

# JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MICHEL POURNIN

## **Utilisation des échantillons des centres de gestion pour l'établissement des revenus régionaux en agriculture**

*Journal de la société statistique de Paris*, tome 108 (1967), p. 245-289

[http://www.numdam.org/item?id=JSFS\\_1967\\_\\_108\\_\\_245\\_0](http://www.numdam.org/item?id=JSFS_1967__108__245_0)

© Société de statistique de Paris, 1967, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

## II

**UTILISATION DES ÉCHANTILLONS DES CENTRES DE GESTION  
POUR L'ÉTABLISSEMENT DES REVENUS RÉGIONAUX  
EN AGRICULTURE**

## INTRODUCTION

*« L'observation du niveau de la rémunération du travail et du capital agricole sera faite par le moyen de comptabilités moyennes d'exploitations représentatives des types d'utilisation du sol, des types d'exploitations et des régions économiques »  
Article 3 de la loi d'orientation agricole du 5 août 1960.*

Cet article de la loi traduit clairement l'attention particulière qui est portée actuellement à l'utilisation des comptabilités d'exploitations agricoles aux fins d'observation économique.

Or depuis plus de dix ans, les Centres de Gestion <sup>(1)</sup> en France établissent les comptes d'exploitation d'un certain nombre d'entreprises agricoles. Ces organismes ont ainsi cherché à apporter aux agriculteurs les éléments de formation et d'information économique dans le but d'améliorer la rentabilité de leur exploitation.

Plus récemment ces comptes d'exploitation ont été utilisés à des fins plus larges et notamment pour obtenir des renseignements sur certaines charges de la branche Agriculture.

Mais la source très importante de statistiques que constituent virtuellement les nombreux comptes d'exploitation accumulés par les Centres de Gestion n'a pas encore été largement exploitée non seulement par manque de moyens financiers et humains mais aussi par manque de méthodes.

Alors que la commission de la Communauté économique européenne demande aux pays membres de créer un « réseau d'information comptable agricole sur les revenus et l'économie des exploitations agricoles » nous avons cherché à savoir si l'échantillon que représentent les comptes d'exploitation disponibles dans un Centre de Gestion pouvait être utilisé à cette fin.

Nous avons donc successivement :

- étudié « l'échantillon » fourni par un Centre de Gestion;
- et cherché à exploiter ce « mauvais échantillon ».

Nous nous sommes efforcés au départ d'appliquer une méthode que M. THIONET a utilisée de 1963 à 1965 sur un échantillon d'entreprises commerciales. Puis nous avons appliqué une méthode dérivée de celle-ci. Enfin nous avons utilisé un calcul de corrélations.

1. Les Centres de Gestion sont des organismes le plus souvent professionnels, régis par la loi de 1961 sur les Associations. Ils sont spécialisés dans les problèmes de gestion des exploitations agricoles.



Nous avons cherché à voir dans quelle mesure, les résultats auxquels nous parvenions pouvaient être généralisés.

Mais dès à présent nous devons signaler que nous nous sommes limités au seul problème de la connaissance des résultats financiers désignés sous le terme de revenus.

Nous avons utilisé les comptes d'exploitation disponibles au Centre de Gestion et d'Économie rurale du Calvados, lequel, créé en 1956, a depuis cette date, accumulé un grand nombre d'informations.

### *Première Partie*

## ÉCHANTILLON ET POPULATION

### I. — OBSERVATIONS PRÉLIMINAIRES

Nous disposons dans le département du Calvados pour l'exercice comptable 1963-1964 de 221 comptes d'exploitation, formant ce que nous appellerons l'échantillon gestion.

Nous nous proposons d'estimer à partir de cet échantillon le revenu global d'une certaine population, celle constituée par les exploitations de polyculture du département.

Nous ne pouvons en effet prétendre appréhender que le revenu des exploitations pratiquant un système représenté dans l'échantillon; cela exclut de notre champ d'investigation les exploitations spécialisées dans les cultures maraîchères, légumières de plein champ, fruitières et florales dont le centre ne suit que de rares unités.

Ces cultures n'occupent d'ailleurs à l'échelon départemental qu'une surface relativement faible : 1 680 ha sur 453 624 d'après l'État VI (1). Éliminer de notre étude ces exploitations ne nous est pas apparu par conséquent comme une grave lacune.

Pour la même raison, les exploitations forestières ont été laissées de côté.

Au sujet de la population étudiée, il faut remarquer encore qu'elle comprend comme chefs d'exploitation des agriculteurs et des non agriculteurs. Par ailleurs la production agricole n'est pas le fait que des seuls exploitants; rappelons les productions sans sol pratiquées par des industriels.

Finalement le revenu global qui pourrait être estimé n'est ni celui de la branche d'activité (produits de l'Agriculture qui englobe tous les producteurs de produits agricoles) ni celui du secteur d'activité qui comprend seulement les producteurs dont l'agriculture est l'activité principale.

Nos ambitions ainsi limitées, nous ferons remarquer que le problème posé est encore intéressant dans la conjoncture économique actuelle avant de passer à l'observation de l'échantillon gestion.

### II. — CARACTÉRISTIQUES DE L'ÉCHANTILLON GESTION

Cet échantillon n'est le résultat ni d'un tirage au hasard ni d'un choix délibéré parmi les exploitations du département. Les agriculteurs viennent adhérer au Centre spontanément montrant par là un souci réel de progrès, d'autres ont adhéré après la visite d'un conseiller qui a obtenu leur nom auprès d'autres organismes professionnels.

1. L'état VI est un document statistique officiel établi jusqu'en 1964 par la Direction des Services agricoles et maintenant par les Directions départementales de l'Agriculture. 453 624 ha est la surface agricole utilisée, bois, forêts et jardins familiaux des non exploitants exclus.

L'échantillon gestion évolue sur les plans technique et économique plus rapidement que l'ensemble de la population. Ceci est le résultat de la collaboration qui s'établit entre l'agriculteur et le conseiller et de l'émulation créée par « les analyses de groupe » qui permettent aux exploitants de comparer leurs résultats à ceux obtenus par d'autres. A titre d'exemples, nous pouvons citer le développement de la production porcine dans le Bocage ou l'intensification de la production laitière dans les 3 régions du Bessin, du Bocage et du Pays d'Auge.

Ces deux « défauts » font qualifier couramment l'échantillon gestion de non représentatif de la population. Ils nous interdisent d'exploiter cet échantillon par les méthodes classiques convenant à un échantillon obtenu par des procédés orthodoxes (tirage au sort ou choix raisonné). Mais ces défauts ne nous semblent pas au départ insurmontables par des méthodes particulières que l'occasion s'offre de mettre au point.

L'échantillon est constitué par un certain nombre d'exploitations par région agricole. Leur répartition par régions et au sein de celles-ci par classe de superficie, telle qu'elle figure dans le tableau ci-dessous, doit être prise en considération. Les limites de classes de superficie retenues pour cette répartition sont celles dont nous disposons pour étudier la population.

*Répartition des exploitations de l'échantillon gestion par régions agricoles et par classes de superficie*

Classes de superficie	Bessin	Bocage	Pays d'Auge	Plaines de Caen-F	Calvados	%
moins de 10 ha	0	0	0	9	0	0
10 à 20 ha	4	14	9	3	30	13,6
20 à 50 ha	20	21	38	23	102	46,2
50 ha et plus	15	6	19	49	89	40,2
<b>TOTAL</b>	<b>39</b>	<b>41</b>	<b>66</b>	<b>75</b>	<b>221</b>	<b>100</b>
<b>Pourcentage</b>	<b>17,7</b>	<b>18,6</b>	<b>29,9</b>	<b>33,8</b>	<b>100</b>	

L'absence d'exploitations de moins de 10 ha est regrettable en raison du grand nombre d'exploitations qui ont moins de 10 ha dans la population.

### III. — CARACTÉRISTIQUES DE LA POPULATION

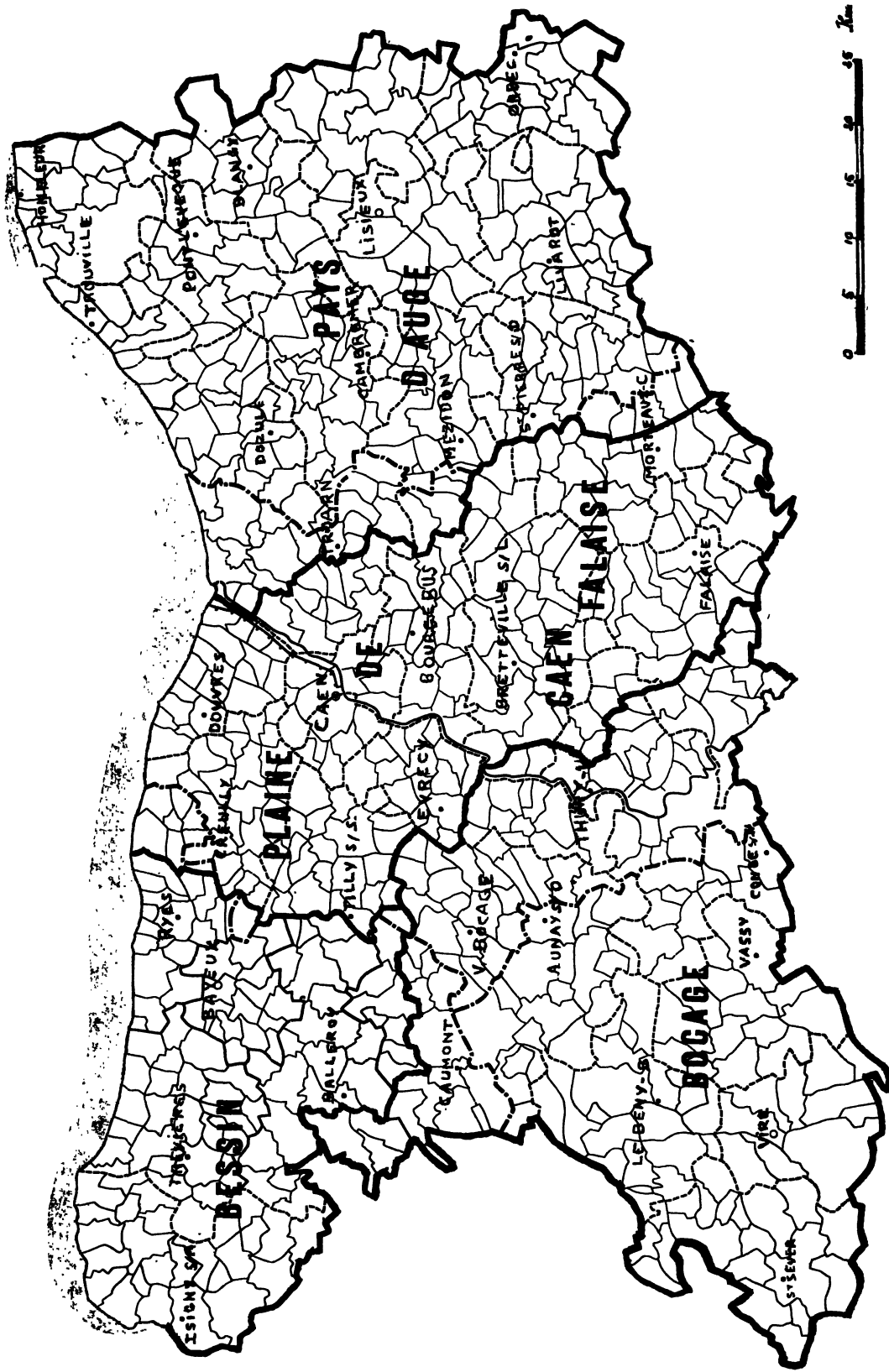
Il n'existe pas de statistiques récentes du nombre d'exploitations départementales par régions agricoles et par classes de superficie.

Mais nous disposons de deux sources d'informations qui nous permettent de les établir approximativement :

- 1° L'enquête « Échantillon Maître » au 1/10 de 1963 du ministère de l'Agriculture.
- 2° Le recensement général agricole de 1955.

L'enquête de 1963 (dite « Échantillon Maître » parce qu'elle doit servir de base de sondage aux enquêtes ultérieures du ministère de l'Agriculture) a porté sur 1/10 des communes rurales. Les communes urbaines et les communes très particulières ont été traitées à part. Elle a comporté :

a) une stratification des communes par région agricole ou regroupement de régions agricoles;



0 5 10 15 20 25 30 Km

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

TOURNEVILLE  
TOURNEVILLE

b) un tirage au sort systématique de 1/10 des communes classées par ordre d'importance (en surface ou en nombre d'exploitations);

c) le recensement de toutes les exploitations agricoles ayant leur siège dans les communes désignées.

Vu le plan et le taux de sondage mis en œuvre, les résultats de cette enquête ne sont considérés comme bons qu'à l'échelon départemental; les échantillons par régions agricoles sont trop petits pour donner des résultats valables à ce niveau.

Le Calvados est l'un des 5 départements (avec la Manche, l'Orne, la Mayenne et l'Indre-et-Loire) dans lesquels le nombre d'exploitations a le moins diminué depuis le recensement de 1955 : moins de 5 %.

Avec 24 382 exploitations en 1955 et 23 330 en 1963 (1) la diminution du nombre d'exploitations agricoles du Calvados est de 4,3 %.

Cette constatation nous a conduit à considérer le recensement de 1955 comme encore utilisable, mais nous avons cherché néanmoins à actualiser le tableau de la répartition des exploitations par régions et par classes de superficie en faisant la synthèse des deux sources ci-dessus.

Pour ceci, nous avons admis que nous pouvions faire les hypothèses suivantes :

1° Les nombres d'exploitations de chaque classe tels qu'ils ressortent de l'enquête de 1963 peuvent être retenus pour l'ensemble du département après comparaisons avec le recensement de 1955.

Voir le tableau ci-dessous.

Classe de superficie	Recensement 1955	Enquête au 1/10 1963
moins de 10 ha. . . . .	11 376	9 870
10 à 20 ha. . . . .	5 675	5 230
20 à 50 ha. . . . .	5 439	6 380
50 ha et plus . . . . .	1 892	1 900
TOTAL . . . . .	24 382	23 330

2° Le nombre total d'exploitations et le nombre d'exploitations de moins de 10 ha par région, correspondant dans tous les cas à un nombre d'exploitations dans l'échantillon de l'enquête supérieur à 100, peuvent être également retenus.

3° Pour les exploitations de 50 ha et plus, peu nombreuses et dont le nombre a probablement peu changé, nous avons retenu les effectifs du recensement.

4° Ce sont surtout des petites exploitations de moins de 10 ha qui ont disparu, venant grossir le nombre d'exploitations de 20 à 50 ha comme le montre la comparaison ci-dessus.

Nous avons retenu comme nombre d'exploitations de 20 à 50 ha par région, les effectifs du recensement corrigés par le pourcentage d'augmentation départemental d'exploitations de cette classe de superficie.

5° Le nombre d'exploitations de 10 à 20 ha par région a été obtenu par différence. Nous pouvons retenir les résultats portés dans le tableau ci-dessous.

1. Ces chiffres sont tirés de la brochure blanche Statistique agricole 1963, Région de programme Haute et Basse Normandie du ministère de l'Agriculture, page 6. Ce sont les résultats provisoires de l'enquête Échantillon Maître au 1/10<sup>e</sup>. Les résultats définitifs connus plus tard ont donné 23 100 exploitations soit une diminution de 5,3 %.

*Répartition des exploitations par régions agricoles et par classes de superficie*

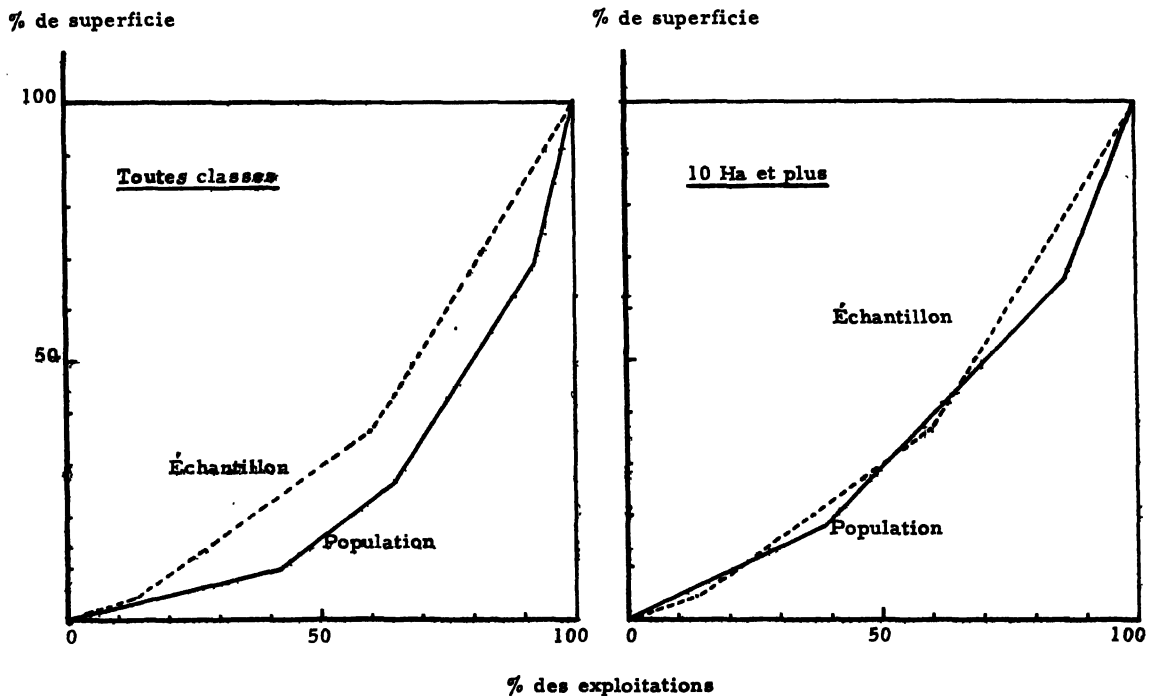
Classes de superficie	Bessin	Bocage	Pays d'Auge	Plaines	Calvados	%
moins de 10 ha . . . . .	1 290	4 170	3 180	1 230	9 870	42,3
10 à 20 ha . . . . .	810	2 080	1 820	520	5 230	22,4
20 à 50 ha . . . . .	1 010	1 900	2 230	1 190	6 330	27,1
50 ha et plus . . . . .	320	270	570	740	1 900	8,2
TOTAL . . . . .	3 430	8 420	7 800	3 680	23 330	100

IV. — COMPARAISON ÉCHANTILLON — POPULATION

La population est composée de 23 330 exploitations; l'échantillon gestion de 221. L'échantillon correspond à un taux de sondage de 1 % : c'est assurément faible et ce, d'autant plus que le milieu agricole étudié n'est pas homogène. Mais ce taux ne peut que croître, le nombre d'exploitations du département diminuant et le nombre d'adhésions au Centre de Gestion augmentant à mesure que les années passent.

