

CAHIER DE ReCHERCHE

JUIN 1996



N° 91

ESTIMATIONS DE LOIS DE CONSOMMATION
ALIMENTAIRE SUR UN PSEUDO PANEL
D'ENQUÊTES DE L'INSEE
(1979, 1984, 1989)

Nilton CARDOSO
François GARDES

CRÉDOC

L'ENTREPRISE DE RECHERCHE

CRÉDOC

**ESTIMATIONS DE LOIS DE CONSOMMATION ALIMENTAIRE
SUR UN PSEUDO PANEL D'ENQUETES
DE L'INSEE (1979, 1984, 1989)**

Nilton CARDOSO,

Université Paris 1 - Panthéon - Sorbonne, LAMIA.

François GARDES,

Université Paris 1 - Panthéon - Sorbonne, LAMIA, CRÉDOC

Remerciements

Nous remercions l'INSEE de nous avoir permis d'utiliser les données des enquêtes Budget des Familles (BDF), spécialement J.M. Hourriez (INSEE) et D. Eneau (INSEE) pour leur collaboration pendant la phase d'homogénéisation de la base de données lors d'un stage à la division 'Conditions de vie des ménages'. Nous avons profité en outre des suggestions et remarques de P. Balestra, D. Bolduc, F. Legendre, P. Gaubert, M. Goaid, J.-L. Madre, J. Mairesse, C. Monmarquette, D. Richaudeau, J.M. Robin, P. Sevestre, A. Trognon et de deux rapporteurs anonymes, lors des séminaires aux Universités de Paris I, Paris XII, Brown, Cornell, à l'INSEE et lors des 4^{èmes} Journées de Jeunes Economètres à Lille, 11^{èmes} Journées de Microéconomie Appliquées (Aix-Marseille) et du Congrès de l'EEA (1995) à Prague.

SOMMAIRE

INTRODUCTION.....	7
Section I - Spécification, procédures d'estimation et regroupement des données.....	11
Spécification du système complet de demande.....	13
Procédure d'estimation.....	17
Correction d'hétéroscedasticité.....	19
Le regroupement des données individuelles dans un pseudo-panel.....	21
Formation du pseudo-panel français.....	23
Le cellulage 'DGC'.....	24
SECTION II - LOIS DE CONSOMMATION POUR L'ENSEMBLE DE BIENS.....	27
Analyse descriptive des données - ensemble de nomenclature fonctionnelle.....	29
Résultats pour l'ensemble de biens.....	36
SECTION III - LOIS DE CONSOMMATION ALIMENTAIRES.....	45
CONCLUSION.....	59
BIBLIOGRAPHIE.....	61

Annexes	63
Annexe 1 : Le cellulage 'DGC'	65
Annexe 2 : Base de données (fichier harmonisé).....	69
Annexe 3 : Détail des postes alimentaires	75
CON111 - Pain et céréales.....	77
CON112 - Viandes	78
CON113 - Poissons.....	79
CON114 - Lait, fromage et oeufs	79
CON115 - Huiles et graisses	79
CON116 - Fruits et légumes sauf pommes de terre	80
CON117 - Pommes de terre et autres tubercules	80
CON118 - Sucre	80
CON119 - Café, thé.....	81
CON11A - Autres produits alimentaires y compris confiserie.....	81
CON12 - Boissons non alcoolisées	82
Annexe 4 : Encadré 1	83
Annexe 5 : Encadré 2 (Décomposition orthogonale de la variance des observations).....	87

INTRODUCTION

L'estimation de systèmes complets de demande sur données de panel doit permettre de prendre en compte l'hétérogénéité des comportements individuels tout en estimant les lois dans leur dimension statique (en cross-section) aussi bien que dans leur dimension dynamique ou temporelle. L'analyse de panels disponibles en France peut être suppléée par la constitution d'un pseudo-panel formé par le regroupement des données individuelles d'enquêtes comparables. Nous utilisons dans ce rapport les enquêtes de Budgets de Famille (BDF) de l'INSEE de 1979, 1984 et 1989, dont nous groupons les données individuelles en cellules aussi homogènes que possible à l'aide de variables qui indiquent des caractéristiques permanentes des ménages, telles la cohorte d'âge du chef du ménage (ou de la personne de référence de la famille) ou la localisation géographique et le niveau d'éducation. Ces données groupées sont alors supposées indiquer les consommations et le revenu en ces trois dates d'individus synthétiques caractérisés par les variables discriminantes utilisées pour le groupement.

Cette méthodologie, proposée formellement par Deaton (1985, mais en fait utilisée implicitement dans nombre de travaux antérieurs effectués sur données groupées), interdit que les variables discriminantes du groupement changent avec le temps : l'âge du chef de famille, par exemple, pourrait être proposé dans le but de contrôler les effets de cycle de vie, mais alors un ménage appartenant à une cellule particulière en 1979 ne pourrait se retrouver dans la

même cellule en 1984 et 1989 (si les tranches d'âges sont de moins de cinq ans). Il en est de même de caractéristiques socio-économiques, telles que le revenu par unité de consommation ou la taille de la famille, qui permettraient d'homogénéiser les cellules (ce qui réduirait les erreurs de mesure liées au fait qu'on ne retrouve pas les mêmes ménages dans une même cellule à deux dates différentes), mais qui interdirait une interprétation des données groupées en terme de panels, puisque par définition on ne retrouverait pas les mêmes ménages dans les mêmes cellules à deux dates différentes.

Nous cherchons à répondre dans ce rapport à la question suivante : les lois de consommation estimées sur séries temporelles sont-elles semblables à celles que l'on peut estimer sur données d'enquêtes ? En particulier, peut on assimiler les élasticités-revenu statiques (i.e. estimées sur données en coupe) des différentes consommations fonctionnelles (qui indiquent les différences des structures de consommations entre des ménages situés à des niveaux différents de revenu relatif en une certaine période) aux élasticités-revenu estimées en séries temporelles, qui indiquent l'effet d'une modification du revenu des ménages au cours du temps (éventuellement sans changement de leur position relative dans la distribution des revenus) sur leurs consommations. Si ces deux types d'élasticité sont égales l'évolution des revenus du ménage au cours de son cycle de vie rapprochera ses consommations de celles des ménages qui sont plus riches que lui à l'origine : on pourra donc parler d'une certaine convergence des structures de consommation au cours du cycle de vie. Si les deux types d'élasticité diffèrent, la croissance des revenus d'un ménage au cours de son cycle de vie ne lui permettra pas d'accéder au mode de consommation des ménages plus riches, et l'on pourra parler d'une différenciation sociale inévitable des modes de consommation de ménages ayant des revenus relatifs différents.

C'est plutôt ce dernier résultat qui ressort des analyses que nous avons effectuées à partir de cinq enquêtes canadiennes (Gardes, Langlois, Richaudeau, 1996). Nous reprenons ici

cette analyse en construisant un pseudo-panel plus conforme à la méthodologie de Deaton, en contrôlant plus complètement les effets des variables socio-économiques essentielles.

On présente les données d'enquête et la spécification retenue dans une première section; ainsi que le mode de regroupement des données. Les résultats obtenus sur l'ensemble de la nomenclature fonctionnelle sont discutés dans la deuxième section, puis les résultats relatifs aux consommations alimentaires désagrégées dans la troisième section.

**SECTION I - SPÉCIFICATION, PROCÉDURES D'ESTIMATION ET
REGROUPEMENT DES DONNÉES**

SPÉCIFICATION DU SYSTÈME COMPLET DE DEMANDE

Deaton et Muellbauer (1980) ont proposé une classe spécifique de préférences qui permet une agrégation exacte des consommateurs. Ces préférences sont représentées par une fonction de coût ou fonction de dépense connues sous le nom de la classe PIGLOG qui définit le minimum de dépense nécessaire pour atteindre un niveau spécifique d'utilité à des prix donnés. La classe PIGLOG peut être défini comme :

$$\log c(u, p) = (1-u) \log \{a(p)\} + u \log \{b(p)\}$$

où $c(u, p)$ est la fonction coût ou dépense pour l'utilité u , et p le vecteur de prix. L'indicateur du niveau de satisfaction u peut varier entre 0 et 1 de sorte que les fonctions linéaires $a(p)$ et $b(p)$ peuvent être interprétées, respectivement, comme le coût du niveau de vie de subsistance et le coût du niveau de satisfaction optimale.

Il s'agit alors de choisir une forme spécifique aux fonctions de coût $a(p)$ et $b(p)$.

Les auteurs proposent de les définir comme :

$$\log a(p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log p_k \log p_j$$

$$\log b(p) = \log a(p) + \beta_0 \prod_k p_k^{\beta_k}$$

de façon à ce que la fonction de coût du système de demande 'almost ideal' (AI) s'écrive

$$\log c(u, p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log p_k \log p_j + u \beta_0 \prod_k p_k^{\beta_k}$$

où α_i , β_j et γ_{ij} sont des paramètres.

Les fonctions de demande peuvent être dérivées directement à partir de cette dernière équation en appliquant le lemme de Shephard. Après quelques transformations les demandes sont présentées de façon à ce que les coefficients budgétaires (w_i) de chaque bien soient liés aux logarithmes des prix et au revenu réel disponible (ici le prix P est l'indice de prix de Stone) :

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log\left(\frac{y}{P}\right)$$

A partir des estimations du modèle on déduit la valeur suivante des élasticités revenu :

$$e_i = 1 + \frac{\beta_i}{w_i}$$

Les restrictions sur les paramètres de ces fonctions sont :

$$\begin{aligned} \sum_{i=1}^n \alpha_i &= 1 & \sum_{i=1}^n \gamma_{ij} &= 0 & \sum_{i=1}^n \beta_i &= 0 \\ \sum_j \gamma_{ij} &= 0 \\ \gamma_{ij} &= \gamma_{ji} \end{aligned}$$

Lorsque ces restrictions sont satisfaites, le système de demande est composé de fonctions qui satisfont à la fois à la propriété d'additivité ($\sum w_i = 1$), d'homogénéité de degré zéro par rapport aux prix et à la dépense totale, et de symétrie de Slutsky.

L'interprétation du système de demande AI indiqué par les auteurs est simple. En absence de changements des prix relatifs et de la dépense totale, les coefficients budgétaires sont constants. Les changements sur les prix relatifs (sous hypothèse d'homogénéité) se reflètent sur les termes γ_{ij} ; chaque γ_{ij} représente l'effet sur le $i^{\text{ème}}$ coefficient budgétaire d'une augmentation relative du $j^{\text{ème}}$ prix, la dépense réelle étant maintenue constante. Les changements sur la dépense réelle s'opèrent à travers les coefficients β_i ; ils sont positifs en

cas de bien de luxe, négatifs pour les biens nécessaires et leur somme pour tous les biens est égale à zéro.

En résumé, le modèle AI a les avantages d'être assez flexible, de satisfaire les axiomes de choix exact et d'agrégation des consommateurs sans qu'il ne soit nécessaire d'imposer le parallélisme de courbes linéaires d'Engel, d'avoir une forme fonctionnelle compatible avec les données connues des budgets des ménages, d'être simple à estimer en évitant des non-linéarités et de permettre de tester aisément les restrictions d'homogénéité et de symétrie à travers des contraintes linéaires sur les paramètres.

Banks et al. (1992) ont proposé une généralisation du système de demande AI (QAI) par l'introduction d'un terme carré du revenu logarithmique pour mieux tenir compte de la non-linéarité des courbes d'Engel. L'élasticité-revenu (e^q) pour le bien i s'écrit alors :

$$e_i^q = 1 + \frac{\beta_1}{w_i} + 2 \frac{\beta_2}{w_i} \log y$$

où β_1 est le coefficient lié au revenu logarithmique et β_2 celui lié au carré du revenu logarithmique.

Dans notre cas, le modèle AI et son homologue quadratique, au niveau du ménage peuvent s'écrire ainsi:

$$\frac{c_{ht}^i}{y_{ht}^*} = w_{ht}^i = a_i + \sum_{m=1}^2 \beta_{im} (\log(y_{ht}/P_t))^m + S_{ht} c_i + \zeta d + u_{ht}^i \quad (1)$$

pour le ménage h et le bien i, au temps t,

m= 1 ou 2,

avec C_{ht}^i = dépenses du ménage h au temps t pour le bien i,

y_{ht}^* = dépenses totales du ménage h au temps t,

y_{ht} = dépenses totales du ménage h au temps t, par unité de consommation
(échelle de l'Insee),

P = indice de prix de Stone¹,

S_{ht} = caractéristiques socio-économiques,

ζ = vecteur des variables indicatrices des enquêtes (deux dans notre cas),

α = terme constant.

On notera qu'on tient compte dans nos estimations des effets-prix et de l'ensemble des variables macroéconomiques communes à tous les ménages par des variables muettes indiquant chaque période. L'absence des prix relatifs dans cette spécification est justifiée par leur faible variabilité sur l'échantillon (pour N cellules, cet échantillon est de taille 3N, les prix relatifs ne variant que deux fois puisqu'ils sont identiques pour tous les ménages d'une enquête : on ne pourrait donc prendre en compte au mieux que des effets-prix directs).

¹ Cet indice de prix s'écrit : $\ln p_t = \sum_{i=1}^n \bar{w}_i \ln p_{it}$, avec \bar{w}_i étant les coefficients budgétaires moyens des n biens. Pondérer par des coefficients budgétaires variables pose des problèmes d'endogénéité à cet indice de prix.

