

Distr.
GENERALE

CES/SEM.37/5 (Summary)
13 novembre 1997

FRANCAIS
Original : ANGLAIS

COMMISSION DE STATISTIQUE et
COMMISSION ECONOMIQUE POUR L'EUROPE

COMMISSION DES COMMUNAUTES EUROPEENNES
(EUROSTAT)

CONFERENCE DES STATISTICIENS EUROPEENS

Séminaire commun CEE/Eurostat sur les
registres des activités industrielles
et commerciales

(Genève, 19-20 février 1998)

**Le traitement des variables dans le cadre d'un registre des activités
industrielles et commerciales : une approche probabiliste**

Résumé du document présenté par l'Istat ¹ (Italie)

1. Le registre des activités industrielles et commerciales nationales consiste en une liste de toutes les entreprises opérant dans un pays, identifiées et enregistrées avec les valeurs individuelles de certaines variables structurelles; il convient de tenir cette liste à jour à des fins statistiques. En Italie, lorsque l'on veut créer et mettre à jour une telle liste, on est obligé, pour diverses raisons (coût, volume du travail dans le domaine statistique, temps disponible...) de tirer la plupart des données des fichiers administratifs.

2. Lorsque l'on établit un registre des activités industrielles et commerciales sur la base des renseignements fournis par plusieurs sources administratives, on doit s'attendre à rencontrer certains problèmes pour attribuer les valeurs des variables aux unités. Dans la pratique, pour une entreprise donnée figurant dans le registre, on observe souvent plusieurs signaux qui ne concordent pas. Quelquefois, on n'observe même pas de signal du tout, mais il faut quand même fournir les données.

3. On se heurte à un problème de choix pour chaque entreprise figurant dans le registre à propos de laquelle les renseignements sont incohérents, et il faut faire des estimations pour chaque unité sur laquelle on ne dispose d'aucune information. Des méthodes ont été mises au point pour faire face à ces principales catégories de problèmes.

¹Andrea Carone et Caterina Viviano, Istat (Italie).

4. Une approche probabiliste a été proposée qui consiste à utiliser une modélisation dépendant de variables. Par conséquent, i) on utilise des variables instrumentales à titre d'approximation des populations lorsque l'on traite les chiffres sur l'emploi selon les règles de la probabilité conditionnelle; ii) on utilise un modèle logistique de calcul des probabilités pour déterminer le type d'activité des entreprises; iii) on applique un processus de décision hiérarchique pour déterminer le code d'activité économique par le biais de choix imbriqués.

5. Ces méthodes pourraient être perfectionnées et adaptées par d'autres utilisateurs parce que le principe d'exploitation des informations administratives semble de plus en plus largement adopté.