La composition de l'échantillon par classe de superficie, comparée à celle de la population, s'avère plus gênante.

Si nous comparons la superficie occupée par les exploitations d'une certaine superficie dans la population et dans l'échantillon, nous obtenons les courbes de Lorenz ci-dessous :



L'échantillon gestion est constitué surtout de grandes exploitations. C'est un avantage dans la mesure où, cherchant à déterminer le revenu global d'une certaine population, il est intéressant de mieux connaître la tranche de population sur laquelle les plus grosses erreurs pourraient être faites.

Mais l'absence des entreprises de moins de 10 ha est un grave défaut, étant donné leur importance numérique.

Encore pourrions nous faire remarquer qu'elles présentent un système d'exploitation simple : surface toute en herbe, matériel excessivement réduit, pas de main d'œuvre salariée. Il suffirait donc de peu de renseignements statistiques sur cette classe de superficie (cheptel vif et rendement laitier essentiellement) pour qu'un conseiller de gestion du département puisse établir un compte type pour ces exploitations et parvenir ainsi à une évaluation approximative du revenu global de cette fraction de la population.

Notons que ces entreprises de moins de 10 ha représentent 42,3 % du nombre d'exploitations mais seulement 10,5 % de la superficie exploitée. La méthode du compte type pourrait par conséquent être admissible, dans la mesure où il s'agirait de déterminer un terme correctif à l'estimation du revenu global des exploitations de 10 ha et plus <sup>(1)</sup>.

Mais nous avons préféré la solution qui consiste à éliminer de notre estimation les exploitations de moins de 10 ha.

Compte tenu de cette élimination les taux de sondage par région agricole de l'échantillon gestion sont les suivants :

Région agricole	Nombre d'exploitation de 10 ha et plus		Taux de sondage
	dans l'échantillon	dans la population	
Bessin . . . . .	39	2 140	1,82
Bocage . . . . .	41	4 240	0,97
Pays d'Auge . . . . .	66	4 620	1,48
Plaines . . . . .	75	2 450	3,07

Vu ces taux de sondage très différents,

Vu le très grand intérêt par ailleurs de pouvoir disposer d'une estimation du revenu par région agricole, nous avons décidé d'essayer la méthode de M. THIONET non pas à l'échelle du département mais à celle de la région agricole.

Nous avons retenu pour cet essai la région du Bessin, pour les raisons suivantes :

1<sup>o</sup> C'est une région géographiquement bien délimitée, homogène, présentant le taux de sondage le plus élevé après les plaines, région agricole hétérogène regroupant la plaine de Caen orientée vers les cultures industrielles et les céréales et la plaine de Falaise orientée vers les productions animales et les céréales.

2<sup>o</sup> Nous jugeons utile de connaître personnellement les exploitations (ce qui est le cas de celles du Bessin) pour interpréter les résultats des méthodes que nous désirons tester.

## Deuxième Partie

### PREMIÈRE MÉTHODE D'ESTIMATION MÉTHODE INITIALE DES POIDS

#### I. — EXPOSÉ DU PROBLÈME

Nous avons vu que les adhésions des agriculteurs au Centre de Gestion ne sont pas le fait du hasard. Mais nous pouvons imaginer que ces entreprises dont la comptabilité

1. Ce que nous venons d'écrire sur les exploitations de moins de 10 ha du Calvados n'est pas vrai dans tous les départements. Il faut bien distinguer dans notre méthode ce qui est propre au cas traité avant même de penser à la généraliser.



est confiée à cet organisme, auraient été tirées au sort avec des probabilités  $q_j$  inégales, en fait inaccessibles. Ces probabilités seraient d'autant plus grandes que l'exploitant cherche une meilleure rentabilité de son entreprise.

Si nous désignons par  $x_j$  le revenu de l'exploitation  $j$ , le revenu global des 2 140 exploitations du Besson de 10 ha et plus est  $\sum_{j=1}^{2140} x_j$ .

Nous supposons que ce revenu peut être estimé par l'estimateur linéaire  $\sum_{j=1}^{39} p_j x_j$ ,  $p_j$  étant un poids affecté à l'entreprise  $j$ .

Ce poids est d'autant plus grand que l'exploitation est « représentative » d'un plus grand nombre d'individus de la population.

Le problème consiste à déterminer les 39 poids  $p_j$  à affecter aux 39 entreprises de l'échantillon.

Ces poids  $p_j$  doivent être positifs puisqu'ils correspondent à l'inverse d'une probabilité. En toute rigueur, ils doivent même être supérieurs ou égaux à 1 car l'exploitation tirée représente au moins elle-même.

Ces poids  $p_j$  doivent respecter certaines contraintes linéaires

a) nous désirons que leur somme par classe de superficie soit égale au nombre d'entreprises par classe dans la population;

b) nous désirons que, pour un certain nombre de variables, la somme (toutes classes) des produits de la valeur de la variable dans chaque exploitation par le poids affecté à celle-ci, soit égale à la valeur de la variable telle qu'elle ressort pour l'ensemble de la population des données statistiques disponibles.

Nous cherchons à déterminer les poids  $p_j$  liés par ces contraintes, de façon qu'une certaine fonction de ceux-ci, caractérisant la qualité de l'estimation soit minimum. Pour simplifier les calculs, nous supposons que c'est la somme des carrés des poids qui est minimum :

$$\sum p_j^2 \text{ minimum}$$

Dans son étude antérieure, M. THIONET a essayé plusieurs autres types de fonctions quadratiques qui n'ont pas semblé donner de meilleurs résultats.

## II. — NOTATIONS

Désignons par :

$c = 1, 2, 3$ , le numéro des 3 classes de superficie : 10 à 20, 20 à 50, 50 ha et plus.

$i = 1, 2, \dots$  le numéro de l'entreprise dans la classe  $c$ .

$N_c$ , le nombre d'entreprises dans la classe  $C$  de la population.

$n_c$ , le nombre d'entreprises dans classe  $c$  de l'échantillon.

$x_{ctk}$ , la valeur de la variable  $x_k$  pour l'entreprise  $i$  de la classe  $c$ .

$\bar{x}_{ck}$ , la valeur moyenne de la variable  $x_k$  dans la classe  $c$  de l'échantillon.

$Y_k$ , la valeur de la variable  $x_k$  pour l'ensemble des 3 classes, telle qu'elle ressort des données statistiques disponibles.

$p_j = p_{ct}$  les poids à déterminer.

Les contraintes formulées précédemment s'écrivent :

a)  $\sum_t p_{ct} = N_c$  et  $\sum_c \sum_t p_{ct} = N$  avec  $N = \sum_c N_c$

b)  $\sum p_{ct} x_{ctk} = Y_k$

## III. — ESTIMATEUR SOUMIS A L'EXPÉRIENCE

Si l'échantillon était tiré au hasard avec d'égales probabilités et était de ce fait représentatif de la population, les poids  $p_j$  seraient par classes les mêmes pour toutes les entreprises et égaux à  $p_{ct} = g_c = \frac{N_c}{n_c}$ , c'est-à-dire à l'inverse du taux de sondage dans la classe.

Ce n'est pas le cas ici, nous allons soumettre à l'expérience l'estimateur  $p_{ct} = \frac{N_c}{n_c} + \sum_k (x_{ctk} - \bar{x}_{ck}) \alpha_k$ , en désignant par  $\alpha_k$  les multiplicateurs de Lagrange correspondants aux contraintes linéaires  $b$ ) quand on rend  $\sum p_j^2$  minimum.

La contrainte linéaire  $\sum p_{ct} x_{ctk} = Y_k$  s'écrit, en remplaçant  $p_{ct}$  pour son estimateur :

$$\begin{aligned} \sum_c \sum_t p_{ct} x_{ctk} &= \sum_c \sum_t \frac{N_c}{n_c} x_{ctk} + \sum_c \sum_t \sum_k (x_{ctk} - \bar{x}_{ck}) \alpha_k x_{ctk}' \\ &= \sum_c N_c \bar{x}_{ck} + \sum_k \alpha_k \sum_c \sum_t (x_{ctk} - \bar{x}_{ck}) x_{ctk}' \\ \sum_k \alpha_k \sum_c \sum_t (x_{ctk} - \bar{x}_{ck}) x_{ctk}' &= Y_k - \sum_c N_c \bar{x}_{ck} = S_k \end{aligned}$$

$S_k$  désignant le second membre, nous obtenons un système de  $k$  équations à  $k$  inconnues permettant de déterminer les  $\alpha_k$ .

IV. — CHOIX DES VARIABLES  $x_k$ 

Notre choix est guidé par les considérations suivantes :

1° Le nombre  $K$  de variables utilisées pourrait être élevé du moment qu'on inverse la matrice ( $K \times K$ ) à la machine. Mais nous nous sommes limités à 4 ou 5 variables pour que les calculs ne soient pas trop compliqués et que la méthode, si elle s'avère donner des résultats, soit applicable dans la pratique. Dans le cas contraire, un petit nombre de variables permettra d'y voir plus clair.

2° Les variables doivent être des grandeurs caractérisant les exploitations et avoir une influence sur le revenu de celle-ci.

3° Il nous paraît souhaitable que les mêmes variables puissent être utilisées, si cela est possible, pour les 4 régions agricoles du département afin d'obtenir des résultats plus comparables entre régions. Mais ce n'est pas impératif dans la mesure où une variable, la production de betteraves à sucre par exemple, peut tenir une place très importante dans une région justifiant pour celle-ci d'être retenue; alors qu'elle est nulle dans d'autres régions.

4° La valeur globale de ces variables pour la population des exploitations de 10 ha et plus doit être obtenue à l'échelon des régions agricoles. Or actuellement les statistiques agricoles sont encore rarement dressées à ce niveau; il nous faut pouvoir les rétablir, avec une précision suffisante.

Nous signalons tout de suite que cette situation a été très contraignante.

Nous avons retenu finalement les 5 variables suivantes qui seront définies en détail plus loin :

- $x_0$  la surface agricole utile (SAU) des exploitations.
- $x_1$  le nombre de bovins présents au 1<sup>er</sup> mai 1963.
- $x_2$  la quantité de blé commercialisée durant la campagne 1963-1964.
- $x_3$  le nombre de tracteurs au 1<sup>er</sup> mai 1964.
- $x_4$  le nombre d'unités travailleurs salariés, employées en 1963.

Schématiquement nous pouvons considérer que :

- les variables  $x_0$ ,  $x_3$  et  $x_4$  caractérisent la structure des exploitations,
- la variable  $x_2$  la production végétale commercialisée,
- la variable  $x_1$  la production animale.

Dans toutes les régions du département, les charges de main-d'œuvre et matériel réunies représentent 40 à 60 % des charges globales.

Le blé est la culture qui occupe, parmi les cultures de ventes, la plus grande superficie. Son importance reflète par le biais des précédents à blé <sup>(1)</sup>, celle des cultures de vente ou des cultures fourragères suivant les régions.

Les productions bovines représentent plus des 2/3 de la production finale des exploitations sur les 9/10 de la superficie du département.

Ce sont les productions les plus développées du département où d'après l'enquête « Bovine 1963 » <sup>(2)</sup> du ministère de l'Agriculture 21 100 exploitations sur 23 100 ont des bovins.

## V. — LES VARIABLES $x_k$

$x_0$  la superficie agricole utile

On appelle surface agricole utile la surface, bois et territoire non agricole non compris (cour, bâtiments, chemins, mares).

a) Dans l'échantillon

Il suffit de lire sur les fiches d'exploitation <sup>(3)</sup> la superficie de celles-ci.

Nous avons retenu la superficie en début d'exercice soit au 1<sup>er</sup> mai 1963.

b) Dans la population

La superficie agricole du Bessin a été déterminée par le Service du Cadastre lors de sa première révision quinquennale au 1<sup>er</sup> janvier 1963. Elle est de 71 675 ha <sup>(4)</sup>.

Il faut déduire de cette superficie la fraction exploitée par les exploitations de moins de 10 ha. Nous avons retenu pour celle-ci le chiffre obtenu par l'enquête EM au 1/10 déjà

1. Le blé est cultivé après certaines cultures, betteraves, maïs, pommes de terre, lin, colza, cultures fourragères, appelées précédents à blé. Une exploitation peut faire d'autant plus de blé qu'elle dispose de plus de précédents.

2. Voir plus loin la note au bas de la page 255.

3. La fiche d'exploitation est le document établi par le centre de Gestion pour chaque exploitation. Ce document rappelle les principales caractéristiques de l'exploitation : moyens de production, résultats techniques et économiques et compte d'exploitation de l'exercice.

4. Cette source nous a paru plus précise pour le Bessin que l'enquête EM (qui a donné 67 234 ha) pour les raisons suivantes :

Le Bessin est une petite région (115 communes) et le taux de sondage de l'enquête EM n'assure pas une grande précision dans ce cas. Ceci nous paraît moins lourd de conséquences quand il s'agit de déterminer un terme correctif. Par contre la superficie indiquée par le Service du Cadastre correspond à une révision exhaustive des baux.

Le Service du Cadastre a utilisé la notion de surface agricole utile comme l'enquête EM, en excluant les landes et friches non productives et la surface bâtiments et cours. Par contre la superficie indiquée par le Cadastre comprend les terres hors exploitation que l'on devrait théoriquement exclure de notre superficie. Nous ne l'avons pas fait, en pensant que ce nombre d'hectares bien que, inconnu, devait être assez faible.

Récemment, nous avons pu obtenir les 3 renseignements suivants :

- la surface totale des exploitations d'après le Recensement agricole de 1955, y compris les landes, friches non productives, les bâtiments et les cours était de 449 664 ha ;
- les inventaires communaux incluant les terres hors exploitation à la même époque était de 455 161 ha ;
- l'enquête EM a donné pour les rubriques 13 « landes et friches non productives » et 14 « sols bâtiments et cours » 6 471 ha.

Nous pouvons en déduire que les terres hors exploitation représentent pour le département 10 000 à 12 000 ha soit au maximum 1 500 ha pour le Bessin.

citée à savoir 6 260 ha. Nous l'avons retenue pour la même raison que le nombre d'exploitations de moins de 10 ha de la région. Cette enquête a employé également la notion de surface agricole utile.

La superficie exploitée par la population des exploitations de 10 ha et plus du Bessin est par conséquent évaluée à 71 675 — 6 260 soit

$$Y_0 = 65\,415 \text{ ha}$$

La superficie occupée par les exploitations spécialisées exclues de notre champ d'investigation est négligeable et n'a pas donné lieu à une correction de la superficie ci-dessus. Encore une fois, il en est différemment dans d'autres régions.

$x_1$  le nombre de bovins présents au 1<sup>er</sup> mai 1963

a) Dans l'échantillon

La fiche d'exploitation indique le cheptel en début et en fin d'exercice. Par ailleurs le mouvement de bêtes établi pour chaque exploitation permet de retrouver aisément l'effectif bovin présent à n'importe quelle date.

b) Dans la population

Le Service central des Enquêtes et Études statistiques du ministère de l'Agriculture a réalisé en 1963 en collaboration avec la section Agriculture de l'I. N. S. E. E. et les Directions départementales des Services agricoles une enquête sur le cheptel bovin (1).

Cette enquête donne un cheptel au 1<sup>er</sup> janvier 1963 dans le Calvados de 473 100 bovins.

Par ailleurs l'état 6, document statistique officiel déjà signalé, indique au 1<sup>er</sup> mai 1963 un cheptel de 483 000 bovins.

Compte tenu de la différence de dates et de modes d'établissement de ces statistiques, et de l'écart constaté entre les effectifs de vaches laitières donnés par ces 2 mêmes sources, respectivement 186 400 et 210 000 (la deuxième évaluation telle que permet de l'approcher la connaissance de la production laitière et des rendements unitaires, semble plus proche de la réalité) nous retenons 480 000 bovins pour le département au 1<sup>er</sup> mai 1963.

Cet effectif est à répartir entre les 4 régions. En l'absence de statistiques récentes, nous disposons des données par région du recensement agricole de 1955. Les agriculteurs à l'époque ont fait de légères sous-déclarations comme l'ont prouvé les réclamations formulées quelques mois plus tard au moment de l'attribution des bons d'essence calculée sur ces déclarations. Mais nous pensons que les sous-déclarations ont été du même ordre de grandeur dans les régions.

Le cheptel bovin s'est accru depuis dans tout le département, nous considérons comme encore utilisable la répartition suivante du cheptel par région, établie en 1955 : Bessin 17 %, Bocage 29 %, Pays d'Auge 37 % et Plaines 17 % soit 81 600 bovins dans le Bessin au 1<sup>er</sup> mai 1963.

Reste à déduire de ce cheptel, celui des exploitations de moins de 10 ha. Nous retenons pour ce terme correctif 1 bovin par hectare soit 6 260 bovins. Ces exploitations sont en effet très

1. Cette enquête a mis en œuvre le plan de sondage suivant :

- 1 — une stratification par région agricole;
- 2 — un tirage au sort de communes avec des probabilités inégales proportionnelles au nombre d'exploitations qu'elles comprenaient au recensement agricole de 1955;
- 3 — un tirage systématique des exploitations des communes tirées au sort après classement des exploitations suivant le nombre croissant de bovins possédés au moment du recensement.

Le taux de sondage a été de 1/100.

Les résultats sont significatifs au niveau du département.

généralement toute en herbe. La norme de 1 bovin/ha est celle de l'enquête sur le cheptel bovin au 1/100 pour cette catégorie très homogène d'exploitations.

