

C. DESROUSSILLES

## **Taille et structure des entreprises étudiées d'après leurs bilans**

*Les cahiers de l'analyse des données*, tome 5, n° 1 (1980),  
p. 45-63

[http://www.numdam.org/item?id=CAD\\_1980\\_\\_5\\_1\\_45\\_0](http://www.numdam.org/item?id=CAD_1980__5_1_45_0)

© Les cahiers de l'analyse des données, Dunod, 1980, 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/>

## TAILLE ET STRUCTURE DES ENTREPRISES

### ÉTUDIÉES D'APRÈS LEURS BILANS

#### [BILANS]

par C. Desroussilles (1)

Les bilans et comptes des entreprises offrent certes matière à bien des critiques. Tels quels, ils sont cependant pour les experts financiers une source d'information très appréciée. Il s'impose donc d'entreprendre par l'analyse des données, la comparaison systématique de grands ensembles de bilans. Le but de ces analyses étant à la fois de répartir les entreprises suivant l'emploi qu'elles font des compétences techniques, du travail, des machines, des matières, des ressources financières... ; et d'améliorer progressivement les bilans eux-mêmes dont la forme est souvent conçue pour rassurer, dissimuler, satisfaire à des règles contingentes... plus que pour être des instruments fidèles de connaissance et d'action.

On trouve dans la thèse de Madame Desroussilles (3° cycle Paris 1980) d'intéressantes analyses fondées sur une masse déjà respectable d'informations : nous voudrions ici rendre compte de quelques unes de ces analyses ; et plus encore suggérer aux statisticiens et aux économistes les voies d'analyses ultérieures.

Au § 1, nous présentons les données et expliquons comment celles-ci se prêtent à la construction de tableaux de correspondance. Le § 2 étudie les entreprises en prenant des variables principales (valeurs brutes ou modalités) qui expriment plus ou moins directement la taille de l'entreprise. Les 4 analyses du § 3 reposent au contraire sur les modalités des variables de formes classiquement appelées *ratios*. Au 6.3 on rend compte de classifications automatiques effectuées afin de partager les entreprises en grandes - moyennes et petites : division conventionnelle certes, mais qui pour cette raison même pose son problème dont l'étude statistique peut intéresser aux méthodes de l'analyse des données, des spécialistes de l'économie les moins disposés à l'aventure multidimensionnelle.

#### 1 Les données issues des bilans d'entreprise

1.1 Les centrales de bilans : Plusieurs organismes ont mis au point des centrales de bilans constituées par la réunion d'un grand nombre de bilans et de comptes d'exploitation préalablement normalisés, (auxquels s'adjoignent ordinairement des informations telles que l'effectif du personnel ; qui gagneraient selon nous à être complétés pour préciser notamment le rôle des connaissances techniques, l'année sociale etc. ; cf *infra* § 1.3). L'objectif est triplé :

1 Etude de situations et rapports financiers individuels.

2 Constitution d'échantillons significatifs pour divers segments (définis par le secteur, la région, la taille...) intéressant

---

(1) Ingénieur C.N.R.S.

l'analyse économique et financière.

3. Publier diverses études et informations, mises à la disposition des entreprises.

Les trois principales centrales de bilans sont en France celles de la Banque de France, du Crédit National, et de la Caisse des Dépôts / SEDES. L'INSEE qui dispose des données des Bénéficiaires Industriels et Commerciaux (BIC) a depuis Mai 1970 rejoint les trois centrales dans un comité de liaison. La présente étude \* repose sur l'échantillon de l'année 1975, des entreprises de la centrale de la Banque de France ; centrale qui, comptant 25.000 entreprises est présentement, la plus importante, après celle de l'INSEE.

1.2 L'échantillon des entreprises : Un renseignement figure en tête de l'enregistrement relatif à chaque entreprise : le code NAP (nomenclature d'activité et de produits), qui est censé définir l'activité économique principale. Les lecteurs du traité sur l'Analyse des Données (cf TI C n° 12) savent d'après les travaux de M. Volle et de ses collaborateurs quelles difficultés on rencontre dans l'établissement d'une nomenclature ; et c'est d'après les séries mensuelles affectées à une centaine de postes de la nomenclature INSEE, que P. Teillard a étudié l'évolution de la production industrielle française de 1963 à 1975 (cf *Cahiers Vol I n° 4 pp 401-417*).

La nomenclature NAP comprend deux niveaux emboîtés ; le niveau 100 et le niveau 600. Au niveau 600 le code comporte 4 chiffres ; dont les deux premiers définissent le niveau 100 ; c'est ce niveau qu'on utilise exclusivement ici, sans tenter d'en critiquer la structure.

Pour satisfaire aux exigences usuelles de la statistique, un échantillon d'entreprise devrait être constitué d'après une stratification tenant compte à tout le moins du secteur (défini e.g. par le code NAP 100), de la taille (définie, provisoirement du moins, par une variable telle que le chiffre d'affaires ; cf § 2), et peut-être aussi de la région (?). L'échantillon même des 24.350 entreprises figurant au fichier 1975 de la Banque de France, n'a pas été construit suivant une telle stratification ; mais principalement selon la disponibilité des informations. C'est ainsi que pour diverses raisons 86 seulement de 99 secteurs (de la NAP 100) sont représentés au fichier B.D.F. . Pour notre recherche, qui devait comporter de nombreux essais, il était raisonnable de limiter le volume des calculs en ne manipulant qu'un sous-fichier : que l'on a arrêté à 2096 entreprises. Y figurent 70 secteurs ; quant à la taille on a tenu compte à la fois de l'intérêt des petites entreprises quant à leur diversité de structure ; et de l'importance économique des grandes entreprises... Quoiqu'il en soit de notre compromis..., il importe de signaler ici quelle souplesse particulière l'analyse des correspondances offre à la constitution d'échantillons.

Avant tout, il faut décider de ce qu'on entend par *poids* d'un individu. Si l'objet de l'étude était une typologie des directeurs d'entreprise, chaque entreprise devrait recevoir le poids 1 ; dans une étude orientée vers les conditions d'utilisation des compétences de la main-d'oeuvre le poids d'une entreprise serait le nombre de salariés ; dans le domaine économique qui est le nôtre ici, une variable comme le chiffre d'affaires est plus appropriée.

Supposons choisi provisoirement au moins, un poids  $p_i$  pour chaque individu (entreprise). Dans une étude exhaustive, portant sur la totalité des entreprises (ou du moins toutes celles rentrant dans un

---

\* est issue d'une thèse de 3° cycle soutenue en 1979 au laboratoire de statistique.

cadre défini *a priori* : e.g. une région *i* une branche...), il faudra constituer un tableau  $I \times J$  ( $J$  ensemble des variables) dont chaque ligne *i* ait pour poids total  $p_i$ . Ceci sera réalisé automatiquement si le poids  $p_i$  accordé à l'entreprise  $p_i$  est le total d'un bilan comportant pour ensemble de postes l'ensemble  $J$  des variables (avec à l'intersection de la ligne *i* et de la colonne *j*, la valeur  $k(i,j)$  de ce poste : e.g. s'il s'agit du personnel, nombre d'employés de la catégorie *j* dans l'entreprise *i* ; s'il s'agit de finances, frais ou ressources de classe *j* pour l'entreprise *i* etc.). Mais il se peut que pour confronter des informations hétérogènes on doive coder celles-ci sous forme disjonctive complète (en découpant en classes les variables numériques ; et affectant une colonne distincte à chaque modalité des variables qualitatives éventuelles...) : dès lors la ligne décrivant l'individu *i*, est formée de 0 et de 1 ; avec un 1 dans chaque groupe de colonnes afférent à une variable (ou une question) ; et dont un poids total égal pour toute ligne au nombre des variables ; afin de respecter les exigences de pondération, il suffira de multiplier cette ligne par le poids  $p_i$  choisi.