PROCÉDURE D'ESTIMATION

On fait l'hypothèse que les erreurs sont homoscédastiques au niveau individuel, d'espérance nulle et que la covariance est égale à zéro entre les erreurs pour deux ménages différents au temps t ou pour le même ménage en deux instants différents. C'est-à-dire :

$$E(u_{ht}) = 0.$$

$$V(u_{ht}) = \sigma^2, \forall h, \forall t.$$

$$\text{cov}(u_{ht}, u_{h't'}) = \text{cov}(u_{ht}, u_{h't'}) = 0, \forall h \neq h', \forall t \neq t'.$$

On a vu que pour travailler en pseudo-panel on a classé les différents ménages en catégories de façon à pouvoir suivre le ménage 'représentatif' de ces cellules. Le coefficient budgétaire d'une cellule pour le bien i est égal au rapport entre la somme de la dépense en bien i de tous les ménages de la cellule, et la somme des dépenses totales de ces mêmes ménages. Il correspond au coefficient budgétaire pour le bien i du ménage représentatif de la cellule :

$$w_{Ht}^i = \sum_{h \in Ht} C_{ht}^i / \sum_{h \in Ht} Y_{ht}^* \quad (2)$$

où H indique une cellule.

Un facteur de pondération doit être utilisé lors de l'agrégation pour calculer le coefficient budgétaire d'une cellule H à partir des coefficients budgétaires des ménages appartenant à cette cellule. Ce facteur appelé γ est le rapport entre la dépense totale du ménage et la somme des dépenses totales des ménages d'une même cellule. Donc,

$$\gamma_{ht} = \frac{\dot{y}_{ht}}{\sum_{h \in H} \dot{y}_{ht}} = \frac{\dot{y}_{ht}}{\bar{\dot{y}}_{Ht} \cdot n_H} \quad (3)$$

où la barre indique la moyenne arithmétique et n_H le nombre des ménages d'une cellule.

$$w_{Ht}^i = \sum_{h \in Ht} (\gamma_{ht} \cdot w_{ht}^i) \quad (4)$$

Si on substitue (3) en (4) on retrouve le même résultat qu'en (2).

Le modèle au niveau de la cellule s'écrit ainsi :

$$w_{Ht}^i = a_i + \sum_{m=1}^2 \beta_{im} \left(\sum_{h \in Ht} \gamma_{ht} (\log(y_{ht}/P_t))^m \right) + c_i \left(\sum_{Ht} \gamma_{ht} S_{ht} \right) + d \left(\sum_{Ht} \gamma_{ht} \zeta \right) + u_{Ht}^i \quad (5)^2$$

²Si on veut développer le terme en logarithme de cette équation, on trouve:

$$\sum_{m=1}^2 \beta_{im} \left(\sum_{h \in Ht} \left[y_{ht}^* (\log(y_{ht}/P))^m \right] / \sum_{h \in Ht} y_{ht}^* \right)$$

CORRECTION D'HÉTÉROSCÉDASTICITÉ

Le groupement des données rend le résidu u_{ht}^i hétéroscédastique, puisque la variance du modèle au niveau de la cellule est égale à $\sigma^2 \cdot (\sum_{Ht} (\gamma_{ht})^2)$. Pour rendre le modèle homoscédastique, on a pré-multiplié toutes les variables par un facteur de correction égal à :

$$(1/\sqrt{(\sum_{Ht} (\gamma_{ht})^2)}).$$

On considère que le terme d'erreur comprend un effet spécifique au ménages et un bruit blanc : $u_{ht} = \alpha_h + \varepsilon_{ht}$. Le modèle est dit à effets fixes ou modèle de la covariance si l'effet spécifique est considéré comme certain, et appelé modèle à erreurs composées si cet effet est supposé aléatoire³.

Le résidu étant composé en un effet spécifique et un effet résiduel : on applique les transformations between et within usuelles en panel à ces données transformées pour estimer les paramètres dans les deux dimensions transversale et longitudinale. La décomposition spectrale ainsi opérée conserve donc toutes les propriétés usuelles de l'analyse des panels (décrite par exemple dans l'ouvrage de Balatgi, pp.13-18). On notera qu'on utilise ici un modèle à composition d'erreur simple (sans composante spécifique temporelle) puisque les effets temporels sont estimés comme des effets fixes par les variables muettes indiquant l'enquête (nous suivons en cela la suggestion de Dormont, 1989).

Il aurait été possible d'effectuer la correction d'hétéroscédasticité et la décomposition spectrale dans l'ordre inverse, en transformant d'abord les variables du modèle (5) par les opérateurs within et between, puis en corrigeant les hétéroscédasticités propres aux deux

³ On reviendra plus tard sur les conséquences de ces deux spécifications sur les estimateurs calculés.

dimensions. On peut prouver alors que le facteur d'hétéroscédasticité en dimension within s'écrit alors⁴ :

$$\text{Var } W u_{Ht}^i = (\sigma_\varepsilon^i)^2 \left[\left(\frac{T-1}{T} \right)^2 \sum_{h \in Ht} \gamma_{ht}^2 + \frac{1}{T^2} \sum_{h \in Ht', t' \neq t} \gamma_{ht'}^2 \right]$$

avec $\text{cov}(Wu_{Ht}, Wu_{Ht'}) \neq 0$ pour $H = H'$; $t \neq t'$

= 0 pour $H \neq H'$.

et la variance de la dimension between :

$$\text{Var}(B\omega_{Ht}^i) = \sigma_{\alpha^i}^2 + \frac{\sigma_{\varepsilon^i}^2}{T^2} \sum_{h \in Ht} \gamma_{ht}^2$$

avec $\text{cov}(Bu_{Ht}, Bu_{Ht'}) = 0$ pour $H \neq H'$.

Ces deux types d'hétéroscédasticité ne peuvent être corrigés simplement par une transformation des données ni même par MCG : ils exigent une procédure itérative puisqu'on doit estimer d'abord le partage de la variance de l'erreur dans les dimensions within et between. Par ailleurs, les erreurs résultantes dans ces deux dimensions ne sont plus nécessairement orthogonales, ce qui complique la comparaison des estimateurs between et within par les tests usuels (d'existence d'un effet spécifique et d'Hausman). On a donc préféré la première procédure consistant à corriger d'abord l'hétéroscédasticité avant d'effectuer la décomposition spectrale.

⁴ M. Gurgand nous a signalé une erreur faite dans la détermination de cette hétéroscédasticité dans une précédente version du texte.

LE REGROUPEMENT DES DONNÉES INDIVIDUELLES DANS UN PSEUDO-PANEL

La technique d'estimation en pseudo panel consiste à estimer des modèles de panel à partir des données transversales indépendantes. Les panels de consommation sont rares et souvent partiels : le Panel Study of Income Dynamics ne contient par exemple que deux postes de consommation alimentaire (estimée par le ménage enquêté, donc probablement surestimée) et le revenu du ménage (connu sans doute avec des erreurs de mesure). Par ailleurs, ils sont soumis à un biais d'attrition et à l'effet d'apprentissage pour les panels longs, et ne contiennent généralement qu'un nombre trop limité de ménages pour qu'on puisse estimer des lois de consommation des sous-populations.

Le choix d'un mode de regroupement des données individuelles est malaisé, dans la mesure où deux principes s'opposent : les cellules doivent contenir un nombre suffisant d'individus pour que les moyennes empiriques des diverses variables constituent un bon estimateur des moyennes théoriques du type d'individus regroupés dans la cellule (Verbeek-Nijman, 1992, montrent que le problème d'erreur de mesure est négligeable lorsque les cellules contiennent plus d'une centaine d'individus), mais, inversement, la taille des cellules diminue, à population enquêtée constante, lorsqu'on multiplie les critères de regroupement pour homogénéiser le plus possible chaque cellule : en augmentant la taille des cellules, on réduit donc la similarité des individus regroupés et l'on augmente la dispersion des valeurs moyennes des variables explicatives et expliquée.

Le premier principe, de regroupement d'individus par catégories les plus homogènes possibles, avec des cellules aussi distinctes que possible des autres, permet de minimiser la perte d'efficacité des estimateurs calculés sur données groupées par rapport à des estimateurs calculés sur données individuelles (encadré 1). Le second principe, de regroupement en des cellules contenant suffisamment d'individus statistiques, permet de réduire les erreurs de

mesure provenant du fait que l'on ne re-interroge pas les mêmes individus à deux dates différentes.

Dans le cadre d'un modèle de régression simple et en supposant vérifiée l'hypothèse de Mundlak, stipulant que les effets spécifiques sont corrélés aux variables explicatives Verbeek-Nijman (1993) (équation 15) montrent que le biais asymptotique de l'estimateur within augmente avec l'erreur de mesure sur la variable explicative, et diminue avec la taille des cellules et la variance within (calculée entre cellules et en moyenne sur toutes les périodes) des moyennes exactes de la variable explicative : l'élargissement des cellules diminue donc ce biais asymptotique pour ces deux derniers déterminants, mais peut accroître en augmentant l'erreur de mesure si les cellules deviennent trop hétérogènes et si les observations individuelles composant ces cellules à diverses périodes risquent d'être plus dispersées que pour des cellules homogènes contenant un type précis d'individu statistique. Les résultats de simulation qu'ils présentent dans leur tables 2 et 3 montrent que la variance asymptotique des estimateurs within diminue fortement lorsque la taille des cellules passe de 10 à 50 individus (division par 4 à 10), puis plus lentement lorsqu'elles atteignent 200 individus (division par 1,5 à 3). Pour des cellules de 50 individus, l'estimateur within n'est pas amélioré par la correction des erreurs de mesure proposée par Deaton (1985), pour des cellules de 100 à 200 individus, la variance asymptotique obtenue par une correction exacte des erreurs de mesure (selon la méthode proposée par ces auteurs), est à peu près égale à celle de l'estimateur within.

Pour tenir compte de l'ensemble de ces résultats (très partiel dans l'état actuel de la littérature), on a choisi homogénéiser les cellules (dans le but de pouvoir dégager des effets spécifiques aux types de ménages correspondant à chaque cellule) en les définissant par trois variables discriminantes (et non par le seul critère de la génération), puis de regrouper les cellules trop petites avec des cellules adjacentes pour réduire la proportion de cellules contenant moins de cinquante individus.

FORMATION DU PSEUDO-PANEL FRANÇAIS

Le critère de cellulaire usuellement rencontré dans la littérature est la simple utilisation des cohortes d'âge. L'idée est de faire des moyennes sur des cohortes d'individus pour retrouver l'individu 'représentatif' de cette cohorte. Ainsi, sur ces données regroupées selon un critère invariant dans le temps nous pouvons faire sortir des effets spécifiques aux individus 'représentatifs'. Pour la formation des pseudo-panels français, nous avons élargi cette procédure classique de regroupement des ménages.

D'une part, pour le cellulaire que nous appelons 'DGC' (diplôme, localisation géographique, cohorte d'âge), nous n'avons appliqué que des critères 'a priori' invariants dans le temps. L'application simultanée de plusieurs critères a pour objet principal la création d'un plus grand nombre des cellules plus homogènes. Dans une autre recherche (Cardoso-Gardes), nous avons appliqué des critères variables dans le temps, le but étant d'homogénéiser le plus possible les cellules selon les classes sociales. L'intérêt de ce second cellulaire est de faire apparaître un effet spécifique aux classes sociales, plutôt qu'un effet spécifique au ménage, mais il ne satisfait pas au critère usuel d'invariance des critères dans le temps.

LE CELLULAGE 'DGC'

La procédure de cellulage DGC a été la suivante : soit n critères de cellulage a priori invariants dans le temps (exemple : 3 variables discriminantes : A, B et C, chacune avec un certain nombre de modalités). Les trois variables choisies ont été :

- *A : localisation géographique du ménage (variable GEO), avec deux modalités : 1 si le ménage habite en région parisienne, 0 sinon.
- *B : Cohorte d'âge de la personne de référence au ménage (variable TRAGE), avec 11 modalités.
- *C : diplôme (variable DIPL), avec les modalités : 1) pas de certificat, 2) certificat d'études primaires, 3) BEP, CAP, BEPC ou niveau équivalent, 4) BAC ou BAC première partie et brevet supérieur, 5) BAC + 2, 6) BAC + 3, 4 ou 5.

Le produit cartésien de ces variables donne :

$$A \times B \times C = \{ (a, b, c) \mid a \in A, b \in B, c \in C \}.$$

On définit la fonction : $f : N^3 \rightarrow N$

$$\begin{aligned} \text{tel que } n = f(A, B, C) &= 1 \text{ si } k_t \in \{(a, b, c)\} \quad k_t = 1, \dots, K_t \quad \text{où } K_t = \text{nb. d'obs. enq.} \\ &= 0 \text{ sinon.} \end{aligned}$$

Ensuite, on définit $f : N^3 \rightarrow N$

$$N = g(a, b, c) = \sum n.$$

Avec cette notation, g représente évidemment le nombre d'effectifs d'une cellule.

- Si $N \geq 100 \forall t$, alors (a, b, c) constitue une cellule ordinaire.
- Si $N < 100$ au moins pour un t , alors il faut essayer d'élargir la cellule dont l'effectif est faible.

