

---

**MODELE DE REGRESSION AVEC DOUBLE CENSURE EMBOITEE ET  
DEPENDANCES : UNE APPLICATION A LA DETECTION DE LA FRAUDE AUX  
COTISATIONS SOCIALES**

Anne ALIGON (\*), Denisa BANULESCU-RADU (\*\*), Sylvain BENOIT (\*\*\*), Raphaël  
BROSSEAU(\*), Christophe HURLIN (\*\*), Cédric VALLEE (\*)

(\*) Caisse Centrale de la MSA, Direction des Etudes, des Statistiques et des Fonds  
(\*\*) Université d'Orléans, Laboratoire d'Economie d'Orléans  
(\*\*\*) Université Paris Dauphine

denisa.banulescu-radu@univ-orleans.fr

**Mots-clés** (6 maximum) : fraude sociale, évaluation, double correction des biais, maximum de vraisemblance

**Domaine concerné** : Économétrie, Modélisation

---

**Résumé** (entre 350 et 900 mots environ)

La maîtrise des risques de fraude sociale et fiscale et la lutte contre le travail illégal sont des problèmes extrêmement importants qui entravent la justice sociale et l'efficacité économique. Cet article se propose ainsi d'évaluer le manque à gagner (MAG) à partir des données issues des contrôles réalisés par les caisses de la Mutualité Sociale Agricole (MSA) auprès des établissements employeurs de main d'œuvre relevant du régime agricole. Le MAG est défini comme la somme potentielle des redressements qui auraient pu être imposés à des entreprises ayant fraudé ou réalisé des déclarations sociales erronées, si elles avaient été effectivement contrôlées alors qu'elles ne l'ont pas été en réalité.

L'objectif de l'étude est double : il s'agit d'une part de définir précisément d'un point de vue théorique et statistique la notion de MAG, et d'autre part de valider les méthodes d'estimation de cette quantité. Pour ce faire, nous recourons à différentes simulations de Monte Carlo. Une fois ces approches validées sur données simulées, elles sont ensuite mises en œuvre sur données réelles fournies par la Mutuelle Sociale Agricole (MSA).

Premièrement, nous proposons une définition théorique des moments de la distribution conditionnelle du MAG. Formellement, le MAG peut être défini au niveau d'une entreprise ou au niveau de la population, cette dernière étant définie comme l'ensemble des entreprises soumises au régime agricole et gérées par la MSA. On note  $MAG_i$ , avec  $MAG_i \in \mathbb{R}^+$ , le manque à gagner associé à l'entreprise  $i = 1, \dots, n$ , où  $n$  désigne la taille de la population, et

$MAG = \sum_{i=1}^n MAG_i$ , le manque à gagner total. Cette définition est obtenue dans le cadre d'un modèle ou processus générateur de données (ou *data generating process*, DGP) spécifiant différentes équations de comportement relatives à la décision de contrôle des caisses de la MSA, à la décision de frauder des entreprises, et de la règle de fixation du montant du redressement, supposé égal au montant de la fraude.

Deux types d'agents sont donc considérés dans l'analyse : (1) les entreprises indicées par  $i = 1, \dots, n$  ; et (2) l'organisme de contrôle représentant les caisses de la MSA. Ces deux agents prennent un premier niveau de décision : i) chaque entreprise décide de frauder ou de ne pas frauder ; et ii) l'organisme de contrôle décide de contrôler ou de ne pas contrôler chaque entreprise. Il est important de noter que ces deux décisions ne sont *ni séquentielles, ni conditionnelles*. Le fait d'être contrôlé ne conditionne en rien le fait de frauder et vice et versa. En revanche, ces *décisions sont liées* : si l'organisme de contrôle remplit sa mission de façon efficace, on doit observer une plus grande probabilité de contrôle pour les entreprises ayant la plus forte probabilité de frauder. Sous des hypothèses de normalité, ce modèle s'apparente à un modèle Tobit avec un mécanisme de double censure (Amemiya, 1984). Dans ce DGP, le MAG est une variable aléatoire dont les réalisations sont par nature inobservables. Toutefois, conditionnellement au DGP, il est possible de caractériser les deux premiers moments (espérance et variance) de sa distribution, qui nous serviront à prévoir le MAG pour chaque entreprise et construire des intervalles de confiance. Du fait de la dépendance entre les processus de contrôle et les décisions de fraude des entreprises, nous montrons que les moments conditionnels du MAG ne dépendent pas de simples ratios de Mills, mais d'espérances et de variances de lois normales multivariées avec double troncature (Manjunath et Wilhelm, 2012).

Deuxièmement, nous procédons à la validité des formules théoriques des moments du MAG grâce à différentes simulations de Monte Carlo. A partir du DGP, des réalisations individuelles du MAG sont simulées pour toutes les entreprises non contrôlées qui ont effectivement fraudé. Une réalisation du MAG agrégé est ensuite déduite pour l'ensemble des entreprises non contrôlées. Cette approche est répétée un grand nombre de fois, et les réalisations sont comparées à l'espérance théorique du MAG agrégé et à des intervalles de confiance construits à partir de la variance théorique du MAG. Les simulations démontrent le bien fondé des formules théoriques proposées lorsque les paramètres du modèle sont connus.

Troisièmement, nous nous intéressons à l'estimation des paramètres du modèle. En effet, l'espérance et la variance de la distribution conditionnelle du MAG dépendent des paramètres du DGP, a priori inconnus. Nous évaluons ici deux stratégies d'estimation : l'approche par maximum de vraisemblance (MV) et l'approche d'Heckman (1976). Nous montrons que cette dernière n'est pas adaptée dans le contexte de notre modèle car non seulement elle ne permet pas d'estimer les corrélations entre les termes d'erreur, mais qu'elle conduit en outre à des estimations biaisées des paramètres d'intérêt (Greene, W., 2006). Finalement, nous estimons le modèle sur données réelles et calculons le MAG.

## Bibliographie

- [1] Amemiya T., « Tobit models: a survey », *Journal of Econometrics*, 24(1-2), 3-61, 1984.
- [2] Greene, W., « A general approach to incorporating selectivity in a model », 2006
- [3] Heckman, J., « The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models », *Annals of Economic and Social Measurement*, Volume 5, number 4, National Bureau of Economic Research, p. 475- 492, 1976.
- [4] Manjunath, B.G. and Wilhelm, S., « Moments calculation for the doubly truncated multivariate normal density », arXiv, 1206.5387.335, 2012