

CH. ARBACHE

## **L'évaluation du prix de biens durables d'une espèce donnée**

*Les cahiers de l'analyse des données*, tome 10, n° 4 (1985),  
p. 401-411

[http://www.numdam.org/item?id=CAD\\_1985\\_\\_10\\_4\\_401\\_0](http://www.numdam.org/item?id=CAD_1985__10_4_401_0)

© Les cahiers de l'analyse des données, Dunod, 1985, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## L'ÉVALUATION DU PRIX DE BIENS DURABLES D'UNE ESPÈCE DONNÉE [BIEN DURABLE]

par Ch. Arbache\*

L'objectif direct de notre travail a été de concevoir une méthode pour prévoir le prix de vente d'un bien durable d'une espèce donnée, en fonction de quelques unes des caractéristiques de ce bien ; et cela en se fondant sur les prix de vente publiés pour un certain nombre de biens de même espèce mis récemment sur le marché. Nous exposons d'abord en termes généraux la méthode de régression par l'analyse des correspondances. Puis nous rendrons compte des résultats obtenus sur une base de données réelles ; sans entrer dans les détails concrets, mais en montrant l'importance des choix faits dans l'application de la méthode générale ; et principalement de la technique du codage de la variable à estimer par fenêtre glissante ; technique nouvelle qui contribue de façon décisive à la précision des résultats.

### 1 La méthode de régression par l'analyse des correspondances

1.1 Régression et validation : Nous supposons pour un ensemble  $I$  d'individus  $i$  de même espèce, d'une part une variable à estimer  $y$  (qui au § 2 sera le prix) d'autre part un ensemble (*a priori* hétérogène) de variables explicatives  $x_1, x_2, \dots, x_n$ . On cherche une formule approchée de validité générale (ou formule de régression) :

$$y \approx F(x_1, x_2, \dots, x_n) ;$$

la formule devant pouvoir être appliquée à des individus nouveaux (non inclus dans  $I$ ) et pour lesquels on connaît seulement les variables explicatives  $(x_1, \dots, x_n)$  mais non la variable à expliquer  $y$ .

Afin de s'assurer d'après le seul ensemble  $I$  des individus pour lesquels  $y$  est connu, de la validité potentielle d'une formule  $F$  pour des individus nouveaux, il convient de partager l'ensemble en deux sous-ensembles  $IB$  et  $IE$ . La formule  $F$  sera établie d'après les valeurs connues des variables pour les seuls individus de l'échantillon de base  $IB$  ; la validation se fera sur l'échantillon d'épreuve  $IE$ , en comparant aux valeurs réelles de  $y$  connues pour ces individus, les valeurs estimées  $F(x_1, \dots, x_n)$  calculées en fonction des variables explicatives.

Il serait en effet illusoire de se satisfaire d'une bonne approximation  $y \approx F(x)$ , vérifiée sur les individus eux-mêmes qui ont servi à ajuster la formule  $F$  ; et cela d'autant plus que l'hétérogénéité des variables  $x$  communément disponibles ne nous permet pas seulement d'imaginer un modèle probabiliste paramétrique plausible pour une loi de distribution conjointe de l'ensemble des variables  $(y ; x_1, \dots, x_n)$ ,

L'efficacité d'une méthode est démontrée quand, dans des cas concrets, sur des données réelles, la différence entre  $y$  et  $F(x)$  n'est pas pour les individus de l'échantillon d'épreuve  $IE$  supérieure à ce

(\*) Docteur 3° cycle. Assistant à l'université de Damas.

qu'elle est pour l'échantillon de la base IB.

Quand Card I, nombre des individus disponibles est élevé, il est facile de choisir Card IB et Card IE de telle sorte que simultanément la formule F soit fondée sur une base de données assez large (Card IB élevé) et que la vérification en soit confirmée (Card IE élevé). Quand les informations sont limitées, on doit faire un compromis : comme il n'y a pas, répétons-le, de modèle probabiliste précis, on doit procéder empiriquement par essais successifs ; les histogrammes des erreurs (c'est-à-dire de différence  $(y - F(x))$ , cf. *infra* § 2) servent à fixer notre choix. En général on prendra Card IB > Card IE ; car la validation, qui est l'étude de la distribution d'une variable réelle unique (l'erreur), peut se faire sur moins d'individus que la construction de la formule F qui se fonde sur un ajustement multidimensionnel.

1.2 Type de formule et codage des données : Il est commun d'utiliser en régression des formules linéaires telles que :

$$F = a_1.x_1 + a_2.x_2 + \dots + a_n.x_n ;$$

éventuellement, on améliore l'ajustement en introduisant des termes de degré supérieur à 1. Plusieurs motifs nous détournent d'utiliser de telles formules. D'une part certaines variables explicatives ne sont pas des nombres réels, mais prennent leur valeur dans un ensemble fini : présence ou absence de tel caractère ; marque de tel équipement ; d'autre part, une variable numérique peut changer de signification au passage de certains seuils ; enfin les formules polynomiales comportent un nombre indéterminé de coefficients ; ceci permet sur l'échantillon de base IB un ajustement parfait, mais illusoire en ce qu'il ne vaut pas sur IE ; on doit alors borner le degré etc. .

