



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

La représentativité du RICA

M. Jean-Marc Boussard, I. Foulhouze

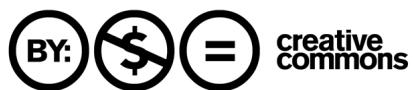
Citer ce document / Cite this document :

Boussard Jean-Marc, Foulhouze I. La représentativité du RICA. In: Économie rurale. N°137, 1980. pp. 29-35;

doi : <https://doi.org/10.3406/ecoru.1980.2704>

https://www.persee.fr/doc/ecoru_0013-0559_1980_num_137_1_2704

Fichier pdf généré le 08/05/2018



Abstract

Can the trench RICA (Réseau d'Information Comptable Agricole) be used for estimating the national values of accounting criteria measured on each of the farms belonging to the sample? In order to answer this question, 18 criteria have been estimated by various methods, and compared to their known value at the national level. It was impossible to find likely and coherent estimations of these 18 criteria together, in the sense that the differences between computed and observed values of the criteria were greater than that which could have resulted from sampling errors. This result casts a doubt upon the representative character of the RICA sample. However, the problem could perhaps have been solved by using a more suitable typology of the farm belonging to the sample.

Résumé

Différentes méthodes ont été utilisées pour retrouver, à partir de l'échantillon RICA français, les valeurs nationales de 18 critères connus à la fois au niveau « France entière » et au niveau de l'échantillon. Aucune de ces méthodes ne permet d'obtenir, pour les extrapolations obtenues, des valeurs vraisemblables, en ce sens que les différences entre les estimations, et les valeurs observées au niveau national ne peuvent être attribuées aux erreurs d'échantillonnage. Ceci jette un doute sur le caractère représentatif de l'échantillon RICA. Cependant, un regroupement des exploitations du RICA différent du regroupement « OTE » permettrait peut-être de résoudre le problème.



LA PRÉSENTATIVITÉ DU RICA

J.M. BOUSSARD et I. FOULHOUZE *

INRA Station d'Economie et de Sociologie Rurales, Paris

Différentes méthodes ont été utilisées pour retrouver, à partir de l'échantillon RICA français, les valeurs nationales de 18 critères connus à la fois au niveau « France entière » et au niveau de l'échantillon. Aucune de ces méthodes ne permet d'obtenir, pour les extrapolations obtenues, des valeurs vraisemblables, en ce sens que les différences entre les estimations, et les valeurs observées au niveau national ne peuvent être attribuées aux erreurs d'échantillonnage. Ceci jette un doute sur le caractère représentatif de l'échantillon RICA. Cependant, un regroupement des exploitations du RICA différent du regroupement « OTE » permettrait peut-être de résoudre le problème.

REPRESENTATIVITY OF THE RICA SAMPLE

Can the french RICA (Réseau d'Information Comptable Agricole) be used for estimating the national values of accounting criteria measured on each of the farms belonging to the sample? In order to answer this question, 18 criteria have been estimated by various methods, and compared to their known value at the national level. It was impossible to find likely and coherent estimations of these 18 criteria together, in the sense that the differences between computed and observed values of the criteria were greater than that which could have resulted from sampling errors. This result casts a doubt upon the representative character of the RICA sample. However, the problem could perhaps have been solved by using a more suitable typology of the farm belonging to the sample.

Le RICA (Réseau d'Information Comptable Agricole) est un échantillon d'environ 3 000 comptabilités d'exploitations françaises. Il est bien connu des économistes ruraux, dont il constitue une source d'information précieuse. Bien qu'il soit exagéré de dire que les mesures effectuées soient sans erreurs, l'affinement des concepts comptables et le sérieux des enregistrements dépassent, dans cette enquête, tout ce qui a été fait dans le genre. Par ailleurs, des réseaux analogues existent dans les autres pays du Marché Commun, et même dans certains pays qui n'adhèrent pas à la CEE, comme l'Espagne. Ceci devrait permettre des comparaisons internationales instructives.