Le cheptel bovin de la population des exploitations de 10 ha et plus du Bessin est par conséquent évalué à 81 600 — 6 260 soit

$$Y_1 = 75\,340 \text{ bovins}$$

$x_2$  la quantité de blé commercialisé durant la campagne 1963-1964

a) Dans l'échantillon

La fiche d'exploitation indique les quantités vendues entre le 1<sup>er</sup> mai 1963 et le 30 avril 1964. Les stocks de blé au 1<sup>er</sup> mai sont très rares dans cette région herbagère. Les quantités qui sont commercialisées le sont immédiatement après la récolte auprès des organismes stockeurs. Il n'y a pratiquement pas de vente au détail. Ceci explique la concordance entre les chiffres portés sur les fiches d'exploitation utilisées et les renseignements obtenus pour la population.

b) Dans la population

Il nous a été possible de relever à la Section départementale de l'Office national interprofessionnel des Céréales (O. N. I. C.) sur les listes de collecte par producteur, les quantités de blé collectées par commune. Par agrégation, nous avons obtenu la collecte de blé dans le Bessin durant la campagne 1<sup>er</sup> juillet 1963-30 juin 1964.

Les exploitations de moins de 10 ha ne cultivent pas de blé.

La quantité de blé commercialisé par la population des exploitations de 10 ha et plus du Bessin est de

$$Y_2 = 61\,836 \text{ qx}$$

$x_3$  le nombre de tracteurs au 1<sup>er</sup> mai 1964

a) Dans l'échantillon

Le nombre de tracteurs est extrait de l'inventaire du matériel fait tous les ans dans les exploitations suivies par le Centre de Gestion.

b) Dans la population

Le Génie rural tient à jour depuis 1961 l'état du parc de tracteurs en service, par commune et par nature de carburant utilisé.

Comme pour la collecte de blé, il suffit d'agrèger ces statistiques par région agricole pour obtenir (à la date voulue) les effectifs suivants : Bessin 1 832 — Bocage 4 050 — Pays d'Auge 3 425 — Plaines 3 427.

Les exploitations de moins de 10 ha possédant un tracteur étant des exceptions, nous avons retenu pour nombre de tracteurs dans les exploitations de 10 ha et plus du Bessin :

$$Y_3 = 1\,832 \text{ tracteurs}$$

$x_4$  le nombre d'unités travailleurs salariés employés en 1963

a) Dans l'échantillon

La fiche d'exploitation indique le nombre de journées de travail de salariés auxquelles l'exploitant a fait appel durant l'exercice 1<sup>er</sup> mai 1963-30 avril 1964. Ce nombre de journées divisé par 300 nous donne par définition le nombre d'unités « travailleur salarié ». Ce nombre calculé, avec 2 décimales, correspond à une notion précise contrairement à celle de salariés permanents employés sur l'exploitation ou tout simplement de nombre de salariés.

b) Dans la population.

Aucune statistique n'est utilisable brute, nous avons confronté les diverses statistiques existantes.

Le Service des Affaires professionnelles et de la Protection sociale du ministère de l'Agriculture nous indique au 1<sup>er</sup> janvier 1964 : 12 030 permanents proprement dits et 4 340 permanents « économiques » définis comme étant des salariés travaillant toute l'année dans l'agriculture mais chez des exploitants différents, soit au total l'équivalent de 16 370 années travail. Nous pouvons confondre ce nombre d'années travail avec le même nombre d'unités « travailleur ».

Mais cet effectif comprend les salariés agricoles permanents des exploitations agricoles et forestières et des entreprises de travaux agricoles.

Après déduction de 300 salariés pour les entreprises de travaux agricoles et de 400 salariés pour les exploitations forestières, renseignements obtenus auprès de l'Inspection des Lois sociales, nous parvenons à un premier effectif de 16 370-700 soit :

*15 670 années travail*

La Mutualité sociale agricole nous indique la masse des cotisations d'Assurance sociale encaissées en 1963 et la cotisation annuelle moyenne pondérée par le sexe et l'âge des salariés. La division du premier nombre par le deuxième nous donne 13 645 années travail pour les salariés des exploitations agricoles et des artisans ruraux regroupés dans les états de la Mutualité.

Après déduction de 400 salariés pour les artisans ruraux (renseignement communiqué par l'Inspection des Lois sociales) nous parvenons à un deuxième effectif de 13 645 - 400 soit

*13 245 années travail*

Les statistiques des Affaires professionnelles et de la Protection sociale pèchent par excès en regard du renseignement que nous recherchons. En effet, un salarié travaillant 6 heures par jour toute l'année chez le même exploitant a été compté pour 1 unité travailleur, de même s'il a été malade pendant un certain temps.

Par contre les statistiques de la Mutualité sociale agricole pèchent par défaut par suite des sous-déclarations qui concernent surtout les temporaires.

Le léger décalage entre les périodes sur lesquelles portent les statistiques des Affaires professionnelles, de la Mutualité et du Centre de Gestion peut être négligé.

Nous avons finalement retenu *14 500 années travail* de salariés pour le département en 1963 (1).

Reste à ventiler ces 14 500 années travail par région agricole.

Nous avons utilisé pour cette ventilation les résultats du recensement agricole de 1955. Il n'y a pas eu de grosses différences dans l'évolution des 4 régions du département entre 1955 et 1962. L'évolution va par contre depuis 1963 en s'accéléralant, comme le prouve l'observation du nombre d'années travail obtenu auprès de la Mutualité sociale pour chaque année de 1961 à 1965.

Nous avons retenu la répartition suivante :

Bessin 15,8 % — Bocage 25,3 % — Pays d'Auge 34,4 % — Plaines 24,5 %  
soit pour le Bessin  $14\ 500 \times 0,158 = 2\ 290$ .

1. Le recensement de population de 1962 s'est avéré inutilisable pour l'établissement de cette statistique. Nous ne disposions pas lors de ce travail des résultats de l'enquête Échantillon Maître au 1/10. Depuis, les résultats publiés nous font penser que l'effectif retenu pourrait être un peu élevé.

Les exploitations de moins de 10 ha n'emploient pas de main-d'œuvre salariée. Nous pouvons donc retenir pour la population des exploitations de 10 ha et plus du Bessin.

$$Y_4 = 2\,290 \text{ unités travailleur salarié}$$

#### Autres variables

La quantité de blé produit serait une variable préférable à la quantité commercialisée mais ne peut être connue par région agricole.

De même, s'il nous avait été possible de l'évaluer par région à partir des statistiques existantes, la quantité de litres de lait produite aurait été retenue à la place du nombre de bovins : les enregistrements du Centre de Gestion ont mis en évidence une corrélation beaucoup plus forte entre le revenu des exploitants et la production laitière qu'entre le revenu et le nombre de bovins.

Parmi les productions végétales, deux autres statistiques peuvent être établies de façon précise à l'échelle de la région agricole ; la collecte de betteraves à sucre et la superficie en lin. Mais ces productions n'intéressent qu'une seule des 4 régions agricoles du Calvados : les plaines de Caen et Falaise avec respectivement 93 % de la collecte et 93 % des emblavements départementaux.

#### VI. — VALEUR DES VARIABLES RETENUES DANS LES ENTREPRISES DE L'ÉCHANTILLON

Classe N° et Nombre d'exploitations	Exploit. N°	Superficie ha $x_0$	Nombre de bovins unités $x_1$	Quantité de blé qz $x_2$	Nombre de tract. unités $x_3$	Nombre d'UTH unités $x_4$
$c = 1$ $n_1 = 4$ $N_1 = 810$	1	13,7	18	57	1	0,03
	2	16,0	23	0	1	0
	3	18,5	25	0	1	0,13
	4	18,2	34	0	1	1,01
$c = 2$ $n_2 = 20$ $N_2 = 1\,010$	5	21,6	31	24	1	0,02
	6	23,0	25	0	1	0,34
	7	24,0	38	0	1	0,33
	8	25,8	37	0	1	1,16
	9	26,0	45	37	1	0,92
	10	26,9	49	0	1	1,98
	11	28,0	50	0	1	0,48
	12	26,3	45	173	1	0,70
	13	30,0	35	232	1	1,43
	14	30,0	39	44	1	1,19
	15	30,0	41	0	1	0,38
	16	31,7	39	115	1	1,00
	17	31,5	45	236	1	1,96
	18	34,0	39	212	1	1,33
	19	34,9	42	50	1	0,66
	20	35,0	45	121	1	1,85
21	35,0	52	117	1	1,65	
22	38,5	40	42	1	2,73	
23	43,0	57	0	0	1,85	
24	41,7	38	357	1	1,76	
$c = 3$ $n_3 = 15$ $N_3 = 320$	25	50,0	67	0	1	1,69
	26	50,3	52	264	1	3,24
	27	52,0	67	275	1	1,54
	28	52,6	88	0	2	1,32
	29	55,0	72	184	1	2,44
	30	55,7	52	151	2	2,42
	31	60,0	67	409	1	4,06
	32	61,3	57	165	1	3,88
	33	65,0	88	76	1	3,31
	34	69,3	82	0	1	2,64
	35	77,5	91	99	1	3,64
	36	90,3	136	163	2	5,14
	37	97,0	137	350	3	3,86
	38	105,3	96	368	2	4,21
	39	181,7	232	164	2	7,44

## VII. — CALCULS ET OBSERVATIONS INTERMÉDIAIRES

 1 — Calcul des 5 seconds membres  $S_k$ 

Nous pouvons résumer ce calcul dans le tableau ci-dessous :

Variable	$x_0$	$x_1$	$x_2$	$x_3$	$x_4$
Calcul de $S_k$					
$Y_k$	65 415	75 340	61 836	1 882	2 200
— $\sum_c N_c \alpha_{ck}$	— 68 741	— 92 052	— 157 497	— 2 177	— 2 514
= $S_k$	— 3 326 ha	— 16 712 bovins	— 95 661 qø	— 345 tract.	— 224 UTH

Les entreprises de l'échantillon sont en moyenne un peu plus grandes que celles constituant la population, mais relativement peu (d'environ 5 %). Par contre elles possèdent plus de bovins (+ 22 %) et commercialisent beaucoup plus de blé (+ 155 %).

Elles sont également plus mécanisées (+ 19 %) et emploient un peu plus de main-d'œuvre salariée (+ 10 %).

Ces écarts sont en partie la conséquence de la collaboration exploitant-conseiller de Gestion à la recherche d'une meilleure rentabilité de l'entreprise.

Développant la production fourragère à l'hectare notamment par des emplois plus grands d'engrais, les adhérents du Centre de Gestion entretiennent davantage de bovins à l'hectare. Ceci leur permet, tout en ayant un troupeau en augmentation de libérer une partie des terres labourables pour faire des céréales. Celles-ci bénéficient des précédents fourragers et apportent la paille nécessaire aux animaux.

Toutes les exploitations, sauf une, possèdent au moins un tracteur alors que dans la population un certain nombre d'exploitations, parmi celles de moins de 30 ha notamment, utilisent uniquement la traction animale.

Enfin les exploitants, adhérents du Centre de Gestion, tout en recherchant à réduire l'emploi de main-d'œuvre salariée par des investissements en matériel et installations, préfèrent conserver une main-d'œuvre encore importante et produire davantage.

Mais c'est surtout par les ventes de blé que les individus de l'échantillon s'éloignent de ceux de la population.

## 2 — Calcul des 25 coefficients des 5 équations

Désignons par  $a_{k'k}$  ces coefficients

$$a_{k'k} = \sum_k \sum_t x_{ctk'} (x_{ctk} - \bar{x}_{ck})$$

Ce sont les numérateurs des 5 variances de la superficie, du nombre de bovins, du nombre de quintaux de blé commercialisé, du nombre de tracteurs et du nombre d'unités travailleurs salariés et des 10 covariances des 10 couples de variables.

Ces coefficients sont symétriques  $a_{k'k} = a_{kk'}$ .

Il n'y a donc que 15 coefficients à calculer, correspondant au système d'équations suivant :

$$\left\{ \begin{array}{l} a_{00} \alpha_0 + a_{01} \alpha_1 + a_{02} \alpha_2 + a_{03} \alpha_3 + a_{04} \alpha_4 = S_0 \\ a_{10} \alpha_0 + a_{11} \alpha_1 + a_{12} \alpha_2 + a_{13} \alpha_3 + a_{14} \alpha_4 = S_1 \\ a_{20} \alpha_0 + a_{21} \alpha_1 + a_{22} \alpha_2 + a_{23} \alpha_3 + a_{24} \alpha_4 = S_2 \\ a_{30} \alpha_0 + a_{31} \alpha_1 + a_{32} \alpha_2 + a_{33} \alpha_3 + a_{34} \alpha_4 = S_3 \\ a_{40} \alpha_0 + a_{41} \alpha_1 + a_{42} \alpha_2 + a_{43} \alpha_3 + a_{44} \alpha_4 = S_4 \end{array} \right.$$



Nous obtenons :

$$\begin{array}{cccccc}
 a_{00} = 17\,432 & a_{01} = 21\,552 & a_{02} = 15\,738 & a_{03} = 250 & a_{04} = 768 \\
 & a_{11} = 31\,925 & a_{12} = 1\,811 & a_{13} = 350 & a_{14} = 923 \\
 & & a_{22} = 458\,842 & a_{23} = 654 & a_{24} = 1\,703 \\
 & & & a_{33} = 9 & a_{34} = 5 \\
 & & & & a_{44} = 46
 \end{array}$$

### 3 — Calcul des multiplicateurs de Lagrange $\alpha_k$

Les multiplicateurs de Lagrange sont les inconnues du système de 5 équations ci-dessus.

Ce système a été résolu en inversant la matrice constituée par les 25 coefficients  $a_{kk'}$  du premier membre. Cette inversion trop lourde pour être effectuée à la main, a été confiée au calculateur électronique I. B. M. de la Faculté des Sciences de Poitiers.

Les coefficients  $\alpha_k$  sont le résultat du produit de la matrice inverse par le vecteur

$$\begin{bmatrix} - & 3\,326 \\ - & 19\,712 \\ - & 95\,661 \\ - & 345 \\ - & 224 \end{bmatrix}$$

à savoir  $\alpha_0 = 7,34$   $\alpha_1 = -1,57$   $\alpha_2 = 0,08$   $\alpha_3 = -137,66$   $\alpha_4 = -81,79$

### 4 — Calcul des poids $p_{ct}$

Nous avons tous les éléments pour calculer les poids à affecter aux entreprises de l'échantillon

$$p_{ct} = \frac{N_c}{n_c} + \sum_k (x_{ctk} - \bar{x}_{ck}) \alpha_k$$

Ces poids devant représenter un nombre d'exploitations ont été calculés à l'unité près. De plus, pour avoir une signification facile à interpréter, ils doivent osciller au sein de chaque classe autour de l'inverse du taux de sondage  $g_c$  dans la classe :

$$g_1 = \frac{810}{4} = 202 \quad g_2 = \frac{1\,010}{20} = 50 \quad g_3 = \frac{320}{15} = 21$$

La classe 1 est si peu représentée que nous devons faire de grandes réserves sur les résultats la concernant.

Rappelons enfin que des poids négatifs n'auraient aucun sens.

Les résultats sont rassemblés dans le tableau ci-dessous :

Classe 1	Classe 2		Classe 3	
$p_1 = 216$	$p_5 = 75$	$p_{15} = 90$	$p_{25} = 58$	$p_{25} = 70$
$p_2 = 223$	$p_6 = 67$	$p_{16} = 64$	$p_{26} = -22$	$p_{26} = -157$
$p_3 = 224$	$p_7 = 55$	$p_{17} = -18$	$p_{27} = 25$	$p_{27} = -130$
$p_4 = 140$	$p_8 = 0$	$p_{18} = 62$	$p_{28} = -63$	$p_{28} = 107$
	$p_9 = 13$	$p_{19} = 106$	$p_{29} = 40$	$p_{29} = 171$
	$p_{10} = -76$	$p_{20} = 10$	$p_{30} = -61$	
	$p_{11} = 53$	$p_{21} = 15$	$p_{31} = -30$	
	$p_{12} = 44$	$p_{22} = -37$	$p_{32} = -6$	
	$p_{13} = 28$	$p_{23} = 177$	$p_{33} = 8$	
	$p_{14} = 38$	$p_{24} = 96$	$p_{34} = 97$	

## VIII — ANALYSE DES RÉSULTATS

L'existence de 10 poids négatifs, 3 dans la classe 2 et 7 dans la classe 3, *ne permet pas d'atteindre le but fixé* : l'estimation du revenu global des exploitations de 10 ha et plus du Bessin. Mais l'analyse de ces résultats est néanmoins très intéressante.

— Si nous n'avions obtenu que 1 ou 2 poids négatifs et petits par classe, la classe 1 exceptée, nous aurions pu éliminer les entreprises correspondantes en leur attribuant un poids nul ou mieux égal à 1 et passer outre.

Remarquons qu'attribuer le poids  $p_j = 0$  à l'entreprise  $j$  l'élimine de l'estimation du revenu  $\sum p_j x_j$  mais la laisse subsister dans les calculs ayant conduit aux  $p_j$ ; on ne peut envisager de recommencer tous les calculs. Ces poids nuls créent des écarts entre  $\sum p_j x_{jk}$  et  $Y_k$  qu'il est possible de calculer.

Par exemple, attribuer à l'entreprise n° 10 un poids  $p'_{10} = 0$  au lieu de  $p_{10} = -76$  crée un écart de  $26,9 \times 76 = 1\,944$  ha, c'est-à-dire de l'ordre de 3 % entre la superficie de la population et son estimation par l'échantillon, c'est encore admissible.

Mais attribuer à l'entreprise n° 36 un poids  $p_{36} = 0$  au lieu de  $p_{36} = -157$  crée un écart de l'ordre de 22 % entre ces mêmes superficies; c'est inadmissible. L'exploitation n° 36 a des caractéristiques trop éloignées des caractéristiques moyennes de la population : elle peut être qualifiée « d'entreprise aberrante ». Le malheur est que, dans ce cas, nous avons dans notre échantillon beaucoup d'entreprises « aberrantes » et nous ne pouvons pas les éliminer toutes.

