielle ou pour délimiter le noyau familial retenu, elles n'ont qu'une importance assez marginale. Le problème est ailleurs ; il vient de ce que l'enquête Revenus Fiscaux est très pauvre en descripteurs sociodémographiques des ménages et des individus : seuls sont disponibles pour l'ensemble de l'échantillon les renseignements portés sur la déclaration de revenus et qui servent à calculer le nombre de parts du quotient familial ou les divers abattements. On connaît donc la liste des membres composant chaque foyer fiscal, leur lien avec le chef de foyer, l'âge de ce dernier, son lieu de résidence. On dispose aussi d'un libellé de profession souvent incomplet, imprécis. On dispose, de par les variables présentes dans la base de sondage, des caractéristiques du logement, telles quelles sont décrites dans le recensement ainsi que des caractéristiques de la commune d'habitat.

Pour les ménages n'ayant pas déménagé entre le recensement et la date de référence de l'enquête (ici entre mars et décembre 1990), soit environ 90 % de l'échantillon, on dispose des descripteurs du ménage et des individus qui le constituent fournis par le recensement, qui sont plus nombreux et forment un ensemble plus riche : meilleure appréhension de la catégorie socioprofessionnelle, connaissance du diplôme des individus...

Enfin pour le tiers de l'échantillon provenant d'Actifs financiers, on dispose des variables - très nombreuses - de l'enquête, parmi lesquelles on peut citer le niveau et la composition du patrimoine, la valeur du logement, la carrière professionnelle suivie par la personne de référence et par son (éventuel) conjoint. Mais, même dans ce dernier cas, on ne dispose pas des variables essentielles pour appliquer les barèmes ou déterminer les ayant-droits de nombreuses prestations. On ne sait pas si, dans le foyer, la présence d'un enfant ou d'un adulte handicapé ouvre droit à la prestation correspondante : même si les montants macroéconomiques sont faibles, au niveau microéconomique il n'en va pas de même. Dans le cas du minimum vieillesse, on ne sait si le ménage a fait valoir ses droits et demandé l'allocation : certains préfèrent n'en rien faire, les montants reçus étant en principe déductibles de l'héritage laissé aux descendants. De plus les délais de latence, même s'ils se sont réduits, restent non négligeables. Enfin, et c'est ce qui va susciter le recours aux méthodes présentées dans ce papier, les variables clefs dans la détermination des aides au logement, à savoir le loyer payé par les locataires ou les charges de remboursement d'emprunt pesant sur les accédants à la propriété, sont totalement indisponibles.

La procédure d'imputation des diverses allocations est assez complexe. Lorsque l'on connaît les variables essentielles au calcul de l'allocation, on impute à chaque ménage le montant théorique de l'allocation à laquelle il a droit, calculé sur les barèmes, parfois légèrement schématisés par éradication de complexités mineures. C'est ce que l'on fait pour les prestations familiales, à savoir les allocations familiales proprement dites, le complément familial, l'allocation pour jeune enfant, les allocations de rentrée scolaire et de parent isolé.

Le nombre d'ayant-droits au minimum vieillesse se trouvant être, dans l'enquête, égal à peu de choses près au nombre réel de ménages percevant l'allocation, on a également imputé cette prestation par application du barème théorique, en négligeant les quelques problèmes évoqués ci-dessus.

A l'opposé de ces cas favorables, on trouve les allocations pour lesquelles on ne saurait arriver à une imputation correcte, par aucune méthode, à la fois parce que manquent dans le fichier les variables essentielles et parce que manque l'information exogène qui permettrait l'application des méthodes économétriques dont on va parler. Parmi ces prestations non imputées, on peut citer l'allocation aux adultes handicapés, l'allocation parentale d'éducation, les allocations de garde d'enfant à domicile, d'éducation spéciale et de soutien familial.

Les aides au logement sont dans une situation intermédiaire : l'indisponibilité des variables clefs rend impossible l'application directe d'un barème fort complexe (cf encadré), mais la disponibilité d'une bonne source statistique permettant de quantifier les liens entre variables sociodémographiques usuelles et montant des aides perçues rend envisageable l'emploi de méthodes économétriques.

I. Les principes de l'imputation et la méthode suivie

Il va s'agir d'affecter à chaque ménage de l'échantillon de Revenus Fiscaux une valeur estimée à partir d'une source exogène (Logement 1988). Deux options s'offrent a priori pour ce faire.

La première consiste à faire porter l'imputation sur le loyer (ou la charge représentée par l'accession) et à calculer l'allocation en appliquant le barème 1990 sur cette valeur estimée. La complexité de la formule, l'existence d'abattements difficiles à connaître ont fait rejeter cette voie.