Malheureusement le RICA français, comme ses homologues étrangers, n'est pas « représentatif »

au sens statistique de ce terme. Les exploitations de l'échantillon RICA ont été choisies sur une base non aléatoire, par un « choix raisonné » (qui, en pratique, dépend largement de la bonne volonté des exploitants pressentis). L'ensemble E des exploitations qui font partie de l'échantillon RICA est donc représentatif d'un ensemble E d'exploitations qui est inclus dans l'ensemble E_T des exploitations agricoles françaises, mais qui n'est pas identique à E_T .

De ce fait, si l'on entreprend d'estimer à partir de E les moyennes de divers indicateurs économiques tels que le revenu par tête ou par ha, capital par tête ou par ha, etc., il faut s'attendre à ce que l'écart entre ces estimations et les véritables moyennes nationales soit supérieur à celui qui résulte d'une simple erreur d'échantillonnage. La question qui se pose alors est de savoir s'il ne serait pas possible d'utiliser quand même l'infor-

* Les auteurs tiennent à exprimer leur reconnaissance à J.P. LEY, du département de biométrie de l'INRA, et à un lecteur anonyme de la revue *Economie Rurale* pour les améliorations qu'ils ont suggérées à une précédente version du présent travail. Les auteurs restent cependant seuls responsables des erreurs subsistantes.

mation fournie par le RICA pour obtenir des estimations raisonnables de ces moyennes nationales.

En effet, on peut penser que le système de sélection par « choix raisonnable » des exploitations qui appartiennent à e fournit une assez grande variété de situations pour qu'en affectant une pondération convenable à chacune d'entre elles, il soit possible de retrouver les moyennes nationales cherchées. Plus précisément, étant donné X_i moyenne nationale du critère i qui prend la valeur x_{ij} sur l'exploitation j de l'échantillon e , d'effectif J , on suppose que l'échantillon est assez diversifié pour qu'il existe J nombres w_j compris entre zéro et un, tels que :

$$[1] \quad x_i = \sum_{j=1}^J w_j x_{ij} + \varepsilon_i, \forall i$$

$$\sum_j w_j = 1$$

où ε_i est une variable aléatoire de moyenne nulle.

Mais comment calculer les w_j ? Et en admettant que nous trouvions un tel système de coefficients d'extrapolation, que vaudraient les résultats obtenus? La variance de ε_i serait-elle assez faible pour que $\hat{X}_i = \sum_j w_j x_{ij}$ puisse être considéré comme une bonne estimation de X_i ? Telles sont les questions auxquelles nous essaierons de répondre dans cette étude.

Extrapolation et stratification

Avant d'aller plus loin, nous remarquerons que J est assez grand, de l'ordre de 3 000. Il ne serait guère possible pratiquement de calculer 3 000 w_j . Il faut donc regrouper les exploitations j en un nombre raisonnable de strates k . Au lieu de rechercher J valeurs de w_j , nous nous bornerons à chercher K valeurs de w_k , telles que :

$$[2] \quad X_i = \sum_{k=1}^K w_k \bar{x}_{ik} + \varepsilon_i$$

(que l'on peut encore écrire $X_i = \hat{X}_i + \varepsilon_i$), avec :

$$[3] \quad \bar{x}_{ik} = \frac{1}{n_k} \sum_{j \in k} x_{ij}$$

où n_k est l'effectif de la strate k dans e .

Il est clair que la façon d'effectuer ce regroupement n'est pas indifférente pour notre problème : plus les strates seront différentes les unes des autres, et plus nous aurons de chances de trouver un système de w_k vérifiant l'équation [2]. En l'occurrence, nous nous sommes bornés à utiliser la stratification en 35 « OTE » (orientation technico-économique) proposée par les auteurs du RICA. Nous étions du reste obligés de nous contenter de cette stratification, du fait que c'est la seule ventilation publiée des résultats du RICA, et que nous n'avions pas accès aux données individuelles, exploitation par exploitation. Comme on le verra, il n'est pas sûr que cette décision ait été heureuse. Ceci étant, sur quelles bases calculer les 35 w_k cherchés ?

Application de la méthode de Thionet pour l'extrapolation de mauvais échantillons

Le problème ainsi posé serait insoluble s'il n'existe pas d'autre sources d'information que le RICA lui-même. Mais de telles sources extérieures existent.