Dans la suite toutes les variables seront considérées comme autant de questions admettant chacune un ensemble fini de modalités de réponse. Pour cela l'intervalle de variation de toutes variables numériques sera découpé en intervalles successifs choisis au vu de l'histogramme des valeurs. Supposons par exemple qu'on distingue pour une variable x trois modalités successives : xp si  $x \leq 2$  ; xm si  $2 < x < 5$  ; xg si  $5 \leq x$ . On posera par exemple pour un individu i si  $x(i) = 3$  :

$$k(i, xp) = 0 ; k(i, xm) = 1 ; k(i, xg) = 0$$

(parce que  $x(i)$  rentre dans la modalité moyenne xm). En d'autres termes, on peut associer aux trois intervalles successifs correspondant aux modalités xp, xm, xg, leurs fonctions caractéristiques  $\varphi_p$ ,  $\varphi_m$ ,  $\varphi_g$  : ou :

$$\begin{aligned} \varphi_p(x) &= \text{si } x \leq 2 \text{ alors } 1 \text{ sinon } 0 ; \\ \varphi_m(x) &= \text{si } 2 < x < 5 \text{ alors } 1 \text{ sinon } 0 ; \\ \varphi_g(x) &= \text{si } 5 \leq x \text{ alors } 1 \text{ sinon } 0 ; \end{aligned}$$

et l'on a :  $k(i, xp) = \varphi_p(x(i))$  et de même pour xm et xg. De par la formule de transition, les facteurs qu'on calculera pour i seront des combinaisons linéaires des  $\varphi_p(x(i))$ . En définitive, au lieu d'utiliser pour fonctions de base, connue en régression polynomiale, les monômes  $x, x^2, x^3, \dots$  ; on utilisera les fonctions caractéristiques d'intervalles  $\varphi_p(x), \varphi_m(x), \varphi_g(x)$ .

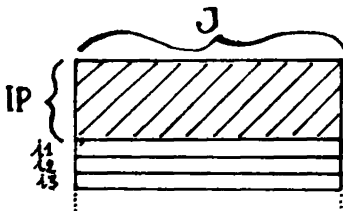
A la limite, en analyse mathématique, il est tout aussi classique d'approcher une fonction par des polynômes (Th. de Weierstrass) ou par des fonctions en escaliers. La question est de savoir laquelle des deux méthodes fournit les meilleures suggestions dans le domaine du fini qui est celui de l'analyse des données réelles.

On peut encore noter que pour résoudre le problème pratique que nous avons en vue, certains préconisent la conception d'un *système expert*: c'est-à-dire en bref un calcul déductif combinant un ensemble de multiples règles simples d'après lesquelles un expert humain estimerait le prix par retouches successives. Tant la collecte des règles de base, que le calcul déductif qui en condamne l'application sont des opérations complexes. Quant à nous, nous ne considérons pas comment le prix est estimé ou se négocie, nous cherchons seulement un plus court chemin mathématique pour passer des variables explicatives  $x$ , à la variable à expliquer  $y$ . En ce qu'elle repose sur un codage de toute variable en  $(0,1)$  (oui - non), cette méthode relève de la logique ; mais en ce qu'elle recourt non à la théorie de la déduction, mais au calcul algébrique usuel le plus simple (faisable sur microordinateur), elle rejoint les méthodes de régression fonctionnelle. Il est clair que seule l'expérimentation (cf. § 2) peut justifier cette position intermédiaire, à première vue séduisante.

1.3 Les tableaux analysés : Notons  $J$  l'ensemble de toutes les modalités des variables explicatives  $x$ , découpées en classes comme on l'a dit au § 1.2. Notons  $IP$  l'ensemble des modalités de la variable à expliquer  $y$  ; ( elle aussi découpée en classes ; éventuellement par la technique de la fenêtre glissante, expliquée au § 2.2). Le tableau principal analysé est le tableau  $IP \times J$  ; avec, quel que soit  $ipr \in IP$  et  $j \in J$  :

$k(ipr, j)$  = nombre des individus de base  $i \in IB$  dont la variable  $y$  (le prix...) rentre dans la modalité  $ipr$ , et dont la description comporte la modalité  $j$  (e.g.  $j$  sera : variable  $x_3$  très faible).