La seconde option consiste à imputer directement l'allocation grâce à l'étude des montants alloués tels qu'ils apparaissent dans l'enquête Logement (situation observée en octobre 1988 et correspondant aux barèmes en usage pour 1988 et début 1989). Bien qu'elle ne soit pas exempte de difficultés, dues à la fois à une certaine imprécision de

l'observation des montants des allocations par l'enquête et, surtout, à l'évolution des barèmes survenue au 1^{er} juillet 1989, c'est cette voie qui a été choisie.

1. L'estimation :

Il s'agit de quantifier le lien entre le montant de l'allocation perçue et les descripteurs du ménage observés dans Revenus Fiscaux. Il ne s'agit donc pas de réaliser le meilleur ajustement économétrique possible compte tenu des descripteurs disponibles dans l'enquête Logement, mais de se restreindre au champ des descripteurs communs aux deux opérations, afin qu'à la fois l'estimation et l'imputation puissent être réalisées. Vu la relative pauvreté des descripteurs de R.F., on est donc dans une situation assez inconfortable pour cette estimation, puisque l'on est sûr que le modèle est incomplet, que des variables potentiellement importantes manquent.

Deuxième point à souligner : la distribution des allocations ayant une forte proportion de valeurs nulles, on ne peut utiliser l'analyse de variance classique.

Une méthode en deux étapes a été choisie, même si elle est sans doute moins pertinente, en théorie, qu'une méthode recourant à un modèle TOBIT, estimé en une seule étape.

Ce choix a été fait parce qu'il semblait que les problèmes dus à l'incomplète spécification rendait hasardeux le recours à un modèle TOBIT, assez délicat à estimer. Cette justification n'est sans doute pas très convaincante et, parmi les développements ultérieurs de ces travaux, devra figurer l'essai d'une méthode TOBIT. Mais revenons à la méthode qui a été employée :

1^{ère} étape : Étude de la probabilité pour un ménage de percevoir une allocation

Estimation d'un modèle Logit, la variable expliquée étant le fait de percevoir l'allocation (AL et APL respectivement), les variables explicatives étant les descripteurs du ménage et du logement communs aux deux enquêtes, à savoir la région de programme, le type de logement (individuel ou collectif, HLM ou non, doté du confort ou non), le nombre de pièces, le milieu social, le niveau de revenus, l'âge de la personne de référence du ménage et le type de ménage. Des régressions indépendantes ont été faites sur la population des locataires et sur celle des propriétaires accédants. (cf Tableau 1 : Résultats de l'estimation pour l'APL des propriétaires).

Tableau 1

Les ménages propriétaires attributaires d'une APL
 (Coefficients obtenus par estimation d'un modèle Logit)
 Variables compatibles avec l'enquête RF 90

La constante est de - 3,47

% paires concordantes 92 %

	Coef	Stud
• REGION		
Ile de France	- 0,49	-2,81
Bassin Parisien hors IDF	0,11	0,88
Nord Pas de Calais	- 0,24	- 1,56
Est	- 0,02	- 0,13
Ouest	0	0
Sud Ouest	0,14	0,96
Centre Est	- 0,09	- 0,60
Méditerranée	0,42	2,80
• HLM : (Appartenance à un organisme de...)		
oui	0,17	0,76
non	0	0
• MILIEU SOCIAL		
Indépendant	- 0,77	- 3,02
Cadre, technicien, prof. intermédiaire	- 1,00	- 2,59
Employé	0	0
Ouvrier	0,42	3,57
Retraité	0,16	0,58
• HABITAT		
Individuel	0,59	4,04
Collectif	0	0
• CONFORT DU LOGEMENT		
Autres cas que ci-après	- 0,40	- 1,32
WC, baignoire ou douche sans chauffage central	0	0
WC, baignoire ou douche avec chauffage central	0,41	3,66
• REVENUS ANNUELS		
Inférieurs à 54 000 F.	- 0,35	- 1,55
entre 54 000 et 72 000 F.	0,17	0,93
entre 72 000 et 96 000 F.	0	0
entre 96 000 et 120 000 F.	- 0,24	- 1,84
entre 120 000 et 180 000 F.	- 0,86	- 7,31
supérieurs à 180 000 F.	- 2,56	- 16,2
• TAILLE DU LOGEMENT		
1 ou 2 pièces	- 0,17	- 0,6
3 pièces	- 0,19	- 1,26
4 pièces	0	0
5 pièces	- 0,16	- 1,66
6 pièces et plus	- 0,39	- 3,66
• TYPE DE MENAGE		
Personne sans famille	0,70	1,26
Famille sans enfant	0	0
Famille avec enfants	0,47	0,89
• NOMBRE D'ENFANTS		
0	- 2,23	- 10,7
1	- 1,09	- 9,65

Tableau 1 (suite)

2	0	0
3	1,02	10,2
4 et plus	1,63	11,2
• AGE		
Moins de 35 ans	2,87	21,1
35 à 44 ans	1,47	11,5
45 à 54 ans	0	0
55 à 64 ans	- 0,80	- 3,30
65 ans et plus	- 1,32	- 3,66