*Nous avons non seulement un nombre trop élevé de poids négatifs mais ceux-ci sont aussi pour la plupart trop importants.*

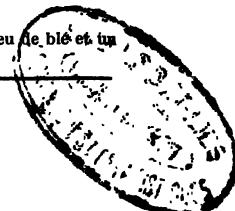
Aucune correction n'est possible.

— Les poids obtenus sont également très dispersés. Certains poids positifs, qui s'éloignent beaucoup des inverses du taux de sondage  $g_2 = 50$  et  $g_3 = 21$  sont une gêne.

Nous hésitons à attribuer à une entreprise, quelle qu'elle soit, un poids prédominant, surtout en agriculture où le revenu d'une exploitation peut varier d'une façon très importante d'une année sur l'autre.

— Après ces 2 observations principales, analysons dans le détail les poids extrêmes. Nous avons résumé cette analyse dans le tableau ci-dessous

Classe	Poids extrêmes		Entreprise N°	Observations Valeur comparée des variables par classe
	Sens	Valeur		
2 $g_2 = 50$	les plus négatifs	— 76	10	Le plus fort chargement de bovins à l'hectare (1,82).
		— 37	22	Le plus de main-d'œuvre salariée à l'hectare.
		— 16	17	Le plus de main-d'œuvre salariée à l'hectare après la n° 10.
	les plus positifs	96	24	Le plus faible chargement de bovins à l'hectare (0,91).
		105	19	Exploitation moyenne pour les 4 variables.
		177	23	La seule exploitation sans tracteur <i>a priori</i> « très moyenne » et toute désignée comme « représentative » dans un choix raisonné.
3 $g_3 = 21$	les plus négatifs	— 157	36	Le plus fort chargement de bovins à l'hectare (1,50) après la n° 28.
		— 130	37	Le plus fort chargement de bovins à l'hectare après les n° 28 et 36 — l'une des exploitations les plus mécanisées (1 tracteur pour 33 ha).
		— 63	28	Le plus fort chargement de bovins à l'hectare (1,67).
	les plus positifs	97	34	Exploitation « très moyenne » jugée représentative dans 1 choix raisonné, contrairement aux 2 suivantes à cause de leur superficie
		107	38	Le plus faible chargement de bovins à l'hectare (0,91).
		171	39	Le moins de tracteur à l'hectare (1 tracteur pour plus de 90 ha) très peu de blé et un chargement bovins moyen.



Nous constatons à l'examen individuel des poids que, lorsque nous connaissons les exploitations sous l'angle économique, ceux-ci sont le plus souvent en accord avec l'idée de « représentativité » que nous pouvons nous faire des exploitations. La méthode devrait donc pouvoir donner des résultats mais l'échantillon est trop « mauvais »; il est constitué d'exploitations dont la production ou la structure s'éloigne trop de celle des individus « moyens » de la population.

Il comprend trop peu d'exploitations « moyennes » ce qui conduit à affecter à celles-ci des poids positifs élevés et ce qui entraîne des poids négatifs pour les exploitations « de tête ».

Ceci provient du fait que ce sont les agriculteurs qui sont déjà « en pointe » ou qui cherchent à le devenir avec la collaboration d'un conseiller, qui font la démarche d'adhérer au Centre de Gestion.

Il faudrait pouvoir disposer pour la région d'une cinquantaine de comptes d'exploitations mieux répartis parmi les 3 classes de superficie retenues et obtenus chez des agriculteurs n'adhérant pas au Centre.

Mais est-ce possible? La tenue d'une comptabilité ou d'un carnet d'exploitation n'est-elle pas en elle-même déjà le signe d'une recherche particulière qui n'est pas le fait courant?

Ne faut-il pas mieux revenir à notre échantillon gestion, et chercher une autre méthode qui permette d'approcher du but fixé : tel est l'objet de notre 2<sup>e</sup> partie.

### *Troisième Partie*

## DEUXIÈME MÉTHODE D'ESTIMATION

### MÉTHODE RÉVISÉE DES POIDS

Nous conservons la même hypothèse de départ, les exploitations de l'échantillon sont tirées au sort avec des probabilités inégales. Nous désirons affecter chacune d'elles d'un poids  $p_j$  qui est l'inverse de la probabilité de la tirer. Mais nous allons chercher à déterminer ces poids d'une façon plus directe à l'aide de la théorie du barycentre.

#### I. — NOTATIONS

Désignons par :

$c = 1, 2, 3$ , les 3 classes de superficie 10 à 20, 20 à 50, 50 ha et plus.

Pour alléger au lieu de  $x_0 x_1 x_2 x_3 x_4 : h, b, q, s, t$ , les variables repères de l'échantillon : nombre d'hectares, nombre de bovins, nombre de quintaux de blé commercialisé, nombre d'unités travailleurs salariés, nombre de tracteurs, H, B, Q, S, T les valeurs de ces variables pour l'ensemble des 3 classes, telles qu'elles ressortent des statistiques disponibles sur la population.

$n_c$  le nombre d'exploitations dans la classe  $c$  de l'échantillon

$N_c$  le nombre d'exploitations dans la classe  $c$  de la population

$p_j$  les poids à déterminer

Nous nous fixons les mêmes contraintes linéaires que précédemment :

a) par classe : la somme des poids est égale au nombre d'exploitations de la classe

$$\sum_c p = N_c$$

b) pour l'ensemble des 3 classes : la somme des produits de la valeur des variables repères dans chaque exploitation, par le poids affecté à celle-ci est égale aux valeurs de ces variables, telles qu'elles ressortent des statistiques disponibles

$$\Sigma ph = H \quad \Sigma pb = B \quad \Sigma pq = Q \quad \Sigma ps = S \quad \Sigma pt = T$$

c) nous nous imposons également des poids au moins égaux à 1

$$p \geq 1$$

S'imposer des poids  $p_j$  au moins égaux à 1 revient à faire le changement de variable  $p = 1 + p'$  avec la condition  $p' \geq 0$ , les équations a)  $\Sigma_c p = N_c$  s'écrivent alors  $\Sigma_c p' = N_c - n_c$  de même les équations b)  $\Sigma ph = H$  s'écrivent  $\Sigma p'h = H - h_1 - h_2 - h_3$ .

## II. — EXPOSÉ DE LA MÉTHODE

Considérons un espace à 5 dimensions rapporté à 5 axes de coordonnées  $(h, b, q, s, t)$  où chaque exploitation de l'échantillon est représentée par un point  $M_j$ . Si nous désignons par  $\mathcal{D}$  l'ensemble des domaines convexes  $D$  tels que tous les points  $M_j$  soient dans  $D$  ou sur la frontière de  $D$ , il existe un domaine convexe  $D_0$  appartenant à  $\mathcal{D}$  ( $D_0 \subset \mathcal{D}$ ) et inclus dans tout  $D$ . C'est un polyèdre dont les sommets sont certains des points  $M_j$ .

L'équation  $\bar{h} = \frac{\Sigma ph}{N}$  et les équations analogues pour les autres variables déterminent les coordonnées d'un point, le barycentre  $G$  des  $M_j$  affectés des poids  $p_j$ , qui, lorsque les  $p$  varient, peut occuper toute position à l'intérieur ou sur la frontière de ce polyèdre  $D_0$ .

Les contraintes  $\Sigma ph = H$  et analogues s'écrivent

$$N_1 \bar{h}_1 + N_2 \bar{h}_2 + N_3 \bar{h}_3 = H = N\bar{H} \text{ en posant } N_1 + N_2 + N_3 = N \text{ et } \frac{H}{N} = \bar{H}$$

Ces contraintes signifient qu'on impose au barycentre  $G$  d'occuper une position repère donnée  $(\bar{H}, \bar{B}, \bar{Q}, \bar{S}, \bar{T})$ .

Notre problème a au moins une solution si ce point repère est à l'intérieur ou sur la frontière de  $D_0$ . Dans ce cas, nous choisirons dans l'ensemble des solutions, celle rendant minimale une certaine fonction quadratique de  $p$  (choisie par ailleurs).

Nous nous sommes inspirés de cette théorie; seulement, vu l'impossibilité de représenter graphiquement un espace à 5 dimensions, nous avons cherché à simplifier le problème pour raisonner dans un plan.

1° L'examen de la variable  $t$  nous a conduit à l'abandonner.

2° Nous avons ensuite donné un rôle privilégié à la variable  $h$ ; nous avons observé dans un plan la position des points repères  $\left(\frac{B}{H}, \frac{Q}{H}\right)$  et  $\left(\frac{B}{H}, \frac{S}{H}\right)$  parmi celles des exploitations.

Pour que notre problème ait une solution, il est nécessaire que ces points repères soient à l'intérieur ou sur le pourtour des plus petits polygones convexes contenant à leur intérieur ou sur leur pourtour les  $n$  points  $\left(\frac{b}{h}, \frac{q}{h}\right)_j$  et  $\left(\frac{b}{h}, \frac{s}{h}\right)_j$

Nous disposons des données suivantes :

pour	$i = 1$	$h_1 = 66,4$	$b_1 = 100$	$q_1 = 57$	$s_1 = 1,17$	$t_1 = 4$
	$i = 2$	$h_2 = 616,7$	$b_2 = 832$	$q_2 = 1\ 760$	$s_2 = 23,68$	$t_2 = 19$
	$i = 3$	$h_3 = 1\ 124,5$	$b_3 = 1\ 384$	$q_3 = 2\ 668$	$s_3 = 50,83$	$t_3 = 22$
		$\Sigma ih = 1\ 807$	$\Sigma ib = 2\ 316$	$\Sigma iq = 4\ 485$	$\Sigma is = 76$	$\Sigma it = 45$

Avant l'abandon de la variable  $t$ , nous avons à résoudre le système d'équations suivant :

$$\left. \begin{array}{l} \Sigma p'h = 65\,415 - 1\,807 = 63\,608 \text{ hectares} \\ \Sigma p'b = 75\,340 - 2\,316 = 73\,024 \text{ bovins} \\ \Sigma p'q = 61\,836 - 4\,485 = 57\,351 \text{ quintaux de blé} \\ \Sigma p's = 2\,290 - 76 = 2\,214 \text{ unités travailleurs salariés} \\ \Sigma p't = 1\,832 - 45 = 1\,787 \text{ tracteurs} \\ \Sigma {}_1p' = 810 - 4 = 806 \text{ exploitations} \\ \Sigma {}_2p' = 1\,010 - 20 = 990 \text{ exploitations} \\ \Sigma {}_3p' = 320 - 15 = 305 \text{ exploitations} \\ \text{avec } p' \geq 0 \end{array} \right\}$$

### III. — ABANDON DE LA VARIABLE $t$

Notre échantillon ne possède qu'une seule exploitation sans tracteur contre 6 exploitations ayant plusieurs tracteurs.

Nous avons  $\Sigma p't = 1\,787$  tracteurs et  $\Sigma p' = N - n = 2\,140 - 39 = 2\,101$ .

A supposer que toutes les exploitations ayant un tracteur ou plus reçoivent le poids  $p = 1$  ou  $p' = 0$ , il y aurait lieu d'affecter l'exploitation sans tracteur du poids  $p'_{23} = 2\,101 - 1\,787 = 314$  soit  $p_{23} = 315$ .

Il s'en suivrait que cette exploitation n° 23 à elle seule compterait pour  $57 \times 315 = 17\,955$  bovins sur les 75 340 de l'ensemble des 3 classes. Cela paraît très dangereux. Par ailleurs, il est facile de constater qu'avec ces poids, le système d'équations ci-dessus ne serait pas résolu.

Nous avons jugé plus sage de laisser de côté cette variable : notre échantillon est trop « mécanisé » comme nous l'avons déjà souligné page 259.

Nous sommes alors ramenés à un système à 7 équations au lieu de 8. Nous avons donc à faire dorénavant à un polyèdre de l'espace à 4 dimensions  $h, b, q, s$ .

### IV. — PROBLÈME POSÉ PAR LA CLASSE 1

A la suite de nombreux tâtonnements à la recherche de poids  $p'$  satisfaisant notre système de 7 équations, nous sommes parvenus à la conclusion qu'il était préférable de se pencher d'abord sur la classe 1 vu sa très mauvaise représentation.

Nous sommes en présence de 4 exploitations dont la somme des poids  $p'$  doit être égale à 806. Nous allons représenter cette classe 1 par une seule des 4 exploitations, celle qui apparaîtra comme la moins défavorable vis-à-vis des contraintes qui en résultent par contre coup pour les classes 2 et 3. A l'observation dans ces exploitations des valeurs des 4 variables retenues, 2 exploitations peuvent être jugées « moyennes », les n°s 2 et 3. Elles sont très proches l'une de l'autre par leur caractéristiques. Nous avons choisi en conséquence d'affecter *a priori* ces exploitations des poids suivants :

$$p'_1 = p'_3 = p'_4 = 0 \quad \text{et} \quad p'_2 = 806$$

Nous avons préféré l'exploitation n° 2 à l'exploitation n° 3 parce que sa surface est plus proche de la surface moyenne des exploitations constituant cette classe de la population.

Nous avons cependant hésité parce que le revenu de cette exploitation doit s'écarter assez largement du revenu moyen des exploitations de cette classe de superficie.

Nous déduisons alors de ce poids la valeur des variables repères pour l'ensemble de la classe 1 :

$$H_1 = 12\,896 \text{ ha} \quad B_1 = 18\,538 \text{ bov.} \quad Q_1 = 0 \text{ q.} \quad S_1 = 0 \text{ UT}$$

Nous avons bien conscience que cette façon de faire n'est pas très bonne, mais nous avons voulu tester la méthode sans être arrêté par cette catégorie d'exploitations.

Les poids des 35 exploitations des classes 2 et 3 doivent alors satisfaire le système d'équations suivant :

$$\left\{ \begin{array}{l} \Sigma_{2,3} p'h = 63\,608 - 12\,896 = 50\,712 \text{ hectares} = H' \\ \Sigma_{2,3} p'b = 73\,024 - 18\,538 = 54\,486 \text{ bovins} = B' \\ \Sigma_{2,3} p'q = 57\,351 - 0 = 57\,351 \text{ qx de blé} = Q' \\ \Sigma_{2,3} p's = 2\,214 - 0 = 2\,214 \text{ unités tra-} \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \text{vail leurs salariés} = S' \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \Sigma_2 p' = 990 \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \Sigma_3 p' = 305 \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \text{avec } p' \geq 0 \end{array} \right.$$

On envisage à présent les 2 polyèdres définis par les exploitations de la classe 2 et par celles de la classe 3 et le polyèdre défini par l'ensemble des 2 classes.

#### V. — UTILISATION DES VARIABLES $h, b, q, s$

Comme il est à peu près impossible de se représenter des polyèdres dans l'espace à 4 dimensions ( $h, b, q, s$ ), on a essayé d'en faire l'étude au moyen de projections sur les plans de coordonnées couples de variables  $hb, hq, hs, bq, bs, qs$ . Cette étude ne renseigne pas suffisamment sur les polyèdres définis par nos exploitations, aussi a-t-on formé les ratios  $\left(\frac{b}{h} \frac{s}{h} \frac{q}{h}\right)$  et étudié les couples de ratios  $\left(\frac{b}{h} \frac{s}{h}\right)$  et  $\left(\frac{b}{h} \frac{q}{h}\right)$  (1). Il s'agit en fait d'une perspective de pôle 0 des 2 polyèdres.

Nous obtenons ainsi des polygones  $P_0$ , projections ou perspectives du polyèdre  $D_0$  défini ci-dessus, mais amputé de la classe 1.

Nous observons les positions de la projection du point repère ( $H' B' Q' S'$ ) par rapport aux polygones  $P_0$ . Cette observation doit permettre de s'assurer d'une condition nécessaire à l'existence d'au moins une solution — si les projections du point repère sont à l'intérieur des polygones  $P_0$ . Nous chercherons alors les 6 exploitations apparemment les mieux situées pour constituer une solution au système des 6 équations précédemment écrites et aussi une solution des inéquations  $p' \geq 0$ .

En fait nous verrons par la suite que la 4<sup>e</sup> équation devra être modifiée sous peine de ne pas obtenir de solution.

1. Nous avons pensé en premier lieu plutôt utiliser les écarts  $\Delta h, \Delta b, \Delta q$  et  $\Delta s$ , calculés au sein des classes 2 et 3 à partir des exploitations « de queue » n° 6 et n° 25, exploitations permettant d'avoir le plus faible nombre d'écarts négatifs. Nous avons attribué à ces 2 exploitations les poids respectifs de 990 et 305. Puis nous avons cherché par des substitutions d'exploitations à l'intérieur de chaque classe à accroître *simultanément* la surface, le nombre de bovins, la quantité de blé commercialisé et le nombre d'unités travailleurs salariés jusqu'à atteindre les valeurs  $H' B' Q'$  et  $S'$ . Mais cela s'est avéré impossible.

Il en a été de même avec les rapports  $\frac{\Delta b}{\Delta h} \frac{\Delta q}{\Delta h} \frac{\Delta s}{\Delta h}$ . De plus on est arrivé à des résultats non réalistes c'est-à-dire des poids négatifs — ce qui conduit à renoncer à cette méthode.

Le lecteur pressé pourra sauter le détail des explorations et passer directement au paragraphe c (p. 268).

a) *Utilisation des ratios  $\frac{b}{h}$  et  $\frac{s}{h}$*

Projetons dans un plan les points représentant les exploitations des classes 2 et 3 d'après le nombre de bovins à l'hectare  $\left(\frac{b}{h}\right)$  et la quantité de main-d'œuvre salariée à l'hectare  $\left(\frac{s}{h}\right)$ . Nous obtenons le graphique ci-après.

Le point de repère  $G_1$  correspondant aux valeurs de ces ratios pour l'ensemble des 2 classes a pour coordonnées :

$$\left\{ \begin{array}{l} \frac{B'}{H'} = \frac{54\,486}{50\,712} = 1,07, \quad \frac{S'}{H'} = \frac{2\,214}{50\,712} = 0,0437 \end{array} \right\}$$

Nous constatons qu'il est à l'intérieur du polygone convexe  $P_0$  défini comme étant la projection du domaine  $D_0$ . Il est même assez nettement à l'intérieur.

