

Combinaison des données au niveau de l'unité et de la région pour l'estimation des petites régions au moyen de modèles multiniveaux pénalisés

Jan Pablo Burgard, Joscha Krause et Ralf Münnich¹

Résumé

Dans les applications sur petits domaines, on dispose souvent de données au niveau de l'unité et de la région. L'utilisation conjuguée de données au niveau de l'unité et de la région aux fins d'une estimation sur petits domaines dans un modèle unique pose plusieurs problèmes méthodologiques. Premièrement, cela implique la nécessité d'estimer un nombre croissant de paramètres de modèle. Il faut sélectionner minutieusement les variables afin d'éviter des prédictions de modèle déstabilisées en présence de petits échantillons. Deuxièmement, les données au niveau de l'unité et de la région peuvent avoir des caractéristiques de distribution différentes pour ce qui est des tendances de dispersion et des structures de corrélation. Troisièmement, les données au niveau de l'unité et de la région sont habituellement sujettes à différents types d'erreurs de mesure. Nous proposons un modèle multiniveau avec une pénalisation selon le niveau pour éviter ces problèmes et utiliser conjointement des données au niveau de l'unité et de la région pour l'estimation sur petits domaines. Dans un de nos exemples, nous combinons les données d'une enquête sur la santé au niveau de l'unité et des dossiers agrégés du microrecensement au niveau de la région pour estimer la prévalence régionale de l'hypertension en Allemagne.

Mots-clés : Descente de gradient par coordonnée; modèle multiniveau; maximum de vraisemblance pénalisé.

1. Introduction

Souvent, on applique l'estimation sur petits domaines pour obtenir des estimations fiables de quantités d'agrégats en particulier (statistiques par domaine) provenant de petits échantillons (Rao et Molina, 2015; Münnich et coll., 2016). Dans ce cas, un estimateur direct qui utilise seulement les données d'un domaine à la fois ne peut pas produire d'estimations statistiques de domaine suffisamment précises. L'estimation sur petits domaines a été conçue pour résoudre ce problème par la combinaison de données de plusieurs domaines dans des modèles de régression appropriés. L'idée générale consiste à accroître l'efficacité des estimations par rapport à un estimateur direct en exploitant la relation fonctionnelle entre la statistique de domaine d'intérêt et des données auxiliaires appropriées. Selon le domaine d'application, des modèles sont proposés au niveau du domaine (Fay et Herriot, 1979) ou au niveau de l'unité (Battese et coll., 1988). Le gain d'efficacité d'un estimateur sur petits domaines par rapport à un estimateur direct est déterminé par le pouvoir explicatif du modèle de régression sous-jacent. Par conséquent, quand on dispose de données auxiliaires au niveau de l'unité et du domaine, on doit prendre en compte les deux niveaux afin de maximiser le pouvoir explicatif et ainsi produire des estimations statistiques de domaine optimales.

Toutefois, l'utilisation conjointe de données au niveau de l'unité et de la région pour l'estimation par petits domaines fondée sur un modèle comporte certains problèmes méthodologiques. Premièrement, elle nécessite d'estimer simultanément les paramètres du modèle aux deux niveaux. En présence de petits échantillons, l'augmentation du nombre de paramètres peut entraîner une variance très élevée dans les estimations des paramètres du modèle en raison du manque de degrés de liberté. Les estimations statistiques de domaine fondées sur un modèle sont ensuite affectées d'une variance élevée et ne sont pas fiables. Deuxièmement, les données au niveau de l'unité et du domaine ont tendance à présenter des caractéristiques de distribution et des structures de corrélation différentes en raison de leurs degrés d'agrégation différents (Clark et Avery, 1976). C'est pourquoi l'on ne doit pas traiter les niveaux de la même manière pour ce qui est de la sélection des variables et de l'estimation des paramètres du modèle. Troisièmement, les données au niveau de l'unité et du domaine sont habituellement sujettes à différents types d'erreurs de mesure. Bien que les données au niveau de l'unité puissent pâtir de réponses imprécises, les données au niveau du domaine peuvent être incertaines parce que ses valeurs sont estimées. Le chercheur doit tenir compte du fait qu'ignorer les erreurs de mesure mène à des estimations sous-optimales des statistiques de domaine (Lohr et Ybarra, 2008). Enfin, en général, la disponibilité des données au niveau de l'unité

¹Jan Pablo Burgard, Université de Trier, Universitätsring 15, Trier (Trèves), Allemagne, D-54296; Joscha Krause, Université de Trier, Universitätsring 15, Trier (Trèves), Allemagne, D-54296; Ralf Münnich, Université de Trier, Universitätsring 15, Trier (Trèves), Allemagne, D-54296 (muennich@uni-trier.de)

et du domaine varie. Souvent, les données au niveau de l'unité ne sont pas disponibles pour des raisons de confidentialité, alors que les données au niveau du domaine sont moins sensibles et plus faciles d'accès, par exemple à partir de registres. Par conséquent, un modèle combiné doit être en mesure de traiter les cas où l'on dispose d'un grand nombre de variables au niveau du domaine, mais seulement de quelques variables au niveau de l'unité.