L'élargissement d'une cellule (a,b,c) dont l'effectif est faible doit obéir à une logique d'agrégation raisonnée. En fait, il est normal que l'on distingue différents degrés d'importance entre les variables discriminantes dans une procédure de cellulaire. Supposons, par exemple, que le critère A soit plus discriminant que B et C, de même pour B par rapport à C. L'élargissement d'une cellule 'faible' (a,b,c) s'effectuera par l'union avec une autre cellule adjacente en c, soit par exemple (a,b,c') , ce qui revient à former une nouvelle cellule (a,b,c^*) . Ensuite, le calcul de l'effectif doit être refait pour cette nouvelle cellule (a,b,c^*) :

$$g(a, b, c^*) = g(a, b, c) + g(a, b, c')$$

- Si $g(a, b, c^*) \geq 100 \forall t$, alors (a,b,c^*) peut être conservée comme une cellule agrégée.
- Si $g(a, b, c^*) < 100$ au moins pour un t , l'opération d'élargissement doit être renouvelée.

Il est clair qu'il y a un arbitrage entre taille de cellules et nombre total de cellules, c'est-à-dire entre variance intra-individuelle et variance inter-individuelle.

Dans le cas présent, si nous renouvelions la procédure d'élargissement des cellules à chaque fois que $N < 100$ pour un t , nous terminerions la procédure avec un nombre très réduit de cellules et certains degrés de discrimination auraient été automatiquement abolis. Ainsi, nous avons parfois préféré avoir des cellules avec un effectif plus petit que 100, pour obtenir le cellulaire décrit succinctement dans l'annexe 1.

**SECTION II - LOIS DE CONSOMMATION POUR
L'ENSEMBLE DE BIENS**

ANALYSE DESCRIPTIVE DES DONNÉES - ENSEMBLE DE NOMENCLATURE FONCTIONNELLE

On trouvera les principes d'harmonisation de données de trois enquêtes de Budget de Familles de l'INSEE (1979, 1984, 1989) en annexe 2.

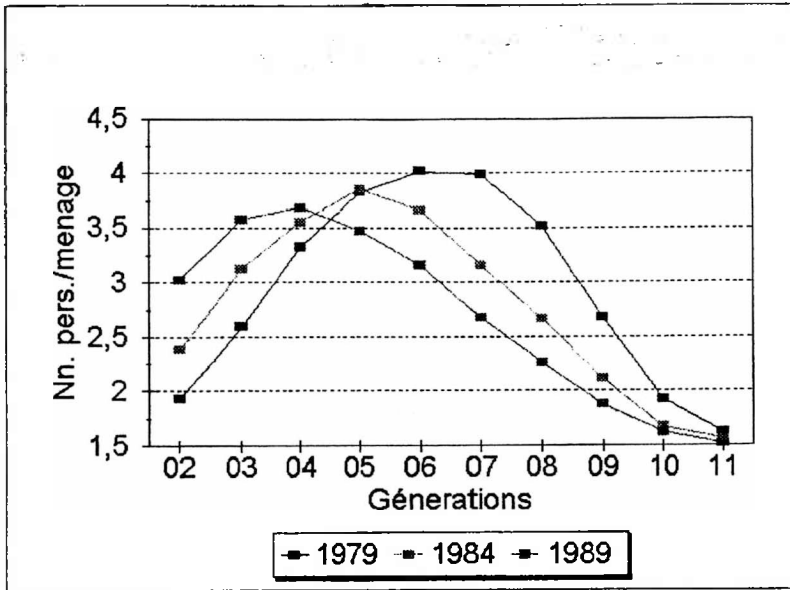
La comparaison des moyennes de variables entre les trois années d'enquête fait apparaître plusieurs traits démographiques et économiques. Le tableau 2 présente le nombre initial des ménages dans chaque enquête, et ensuite le nombre de ménages après filtrage (selon le critère de fiabilité du revenu déclaré) et, finalement, le nombre de ménages en conservant les mêmes cellules pour les calculs de régressions.

Tableau 1 : Nombre de ménages

Version BDF	Initial	Après filtrage	Avec les mêmes cellules
1979	10.645	9.313	9.313
1984	11.976	10.746	10.746
1989	9.038	8.462	7.722

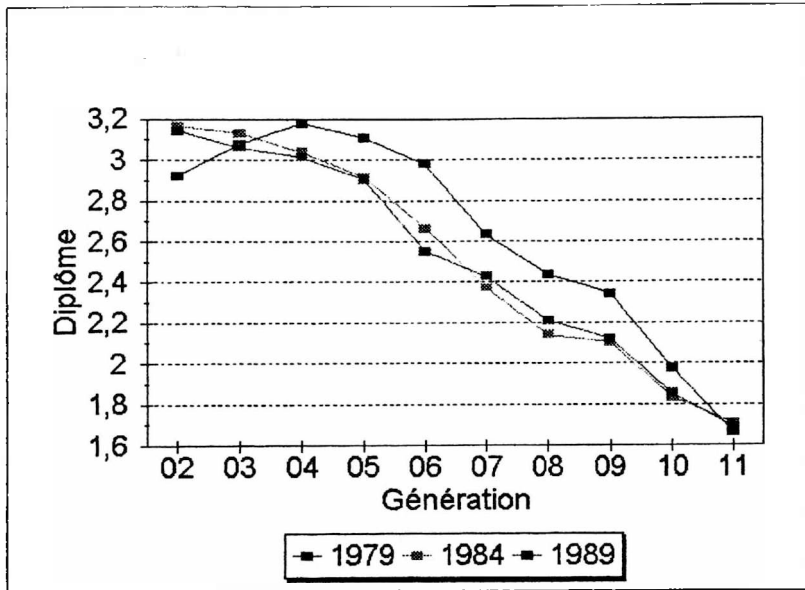
Le nombre moyen de personnes composant un ménage (pour le fichier des ménages dont le revenu est considéré comme fiable) varie de 3,05 en 1979 à 2,76 en 1984 et à 2,72 en 1989. Le graphique 1 montre que cette diminution concerne toute les générations. On constate que les trois courbes du graphique 2 atteignent leur maximum toujours au point où les personnes de référence sont âgées entre 37 et 42 ans. Ces maxima sont progressivement plus petits au fur et à mesure qu'on avance dans le temps (ils passent de 4,01 en 1979 à 3,68 en 1989). Ceci montre la rapidité des changements démographiques durant cette période, et l'intérêt de les contrôler dans les estimations.

Figure 1 : Nombre de personnes par ménage



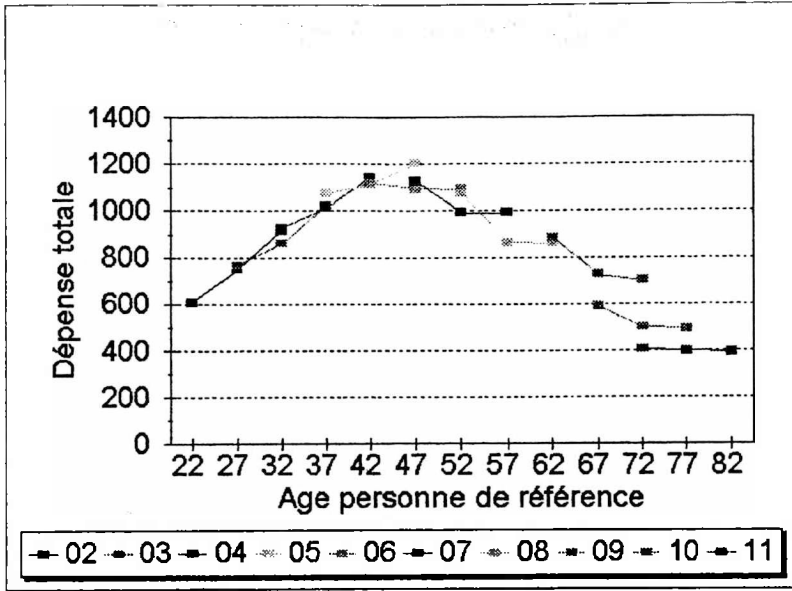
L'éducation des individus dans l'enquête est mesurée par la variable diplôme créée lors d'homogénéisation des enquêtes. Le graphique 2 montre l'existence d'une relation inverse entre génération et diplôme.

Figure 2 : Relation entre la génération et le diplôme



Comme on part de l'hypothèse qu'une fois le diplôme acquis, l'individu le garde pour toujours, on s'attend à ce que les courbes pour les différentes années soient pratiquement superposées. Cela est précisément ce qui arrive pour les enquêtes 1984 et 1989 qui ont une méthodologie très similaire. Par contre, le petit écart entre celle de 1979 et les deux autres ne peut être expliqué que par une différence de classification des réponses lors des interviews.

Figure 3 : Relation entre la dépense totale et le cycle de vie

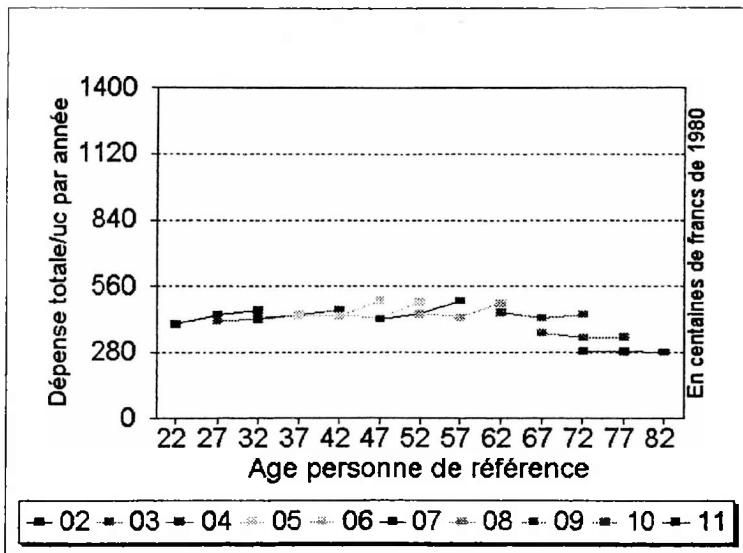


Attanasio et Browning (1996) ont utilisé la technique de pseudo-panel pour tester l'hypothèse de cycle de vie selon laquelle l'utilité marginale de la monnaie est constante dans le temps (la conséquence étant que les ménages lissent leur consommation face à des changements de profil de revenu). Ils ont réalisé deux types de cellulage un basé seulement sur le critère d'année de naissance de la personne de référence (cohortes) et un autre croisant cette variable discriminante avec la vague d'enquête dans laquelle le ménage a été interviewé.

En régressant le logarithme de la consommation et le logarithme du revenu net sur des variables muettes indiquant les cohortes d'années et les trimestres, ils obtiennent deux courbes en forme de U inversée⁵. Dans notre cas, la disposition de la dépense totale contre l'âge de la personne de référence, comme dans le graphique 4, donne déjà ce format familier d'un U inversé.

Jusqu'à la génération '05' il y a augmentation de la dépense totale en francs constants d'une enquête à l'autre. Les générations '06' et '11' gardent à peu près le même niveau de dépense au fil du temps. Enfin, de la génération '07' à la '10' incluses, les dépenses totales diminuent d'une année à l'autre.

Figure 4 : Relation entre la dépense totale par UC et le cycle de vie



⁵ Ils notent que deux autres études qui portent sur le même sujet trouvent des profils similaires pour ces courbes, mais que Carrol et Summers (1990) interprètent ces résultats comme une manque d'évidence de lissage de consommation tandis que Ghez et Becker (1975) interprètent cela comme étant plutôt une évidence de la validité du modèle cycle de vie.

Il est clair que si on ne contrôle pas ce type d'analyse par l'introduction des variables démographiques (par exemple, la taille de la famille), on a naturellement une augmentation (ou une baisse) de la dépense totale d'un ménage lorsqu'il y a la présence (ou le départ) des enfants. On constate en effet dans le graphique 5 que la dépense totale par unité de consommation n'a plus qu'une forme en U très atténuée. On peut donc conclure qu'une fois la taille de la famille contrôlée par le nombre d'unités de consommation⁶, l'évolution de la dépense totale est assez stable dans le temps et ne peut être expliquée que par des effets autres que le cycle de vie, selon Attanasio-Weber : des effets de revenu, des effets de cohorte, des effets cycliques et d'hétérogénéité des ménages.

Les premiers effets, dans ce contexte, sont liés au niveau, à l'évolution et au risque du flux de revenu. Les effets de cohorte correspondent à l'idée que les changements sociaux d'une génération à l'autre peuvent avoir un effet sur leur comportement envers le risque, les préférences et les facteurs d'escompte dans le choix de consommation. Les effets cycliques sont ceux communs à tous les agents qui vivent les mêmes cycles économiques. L'hétérogénéité existe forcément dans ce type de données puisqu'aucune famille n'est identique à une autre sur toutes les caractéristiques observables. L'ensemble de ces variables explicatives et les effets spécifiques aux cellules permettent de prendre correctement en compte les effets de cohorte et les effets cycliques et partiellement les effets de revenu (par le terme quadratique) et l'hétérogénéité des ménages.

Les statistiques descriptives des données BDF, présentées dans plusieurs ouvrages de l'INSEE, diffèrent légèrement de nos statistiques descriptives. Cela est dû non seulement à des considérations de filtrage, mais aussi parce que les postes de dépense que nous utilisons ont été organisés d'une façon distincte de ceux originels des enquêtes. Par exemple, nous n'avons

⁶ Le nombre d'unités de consommation d'un ménage est la somme des coefficients suivants: 1 pour la personne de référence, 0,7 pour chaque autre personne du ménage âgée de 14 ans ou plus, et 0,5 pour ceux de moins de 14 ans.

pas tenu compte du changement de la fonction 'transports' en 1989 pour inclure les dépenses d'assurance automobile puisque dans notre cas ces dépenses sont considérées comme faisant partie du groupe 'postes hors champ comptabilité nationale'. De même, si avant 1989, la fonction hygiène-santé comprenait les dépenses de coiffure et de produits de toilette et de parfumerie, dans notre cas, ces mêmes dépenses ne sont jamais comptées sous la rubrique santé, et font partie des services divers. Egalement, dans notre nomenclature à 9 postes, l'alimentation ne prend jamais en compte les repas pris à l'extérieur, qui pour nous sont inclus sous la rubrique services divers.