Il apparaît ainsi qu'il peut être le barycentre aussi bien d'un ensemble d'exploitations de la classe 2 que de la classe 3, ces exploitations étant affectées des poids adéquats. Considérons par exemple les ensembles  $\{6\ 12\ 22\ 24\}$  et  $\{30\ 34\ 35\ 38\}$ . Il est important également de remarquer que les superficies de ces exploitations encadrent les superficies moyennes des exploitations des classes 2 et 3 de la population. La condition portant sur les superficies doit pouvoir être satisfaite en même temps que celles portant sur les ratios.

Ces observations laissent à penser qu'il doit être aisé de trouver un ensemble de 5 poids positifs, solution du système :

$$\left\{ \begin{array}{l} \Sigma p'h = H' \\ \Sigma p'b = B' \\ \Sigma p's = S' \\ \Sigma_2 p' = 990 \\ \Sigma_3 p' = 305 \end{array} \right.$$

b) *Utilisation des ratios  $\frac{b}{h}$  et  $\frac{q}{h}$*

Comme précédemment, projetons dans un plan les points représentant les exploitations des classes 2 et 3 d'après le nombre de bovins à l'hectare  $\left(\frac{b}{h}\right)$  et la quantité de blé commercialisé à l'hectare  $\left(\frac{q}{h}\right)$ . Nous obtenons le graphique de la page 267.

Le point repère  $G_{II}$  correspondant aux valeurs de ces ratios pour l'ensemble des 2 classes a pour coordonnées :

$$\left\{ \begin{array}{l} \frac{B'}{H'} = \frac{54\,486}{50\,712} = 1,07, \quad \frac{Q'}{H'} = \frac{57\,351}{50\,712} = 1,13 \end{array} \right\}$$

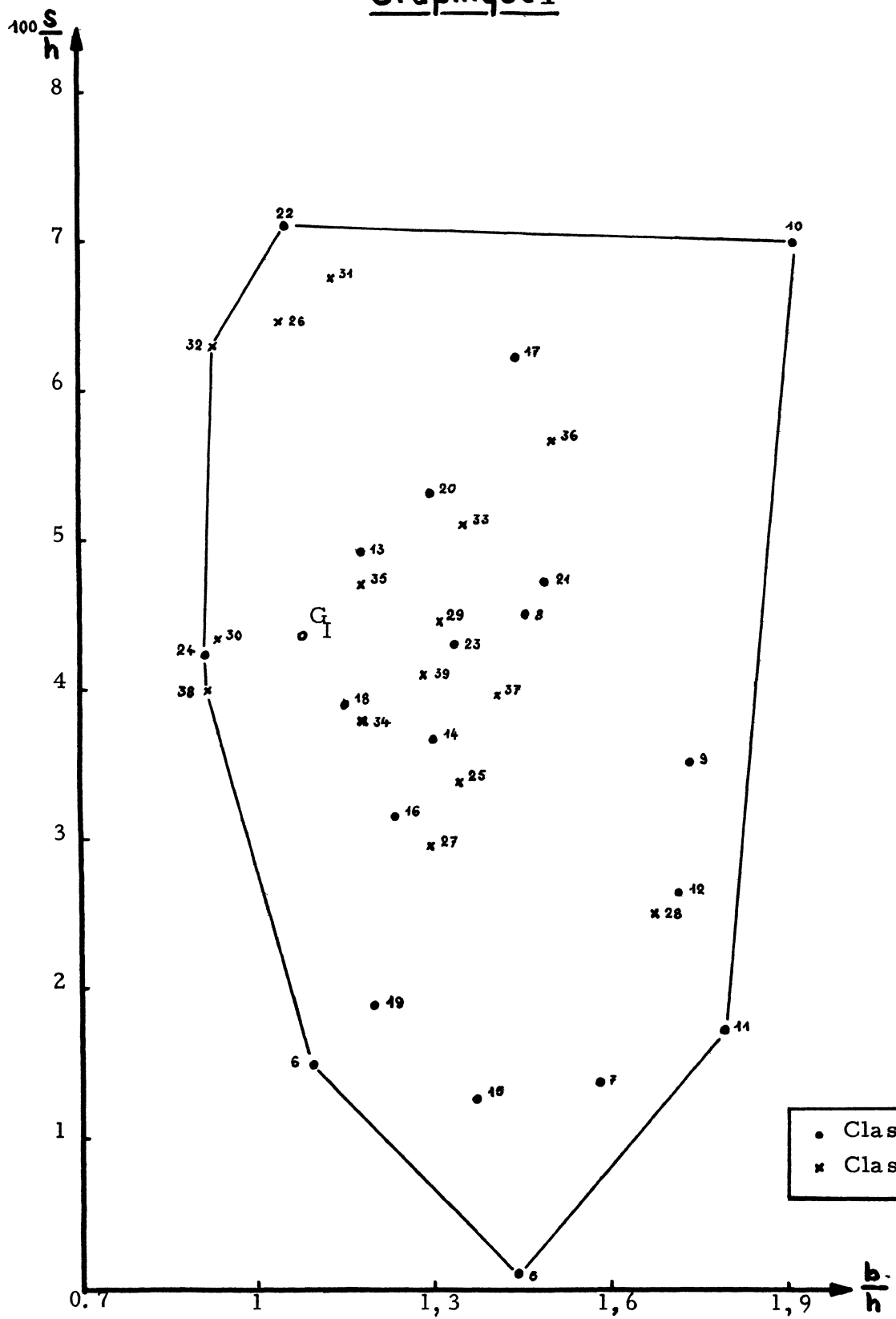
Il est à l'intérieur du polygone convexe  $P_0$ . Mais à l'examen du graphique, nous constatons :

1° qu'il est proche du pourtour;

2° qu'il est possible de le situer à l'intérieur d'un polyèdre composé de points exploitations appartenant à la classe 2 mais que la même opération est impossible avec la classe 3;

3° qu'en l'absence des exploitations n° 6 et 22, notre problème serait sans solution;  $G_{II}$  serait à l'extérieur du polygone convexe  $P_0$ .

# Graphique I



• Classe 2  
x Classe 3



Nous pouvons en déduire que l'exploitation n° 22 doit entrer dans la solution, s'il en existe une, avec un poids important.

Ce poids sera d'autant plus important que les autres exploitations choisies pour encadrer  $G_{II}$  seront plus éloignées de l'exploitation 22.

Nous avons donc intérêt à choisir les exploitations le plus groupées possible autour de  $G_{II}$  de façon à ne pas avoir au départ de poids trop différents.

Le graphique semble nous indiquer les exploitations 19 et 22 dans la classe 2 et les exploitations 30 et 34 dans la classe 3.

De plus pour répondre à la condition  $\Sigma p'h = H'$  soit

$$\frac{\Sigma p'h}{N'} = \frac{H'}{N'} = \frac{50\,712}{1\,295} = 39,1 \text{ hectares où } N' = \Sigma {}_2p' + \Sigma {}_3p'$$

une petite exploitation est nécessaire et la n° 6 est toute désignée.

Finalement la difficulté semble se trouver dans la « déformation » de l'échantillon que reflète le graphique II.

c) *Conclusion de l'exploration faite en a et b*

Cela nous a amené à chercher d'abord une solution  $p' \geq 0$  au système suivant :

$$\left\{ \begin{array}{l} \Sigma p'h = H' \\ \Sigma p'b = B' \\ \Sigma p'q = Q' \\ \Sigma {}_2p' = 990 \\ \Sigma {}_3p' = 305 \end{array} \right.$$

laissant de côté la variable  $s$ , quitte à en tenir compte par la suite si, une solution au système existant, cette solution était loin de respecter la contrainte  $\Sigma p's = S'$ .

Nous aurons donc l'espoir d'une solution à notre problème s'il est possible de résoudre en premier lieu le système suivant de 5 équations à 5 inconnues :

$$\left\{ \begin{array}{lllll} 23 p'_6 + 34,9 p'_{19} + 38,5 p'_{22} + 55,7 p'_{30} + 69,3 p'_{34} = 50\,712 \text{ ha} \\ 25 p'_6 + 42 p'_{19} + 40 p'_{22} + 52 p'_{30} + 82 p'_{34} = 54\,486 \text{ bov.} \\ \quad \quad \quad 50 p'_{19} + 42 p'_{22} + 151 p'_{30} = 57\,351 \text{ qx} \\ \quad \quad \quad p'_6 + p'_{19} + p'_{22} = 990 \text{ exp.} \\ \quad \quad \quad \quad \quad \quad \quad p'_{30} + p'_{34} = 305 \text{ exp.} \end{array} \right.$$

avec  $p'_j \geq 0$

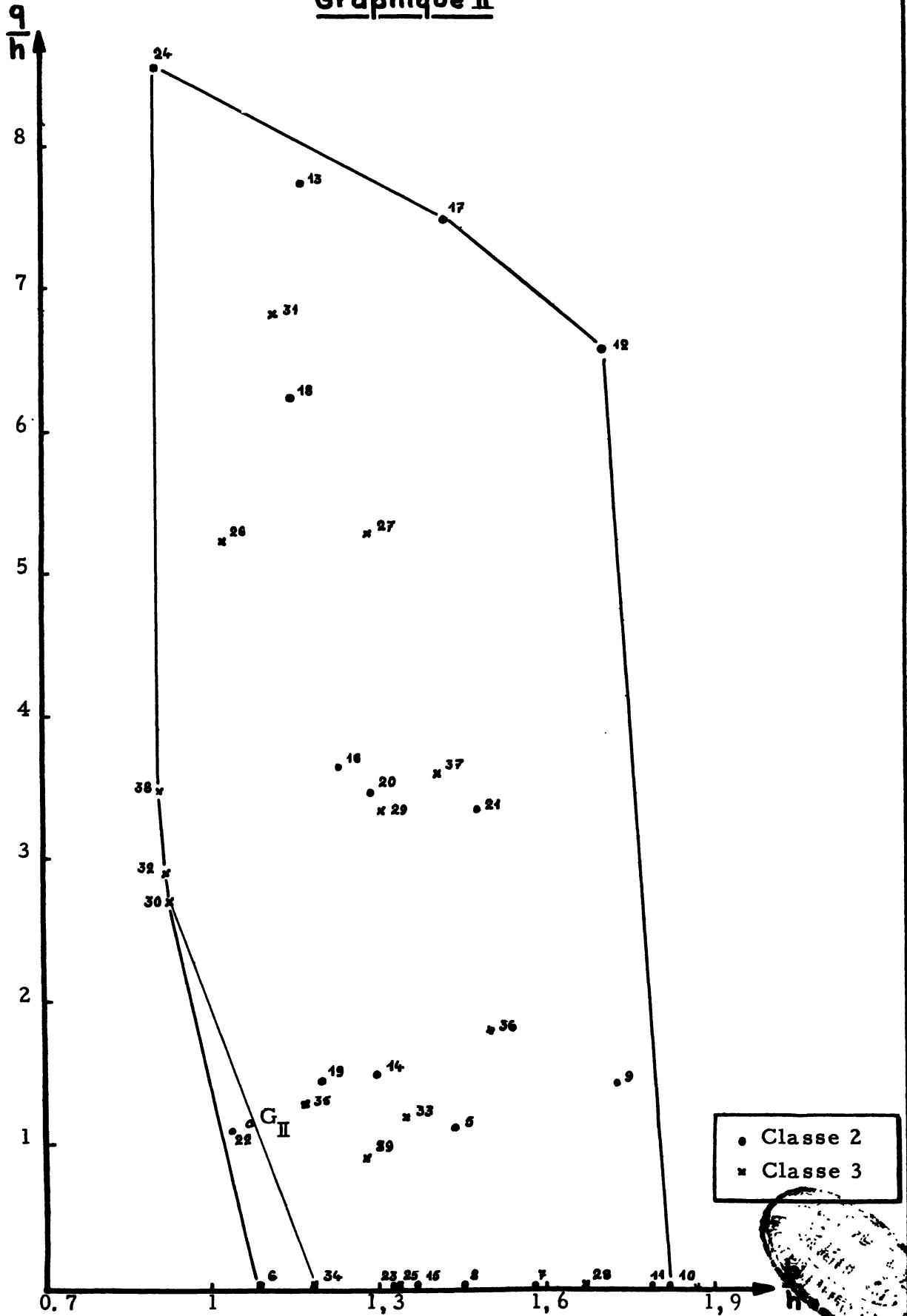
La résolution de ce système s'est avérée possible et nous a donné la solution suivante :

$$p'_6 = 336 \quad p'_{19} = 246 \quad p'_{22} = 408 \quad p'_{30} = 185 \quad p'_{34} = 120$$

que nous avons ainsi vérifiée :

N° Exploit.	$p'$	$h$	$b$	$q$	$s$
2	808	12 896	18 538	0	0
6	336	7 728	8 400	0	114,24
19	246	8 585	10 332	12 300	162,36
22	408	15 708	16 320	17 136	1 113,84
30	185	10 304	9 620	27 035	446,70
34	120	8 316	9 640	0	316,80
TOTAL . . . . .	2 101	63 537	72 850	57 371	2 153,94
Au lieu de . . . . .	2 101	63 608	73 024	57 351	2 214
ÉCART . . . . .	—	— 71	— 174	+ 20	— 60

# Graphique II



• Classe 2  
 x Classe 3



Les écarts en effet négligeables sont dus à l'arrondissement dans les calculs.

Par ailleurs nous avons déjà abandonné la variable  $t$  (tracteurs) par ce que notre échantillon était « trop mécanisé ». Nous pouvons en tenir compte maintenant en posant l'hypothèse d'une certaine compensation entre le nombre de tracteurs plus élevé et le nombre d'unités travailleurs salariés plus faible dans la « reconstitution de la population » (en partant des exploitations de notre échantillon affectées de leurs poids) que dans la population.

Acceptant en somme de remplacer  $S' = 2\,214$  par  $S'' = 2\,153$  nous avons finalement la solution :

$p_2 = 807$	$p_6 = 337$	$p_{19} = 247$	$p_{22} = 409$	$p_{30} = 186$	$p_{34} = 121$
$p = 1$ pour les 33 autres exploitations					

### Remarques

1° Nous percevons l'inconvénient de cette méthode qui donne une grande importance à un très petit nombre d'exploitations et réduit le rôle des autres à la représentativité d'elles-mêmes. Mais il n'est pas dit qu'on en restera à cette solution (voir VI ci-après).

2° Les difficultés rencontrées pour trouver cette solution sont la raison pour laquelle nous n'avons pas directement cherché à résoudre le système de 6 équations à 6 inconnues posé à la fin du IV) ci-dessus.

#### d) Retour à la variable $s$

La solution ci-dessus donne un nombre d'unités travailleurs salariés de 2 153 au lieu de  $S' = 2\,214$ . La différence est faible, de l'ordre de 2,7 %.

Nous avons cependant recherché, par des transports de poids de ces exploitations sur d'autres, à réduire cet écart, tout en respectant les autres contraintes le plus possible. Mais toutes les substitutions que nous avons essayées en vue d'augmenter  $\sum ps$ , augmentaient également la valeur moyenne des rapports  $\frac{Q'}{S'}$  ou  $\frac{B'}{S'}$

## VI. — LA SOLUTION TROUVÉE EST-ELLE LA MEILLEURE?

Puisque nous avons trouvé une solution, il peut en exister théoriquement une infinité dans un espace à 39 dimensions. (Nous avons conservé finalement 6 équations :

$$\begin{aligned} \sum ph &= H' & \sum pb &= B' & \sum pq &= Q' \\ \sum_1 p &= 810 & \sum_2 p &= 990 & \sum_3 p &= 305 \end{aligned}$$

avec 39 inconnues.)

Le problème se pose de chercher la meilleure solution que nous définissons comme étant celle qui rend minimum la fonction  $\sum p_j^2$ .

Nous avons en effet trouvé un sommet  $S$  de l'hyperpolyèdre qui satisfait à l'ensemble des conditions rappelées ci-dessus.

Il existe dans cet hyperpolyèdre une région  $R$  qui comprend  $S$  et vérifie aussi les conditions fixées. Cette région peut se réduire au point  $S$  mais il n'y a *a priori* aucune raison pour qu'elle se réduise à  $S$ .

Il nous faut explorer cette région R. Cette exploration peut consister à voir, si en s'écartant un tout petit peu de S dans diverses directions on se trouve encore dans R. S'il se trouve que R se réduise au point S il n'y a plus de problème,  $\Sigma p^2$  est déterminé et est son propre minimum. Si nous parvenons par contre à trouver un point de R autre que S nous en déduisons que R ne se réduit pas à S et que notre solution n'est probablement pas la meilleure.

Ce problème pourrait être traité complètement avec un ordinateur, nous nous proposons ici simplement une exploration numérique par tâtonnements.

Cette exploration numérique consiste à reporter une fraction des poids des entreprises nos 2, 6, 19, 22, 30 et 34 sur d'autres afin de diminuer  $\Sigma p^2$  tout en respectant les contraintes linéaires que nous nous sommes imposées.

