

I PB et HV / 23/11/01

INTEGRATION DES REVENUS IMPUTES DANS LES TABLES MENAGES ET REVENUS : PROBLEMES ET CHOIX RETENUS

Des imputations de revenus sur des champs variables et uniquement sur les trois premières vagues

Dans la précédente version des fichiers de diffusion (CD-ROM du 31 décembre 1999), les revenus déclarés par les ménages et les revenus imputés par la division revenus et patrimoine figuraient en général dans des fichiers différents, en nombre variable selon les vagues (cf. annexe 2) :

- En première vague, seules les imputations sur les revenus d'activité (traitements et salaires d'une part, et revenus d'indépendants d'autre part) et les imputations sur les revenus du patrimoine ont été conservées dans des fichiers séparés. En revanche, les imputations réalisées sur cette vague sur les prestations familiales ont été considérées comme de meilleure qualité que les déclarations des enquêtes (beaucoup sont attribuées sur barème ou selon des conditions d'attribution simple), et les prestations simulées ont donc été conservées dans les fichiers REVENUVx.sd2 en remplacement des prestations déclarées (cf. pour plus de détail le document de travail n°9705, p 17). Les autres revenus sociaux n'avaient pas fait l'objet d'imputations.
- En deuxième vague, toutes les imputations ont été conservées séparément des données déclarées, et le champ des imputations a été élargi à l'ensemble des revenus sociaux (chômage, retraite, maladie-invalidité...). Au total, les données imputées faisaient l'objet de 11 fichiers différents (un par type de revenu).
- En troisième vague, le champ des imputations a été à nouveau réduit : parmi les revenus sociaux, seules les prestations familiales et les allocations logement ont continué à faire l'objet d'imputations, dont les résultats étaient diffusés dans 5 fichiers.

Pour plus de détail, le lecteur intéressé par les méthodes d'imputations employées et les choix effectués pourra se reporter aux documents de travail de la Direction des Statistiques Démographiques et Sociales sur le panel européen (n°F9705, F9715 et F2001).

A partir de la quatrième vague, il a été décidé d'arrêter de procéder à des imputations de revenus ; notamment car :

- Il apparaissait clair que le panel allait être de plus en plus utilisé dans sa dimension longitudinale, et que les imputations essentiellement transversales menées jusqu'à présent¹ seraient inexploitable dans ce contexte².
- La non-réponse partielle devenait de plus en plus réduite d'une vague sur l'autre (cf. exemple des revenus d'activité ci-dessous). On peut notamment penser que les enquêtés continuant à accepter l'enquête étaient les mieux disposés à y répondre correctement, que les enquêteurs avaient acquis une plus grande expérience sur le panel et étaient de mieux en mieux formés aux problèmes qu'il pose, et enfin, que les contrôles de cohérence lors de la collecte rendus possibles

¹ Même si dans les imputations économétriques, la valeur de l'année précédente intervient en général comme variable explicative.

² Alors que les variations de revenus sont en général faibles d'une année sur l'autre pour un individu donné, ce n'est pas le cas pour les revenus imputés compte tenu de l'inévitable imprécision des imputations (les régresseurs n'expliquent qu'une faible part de la variance). Etudier les variations de revenu en tenant compte de données imputées fait donc courir le risque d'analyses fortement biaisés par ces données.

par le passage à une collecte assistée par ordinateur (CAPI) à partir de la vague 5 ont également contribué à limiter les oublis.

Le fait que le champ des revenus imputés ait beaucoup varié d'une vague à l'autre rend l'usage de ces imputations délicat. Néanmoins, il nous a semblé important d'entreprendre un travail supplémentaire sur les données pour faciliter l'usage de ces imputations : il s'agissait de réunir pour une vague donnée toutes les données de revenus déclarés et imputés dans le même fichier (REVENUVx.sd2 pour les revenus déclarés au niveau individuel ou MENAGEVx.sd2 pour les revenus déclarés au niveau du ménage), afin que l'utilisateur de ce fichier puisse aisément opter entre des données éventuellement imputés ou non (sans avoir à gérer des données diffusés dans des formats complètement différents). Ce travail s'est en fait révélé assez long, pour des raisons qui seront exposées ci-dessous. Il permettra aux utilisateurs de tester facilement l'incidence de la non-réponse partielle sur leurs résultats, et surtout, a permis de souligner les multiples difficultés qui se posent à la personne souhaitant réaliser des imputations (et donc les limites de ces dernières). En gros, pour une enquête comme le panel où la non-réponse partielle est très réduite, il n'est pas sûr que les imputations (en tout cas telles qu'elles sont réalisées actuellement) améliorent systématiquement la qualité de l'information statistique. Si l'on peut penser qu'elles améliorent pour des prestations sur barème comme les allocations familiales, c'est beaucoup moins évident pour les revenus d'activités.