2^{ème} étape : Étude du montant perçu par les ménages allocataires

Estimation d'un modèle d'analyse de variance sur la population des allocataires. Quatre champs différents ont été retenus : les propriétaires percevant l'AL, les propriétaires percevant l'APL, les locataires touchant l'AL et les locataires touchant l'APL. Pour chaque analyse, la variable expliquée était le logarithme du montant d'allocation perçu, les variables explicatives étant les mêmes descripteurs du ménage et du logement que pour le modèle Logit (cf Tableau 2 : résultats de l'estimation pour l'APL des propriétaires).

2. L'imputation :

Deux phases à nouveau :

* La première phase consiste à déterminer, dans l'échantillon de l'enquête, qui on va considérer comme allocataire, parmi les propriétaires et parmi les locataires, pour l'AL et pour l'APL.

Pour chaque ménage propriétaire (respectivement locataire), on calcule la probabilité de percevoir l'AL (et la probabilité de percevoir l'APL) par la formule :

$$\Pi_{ALP} = \frac{1}{1 + \exp \left[-\hat{\alpha}_{O,ALP} - \sum_i \hat{\alpha}_{i,ALP} X_i \right]} \quad (\text{et formule similaire pour l'APL})$$

où $\hat{\alpha}_{O,ALP}$ est la constante obtenue par l'estimation du modèle Logit, sur l'enquête Logement, pour la variable AL sur le champ des propriétaires

et $\hat{\alpha}_{i,ALP}$ est le coefficient estimé sur le même champ relatif à la variable X_i .

Tableau 2
Évaluation d'une APL fictive aux propriétaires
 (Coefficients obtenus par GLM)
 Variables harmonisées avec RF 90

La constante est de 7,11

$R^2 = 0,22$; RMSE : 0,70

	Coef	Stud
• REGION		
Ile de France	- 0,15	- 1,35
Bassin Parisien hors IDF	- 0,21	- 2,68
Nord Pas de Calais	- 0,24	- 2,59
Est	- 0,22	- 2,36
Ouest	- 0,19	- 2,32
Sud Ouest	- 0,12	- 1,42
Centre Est	- 0,17	- 1,98
Méditerranée	0	0
• HLM (Appartenance à un organisme de...)		
oui	0,0	0,01
non	0	0
• MILIEU SOCIAL		
Indépendant	0,04	0,26
Cadre, technicien, prof. intermédiaire	- 0,14	- 1,48
Employé.	- 0,12	- 2,27
Ouvrier	- 0,02	- 0,35
Retraité	0	0
• HABITAT		
Individuel	- 0,21	- 2,42
Collectif	0	0
• CONFORT DU LOGEMENT		
Autres cas que ci-après	- 0,27	- 1,54
WC, baignoire ou douche sans chauffage central	- 0,05	- 0,83
WC, baignoire ou douche avec chauffage central	0	0
• REVENUS ANNUELS		
Inférieurs à 54 000 F.	0,85	5,74
entre 54 000 et 72 000 F.	0,65	5,18
entre 72 000 et 96 000 F.	0,56	5,67
entre 96 000 et 120 000 F.	0,48	5,16
entre 120 000 et 180 000 F.	0,30	3,47
supérieurs à 180 000 F.	0	0
• TAILLE DU LOGEMENT		
1 ou 2 pièces	- 0,22	- 1,13
3 pièces	- 0,02	- 0,23
4 pièces	- 0,15	- 0,26
5 pièces	0,04	0,64
6 pièces et plus	0	0
• TYPE DE MENAGE		
Personne sans famille	- 0,26	- 1,31
Famille sans enfant	0,10	0,25
Famille avec enfants	0	0
• NOMBRE D'ENFANTS		
0	- 1,18	- 7,35
1	- 1,05	- 11,09

Tableau 2 (suite)

2	- 0,33	- 11,04
3	- 0,43	- 5,85
4 et plus	0	0
• AGE		
Moins de 35 ans	0,30	1,23
35 à 44 ans	0,13	0,51
45 à 54 ans	0,0	- 0,02
55 à 64 ans	0,02	0,08
65 ans et plus	0	0

La situation de référence est celle qui est choisie automatiquement par la procédure (dernière modalité de chaque code). Elle présente un manque de cohérence qui n'est pas gênant pour l'usage qui est fait des résultats

On classe les ménages par probabilité décroissante et les NALP premiers ménages sont réputés percevoir l'allocation, NALP étant l'effectif de propriétaires percevant l'AL, au 31/12/1990 d'après les comptes publiés par la CNAF.