Nos tâtonnements sont reportés dans le tableau suivant :

Entreprise N°	Report des poids	Modifications des écarts			
		<i>h</i>	<i>b</i>	<i>a</i>	<i>s</i>
Écart initial . . . . .		— 70	— 174	+ 20	— 59,06
34	— 1	— 69,3	— 82	0	— 2,64
35	+ 1	77,5	91	99	3,64
		— 61,8	— 165	119	— 58,06
22	— 1	— 38,5	— 40	— 42	— 2,73
14	+ 1	80	39	44	1,10
		— 70,3	— 166	121	— 59,69
22	— 1	— 38,5	— 40	— 42	— 2,73
20	+ 1	85	45	121	1,85
		— 73,8	— 161	200	— 60,57
19	— 1	— 84,9	— 42	— 50	— 0,66
15	+ 1	30	41	0	0,38
		— 78,7	— 162	150	— 60,75
19	— 1	— 34,9	— 42	— 50	— 0,66
22	+ 1	38,5	40	42	2,73
		— 75,1	— 164	142	— 58,78
34	— 1	— 69,3	— 82	0	— 2,64
25	+ 1	50	67	0	1,69
		— 94,4	— 179	142	— 59,73
30	— 1	— 55,7	— 52	— 151	— 2,42
32	+ 1	61,8	57	165	3,88
		— 88,3	— 174	156	— 58,27
6	— 10	— 230	— 250	0	— 3,40
15	+ 10	300	410	0	3,80
		— 18,3	— 14	156	— 57,87
15	— 2	— 60	— 82	0	— 0,76
22	+ 2	77	80	84	5,46
		— 3,3	— 16	240	— 58,17
30	— 1	— 55,7	— 52	— 151	— 2,42
33	+ 1	65	88	76	3,31
		+ 8	20	165	— 52,27
19	— 3	— 104,7	— 126	— 150	— 1,98
8	+ 3	76,8	111	0	3,48
		— 19,9	5	15	— 50,77
19	— 5	— 174,5	— 210	— 250	— 3,30
22	+ 5	192,5	200	210	13,65
		— 1,9	— 5	— 25	— 40,42
8	— 1	— 25,6	— 37	0	— 1,16
15	+ 1	80,0	41	0	0,38
Écart final . . . . .		2,5	— 1	— 25	— 41,20

Ces tâtonnements nous conduisent à une autre solution acceptable comme le montre sa vérification ci-dessous :

Exploitat. N°	$p'$	$h$	$b$	$a$	$s$
2	806	12 806	18 538	0	0
6	326	7 498	8 150	0	110,84
8	2	51,2	74	0	2,32
14	1	30	39	44	1,10
15	10	300	410	0	3,80
19	236	8 236	9 912	11 800	155,76
20	1	35	45	121	1,85
22	414	15 939	16 560	17 388	1 130,22
25	1	50	67	0	1,69
30	183	10 192,6	9 516	27 633	441,86
32	1	61,8	57	165	3,88
33	1	65	88	76	3,31
34	118	8 176,4	9 476	0	311,52
35	1	77,5	91	99	3,64
TOTAL . . . . .		63 608,5	73 023	57 326	2 171,79
Au lieu de . . . . .		63 608	73 024	57 351	2 214
Écart . . . . .		+ 0,5	— 1	— 2,5	— 42,21

R ne se réduit pas au point S, puisque nous avons trouvé par tâtonnement une seconde solution il en existe donc un grand nombre. Il est impossible de trouver manuellement la meilleure; tout au plus pouvons nous préférer la seconde à la première suite aux calculs des  $\Sigma p^2$  correspondants.

Pour la 1<sup>re</sup> solution  $\Sigma p^2 = 461\ 335$

Pour la 2<sup>e</sup> solution  $\Sigma p^2 = 453\ 728$

## VII. — RELATION AVEC UN CERTAIN PROGRAMME LINÉAIRE

Un problème complémentaire du précédent aurait pu être traité avec un calculateur électronique et aurait apporté de façon pratique des informations très utiles.

On aurait pu transformer notre problème en un programme linéaire.

La fonction économique serait l'estimation du revenu global de la population, les variables seraient les poids à affecter à chaque entreprise de l'échantillon — ou plus exactement les poids diminués de 1 —; les coefficients techniques de la matrice seraient les caractéristiques des entreprises.

La matrice comporterait autant de lignes que de contraintes linéaires retenues et autant de colonnes que d'entreprises.

Les exploitations jugées aberrantes seraient affectées du poids unité et ne seraient pas portées dans la matrice.

Cette méthode ne donnerait pas une estimation du revenu de la population. Mais la maximisation et la minimisation de la fonction économique nous fourniraient deux valeurs encadrant l'estimation recherchée. Il est évident que le résultat ne serait intéressant que si la fourchette obtenue était assez étroite, mais la fourchette pourrait venir à disparaître si les contraintes se révèlent incompatibles.

Nous présentons dans la page 273 la matrice correspondant à notre problème, prête à être codifiée pour passer sur ordinateur.

Le niveau retenu pour chaque ressource correspond à la détermination des poids diminués de 1.

Cette opération simple et rapide, le programme existant déjà, pourrait venir en complément des méthodes précédentes pour nous renseigner sur la précision des résultats obtenus par celles-ci.

Fonction économique				9 081	30 573	12 458	29 640	16 309	— 2 500	14 810		43 216	33 872	31 836
Ressources	Niveau	Unités	Signe	1	2	3	4	5	6	7		85	36	37
Surface	63 895	Hectare	=	13,6	16,0	18,5	18,2	21,6	23,0	24,0		77,5	90,8	97
Blé commercialisé	57 883	Quintal	=	57				24				99	163	350
Bovins présents	72 352	Tête	=	18	23	25	34	31	25	38		91	186	137
Main-d'œuvre salariée	2 226	Années travail	=	0,03		0,13	1,01	0,02	0,34	0,33		3,64	5,14	3,86
Nombre d'entreprises dans classe 1	806	Entrep.	=	1	1	1	1							
Nombre d'entreprises dans classe 2	990	Entrep.	=					1	1	1				
Nombre d'entreprises dans classe 3	307	Entrep.	=									1	1	1

#### VIII. — CALCUL DU REVENU DES EXPLOITATIONS DE 10 HA ET PLUS DU BESSIN

Il ne nous reste qu'à appliquer les poids ainsi déterminés aux revenus des exploitations de notre échantillon. Seulement nous ne devons peut être pas le faire « aveuglément » en ce qui concerne les 6 exploitations nos 2, 6, 19, 22, 30 et 34 ayant un poids différent de 1. Il faut s'assurer au préalable que ces revenus ne sont pas fortement influencés par une situation particulière ou une production sans sol très développée.

A l'observation des revenus, nous constatons que seule l'entreprise n° 2 de la classe 1 a un revenu nettement « déformé » par une production porcine de beaucoup plus importante que celles couramment rencontrées dans cette classe d'exploitations de la population. Il nous faut rectifier le revenu de cette exploitation pour le rendre « représentatif » de ceux de sa classe. Cette rectification en elle-même n'est pas « heureuse ». Il aurait peut être mieux valu essayer de se servir de l'exploitation n° 3, mais comme nous l'avons déjà souligné, la surface et le nombre de bovins plus élevés de celle-ci rendaient plus aléatoire l'obtention d'une solution. Toutefois, le fait d'avoir conseillé les exploitants pendant cinq ans nous autorise à envisager cette correction.

Pour ce faire, nous avons retranché du revenu de l'entreprise n° 2 la fraction de revenu tiré de sa production porcine, puis réintroduit un certain « revenu porc » correspondant à une évaluation du revenu moyen obtenu dans les exploitations de cette classe de superficie à partir de la production porcine.

Établissant pour les exploitations suivies par le Centre de Gestion, en plus du compte d'exploitation un compte par marges <sup>(1)</sup>, il nous est facile de connaître le revenu de la production porcine et de le soustraire du revenu global de l'exploitation.

Nous avons le compte suivant :

« produit porc » . . . .	16 843 F
« charges porc » . . . .	5 882 F
« revenu porc » . . . .	10 961 F

1. On appelle marge la différence entre la valeur d'une production et les charges variables affectables à cette production.

Il est moins facile d'évaluer le « revenu porc moyen » des exploitations de 10 à 20 ha de la population. Nous avons deux renseignements :

- le nombre de truies à l'hectare relevé dans l'exploitation n° 2 à savoir 0,41;
- le nombre moyen de truies à l'hectare de 85 exploitations de 10 à 20 ha du Bessin, tiré de l'enquête « Échantillon Maître », à savoir 0,04.

Cette différence très grande nous apparaît refléter la réalité. Partant de là, nous avons évalué le revenu moyen recherché au 1/10 du « revenu porc » de l'exploitation n° 2.

Nous pouvons penser que c'est une évaluation par excès vu que dans l'exploitation n° 2, l'exploitant possède une forte compétence technique qui doit lui permettre des résultats financiers meilleurs par truie que ceux obtenus couramment par les producteurs de porcs.

Nous pouvons penser aussi que c'est une évaluation par défaut, vu que nous ne tenons pas compte par ce mode d'évaluation des revenus obtenus par les agriculteurs qui produisent des porcs gras en partant de porcelets achetés.

Admettant finalement notre évaluation, nous réintroduisons un revenu porc de 1 096 F, ce qui revient à diminuer le revenu global de l'exploitation n° 2 de

$$10\ 961 - 1\ 096 = 9\ 865\ \text{F.}$$

*Remarque.* Nous avons bien conscience que cette « opération » n'est qu'un pis aller. Heureusement que la classe 1 n'a qu'une importance relativement faible par rapport aux 2 autres classes.

Une fois ce revenu « rectifié », l'application des poids  $p$  aux revenus des 39 exploitations de notre échantillon nous donne un *revenu global pour les exploitations de 10 ha et plus du Bessin de 33 709 547 F.*

Que penser de cette estimation?

À l'échelon des régions agricoles, nous ne disposons pas de comptabilités régionales auxquelles nous pourrions nous reporter : mais l'Association normande d'Économie rurale appliquée (A. N. E. R. A.) dresse les comptes du département depuis 1961, suivant le plan de la comptabilité nationale (1). Nous pouvons nous reporter à ces comptes départementaux pour essayer de porter un jugement sur notre estimation. Pour ce faire, nous comparerons les surfaces et le revenu estimé des exploitations de 10 ha et plus du Bessin aux mêmes critères départementaux.

Il faut auparavant rendre ces critères comparables et faire disparaître les différences suivantes :

1° Les comptes ANERA indiquent un revenu brut, les comptes Gestion un revenu net, amortissements déduits.

2° Les comptes ANERA indiquent le revenu de la branche agriculture comprenant le revenu des bois et forêts et des jardins familiaux des non exploitants; les comptes Gestion sont rapportés à la surface, bois et forêts exclus.

3° Les comptes ANERA, comme la comptabilité nationale, pour la branche Agriculture, ne prennent en considération que les fermages versés. Les comptes Gestion comprennent en charges d'exploitation, la valeur locative des terres exploitées en propriété sous le vocable de fermages calculés.

4° Les comptes Gestion ne prennent pas en considération la contribution foncière des propriétés bâties contrairement aux comptes ANERA.

1. De même que la commission des comptes de la nation cherche à décontracter les comptes nationaux en comptes régionaux et départementaux, l'ANERA pense ultérieurement décontracter ses comptes départementaux à l'échelon des régions agricoles I. N. S. E. E.

5° Enfin il existe une dernière différence dans le calcul du revenu : les comptes ANERA comportent un poste intérêts versés qui ne figurait pas en 1963 dans les comptes d'exploitation établis par le Centre de Gestion (1).

Partant du revenu brut de l'agriculture établi par l'ANERA, il est nécessaire de lui apporter les modifications suivantes pour obtenir un revenu comparable au revenu agricole du Centre de Gestion.

Revenu brut de l'agriculture ANERA en milliers de francs	290 824
— Revenu brut des bois et forêts . . . . .	2 900 287 924
— Revenu brut des jardins familiaux des non exploitants	20 400 267 524
— Amortissements du matériel et des bâtiments . . .	43 708 223 816
— Fermages calculés . . . . .	38 737 185 079
+ Contribution foncière des propriétés bâties . . . . .	3 865 188 944
+ Intérêts versés . . . . .	9 125 198 069

Nous constatons alors que les exploitations de 10 ha et plus du Bessin occupent 14,4 % de la surface agricole départementale (65 415 sur 453 624 ha) et obtiendraient 17 % du revenu agricole (33 709 sur 198 069 milliers de francs).

La connaissance du département nous laisse penser que le revenu moyen à l'hectare est différent d'une région à l'autre et que la constatation ci-dessus est vraisemblable.

Le revenu moyen à l'hectare dans le Pays d'Auge, région totalement herbagère, au relief défavorable et jusqu'à une époque récente, peu perméable au progrès technique est inférieur au revenu moyen départemental.

Le revenu moyen à l'hectare dans le Bocage, région de petites exploitations familiales, supérieur à celui du précédent doit être voisin du revenu moyen départemental.

Le revenu moyen à l'hectare dans la plaine de Caen-Falaise est supérieur à la moyenne, influencé par le revenu des exploitations de « grande culture » de la plaine de Caen.

Le revenu moyen à l'hectare du Bessin peut être supérieur au revenu moyen départemental, mais le résultat obtenu ci-dessus correspond à un écart (515 F contre 437 F) qui nous paraît fort (même compte tenu que dans notre raisonnement — assez grossier par ailleurs — nous négligeons les exploitations de moins de 10 ha du Bessin). Que cette méthode nous fournisse une estimation un peu forte du revenu de la région ne nous surprend pas.

Nous « redressons » l'échantillon pour ce qui est de la structure des exploitations surface et main-d'œuvre salariée, des ventes de blé et du cheptel bovin, mais la différence de rendement par bovin qui peut exister entre les entreprises de l'échantillon et celles de la population subsiste.

Le lecteur peut se reporter à ce que nous avons déjà écrit page 258 sur le choix d'autres variables qui permettraient de tenir compte à la fois du nombre d'unités de production et du rendement par unité.

#### *Quatrième partie*

### TROISIÈME MÉTHODE D'ESTIMATION MÉTHODE DES CORRÉLATIONS

Nous n'allons plus chercher à redresser l'échantillon en affectant chacun des individus d'un certain poids mais plutôt à en tirer une « loi » applicable aux individus de la population.

1. Depuis 1964, devant l'intérêt de suivre l'endettement des agriculteurs et la nécessité de connaître leurs emprunts pour pouvoir les conseiller, ce poste a été introduit dans les comptes d'exploitation.



## I. — EXPOSÉ DE LA MÉTHODE

Nous avons implicitement supposé dans la mise en œuvre des 2 méthodes précédentes l'existence d'une relation entre le revenu global des exploitations et les 4 variables considérées.

Nous cherchons maintenant à établir cette relation en essayant d'ajuster un hyperplan d'estimation au nuage de points constitué par les exploitations de notre échantillon.

Existe-t-il une relation linéaire du type

$$r = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + e$$

entre le revenu  $r$  des exploitations et la valeur des 4 variables

$x_1$  leurs superficies

$x_2$  le nombre de bovins qu'elles possèdent au 1<sup>er</sup> janvier

$x_3$  le nombre de quintaux de blé commercialisé durant la campagne

$x_4$  le nombre d'unités travailleurs salariés employés?

Si oui, cette relation est-elle significative?

Pour chaque exploitation  $j$ , les variables  $(x_1 x_2 x_3 x_4)_j$  sont connues avec précision. Nous faisons l'hypothèse, vraisemblable, que la valeur observée du revenu  $r_j$  est susceptible de variations aléatoires  $e$  autour de sa valeur théorique moyenne  $r'_j$  et que ces fluctuations aléatoires suivent la même loi normale (ce qui est moins vraisemblable).

Dans ces conditions, la méthode des moindres carrés nous permet de déterminer les coefficients  $a_0, a_1, a_2, a_3, a_4$  tels que

$$\Sigma (r_j - r'_j)^2 \text{ soit minimum}$$

Si nous parvenons à une régression significative, nous pourrons alors estimer le revenu de l'« exploitation moyenne » de la population en donnant aux variables  $x$  les valeurs moyennes telles qu'elles ressortent des statistiques déjà vues sur la population.

Ce revenu moyen multiplié par le nombre d'exploitations de 10 ha et plus du Bessin nous fournira l'estimation recherchée du revenu global de celles-ci.

Il est possible qu'il existe une bonne corrélation par classe de superficie 10-20, 20-50, 50 ha et plus, et qu'il n'en existe pas pour l'ensemble des 3 classes. Mais comme nous ne disposons pas de statistiques par classe concernant  $x_1 x_2 x_3$  et  $x_4$ , nous ne pourrions pas en tirer une estimation du revenu.

De plus les exploitations de la classe 10-20 ha sont trop peu nombreuses pour permettre d'établir une corrélation pour celle-ci. Pour ces 2 raisons, nous regroupons les 3 classes pour établir la corrélation cherchée.

## II. — CALCUL DE LA CORRÉLATION — SA SIGNIFICATION

La relation cherchée est du type

$$r = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + e$$

Posons  $R = r - \bar{r}$      $X = x - \bar{x}$ . Cette relation devient

$$R = b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + e'$$

Appliquons la méthode des moindres carrés, les 4 coefficients d'estimation  $b$  (ou de régression) sont solutions du système

$$\begin{cases} b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + b_3 \Sigma X_1 X_3 + b_4 \Sigma X_1 X_4 = \Sigma X_1 R \\ b_1 \Sigma X_2 X_1 + b_2 \Sigma X_2^2 + b_3 \Sigma X_2 X_3 + b_4 \Sigma X_2 X_4 = \Sigma X_2 R \\ b_1 \Sigma X_3 X_1 + b_2 \Sigma X_3 X_2 + b_3 \Sigma X_3^2 + b_4 \Sigma X_3 X_4 = \Sigma X_3 R \\ b_1 \Sigma X_4 X_1 + b_2 \Sigma X_4 X_2 + b_3 \Sigma X_4 X_3 + b_4 \Sigma X_4^2 = \Sigma X_4 R \end{cases}$$

Avant d'effectuer les calculs, nous avons jugé bon de laisser de côté les 2 plus grandes exploitations nos 38 et 39 et de leur attribuer un poids  $p = 1$ , pour les 2 raisons suivantes :

1° Il semble trop hardi de rechercher une relation unique valable de 10 à 180 ha. En se limitant à l'intervalle 10 à 97 ha, nous englobons la quasi-totalité des exploitations de la population; les exploitations de plus de 100 ha sont de rares exceptions dans le Bessin.

2° Ces 2 exploitations ont un revenu extrêmement faible, « aberrant » face à leur superficie et risque de fausser la corrélation établie sur l'échantillon.

Remarquons qu'il n'était pas gênant dans la 2<sup>e</sup> méthode de les conserver dans l'échantillon, à condition de leur affecter un poids unité, ce que nous avons fait.

Nous pouvons également prévoir les signes des coefficients  $b_1, b_2, b_3, b_4$ .

De toute évidence, le revenu doit croître avec la surface et le nombre de bovins :  $b_1$  et  $b_2$  doivent être positifs, il en est de même du signe de  $b_3$ , si le blé est « rentable » dans cette région herbagère.

Par contre la main-d'œuvre salariée crée une charge importante et par conséquent un signe négatif pour le coefficient  $b_4$  semble normal.