RÉSULTATS POUR L'ENSEMBLE DE BIENS

Les variables explicatives de la consommation, outre la dépense totale et une constante, comprennent deux variables muettes indiquant l'année des enquêtes pour prendre en compte les effets-prix, supposés identiques pour tous les ménages d'une enquête, et toutes les influences macroéconomiques liées à la période, cinq variables muettes indiquant la structure démographique du ménage (pour corriger l'éventuel biais lié au choix d'une échelle de consommation particulière), sept variables muettes pour la catégorie socio-professionnelle de la personne de référence au ménage, l'âge de la personne de référence au ménage, une variable muette pour la propriété du logement principal et une autre pour la propriété d'une résidence secondaire, une variable muette pour indiquer si la personne de référence au ménage et le conjoint étaient actifs ou non lors de l'enquête.

L'inclusion de toutes ces variables entraîne une forte multicolinéarité aggravée par la présence du facteur correcteur d'hétéroscédasticité dans le modèle. De plus, la précision des estimations était pénalisée par une perte non négligeable des degrés de liberté.

Des tests de Fisher sur les modèles linéaires quadratiques montrent que les variables indiquant la PCS des ménages ne sont pas globalement significatives pour sept postes sur quinze, mais, de plus, on constate une très forte colinéarité des PCS avec la dépense totale ($R^2 = 0,9992$). On a donc choisi de les exclure. La propriété du logement s'avère significative à la différence de la propriété d'une résidence secondaire, qui est donc exclue du système d'équations. Toutes les autres variables explicatives ont été conservées comme nécessaires à une bonne spécification du modèle.

Tableau 2 : Estimations sur données empilées.

Modèle	AI	AI	AI	AI	QAI	QAI	QAI	QAI
	Coupe empilée *	Between	Within	MCG	Coupe empilée *	Between	Within	MCG
BIENS	élasticité - revenu (t)	élasticité - revenu (t)	élasticité - revenu (t)	élasticité - revenu (t)	élasticité - revenu (t)	élasticité - revenu (t)	élasticité - revenu (t)	élasticité - revenu (t)
1. Alimentation à l'extérieur	1,580 (29,07)	1,527 (16,69)	1,306 (8,55)	1,569 (27,16)	1,684 (23,44)	1,659 (13,62)	1,474 (8,12)	1,672 (22,56)
2. Alimentation à domicile	0,313 (10,80)	0,374 (6,79)	0,370 (5,09)	0,295 (9,07)	0,197 (5,43)	0,310 (4,18)	0,182 (2,26)	0,167 (4,36)
3. Alcools et tabac	0,605 (8,18)	0,521 (4,85)	0,819 (3,54)	0,606 (8,36)	0,448 (4,60)	0,487 (3,30)	0,556 (2,03)	0,451 (4,66)
4. Habillement	1,153 (24,29)	1,238 (14,14)	1,117 (7,92)	1,148 (22,40)	1,206 (19,04)	1,305 (10,95)	1,277 (7,63)	1,199 (18,04)
5. Logement	1,381 (23,75)	1,248 (11,68)	0,786 (5,26)	1,348 (20,80)	1,373 (17,59)	1,093 (7,69)	0,879 (4,90)	1,371 (16,66)
6. Électricité	0,233 (4,67)	0,103 (1,48)	0,384 (2,52)	0,229 (4,71)	0,172 (2,58)	0,007 (0,08)	0,230 (1,27)	0,164 (2,51)
7. Articles de ménage	1,164 (20,95)	1,131 (11,99)	1,574 (10,04)	1,192 (20,08)	1,122 (15,08)	0,971 (7,82)	1,483 (7,89)	1,148 (14,96)
8. Santé	0,740 (8,08)	0,678 (4,06)	1,075 (4,21)	0,770 (7,70)	0,821 (6,70)	0,713 (3,12)	1,001 (3,26)	0,823 (6,37)
9. Achat de véhicules	1,038 (10,97)	1,251 (7,98)	1,546 (5,44)	1,044 (10,58)	1,277 (10,33)	1,523 (7,40)	1,953 (5,85)	1,279 (10,11)
10. Utilisation de véhicules	0,808 (11,46)	1,011 (8,85)	0,999 (5,20)	0,814 (10,84)	1,103 (11,74)	1,607 (10,71)	1,062 (4,60)	0,902 (9,32)
11. Transport public	2,245 (19,22)	2,450 (10,66)	1,877 (5,34)	2,400 (17,24)	2,416 (14,11)	2,397 (7,61)	2,175 (5,19)	2,397 (13,31)
12. Téléph. télécom..	1,365 (23,33)	1,447 (19,47)	0,778 (4,29)	1,379 (25,55)	1,530 (20,15)	1,582 (16,26)	0,951 (4,40)	1,543 (21,75)
13. Loisirs et éducation	1,401 (29,15)	1,353 (19,01)	0,921 (6,55)	1,399 (28,72)	1,479 (23,19)	1,454 (15,32)	1,071 (6,40)	1,479 (23,05)
14. Services	1,436 (22,39)	1,465 (14,12)	1,327 (6,59)	1,433 (21,92)	1,474 (17,15)	1,537 (10,87)	1,394 (5,76)	1,472 (16,92)
15. Divers	1,149 (23,00)	1,037 (11,41)	1,135 (8,91)	1,159 (21,04)	1,047 (15,89)	0,927 (7,59)	0,975 (6,48)	1,060 (15,34)

* Ensemble des trois enquêtes.

Tableau 3 : Tableau des élasticités-revenu avec instrumentation pour la spécification A1

Biens	1	2	3	4	5	6	7
Between	1,574 (16,22)	0,331 (5,53)	0,471 (4,14)	1,242 (13,10)	1,236 (10,64)	0,079 (0,92)	1,119 (10,98)
Within	1,205 (9,57)	0,578 (8,69)	0,879 (4,64)	1,078 (9,33)	0,857 (6,98)	0,587 (4,60)	1,384 (10,56)

Biens	8	9	10	11	12	13	14	15
Between	0,651 (3,64)	1,233 (7,27)	1,006 (8,20)	2,541 (10,21)	1,446 (17,16)	1,379 (18,04)	1,509 (13,67)	1,080 (11,14)
Within	1,050 (5,03)	1,366 (5,83)	0,999 (6,35)	1,587 (5,47)	0,851 (5,72)	0,947 (8,23)	1,219 (7,37)	1,090 (10,44)

Le test du modèle quadratique contre le modèle linéaire n'est significatif que pour six consommations sur quinze. On a néanmoins présenté dans le tableau l'estimation des deux modèles pour ne pas négliger la non-linéarité des lois de consommation de ces six postes.

Tableau 4 : Test effet spécifique (Dormont, 1989) et Test d'Hausman

Modèle variable dépend.	Test effet spécifique (Dormont, 1989)		Test d'Hausman	
	AI	QAI	AI	QAI
	Statistique F(43,99)	Statistique F(44,98)	Statistique χ^2 (13)	Statistique χ^2 (14)
1. Alimentation à l'extérieur	1,50*	1,49	34,81*	35,83*
2. Alimentation à domicile	2,40*	2,80*	37,04*	26,50*
3. Alcools, tabac	0,90	0,95	32,16*	28,91*
4. Habillement	1,61*	1,67*	13,37	15,60
5. Logement	2,13*	2,07*	40,25*	44,21*
6. Électricité	0,87	0,87	42,70*	45,63*
7. Articles ménage	1,52*	1,44	34,19*	38,98*
8. Santé	1,79*	1,83*	26,56*	25,95*
9. Achat véhicules	1,27	1,25	25,18*	26,38*
10. Utilisation véhicules	1,48	1,39	48,84*	51,68*
11. Transports publics	1,79*	1,86*	29,97*	31,43*
12. Télécom.	0,70	0,67	45,48*	37,89*
13. Loisirs, éducation	1,07	1,06	45,93*	47,81*
14. Services	1,11	1,13	19,15	19,31
15. Hors compt. nationale	2,13*	2,17*	35,57*	39,59*

* Statistiques significatives à 5% de confiance.

$F(0,95,43,99)_{critique} = 1,503$ $F(0,95,44,98)_{critique} = 1,501$ $\chi^2(13)_{critique} = 22,36$ $\chi^2(14)_{critique} = 23,68$

L'existence d'effets spécifiques peut être testée de plusieurs manières : pour un test de Fisher entre le modèle en niveau avec des variables muettes indiquant les cellules (on teste alors la présence d'effets fixes certains), ou par la comparaison des variances between et within qui permettent de juger de l'importance des deux composantes de l'erreur (on teste alors la présence d'effets spécifiques aléatoires). Ce dernier test (opéré comme l'indique Dormont,

1989) montre qu'un effet spécifique existe pour neuf postes sur quinze (pour un seuil de confiance de 5%, tableau 4).

On a donc estimé le système complet sous une hypothèse de composition d'erreur pour l'ensemble des équations, par la méthode SURE (qui donne ici les mêmes estimations que les MCO puisque les variables explicatives sont communes à toutes les équations) et sous l'hypothèse d'additivité (il n'y a pas lieu de tenir compte ici des contraintes liées aux effets-prix).

Les élasticités estimées sont toutes très significatives, positives, de valeurs assez dispersées (de 0,3 à 2,4); on note que la transformation within fournit des élasticités très proches de la transformation en différence première, (toutes deux, en éliminant les effets spécifiques, permettent de mesurer l'effet des changements temporels de la dépense totale). Les valeurs des élasticités (en particulier pour l'estimation within) sont proches de celles qu'obtiennent Nichèle-Robin à partir des trois mêmes enquêtes en pooling (les faibles différences peuvent être pour partie attribuées à l'harmonisation des trois enquêtes que nous avons effectuée). On note quelques différences de ces estimations avec celles obtenues, sur les mêmes enquêtes, à partir d'un cellulaire différent (Cardoso-Gardes, 1996).

Les élasticités within et between sont généralement positionnées identiquement pour tous les postes par rapport à 0 et à un : les postes sont donc classés comme des consommations normales ou de luxe par la comparaison, sur une enquête, des consommations et dépense totale de ménages classés différemment sur l'échelle des revenus, ou par l'effet sur les consommations des variations des revenus au cours du temps. Ceci n'est néanmoins pas vérifié pour les dépenses de téléphone et de logement, classées de luxe transversalement mais biens nécessaires longitudinalement (à élasticité légèrement inférieure à un). Cette différence des estimations between et within est en fait assez générale, puisque un tiers des postes seulement a des élasticités comparables dans les deux dimensions (il s'agit de l'alimentation à

domicile, de l'habillement, des services, des transports publics et des dépenses diverses), alors qu'un autre tiers est plus élastique transversalement et le tiers restant plus élastique temporellement.

Cette différence des estimations within et between est jugée fortement significative pour treize postes sur quinze par le test d'Hausman (tableau 4), tant pour le système linéaire AI que pour le système quadratique. La réponse à la question posée au début de cet article est donc positive : pour une même spécification, sur un même ensemble de données, les estimations between diffèrent significativement des estimations en within.

Cette différence peut provenir de diverses causes :

- (i) un *biais d'agrégation* des données, mais dans cet article, l'agrégation des données individuelles en pseudo-panel est opérée de manière exacte pour les systèmes AI et QAIDS ;
- (ii) des *erreurs de mesure* sur les variables endogènes corrélées à celles des variables explicatives : une telle corrélation pourrait exister entre les erreurs sur les consommations partielles et celles de la dépense totale. L'agrégation des données individuelles dans des cellules assez nombreuses constitue un mode d'instrumentation qui devrait réduire cette corrélation. On a, au surplus, instrumenté la dépense totale par les CSP, le niveau d'éducation, l'âge et l'âge au carré de la personne de référence au ménage, le fait de posséder son logement ou une résidence secondaire : les paramètres estimés en between et within demeurent tout aussi éloignés. Cette cause, invoquée en particulier par Mairesse, ne semble donc pas explicative ici ;
- (iii) une erreur de spécification par *omission d'un facteur dynamique* : l'éloignement des enquêtes interdit d'estimer un modèle dynamique, et l'on peut penser que les ajustements partiels aux modifications de la dépense totale entre deux enquêtes sont,

sur une période de cinq ans, négligeables. Cette cause possible justifie néanmoins une analyse approfondie⁷;

(iv) *l'endogénéité des effets spécifiques* (prouvée par le test d'Hausman) biaiserait les estimations between, qui différeraient des estimations within qui éliminent ces effets fixes.