Mise en cohérence des fichiers d'imputations et des fichiers de revenus déclarés.

Regrouper dans un même fichier les revenus individuels imputés et déclarés, et les mettre en cohérence s'est révélé être un travail lourd. En effet, un examen préalable des données déclarées et imputées a montré que l'on ne pouvait se contenter de faire une simple fusion des fichiers des données déclarées et imputées, pour plusieurs raisons :

- Une même personne peut avoir effectivement un revenu dans le fichier REVENUVx.sd2, et s'en être vu cependant également imputer un. En effet, en premier lieu, les imputations ont été réalisées sur des fichiers de données peu apurés. Or le lourd travail d'apurement réalisé par la suite sur les premières vagues a souvent permis de récupérer ou d'estimer des montants préalablement manquants, qu'il n'aurait donc pas été nécessaire d'imputer³. En second lieu, dans certains cas, des imputations ont été faites de manière systématique (i.e. qu'il y ait ou non un montant déclaré), par exemple à l'aide de barèmes ou pour imposer le respect de contraintes particulières (par exemple, le fait qu'un retraité ne devait pas voir sa pension baisser d'une année sur l'autre) : dans ces cas, il apparaissait logique que les montants déclarés et imputés puissent différer, et il fallait s'interroger pour savoir si nous souhaitions ou non que les fichiers « définitifs » reflètent ces divergences⁴.
- Pour un type de revenu donné, une personne peut figurer dans le fichier des revenus déclarés mais pas dans le fichier des revenus imputés (ou vice versa). En effet, les fichiers de revenus déclarés et imputés ne comportent des enregistrements que pour les personnes qui perçoivent le

³ En effet, lors de chaque vague, en automne, les enquêtés sont interrogés non seulement sur leurs revenus depuis le début de l'année civile, mais également sur leurs revenus de l'année précédente : il est donc possible de compléter les données de revenus relative à l'année complète dans une vague donnée par les données de revenus relatives à cette même année partiellement observée dans la vague précédente. Cela permet notamment de récupérer des revenus qui n'ont été perçus que quelques mois en début d'année, qui sont bien déclarés l'automne suivant, mais sont souvent oubliés par l'enquêté lors de l'enquête d'après. Par ailleurs, le fait de disposer de vagues précédentes, mais aussi suivantes, permet dans de nombreux cas d'estimer un revenu manquant par interpolation, ce qui conduit à des résultats plus fiables que par des méthodes économétriques.

⁴ En effet, disposant d'un nombre de vagues plus élevés, nous avons pu remettre en question certains choix effectués à partir d'une information plus limitée à l'époque où les imputations ont été effectuées. Par exemple, un retraité qui déclarait une pension en diminution entre les vagues 1 et 2 a pu se voir imputer une pension de vague 2 identique à celle de première vague, alors que maintenant, en examinant les vagues 3 et 4, on a l'impression que le niveau de pension déclaré en vague 2 était en fait le plus conforme à la réalité.

revenu étudié. Il est donc apparu nécessaire de se demander quelle information conserver dans ces cas. Ainsi, lorsqu'une personne avait un revenu déclaré dans le fichier REVENUVx.sd2, mais était absente du fichier d'imputations correspondant, devait-on en déduire qu'il y avait une erreur dans le fichier de déclaration (par exemple, personne n'ayant pas d'enfant ayant déclaré des allocations familiales), ou au contraire que cette incohérence apparente était liée à des degrés d'apurement différents ou encore au fait que les imputations n'ont parfois été effectuées que sur un sous-échantillon d'individus (par exemple, les personnes n'ayant que partiellement répondu ou celles ayant quitté le ménage en cours d'année ne semblent pas avoir été prises en compte pour les imputations, alors que l'on disposait parfois de revenus déclarés pour elles) ? En d'autres termes, devait-on plutôt faire confiance au fichier de revenus déclarés ou au fichier de revenu imputés ?