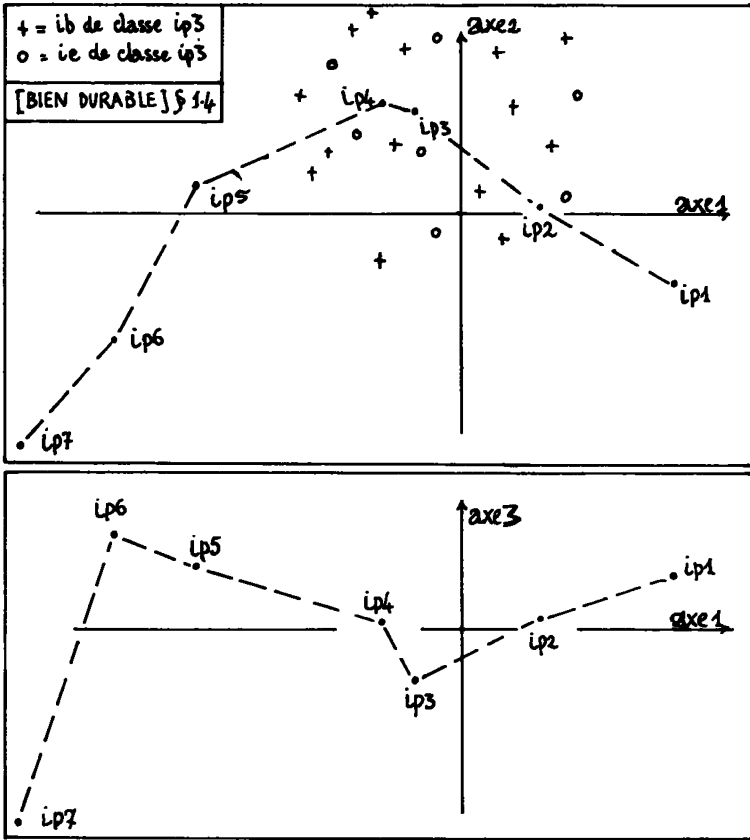
A ce tableau sont adjoints en lignes supplémentaires les vecteurs de description en  $(0,1)$  de tous les individus de l'échantillon de base  $IB$ , ou de l'échantillon d'épreuve  $IE$ .



Le tableau principal  $IP \times J$  est un tableau de contingence ; qui peut être regardé comme un sous-tableau rectangulaire du tableau de Burt  $(IP \cup J) \times (IP \cup J)$ , croisant avec lui-même l'ensemble de toutes les modalités des variables explicatives ou de la variable à expliquer. On peut encore dire que chaque ligne principale  $ipr$  est la somme des lignes supplémentaires  $ib$  décrivant des individus de base pour lesquels la variable  $y$  est comprise dans  $ipr$ . Ce tableau conformément aux principes posés au § 1.1, est construit d'après les seules informations relatives aux individus de l'échantillon de base  $IB$ . Si, comme il est de règle,  $Card IP < Card J$ , l'analyse fournit un nombre de facteurs égal à  $Card IP - 1$  : par exemple 6 facteurs si la variable à expliquer  $y$  est découpée en 7 modalités. C'est par l'intermédiaire de ces facteurs, ou seulement des premiers de ceux-ci, qu'on estimera  $y$ . On voit qu'à la différence de ce qu'est le cas de régression polynômiale, où pour améliorer l'ajustement on est entraîné à accroître sans limite précise le nombre des termes, jusqu'à rendre la régression illusoire (cf. § 1.2), ici le format du même tableau analysé assigne une limite claire à la complexité de l'estimation.

1.4 Résultats d'analyse d'estimation de la variable à expliquer : Ce que nous dirons ici des résultats n'est pas universellement réalisé, comme en vertu d'un théorème mathématique (qui ne peut exister faute d'hypothèses *a priori* acceptables sur la structure des données). Mais ce qu'on obtient dans la pratique, quand (comme ce sera le cas au § 2), l'ensemble des variables explicatives choisies  $x$  suffit à reconstituer approximativement la variable à expliquer  $y$ .

Dans le plan (1,2), l'ensemble IP des modalités de  $y$  se projette



en dessinant une ligne d'allure parabolique ; les modalités étant rangées dans leur ordre naturel. Dans le plan (1,3), la courbe des modalités a la forme d'un S : comme si sur IP le facteur  $F_3$  était approximativement égal à un polynôme de  $o_3$  en  $F_1$  ; etc. . C'est le classique effet Guttman.

L'ensemble  $J$  des modalités des variables explicatives ne peut dans son ensemble être décrit par un schéma aussi simple. Toutefois, certaines variables numériques  $x$  découpées en classes dessinent des lignes voisines de celles des modalités de  $y$ . Ces variables sont précisément celles auxquelles est dévolu le plus grand rôle dans la détermination approchée de  $y$ .

Quant au nuage  $I$  des individus  $i$  (placés en éléments supplémentaires), il se projette dans le plan  $(1,2)$  avec pour zone de densité maxima un croissant encadrant la courbe des modalités de  $y$ . Chaque modalité  $ipr$  de  $y$  (par exemple  $ip3$ ) est exactement au centre de gravité du sous-ensemble des individus  $ib$  de l'échantillon de base  $IB$  pour lesquels  $y$  rentre dans  $ipr$  : ceci résulte de la construction même du tableau principal  $IP \times J$ , dont chaque ligne est somme de lignes du tableau supplémentaire  $IB \times J$  (en  $0,1$ ) décrivant les individus de base. Mais l'important pour nous, est que ce sous-nuage est modérément dispersé autour de son centre ; et que s'y mêlent les individus  $ie$  de l'échantillon d'épreuve rentrant dans la même modalité  $ipr$ .