Les tests d'Hausman prouvent en effet, pour treize des quinze postes, que les effets spécifiques ne sont pas indépendants des variables explicatives : $E(\alpha_H | X_{ht}) \neq 0$. Si l'on suppose, selon la démarche de Mundlak, l'existence d'une relation entre ces effets spécifiques et les transformées between des variables explicatives :

$$\alpha'_H = BX_{Ht} \pi_i + \zeta_H$$

$$\text{avec } \zeta_h \sim N(0, \sigma_\zeta^2), \text{ i.i.d.}$$

$$E(\zeta_h, BX_{ht}) = 0$$

alors le modèle en between s'écrit :

$$BW'_{Ht} = BX_{Ht} \beta_i + BX_{Ht} \pi_i + \zeta_H + Bu_{Ht}$$

avec X_{Ht} les variables explicatives du système complet de demande (5), ce qui implique pour les paramètres β_i des X_{Ht} estimés en between :

$$E\left(\hat{\beta}_i\right) = E\left[(X' BX)^{-1} X' Bw_i\right] = \beta_i + \pi_i.$$

En within, l'estimateur de β_i est tel que : $E\left(\hat{\beta}_i\right) = \beta_i$.

Le biais π_i de l'estimation between peut provenir par exemple de l'existence des contraintes qui jouent différemment selon la richesse des ménages (l'existence de ces contraintes étant alors corrélés à leur revenu) ou des changements des prix relatifs entre classes sociales qui, n'étant pas prises en compte dans les estimations, en cross-section, s'incorporeraient dans les élasticités-revenus (l'existence de contraintes pouvant d'ailleurs se traduire par des

⁷ En cours sur le Panel Study of Income Dynamics et les panels polonais (1988-1994).

changements des prix virtuels qui en découlent). Si par exemple le prix relatif complet d'un bien (intégrant le coût de toutes les ressources du ménage et les prix virtuels liés aux contraintes qu'il subit) dépend du revenu relatif des ménages, avec une élasticité constante ζ_i : $p_i = a_i y^{\zeta_i}$, on montre facilement que l'élasticité-revenu en cross-section diffère de l'élasticité-revenu estimée sur série temporelle (pour une série des ménages conservant leur position dans l'échelle des revenus relatif) en fonction des élasticités prix croisées du bien i par rapport aux prix relatifs du bien j : $e_{cs} - e_{ct} = \sum_j \zeta_j e_j$. Si l'on suppose de plus que $\zeta_j = 0$ pour $j \neq i$, la différence entre les deux élasticités indique l'évolution inverse des prix complets le long de l'échelle des revenus.

SECTION III - LOIS DE CONSOMMATION ALIMENTAIRES

Cette partie de l'étude concerne l'analyse de l'évolution des postes alimentaires détaillés⁸ par rapport aux dépenses alimentaires à domicile. Nous étudions le mouvement des coefficients budgétaires intra-alimentaires selon trois classes d'âge (moins de 37 ans, de 37 à 57 ans et plus de 57 ans), et selon six niveaux de diplôme de la personne de référence au ménage et également selon la localisation géographique du ménage (Paris x province).

Les biens alimentaires détaillés peuvent être divisés en deux groupes selon leur importance dans le budget alimentaire des ménages. Un premier groupe de biens qui pèse le plus lourdement sur ce budget est composé, par ordre de poids dans la dépense alimentaire à domicile, par : **viandes, fruits et légumes, pains et céréales, produits laitiers** (ces deux derniers dans le même ordre de grandeur) et finalement **autres produits**.

Un deuxième groupe de biens comportant les biens moins onéreux pour le budget alimentaire à domicile est composé, par ordre décroissant de poids, par **poissons, huiles et graisses, boissons non alcoolisées, café et thé, pommes de terre** et en dernier **sucre**. La moyenne de ces coefficients budgétaires pour la période 1979 - 1989 est illustrée au graphique 5 en annexe et l'on peut remarquer également cette hiérarchie des dépenses aussi dans tous les autres graphiques qui suivent.

Pour l'ensemble de l'échantillon, les coefficients budgétaires d'alimentation à domicile et d'alcool et tabac ont varié négativement pour la décennie étudiée de pratiquement un cinquième, tandis que celui de l'alimentation à l'extérieur n'a eu qu'une légère augmentation positive.

⁸ Les postes alimentaires détaillés sont décrits succinctement en annexe 3.

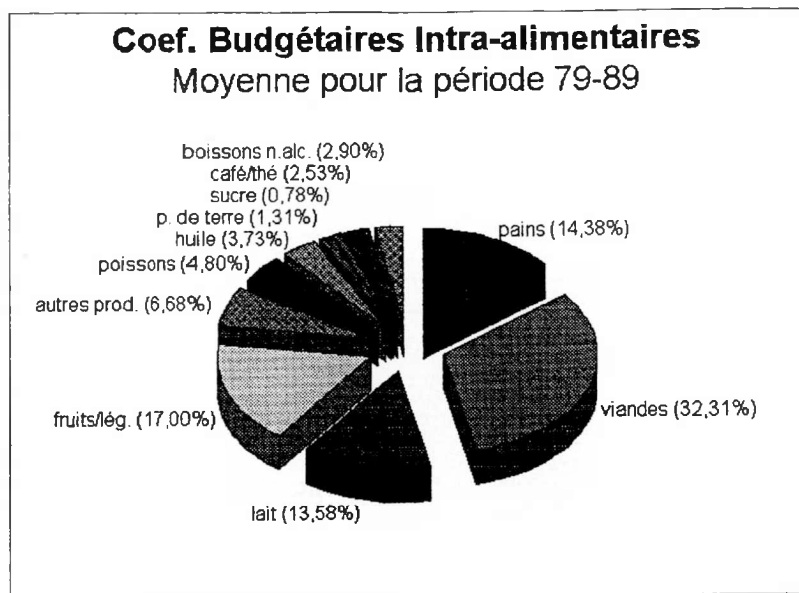
Ce mouvement de chute du coefficient budgétaire d'alimentation à domicile a été vérifié pour toutes les trois classes d'âge : de 20% pour ceux dont la personne de référence au ménage avait jusqu'à 37 ans au moment de l'enquête, de 22% pour ceux dont la personne de référence au ménage avait de 37 à 57 ans, et de 18% pour les autres ménages. Cette baisse relativement plus petite pour le troisième groupe de ménages montre donc une certaine rigidité de la structure de consommation de ces ménages par rapport aux biens alimentaires. Un tel phénomène peut être lié à une certaine stabilité du nombre de personnes au sein du ménage (pas de nouveau-nés et pas de départ des enfants) ou encore à une inélasticité des habitudes alimentaires de personnes âgées.

La réduction de la part budgétaire concernant l'alcool et le tabac s'est avérée plus grande pour les ménages dont la personne de référence est âgée, 28%. Les ménages dont la personne de référence est jeune ont eu une baisse de ce coefficient budgétaire de 12% contre 22% pour les ménages dont la personne de référence était au milieu du cycle de vie.

Contrairement à ces mouvements des coefficients budgétaires d'alimentation à domicile et d'alcool et tabac, le coefficient d'alimentation à l'extérieur a augmenté légèrement pour les trois classes d'âge. Il y a eu une variation autour de 5% pour les deux classes d'âge extrêmes et d'à peine 2% pour celle du milieu.

Les biens **fruits et légumes, produits laitiers, pommes de terre et poissons** (excepté ce dernier pour la dernière classe d'âge) ont subi une variation dans le temps peu importante (au maximum de 14%) de leurs coefficients budgétaires intra-alimentaires. Ces biens apparaissent donc comme les moins dynamiques temporellement et curieusement leurs coefficients sont toujours, en niveau, plus élevés pour les ménages dont la personne de référence est âgée que pour les autres ménages (voir les graphiques 6 et 7).

Figure 5 : Coefficients budgétaires intra-alimentaires - Moyenne pour la période 1979-1989



Les postes **pains et céréales**, **viandes**, **sucres**, **thé et café** ont présenté une évolution plus importante que les biens précédents. Parmi ceux-ci seul le coefficient de **pains et céréales** a connu une augmentation. Pour **thé et café**, toutes les classes d'âge présentent une même variation négative de 29% de leur coefficient budgétaire pour l'ensemble de la période, ce qui peut indiquer un choc de prix pour ce poste, ressenti de la même manière par toutes les classes d'âge confondues. Voir également les graphiques 6 et 7.

Tableau 5 : Coefficients Budgétaires Intra-alimentaires 1 par classes d'âge

Biens alimentaires	Classe d'âge 1 moins de 37 ans			Classe d'âge 2 de 37 à 57 ans			Classe d'âge 3 plus de 57 ans		
	1979	1984	1989	1979	1984	1989	1979	1984	1989
Viandes	34,32	30,82	27,67	35,36	33,38	30,59	34,67	32,10	30,06
Fruits et Légumes	15,77	15,69	15,00	16,71	16,88	16,65	18,47	18,92	19,08
Pains et Céréales	14,39	15,46	17,32	13,41	14,00	15,13	12,79	13,36	14,21
Produits Laitiers	13,85	14,61	15,56	13,14	13,50	13,98	12,17	12,70	12,96
Autres Produits	6,18	7,93	10,42	4,84	6,35	8,77	4,02	5,21	6,78

Figure 6 : Coefficients budgétaires intra-alimentaires 1 - Niveau par classe d'âge et par enquête

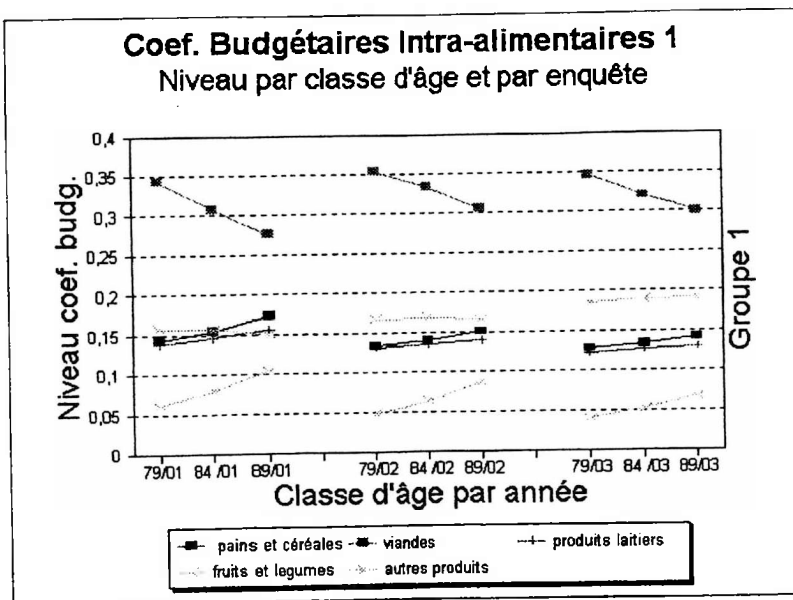
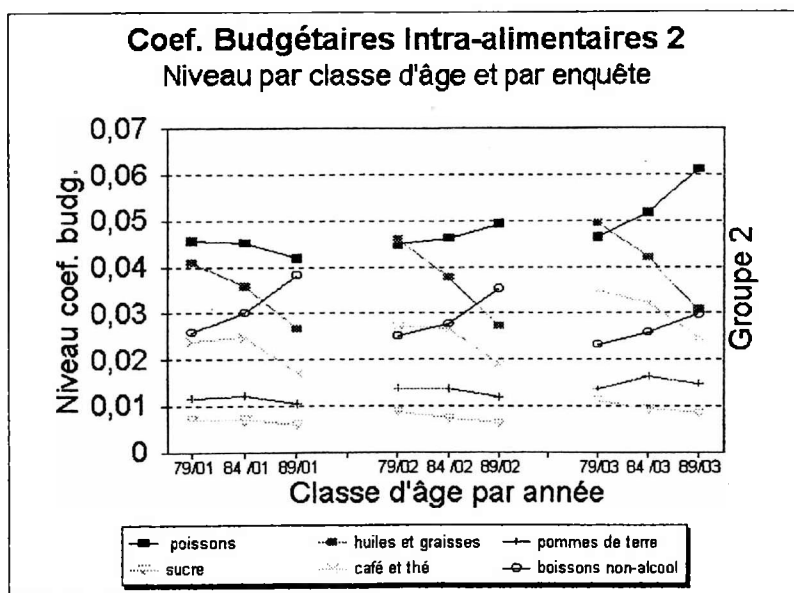


Tableau 6 : Coefficients Budgétaires Intra-alimentaires 2 par classes d'âge

Biens alimentaires	Classe d'âge 1 moins de 37 ans			Classe d'âge 2 de 37 à 57 ans			Classe d'âge 3 plus de 57 ans		
	1979	1984	1989	1979	1984	1989	1979	1984	1989
Poissons	4,56	4,52	4,19	4,48	4,62	4,93	4,64	5,17	6,10
Huile	4,10	3,58	2,67	4,59	3,78	2,72	4,95	4,20	3,07
Boissons non alcoolisées	2,58	3,00	3,83	2,50	2,76	3,53	2,32	2,58	2,97
Café et thé	2,39	2,47	1,70	2,72	2,66	1,90	3,48	3,21	2,47
Pommes de terre	1,16	1,21	1,04	1,37	1,35	1,17	1,37	1,63	1,47
Sucre	0,70	0,70	0,60	0,87	0,73	0,63	1,11	0,93	0,83

Figure 7 : Coefficients budgétaires intra-alimentaires 2 - Niveau par classe d'âge et par enquête



Finalement, les biens **huiles et graisses, boissons non alcoolisées et autres produits** ont présenté les plus fortes variations relatives des coefficients budgétaires intra-alimentaires, ce qui les caractérise comme des biens temporellement dynamiques. Pour le premier bien cette variation a été négative et pour les deux autres positive. On note que les coefficients de ces biens sont pratiquement les plus grands, en niveau, pour les ménages qui sont dans le premier tiers du cycle de vie. Ce sont les aliments services à forte valeur ajoutée (BRSA, plats préparés, ...).