Il est maintenant facile d'alléger par échantillonnage les calculs (d'un coût souvent prohibitif) que requiert un tableau exhaustif. Pour cela on définira en toute liberté

1°) Une stratification de l'ensemble  $I$  : c'est-à-dire une partition  $S$  de  $I$  en strates  $s$  deux à deux non empiétantes.

2°) Un taux de sondage  $t(s)$  propre à chaque strate.

Soit donc  $Card_s$  le nombre des individus de la strate  $s$  ; on traitera seulement les informations relatives à une partie  $e(s)$  de  $s$  : avec  $Card_{e(s)} \approx E(s) Card_s$ , mais on multipliera les lignes conservées par un coefficient  $co(s)$  choisi pour conserver à la strate  $s$  son poids total :

$$co(s) = \Sigma\{p_i | i \in s\} / \Sigma\{p_i | i \in e(s)\}.$$

Eventuellement, le coefficient  $co(s)$  sera estimé approximativement.

C'est le principe d'équivalence distributionnelle qui rend légitimes ces manipulations où l'on s'est seulement astreint à respecter d'une part le *profil* des lignes individuelles et d'autre part le *poids* des strates. Ainsi on pourra choisir un taux de sondage élevé  $t(s)$  pour la strate des petites entreprises, (afin d'en bien détailler la complexité), sans pour autant leur attribuer un poids excédant leur importance économique. Quant aux très grandes entreprises, elles ont en principe un poids considérable : ce qui laisse craindre qu'elles ne définissent à elles-seules les axes de certaines analyses : on pourra être contraint de les mettre en éléments supplémentaires ; ou seulement de les affecter d'un coefficient d'amortissement...

. Eventuellement, on adjoindra au tableau  $I \times J$  un tableau  $S \times J$ , dont chaque ligne est le total des lignes des individus *i* rentrant dans la strate  $S$  (e.g. = ensemble des postes de la nomenclature NAP 100, ou NAP 600). Si le tableau  $S \times J$  est mis en élément supplémentaire, chaque point  $s$  (strate) apparaîtra sur les graphiques au barycentre des individus *i* de  $s$  ; et un graphique partiel permettra d'apprécier l'homogénéité de la strate  $s$ . Parfois on préférera traiter  $S \times J$  en tableau principal ; et  $I \times J$  en tableau supplémentaire.

On n'insistera jamais assez sur l'importance des éléments supplémentaires : c'est par eux que sur un fond général (où le spécialiste s'irrite souvent de retrouver seulement les grandes lignes de

ce qu'il sait) des chapelets ordonnés de points signaleront des relations précises insoupçonnées. Par exemple (cf *infra* § 3) on ne s'étonnera guère de voir de nombreuses entreprises de bâtiment présenter un caractère de petites entreprises : elles le sont en effet ! et cela attirera aussi dans ce sens le point moyen du secteur bâtiment. Mais il reste une question : à taille égale (e.g. à chiffre d'affaire égal ; ou à effectif égal etc.), les entreprises de ce secteur ont-elles plus que celles des autres un comportement économique de petite entreprise ? ou encore l'incidence de la taille sur la structure diffère-t-elle de branche à branche ? Pour préciser cela, il suffit de placer sur les graphiques des points (bt) représentant la moyenne des entreprises d'une branche donnée b rentrant dans une modalité donnée t de taille : on verra alors l'espacement des points bt de telle branche b, plus ou moins espacés entre eux ; chacun en retard ou en avance sur la moyenne générale afférente à la taille t etc. .

Lorsqu'on prétend faire des classifications automatiques, il est crucial de respecter les points relatifs des strates. En effet il est rare que les coupures établies par l'algorithme de classification correspondent à des zones absolument blanches séparant les classes : les coupures tombent (au mieux) dans des zones de faible densité. Alléger une strate, c'est donc créer artificiellement une zone de rupture, suivant laquelle l'algorithme de classification partagera sans doute I !

1.3 Informations disponibles et variables retenues : Pour chaque entreprise, les informations contenues dans le fichier de la Banque de France sont au nombre de 400. Ce sont principalement les postes détaillés du bilan (actif et passif) ; ceux du compte d'exploitation générale (crédit et débit) ; et ceux du compte des pertes et profits. La confiance que méritent ces chiffres est malheureusement inégale ; soit parce que les petites entreprises (agriculteurs, artisans, petits commerçants) ne sont astreints à garder trace que d'un petit nombre d'articles et d'activités ; soit parce que la complexité des opérations financières (sous-traitance ; recours à des entreprises de travail temporaire) empêche de retrouver dans un poste la totalité de ce que désigne son nom (e.g. frais de personnels...) : nous croyons, répétons-le, que l'analyse des données en montrant à l'évidence l'ambiguïté, voire l'absurdité de certains usages, suscitera des remaniements que les spécialistes savent être nécessaires. Ici on ne tentera pas d'analyser l'ensemble de tous les postes financiers ; on traitera seulement de quelques agrégats, usuels pour la plupart ; et de certains rapports de base (ou *ratios*) ; dont voici la définition avec les sigles (à quatre lettres) par lesquels nous les désignerons.

1.3.1 Agrégats calculés à partir des postes du fichier initial : Outre les trois postes *chiffre d'affaire* (CA), *chiffre d'affaire* (CE) et *investissement* (INV) on a considéré les agrégats suivants :

RHT : Ressources hors-taxe = chiffre d'affaire - TVA incluse dans les affaires - produits accessoires + TVA incluse dans les achats.

VA : Valeur ajoutée = ressources hors taxe + stock fin brut - TVA incluse dans les stocks fins - stocks fins provisions + RRRO - stock début brut + stock début provision + TVA incluse dans stocks début - achats marchandises + TVA incluse dans achats - TFSE - transport - déplacements et frais de gestion.

ERE : Excédent brut d'exploitation = valeur ajoutée - salaires - charges sociales - impôts et taxes patentes et taxes annexes - TCA - autres dépôts et taxes + TVA incluse dans affaires - TVA incluse dans achats.

RP : Ressources propres = capital social + réserves + report à nouveau + subvention d'équipement + provisions sur pertes et charges + bénéfices

de l'exercice - pertes de l'exercice. .

AFN : Actif fixe net = frais d'établissement net + immobilisations net + autres valeurs immobilisées net.

RTC : Ressources toutes taxes comprises = chiffre d'affaires + produits accessoires + TFLEM + TVA exclue des affaires.

SM ou SMIG = SMIG annuel  $\times$  effectif.

HSM ou HSMIG = salaires - SMIG.

### 1.3.2 Signification approchée des variables calculées comme des agrégats calculés

Ressources hors taxe : aux produits accessoires près, elle se confond avec le chiffre d'affaire hors taxe ; cette variable correspond à la partie vendue de la production et aux ventes de marchandises en l'état ; les ressources hors taxe demeurent une référence traditionnelle en matière d'appréciation de l'importance de l'activité.

Valeur ajoutée : elle exprime l'apport spécifique de l'entreprise dans la production d'un bien, dans le poids économique de l'entreprise, puisqu'elle se détermine en principe en retirant du chiffre d'affaire, la production stockée et le montant des biens et services consommés.

Excédent brut d'exploitation : il s'obtient essentiellement en soustrayant les charges de personnel de la valeur ajoutée ; c'est la ressource fondamentale que l'entreprise tire régulièrement de son exploitation cyclique.

Ressources propres : l'importance des fonds dont dispose l'entreprise (le capital social augmenté du bénéfice net et des subventions).

Actif fixe net : c'est le montant chiffré de ce que représente le montant des immobilisations, des créances à long terme et des frais d'établissement.

SMIG : détermine le montant de la masse salariale payée par l'entreprise en supposant que tout le monde est payé au SMIG ; on dira qu'il s'agit de la masse salariale minimum.

HSMG : détermine la part des salaires payés en surplus du SMIG. On peut retenir l'idée qu'il représente la rémunération de la qualification du travail.

Ces dernières variables ont été calculées en l'absence de renseignements sur l'effectif et le niveau de rémunération de chacune des catégories professionnelles employées par l'entreprise ; nous y reviendrons (cf § 1.3.5).