Bien que les sous-nuages de  $I$  afférents aux modalités successives  $ip1, ip2, ip3$  de la variable à expliquer  $y$  se recouvrent notablement entre eux une méthode d'estimation de  $y$  est immédiatement suggérée par le graphique. Tout individu nouveau  $in$  sera projeté sur le plan  $(1,2)$  et attaché à la modalité  $ipr$  de  $y$  dont il est le plus proche : l'intervalle définissant cette modalité donnera à la fois une estimation de  $y(in)$  et l'ordre de grandeur de la précision espérée. Comme on l'a dit au § 1.1, la validité de cette méthode sera estimée d'après l'échantillon d'épreuve  $IE$ .

1.5 Variantes et améliorations : Dans son application à un problème concret, la méthode esquissée ci-dessus admet en fait de multiples variantes. Il faut choisir des variables pertinentes : des analyses préliminaires signalent les variables qui contribuent le plus à créer les premiers facteurs (particulièrement  $F1$  et  $F2$ ). L'information pertinente elle-même, avant d'être codée en  $(0,1)$  : par exemple s'il  $y$  a une variable de taille (telle que la surface habitable vu le nombre de pièces pour un logement) on pourra diviser par celles-ci certaines des autres variables ; et, en particulier la variable à expliquer elle-même. Il n'est pas en effet équivalent d'appliquer une méthode donnée à l'estimation directe du prix total, ou à celle du prix au mètre carré.

Le découpage de la variable à expliquer joue un rôle critique : finalement, on a choisi d'utiliser simultanément plusieurs découpages dont les classes empiètent entre elles (cf. § 2,2 : Fenêtre glissante).

Au lieu d'attacher un individu nouveau  $in$  au centre  $ipr$  dont il est le plus proche, on peut tenir compte de deux centres  $ipr$  et  $ipr'$  entre lesquels il se projette sur la ligne des modalités ; ou au lieu de centres, chercher les individus de base  $ib$  les plus proches de  $in$  (cf. A. Hathout *in* CAD Vol VII n° 1 pp 19-26 (1983)).

Enfin, pour la clarté de l'exposé géométrique, nous nous sommes placés dans le plan engendré par les deux premiers axes ; où la ligne  $IPR$  a le dessin le plus simple. Mais, à la vérité, c'est dans l'espace engendré par les 15 premiers axes qu'on trouve au § 2 l'approximation la plus satisfaisante ; ainsi que l'atteste l'échantillon d'épreuve.

## 2 Un exemple d'estimation de prix

2.1 Choix des données : Il fallait à notre étude une base concrète. Celle-ci est constituée par la description d'un certain nombre d'individus dont on connaît le prix. Bien que les données soient du domaine public (elles nous ont été communiquées gracieusement par la Société FRETOIL) nous ne les décrivons pas explicitement. Le lecteur pourra concevoir qu'il s'agit de logements, de navires, d'avions ; ... la méthode appliquée ici à l'estimation du prix de vente pouvant d'ailleurs servir pour un rendement agricole, un prix de revient etc. .

La base de données comprend 217 individus ayant fait, au cours d'une même année l'objet d'une vente à un prix publié (prix que nous admettrons être le prix réel...). Selon un choix guidé d'abord par l'usage des experts, on a retenu 9 variables explicatives dont les unes sont des dimensions continues et les autres sont données avec un nombre fini de modalités. Toutes ces variables ont été codées suivant un format homogène en (0,1), comme on l'a expliqué au § 1.2 ; d'où finalement un ensemble J de 63 modalités, représentées sur une figure du § 2.3.

Dans les traitements dont on rend compte ici, l'ensemble I des individus est partagé en un échantillon de base IB (173 individus ≈ 80%) ; et un échantillon d'épreuve JE (44 individus ≈ 20%).

Quant au prix, les analyses ont été faites deux fois : d'une part, en prenant pour variable à expliquer le *prix total* du bien enregistré à la vente ; d'autre part en considérant le *prix unitaire* ou quotient du prix total par la capacité (ou surface), obtenant ainsi le prix du m<sup>3</sup> (ou du m<sup>2</sup>). Dans tous les cas, la variable capacité (ou surface) est prise parmi les variables explicatives, car son rôle est très important.