Les graphiques 6 et 7 permettent de faire quelques remarques supplémentaires. Les variations des coefficients budgétaires intra-alimentaires révèlent des effets de substitution, des effets-prix et des différents degrés d'élasticité-revenu qui seront l'objet de l'étude. Les ménages dont la personne de référence est âgée sont ceux qui consomment relativement le plus de **poissons**, surtout en 1989.

Pour les trois enquêtes étudiées, le coefficient budgétaire de l'alimentation à domicile baisse au fur et à mesure que le niveau de diplôme de la personne de référence du ménage augmente (sans contrôler le revenu). Il diminue de 27% en 1979 (et de 24% et de 21% respectivement en 1984 et 1989) pour les ménages dont la personne de référence n'a pas de diplôme jusqu'à 14% en cette même année (et 13% et 11% pour 1984 et 1989) pour les ménages dont la personne de référence a BAC + 5 ou plus. Cette évolution est probablement causée par plusieurs facteurs corrélés au niveau de diplôme, comme le niveau de revenu. Ainsi, à ce stade on ne peut tirer des conclusions fermes sur les facteurs déterminants. Les statistiques descriptives permettent néanmoins de détecter quel est le sens de l'évolution des dépenses relatives des ménages au fil du temps et aussi par des coupes transversales selon des caractéristiques sélectionnées, le cas échéant le niveau d'éducation formelle de la personne de référence du ménage.

L'évolution transversale du coefficient budgétaire d'alcool et tabac suit un profil similaire à celui de l'alimentation à domicile en ce qui concerne les différents niveaux de diplôme. En contrepartie, celle de l'alimentation à l'extérieur varie dans le sens opposé à ces

deux premiers coefficients : plus le niveau de diplôme est élevé plus le coefficient budgétaire d'alimentation à l'extérieur est grand.

En regardant l'évolution temporelle des coefficients budgétaires intra-alimentaires, on remarque difficilement une évolution différenciée pour deux biens. Le coefficient de la **viande** a baissé plus fortement au fil du temps pour les ménages avec des diplômes plus élevés (jusqu'à 40% pour bac+ 3 ou 4) tandis que chez les autres ménages cela a été moins fort (de 15% pour les sans diplôme). En ce qui concerne la consommation de **poissons**, son coefficient budgétaire a augmenté pour les deux premières et pour la dernière catégories de la variable diplôme de près de 20%, tandis que pour les autres ce mouvement n'a pas été accentué.

Figure 8 : Coefficients budgétaires intra-alimentaires 1 - Niveau par localisation géographique et par enquête

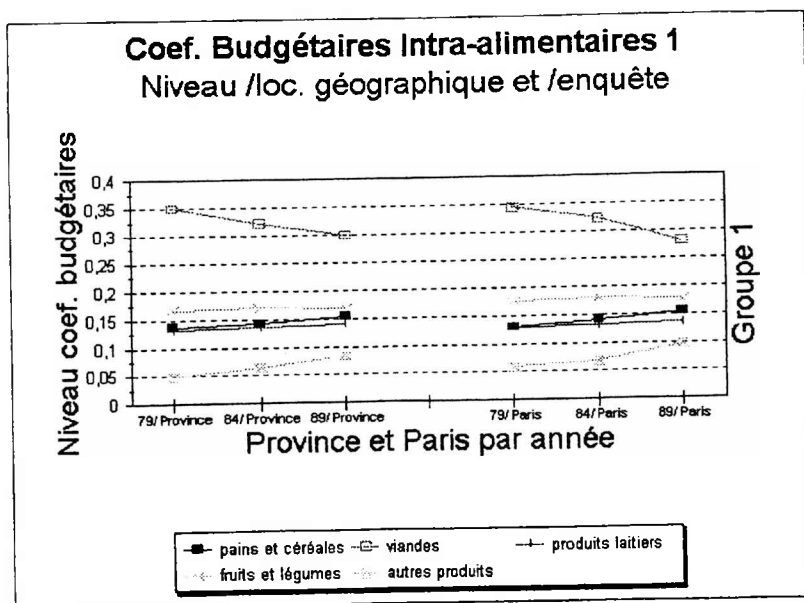


Tableau 7 : Coefficients Budgétaires Intra-alimentaires 1 par localisation géographique

Biens alimentaires GROUPE 1	Paris			Province			Toute la France		
	1979	1984	1989	1979	1984	1989	1979	1984	1989
Viandes	34,54	32,33	28,14	35,00	32,29	29,97	34,92	32,30	29,71
Fruits et Légumes	17,51	17,94	17,70	16,72	16,99	16,91	16,85	17,14	17,02
Produits Laitiers	13,01	13,09	13,47	13,14	13,67	14,13	13,12	13,58	14,04
Pains et Céréales	12,97	14,05	15,41	13,65	14,25	15,36	13,54	14,22	15,37
Autres Produits	5,88	6,80	9,64	4,85	6,40	8,36	5,02	6,46	8,54

Finalement, l'étude des coefficients budgétaires intra-alimentaires selon la localisation géographique du ménage est résumé visuellement dans les graphiques 8 et 9. On remarque immédiatement qu'il n'y a aucune différence en niveau très marqué pour aucun bien. Toutes les évolutions temporelles de coefficients gardent aussi pratiquement le même profil indépendamment du bien en question.

Figure 9 : Coefficients budgétaires intra-alimentaires 2 - Niveau par localisation géographique et par enquête

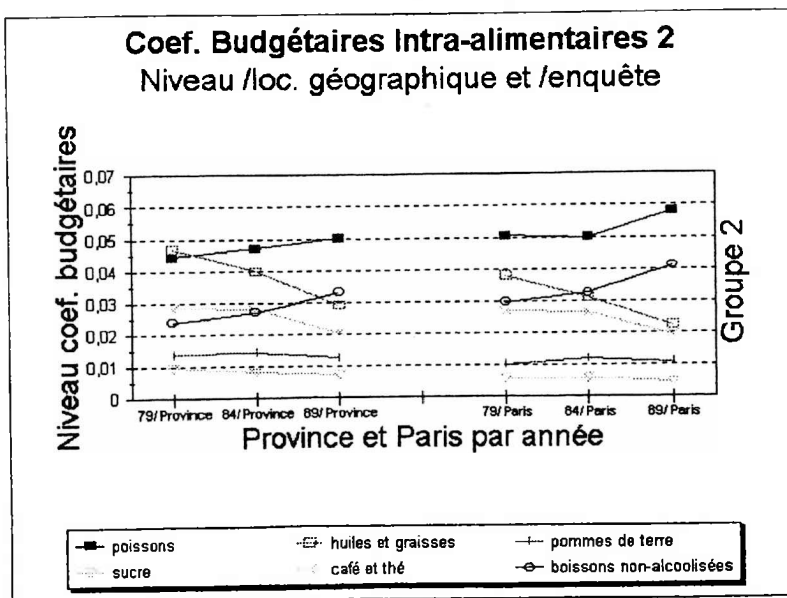


Tableau 8 : Coefficients Budgétaires Intra-alimentaires 2 par localisation géographique

Biens alimentaires GROUPE 2	Paris			Province			Toute la France		
	1979	1984	1989	1979	1984	1989	1979	1984	1989
Poissons	5,04	4,99	5,81	4,44	4,70	5,01	4,54	4,75	5,12
Huiles	3,81	3,13	2,24	4,68	3,98	2,91	4,54	3,85	2,82
Café et Thé	2,68	2,64	1,94	2,82	2,79	2,05	2,80	2,77	2,03
Boissons non alcoolisées	2,96	3,25	4,10	2,39	2,69	3,31	2,48	2,78	3,42
Pommes de terre	1,02	1,20	1,09	1,37	1,43	1,26	1,31	1,39	1,23
Sucre	0,57	0,59	0,46	0,94	0,81	0,72	0,88	0,78	0,69

Tableau 9 : Crière synthétique (CGR)

Classes Sociales	Année	Pauvre (PP)	Quasi pauvre (MP)	Classe Moyenne (MM)	Quasi riche (MR)	Riche (RR)
% effectifs	1979	6,12	12,01	66,31	10,11	5,43
	1984	5,94	11,40	67,35	10,11	5,20
	1989	6,20	11,90	66,53	9,89	5,47
C. BUD. Alimentation à domicile	1979	50,55 (14,05)	38,52 (12,78)	23,61 (10,38)	12,54 (6,26)	10,03 (4,96)
	1984	43,70 (12,81)	34,31 (12,15)	21,35 (9,37)	11,65 (5,45)	9,48 (4,41)
	1989	39,55 (11,49)	30,67 (11,37)	19,11 (8,72)	10,60 (5,02)	8,86 (4,05)
C. BUD. Alimentation à l'extérieur	1979	2,66 (4,18)	3,50 (4,89)	5,95 (7,33)	8,60 (9,03)	9,20 (9,17)
	1984	2,60 (4,43)	3,58 (5,41)	5,73 (7,02)	8,27 (8,79)	8,27 (8,17)
	1989	3,00 (4,14)	4,12 (5,67)	6,15 (7,19)	8,28 (7,63)	7,96 (7,32)
C. BUD. Alcools et Tabac	1979	4,34 (4,92)	3,55 (4,05)	3,08 (3,89)	2,58 (4,16)	2,67 (5,00)
	1984	3,78 (4,41)	3,27 (4,02)	2,66 (3,50)	2,14 (3,08)	1,92 (3,36)
	1989	3,27 (3,95)	2,93 (4,10)	2,47 (3,40)	1,94 (2,92)	1,71 (2,79)

RÉSULTATS POUR LES POSTES ALIMENTAIRES DÉTAILLÉS

Les systèmes complets de demande sont d'abord estimés pour une répartition de la demande totale en dépenses partielles d'une nomenclature en grand groupe fonctionnelles, comprenant l'alimentation à domicile, l'alimentation à l'extérieur et les dépenses d'alcool et tabac.

Tableau 10 : Modèle AIDS - Cellulage DGC

Biens **	Elasticité-revenu alimentaire par rapport							
	à l'alimentation à domicile				à la dépense totale			
	OLS	Between	Within	GLS	OLS	Between	Within	GLS
Pains / céréales	0,769* (15,82)	0,757* (9,04)	0,891* (8,83)	0,772* (15,35)	0,241 [0,17;0,32]	0,283 [0,16;0,44]	0,330 [0,16;0,56]	0,228 [0,15;0,31]
Viandes	0,946* (18,77)	1,004* (9,74)	0,998* (12,51)	0,955* (17,23)	0,296 [0,22;0,38]	0,376 [0,21;0,58]	0,369 [0,19;0,59]	0,282 [0,20;0,38]
Poissons	2,006* (16,00)	1,676* (9,60)	1,492* (5,97)	2,004* (16,29)	0,628 [0,45;0,83]	0,627 [0,35;0,97]	0,552 [0,23;1,02]	0,591 [0,41;0,81]
Laitiers	0,776* (17,67)	0,748* (9,06)	0,827* (9,81)	0,787* (16,82)	0,243 [0,17;0,31]	0,280 [0,15;0,44]	0,306 [0,15;0,51]	0,232 [0,16;0,31]
Huiles	0,136 (1,62)	0,286 (1,75)	0,496* (3,32)	0,176 (1,94)	0,043 [-0,01;0,1]	0,107 [-0,01;0,3]	0,183 [0,05;0,40]	0,052 [-0,0;0,13]
Fruits/ Légumes	1,368* (27,23)	1,229* (14,63)	1,238* (13,63)	1,366* (25,94)	0,428 [0,32;0,54]	0,460 [0,28;0,67]	0,458 [0,24;0,73]	0,402 [0,29;0,53]
Pommes de terre	0,303* (2,05)	0,407 (1,71)	0,795* (2,93)	0,328* (2,12)	0,095 [0,00;0,21]	0,152 [-0,01;0,4]	0,294 [0,06;0,68]	0,097 [0,01;0,23]
Sucre	-0,571* (-3,40)	-0,411 (-1,35)	0,321 (1,01)	-0,502* (-2,82)	-0,179 [-0,2;-0,1]	-0,154 [-0,27;0,1]	0,119 [-0,07;0,5]	-0,148 [-0,2;-0,1]
Café et thé	0,514* (5,23)	0,556* (4,00)	0,884* (4,09)	0,510* (5,46)	0,161 [0,08;0,26]	0,208 [0,07;0,40]	0,327 [0,10;0,67]	0,150 [0,07;0,25]
Autres produits	1,319* (10,53)	1,482* (7,41)	1,075* (5,21)	1,267* (9,47)	0,413 [0,27;0,58]	0,554 [0,29;0,90]	0,398 [0,15;0,76]	0,374 [0,23;0,55]
B. non alcoolisés	1,507* (13,01)	1,717* (9,90)	1,013* (4,07)	1,519* (13,36)	0,472 [0,33;0,64]	0,642 [0,37;0,99]	0,375 [0,12;0,77]	0,448 [0,30;0,62]

* Estimations significatives au seuil de 5%.