Les coefficients d'estimation sont solution du système

$$\begin{cases} 15\ 091\ b_1 + 18\ 635\ b_2 + 39\ 715\ b_3 + 846\ b_4 = 4\ 025\ 009 \\ 18\ 635\ b_1 + 26\ 518\ b_2 + 36\ 801\ b_3 + 986\ b_4 = 5\ 420\ 145 \\ 39\ 715\ b_1 + 36\ 801\ b_2 + 504\ 197\ b_3 + 2\ 958\ b_4 = 13\ 469\ 167 \\ 846\ b_1 + 986\ b_2 + 2\ 958\ b_3 + 61\ b_4 = 204\ 078 \end{cases}$$

Ce système a été résolu en inversant la matrice constituée par les coefficients des premiers membres des 4 équations :

Nous avons obtenu :

$$\begin{aligned} b_1 &= 137,70 \\ b_2 &= 151,52 \\ b_3 &= 14,96 \\ b_4 &= -1\ 743,25 \end{aligned}$$

Ces coefficients expriment la variation moyenne du revenu entraînée par une variation élémentaire — petite — d'une des variables, les autres étant supposées garder la même valeur.

Autrement dit,

— 1 hectare de plus procure un revenu supplémentaire de . . . . .	137 F
— 1 bovin de plus au 1 <sup>er</sup> janvier . . . . .	151 F
— 1 quintal de blé commercialisé de plus . . . . .	15 F
— 1 unité travailleur salarié entraîne par contre une baisse du revenu de	1 743 F

Ces variations nous paraissent cohérentes et très vraisemblables, connaissant l'économie de ces exploitations.

Nous avons donc la relation suivante :

$$R = 137,70 X_1 + 151,52 X_2 + 14,96 X_3 - 1\ 743,25 X_4$$

Dans la mesure où nous avons le droit de faire les hypothèses ci-dessus sur les fluctuations aléatoires du revenu autour d'une valeur moyenne et l'indépendance des variables (1), le test F nous renseigne sur la signification de la régression calculée.

L'analyse de la variance donne les résultats suivants :

Source de variation	Somme des carrés	Degrés de liberté	Variance
Liaison . . . . .	1 221 525 625	4	$s_1^2 = 305\ 381\ 406$
Résiduelle . . . . .	3 227 067 294	32	$s_2^2 = 100\ 845\ 852$
Variance totale . . . . .	4 448 592 919	36	

$$F = \frac{s_1^2}{s_2^2} = 3,05 > F_{32}^4(0,05) = 2,56$$

La variance expliquée par la liaison est significativement supérieure à la variance résiduelle. Nous dirons que la corrélation linéaire multiple établie est significative.

Revenons aux variables  $x$ . La relation permettant d'estimer le revenu moyen d'une exploitation connaissant pour celle-ci la valeur des 4 variables considérées est

$$r' = 6\ 374 + 137,70 x_1 + 151,52 x_2 + 14,96 x_3 - 1\ 743,26 x_4$$

### III. — ESTIMATION DU REVENU GLOBAL DES EXPLOITATIONS DE 10 HA ET PLUS DU BESSIN

Remplaçons dans l'équation ci-dessus,  $x_1 x_2 x_3 x_4$  par les valeurs moyennes de ces variables pour l'ensemble des 3 classes de la population;

$$\left. \begin{aligned} \bar{x}_{1 \text{ pop}} &= \frac{65\ 415}{2\ 140} = 30,6 \text{ hectares} \\ \bar{x}_{2 \text{ pop}} &= \frac{75\ 340}{2\ 140} = 35,2 \text{ bovins} \\ \bar{x}_{3 \text{ pop}} &= \frac{61\ 836}{2\ 140} = 28,9 \text{ qx de blé} \\ \bar{x}_{4 \text{ pop}} &= \frac{2\ 290}{2\ 140} = 1,07 \text{ unité travailleur salarié} \end{aligned} \right\}$$

nous obtenons une estimation du revenu moyen des exploitations de 10 ha et plus du Bessin.

$$\bar{r}_{\text{pop}} = 16\ 181 \text{ F}$$

Il est intéressant de comparer cette estimation au revenu moyen observé dans l'échantillon gestion à savoir :

$$\bar{r} = 18\ 807 \text{ F}$$

Nous reviendrons plus loin sur cette comparaison.

Nous en déduisons une estimation du revenu global R des exploitations de 10 ha et plus du Bessin en multipliant le revenu  $\bar{r}_{\text{pop}}$  par le nombre d'exploitations de la population.

$$R = 16\ 181 \times 2\ 140 = 34\ 627\ 340 \text{ F}$$

$$R = 34\ 627\ 340 \text{ F}$$

Si nous rapprochons cette estimation de la précédente, obtenue par la méthode révisée des poids  $R = 33\ 709\ 547 \text{ F}$  nous constatons des résultats similaires.

1. Les variables choisies ne sont pas en réalité indépendantes. Une analyse factorielle préalable pourrait permettre un choix plus judicieux. Mais n'oublions pas la contrainte actuelle des statistiques connues à l'échelon de la région agricole.

Il ne nous est pas possible de calculer la précision de cette estimation, l'échantillon gestion n'étant pas tiré au hasard.

Mais, étant donné la grande dispersion des revenus dans l'échantillon à l'image de la dispersion que l'on peut probablement observer dans la population, nous pensons qu'il n'est pas sans intérêt de calculer cette précision dans l'hypothèse — que l'on ne peut pas faire — d'un échantillon tiré au hasard.

Ce calcul nous donne un écart type de 673 F. Autrement dit, une erreur de 2 écarts types sur l'estimation du revenu moyen  $\bar{r}_{pop}$  correspondrait à une précision de 8 %. Cela laisse à penser que nos 2 estimations de R ne sont pas significativement différentes (1).

L'échantillon gestion est trop réduit et trop mal distribué pour pouvoir espérer une précision supérieure. Il nous permet néanmoins une approche intéressante du revenu global de la population considérée.

#### IV. — REMARQUES IMPORTANTES

Le revenu moyen estimé par exploitation est de 16 181 F pour une superficie de 30,6 ha.

Le revenu moyen des exploitations suivies par le Centre de Gestion est de 18 807 F pour une superficie de 41,6 ha.

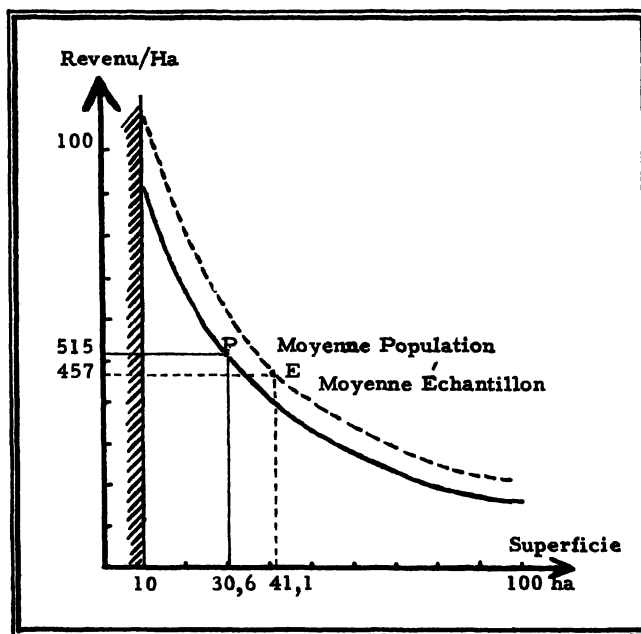
Autrement dit, pour une superficie moyenne supérieure de 33 %, le revenu moyen dans l'échantillon gestion n'est supérieur que de 16 %. Il pourrait en être déduit trop hâtivement que les résultats obtenus par les exploitants « en gestion » sont en fait moins bons à superficie égale : ce serait ignorer ou oublier que dans la région, le revenu à l'hectare diminue quand la superficie augmente.

La diminution de l'intensification en région herbagère et la part croissante du travail fait par des salariés par rapport au travail total fourni, entraînent, quand la superficie

Si nous ne regardions que ces seuls critères, superficie et revenu à l'hectare, nous pourrions ajuster la population et l'échantillon à 2 courbes dont les positions l'une par rapport à l'autre seraient celles représentées ci-contre; ces courbes passent notamment par les points moyens

$$P \left( 30,6 - \frac{16\,181}{30,6} \right)$$

$$\text{et } E \left( 41,1 - \frac{18\,807}{41,1} \right)$$



1. En revanche  $r_{pop}$  est significativement plus petit que  $r$ .

augmente, une baisse du revenu à l'hectare <sup>(1)</sup>. La baisse de production et les charges de salaires plus élevées ne sont pas compensées par la diminution du montant de charges fixes ramenées à l'hectare (main-d'œuvre, matériel, frais généraux).

Ceci nous conduit à faire les 2 remarques suivantes qui nous paraissent très importantes.

1° Il est impossible d'estimer le revenu de la population à partir de la connaissance du revenu moyen à l'hectare d'un groupe d'exploitations.

2° Il est regrettable que, faute de statistiques et d'un échantillon plus important, il ne soit pas possible d'utiliser cette méthode par classe de superficie, ce qui pourrait augmenter de beaucoup la précision de notre estimation.

### *Cinquième partie*

## COMPARAISON DES 3 MÉTHODES

Chacune des méthodes mises en œuvre dans les chapitres précédents présente des avantages et des inconvénients; nous pouvons essayer de les résumer brièvement dans le tableau synthétique ci-dessous :

Méthodes	Avantages	Inconvénients
I — Méthode initiale des poids	— Pas de tâtonnement  — Tient compte dans sa solution de tous les individus de l'échantillon	— N'est pas applicable dans le cas d'échantillon présentant des individus très « en pointe » — Nécessite des calculs relativement longs
II — Méthode révisée des poids	— La représentation graphique de certains aspects du problème renseigne sur les chances de trouver une solution acceptable	— Tâtonnements nécessaires — Rôle important donné à un petit nombre d'individus
III — Méthode des corrélations	— Méthode simple — Calculs simples	— Cerne de moins près la réalité en ne tenant pas compte du nombre d'exploitations par classe de superficie

Cette comparaison ne permet pas de choisir une méthode comme étant meilleure que les autres. Si la simplicité est un avantage important, la précision des résultats ne doit pas trop en souffrir.

Il peut être conseillé de les appliquer toutes les trois, ou deux des trois si cela est possible, afin de retenir comme résultat la synthèse des résultats obtenus. Si on ne peut en appliquer qu'une seule, le choix de celle-ci dépendra de l'échantillon et des statistiques disponibles.

Nous pouvons ainsi envisager 4 cas :

1° L'échantillon ne comprend pas trop d'exploitations « en pointe »; les statistiques nécessaires ne sont pas disponibles par classe de superficie : la première méthode sera préférée.

1. Il n'en est pas de même dans les exploitations de grande culture de la plaine de Caen.

2° L'échantillon comprend trop d'exploitations « en pointe » mais un nombre suffisant d'exploitations « moyennes »; les statistiques ne sont pas disponibles par classe de superficie : la deuxième méthode sera préférée.

3° L'échantillon comprend quelques exploitations très « en pointe »; les statistiques ne sont pas disponibles par classe de superficie.

Si le revenu à l'hectare dépend beaucoup de la superficie, la deuxième méthode sera préférée.

Si le revenu à l'hectare dépend peu de la superficie, la troisième méthode sera préférée.

4° Si nous avons enfin à déterminer le revenu d'une population dans laquelle les systèmes d'exploitation sont tels que le revenu à l'hectare dépende beaucoup de la superficie, si les statistiques nécessaires sont disponibles par classe et si l'échantillon possède un nombre minimum d'environ 10 exploitations par classe, la troisième méthode dite des corrélations pourra encore être préférée.

Ce nombre minimum d'exploitations, nécessaire pour que la corrélation établie par classe ait un sens, varie en toute logique, suivant que cette classe de la population est plus ou moins homogène.

## CONCLUSION

La première conclusion qui s'impose est que *notre échantillon était un « très mauvais » échantillon*. Il n'était même pas « représentatif » des exploitations de 10 hectares et plus comme l'a prouvé la comparaison des valeurs moyennes des variables considérées dans l'échantillon et dans la population.

*C'est un échantillon tronqué et biaisé; ce n'est pas pour autant un échantillon inutilisable pour l'estimation du revenu global d'une certaine population, à condition de le traiter d'une façon particulière appropriée.*

Sur 3 méthodes à l'essai, deux ont donné le même résultat, une est apparue inutilisable dans le cas d'échantillon possédant un trop grand nombre d'exploitations « de pointe ». Le résultat est donc finalement positif, non pas tant par l'estimation obtenue — qui nous paraît par ailleurs acceptable — que par les enseignements à en tirer pour généraliser l'emploi de ces méthodes à d'autres régions.

1° L'estimation ne peut concerner que les classes de superficie suffisamment représentées dans l'échantillon Gestion.

C'est la raison pour laquelle dans de nombreuses régions de polyculture élevage, les exploitations de moins de 10 ha seront laissées de côté.

Il est nécessaire, avant de chercher à estimer le revenu des autres exploitations, de s'assurer que ces petites exploitations n'ont pas une importance relative trop grande par leur nombre ou la superficie qu'elles occupent.

2° Néanmoins un effort doit être fait pour augmenter le nombre d'exploitations suivies en Gestion dans les classes de superficie insuffisamment représentées. Il s'agira assez souvent de la classe des 10 à 20 ha.

3° Un nombre minimum de statistiques connues avec une bonne précision par région agricole et autant que possible par classe de superficie est nécessaire. Ces statistiques doivent

correspondre à des variables caractéristiques des systèmes d'exploitation et ayant une influence sur le revenu.

4° Il est préférable de soustraire de l'échantillon les exploitations trop aberrantes qu'il pourrait contenir.

5° L'extrapolation du revenu moyen de l'échantillon gestion au prorata de la superficie qu'il occupe est sans valeur.

*Ces conditions respectées, il serait possible d'utiliser les échantillons des Centres de Gestion pour estimer le revenu des populations dont ils sont issus.*

Les échantillons restant à peu près les mêmes d'une année sur l'autre, il pourrait même être envisagé de suivre l'évolution de ces revenus. Pour cela, il est nécessaire que les statistiques régionales utilisées soient précises. Dans la mesure où elles seront obtenues par sondage, il semble préférable de disposer de sondages moins fréquents mais faits avec un taux de sondage plus élevé, surtout pour les enquêtes de structure; mieux vaudra par exemple un sondage tous les 3 ans au 1/3 que tous les ans au 1/10. Cela ne paraît pas utopique, vu le petit nombre de statistiques nécessaires. Dans ce cas, l'évolution des revenus sera suivie en conservant pendant cette courte période — de 3 ans — les mêmes coefficients de pondération ou la même corrélation — ce qui suppose également que la structure des revenus n'ait pas subi une évolution brutale. Mais ceci est un autre problème.

par Michel POURNIN

*Ingénieur d'Études*

*au Centre de Gestion et d'Économie rurale  
du Calvados (1)*

## DOCUMENTS ET STATISTIQUES UTILISÉS

P. THIONET. — Une façon d'exploiter un mauvais échantillon. 35<sup>e</sup> Session de l'Institut international de Statistique, Belgrade, 1965.

P. THIONET. — Séminaire Statistique-Économétrie. Université de Poitiers, janvier 1966 (Ronéo). Statistique agricole 1963, Ministère de l'Agriculture.

Premiers résultats de l'enquête au 1/10 pour les structures agricoles en 1963 (Échantillon Maître), Ministère de l'Agriculture. Supplément « Série Études » n° 5 Série jaune, juin 1965.

Enquête au 1/10 sur les exploitations en 1963. II Inventaires départementaux et éléments de structure, Ministère de l'Agriculture. Supplément « Série Études » n° 7, Série Jaune, septembre 1965.

Étude sur le cheptel bovin de la Normandie au 1<sup>er</sup> janvier 1963, Ministère de l'Agriculture.

Cahiers de l'Association normande d'Économie rurale appliquée. Cahier n° 2, septembre 1965.

Résultats du Bessin, exercice 59/60 - 60/61 - 61/62 - 62/63 - 63/64 - 64/65. Centre de Gestion et d'Économie rurale du Calvados.

Études et Conjoncture. Les comptes de l'Agriculture P. U. F., mai 1965.

Rapports et annexes présentés le 19 novembre 1964 sur l'utilisation des comptabilités d'exploitations à des fins d'observation économique. Commission des comptes de l'Agriculture de la Nation.

Études et Conjoncture. Méthode de la comptabilité nationale. Cadre et définition de la base 1959, n° 3, mars 1966.

1. Thèse présentée à l'Institut de Statistique de l'Université de Paris.

## DISCUSSION

M. G. THÉODORE.

## 1 — REMARQUES GÉNÉRALES SUR LE PLAN TECHNIQUE

L'exploitation du « mauvais échantillon gestion » repose sur une méthode dont le principe consiste à imaginer une procédure aléatoire qui aurait pu conduire à obtenir l'échantillon dont on dispose :

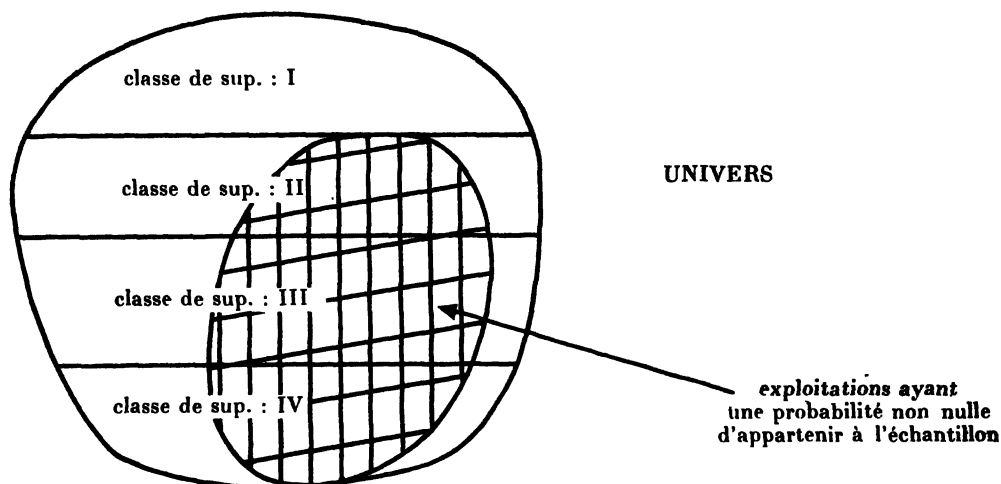
Chaque exploitation de l'échantillon est le résultat d'une tirage au sort, la probabilité de désignation de chacune étant inconnue. La détermination de ces probabilités fait l'objet du procédé préconisé et doit permettre d'utiliser l'échantillon malgré sa non « représentativité ». Cette notion de représentativité mérite d'être précisée.