On aurait pu, et ceci peut fournir aussi l'occasion de travaux ultérieurs, tirer pour chaque ménage le fait d'être allocataire ou non dans une loi, compte tenu de la probabilité calculée pour ce ménage. Par exemple si la probabilité vaut 0,8 on tirera le fait d'être allocataire dans une loi uniforme, avec le seuil de 0,8. (80 % de chance d'être allocataire).

On respecterait sans doute mieux ainsi le caractère aléatoire de la méthode. L'affectation choisie est, à ce stade, sans doute un peu trop déterministe ; le caractère aléatoire a par contre été respecté dans l'imputation des montants, qui fait l'objet de la deuxième phase.

* Pour cette deuxième phase, pour chaque ménage "estampillé" allocataire à la première phase, on calcule un montant d'allocation par la formule :

$$\hat{A}L_P = \exp \left[\hat{\beta}_{O,ALP} + \sum_i \hat{\beta}_{i,ALP} X_i + v \hat{\sigma} \right]$$

$\hat{\beta}_{O,ALP}$ et $\hat{\beta}_{i,ALP}$ étant les coefficients obtenus par l'analyse de variance,

v aléa tiré dans une loi normale $\mathcal{N}(0,1)$,

$\hat{\sigma}$ étant l'écart type estimé de la régression.

Afin que le tirage du résidu ne conduise pas à des valeurs aberrantes, compte tenu des divers plafonds légaux, on ne retient la valeur ainsi calculée que si elle ne dépasse pas un certain seuil plafond, défini à partir des valeurs réelles observées dans l'enquête Logement. Sinon, on fait un second tirage.

L'avantage du tirage du résidu aléatoire est de ne pas reconstituer une distribution trop peu dispersée. La distribution reconstituée a la même variance que la distribution initiale (pour les propriétés des résidus simulés, voir Gouriéroux, Monfort). Evidemment au niveau individuel, la méthode se justifie moins.

II. Validation de la démarche

Par construction le nombre d'allocataires est bon. Un seul bémol doit être introduit, qui provient du fait que l'enquête est pondérée et que la distribution des poids, reflet de la forte surreprésentation des catégories aisées dans l'échantillon, est très dispersée (poids allant de 53 à 4942, avec des agrégats sur certaines valeurs). Si, dans la zone limite entre ceux que l'on qualifiera d'allocataires et les autres, les pondérations sont fortes, on ne peut éviter toutes les discontinuités. Un certain écart avec les données exogènes subsistera, mais il ne s'agit ici que d'un problème mineur. On peut toutefois retenir cet aspect pour des travaux futurs : quand les deux enquêtes utilisées dans un tel processus d'imputation sont toutes deux pondérées, avec des variables de pondération très différentes, quelle est, en théorie, la meilleure façon de procéder ? Ne faudrait-il pas, par exemple, réaliser physiquement sur le fichier le processus de "géméllisation" que représente une pondération et appliquer à chaque individu "clone" ainsi obtenu indépendamment l'imputation, plutôt que d'affecter, à toute la grappe d'individus, la même valeur ?

La validation de la méthode telle qu'elle a été appliquée vise donc davantage la seconde phase que la première. Elle consistera à vérifier que les montants globaux obtenus par agrégation et extrapolation des imputations microéconomiques ont des ordres de grandeur compatibles avec les montants versés, tels qu'ils sont connus aux travers des statistiques de la CNAF. Soulignons que cette validation ne pourra être qu'approximative, le champ des enquêtes et le champ CNAF différant légèrement (ce n'est pas dans les deux cas le champ des "ménages ordinaires").

On peut penser que plusieurs conditions sont nécessaires (mais sans doute pas suffisantes) pour obtenir une imputation acceptable :

- La source de base doit être de qualité acceptable
- L'ajustement économétrique doit être satisfaisant dans "l'absolu", soit avec le jeu le plus complet de descripteurs théoriquement pertinents et pratiquement disponibles dans l'enquête Logement et il doit ne pas trop perdre de qualité quand on se limite au modèle réduit, incomplètement spécifié, n'introduisant que les variables disponibles dans Revenus Fiscaux.

Aucune de ces conditions n'est a priori satisfaite. Malgré tout le soin apporté à la collecte de l'enquête Logement, il faut reconnaître que les montants d'aide perçus

restent difficiles à obtenir de façon précise (en particulier parce que les évolutions en cours d'années peuvent échapper, ou surtout parce que les aides versées directement au propriétaire peuvent être mal connues). Si les régressions mettent en évidence des phénomènes qui paraissent vraisemblables, la variance expliquée dans la régression portant sur les montants, même dans le modèle le plus complet, reste faible, même comparée à ce que l'on obtient habituellement dans une régression sur données individuelles (le R^2 est ici de 0,35 pour l'APL des propriétaires, de 0,22 pour le modèle réduit aux variables RF).