** Voir annexe 5 pour le détail de la nomenclature des postes alimentaires.

T de Student entre parenthèses et plage de variation à 95% entre crochets

Dans une seconde étape un système complet est estimé sur une décomposition alimentaire à domicile sur 11 postes détaillés et les paramètres d'élasticité à la dépense totale alimentaire sont combinés à l'élasticité alimentaire à cette dépense totale pour obtenir les élasticité de dépenses alimentaires partielles par rapport à la dépense totale. On indique également une plage de variation à 95% de confiance de ces élasticité, tableau 10. Cet intervalle de confiance a été calculé de la façon suivante :

$$\left[e_{AI,AD}^{inf} \cdot e_{ADI,DT}^{inf} \cdot e_{AI,AD}^{sup} \cdot e_{ADI,DT}^{sup} \right]$$

où

- $e_{AI/AD}^{inf(sup)}$ est la limite inférieure (supérieure) de l'intervalle de confiance à 95% de l'élasticité-revenu alimentaire par rapport à l'alimentation à domicile, et
- $e_{AD/DT}^{inf(sup)}$ est la limite inférieure (supérieure) de l'intervalle de confiance à 95% de l'élasticité-revenu alimentaire à domicile par rapport à la dépense totale.

Les résultats de cette estimation sur données transformées within ne sont pas fondamentalement différentes de celles que présente P. Combris (1992) dans ses estimations sur séries temporelles agrégées selon plusieurs spécifications. Mais ses résultats sont très différents selon la spécification qu'il a adopté. On peut surtout comparer nos estimateurs within à ceux de sa spécification AIDS statique ou dynamique. Les élasticités-revenu des pseudo-panels sont de même ordre de grandeur que les siennes, sauf pour les boissons non alcoolisées. Mais pour ce dernier poste seule l'estimation AIDS statique de Combris est très différente de l'AIDS dynamique sur série temporelle macro. Pour ce poste, ses calculs fournissent une élasticité de 0,35 comparable à notre estimation de 0,38.

On peut considérer de manière générale que les élasticités-revenu de pseudo-panel ont un ordre de grandeur plausible. Par ordre décroissant de dynamisme, on a trois groupes de biens :

- 1) poissons, fruits et légumes, autres produits, boissons non alcoolisées, viandes,
- 2) pains et céréales, produits laitiers, pommes de terre, thé et café,
- 3) huile et sucre.

A propos des différences transversales et longitudinales, les conclusions que l'on peut tirer ne diffèrent pas de celles que l'on a présentées pour l'ensemble des dépenses fonctionnelles.

CONCLUSION

L'estimation d'un système complet de demande Qaids sur un pseudo-panel de trois enquêtes de budgets de famille de l'Insee fournit des élasticités-revenu transversales (estimées en coupe) souvent très différentes des élasticités temporelles (estimées en séries temporelles des ménages regroupés en cellules homogènes). Ce résultat peut refléter l'endogénéité des effets fixes, qui impliquerait que les estimations en coupe sont biaisées et que les estimateurs within sont seuls convergents. Dans ce cas, on ne pourrait utiliser les estimations en coupe pour prévoir les évolutions des consommations au cours du temps.

Une seconde explication de cette différence tiendrait à l'omission de variables dynamiques dans les lois de consommation, qui proviendraient de processus d'ajustement partiels, de comportements myopes ou d'effets d'habitude ou de stock. Une recherche en cours sur les panels polonais (1987-90) et le Panel Study of Income (1983-87)⁹ permettra de juger si la prise en compte des effets dynamiques permet de réduire totalement la différence des deux types d'élasticités.

⁹ En collaboration avec G. Duncan, P. Gaubert et C. Starzec.

BIBLIOGRAPHIE

- ATTANASIO, O, BROWNING, M. (1995) - 'Consumption over Life Cycle and over Business Cycle', *American Economic Review*, December, pp. 1118-1137.
- BALTAGI, B. H.(1995) - 'Econometric Analysis of Panel Data', pp.257, J. Wiley & Sons.
- BANKS, J., BLUNDELL, R., LEWBEL, A. (1992) - 'Quadratic logarithmic Engel curves and consumer demand.' Mimeo.
- BLUNDELL, M, BROWNING, M, MEGHIR, C. (1989) - 'A microeconomic model of intertemporal substitution and consumer demand.' Discussion Paper In Economics 89 - 11, UCL, London.
- BROWNING, M, DEATON, A, IRISH, M. (1985) - 'A profitable approach to labor supply and commodity demands over life cycle.' *Econometrica*, Vol. 53, pp. 503-543.
- CARDOSO, N., GARDES, F. (1996) - 'Caractérisation et analyse des comportements de consommation des ménages pauvres sur données individuelles françaises', *Revue Économique*, Mai.
- COMBRIS, P. (1992) - 'Changements structurels : le cas des consommations alimentaires en France de 1949 à 1988', *Économie et Prévision*, numéro 102-103, pp. 221-246.
- DEATON, A (1985) - 'Panel data from time series of cross-sections', *Journal of Econometrics* - North Holland.
- DEATON, A. et MUELLBAUER, J. (1980) - 'An almost ideal demand system', *American Economic Review*, 70, n° 3, pp. 312-326.

-
- DORMONT, B. (1989) - *Introduction à l'analyse de données de panel*. Monographie d'Économétrie, Editions du CNRS.
- GARDES, F. (1983) - 'L'évolution de la consommation marchande en Europe et aux USA depuis 1960.' *Consommation -Revue de Socio-Economie*.
- GARDES, F. et LOUVET (1986) - 'La convergence internationale des structures de consommation. Cahiers.' CREDOC.
- GARDES, F., LANGLOIS, S. et RICHAUDEAU, D. (1996) - 'Cross-section versus time-series income elasticities: an estimation on a pseudo-panel of Canadian Surveys.', *Economic Letters*.
- MADDALA, G.S. (1988) - *Econometrics*. McGraw-Hill International Editions, 5^e ed.
- NICHELE, V. et ROBIN, J. M. (1993) - 'Evaluation des effets budgétaires et redistributifs de reformes de la fiscalité indirecte française.' Document de travail n° 9303, INRA.
- POLLAK, R et WALES, T. (1992) - *Demand system specification and estimation*. Oxford University Press.
- ROSSI (1983) - 'Elasticidades de Engel para dispendios familiares nas principais capitais brasileiras.' *RBE*, 44 (176): 381-451.
- VERBEEK, M. (1992) - 'Pseudo panel data' in *The econometrics of panel data*. Kluwer Academic Publishers.
- VERBEEK, M. et NIJMAN, T. (1992) - 'Can cohort data be treated as genuine panel data?' in *Panel Data Analysis*, 10 - 23, Physica-Verlag Heidelberg.

ANNEXES

ANNEXE 1 : LE CELLULAGE 'DGC'

Tableau 11 : Structure des cellules 'DGC'.

geo	trage \dip1	02	03	04	05	06	07	08	09	10	11
0	1										
0	2										
0	3										
0	4										
0	5										
0	6										

geo	trage \dip1	02	03	04	05	06	07	08	09	10	11
1	1										
1	2										
1	3										
1	4										
1	5										
1	6										

Les modalités des trois variables discriminantes sont:

Geo = 0, le ménage n'habite pas en région parisienne,

= 1, le ménage habite en région parisienne.

Dipl (cette variable concerne la personne de référence au ménage),

=1, pas de certificat,

=4, BAC, Brevet supérieur,
BAC 1^{ère} partie

=2, certificat d'études primaires,

=5, BAC + 2,

=3, BEP, CAP, BEPC ou niveau
équivalent,

=6, BAC + 3, 4 ou 5.

Trage (cette variable concerne également la personne de référence au ménage)

- | | |
|--------------------------|--------------------------|
| = 02, né de 1957 à 1962, | = 07, né de 1932 à 1937, |
| = 03, né de 1952 à 1957, | = 08, né de 1927 à 1932, |
| = 04, né de 1947 à 1952, | = 09, né de 1917 à 1927, |
| = 05, né de 1942 à 1947, | = 10, né de 1907 à 1917, |
| = 06, né de 1937 à 1942, | = 11, né avant 1907. |

Tableau 12 : Statistiques Descriptives des Cellules 'DGC'.

Version BDF	Nb. cellules	Pourcentage des cellules avec			Taille de cell. moy.	Taille min cellule	Taille max cellule
		<50 mén.	50<mén.<100	>100 mén.			
1979	56	7 %	27 %	66 %	166,30	18	506
1984	56	3 %	18 %	79 %	191,89	27	549
1989	56	11 %	36 %	53 %	137,89	24	529

On trouvera exposé un autre type de cellulage, basé sur l'appartenance des ménages à une cohorte d'âge et à une classe sociale définie par des critères de pauvreté ou de richesse, dans Cardoso-Gardes (1996).

ANNEXE 2 : BASE DE DONNÉES (FICHER HARMONISÉ)

PRINCIPES D'HARMONISATION DES ENQUÊTES¹⁰

Le but a été l'homogénéisation d'un sous-ensemble de variables issu des enquêtes 'Budget de Famille' en France pour les années 79, 84 et 89. Les bases de données de l'INSEE à partir desquelles ces variables ont été sélectionnées sont :

- "GN90.L03.SASBDF89" et "GN90.L38.BDF89.MENDEP" en 1989 ;
- "GN90.L02.BDF84.CERC1" pour les variables originaires du niveau "ménage" et "GN90.L02.BDF84.CERC2" pour celles issues du niveau "individu", en 1984 ;
- et "GN90.L38.BDF79.MENDEP" et "GN90.L02.BDFT79" en 1979.

Le nombre d'observations varie d'une enquête à l'autre, soit :

- 9 038 ménages et 24 595 individus en 1989,
- 11 976 ménages et 33 236 individus en 1984,
- 10 645 ménages et 32 475 individus en 1979.

Il a fallu recoder et traduire plusieurs variables originelles de façon à rendre comparables les trois enquêtes. Cette opération s'est révélée assez longue surtout pour l'année 79, puisqu'il y a eu une rupture de méthodologie cette année là.

Le résultat de cette procédure a été la création de cinq tableaux de base pour chaque enquête, appelés désormais fichier harmonisé.

¹⁰ La construction de la base de données a été réalisée dans le cadre d'un stage du 19 juillet 1993 au 30 septembre 1993 auprès de l'INSEE (Institut National de la Statistique et des Etudes Economiques).

Tableau 13: Les tableaux de base du fichier harmonisé.

Nom de tableau	Libellé du tableau	Nombre d'observations	Nombre de variables
CONS79	Dépenses détaillées du ménage en 79, selon la nomenclature approchée de celle de la Compta. Nat.	10645	70
CONS84	Idem CONS79, mais pour l'année 84.	11976	70
CONS89	Idem CONS79, mais pour l'année 89.	9038	70
INDI79	Variables au niveau individuel pour 79.	32475	7
INDI84	Idem INDI79, mais pour l'année 84	33236	7
INDI89	Idem INDI79, mais pour l'année 89	24595	7
MENA79	Variables au niveau ménage, pour l'année 79	10645	19
MENA84	Idem MENA79, mais pour l'année 84	11976	19
MENA89	Idem MENA79, mais pour l'année 89	9038	19
REV79	Revenus fiables ¹¹ des ménages en 79	9371	7
REV84	Revenus fiables et détaillés des ménages en 84	11076	20
REV89	Idem REV84, mais pour l'année 89	8487	20
SALAIR79	Salaires des ménages en 79	10645	3
SALAIR84	Salaires des ménages en 84	11977	3
SALAIR89	Salaires des ménages en 89	9038	3

Les tableaux CONS79, CONS84 et CONS89 contiennent chacun 70 variables dont 68 commencent par CON (indiquant consommation d'un bien). Ces variables sont le résultat d'une réorganisation des dépenses des enquêtes BDF en des postes compatibles avec la nomenclature de la Comptabilité Nationale. La différence d'optique entre les enquêtes BDF

¹¹ Dans les enquêtes BDF de l'INSEE, les revenus déclarés sont codés comme fiables ou non fiables. Les ménages dont les revenus déclarés n'est pas considéré comme fiables ont été éliminés du fichier.

(optique de dépenses) et les Comptes Nationaux (optique de consommation) implique un changement de nomenclature¹² qui s'est révélée difficile et a parfois exigé l'adoption d'hypothèses ad hoc (ventilation de quelques dépenses génériques, suppression des postes non initialisés).

A titre d'exemple, quelques variables des tableaux CONS84 et CONS89 sont présentées en annexe I B avec leur postes de dépenses originaux des enquêtes 84 et 89 .

La traduction de l'enquête BDF de 79 vers la nomenclature de Comptes Nationaux a été plus complexe, car les postes de dépenses ont subi une évolution sensible entre 79 et 84. Plusieurs postes ont été créés en 84. Les pondérations et éclatements des postes en 79 ont eu 84 comme base d'appui.

¹² A posteriori, il y aura un développement de cette section pour inclure : une description de la façon dont les enquêtes BDF sont menées, quels biais peuvent exister (non réponse et auto-sélection) et de quelle façon la technique de pseudo-panel peut tenir compte de ce problème.

ANNEXE 3 : DÉTAIL DES POSTES ALIMENTAIRES

CON1 - Dépense en produits alimentaires, boissons et tabacs**CON111 - Pain et céréales**

- 1000 - Pain ordinaire, bâtard, boule, ficelle
- 1001 - Biscuits, biscottes
- 1002 - Pâtisserie
- 1003 - Pâtisserie surgelée
- 1004 - Riz
- 1005 - Pâtes alimentaires
- 1006 - Farine de blé
- 1007 - Couscous et autres semoules
- 1008 - Autres céréales
- 1094 - Desserts à préparer et produits à pâtisserie
- 1114 - Cacao, banania, chocolat en poudre et autres produits cacaotés
- 1120 - Produits pour enfants et bébés
- 1130 - Produits diététiques
- 1300 (x 0,13) - Dépenses alimentaires non détaillées (pour 84)
- 1300 (x 0,14) - Dépenses alimentaires non détaillées (pour 89)

Dépenses incluses à tort: une partie de dep.1094 correspondant à levure est incluse dans le poste 11A.