### 1.3.3 Rapports et taux

V/E : Rendement apparent de la main-d'oeuvre = valeur ajoutée/effectifs.

TMB : Taux de marge brut d'exploitation = excédent brut d'exploitation/ressources hors taxe.

TEX : Taux d'exportation = chiffre d'affaires à l'exportation/ressources hors taxe.

TVA : Taux de valeur ajoutée = valeur ajoutée/ressources hors taxe.

TND : Taux d'endettement = capitaux empruntés/ressources propres.

TNV : Taux d'investissement = investissements/valeur ajoutée.

PP : Part de la valeur ajoutée consacrée à rémunérer le personnel = salaires + CS/valeur ajoutée.

PSM : Part de la valeur ajoutée consacrée à rémunérer le personnel au SMIG = SMIG + CS/valeur ajoutée.

PASM : Part de la valeur ajoutée consacrée à rémunérer la compétence du personnel = HSMIG + CS/valeur ajoutée.

#### 1.3.4 Signification des rapports et taux (ratios)

Taux de marge brut : mesure la performance industrielle et commerciale.

Taux d'exportation : mesure les ventes à l'étranger relativement au total des ventes.

Taux de valeur ajoutée : importance de la valeur ajoutée dans le montant des ressources.

Taux d'endettement : mesure le degré de dépendance de l'entreprise vis-à-vis des tiers.

Taux d'investissement : importance des besoins financiers, exigé par le cycle d'exploitation.

Part de la valeur ajoutée consacrée à rémunérer le personnel : rémunération relative du travail.

Part de la valeur ajoutée consacrée à payer le personnel au SMIG : part de la rémunération minimum du travail dans la valeur ajoutée.

Part de la valeur ajoutée consacrée à rémunérer la compétence : part de la rémunération de la qualification du travail dans la valeur ajoutée.

1.3.5 Intégration du profil du personnel au bilan financier : Il est certain que la structure du personnel par âge, sexe, niveau de qualification et nationalité nous informe sur le degré de technicité d'une entreprise, sa stabilité, ses potentialités d'expression et de repli par embauche et débauche etc. On pourrait par exemple considérer 3 partitions du personnel : 1° par tranches d'âge - sexe (e.g. femmes de moins de 30 ans ; hommes de 30 à 45 ans etc.) ; 2° par qualification (manoeuvres, ouvriers, maîtrise ; cadres ; éventuellement scindés par sexe) ; 3° par nationalité ou origine régionale. D'où pour chaque entreprise (ou pour chaque branche, strate etc.) une ligne mettant bout à bout ces 3 partitions. Mais il est aussi pénible de convertir en bilan de dépenses, ce recensement humain : au lieu d'inscrire en  $k(i, j)$  le nombre de travailleurs de la classe  $j$  que comporte l'entreprise  $i$ , on inscrira le total des dépenses en salaires versés par l'entreprise  $i$  à des travailleurs de la classe  $j$ . On ne dispose pas de données pour construire de tels tableaux ; tout ce qu'on a pu faire est de distinguer le salaire en deux pentes : SMIG (ce qu'aurait été le salaire si tous les employés étaient au SMIG) et HSMIG (complément de salaire d'autant plus élevé que les cadres sont plus nombreux... ; ou que les ouvriers font plus d'heures...).

## 2 Analyse des variables exprimant la taille de l'entreprise

2.0 A la recherche d'un indice de taille : Ainsi que le notent P.Y.Hémin, B. Paulre et M. Ponchain (rapport CORDES 1976) la variable de revenu

a joué un grand rôle dans les statistiques relatives aux budgets des ménages : en cumulant les budgets observés dans des ménages rentrant dans une même tranche de revenus, on ne simplifie pas seulement les calculs, on estompe les erreurs affectant les données individuelles, et enfin on protège le secret statistique. Or pour les budgets d'entreprises les cumuls par classe de taille seraient également possibles, mais selon les auteurs cités on n'en a pas fait grand usage. Cette remarque révèle les difficultés que rencontrent les économistes quand ils tentent d'agréger les entreprises dans des classes de taille dont ils espèrent qu'elles recouvrent une relative homogénéité de comportement. Assurément, l'unité "entreprise" représente un univers beaucoup plus complexe que l'unité "ménage". Et à supposer que l'on décide de classer les entreprises suivant leur taille, il reste à déterminer l'indice qui mesurera celle-ci : plusieurs des variables introduites au § 1.3.1 (notamment le chiffre d'affaire et la valeur ajoutée ; ainsi que les effectifs du personnel ayant été utilisés pour cela : la taille d'une entreprise n'est certes pas une grandeur unidimensionnelle !

Ce serait d'ailleurs une illusion de croire que le revenu d'un ménage fournisse de son niveau de vie une expression unidimensionnelle valable. Au seuil d'un § consacré à exprimer la taille des entreprises, il convient de signaler à l'horizon des recherches présentes portant sur les budgets d'entreprise, les conceptions auxquelles on est parvenu dans ce domaine sans doute moins ardu de l'analyse des budgets familiaux ; conceptions qui devraient s'imposer aussi aux analystes financiers.

Au terme de leur étude sur la "*Structure des budgets familiaux et des impôts indirects*" (in. *Cahiers* ; Vol IV n° 4 ; pp 443-463), P. Clavier & J.L. Madre écrivent".

"*Tant pour l'étude des consommations que pour celles des taxes, l'analyse de données individuelles devrait apporter une information beaucoup plus riche que celle des données agrégées par C.S.P. publiées par l'INSEE (Soulignons au passage que le critère de taille est déjà vite dépassé...). L'importance d'autres critères socio-démographiques (taille du ménage, région,...) pourrait ainsi apparaître. Il semble cependant que cette analyse de données individuelles soit insuffisante. En effet, les travaux de Nicole Tabard montrent que la tenue imparfaite des carnets de compte (l'analogie avec les comptabilités des petites entreprises est ici évidente ; cf § 1.3 début) est un facteur important de différenciation des données individuelles. Et après avoir mis en doute la représentativité individuelle de certains postes (investissements durables des ménages notamment, les auteurs concluent : Il semble donc que dans un second temps il y ait avantage à agréger les ménages dans des catégories dont l'analyse des données individuelles aura montré l'homogénéité. Il pourrait s'agir par exemple d'un croisement entre le niveau de revenu ou la C.S.P. du chef de ménage et la taille du ménage, ou plutôt de catégories plus complexes (composition du ménage). Ici encore la traduction en terme de budgets d'entreprise s'impose : constituer des catégories croisant la taille de l'entreprise avec sa branche ; ou plutôt tenir compte de la structure du personnel révélée par des analyses telles que celles suggérées au § 1.3.5...*

2.1 Analyse des corrélations entre variables de taille : Proposons d'abord en le commentant brièvement, le tableau des corrélations mutuelles entre neuf variables qui toutes croisent avec la taille de l'entreprise, et seraient donc susceptibles d'exprimer celle-ci...

Les 3 corrélations les plus fortes sont entre *Salaire hors SMIG* et *SMIG × effectif* (ou ce qui revient au même *effectif* !) puis *chiffre d'aff. à l'export.* et *Effectif* ; et *Investissement* et *Salaire hors SMIG* :

Matrice des corrélations entre les variables de taille

Ch. d'af. à l'exp.	CA EX	CA EX	SMIG	INVT	RET	VA	EBE	RP	AFN	HSMIG
Effectif	SMIG	SMIG	0.9747							
Investissements	INVT	0.9431	0.9453							
Ressources h.-taxe	RHT	0.8378	0.8424	0.8579						
Valeur ajoutée	VA	0.9050	0.9153	0.9276	0.9576					
Excédent brut	EBE	0.5769	0.5536	0.6220	0.6315	0.7396				
Ressources propres	RP	0.5281	0.5430	0.6040	0.5727	0.6563	0.7169			
Actif fixe net	AFN	0.6737	0.6932	0.7495	0.7097	0.7793	0.6822	0.9595		
Sal. hors SMIG	HSMIG	0.9533	0.9812	0.9620	0.8604	0.9345	0.5636	0.5840	0.7381	

ce qui confirme que les HSMIG est un bon indicateur de la technicité d'une entreprise. Quant à la forte corrélation des exportations avec les variables de taille traditionnelle, elle ne semble pas pouvoir être extrapolée à l'ensemble des entreprises françaises par rapport auxquelles l'échantillon de la Banque de France comporte un biais en ce qui concerne précisément l'exportation.