La statistique agricole française publie les chiffres sur les valeurs nationales de certains des critères agrégables, mesurés également sur les exploitations du RICA. Le concept de « surface en blé » est identique dans la statistique agricole annuelle et dans la RICA, et on admettra que pour ce critère, comme pour les autres, les deux valeurs sont mesurées avec une erreur négligeable [4].

On peut donc répartir l'ensemble S des critères agrégables mesurés par les exploitations du RICA en deux sous-ensembles S_1 et S_2 : pour $i \in S_1$, on connaît à la fois x_{ij} et X_i . Pour $i \in S_2$, on ne connaît que x_{ij} . On cherche à obtenir \hat{X}_i pour $i \in S_2$ et on ne s'intéresse aux \hat{X}_i pour $i \in S_1$, que parce qu'ils vont nous fournir l'information qui nous manque pour calculer les coefficients d'extrapolation.

Il faut choisir les critères i convenables de S_1 , c'est-à-dire tels qu'il existe des mesures fiables au niveau de E_T pour ces critères. Le nombre de critères doit être en rapport avec le nombre d'inconnues w_k . La liste des 18 critères retenus a été établie en fonction de ces considérations, et figure dans les tableaux qui suivent. Ces critères ont été rapportés à l'hectare SAU pour éliminer l'effet « taille » [2].

Cela fait, la connaissance de S_1 nous permet d'utiliser les travaux de Thionet (1969, 1978), sur l'extrapolation des « mauvais échantillons » [3].

En posant $\varepsilon_i = 0 \forall i$, l'équation [2] écrite pour tout $i \in S_1$, définit pour les w_k inconnus, un domaine réalisable à l'intérieur duquel on peut choisir un point optimal $w^* = \{w_k\}$ sur la base d'un critère secondaire : par exemple, on minimise la somme des variances des \hat{X}_i ou l'entropie des \hat{X}_i .

Cette méthode conduit à résoudre le problème A :

Optimiser $f(w_k)$ sous :

$$[4] \quad \sum_k w_k \bar{x}_{ik} = X_i \text{ pour } i \in S_1$$

$$\sum_i w_k = 1$$

Bien évidemment, ce n'est pas parce que [4] est vérifié pour $i \in S_1$ que $\hat{X}_i = X_i$ pour $i \in S_2$. Mais s'il existe un vecteur w qui vérifie [2] avec $\varepsilon_i = 0$ pour tout i , il doit en particulier vérifier [4], et c'est ce qui justifie le choix de cette contrainte, laquelle incorpore toute l'information disponible sur le problème.

Les résultats du problème A obtenus dans ces conditions n'ont pas répondu à notre attente. Le domaine Z défini par [4] pour $i \in S_1$ est vide (du moins pour l'année 1974, sur laquelle nous avons travaillé). Il n'est pas possible de satisfaire simultanément l'ensemble des égalités [4] pour les 18 critères considérés. Il n'y a donc pas lieu de se poser de questions sans objet sur la forme de la fonction

$f(w_k)$. D'autre part, ce résultat laisse déjà planer un doute sur la possibilité de résoudre notre problème.

Une généralisation de la méthode de Thionet

Cependant, les égalités [4] sont peut-être trop contraignantes. En effet même si le RICA était un échantillon aléatoire, il n'y aurait pas de raisons pour que [4] soit satisfait. Du fait des erreurs d'échantillonnage il est vraisemblable que

$$\hat{X}_i - X_i \neq 0.$$

On va donc admettre qu'il n'y a pas égalité stricte entre \hat{X}_i et X_i et on va s'intéresser à la vraisemblance de $\hat{X}_i - X_i$. Ceci suppose que l'on dispose de quelques informations sur la loi de probabilités de X_i .

Si l'on admet que les x_{ij} sont gaussiens, on peut obtenir une estimation de $\hat{\sigma}_{ik}$ de l'écart-type σ_{ik} de \bar{x}_{ik} . Pour cela on calcule l'estimateur S_{ik} de la variance de x_{ij} par :

$$[5] \quad S_{ik} = \frac{1}{n_k - 1} \sum_{j \in k} (x_{ij} - \bar{x}_{ik})^2$$

puis :

$$[6] \quad \hat{\sigma}_{ik} = \frac{1}{\sqrt{n_k}} S_{ik}$$

Alors une estimation de la variance de \hat{X}_i est donnée par :

$$[7] \quad V(\hat{X}_i) = \sum_k w_k^2 \hat{\sigma}_{ik}^2 = \hat{\sigma}_i^2$$

De ce fait, la variable :

$$[8] \quad \frac{\hat{X}_i - X_i}{\hat{\sigma}_i} = t_i$$

est une variable de Gauss centrée réduite.