- Les imputations ont parfois été réalisées sur des types de revenus agrégés : les salaires, primes, indemnités de congés payés, pourboires n'ont pas été imputés séparément, mais globalement, ce qui a compliqué la comparaison entre données déclarées et simulées⁵.
- Un premier examen des données imputées et déclarées a montré que dans certains cas, des revenus avaient été imputés alors qu'il n'y avait sans doute pas lieu. En effet, en particulier dans les toutes premières vagues, les montants étaient très peu fréquemment codés à 'XXXXXXX' lorsque les revenus correspondants existaient mais n'étaient pas déclarés. Il n'était donc pas possible de distinguer facilement les revenus manquants des revenus non perçus. Ce problème a incité à examiner plus en détail les imputations effectuées, afin de s'assurer qu'elles ne conduisaient pas à surestimer les revenus concernés.
- Les imputations de revenus ont consisté dans de nombreux cas (pour les revenus sociaux notamment) en de l'apurement. Or, en reprenant pour l'année courante un revenu déclaré l'année précédente, on risque de compter deux fois ce revenu s'il a déjà été déclaré pour l'année courante dans une autre rubrique, ou s'il a été inclus dans un autre (cas des retraites complémentaires parfois déclarées avec la retraite principale, parfois déclarées seules). En cours et en fin du processus de validation des valeurs imputés, nous avons donc pris un soin particulier à vérifier que les choix effectués ne se traduisaient pas par des doubles comptes (notamment, une analyse systématique des montants de revenus apparaissant deux fois pour un même individu a été menée, afin de vérifier si cela ne résultait pas d'une intervention dommageable de notre part).

Les nouveaux fichiers REVENUVx.sd2

Le résultat de ce travail est la production d'un unique fichier de revenus individuels pour chaque vague, dans lequel les enregistrements sont repérés par le numéro d'individu concerné (NIDENT⁶), le type de revenu (CODFR) et l'année (AN). Pour chaque enregistrement de ce type, la variable MONT donne le montant dit « déclaré », qui en fait a pu être modifié lors de l'apurement, et la variable MONTS le montant « après simulation éventuelle » correspondant. Les montants simulés ne sont toutefois remplis

⁵ En effet, reconstituer la somme des revenus du bloc « traitements et salaires » implique de faire des choix, comme le traitement des revenus inclus dans un autre. Normalement, selon la documentation aux enquêteurs, les codes inclusions n'auraient dû être utilisés que pour indiquer que des revenus dont la personne ne connaissait pas le détail avaient bien été comptabilisés ailleurs. En fait, il arrive parfois qu'une personne déclare (par exemple) un montant d'indemnité de congés payés, mais déclare simultanément que ce montant est déjà inclus dans le salaire qu'elle a précédemment donné. Dans ce cas, faut-il ou non ajouter cette indemnité au salaire ? Après analyse, nous avons décidé de ne pas le faire, alors que le choix inverse avait été fait lors de la mise au point des imputations et dans le CD-ROM du 31 décembre 1999. En effet, les évolutions de revenu d'une année sur l'autre semblaient légèrement plus vraisemblables lorsque l'on ne comptabilisait les revenus déclarés « inclus » dans d'autres. De toute façon, le nombre d'individus concernés est faible, ainsi en général que les revenus impliqués. Ces points sont évoqués plus en détail dans le fichier des « questions réponses ».

⁶ En première vague, les revenus des personnes ayant quitté le ménage ont également été recueillis. Ces personnes n'ayant pas de numéro d'identification (NIDENT), les lignes correspondantes peuvent être repérées par le numéro du ménage qu'elles ont quitté (IDENT) et leur numéro d'individu dans ce ménage (NOI).

que pour les « années complètes » (ils valent ' ' pour les années partielles, car aucune imputation n'a été faite sur ces années).

Pour les années complètes, le montant simulé est dans la grande majorité des cas égal au montant déclaré (qu'en cas de doute nous avons privilégié). Il est différent lorsqu'il n'y avait pas de montant déclaré (la variable MONT a alors été recodée à 'XXXXXXX') et que nous avons jugé qu'il y avait bien lieu d'en imputer un, ou bien dans le cas des prestations familiales (CODFR XX à XX), pour lesquelles nous avons estimé que les données simulées pouvaient être considérées comme plus fiables que les données déclarées (ce sont donc les données simulées qui ont été conservées dans MONTS). Dans un nombre très faible de cas, la variable simulée vaut 'XXXXXXX' : cela se produit quand une analyse au cas par cas a montré qu'il y aurait sans doute eu lieu de faire une imputation de revenu, qui n'avait pas été faite.