Au contraire, on trouve dans le bloc  $2 \times 2$  croisant *Excédent brut d'exploitation* et *Ressources propres* d'une part avec *chiffre à l'exportation* et *Effectif* d'autre part, les trois coefficients inférieurs à 0,56, (et un autre qui n'atteint pas 0,58) : sans être analyste financier on conclura de ces très faibles corrélations que l'aisance de trésorerie d'une entreprise n'a pas avec la taille de celle-ci un parallélisme rigoureux.

On peut poursuivre l'étude des corrélations en diagonalisant la matrice présentée ici : ce qui revient à effectuer une analyse en composantes principales sur les 9 variables normées et centrées. Comme on l'attendait, le premier facteur  $F_1$  est une combinaison linéaire à coefficients positifs de 9 variables. Si les corrélations mutuelles des 9 variables étaient toutes égales et très voisines de 1, les 9 coefficients seraient égaux entre eux : en fait 6 des coefficients vont de 0,32 et 0,34 ; mais le coefficient de AFN est en dessous de 0,2 et ceux de RP et de EBE (variables dont la singularité apparaissait sans analyse) sont à 0,255. Et justement le 2° facteur  $F_2$  oppose les trois variables RP, EBE, et AFN, à tout le reste, particulièrement à CA EX et SMIG (i.e. effectifs).

Tel quel, le premier facteur  $F_1$  peut fournir un indice de taille combinant les variables primaires ; et il est facile à calculer (simple combinaison linéaire). On l'a donc utilisé pour répartir les entreprises de notre échantillon en 8 classes d'égal effectif, par taille croissante (cf § 2.2.1).

Il faut encore noter que l'analyse des variables de taille découpées en classes et codées en (0,1), sous forme disjonctive complète (cf ci-dessous § 2.2) est apte également à fournir un indice de taille d'un calcul très facile et qui a l'avantage de faire intervenir les variables de base sous une forme non linéaire. En effet comme il apparaît dans le travail de C. Sabaton (Pollution : *in Cahiers Vol II n° 1 p.87 1977*), le calcul d'un facteur  $F_\alpha$ , issu d'un tableau en (0,1) se fait très simplement pour un individu supplémentaire par simple *addition* des coefficients afférents aux modalités des variables propres à cet individu :

$$F_{\alpha}(is) = \lambda_{\alpha}^{1/2} \Sigma \{k(is,j)/k(is) G_{\alpha}(j) | j \in J\}$$

$$= \Sigma \{k(is,j) \text{ Coeff}_{\alpha}(j) | j \in J\} ; \text{ où } :$$

$\text{Coeff}_{\alpha}(j) = \lambda_{\alpha}^{1/2} (1/\text{Card}V) G_{\alpha}(j)$ , ( $\text{Card}V$  nombre des variables) ; et où les  $k(is,j)$  à prendre en compte sont seulement ceux qui valent 1 (les autres étant à zéro).

Toutefois, comme on le verra au § 2.2, la valeur ajoutée semble être telle quelle un excellent indice ; qu'on pourra seulement nuancer en considérant la place du point (i) dans le plan  $1 \times 2$  issu de l'analyse des variables de tailles codées en (0,1).

On peut également appliquer ici les méthodes de codage par rang appliquées par D. Maïti dans l'étude des variables fortement liées (*in Cahiers Vol IV n° 3*, pp 266 sqq ; 1979). L'intuition et l'expérience affirment en effet que si l'on fait un tri hiérarchique (i.e. un rangement) des entreprises suivant leurs effectifs, leur chiffre d'affaire, leur valeur ajoutée... la même entreprise n'aura pas le même rang suivant la variable choisie. Ainsi si à chiffre d'affaire égal, on compare une entreprise commerciale à une entreprise industrielle d'après leurs effectifs, la première se place bien avant la deuxième. Réciproquement si l'on se sert du chiffre d'affaire pour classer des entreprises de même effectif, un commerce va arriver bien avant une petite usine. Ceci suggère de calculer relativement à chaque variable  $v$  le rang  $r(i,v)$  de tout individu  $i$  de notre échantillon  $I$ . Plus précisément, le calcul du rang doit incorporer les pondérations ; on posera donc :

$$r(i,v) = \Sigma \{p(i') | i' \in I ; v(i') > v(i)\} / \Sigma \{p(i') | i' \in I\}.$$

(i.e. le rang  $r(i,v)$  est la somme des poids relatifs des individus  $i'$  qui quant à la variable  $v$ , devancent  $i$ ). La taille d'une entreprise serait alors définie par son *rang moyen* relativement à quelques variables de taille. (Ici encore l'insertion d'un individu supplémentaire est facile, si l'on connaît les bornes des quantités, i.e. les valeurs des variables  $v$  correspondant à des valeurs typiques du rang  $r(i,v)$ , e.g. 0,125 ; 0,250 ; 0,375 ... si l'on choisit des octiles). Et comme dans l'article cité de D. Maïti, on peut après la définition du rang moyen caractériser de façon précise tout individu par l'écart (en avance ou en retard) de ses divers rangs relativement à son rang moyen (e.g. cf *supra*, un commerce aurait un rang en chiffre d'affaire en avance sur son rang moyen :  $2(i,CA) < r_{\text{moy}}(i)$  ; et en rang en effectifs en retard :  $r(i,SM/G) > r_{\text{moy}}(i)$  ; le contraire étant vrai pour une usine) ; et faire enfin l'analyse factorielle de ces écarts.

Mais quoiqu'il en soit de ces projets, on rendra compte ci-dessous d'une analyse effectuée sur tableau disjonctif complet codant les variables de taille.

## 2.2 Analyse de variables de taille codées sous forme disjonctive complète

2.2.1 Codage logique des variables : Les analyses dont nous rendrons compte dans la suite portent sur les 11 variables de taille introduites au § 1.3.1 (3 postes de compte ; et 8 agrégats) ; et sur 9 variables de forme (rapports) définis au § 1.3.3. Enfin comme on l'a dit au § 2.1 l'analyse en composantes principales a fourni un premier facteur qui servant d'indice de taille sera ici noté TA.

Toutes ces variables ont été découpées en huit modalités consécutives de poids approximativement égal et numérotées dans l'ordre décroissant de 1 à 8 (e.g. pour la valeur ajoutée, de VA1, très faible

valeur ajoutée à VA8, très forte valeur ajoutée). Toutefois le rapport TEX (taux d'exportation étant nul (ou quasi-nul) pour la moitié environ des entreprises on a créé une modalité très lourde TEX1 pour ce cas ; puis quatre autres modalités numérotées de TEX5 à TEX8 pour exportations réelles (les sigles TEX2, TEX3, TEX4 restant ainsi inemployés...) Notons encore que vu le codage