Sans doute l'assimilation de x_{ij} (et donc de \bar{x}_{ik}) à une variable gaussienne est-elle une hypothèse assez forte. Pourtant, beaucoup d'études de théorie des sondages reposent sur des hypothèses

analogues, fondées sur le fait que \bar{x}_{ik} est la somme d'un assez grand nombre de variables dont on peut admettre qu'elles sont indépendantes.

Par ailleurs, les t_i sont indépendants si les \bar{x}_{ik} sont indépendants. Pour le démontrer, on remarque d'abord que si les variables non réduites $z_i = \bar{X}_i - X_i$ sont indépendantes, les variables réduites $t_i = z_i / \sigma_i$ sont également indépendantes. Il faut donc montrer que les z_i sont indépendants.

Par définition :

$$E(z_i z_{i'}) = (\sum_k w_k \bar{x}_{ik} - X_i) (\sum_k w_k \bar{x}_{i'k} - X_{i'})$$

$$\text{Si : } E\left[\sum_k w_k \bar{x}_{ik}\right] = \sum_k w_k E(\bar{x}_{ik}) = X_i$$

on a pour tout i et i' :

$$\begin{aligned} E(z_i z_{i'}) &= -X_i X_{i'} + E\left[(\sum_k w_k \bar{x}_{ik})(\sum_k w_k \bar{x}_{i'k})\right] \\ &= -X_i X_{i'} + A \end{aligned}$$

avec :

$$A = E\left[(\sum_k w_k \bar{x}_{ik})(\sum_k w_k \bar{x}_{i'k})\right] = E\left[\sum_k \sum_{k'} w_k w_{k'} \bar{x}_{ik} \bar{x}_{i'k'}\right]$$

Si \bar{x}_{ik} et $\bar{x}_{i'k}$ sont indépendants en probabilité,

$$\begin{aligned} A &= \sum_k \sum_{k'} w_k w_{k'} E(\bar{x}_{ik}) E(\bar{x}_{i'k'}) \\ &= \left[\sum_k w_k E(\bar{x}_{ik}) \right] \left[\sum_{k'} w_{k'} E(\bar{x}_{i'k'}) \right] = X_i X_{i'} \end{aligned}$$

Il en résulte que $E(z_i z_{i'}) = 0$, et par conséquent, $E(t_i t_{i'}) = 0$, ce qu'il fallait démontrer.

Ici encore, l'hypothèse que x_{ij} est indépendant de x_{ij} est discutable. Pour surmonter cette difficulté il aurait fallu rendre les \bar{x}_{ik} indépendants, en effectuant le changement de base défini par la matrice des vecteurs propres de la matrice de variance covariance des \bar{x}_{ik} . Faute de disposer de cette information, nous avons dû faire comme si les \bar{x}_{ik}

étaient indépendants, ce qui rend nos résultats un peu suspects. D'un autre côté, comme les corrélations négatives entre les x_{ij} sont probablement tout aussi nombreuses que les corrélations positives, il est probable que les ordres de grandeur obtenus pour les $\hat{\sigma}_{ik}$ sont corrects.

Dans ces conditions, un système quelconque, W , de coefficients d'extrapolation est d'autant plus vraisemblable que les t_i calculés par [8] sont plus semblables à des variables de Gauss centrées réduites.

En fait, l'ensemble des t_i représente une épreuve d'une loi de Gauss à l_1 dimensions, si l_1 est le cardinal de S_1 (c'est-à-dire, le nombre de critères appartenant à l'ensemble S_1 , sur lequel se fait la comparaison entre X_i et \hat{X}_i). Dans ces conditions, si t_i est indépendant de t_j , $\sum_i t_i^2$ suit une loi de χ^2 à l_1 degrés de liberté.