2.2 Fenêtres glissantes et codage de la variable à expliquer : Cette technique qui fait l'originalité de notre travail est simple et efficace : elle consiste à effectuer plusieurs découpages sur la variable à expliquer, en changeant les limites des classes d'un découpage à l'autre. Dans l'exemple qu'on traite, la variable à expliquer (prix total ou prix unitaire) a été découpée 5 fois en 7 classes. Voici, à titre d'exemple le schéma des limites des classes avec les effectifs des individus qui y sont compris, pour les cinq partitions du prix unitaire.

|         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|
|         | 7       | 12      | 17      | 22      | 27      | 32      |
| IA1(39) | IA2(30) | IA3(49) | IA4(26) | IA5(15) | IA6(8)  | IA7(6)  |
|         | 6       | 11      | 16      | 21      | 26      | 31      |
| IB1(29) | IB2(32) | IB3(47) | IB4(28) | IB5(21) | IB6(10) | IB7(6)  |
|         | 5       | 10      | 15      | 20      | 25      | 30      |
| IC1(26) | IC2(28) | IC3(44) | IC4(34) | IC5(20) | IC6(14) | IC7(7)  |
|         | 4       | 9       | 14      | 19      | 24      | 29      |
| ID1(11) | ID2(34) | ID3(43) | ID4(40) | ID5(21) | ID6(15) | ID7(9)  |
|         | 3       | 8       | 13      | 18      | 23      | 28      |
| IE1(9)  | IE2(35) | IE3(37) | IE4(45) | IE5(24) | IE6(13) | IE7(11) |

[BIEN DURABLE] : § 2.2 : Fenêtre glissante .

Les cinq découpages sont désignés par les lettres A, B, C, D et E. On lit par exemple, sur la quatrième ligne du schéma, que la 3-ème classe ID3 du 4-ème découpage est définie par les bornes 9 et 14 (bornes exprimées dans une unité convenable...), et comprend 43 individus. Les 30 bornes (30 = 5 × 6) des 5 découpages sont disposées ici en progression arithmétique de raison 1. Si l'on met à part les classes initiales et finales (IA1... IE1 ; et IA7... IE7), toutes les classes ont même amplitude et constituent les positions successives d'une "fenêtre glissante" de longueur 5 :

IE2, ID2, IC2, IB2, IA2, IE3, ID3,....., IB6, IA6

Eventuellement on pourra prendre une fenêtre glissante d'effectif constant (i.e. contenant un nombre donné d'individus) et définir le décalage de celle-ci par le nombre des individus qu'elle perd à gauche et gagne à droite en se décalant. Les valeurs choisies importent peu, mais on se guidera sur l'histogramme de la variable à expliquer pour concevoir un codage qui ne comporte pas de classes d'effectif trop faible, (au profil mal défini), ni de décalage trop rapide (imprécision ; cf. *infra*).

Avec ce codage, on ne sort pas du format disjonctif complet, car chaque découpage définit une variable à 7 modalités. Ainsi un individu dont le prix unitaire est 14,5 rentre dans les modalités IA3, IB3, IC3, IA4, IE4 ; sur l'ensemble IP des 35 modalités, il est représenté comme suit :

0010000 0010000 0010000 0001000 0001000

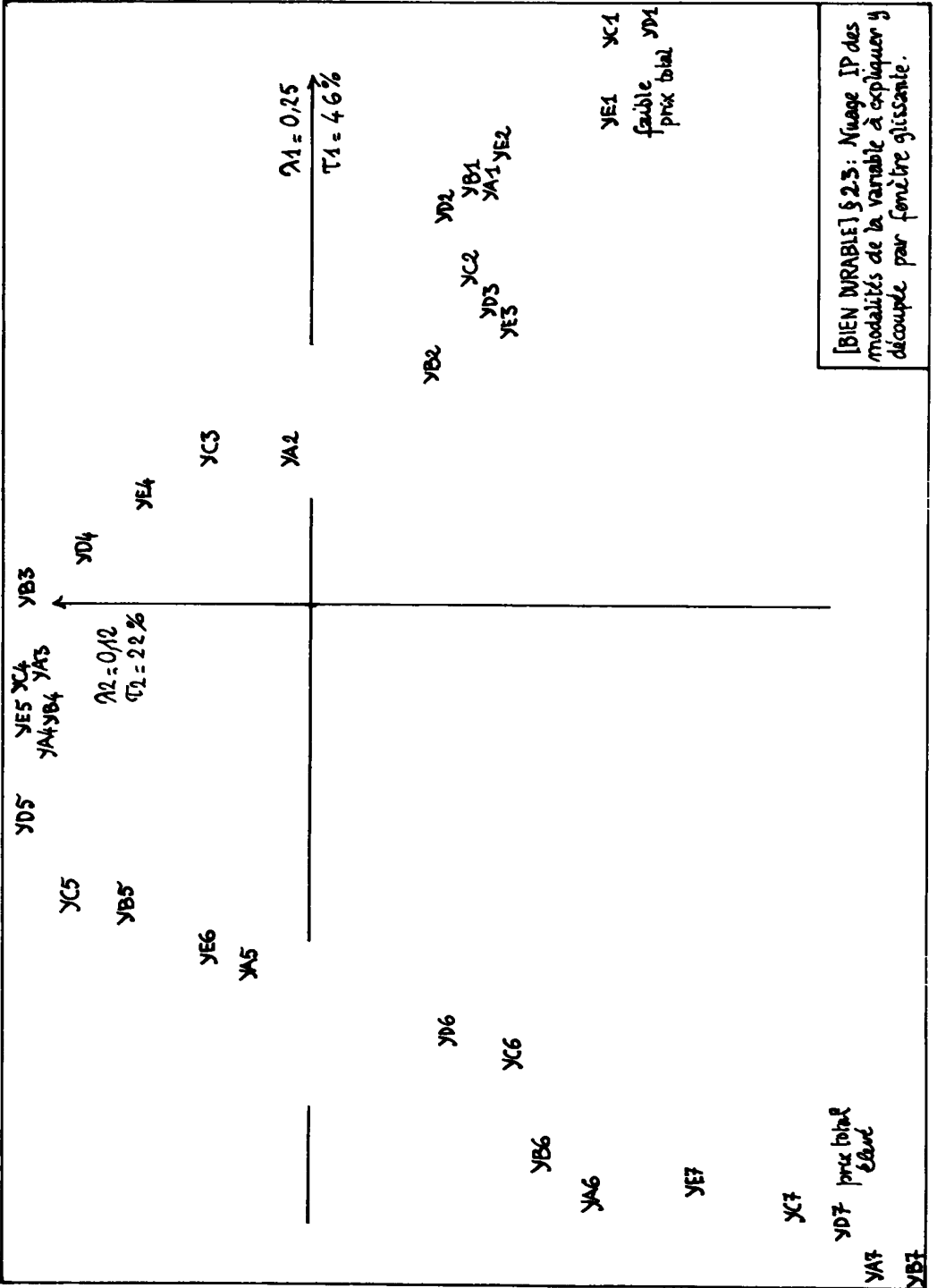
Le tableau principal analysé est la tableau IP x J (35 x 63) ; où  $k(iXr, j)$  = nombre des individus de l'échantillon de base IB rentrant à la fois dans la r-ème modalité du X-ème découpage du prix, et de la modalité descriptive j. Comme dans le cas d'un seul découpage du prix (cf. § 1.3) c'est un rectangle extrait d'un tableau de Burt.