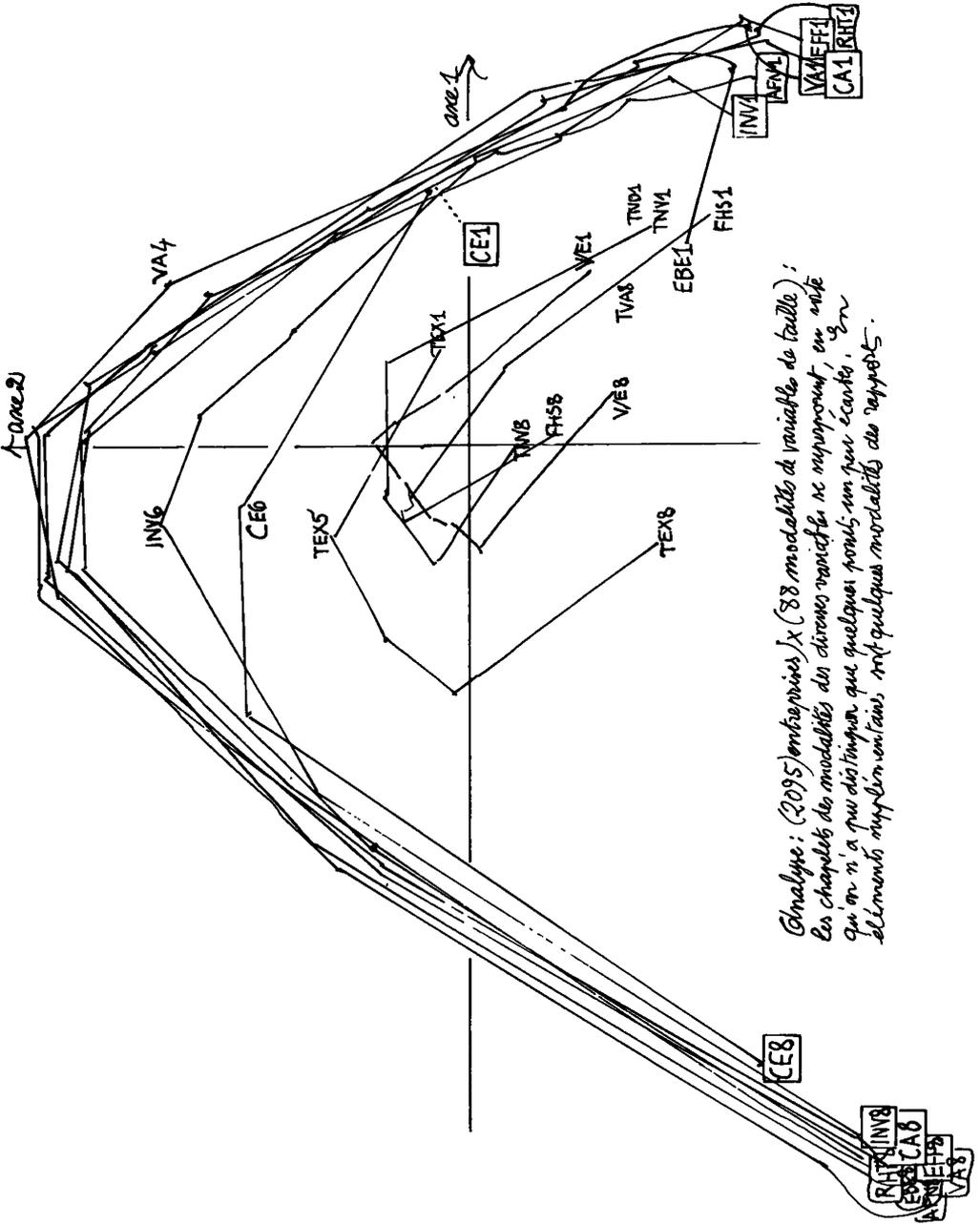
	VA1	VA2	VA3	VA4	VA5	VA6	VA7	VA8
TA1	243	117	1					
TA2	9	128	109					
TA3	4	13	138	51				
TA4	1	1	14	186	49			
TA5	1	1	3	22	187	44		
TA6			1	2	21	191	41	
TA7	2	1		1	6	25	200	26
TA8	1					2	21	233

Tableau donnant à l'intersection de la ligne TA<sub>x</sub> et de la colonne VA<sub>y</sub>, le nombre d'entreprises ayant une taille  $x$  (au sens du § 2.1) et une valeur ajoutée  $y$  (suivant le découpage en classes d'égal effectif)

Dans une première analyse, objet du présent §, figurent en éléments principaux les 88 modalités des 11 variables de taille ; le reste étant en supplémentaire. Le tableau principal a donc 2095 lignes (les entreprises) et 88 colonnes (les modalités), avec à l'intersection de la ligne  $i$  et de la colonne  $j$ ,  $k(i,j) = 1$  si  $i$  possède la modalité  $j$ , et zéro sinon. On sait que l'analyse d'un tel tableau en  $(0,1)$  est équivalente (à un changement d'échelle près sur les facteurs) à celle du tableau de Burt ou tableau de cooccurrence des modalités :

$k(j,j')$  = nombre des individus possédant à la fois les modalités  $j$  et  $j'$  ; (ou plus généralement, si les individus sont affectés de poids, comme on l'a suggéré au § 1.2,  $k(j,j')$  sera la somme des poids des individus possédant  $j$  et  $j'$ ). Ce tableau de Burt mérite d'être examiné en lui-même avant toute analyse, particulièrement dans le cas de variables fortement liées entre elles, où il se compose de blocs quasi-diagonaux (cf D. Maïti *loc. cit. in Cahiers* Vol IV n° 3 pp 261 sqq). Voici par exemple le bloc  $8 \times 8$  croisant les modalités de VA (valeur ajoutée) et TA (indice de taille) calculé au § 2.1

2.2.2 Histogramme des valeurs propres dans l'analyse des variables de taille : On sait que dans l'analyse d'un tableau sous forme disjonctive complète les valeurs propres obtenues ne donnent pas une juste idée des pourcentages d'information afférents aux facteurs : on doit plutôt considérer les v. p. issues du tableau de Burt (lesquelles sont le carré des v. p. issues du tableau en  $(0,1)$ ). Mais de plus, ainsi qu'on l'a récemment constaté, l'histogramme des v. p. (issues du tableau en  $(0,1)$ ) présente généralement une discontinuité autour de  $1/\text{Card}V$  inverse du nombre des variables codées : ici  $\text{Card}V = 11$  ;  $1/\text{Card}V = 0,091$  ; les facteurs interprétables étant ceux pour lesquels  $\lambda_\alpha$  dépasse ce seuil ; avec des postes d'information de l'ordre de  $(\lambda_\alpha (1/\text{Card}V))^2 \dots$  (Pour plus de détail cf *Cahiers* Vol IV n° 3 p. 377, 1979). Si l'on applique ici ce principe on trouve 60% d'inertie pour l'axe 1 ; 25% pour l'axe 2, et moins de 10% pour l'axe 3. On ne considérera donc ici que le plan  $1 \times 2$  ; et cela d'autant plus que le facteur 3 est lié aux premiers par effet Guttman.



*Analyse : (2095) entreprises, x (88 modalités de variables de taille) ;  
 Les chapitres des modalités des divers variables ne regroupent, en note  
 qui on n'a pu distinguer que quelques points un peu écartés. En  
 éliminant simultanément, ont quelques modalités des rapports.*

2.2.3 Place des modalités des variables de taille : Ces modalités dessinent une parabole, dans le plan  $1 \times 2$  ; et les modalités de même rang de toutes les variables sont étroitement groupées, au point qu'on a peine à les placer sur le graphique car elles se superposent. Toutefois les modalités INVx (investissements) et plus encore CEX (chiffre à l'exportation) se placent nettement à l'intérieur de la parabole, du côté des faibles modalités (petite taille). Au contraire la ligne des VAX (valeur ajoutée) semble être de toutes la plus enveloppante. Tandis que les modalités TAX (indice de taille du § 2.1) qui sont en élément supplémentaire, se placent assez près de la périphérie.

2.2.4 Place des modalités des rapports : En bref les modalités intermédiaires (3,4,5,6) de tous les ratios sont au centre du plan (légèrement décalées vers les tailles 6, 7). Seules se déplacent les valeurs extrêmes (en général 8 et 1 ; parfois 2) attirées vers les modalités 1 des variables de taille : ce qui témoigne que les petites entreprises sont susceptibles de présenter des profils très accusés. Le cas du taux d'exportation TEX est particulier. Nous rappelons qu'il n'a que 5 modalités : TEX1 (quasi-nul) TEX5, TEX6, TEX7, TEX8. Les points 5, 6, 7 se suivent régulièrement à l'intérieur de la parabole, ce qui montre une certaine corrélation du taux des exportations avec la taille de l'entreprise ; d'autant plus que TEX1 est sur le 1<sup>o</sup> axe positif (i.e. avec les faibles tailles). Mais le point TEX8, montre qu'à ce niveau joue une conjonction des extrêmes : un tel taux pouvant être le fait non seulement de très grandes entreprises, mais aussi de quelques petites firmes sans doute spécialisées...