Ceci conduit au **problème B** :

Minimiser $\sum_i t_i^2 - l_1$, t_i étant donné par [8], sous la contrainte : $\sum_k w_k = 1$.

C'est ce qui a été fait en utilisant divers programmes (4) de minimisation mis à notre disposition par l'IRIA (5).

Nous avons obtenu pour le $\sum_i t_i^2$ minimal (6) une valeur de 149 (tableau 1). C'est beaucoup trop pour un χ^2 à 18 degrés de liberté. Ce résultat nous conduit encore une fois à rejeter l'hypothèse selon laquelle les écarts entre X_i et \hat{X}_i résultent d'erreurs d'échantillonnage. Les valeurs obtenues pour chacun des t_i confirment cette conclusion : le tableau 1 montre que les t_i ne prennent des valeurs acceptables, de l'ordre de 2, que pour un petit nombre de valeurs, telles que les surfaces en blé, en pommes de terre, en betteraves, ou les effectifs d'ovins, de porcins ou de bovins. Par contre, des valeurs de l'ordre de 15, comme pour la surface en faire-valoir direct, ou en vigne, ou les consommations intermédiaires, sont tout à fait innacceptables.

Tableau 1. – Résultat de la minimisation de t_i et t_i^2 pour les chiffres appartenant à S_1

Critère	Valeur des X_i par ha Estimation de la statistique agricole	Valeur des \hat{X} par ha Estimation tirée du RICA	Valeur des t_i^2	Valeur des t_i
Surface en faire-valoir direct49810855	.78016357	15.60683963	3.95054928
Surface en fermage44144494	.36538971	2.12526920	1.45782996
Surface en métayage02186630	.00586773	7.46897101	2.73294182
Consommations intermédiaires	1423.15151515	5643.22588840	27.28895761	5.22388388
Surface en blé11557576	.12663487	.71670746	.84658577
Surface en orge01921212	.06928594	18.80705007	4.33670959
Surface en maïs05778788	.24237834	9.25616362	3.04239439
Surface en pommes de terre00948485	.00868266	.10193297	.31911279
Surface en betteraves01618182	.01400860	1.72702350	1.31416266
Surface en pommiers03381818	.02625610	2.46331928	1.56949651
Surface en poiriers00169697	.01020629	4.86815372	2.20638930
Surface en pêchers00066667	.01350106	9.67617414	3.11065494
Surface en vigne00127273	.09215703	32.50218280	5.70106857
Surface en jachère00630303	.01162211	2.13415370	1.46087429
Effectif de chevaux01212121	.00443230	9.71674784	3.11716985
Effectif de bovins73472727	.88283389	2.22705756	1.49233293
Effectif d'ovins32444242	.32359962	.00072308	.02689023
Effectif de porcins45248485	1.08131888	2.20357887	1.48444564

Nous avons du reste cherché encore d'autres méthodes apparentées à la précédente, comme par exemple de minimiser la valeur du t_i maximum. Les résultats n'ont pas été plus satisfaisants : il est impossible de faire descendre la plus grande valeur de t_i en dessous de 5.

La conclusion à tirer de ces résultats est que l'échantillon e est trop éloigné de E_T pour qu'il soit possible, même en recherchant les pondérations les plus favorables possibles des strates de e , de retrouver, à partir de e , l'ensemble des grandeurs caractéristiques de la population E_T .

Naturellement, en diminuant le nombre des critères qui figurent dans l'ensemble S_1 , nous aurions pu finir par obtenir des résultats plus satisfaisants. Mais nous n'avons pas le droit intellectuel de pratiquer de telles manipulations. A ce propos, il faut se souvenir que l'obtention de résultats « satisfaisants » pour l'ensemble S_1 considéré ici est une condition nécessaire, mais non suffisante pour que

les estimations correspondantes des éléments de S_2 soient « bonnes ». Si même cette condition n'est pas satisfaisante, il est clair que tout autre système d'extrapolation moins contraignant n'a aucune chance d'être « bon ».