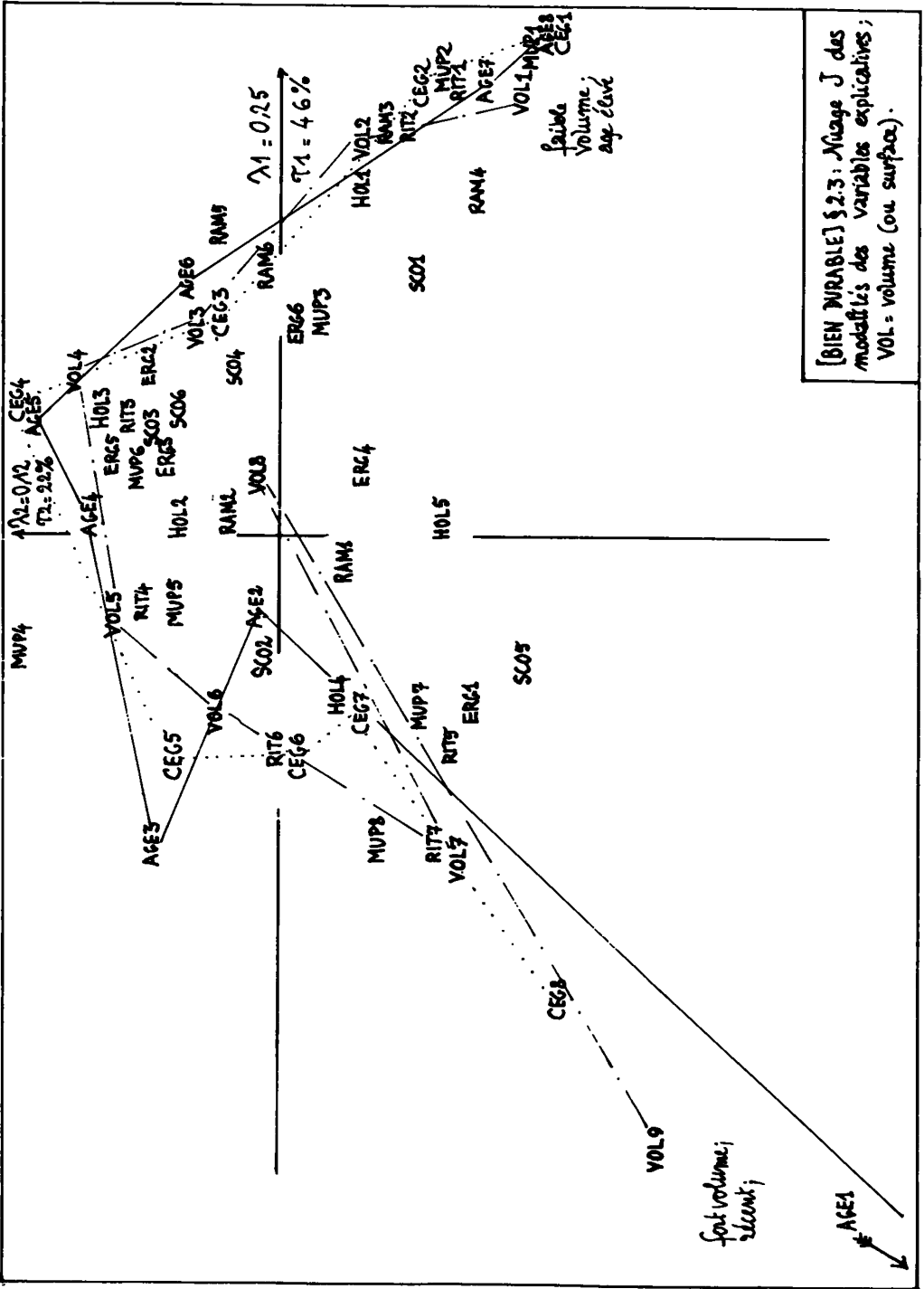
Cette technique permet d'éliminer à la fois : d'une part, l'instabilité inhérente à un découpage unique très fin, avec un grand nombre de classes aux profils peu significatifs, car leurs effectifs sont faibles ; et d'autre part, la perte d'information qu'entraîne un codage grossier où les classes sont larges et peu nombreuses. Ici les classes sont généralement d'effectif assez élevé (plus de 10) (à l'exception de certaines classes extrêmes ; le prix des individus extrêmes étant, de toute manière difficile à estimer) ; et leur largeur ne fait pas perdre d'information, parce que le codage sur IP d'un individu i change chaque fois que l'on traverse une des 30 bornes de l'ensemble des découpages : par exemple, le codage donné ci-dessus pour la valeur de 14,5 vaut seulement entre 14 et 15.