L'annexe 1 de cette note détaille type de revenu par type de revenu, les problèmes rencontrés lors du rapprochement des données imputées et des données déclarées au niveau individuel, et les choix qui ont été faits. Ces choix l'ont en général été au cas par cas, ou par application de règles de décisions définies à partir d'analyse de plusieurs cas.

Annexe 1

Imputations des revenus d'activité

Voici quelques chiffres sur l'importance des données imputées dans le panel actuel, pour les revenus salariaux et les revenus d'indépendants.

1) Après examen (quasiment au cas par cas) des imputations de revenus salariaux et indépendants qui avaient été faites pour les vagues 1 à 3 du panel, nous avons considéré qu'il y avait lieu de faire une imputation des revenus salariaux dans 3,3% des cas en vague 1, 1,8% des cas en vague 2, et 0,7 % des cas en vague 3 imputation des revenus d'indépendants dans 13,7% des cas en vague 1, et dans 8,1% des cas en vague 2, et 9,2% en vague 3. Les chiffres des vagues 2 et 3 sont beaucoup plus faibles : d'abord, la qualité des données est meilleure (formation des enquêteurs...), et ensuite, il a été possible par un apurement préalable de récupérer certaines informations manquantes en vague 2 dans la vague 1 (ex : rémunération liée à un job de début d'année, déjà déclarée en V1, et "oubliée " en V2). N.B. les pourcentages se rapportent au nombre de revenus salariaux ou d'indépendants, et non au nombre d'individus (ils seraient plus faibles)

2) Or, les imputations d'origine étaient beaucoup plus nombreuses : n'ont été conservées que les 2/3 des imputations initialement faites en v1, et moins de la moitié des imputations v2. En effet, il est parfois particulièrement délicat de savoir si la personne aurait dû ou non percevoir une rémunération. Par exemple, dans de nombreux cas, les revenus d'indépendants imputés l'ont été pour des femmes qui se déclaraient 'aide familiale' et exerçaient la même profession que leur mari, qui déclarait des revenus d'indépendant : dans ces cas, il nous a semblé que les revenus du mari étaient ceux du couple, et qu'il n'y avait pas lieu d'en imputer à la femme. Parmi les autres cas d'imputations "erronées" : les personnes qui déclaraient d'autres types de revenus (par exemple des salariés déclarant des indemnités journalières de maladie ou de maternité, de jeunes thésards déclarant des bourses de thèse et non des salaires, des personnes déclarant leurs revenus salariaux comme revenus d'activité secondaire ; pour les indépendants, c'est le cas des indépendants salariés de leur entreprise : ils déclarent des salaires et non des revenus d'indépendants...).

Bien sûr, on peut dire que certains de ces problèmes sont liés au fait que le questionnement papier des vagues papier n'assurait pas de récupérer une variable fiable indiquant si la personne a ou non touché des salaires (ou des revenus d'indépendants). L'essentiel du travail restait alors à faire pour déterminer à qui il fallait imputer les revenus. Toutefois, ces problèmes risquent de ne pas disparaître complètement avec CAPI. Que déclarent par exemple les aides familiales si on leur demande si elles ont touché des revenus d'indépendants ? Tout dépend également des contrôles de cohérence effectués avec CAPI avec le calendrier d'activité : rappelle-t-on à la personne qui a été salariée deux mois en début d'année qu'elle doit déclarer un salaire ? Si non, se baser uniquement sur le fait que la personne déclare ou non avoir perçu un revenu pour l'imputer est insuffisant. Et faire autrement est très coûteux. Dans tous les cas, espérer se servir d'une variable issue de CAPI pour déterminer à qui imputer un revenu nécessite que soient prévues sous CAPI toutes les possibilités de non-perception de revenu. En bref :

- soit on se contente d'imputer pour les valeurs déclarées manquantes, sans se poser trop de questions sur le nombre de mois à imputer, et le travail peut être relativement réduit, mais à notre sens, le faible nombre de cas de revenus non déclarés le rend tout de même trop coûteux ;

- soit on fait un examen plus complet des situations individuelles pour rattraper les oublis, et cela devient très coûteux.