AUTRES POSSIBILITES

Les résultats précédents laissent peu d'espoir de trouver un système de coefficients d'extrapolation des moyennes relatives aux OTE qui garantisse la plausibilité des résultats obtenus. Faut-il abandonner tout espoir de se servir des résultats du RICA pour obtenir des estimations « France entière » ? Nous ne le croyons pas. Mais le travail sera beaucoup plus difficile et plus coûteux que celui qui a déjà été fait. Et le résultat n'est en tout cas pas garanti. Deux approches peuvent être envisagées :

a) Il se peut que le regroupement des exploitations au niveau des OTE ne soit pas heureux. Pour que les méthodes qui viennent d'être exposées

réussissent il faut que les valeurs de X_i se trouvent dans l'espace R^K , assez proche du polyèdre convexe défini par $X_i = \hat{X}_i$, $i \in S_1$. Ce n'est manifestement pas le cas. Mais le regroupement des exploitations en OTE a pour conséquence d'« aplatisir » ce polyèdre, si bien que X_i pourrait se trouver à l'intérieur du polyèdre défini par

$$\sum_{j \in e} w_j x_{ij} = X_i$$

sans se trouver à l'intérieur du polyèdre défini par

$$\sum_k w_k \bar{x}_{ik} = X_i$$

Cette observation conduit à reprendre les mêmes calculs que précédemment, mais sur la base d'un regroupement des exploitations de e différent de celui des OTE. On peut songer pour cela à l'emploi de l'analyse discriminante. Ceci supposerait la possibilité de partir des données originales du RICA, exploitation par exploitation, au lieu de partir des données publiées, OTE par OTE.

b) On peut également songer à rechercher des relations entre les valeurs de x_{ij} pour des $i \in S_1$, et de x_{ij} pour $i \in S_2$. Si de telles relations existent, et peuvent être établies à l'aide, par exemple, d'un calcul de régression linéaire sur les observations fournies par le RICA, il sera possible de calculer des estimations \hat{X}_i de X_i pour $i \in S_2$ à partir de la connaissance de X_i pour $i \in S_1$ dans n'importe quelle exploitation, même n'appartenant pas au RICA. Cette approche est celle de MOREL et VENNAT. Ils supposent que pour une exploitation de surface s_j , la valeur de x_{ij} / s_j est une fonction linéaire de s_j . Ainsi, pour tout i on a :

$$[9] \quad \frac{x_{ij}}{s_j} = a + b s_j + \varepsilon_j, \quad E(\varepsilon_j) = 0, \quad E(s_j \varepsilon_j) = 0$$

La valeur de ce type d'estimation dépend entièrement de la validité de l'hypothèse selon laquelle une équation du genre de [9] correspond bien à une régression de la variable expliquée sur la ou les variables explicatives. De ce point de vue, la relation [9] paraît du moins discutable. Il est regret-

table que MOREL et VENNAT n'aient fourni aucun test statistique de la validité de leurs estimations.

Il serait probablement plus justifié d'un point de vue théorique d'estimer une véritable fonction de production liant le vecteur y_j des outputs de l'exploitation j aux vecteurs des inputs q_j de cette même exploitation :

$$[10] \quad y_j = f(q_j) + \varepsilon_j, \quad E(\varepsilon_j) = 0, \quad E(\varepsilon_j q_j) = 0$$

La théorie économique pourrait fournir des informations sur la forme de la fonction f à retenir. A partir de là, il serait sans doute possible, connaissant certains des y_j et des q_j , d'en déduire les autres.

En tout état de cause, une telle entreprise serait une œuvre de longue haleine, et exigerait l'accès aux informations comptables individuelles – ce que nous n'avons malheureusement pas obtenu des administrations qui ont la charge du RICA.

CONCLUSION

Nous avons cherché à obtenir des estimations « France entière » d'un certain nombre de grandeurs agrégeables mesurées au niveau de chaque OTE du RICA, en multipliant chaque valeur relative à une OTE donnée par un même coefficient d'extrapolation attaché à l'OTE, et en effectuant la somme des résultats obtenus de cette façon. Pour cela, nous avons cherché à estimer les coefficients d'extrapolation par diverses méthodes, en optimisant un critère lié à la vraisemblance des extrapolations obtenues, lorsque la véritable valeur de l'agrégat national était connu. En dépit du faible nombre de critères utilisés dans cet exercice, au nombre de 18 seulement, les résultats ont toujours donné à notre critère de vraisemblance une valeur incompatible avec l'hypothèse que les différences entre les estimations et les véritables valeurs pouvaient s'expliquer par des erreurs d'échantillonnage. Il faut donc admettre que les résultats publiés par le RICA ne sont pas extrapolables au niveau national.

