

The rise of small area estimation and R in the Romanian official statistics

Ana Maria Dobre

National Institute of Statistics
Romania

8ème Colloque Francophone sur les Sondages
Université de Bourgogne
19 Novembre 2014

① Introduction

Analyse de littérature

② La méthodologie basée sur l'estimation sur petits domaines et R

Description du modèle

Diagnostic du modèle

③ Conclusions

Autres possibilités de recherche

Introduction

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle
Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

- Le manque de disponibilité des chiffres exacts sur l'émigration a conduit à la nécessité d'une nouvelle réflexion statistique basée sur des modèles économétriques.
- Etant donné la difficulté de mesurer le phénomène de la migration internationale, les statistiques officielles roumaines ont élaboré une méthodologie basée sur les techniques d'estimation sur petits domaines.

Solution informatique

La solution informatique pour l'application des techniques d'estimation sur petits domaines a été R.

Analyse de littérature

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction
Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle
Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

- Rao, J.N.K. (2003), Small Area Estimation
- Sarndal, C. (1984), Design-consistent versus model-dependent estimation for small domains Journal of the American Statistical Association

Ces méthodes ont été mises en œuvre par Breindanbach (2011) dans le paquet JoSAE en R.

Description du modèle (I)

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

- Les estimations sur petits domaines visent à améliorer la précision des estimations lorsque les méthodes standard ne sont pas assez précises.
- En Roumanie, la méthodologie appliquée de l'Estimation sur petits domaines est utilisée pour produire des données sur les sorties annuelles sur la migration internationale.

Description du modèle (II)

Deux sources de données sont combinées:

- *l'Enquête sur la population active (EPA)*, une enquête statistique contenant la variable d'intérêt et un ensemble de covariables
- *le Recensement de la population et des ménages* contenant les mêmes variables et une variable auxiliaire (population résidente).

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

Description du modèle (III)

Les variables utilisées dans le modèle:

- Variable indicateur binaire pour la migration (variable de l'enquête): absent
- Covariables: sexe; groupe d'âge; zone de résidence; niveau d'éducation; activité; état civil.

Les données sur la migration de l'EPA se composent de variables binaires au niveau de l'unité.

La variable d'intérêt pour le modèle est absent(e)

- 1 - Si les gens sont absents de chez eux pendant plus de 12 mois;
- 0 - Sinon.

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

Description du modèle (IV)

Solution informatique

La méthode de calcul utilisée est le paquet *JoSAE* et le paquet *nlme* de logiciel R.

- Le modèle est basé sur des modèles de régression linéaire avec des effets mixtes.
- Les coefficients de régression sont calculés par la méthode de maximum de vraisemblance restreint (Restricted Maximum Likelihood - REML).
- Le paquet *JoSAE* en R produit trois types d'estimateurs: l'estimateur **GREG**, l'estimateur **Synth** et l'estimateur le meilleur prédicteur linéaire sans biais empirique (**MPLSB** ou **EBLUP**).

Diagnostic du modèle (I)

Le paquet *JoSAE* produit une trame de données avec de nombreuses entrées pour chaque domaine:

- moyenne des résidus d'un modèle linéaire et le modèle à effet linéaire mixte;
- la variance des moyens pour l'échantillon;
- les composants de la variance de EBLUP et les erreurs types dérivés des écarts.

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

Diagnostic du modèle (II)

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie

basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

**Diagnostic du
modèle**

Conclusions

Autres
possibilités de
recherche

En outre, ont été calculés:

- les mesures modèle de sélection: le Critère d'Information Bayésien (BIC);
- le Critère Akaike Information (AIC);
- les coefficients de variation.

Conslusions

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

Diagnostic du
modèle

Conslusions

Autres
possibilités de
recherche

- Les techniques d'estimation sur petits domaines ont été utilisées pour obtenir des estimations au niveau des départements (NUTS 3), avec des coefficients de variation fiables.
- En fait, le modèle est bien sûr perfectible; il reste certains défis tels que le caractère non-migration de l'EPA, la disponibilité des données ou l'impossibilité d'estimer les niveaux de désagrégation.

Autres possibilités de recherche

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

Les techniques d'estimation sur petits domaines pourraient être appliquées avec succès dans de divers domaines de statistiques sociales et économiques:

- la mesure de la pauvreté
- l'état du travail
- le niveau d'éducation

Reconnaissance

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle

Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

Le présent document fait partie du projet de recherche de l'équipe roumaine qui travaille sur R - "Romanian R Team". Je suis reconnaissant à l'équipe pour la grande collaboration. Aussi, je tiens à donner une reconnaissance particulière à Nicoleta Caragea, qui est l'auteur de la méthode et a exprimé son orientation sincère pour ce projet.

The rise of
small area
estimation
and R in the
Romanian
official
statistics

Ana Maria
Dobre

Introduction

Analyse de
littérature

La
méthodologie
basée sur
l'estimation
sur petits
domaines et R

Description du
modèle
Diagnostic du
modèle

Conclusions

Autres
possibilités de
recherche

Merci pour votre attention!

anamaria.dobre@insse.ro

www.r-project.ro