Enfin, dernière remarque : les personnes qui ne déclarent pas leurs revenus ont toutes les chances d'avoir des revenus faibles) : il y a des cas d'étudiants qui se sont déclarés salariés une partie de l'année, de femmes au foyer qui ont également déclaré un salaire (qui doit correspondre à un nombre d'heures restreint), de RMI-stes qui eux non plus n'ont pas dû percevoir beaucoup, ou d'indépendants ayant juste

démarré leur activité. Imputer des revenus "moyens" à ces personnes revient sans doute à sur-estimer leurs revenus, alors qu'une imputation d'un revenu nul serait sans doute plus proche de la réalité... Dans le même ordre d'idée, les imputations d'origine ont souvent été faites sur 12 mois (en particulier lorsqu'il n'y avait pas d'information sur le nombre de mois de perception), alors qu'une analyse au cas par cas a souvent montré que la période de perception des revenus était en fait plus courte (et les revenus imputés ont été réduits d'autant).

Imputations des revenus sociaux (niveau individu)

Pour chaque ligne de revenu concernée de la table REVENUVx, on regarde les 3 situations possibles;

- 1) il existe une valeur déclarée (MONT), et une valeur imputée à valeur manquante
- 2) il n'existe pas de valeur déclarée (MONT= blanc ou ne sait pas), et on devrait en avoir une. Donc il y a valeur imputée MONTS.
- 3) une valeur avait été déclarée, une autre a été simulée.

Pour certains revenus la valeur simulée résulte de l'application de barème (cas des prestations familiales), pour d'autre la valeur simulée pour une année est calculée à partir de la déclaration de l'année précédente.

A l'occasion de ces analyses, des erreurs manifestes de saisie ont été corrigées sur les 4 années.

Chômage.

Après examen de quelques cas, on a validé toutes les prestations liées au chômage.

Retraites.

Après une analyse détaillée, les montants déclarés ont été systématiquement privilégiés : ils étaient conservés même s'ils ne figuraient plus dans le fichier d'imputation ; en cas de divergence entre les montants déclarés et imputés, ils ont été conservés dès lors que l'évolution des revenus totaux du bloc retraite par rapport aux vagues précédentes et suivantes paraissait réaliste, ou que la différence (relative ou absolue) avec les revenus imputés était faible. De manière générale, le traitement de ce bloc a été rendu particulièrement complexe par le fait que les différentes composantes des retraites peuvent être déclarées par le même individu une année de manière globale, et l'année suivante de manière détaillée, et qu'il ne faut pas dans ce cas bien sûr rajouter les composantes détaillées lors de l'apurement. Pour limiter ce type de problème, nous avons basé nos règles de choix autant que faire se peut sur le montant total des retraites et non sur chaque ligne retraite particulière.

Prestations familiales.

Après analyse, la règle retenue pour les prestations familiales a été de valider systématiquement les données imputées sur barème. En effet, par rapport aux autres composantes de revenus, les prestations familiales sont plus faciles à imputer, et les imputations moins entachées d'erreur que les imputations économétriques. Une attention particulière a été portée aux éventuels transferts de déclaration 'un individu du ménage sur l'autre (par exemple, le père déclare les allocations familiales une année, la mère l'année suivante).

Plus précisément, les analyses et choix faits ont été les suivants :

Dans le cas 1) ; des prestations étaient déclarées par le ménage et la simulation les annulaient. Après vérification il semble que des ménages déclarent à tort des allocations auxquelles ils ne peuvent prétendre. Mis à part quelques exemples pour lesquels la suppression de la déclaration du ménage semble injustifiée, les imputations sont validées dans tous les cas.

Dans le cas 2) ; des prestations n'étaient pas déclarées par le ménage et la simulation est validée dans tous les cas.

Dans le 3ème cas, les valeurs déclarées et les simulées diffèrent ; 704 individus étaient concernés en vague 2. La moitié des divergences résultent d'une déclaration approximative de l'enquêté (valeur arrondie du barème). La valeur retenue est celle de l'imputation qui est l'application du barème. Pour l'autre moitié les différences peuvent être beaucoup plus importantes.

Il a été vérifié systématiquement que l'application de ces règles ne conduisait pas à comptabiliser deux fois le même revenu au sein du ménage.