Le sens que l'on donne couramment au terme « représentativité » de l'échantillon est : chaque unité de l'univers a la même probabilité (donc connue) d'appartenir à l'échantillon, et les coefficients d'extrapolation sont les mêmes pour toutes les exploitations échantillons. En ce sens l'échantillon « gestion » n'est pas représentatif mais cette seule raison ne suffit pas à en interdire l'exploitation.

Pour être cohérent avec le modèle proposé, il faut utiliser une autre notion de représentativité qui consiste à se demander quelles sont les unités que l'on peut considérer comme « représentées » dans l'échantillon, c'est-à-dire se demander quelles exploitations agricoles de l'ensemble de la population peuvent être considérées comme ayant eu une probabilité non nulle <sup>(1)</sup> d'appartenir à l'échantillon :

- de façon d'une part à conserver l'homogénéité des données externes nécessaires au calcul des coefficients mis en œuvre avec la population « représentée » par l'échantillon,
- de façon, d'autre part, à préciser le domaine pour lequel seraient valables les estimations ultérieures.

Cet aspect a été abordé et a conduit à éliminer les exploitations agricoles de moins de 10 ha : ce qui introduit implicitement l'hypothèse que la superficie est déterminante en ce qui concerne la probabilité d'appartenir à l'échantillon gestion, hypothèse qui reste à vérifier. En effet, si cette hypothèse n'était pas vérifiée <sup>(2)</sup>, cela reviendrait à mettre en cause toutes les données extérieures utilisées dans les calculs. Or, on peut fort bien imaginer la situation résumée par le schéma suivant :



· 1. ou plutôt non négligeable. — 2. au moins approximativement.



Les conclusions tirées de la méthode I vont dans ce sens : l'échantillon n'est pas « représentatif » des exploitations agricoles de 10 ha et plus. Ce qui signifie que l'information que fournit cet échantillon concerne une population plus restreinte. Il eût alors peut-être été utile d'essayer de quel critère permettrait de délimiter cette population. Mais, dans la mesure où l'on estime que la méthode I utilisée est valable, donc que l'on peut en retenir les conclusions, il paraît inacceptable de chercher à utiliser une autre méthode qui « permettrait » de fournir des estimations concernant un domaine plus vaste que celui que « représente », même mal, l'échantillon. Une seule méthode valable, pour être cohérent, pourrait consister à chercher une meilleure façon d'exploiter l'information contenue dans l'échantillon, mais en aucun cas il ne sera possible d'exploiter une information non contenue dans l'échantillon. Si l'on passe outre aux conclusions de la 1<sup>re</sup> méthode on se trouve conduit, comme dans la méthode II, à ne plus tenir compte de l'observation réelle, mais à introduire comme une donnée d'observation l'idée subjective du statisticien sur le sujet : ceci apparaît dans le calcul exposé sur le revenu des exploitations de 10 hectares et plus du Bessin, où l'on modifie les données d'une exploitation agricole avant de les utiliser dans l'estimateur bâti. La remarque du peu d'importance de la classe 1 ne justifie pas le procédé : en effet on peut admettre que si les données des exploitations des autres classes n'ont pas été modifiées, c'est qu'elles correspondaient à l'idée que se faisait le statisticien de ces classes.

Il faut ajouter que les tests prennent en compte quelques critères et que si la représentativité peut être correcte à leur égard, elle peut ne pas exister au regard d'autres aussi majeurs. On remarquera que les critères retenus dans le modèle sont caractérisés par un degré de corrélation élevé (notamment S. A. U. et bovins).

Enfin, en évolution, la variation de l'échantillon gestion peut être très différente de celle du domaine tout entier puisque la gestion tend à retenir les exploitations les plus dynamiques.

## II — DEMANDE DE PRÉCISIONS

— Les coefficients proposés doivent satisfaire à certaines contraintes (linéaires) dont on conçoit intuitivement la signification. Ces contraintes déterminent un domaine de solutions possibles. La solution retenue correspond à la minimisation d'une forme quadratique, mais il demeure à justifier le choix de cette expression.

— Peut-on donner des éléments sur la validité de la méthode dans les deux cas suivants :

- La probabilité d'appartenir à l'échantillon est liée à la valeur de la variable pour chaque unité.
- La probabilité d'appartenir à l'échantillon est indépendante de la valeur de la variable pour chaque unité.

## III — QUELQUES CRITIQUES DE DÉTAIL

### *Comparaison échantillon — population*

Si le nombre d'adhérents aux centres de gestion croît, l'information (au sens de l'auteur) ne croît pas obligatoirement beaucoup, en poussant le raisonnement, si l'échantillon

est fabriqué à partir des  $n$  meilleures exploitations, aucune manipulation ne reconstitue l'ensemble de la population mère.

*La variable  $x_0$  superficie agricole utile*

La limite inférieure de taille des exploitations retenue soit 10 ha, correspond à une coupe souvent utilisée mais est arbitraire.

*Poids  $p_{ci}$*

Les poids n'ont pas obligatoirement à être entiers ils ne « représentent » pas  $n$  exploitations, mais correspondent à l'inverse d'une probabilité.

*Exposé de la 2<sup>e</sup> méthode*

L'abandon de la variable  $t$  aurait pu être justifié.

*Calcul de la corrélation 3<sup>e</sup> méthode*

Un coefficient de régression négatif en ce qui concerne la main-d'œuvre n'est pas obligatoire, si, par exemple, cette variable main-d'œuvre est corrélée positivement avec des activités de production non prises en compte par les autres variables.

*Estimation du revenu global Bessin 3<sup>e</sup> méthode*

Une fois les coefficients déterminés (c'est-à-dire les inverses des probabilités) il n'est pas plus critiquable de former un estimateur de la variance qu'il ne l'est d'en former un de la moyenne (ce qui est l'objet du travail).

*Comparaison des trois méthodes*

En règle général, en statistique, on ne choisit pas diverses méthodes pour effectuer ensuite la synthèse des résultats. On recherche plutôt une méthode synthétique prenant en compte les trois méthodes initiales.

#### IV — PROBLÈMES POSÉS SUR UN PLAN PLUS ADMINISTRATIF DE CETTE RECHERCHE

On sait que l'Institut de Gestion et d'Économie rurale s'est particulièrement intéressé à ce problème parce que les centres de gestion disposent effectivement d'une documentation extrêmement riche. Il faut toutefois ne pas oublier qu'il s'agit soit d'exploitations de pointe, soit d'exploitations en perte de vitesse et il y a lieu également de se rappeler qu'en matière de gestion la doctrine entre la promotion d'un ensemble d'exploitations agricoles, ou l'intérêt à porter vers les seules exploitations de tête, n'est pas tranchée. Quoi qu'il en soit, il y a là l'existence d'une documentation extrêmement tentante, il se pose alors un problème d'utilisation de ce matériel numérique atypique.

Pour les raisons techniques précédemment indiquées, il convient d'être extrêmement prudent :

- en premier lieu avant d'entreprendre des calculs de produit ou de revenu agricole par petite région agricole, il paraîtrait plus urgent et plus positif d'améliorer les comptes au niveau national, de rechercher les mouvements interrégionaux de produits, de mettre en place un système d'observation de prix agricoles à la

production, de telle manière que la décontraction des comptes nationaux soit par grande région, soit par sous-branche, puisse être envisagé dans les moins mauvaises conditions possibles ;

- en second lieu, les comptes de l'agriculture ont fait l'objet d'une politisation grandissante durant ces dernières années et s'il est bon et réconfortant que la documentation statistique fasse l'objet d'une utilisation intensive, il est fondamentalement mauvais que la polémique soit alimentée par le heurt de sources discordantes, ou de « contre-chiffres » s'opposant au chiffre initialement présenté.

Si au niveau national la diversité des sources, le frottement entre techniciens, permettent néanmoins de conclure le dialogue, il serait imprudent et inutile de transposer ces discussions aux niveaux départemental et régional. L'objectif même « de généraliser l'emploi de ces méthodes à d'autres régions » est incontestablement intéressant sur le plan de la recherche, mais sur le plan de la politique économique générale les centres de décision sont placés à un autre niveau et il est peu rentable de rechercher par région agricole des séries numériques en vue de suivre annuellement ou tous les trois ans l'évolution du revenu de la petite région agricole. Cette voie de recherche de l'A. N. E. R. A. ne peut être que source de confusion et de gâchis de moyens.

Par contre nous paraît beaucoup plus prometteuse l'utilisation de ces méthodes pour analyser les réactions traduites par des fonctions de production à diverses mutations de prix, d'investissements, de transferts de spéculation, etc., mais il s'agit d'objectifs assez différents de ceux exposés dans la communication.

M. H. GUITTON. — Je crois que votre étude a une portée plus large que le domaine agricole où vous l'avez menée, ne pourrait-on pas dire que tout échantillon est mauvais, et que la vocation même de la Statistique est d'utiliser les notations imparfaites pour en tirer cependant la part de vérité qu'elles recouvrent, pour s'approcher le plus possible de l'expression parfaite, donc irréalisable, des réalités ?

Sans doute votre échantillon est-il particulièrement mauvais ; bien qu'il soit tronqué et biaisé, vous nous montrez comment, convenablement traité, il est tout de même utilisable. L'exploitation que vous faites ainsi des comptes des Centres de Gestion du Calvados me paraît d'un très grand intérêt.

J'ai scrupule à parler de la technique même des trois méthodes que vous proposez, devant celui qui a inspiré vos recherches, le professeur THIONET. Je suis cependant frappé de la très faible différence entre les résultats de la méthode classique et relativement simple des corrélations, qui, comme vous le dites « cerne de moins près la réalité » et des méthodes des poids. Faudrait-il donc désespérer de vouloir toujours mieux serrer cette réalité ?

Ce qui me tourmenterait davantage serait d'ordre économique plus que statistique. C'est le fait que vous ne posiez pas le problème de la définition du revenu agricole, et de la valeur même de la définition que l'on adopte. Je sais bien que le but de la loi d'orientation agricole du 5 août 1960 est d'arriver à connaître « le niveau de la rémunération du travail et du capital agricole ». Mais si jamais on avait l'idée d'utiliser vos travaux pour comparer le revenu agricole au revenu non agricole, et à fonder sur ces résultats des mesures de « politique des revenus » si à l'ordre du jour, j'assume que j'aurais quelque inquiétude.

La notion de revenu est une des plus délicates de l'analyse économique. C'est pourquoi, avant d'utiliser la technique des estimations, il faudrait être sûr que la matière première sur

laquelle on travaille n'est pas elle-même tronquée, qu'elle concerne en tout cas la même valeur relative à travers les périodes où on utilise les méthodes d'estimation.

Au fond, le traitement statistique des données, quoique toujours délicat, est moins préoccupant que la définition même de ces données. Mais, comme vous le dites vous-même dans votre dernière phrase à propos de la structure des revenus, ceci est un autre problème, et ce n'est pas celui que vous aviez en vue.

M. THIONET. — J'entends certains d'entre nous parler entre eux de programme linéaire; et je désire faire très rapidement le point sur l'aspect purement mathématique du problème auquel M. POURNIN s'est heurté.

Il s'agit d'un programme, non pas linéaire mais quadratique, et de ce fait non traitable par une technique banale. On considère des inconnues (les poids) soumises à un système :

— d'équations linéaires (L)

— d'inéquations linéaires (I) de la forme :  $p - 1 \geq 0$  cependant qu'on entend minimiser une fonction quadratique (convexe) (Q) desdites inconnues  $p$ .

Il a été possible de représenter les  $p$  par un point, astreint à se trouver dans un certain corps convexe (C), dont on ignore à l'avance s'il est vide ou non. Ayant pu trouver un point  $p^0$  de (C), c'est que (C) n'est pas vide; et il reste à se déplacer dans (C) de telle sorte que la valeur de (Q) descende de  $Q^0$  jusqu'à son minimum.

En fait, cette dernière opération n'a été que très partiellement menée, en tâtonnant, sans moyens de calcul. Mais la recherche de  $p^0$  demanda beaucoup de travail : sur les bases initiales du problème (C) était bel et bien vide. Autrement dit le programme était impossible, car il y avait *trop de tout* dans l'échantillon.

On avait surtout trop de tracteurs, et il a fallu renoncer à imposer la contrainte « tracteurs ». Mais on avait aussi (c'est paradoxal) *trop de salariés*, et il a fallu *adoucir* la contrainte : ce qui signifie : remplacer les salariés par une variable plus réaliste : le nombre d'années de travail salarié, englobant les travailleurs à temps partiel, les travailleurs saisonniers, et pas seulement ceux à temps plein.

M. POURNIN a parlé de 3 méthodes. La dernière (celle de corrélations) est classique et je ne m'y attarderai pas. Il a ajusté le plan des moindres carrés sur ses données. Guidé par le livre d'Analyse statistique de MM. MORICE et CHARTIER, il s'est demandé si ce plan de régression était significatif ou non; cela n'avait de sens que dans le cadre d'hypothèses stochastiques à préciser. Celles d'indépendance qu'il a faites sont d'usage courant et (bien entendu) sont très contestables; précisons bien qu'il ne s'agissait pas d'un travail d'économétrie, non plus que sur l'analyse « multivariate ».

De toute façon je crois que la *robustesse* des estimations des moindres carrés (alors que les hypothèses classiques cessent d'être vérifiées) n'est plus guère contestée par les économètres.

Mais revenons sur la partie originale du travail, c'est-à-dire sur les deux premières méthodes.

La première méthode était celle que j'avais expérimentée sur des comptabilités de « quincailleurs » avec la Caisse des Dépôts. Elle consiste à traiter par le mépris les inéquations (I) dont j'ai parlé plus haut.

Alors les contraintes (L) font qu'on recherche l'extremum lié de (Q), ce qui, — en

utilisant les multiplicateurs de Lagrange, — conduit à un système d'équations du 1<sup>er</sup> degré. On le résout par inversion de matrice, — ou autrement, cela va de soi.

Pour les quincailliers, on avait trouvé *a posteriori* que les inéquations (I) étaient vérifiées, à l'exception de quelques cas marginaux sans portée pratique. Avec les exploitations agricoles de M. POURNIN, il a bien fallu reconnaître que ce n'était pas le cas. Les contraintes (I) passent au premier rang de nos soucis, tandis que, pour la fonction (Q), on tire profit des enseignements du travail antérieur en choisissant (Q), la plus simple possible.

On voudra bien noter que les conditions (I + L) peuvent se traduire par un modèle constitué par un hyperpolyèdre (H) et un point donné (M) : tout se ramène à la question de savoir si (M) est ou n'est pas dans (H). Le problème est résoluble en théorie en calculant un nombre fini de déterminants. En pratique, faute d'une ordinateur, c'était inextricable à la main; on a donc cherché à voir (H) par ses projections sur des plans.

Il se trouve que, chaque fois (M) se projette à l'intérieur du contour apparent de (H), ce qui est nécessaire mais non suffisant pour que (M) soit vraiment dans (H). En fait (M) était très voisin d'une face de (H), et initialement à l'extérieur; en substituant la contrainte : Salaires, à la contrainte : Salariés permanents, le nouvel (M) est bien dans le nouvel (H).

\*  
\* \*

Un auditeur s'est étonné du choix de  $Q = Sp^2$ ; je réponds que  $S(p - \bar{p})^2 = Sp^2 - (Sp)^2/n$ . La somme des poids étant imposée (c'est le nombre d'exploitations agricoles au total), rendre (Q) minimum équivaut à rendre minimum la somme des carrés des écarts des poids autour de leur moyenne (ou plutôt de leurs moyennes de classe), moyen commode d'écrire que les poids ne sont pas très différents les uns des autres.

M. DEVRON a suggéré (trop tardivement) la programmation linéaire pour explorer (C) et voir si (C) est vide ou non; c'est extrêmement pertinent et à retenir pour la prochaine fois.

Il faut insister sur le fait qu'on a imité la théorie des vrais sondages; je reconnais que ce sondage fictif n'est pas d'une logique éblouissante, s'il a l'avantage de fournir des résultats concrets.

M<sup>lle</sup> de MENDITTE

1) A-t-il été tenu compte de l'*auto-consommation* dans l'appréciation des revenus agricoles et quelle en a été sa mesure?

Il semble, à la réponse de M. POURNIN, que l'*auto-consommation* pouvait être négligée dans le cadre de son étude en raison de sa pénible valeur dans le cas précis.

2) Dans le cadre de comparaisons dans le temps, comment est-il envisagé de tenir compte des variables suivantes :

— *variable du rendement* de la production (très influencée par les conditions atmosphériques en dehors de toute autre considérations telles que emplois d'engrais);

— *variable « surface de culture »* particulièrement importante dans le cas de cultures annuelles qui sont étendues ou réduites suivant le prix de vente de la dernière récolte;

— *variable « prix de vente »* très sensible à la loi de King : l'inélasticité de la demande de produit agricole entraîne l'effondrement des cours dans le cas de récoltes excédent les besoins du marché traditionnel.

3) Ne pourrait-on introduire une sorte de *radio composite du type UGB* <sup>(1)</sup> permettant d'établir des équivalences calculées entre mécanisation agricole et animaux de traits? Cela eut permis — peut-être — de résoudre les problèmes de coûts comparatifs, entre les exploitations sur le plan équipement.

4) L'intérêt de la méthode exposée par M. Pournin serait — me semble-t-il — plus grand pour le calcul des prix de revient agricoles. Les bases des variables y paraissent plus stables.