Comme c'est souvent le cas, les valeurs estimées s'ajustent mal. On observerait d'ailleurs le même phénomène si l'on avait essayé d'ajuster le loyer perçu : aucun jeu de variables explicatives, dans la limite de ce qui est disponible dans l'enquête, ne permet d'ajuster les loyers les plus bas observés.

De plus rappelons que les barèmes ont changé mi-1989, ce qui crée une autre raison de divergence.

Pour 1990, la CNAF fournit, en ce qui concerne l'APL des propriétaires, les chiffres suivants : 977 000 allocataires, recevant une masse globale de $12,8 \times 10^9$ F, soit un montant moyen annuel de l'ordre de 13 100 F.

Le processus d'imputation décrit conduit à retenir, pour Revenus Fiscaux, les chiffres suivants : 966 000 allocataires, recevant $12,7 \times 10^9$ F, soit un montant moyen annuel de l'ordre de 13 150 F. Le caractère macroéconomiquement acceptable de l'imputation ne fait aucun doute. Il en va de même pour les quatre imputations réalisées, avec toutefois des écarts parfois un peu plus forts.

Toutefois, on peut considérer que la validation précédente est peu conclusive, le résultat, assez bon, pouvant n'être que le fruit du hasard, des écarts forts, en sens contraire, aux diverses étapes pouvant se compenser accidentellement.

Afin de mieux comprendre la méthode, de mieux en cerner les forces et les limites, on s'est donc livré à une seconde validation, en appliquant, la méthode à l'enquête Logement elle même : on pourra ainsi comparer sous divers angles la distribution "simulée" et la distribution "observée".

Plusieurs étapes peuvent être distinguées dans le processus de validation, que l'on va conduire dans le cas particulier de l'APL des propriétaires :

1. On estime le modèle Logit sur l'ensemble des propriétaires, la variable expliquée étant le fait de percevoir l'APL, les variables explicatives étant toutes celles disponibles dans Logement, a priori pertinentes (tableau 3). Par rapport aux variables finalement retenues, on a ici, en plus, le type de commune d'habitat, l'ancienneté d'occupation du logement, et des codes "confort" et "taille du logement" plus détaillés.

Tableau 3

Les ménages propriétaires attributaires d'une APL*(Coefficients obtenus par estimation d'un modèle Logit sur propriétaires)*

La constante est de - 3,08

% paires concordantes 95 %

	Coef	Stud
• RÉGION		
Ile de France	- 0,15	- 0,57
Bassin Parisien hors IDF	0,01	0,04
Nord Pas de Calais	- 0,14	- 0,80
Est	- 0,03	- 0,21
Ouest	0	0
Sud Ouest	0,05	0,32
Centre Est	- 0,07	- 0,44
Méditerranée	0,10	0,64
• HLM (Appartenance à un organisme de...)		
oui	0,34	1,39
non	0	0
• MILIEU SOCIAL		
Indépendant	- 0,87	- 3,22
Cadre, technicien, prof. intermédiaire	- 1,22	- 3,10
Employé	0	0
Ouvrier	0,35	2,71
Retraité	0,04	0,12
• TRANCHE D'UNITE URBAINE		
Commune rurale	0,13	1,09
Unité urbaine de moins de 2 000 hab.	- 0,05	- 0,33
Unité urbaine de 20 000 à moins de 100 000 hab.	- 0,42	- 2,60
Unité urbaine de plus de 100 000 hab. (sauf agglomération de Paris)	0	0
Agglomération de Paris	- 0,94	- 2,83
• HABITAT		
Individuel	0,50	3,02
Collectif	0	0
• CONFORT DU LOGEMENT		
Logement avec eau, WC, sans sanitaire	- 0,33	- 0,79
Eau, baignoire ou douche sans WC	- 0,38	- 0,75
WC, douche ou petite baignoire sans chauffage central	- 1,20	- 3,62
WC, douche ou petite baignoire avec chauffage central	0	0
WC, grande baignoire sans chauffage central	0,02	0,11
WC, grande baignoire avec chauffage central	0,26	1,59
• REVENUS ANNUELS		
Inférieurs à 54 000 F.	0,06	0,24
entre 54 000 et 72 000 F.	0,38	1,86
entre 75 000 et 96 000 F.	0	0
entre 96 000 et 120 000 F.	- 0,42	- 1,83
entre 120 000 et 180 000 F.	- 1,13	- 8,35
supérieurs à 180 000 F.	- 3,12	- 17,50
• TAILLE DU LOGEMENT		
Petit studio	- 0,74	- 0,64
Grand studio	- 1,37	- 0,92
Petit 2 pièces	1,08	2,12

Tableau 3 (suite)