2.3 Résultats d'analyse : Afin d'illustrer par un exemple concret le schéma général d'effet Guttman proposé au § 1.4, on publie le plan (1,2) issu de l'analyse fondée sur des découpages du prix total ; avec sur des pages séparées le nuage N(IP) des 35 modalités du prix total ; et le nuage N(J) des 63 modalités des variables explicatives. Parmi les variables dont la suite des modalités se dispose à peu près régulièrement sur la même ligne que celles du prix total, on ne s'étonnera pas de voir le VOLUME (qui varie dans le même sens que le prix) et l'AGE (qui varie en sens opposé) ; une troisième variable est celle désignée par le sigle MUP.

La liaison entre VOLUME et prix total est *a priori* si forte, qu'on peut craindre qu'elle ne masque des relations plus subtiles rendant compte des différences de prix entre biens de même volume. C'est pourquoi (cf. § 2.1) on a fait des analyses avec pour variable à expliquer le prix unitaire (prix du m<sup>3</sup>). Plus précisément, puisqu'il s'agit d'éliminer le lien entre volume et prix total, on a d'abord tracé la courbe régularisée de la variation du prix total y en fonction du volume v :  $y \approx \varphi(v)$ . Notre intention était de prendre pour variable à expliquer l'écart par rapport à cette relation, c'est-à-dire le quotient  $y/\varphi(v)$ . Mais la fonction  $\varphi(v)$  étant à peu près linéaire, on a simplement pris  $(y/v)$  pour variable à expliquer.

