



Small Area Estimation-personnes

P. Eichenberger, A. Massiani, J.-M. Nicoletti

Section des méthodes statistiques METH

Office fédéral de la statistique

*Séance d'information RFP
10 octobre 2013*



Le projet Small Area Estimation-personnes

- But : Améliorer les estimateurs dans les petits domaines. On a décidé de commencer par l'estimation du nombre de personnes actives dans les communes ou groupe de communes en Suisse.
- Les petits domaines sont donc ici les communes ou groupe de communes.
- Difficulté : estimation à l'aide de méthodes d'estimation directes ne sont pas efficaces pour de petites communes.
- Solution : utiliser des méthodes d'estimation basées sur les petits domaines.



Mandat à l'université de Trier

- Mandat a débuté en 2010 et pris fin début 2012.
- Données considérées : le recensement 2000.
- Différents modèles étudiés, dont un en particulier :
La variable dépendante (variable réponse)

$$Y_i = \begin{cases} 1 & \text{si l'individu } i \text{ est actif,} \\ 0 & \text{sinon.} \end{cases}$$

Le modèle linéaire :

$$Y_i = \alpha + Age_{i,1} + Age_{i,2} + Age_{i,3} + Genre_i + \\ GrandesRegions_{i,1} + \dots + GrandesRegions_{i,8} + \varepsilon_i$$



Mandat à l'université de Trier

- Résultats : Une nette amélioration de la précision dans un grand nombre de communes, cependant :
 - ▶ Pour certaines communes, le biais est grand, pas d'explication sur la cause.
 - ▶ Aucune solution sur la détection de ces communes problématiques n'a été proposée.
 - ▶ Aucun diagnostic du modèle.
- Risqué de publier des résultats pouvant être fortement biaisés et peu vraisemblables dans certains cas.



1er essai d'amélioration du modèle

- Amélioration du modèle : Proposition de méthodes de diagnostic du modèle et ajout des interactions entre les variables.
- On se base toujours sur les données du recensement 2000.
- Résultats : Les variables explicatives disponibles dans le recensement 2000 ne permettent pas d'améliorer significativement le modèle.
- Conclusion : recherche de nouvelles variables auxiliaires.



Comparaisons des évaluations des modèles

Modèle de l'Uni Trier	Nouveau modèle
82.26%	82.72%

TABLE : Pourcentage des d'individus correctement détectés
comme actifs ou non actifs.

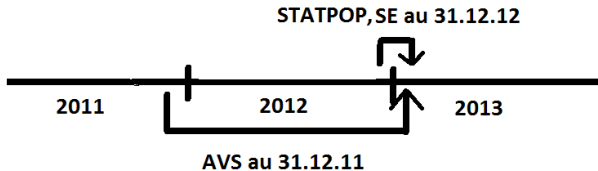


Utilisation des données AVS

- Développer un modèle sur les données de l'enquête structurelle (SE2012), utilisant les données STATPOP2012.
- A disposition : une variable provenant de l'AVS, indiquant si un individu a cotisé à l'AVS dans l'année 2011 et une variable indiquant le montant du salaire soumis à l'AVS.
- Délai → avril 2014.

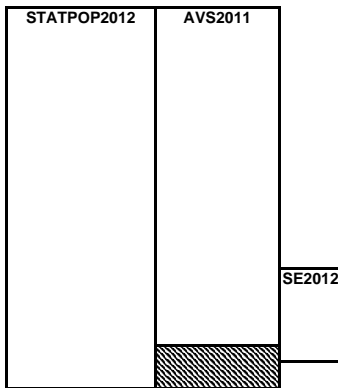


Utilisation des données AVS





Utilisation des données AVS





Utilisation des données AVS

- Modèles considérés pour l'instant :

$$M_1 : Y_i = \alpha + \text{existenceCotisationAVS}_i + \varepsilon_i$$

$$M_2 : Y_i = \alpha + \text{existenceCotisationAVS}_i + \text{Age}_{i,1} \\ + \text{Age}_{i,2} + \text{Age}_{i,3} + \text{Genre}_i + \text{etatCivil}_i \\ + \text{nationalite}_i + \varepsilon_i,$$



1er résultats

M_1	M_2
90.73%	90.74%

TABLE : Pourcentage des d'individus correctement détectés
comme actifs ou non actifs.



Conclusion

- Un modèle n'utilisant que les variables usuelles n'est pas suffisant.
- La variable existence d'une cotisation AVS améliore clairement le modèle. Cependant cette variable peut être améliorée, en tenant compte du salaire, par exemple.