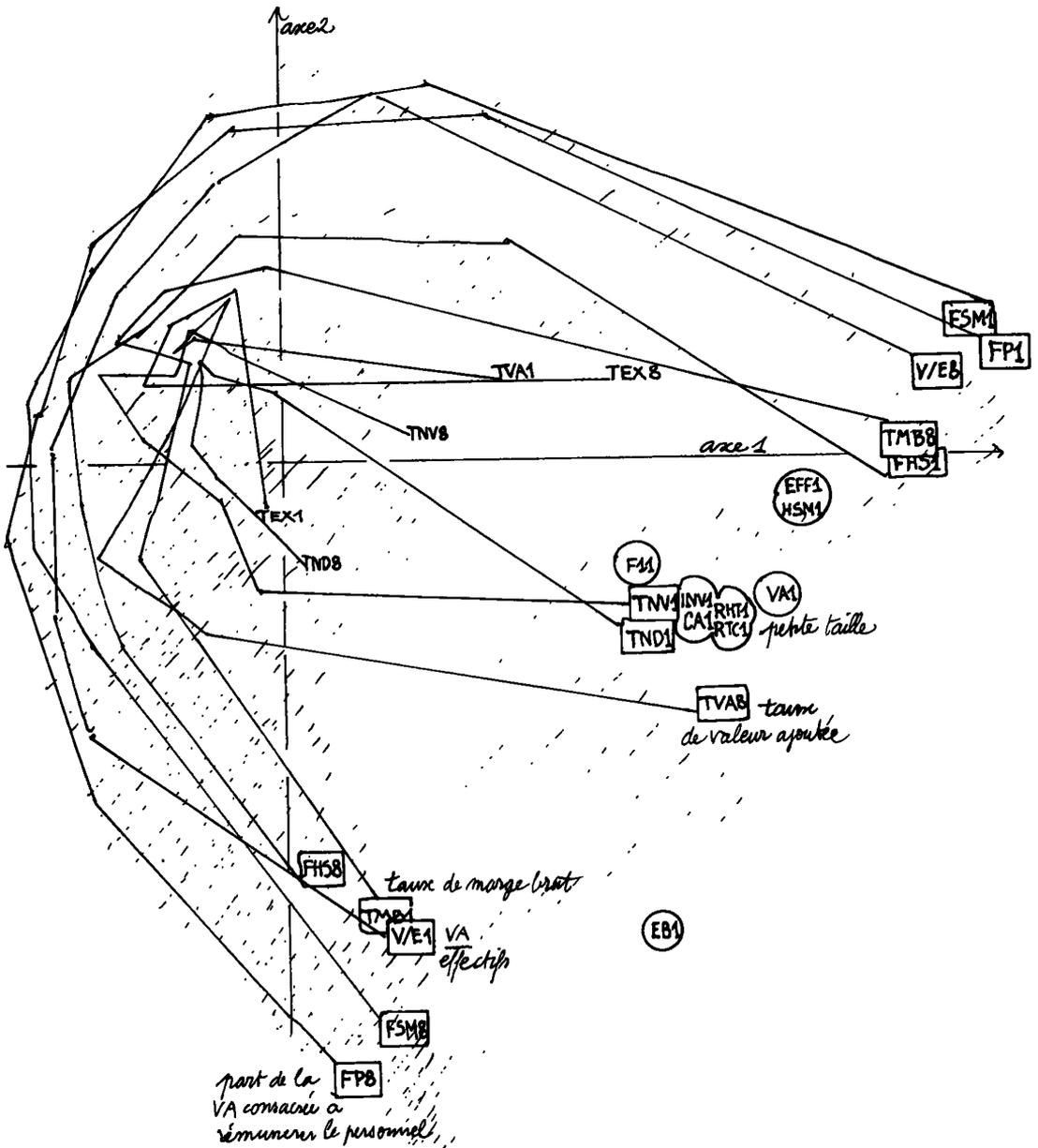
2.2.5 Nuage des entreprises et indices de taille : Ce nuage est nettement parabolique ; toutefois on trouve éparpillés autour de l'origine quelques dizaines de points aberrants, qui participant de modalités hétérogènes ont une place typique. Sur la figure on a suggéré de définir par des classes de taille en partageant le plan par des droites rayonnantes, après avoir éliminé la zone centrale. Ainsi qu'on l'a rappelé au § 2.1, il n'est pas difficile de calculer les coordonnées d'une entreprise si l'on connaît pour elles les valeurs des variables de taille.

Quant aux indices de tailles, on peut dire qu'un indice est d'autant meilleur qu'il définit un découpage en classes aussi resserrées que possible, donc ayant des centres de gravité (points-modalités) aussi excentriques que possible. De ce point de vue, la valeur ajoutée, agrégat certes complexe, apparaît comme le meilleur des indices disponibles : on le conçoit, car comportant à la fois l'excédent brut d'exploitation et les charges de personnel (cf § 1.3.2), il rend compte de deux aspects complémentaires de l'activité d'une entreprise...

En éléments supplémentaires, ont été placés les points moyens (barycentres) des entreprises des divers secteurs : à quelques détails près, la place de ces points apprend peu : on savait en effet *a priori* que les commerces de détails alimentaires sont de petites entreprises ; hormis les commerces de grande surface, que la nomenclature a opportunément mis à part ; etc.

### 3 Analyse des variables de structure

3.0 Il conviendrait de reprendre à la base l'analyse des comptes et bilans des entreprises : l'analyse d'un tableau croisant un ensemble d'entreprises avec un ensemble quasi-exhaustif de postes comptables, doit révéler toute l'information structurelle ordinairement demandée à l'analyse des ratios et la compléter grandement. Toutefois dans la présente étude on se bornera à une étape intermédiaire : analyser simultanément un ensemble de ratios classique codés sous forme booléenne (en 0,1). Cette analyse tentée d'abord sur un ensemble comprenant beaucoup



Analyse (2095 entreprises) x (72 modalités de rapports): les éléments principaux totalisant avec les deux premiers axes des CB<sup>2</sup> de plus de 0,200, sont seuls encadrés d'un rectangle. On a dans un cercle quelques éléments supplémentaires. Des lignes brisées relient les chapiteaux de modalités principales. Le nuage des entreprises est suggéré par des hachures.

de petites entreprises a paru dominé par les valeurs aberrantes des rapports calculés pour beaucoup de celles-ci (§ 3.1). L'analyse a donc été reprise en s'affranchissant des valeurs aberrantes (§ 3.2).

3.1 Première analyse : 2095 entreprises  $\times$  72 modalités de rapports (*ratios*) : On voit sur la figure que les modalités de la plupart des rapports se placent sur une même courbe, dans le plan  $1 \times 2$ . A l'extrémité positive de l'axe 1 sont les modalités définissant une faible part des salaires dans les dépenses de l'entreprise (FP1 ; FSM1 ; FHS1) ; et corrélativement un fort rendement apparent de la main-d'oeuvre (V/E8 ; fort quotient : Val. Aj. sur effectifs ; taux de marge brut : TMB8). A l'extrémité négative de l'axe 2, on trouve les modalités opposées ; les modalités intermédiaires étant pour la plupart dans le quadrant (F2 < 0 ; 0 < F1) au voisinage de l'origine. Mais le taux d'exportation fort (TEX8) a une place à part, à mi-chemin sur le premier axe positif ; et les rapports TND, TNV, TVA sont mal représentés dans le plan  $1 \times 2$ .