Les valeurs simulées des nombres mois n'ayant pas été restituées dans les bases d'imputation, on a dû les réaffecter.

*pour 1 ou 2 enfants déclarés dans le ménage (pas de majoration des allocations familiales), après application des barèmes, on peut aisément affecter un nombre de mois. Pour les montants différents des barèmes, quand ils existent, on affecte le nombre de mois déclarés. Sinon, si le montant simulé est supérieur au montant annuel, on prend 12 mois. Pour les cas restants n'entrant pas dans cette configuration, le nombre de mois est laissé à blanc

*pour 3 enfants ou plus déclarés dans le ménage, il existe des majorations selon les prestations et selon l'âge de l'enfant et les revenus des parents. Après application des barèmes et calcul d'un montant annuel minimum (sans majoration), on peut aisément affecter un nombre de mois (en l'occurrence 12 mois dès que le montant simulé est supérieur au seuil minimum). Pour les cas restants, le nombre de mois est laissé à blanc ou celui déclaré a été conservé .

Invalidité

Très peu de cas à traiter. On valide toutes les simulations (qui relèvent de l'apurement).

Education, formation

L'analyse a souvent été menée au cas par cas, notamment car il a été nécessaire de vérifier que les prestations imputées (par report de prestations déclarées une autre année ou reclassement de certains revenus dans une autre rubrique) n'entraînaient pas de double compte, soit au niveau de l'individu (le partage entre les lignes 81 et 83 n'étant pas clair), soit au niveau du ménage (les bourses d'études pouvant être déclarées par les parents ou les enfants). En général, les imputations ont été conservées.

RMI

L'imputation consiste à reporter en V2 pour 94 ce qui avait été déclaré en V1. En cas de différence entre les 2 années, on peut supposer que la mémoire joue négativement en V2 pour l'année écoulée. On privilégie donc le montant simulé (estimé à partir de la déclaration de l'année en cours)

Imputations des revenus déclarés au niveau ménage

Allocations logement

Exemple de la vague 2 ;

Quand une valeur est déclarée , la valeur retenue dans MONTS est égale à la déclarée (230 ménages)
Dans les cas où les valeurs simulées diffèrent des valeurs déclarées, plusieurs configurations sont possibles :

* pour 236 ménages, la valeur en Vague 2 est absente et elle existait pour la même année en Vague 1 (225). Dans ce cas la simulation consiste à reporter le montant déclaré à la vague précédente.

*pour 1033 ménages, la valeur simulée est différente de la valeur déclarée en vague 2. Dans tous les cas sauf un, la simulation a consisté à reporter la valeur déclarée en première vague pour l'année 94.

Les imputations ont donc toutes été validées.

Les placements financiers

Pour chaque type de produits financiers, les déclarations des ménages se font dans un système de tranches. En appliquant la méthode des résidus simulés, on simule un montant pour chaque ménage détenteur. Ensuite, en appliquant un taux de rendement (source Compte de patrimoine), les revenus par groupe de produits financiers sont estimés.

Pour la variable « revenus des placements financiers » déclarée également dans une variable à réponse en tranches, on estime un montant en clair à partir de la réponse quand elle existe. Sinon, le revenu est estimé à partir des déclarations faites produits par produits.

Les placements immobiliers

La simulation est effectuée quand le ménage déclare posséder des biens immobiliers et ne déclare pas les loyers perçus (brut et/ou net). Pour les vagues 2 et 3 quand l'un des deux est déclaré, l'autre est simulé en appliquant un taux de passage de 57.45% (taux qui avait été appliqué en vague 2 d'après les déclarations de la vague, et appliquer également en vague 3). Quand les deux montants sont déclarés, les valeurs simulées sont égales aux déclarées.

Pour la vague 2 par exemple, dans 9 cas, on n'a observé aucun montant déclaré de loyers pour des ménages déclarant détenir des biens immobiliers. La simulation, pour ces ménages est également nulle et restera en l'état.

A l'inverse, 28 ménages déclarent ne pas posséder de biens, ne déclarent aucun montant de loyer, et la simulation leur affecte des loyers bruts et nets. Tous ces ménages sont des ménage éclatés (EC=2). Par ailleurs, dans leur ménage initial (EC=1), des biens immobiliers et des loyers perçus sont déclarés pour 20 d'entre eux. Pour les 8 autres, il n'y a apparemment pas de disposition de biens.