Il est possible cependant, que ce résultat négatif soit dû à la façon dont les résultats publiés par le RICA sont ventilés en « OTE ». Un autre regroupe-

ment des exploitations, selon une typologie différente de celle des « OTE » nous aurait peut-être permis d'obtenir de meilleurs résultats. Il aurait fallu, pour cela, avoir accès aux résultats individuels du RICA. De même, il n'est pas impossible de songer à utiliser les observations du RICA pour estimer les paramètres d'une fonction liant les critères inconnus aux critères connus. La validité d'une telle approche repose sur la validité de la spécification adoptée pour cette fonction à estimer.

De toute façon, les résultats obtenus jusqu'ici laissent planer un doute sur la valeur des politiques agricoles fondées sur les informations fournies par le RICA : même si ces politiques produisent des résultats escomptés sur les exploitations dont l'échantillon RICA est représentatif, elles produisent sûrement des résultats différents sur les exploitations qui ne sont pas représentées par le RICA, et nos résultats semblent indiquer qu'il ne s'agit pas d'un ensemble négligeable.

RÉFÉRENCES BIBLIOGRAPHIQUES

LEMARECHAL (C). – Non smooth optimisation and descent methods. *Research report*, N° RR-78-004, IASA (Vienne), 1978.

LEMARECHAL (C) et MIFFLIN (R). – *Non smooth optimisation*. Pergamon Press, Londres, 1978.

MOREL (B) et VENNAT (F). – Du réseau d'information comptable agricole aux comptes nationaux de l'agriculture. Un essai d'extrapolation de 1970 à 1975. Document de travail, INSEE, octobre 1978.

POURNIN (M). – Utilisation des échantillons des centres de gestion pour l'établissement des revenus régionaux en agriculture. *Journal de la Société de Statistique de Paris*, 1967, 4^e trimestre, pp. 245-289.

THIONET (P.). – Item analysis and reweighting in : new developments in survey sampling, JOHNSON and SMITH, ed. Wiley, 1969, pp. 282-304.

THIONET (P.). – *Quelques problèmes concernant les sondages*. Vandenhoeck et Ruprecht, Göttingen, 1978, (chapitre 8).

NOTES

(1) Les concepts comptables du RICA ne coïncident pas toujours avec ceux de la statistique agricole, ou des comptes de l'agriculture. De ce point de vue, MOREL et VENNAT (1978) fournissent d'excellentes tables montrant les correspondances entre les critères RICA et les agrégats nationaux.

(2) Ceci oblige du reste à modifier la contrainte $\sum_k w_k = 1$: l'ensemble des critères étant rapportés à l'hectare, la somme des coefficients de pondération est égale à N/S c'est-à-dire l'inverse de la surface moyenne (S) par exploitation (N étant le nombre d'exploitations françaises).

En outre, une autre contrainte également imposée dans les essais dont nous rendons compte, mais négligée dans l'exposition de la méthode pour ne pas l'alourdir exagérément, impose à la somme des surfaces par ha en faire-valoir direct, métayage et fermage, d'être égale à 1. Une telle contrainte est évidemment nécessaire dans un travail comme celui-ci.

Ces complications, qui ne changent pas la philosophie de la méthode, sont négligées dans l'exposé qui suit.

(3) Ces travaux ont déjà fait l'objet d'une application aux problèmes agricoles. Cf POURNIN (1967).

(4) Ils sont décrits dans LEMARECHAL (1978) ainsi que dans LEMARECHAL et MIFFLIN (1978).

(5) Nous tenons à cette occasion à exprimer notre reconnaissance à M. LEMARECHAL, ingénieur de recherche à l'IRIA.

(6) Avec cependant les réserves liées au fait que nous avons négligé la co-linéarité éventuelle des x_{ij} , et aussi au fait que la fonction $\sum_i t_i^2$ n'étant pas concernée, nous ne pouvons pas garantir que l'optimum trouvé n'est pas un optimum local.