CON112 - Viandes

- 1030 - Boeuf frais
- 1031 - Veau frais
- 1032 - Mouton, agneau, chevreau (frais)
- 1033 - Cheval frais
- 1034 - Porc frais, salé, fumé, demi-sel
- 1035 - Boeuf surgelé
- 1036 - Veau surgelé
- 1037 - Mouton, agneau, chevreau(surgelés)
- 1038 - Cheval surgelé (seulement pour 84)
- 1039 - Porc surgelé
- 1040 - Jambon frais
- 1041 - Autres produits de charcuterie
- 1042 - Triperie et abats frais
- 1043 -Triperie et abats surgelés
- 1044 - Conserves de viande
- 1060 - Volailles
- 1061 - Lapins
- 1062 - Gibiers
- 1063 - Volailles surgelées
- 1064 - Lapins surgelés
- 1065 - Gibiers surgelés
- 1300 (x 0,29) - Dépenses alimentaires non détaillées (pour 84)
- 1300 (x 0,25) - Dépenses alimentaires non détaillées (pour 89)

Dépenses omises: le poste CON112 inclut une partie des dépenses.:

- 1046 - Plats préparés à base de : viande, poisson, légumes
- 1047 - Plats préparés en conserve à base de : viande, poisson, légumes

CON113 - Poissons

- 1045 - Conserves de produits de la pêche
- 1050 - Poissons frais, salés, fumés
- 1051 - Crustacés, mollusques, coquillages frais
- 1052 - Poissons et crustacés, coquillages surgelés
- 1300 (x 0,04) - Dépenses alimentaires non détaillées (pour 84)
- 1300 (x 0,05) - Dépenses alimentaires non détaillées (pour 89)

Dépenses omises: le poste con113 inclut une partie des dépenses.

- 1046 - Plats préparés à base de : viande, poisson, légumes
- 1047 - Plats préparés en conserve à base de : viande, poisson, légumes

CON114 - Lait, fromage et oeufs

- 1066 - Oeufs
- 1070 - Lait frais
- 1071 - Lait concentré, en poudre
- 1072 - Fromages(non frais)
- 1073 - Yaourts
- 1074 - Autres fromages frais
- 1075 - Crème fraîche
- 1300 (x 0,12) - Dépenses alimentaires non détaillées (pour 84)
- 1300 (x 0,13) - Dépenses alimentaires non détaillées (pour 89)

CON115 - Huiles et graisses

- 1080 - Beurre
- 1081 - Huiles alimentaires
- 1082 - Margarines et autres graisses
- 1300 (x 0,04) - Dépenses alimentaires non détaillées (pour 84)
- 1300 (x 0,03) - Dépenses alimentaires non détaillées (pour 89)

CON116 - Fruits et légumes sauf pommes de terre

- 1011 - Légumes frais
 - 1012 - Légumes secs
 - 1013 - Légumes en conserve (boîtes ou bocaux)
 - 1014 - Légumes surgelés
 - 1015 - Légumes déshydratés
 - 1020 - Fruits frais non tropicaux
 - 1021 - Fruits tropicaux
 - 1022 - Fruits secs, à coque
 - 1023 - Conserves de fruits
 - 1024 - Confitures
 - 1025 - Fruits surgelés
 - 1300 (x 0,16) - Dépenses alimentaires non détaillées(pour 84 et 89)
- Dépenses omises : le poste CON116 inclut également une partie des dépenses :
- 1046 - Plats préparés à base de : viande, poisson, légumes
 - 1047 - Plats préparés en conserve à base de : viande, poisson, légumes

CON117 - Pommes de terre et autres tubercules

- 1010 - Pommes de terre
- 1300 (x 0,01) - Dépenses alimentaires non détaillées(pour 84 et 89)

CON118 - Sucre

- 1090 - Sucre
- 1300 (x 0,01) - Dépenses alimentaires non détaillées(pour 84 et 89)

CON119 - Café, thé

1112 - Café

1113 - Thé et plantes d'infusion

1300 (x 0,03) - Dépenses alimentaires non détaillées (pour 84)

1300 (x 0,02) - Dépenses alimentaires non détaillées (pour 89)

Dépenses omises: le poste CON119 inclut le dép. :

1096 - Assaisonnements, concentrés pour les plantes aromatiques

CON11A - Autres produits alimentaires y compris confiserie

1046 - Plats préparés à base de : viande, poisson, légumes

1047 - Plats préparés en conserve à base de : viande, poisson, légumes

1048 - Plats préparés surgelés

1091 - Miel

1092 - Chocolat en tablettes et confiseries à base de chocolat

1093 - Confiseries, sucreries

1095 - Crèmes glacées surgelées

1096 - Assaisonnements, concentrés

1097 - Potages déshydratés

1098 - Potages en boîtes et en conserves

1300 (x 0,06) - Dépenses alimentaires non détaillées (pour 84)

1300 (x 0,08) - Dépenses alimentaires non détaillées (pour 89)

Dépenses incluses à tort: une partie des dép. 1046 et 1047 est incluse dans les postes :

CON112 - Viandes

CON113 - Poissons

CON116 - Fruits et légumes sauf pommes de terre

Une partie de dép.1096 correspondant aux plantes aromatiques est incluse dans le poste
CON119 - Café, thé

Dépenses omises : le poste CON11A inclut une partie de dép.1094 - Desserts à préparer et
produits à pâtisserie pour la levure

CON12 - Boissons non alcoolisées

1110 - Eaux minérales

1111 - Autres boissons non alcoolisées

1300 (x 0,03) - Dépenses alimentaires non détaillées(pour 84 et 89)

ANNEXE 4 : ENCADRÉ 1

Soit une équation à une seule variable explicative: $y = \alpha + \beta x + u$.¹³ Nous allons comparer l'efficacité de deux estimateurs : b qui désignera l'estimateur de β si les données utilisées sont individuelles et β_b dans le cas où les données employées sont des moyennes de groupes. La variance des erreurs sur données individuelles est égale à $\text{var}(u_i) = \sigma^2$, tandis que la variance des erreurs sur données groupées équivaut à $\text{var}(\hat{u}_i) = \sigma^2 / n$.

Nous considérons x_{ij} ($j = 1, 2, \dots, n_i$), les observations individuelles de la $i^{\text{ème}}$ cellule; la valeur de cette cellule est la moyenne sur j de ces observations :

$$\bar{x}_i = \left(\frac{1}{n_i}\right) \sum_j x_{ij} .$$

La moyenne générale des cellules et des individus j est:

$$\bar{x} = \frac{\sum_i n_i \bar{x}_i}{\sum_i n_i}$$

La somme totale de carrés pour x ,

$$s_{xx} = \sum_i \sum_j (x_{ij} - \bar{x})^2 ,$$

peut être décomposée en deux parties : la SC between S_{xxb} et la SC within S_{xxw} où

$$s_{xxb} = \sum_i n_i (\bar{x}_i - \bar{x})^2 \text{ et } s_{xxw} = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 .$$

Les variances de b et de β_b sont respectivement :

¹³ La démonstration qui suit est tirée de Maddala (1988). Voir aussi Cramer op.cit. pp. 144-145 pour une discussion à propos de comment l'agrégation augmente la variance des coefficients de la régression.

$$v(b) = \frac{\sigma^2}{S_{xx}} \text{ et } v(\beta_b) = \frac{\sigma^2}{S_{xxb}}.$$

Faisons le quotient de variance de ces deux estimateurs:

$$\frac{v(b)}{v(\beta_b)} = \frac{S_{xxb}}{S_{xx}} = \frac{1}{\left(1 + \frac{S_{xxw}}{S_{xxb}}\right)}.$$

Ainsi, nous arrivons à la conclusion que pour réduire la perte d'efficacité des estimateurs calculés sur des données groupées, par rapport à ceux mesurés à partir des données individuelles, il faut minimiser la variation de la dimension intra-individuelle et maximiser la variation inter-individuelle. Comme le remarque Maddala (1988), cette discussion est valide pour le cas d'une seule variable explicative, ayant celle-ci comme critère de cellulaire.

ANNEXE 5 : ENCADRÉ 2
(DÉCOMPOSITION ORTHOGONALE DE LA VARIANCE DES OBSERVATIONS)

L'intérêt de procéder à des décompositions orthogonales est de dégager, pour les observations, ce qui est comportement moyen de ce qui est comportement transitoire. Et ensuite d'utiliser tout l'arsenal des estimateurs spécifiques aux données de panel pour en retirer des leçons.

Les opérateurs between (ou inter) et within (ou intra) permettent de projeter l'information statistique en deux composantes orthogonales.

Dans l'espace \mathcal{RT} , l'opérateur between s'écrit:

$$B_{(NT,NT)} = I_N \otimes \frac{J_T}{T}$$

où I_N est la matrice identité de dimension (N, N) et J_T/T est une matrice de dimension (T, T) dont les éléments sont tous $1/T$.

L'opérateur within, dans cet espace, est égal à:

$$W_{(NT,NT)} = I_N \otimes \left(I_T - \frac{J_T}{T} \right)$$

où I_T est la matrice identité de dimension (T, T) et les autres matrices sont telles comme définies auparavant (voir Baltagi, 1995).

Si on pré-multiplie un vecteur y avec NT observations par l'opérateur between, on calcule T fois la moyenne des y_i . Si, par contre, l'on pré-multiplie par l'opérateur within, on calcule les écarts à la moyenne de chacune des composantes y_i de y .

La structure de la variance des observations varie selon leur disposition en niveau, en ratio ou en taux de croissance. Dormont (1989) explique que pour des variables en niveaux la variance between prédomine sur la variance within. De ce fait, les estimateurs calculés sur données empilées sont proches de ceux calculés sur données transformées en between. Ces estimateurs sont en principe robustes, mais ne sont pas capables de prendre en compte des effets dynamiques. Les estimateurs sur données transformées en within souffrent pour leur part d'une fragilité vis-à-vis d'une omission de variables conjecturelles. Cette prédominance de la variance between sur la variance within est aussi valable sur des données présentées en

ratio, mais dans une proportion moins marquée. En ce qui concerne les variables en taux de croissance, la variance within devient plus importante que celle en between et du coup les estimateurs calculés sur données empilées et sur données transformées en between diffèrent entre eux. Dans ce cadre, plus la dimension temporelle est importante, plus la variance between diminue et la variance within augmente.

Le tableau ci-après présente cette décomposition de la variance:

Tableau 14 : Décomposition de la variance

Variable dépend.	Cellulage 'DGC'			%var bet/tot
	total	between	within	
Alimentation à l'extérieur	0,2086	0,1706	0,1201	66,88
Alimentation à domicile	1,1061	1,0322	0,3974	87,10
Alcools, tabac	0,1293	0,1115	0,0655	74,36
Habillement	0,2774	0,2229	0,1651	64,57
Logement	0,2884	0,2464	0,1499	72,99
Electricité	0,3269	0,2872	0,1560	77,19
Articles ménage	0,3400	0,2906	0,1765	73,05
Santé	0,3455	0,3191	0,1323	92,36
Achat véhicules	0,3692	0,3344	0,1567	82,04
Utilisation véhicules	0,3410	0,3020	0,1585	78,43
Transports publics	0,0741	0,0590	0,0449	63,40
Télécom.	0,0681	0,0540	0,0416	62,69
Loisirs éducation	0,2160	0,1839	0,1133	72,49
Services	0,1149	0,0981	0,0598	72,89
Hors comp. nationale	0,7061	0,5877	0,3914	69,27

Il est clair qu'on se trouve dans un cadre où la variance between est beaucoup plus importante que la variance within. Ceci montre que le cellulaire permet de conserver une partie de la forte hétérogénéité des données individuelles, et donc que l'interprétation des moyennes de cellules comme indicatrice d'un type homogène de ménage n'est pas absurde a priori.

Dépôt légal : Juin 1996

ISSN : 1257-9807

ISBN : 2-84104-066-6

CAHIER DE RECHERCHE

Récemment parus :

Un tour d'horizon des aspirations et des valeurs des Français

Franck BERTHUIT - n°84 (1996)

Visualisation graphique pour l'analyse exploratoire des données

Christine HENRIOT et Marie-Odile SIMON - n°85 (1996)

Construction et analyse des compétences dans le secteur éducatif et social

Patrick DUBÉCHOT - n°86 (1996)

La construction des objectifs d'une mesure de politique d'emploi : l'histoire de l'aide aux chômeurs créateurs d'entreprise

Anne-Lise AUCOUTURIER - n°87 (1996)

Consommateurs et préférences de consommation en 1996

Aude COLLIERIE DE BORELY - n°88 (1996)

Deux articles sur la localisation industrielle

Philippe MOATI - n°89 (1996)

Les inégalités en France : les différentes façons de "penser" en haut et en bas de l'échelle sociale

Georges HATCHUEL, Anne-Delphine KOWALSKI et Jean-Pierre LOISEL - n°90 (1996)

Président : Bernard SCHAEFER Directeur : Robert ROCHEFORT
142, rue du Chevaleret, 75013 PARIS - Tél. : (1) 40.77.85.01

ISBN : 2-84104-066-6

CRÉDOC

Centre de recherche pour l'Étude et l'Observation des Conditions de Vie