Moyen 2 pièces	- 0,83	- 1,55
Grand 2 pièces	0,39	0,85
Petit 3 pièces	- 0,14	- 0,33
Moyen 3 pièces	- 0,32	- 1,47
Grand 3 pièces	0,09	0,32
Petit 4 pièces	- 0,41	- 1,62
Moyen 4 pièces	0	0
Grand 4 pièces	- 0,07	- 0,43
Petit 5 pièces	0,08	0,40
Moyen 5 pièces	- 0,18	- 1,26
Grand 5 pièces	- 0,33	- 1,79
Petit 6 pièces et plus	- 0,31	- 1,55
Moyen 6 pièces et plus	- 0,37	- 2,44
Grand 6 pièces et plus	- 0,76	- 3,02
• ANCIENNETÉ D'OCCUPATION		
Moins d'un an	- 0,60	- 3,43
de 1 à 3 ans	0,43	3,82
de 3 à 6 ans	0	0
de 6 à 10 ans	- 1,32	- 11,72
plus de 10 ans	- 3,57	- 20,57
• TYPE DE MÉNAGE		
Ménage sans famille	0,36	1,30
Famille principale monoparentale :		
avec 1 enfant	1,84	5,11
avec 2 enfants	2,38	6,47
avec 3 enfants et plus	2,69	6,20
Famille principale en couple : sans enfants	0	0
avec 1 enfant	1,32	5,58
avec 2 enfants	2,69	11,73
avec 3 enfants	4,03	16,78
• AGE		
Moins de 35 ans	1,14	7,71
35 à 44 ans	0,58	4,24
45 à 54 ans	0	0
55 à 64 ans	- 0,87	- 3,44
65 ans et plus	- 1,16	- 3,12

L'estimation semble d'assez bonne qualité.

2. On estime le modèle d'analyse de variance (GLM) sur les propriétaires percevant l'APL, tels qu'ils sont observés dans l'enquête. La variable expliquée est le logarithme du montant de l'APL, les variables explicatives toutes celles disponibles, comme pour le Logit (cf tableau 4).

La qualité de l'ajustement apparaît assez moyenne.

3. Ajustement d'un Logit analogue à celui de l'étape 1, avec les seules variables disponibles dans RF (tableau 4). La comparaison des ajustements 1 et 3 fait apparaître une perte de qualité somme toute acceptable.

Tableau 4

Évaluation d'une APL fictive aux propriétaires*(Coefficients obtenus par modèle GLM sur propriétaires percevant l'APL)*

La constante est de 6,70

 $R^2 : 0,35 ; RMSE : 0,64$

	Coef	Stud
• REGION		
Ile de France	- 0,24	- 1,78
Bassin Parisien hors IDF	- 0,14	- 2,02
Nord Pas de Calais	- 0,21	- 2,40
Est	- 0,15	- 1,75
Ouest	- 0,12	- 1,63
Sud Ouest	- 0,11	- 1,42
Centre Est	- 0,15	- 1,85
Méditerranée	0	0
. HLM (Appartenance à un organisme de...)		
oui	0,01	0,05
non	0	0
• MILIEU SOCIAL		
Indépendant	- 0,06	- 0,42
Cadre, technicien, prof. intermédiaire	- 0,24	- 2,67
Employé	- 0,11	- 2,26
Ouvrier	- 0,01	- 0,15
Retraité	0	0
• TRANCHE D'UNITE URBAINE		
Commune rurale	- 0,13	- 0,80
Unité urbaine de moins de 2 000 hab.	- 0,22	- 1,29
Unité urbaine de 20 000 à moins de 100 000 hab.	- 0,11	- 0,61
Unité urbaine de plus de 100 000 hab. (sauf agglomération de Paris)	- 0,10	- 0,56
Agglomération de Paris	0	0
• HABITAT		
Individuel	- 0,22	- 2,53
Collectif	0	0
• CONFORT DU LOGEMENT		
Logement avec eau, WC, sans sanitaire	- 0,12	- 0,60
Eau, baignoire ou douche sans WC	- 0,47	- 1,84
WC, douche ou petite baignoire sans chauffage central	- 0,20	- 1,28
WC, douche ou petite baignoire avec chauffage central	- 0,02	- 0,22
WC, grande baignoire sans chauffage central	- 0,05	- 0,80
WC, grande baignoire avec chauffage central	0	0
• REVENUS ANNUELS		
Inférieurs à 54 000 F.	1,02	7,19
entre 54 000 et 72 000 F.	0,81	6,84
entre 75 000 et 96 000 F.	0,68	7,34
entre 96 000 et 120 000 F.	0,57	6,58
entre 120 000 et 180 000 F.	0,36	4,50
supérieur à 180 000 F.	0	0
• TAILLE DU LOGEMENT		
Petit studio	- 0,30	- 0,42
Grand studio	- 1,24	- 1,39

Tableau 4 (suite)