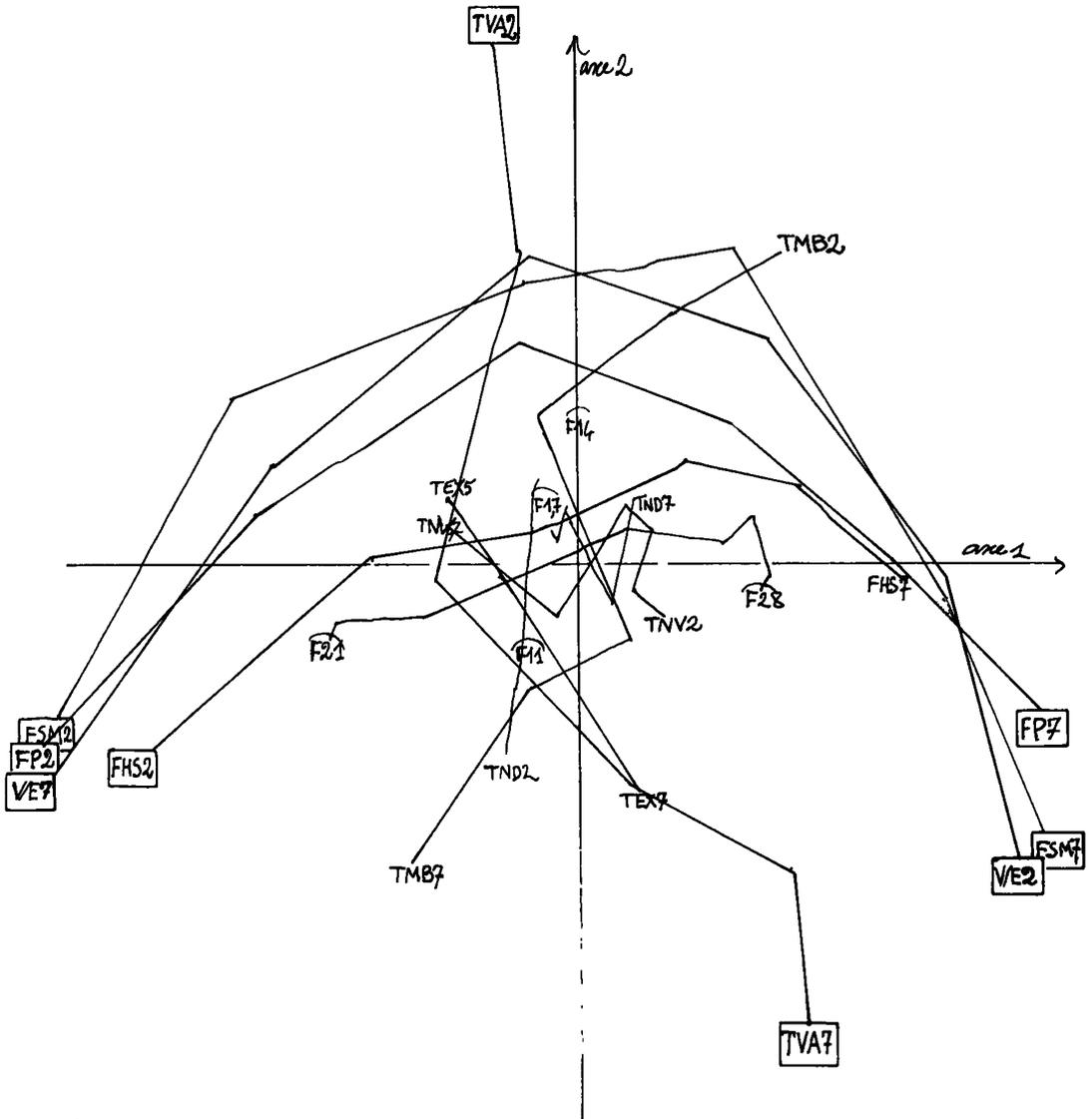
A considérer les listes de corrélations, on voit d'ailleurs que même pour les *ratios* dont les modalités suivent la ligne régulière, seuls les points extrêmes sont en fait bien représentés dans le plan  $1 \times 2$  : ce qui incite à critiquer l'interprétation apparue d'abord. Les modalités de la taille ayant été placées en éléments supplémentaires, sont toutes accumulées au voisinage de l'origine ; à l'exception toutefois des modalités 1 qui sont bien groupées à mi-chemin entre les valeurs extrêmes des rapports FP, VE etc. : ceci indique, comme on l'a pu vérifier, que tant sur l'axe 1 positif que sur l'axe 2 négatif on trouve en position extrême beaucoup d'entreprises de très petite taille. Or souvent de telles entreprises ont une valeur ajoutée ou des ressources hors taxes négatives, voire nulles ; ou encore des effectifs de salaires nuls : par le fait même, l'un ou l'autre des *ratios* ayant un dénominateur nul, est affecté à la modalité 8 même si le numérateur est très faible : finalement pour un tiers environ de très petites entreprises de notre échantillon, la notion de *ratio* n'est pas sûre. C'est par de telles entreprises qu'est exclusivement occupée l'extrémité positive de l'axe 1 ; et on en trouve aussi (mêlées à des entreprises de taille moyenne dont les *ratios* sont fiables) à l'extrémité négative de l'axe 2.

Pour préciser cela, on fait une analyse partielle  $361 \times 72$  (361 entreprises de petite taille)  $\times$  (72 modalités de *ratios*) : le nuage des modalités obtenues dans le plan  $1 \times 2$ , diffère peu de celui issu de l'analyse  $2095 \times 72$  ; et le nuage des 361 entreprises dessine un croissant, (comme le font les 2095 entreprises) : d'après cette analyse on a caractérisé les entreprises aux *ratios* non fiables, pour les éliminer de nos essais ultérieurs de classification (cf § 4).

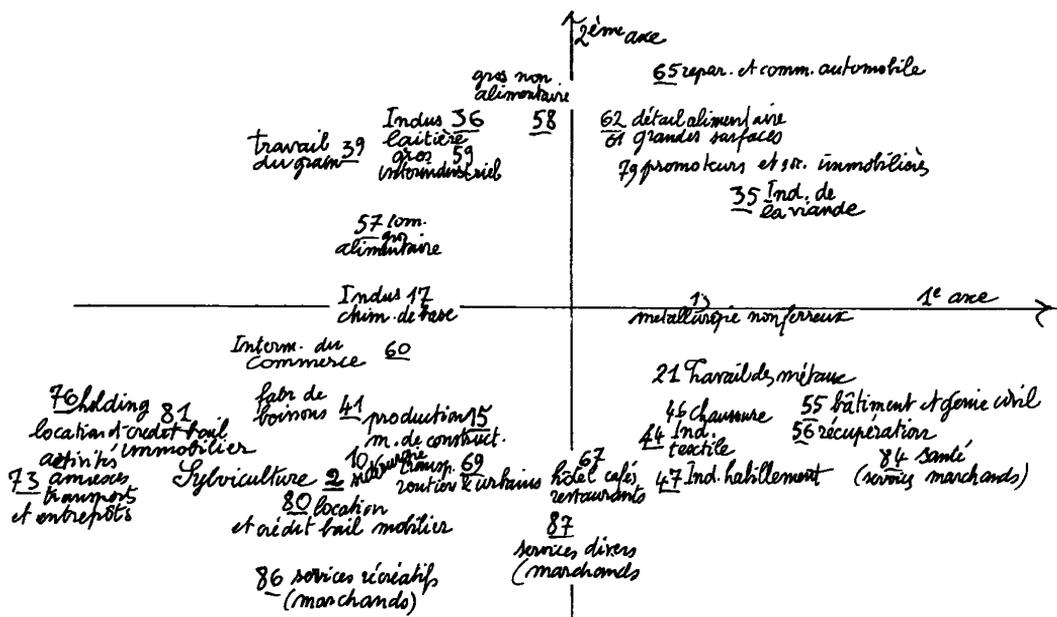
3.2 Deuxième analyse : 2095 entreprises  $\times$  54 modalités de rapports

Il s'impose donc de reprendre l'analyse des *ratios* en s'affranchissant des perturbations causées par les valeurs extrêmes. Le mieux aurait été sans doute d'éliminer les petites entreprises (ou du moins celles qui sont écartées du centre) et de recalculer sans elles les bornes des modalités des rapports (définir à nouveau des classes d'égal effectif). On s'est contenté de mettre en élément supplémentaire les modalités extrêmes (1 et 8) de tous les rapports : ce qui a suffi à faire apparaître une représentation pleinement interprétable non seulement pour le nuage des modalités ; mais aussi pour le nuage des entreprises ; ou, plus exactement, (car il serait difficile de suivre celles-ci une à une...) pour le nuage des centres de gravité des secteurs.