On considère donc pour ces 28 ménages que la simulation ne doit pas être retenue.

L'impôt sur le revenu en vague 2

L'imputation de l'impôt est effectuée d'après barème, pour tous les ménages, à partir d'un fichier foyer fiscal. Ainsi on retiendra, pour un ménage, la somme des montants payés séparément par individus (SIRP). La simulation peut imputer un impôt nul, y compris pour un ménage déclarant payer de l'impôt (dans ces cas on corrige la valeur de IMPR).

- 1) Si un ménage déclare devoir payer l'impôt sur le revenu et ne déclare aucun des 2 montants (MIMPR et MSEPAR), on retient en valeur imputée la somme SIRP des différents montants imputés pour le calcul de l'impôt pour le ménage. **IRT94 ?**
- 2) Pour les autres ménages, qui déclarent bien un (ou plusieurs) montant, cette déclaration est privilégiée dans le calcul de l'impôt du ménage

Annexe 2 - Les bases d'imputations de revenus figurant dans le CD-ROM du 31 décembre 1999

Les 3 vagues contiennent les imputations de revenus d'activité et des revenus du patrimoine :

- les bases RPATVxS.SD2 rassemblant les imputations de revenus du patrimoine pour les ménages

- les bases RACTVxS.SD2 rassemblant les imputations de revenus d'activité pour les individus. Les variables imputées contiennent soit la valeur déclarée lors de l'enquête, soit la valeur simulée en cas de non-réponse.

En ce qui concerne les revenus sociaux, les 3 vagues n'ont pas été traitées de façon identique

1) Pour la première vague, les valeurs simulées remplacent celles déclarées à l'enquête quand elles étaient manquantes ou peu probables

2) Pour la deuxième vague, les bases générales contiennent les valeurs déclarées à l'enquête : sont donc également mises à disposition les huit bases d'imputation .

- la base SIMUFAMI.sd2 rassemble les imputations liée à la famille
- la base SIMULOG.sd2 rassemble les imputations sur l' allocation logement
- la base SIMUTEM.sd2 rassemble les imputations concernant les transferts entre ménages
- la base SIMUFORM.sd2 rassemble les imputations liées aux indemnités de formation
- la base SIMUINV.sd2 rassemble les imputations liées à l'invalidité
- la base SIMURMI.sd2 rassemble les imputations du RMI
- la base SIMURET.sd2 rassemble les imputations sur les retraites
- la base SIMUCHO.sd2 rassemble les imputations concernant les allocations chômage .

3) Pour la vague 3, les bases d'imputation se limiteront à

- la base SIMFAMV2.sd2 rassemble les imputations liée à la famille
- la base SIMLOGV2.sd2 rassemble les imputations sur l' allocation logement

Liste des tables SAS du CD-ROM du 31/12/ 1999 (en italiques, bases dont le contenu est désormais inclus dans les tables REVENUVx, et qui ne seront donc plus diffusées par la suite).

	Vague 1	Vague 2	Vague 3
Base générale :	MENAGEV1.sd2 (y c. pondération transversale) INDIVIV1.sd2 REVENUV1.sd2	MENAGEV2.sd2 (y c. pondération transversale 'transv2c') INDIVIV2.sd2 (y c. pondération longitudinale 'basev2')	MENAGEV3.sd2 (y c. pondération transversale 'transv3') INDIVIV3.sd2 (y c. pondération longitudinale 'basev3')

	ACTIVIV1.sd2	REVENUV2.sd2 ACTIVIV2.sd2	REVENUV3.sd2 ACTIVIV3.sd2
Imputations :	<i>RACTV1S.sd2</i> <i>RPATV1S.sd2</i>	<i>REVACV2S.sd2</i> <i>RPLAFIV2.sd2</i> <i>RPATV2S.sd2</i> <i>SIMUFAMI.sd2</i> <i>SIMULOG.sd2</i> <i>SIMUTEM.sd2</i> <i>SIMUFORM.sd2</i> <i>SIMUINV.sd2</i> <i>SIMURMI.sd2</i> <i>SIMURET.sd2</i> <i>SIMUCHOM.sd2</i>	<i>REVACV3S.sd2</i> <i>RPLAFIV3.sd2</i> <i>IMMOV3S.sd2</i> <i>SIMFAMV3.sd2</i> <i>SIMLOGV3.sd2</i>