Petit 2 pièces	- 1,24	- 3,76
Moyen 2 pièces	- 0,40	- 1,13
Grand 2 pièces	0,06	0,23
Petit 3 pièces	- 0,73	- 2,86
Moyen 3 pièces	- 0,10	- 0,64
Grand 3 pièces	- 0,08	- 0,45
Petit 4 pièces	- 0,10	- 0,63
Moyen 4 pièces	- 0,11	- 0,91
Grand 4 pièces	- 0,15	- 1,13
Petit 5 pièces	- 0,09	- 0,65
Moyen 5 pièces	- 0,11	- 0,88
Grand 5 pièces	0,02	0,19
Petit 6 pièces et plus	- 0,06	- 0,44
Moyen 6 pièces et plus	- 0,10	- 0,81
Grand 6 pièces et plus	0	0
• ANCIENNETE D'OCCUPATION		
Moins d'un an	0,68	5,38
de 1 à 3 ans	0,73	7,16
de 3 à 6 ans	0,72	6,98
de 6 à 10 ans	0,04	0,45
plus de 10 ans	0	0
• TYPE DE MENAGE		
Ménage sans famille	- 1,06	- 7,04
Famille principale monoparentale :		
avec 1 enfant	- 0,59	- 3,51
avec 2 enfants	- 0,36	- 2,51
avec 3 enfants et plus	0,23	1,29
Famille principale en couple : sans enfant	- 0,91	- 6,35
avec 1 enfant	- 0,85	- 11,7
avec 2 enfants	- 0,55	- 11,7
avec 3 enfants	0	0
• AGE		
Moins de 35 ans	- 0,09	- 0,39
35 à 44 ans	- 0,13	- 0,56
45 à 54 ans	- 0,12	- 0,49
55 à 64 ans	- 0,02	- 0,07
65 ans et plus	0	0

La situation de référence est celle qui est choisie automatiquement par la procédure (dernière modalité de chaque code). Elle présente un manque de cohérence qui n'est pas gênant pour l'usage qui est fait des résultats.

4. On procède de même à une estimation du GLM "réduit". La comparaison de 2 et 4 montre que la qualité finale est moyenne, la perte supplémentaire de qualité n'étant pas catastrophique.

Le résultat obtenu par le processus enchaîné 3 + 4 peut être comparé au résultat CNAF de 1988.

Tableau 5
APL (propriétaires) (annuelle)

	Moyenne (88 ; F)	Médiane	Masse (109 F)	Nombre alloc.
CNAF	11 600	-	11,4	990 000
Logement (observé)	11 800	10 600	10,8 -	907 000 -
Estimation toutes var. exp., processus complet	12 700 +	10 300	12,5 +	988 000
Estimation toutes var. exp., sur champ "réel"	12800 +	9 300	11,7 =	907 000
Estimation var. "R.F.", processus complet	12 900 +	9 800	12,6 +	985 000
Estimation var. "R.F.", champ "réel"	11 800	8 900	10,8 -	907 000 -

(validation 88 : enq. Logement)

Tableau 6
Quelques éléments pour comprendre les écarts

	Logt (obs)	Estimations	
		ttes var.	var "RF"
min	50	67	74
D ₁	250	320	324
Q ₁	486	508	487
M	890	871	814
Q ₃	1 370	1 428	1 350
D ₉	1 860	2 121	2 345
Max	3 752	3 800	3 500
(valeurs mensuelles)			
D ₉ /D ₁	7,4	6,6	7,2
Q ₃ /Q ₁	2,8	2,8	2,8

Tableau 7

Quelques données sur les masses des prestations par rapport aux divers types de revenus

1. Passage du revenu fiscal au revenu disponible

- Revenu fiscal : 148 000 F. (moyenne)
- Apport des Prestations sociales : (en %)
 - 5,3
 - dont
 - 2,8 prestations familiales
 - 1,5 prestations logement
 - 1,0 autres prestations
- Prélèvement fiscal : (en %)
 - 9,6
 - dont
 - 8,4 Impôt sur le revenu
 - 1,1 Taxe d'habitation
- Revenu disponible en % du revenu fiscal : 95,5 %
(selon le type de ménage, ce chiffre varie de 87,5 % à 111,6 %)

2. Les masses des diverses allocations imputées :

ALLOCATION	REEL		RF 90	
	MASSES (x106 F)	NB D'ALLOCATAIRES (x103)	MASSES (x106 F)	NB D'ALLOCATAIRES (x103)
ALLOC. FAM.	62 800	4 500	59 600	4 440
COMPL. FAM.	8 360	880	7 800	850
APJE	19 300	1 900	19 400	1 950
ALLOC. REN. SCO.	1 930	2 700	1 960	2 420
API	3 750	130	2 900	121
AL	21 200	2 150	21 000	2 200
APL	27 500	2 400	27 150	2 360
MIN. VIEIL.	≈ 24 000	≈ 1 500	20 000	1 200
RMI	8 000	520	8 700	470

3. Masses des prestations non imputées

- AAH ; 14 Mds de F.
- APE ; 5,9 Mds de F.
- AGED ; 0,2 Mds de F.
- A.E.S. ; 1,2 Mds de F.