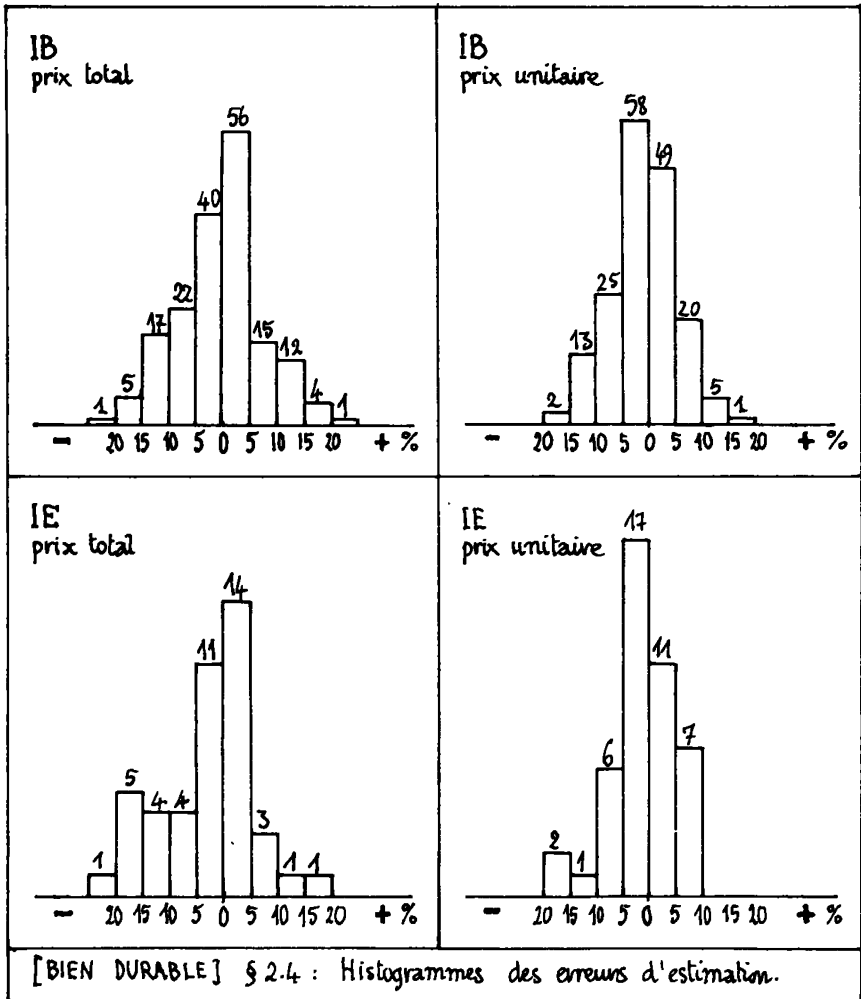






2.4 Résultats de l'estimation du prix : Un individu  $i$  est d'abord affecté à la modalité  $iX_r$  du point représentatif de laquelle il est le plus proche (cf. § 1.4) ; et on donne pour estimation de  $y(i)$  la valeur correspondant au milieu de l'intervalle  $iX_r$  ; cette règle est toutefois à modifier pour les classes extrêmes : on admet alors qu'est affectée à une classe, toute valeur du prix qui est plus proche du milieu de cette classe que de celui d'aucune autre classe. (Une autre possibilité est d'attribuer pour estimation à chaque modalité  $iX_r$  la moyenne des prix des individus  $i$  de IB qui sont rattachés au point  $iX_r$  : l'expérience a montré que les résultats ne diffèrent pas sensiblement selon la voie suivie.

Jusqu'ici on a parlé de "point représentatif" de la modalité  $iX_r$  et de proximité ou de distance de  $i$  à ce point, sans préciser dans quel espace on se plaçait. Les résultats diffèrent grandement selon qu'on se place dans le plan des deux premiers axes, ou un espace engendré par plus de deux axes. Dans le cas présent, avec Card IP = 35,



on a 34 facteurs non triviaux. Ayant commencé par une estimation d'après le plan (1,2), qui ne fournissait de résultats acceptables (erreur < 20%) que dans un peu plus de la moitié des cas, nous avons eu la surprise de voir la précision de l'estimation s'accroître avec le nombre des facteurs pris en compte ; et nous nous sommes arrêtés à une affectation dans l'espace engendré par les 15 premiers axes.

Les histogrammes publiés ici montrent que la précision sur l'échantillon d'épreuve IE n'est pas inférieure à celle obtenue sur l'échantillon de base IB ; ce qui atteste la stabilité de la méthode.

Les résultats sont meilleurs quand on prend pour variable à expliquer le prix unitaire (i.e. le quotient prix/vol, ou prix/surf), qu'ils ne sont avec le prix total. (De toute façon le prix total peut être estimé si on a estimé (prix/vol) en multipliant par vol cette estimation!). Eventuellement, on peut faire la moyenne de deux estimations obtenues par deux voies différentes... Tels quels, les résultats, comparés aux performances des experts, sont parfaitement satisfaisants.

On a tenté d'obtenir un ordre de grandeur de l'erreur commise dans chaque cas d'après la distance (i, iXr) entre l'individu et le centre auquel il a été affecté : il semblait normal que l'erreur fût d'autant plus grande que la distance l'était. Mais nos calculs n'ont pas confirmé l'existence d'une telle relation.

2.5 Evolution temporelle des prix du marché : Les prix des biens durables varient non seulement avec le pouvoir d'achat moyen de l'unité monétaire utilisée (Franc, Dollar...), mais avec l'état du marché des biens de l'espèce considérée. D'ailleurs en renchérissement où une dépréciation peuvent affecter inégalement des biens de taille différente. Présentement, l'appréciation de l'Etat du marché est laissée au jugement des experts, qui observent attentivement les ventes. Or la même technique donnée ici pour estimer le prix d'un bien individuel peut aussi servir à apprécier les tendances du marché. Voici comment.

Supposons qu'on a mis au point un algorithme d'estimation fondé sur une base de données contenant les ventes d'une année. Cet algorithme peut s'appliquer à toute vente nouvelle dont le prix réel est connu. Si le pris est stationnaire, le pourcentage d'erreur affectant l'estimation sera en moyenne nul ; si la moyenne des erreurs affectant un certain nombre de ventes récentes dépasse nettement la moyenne des erreurs obtenues sur des échantillons d'un même nombre de cas extraits de la base des données, on conclura que ces ventes récentes manifestent une tendance du marché. Plus exactement, on fera ces calculs séparément pour 3 ou 4 tranches de taille de biens ; car, comme on l'a dit, la tendance n'est pas uniforme sur l'échelle des tailles.