Considérons d'abord le nuage des modalités des rapports : les variables exprimant le taux de frais de personnel (FP, FSM, FHS) décrivent



Analyse: 2095 entreprises) x (54 modalités de rapports): Les modalités extrêmes (1 et 8) sont en éléments supplémentaires. Comme pour l'analyse (2095 x 72), on a tracé les chapelets de modalités, en encadrant celles qui font avec les axes 1 et 2 des cos 2 totalement plus de 0,200.



Analyse (2095 entreprises) (54 modalités de rapports) : quelques points de la nomenclature représentés en éléments supplémentaires. Seuls sont soulignés les nos des points dont  $\cos^2$  avec les axes 1 et 2 totalisent plus de 0,200.

une courbe à peu près parabolique dont la concavité est orientée vers le bas, c'est-à-dire vers la partie négative de l'axe 2 : en sens inverse mais sur la même courbe s'inscrivent les modalités du quotient V/E (valeur ajoutée sur effectifs). Cependant, la variable TVA (taux de valeur ajoutée = valeur ajoutée/ressources hors taxes) se développe dans le sens opposé à l'axe 2 (accompagnée, approximativement de TM3 : taux de marge brut) : toutefois à l'extrémité négative de l'axe 2 (fortes valeurs de TVA) la ligne s'infléchit vers la direction positive de l'axe 1 : le point TVA7 approche des forts taux de salaires. Les valeurs élevées du taux d'exportation (TEX7) vont nettement avec F2 négatif. Les variables TND, TNV endettement et investissements sont mal représentées ce qu'on expliquera par le caractère conjoncturel de ces ratios non directement déterminés par le type d'activité d'entreprise. Quant aux éléments supplémentaires, les variables de taille (sur le graphique figurent seulement les modalités du facteur F1 de l'analyse en composantes principales rapportées au § 2.1) sont maintenant bien resserrées autour de l'origine ; ce qui indique que les facteurs structurels obtenus sont assez peu liés à la taille ; en revanche les modalités du facteur F2 issu de l'analyse du § 2.1, s'alignent régulièrement sur l'axe 1 ; ce qui n'étonnera pas puisqu'il s'agit d'un facteur de structure.

Le nuage des activités des entreprises offre une image particulièrement cohérente\* A l'extrémité positive de l'axe 2 sont les secteurs commerciaux, caractérisés par un faible taux de valeur ajoutée (TVA2) : au commerce, s'associe le travail mécanique du grain et l'industrielleitière, qui ont des rotations rapides de capitaux.

A l'extrémité inférieure gauche du nuage, on trouve des activités qui utilisent relativement peu de main-d'oeuvre mais requièrent des

\* (qu'on comparera aux conjectures de [A. D. ECO.] § 4 in fine: cf Cahiers Vol V n° 1).



investissements lourds de diverse nature : sidérurgie ; mais aussi location d'immeubles ; forêts ; entrepôts ; ou capitaux investis (notre échantillon comporte un *holding*).

A l'extrémité inférieure droite, vont avec les forts taux de main-d'oeuvre les industries du vêtement, le bâtiment... Il est intéressant de noter que le bâtiment figure ici en trois positions, aux trois sommets du nuage :

comme secteur à capital tournant : promoteurs et sociétés immobilières :  $0 < F2 > 0$ .

comme investissement immobilisé : location et crédit bail :  $F2 < 0$  ;  $0 < F1$  ;

comme travail humain : bâtiment proprement dit :  $F2 < 0$  ;  $0 < F1$ .

La place des hôtels, cafés, restaurants est également à noter :  $F2 < 0$  ;  $F1 \approx 0$  ; il y a un important investissement ; et aussi des frais de main-d'oeuvre ce qui équilibre les effets sur l'axe 1 ; l'aspect commercial apparaît secondaire ( $F2 < 0$ ).

#### 4 Correspondance entre structure et taille

On a construit un tableau  $32 \times 72$  (les 32 modalités de quatre variables de taille)  $\times$  (les 72 modalités de 9 variables de structure). Les 9 variables de structure sont celles retenues au § 3 ; les variables de taille sont, parmi celles considérées au § 2 : la valeur ajoutée ; l'effectif, les ressources hors taxe, et l'indice construit au § 2.1. Comme il est naturel on a, à la croisée de la ligne  $i$  et de la colonne  $j$  le nombre des entreprises de notre échantillon rentrant à la fois dans les modalités  $i$  et  $j$  : toutefois un certain nombre d'entreprises de petite taille dont les *ratios* ne sont pas fiables ont été éliminées ainsi qu'on l'a expliqué au § 3.1 (*in fine*). Le tableau a été soumis à l'analyse factorielle et à la classification automatique (C.A.H.).

Un premier résultat visible tant à l'analyse factorielle qu'en C.A.H., est que les modalités de même rang des quatre variables de taille sont généralement groupées (i.e., cf graphiques V2, E2, R2, F2 vont ensemble ; etc.). Plus précisément, on a trois groupes de taille : 1-2-3 (moins de 16 salariés) ; 4-5-6 (de 16 à 116) ; et 7-8 (plus de 116 salariés) ; la frontière entre ces deux derniers groupes étant toutefois moins nette qu'entre les deux premiers. Nous n'insisterons d'ailleurs pas sur ces divisions qui peuvent dépendre du choix des variables et de la composition de l'échantillon. L'accord entre les quatre variables de taille et leur échelonnement régulier dans le plan  $1 \times 2$  est, en revanche, un fait important : il témoigne d'une évolution graduelle de structure en fonction de la taille ; sinon pour les entreprises considérées individuellement, au moins pour les classes de taille (quelle qu'en soit la définition).

Les grandes entreprises se distinguent en ce que parmi elles, on trouve souvent les valeurs les plus élevées du taux d'exploitation TEX. Les plus petites entreprises ont un très faible taux de salaire hors-SMIG/Valeur ajoutée : en fait dans ces entreprises il n'est même pas sûr que la règle du SMIG soit toujours respectée, en sorte que la masse des salaires distribuée au-dessus du SMIG est négligeable, vis-à-vis de la V.A. où rentre aussi le fruit du travail des non salariés (dont le rôle est particulièrement important dans une petite entreprise).

A regarder dans le détail (grâce aux contributions des variables aux subdivisions des noeuds, calculées en C.A.H.) on voit que les entreprises de taille 3 se distinguent de celles de classe 1 et 2 en

ce que celles-là (i.e. taille 3) prennent moins fréquemment les modalités 1 du rapport HSM (salaire hors SMIG/VA), du taux d'endettement et du taux d'investissement : C'est donc que pour la taille 3 existent déjà, des investissements de l'emprunt, un coût de la qualification ; par contre la modalité TVA8 est plus rarement atteinte en classe 3 qu'en classes 1 et 2.

## 5 Conclusion

Au terme de cete étude préliminaire semi-classique (les *ratios* y ont conservé un grand rôle) on peut conclure que comme les bilans, des dépenses domestiques déjà étudiés de près au CREDOC, les bilans et comptes des entreprises se prêtent à de fructueuses analyses. Après la corrélation des variables de taille étudiée au § 2, la répartition des secteurs d'activités selon la structure des comptes et bilans (§ 3.2), et la correspondance entre variable de structure et variables de taille (§ 4) sont clairement appariés. A noter aussi que pour beaucoup de très petites entreprises les calculs de rapport sont illusoi- res (§ 3.1).