Afin de mieux comprendre les différences, on peut simuler une estimation des montants sur champ réel (étape 4 seulement, appliquée aux bénéficiaires observés et non aux bénéficiaires estampillés à la phase 3). (cf tableau 5)

L'écart entre l'enquête Logement et les statistiques CNAF provient sans doute en partie de la différence de champ, mentionnée plus haut. Le "bon" chiffre correspondant à l'ensemble des ménages ordinaires au sens de l'INSEE est sans doute intermédiaire. Des travaux plus précis devraient sans doute s'attacher à affiner les statistiques exogènes utilisées pour le calage du modèle. La principale information du tableau est de faire apparaître une tendance assez nette à la surestimation des montants (environ 10 %) qui provient d'une incapacité du processus à reconstituer le bas de la distribution et d'une certaine tendance à surestimer le haut.

Le premier décile de la vraie distribution (celle de l'enquête) est en effet égal à 250 F. par mois, alors qu'il est de l'ordre de 320 F. dans les distributions reconstruites (324 F. lorsque seuls les variables RF sont utilisées). De même D_9 est trop fort (2 345 F. contre 1 860 F.) alors que les déciles intermédiaires sont plus faibles. D_9/D_1 semble par contre plus faible que sur la vraie distribution (7,2 et 6,6 selon que l'on utilise seulement les variables RF ou le modèle le plus riche, contre 7,44).

La distribution reconstruite est sans doute moins dispersée, mais curieusement la situation semble meilleure sur le modèle le plus incomplet ! Ces résultats demandent toutefois à être confirmés par d'autres variantes. Ils sont donnés ici à titre illustratif et ne sauraient être pris pour des résultats archétypiques alors que rien ne prouve qu'ils ne sont pas qu'anecdotiques.

Quelques précisions sur les conditions d'attribution des aides au logement et la détermination de leur montant (situation en 1990)

Deux familles d'allocations : l'allocation logement et l'aide personnalisée au logement.

1. L'allocation logement

a) L'allocation à caractère familial

- Cinq catégories de bénéficiaires :
 - bénéficiaires d'une prestation familiale,
 - non bénéficiaires d'une telle prestation, mais avec enfant à charge,
 - jeunes ménages sans enfant, mariés depuis moins de 5 ans, le conjoint ayant moins de 40 ans au moment du mariage,
 - ménages avec ascendant à charge, de plus de 65 ans ou de plus de 60 ans si inapte au travail, le revenu de l'ascendant étant inférieur à 34 050 F,
 - ménages avec ascendant, descendant ou collatéral à charge, présentant une infirmité supérieure à 80 %.
- Formule de calcul :
 - $AL = K \cdot (L + C - L_0)$,
avec :
 - AL : montant mensuel
 - K : coefficient de prise en charge $k = 0,9 -$
 - L : loyer principal plafonné (par zone géographique), ou charge accession
 - C : majoration forfaitaire pour charges
 - L_0 : loyer minimum à charge, dépendant des ressources et de la composition familiale

b) L'allocation à caractère social

- Quatre catégories de bénéficiaires :
 - les personnes âgées (de plus de 65 ans ou de plus de 60 ans),
 - les invalides,
 - les jeunes travailleurs (de moins de 25 ans),
 - les chômeurs.
- La formule de calcul est la même que ci-dessus, avec une utilisation d'un loyer forfaitaire pour les résidents en foyer.

2. L'aide personnalisée au logement

- Les bénéficiaires :
 - logements occupés par leur propriétaire, construits ou améliorés depuis le 4/01/1977, grâce à un prêt PAP ou conventionné,
 - logements à usage collectif conventionné.
- Formule de calcul :
 - $APL = K \cdot (L + C - L_0)$
 - Diverses formules existent pour K et L_0 selon le cas donnant droit à prestation. A titre d'exemples :
 - $K_1 = 0,95 -$
(avec N = Nombre de parts)

- $K_A = 0,95$
- L : Loyer (ou charge d'accession) plafonné.
- Le plafond dépend de la date d'accession et du type de logement (neuf ou ancien...)
- C : majoration forfaitaire pour charges.
- L_0 : loyer minimum à charge.

BIBLIOGRAPHIE

GOURIEROUX, C., MONFORT, A., RENAULT, E., TROGNON, A., (1987), "Simulated Residuals", *Journal of Econometrics*, 34, pp. 202-252.

LOLLIVIER S., VERGER, D., (1988), "D'une variable discrète à une variable continue : une application de la méthode des résidus simulés", dans *Mélanges Economiques, essais en l'honneur de E. Malinvaud*, *Economica*.