Nous proposons de combiner les données au niveau de l'unité et de la région pour l'estimation sur petits domaines dans un modèle multiniveau avec une pénalisation selon le niveau. La pénalisation selon le niveau désigne l'estimation par la méthode du maximum de vraisemblance des paramètres du modèle avec pénalisation, où les coefficients des effets fixes sur chaque niveau sont réduits par une pénalité individuelle. À cette fin, on tient compte du Lasso (*Least Absolute Shrinkage and Selection Operator*, opérateur de sélection et réduction par moindres valeurs absolues), de la pénalité ridge et de la méthode en filet élastique. L'utilisation d'une pénalisation selon le niveau dans des modèles multiniveaux résout les problèmes méthodologiques mentionnés plus haut. Premièrement, elle permet une inférence de grande dimensionnalité. Par conséquent, même si le nombre de paramètres du modèle est supérieur au nombre d'observations, le problème d'optimisation sous-jacent pour l'estimation des paramètres du modèle est toujours bien posé. Cela est particulièrement intéressant en présence de petits échantillons. Deuxièmement, la pénalisation selon le niveau est une façon simple de traiter différemment les données au niveau de l'unité et celles au niveau du domaine aux fins de l'estimation des paramètres du modèle. On peut définir les pénalités en fonction des caractéristiques de distribution des données auxiliaires correspondantes. De plus, si l'on choisit une pénalité conduisant à la parcimonie (p. ex. Lasso), une sélection automatique de variables selon le niveau est effectuée. Troisièmement, la régularisation basée sur la norme implique une robustification contre les erreurs de mesure dans les données auxiliaires (Bertsimas et Copenhaver, 2018; Burgard et coll., 2019a). Par conséquent, la pénalisation selon le niveau permet une estimation robuste malgré les erreurs de mesure différentes à chaque niveau. Enfin, on peut modifier le degré de pénalisation à chaque niveau en fonction du nombre de variables disponibles aux fins de prédiction.

On effectue l'estimation par la méthode du maximum de vraisemblance avec pénalisation des paramètres du modèle à l'aide d'un algorithme de descente de gradient par coordonnée stochastique en s'appuyant sur les données de Tseng et Yun (2009) et de Schelldorfer et coll. (2011). La prédiction de l'effet aléatoire est effectuée à partir d'une approche bayésienne utilisant le maximum a posteriori, comme l'ont proposé Schelldorfer et coll. (2011). Nous en présentons une application dans l'exemple de la mesure de la santé en Allemagne. Nous combinons les données au niveau de l'unité de l'enquête allemande sur la santé *Gesundheit in Deutschland aktuell (GEDA)* avec des données au niveau de la région provenant de dossiers agrégés de microrecensement pour estimer la prévalence régionale de l'hypertension. Le plan de l'article est le suivant. Au chapitre 2, nous décrivons le modèle multiniveau et l'estimation des paramètres du modèle. Au chapitre 3, nous présenterons l'application à la mesure de la santé. Au chapitre 4, nous terminerons l'article sur des remarques concluanes. La présente contribution est une version abrégée du document de travail sur les *Modèles sur petits domaines pénalisés pour la combinaison de données au niveau de l'unité et de la région* rédigé par les mêmes auteurs. Elle donne un aperçu de la méthodologie. Pour en savoir plus sur notre approche et consulter le détail des calculs, nous renvoyons à Burgard et coll. (2019b).

2. Méthodes

2.1 Modèle multiniveau

Soit U une population finie de taille N consistant en m domaines disjoints par paires de taille N_i avec $i = 1, \dots, m$ et $\sum_{i=1}^m N_i = N$. Supposons qu'un échantillon aléatoire S de taille n soit tiré de U de sorte qu'il y ait m sous-échantillons du domaine S_i de taille $n_i > 1$ avec $\sum_{i=1}^m n_i = n$. Soit $\mathbf{y}_i \in \mathbb{R}^{n_i \times 1}$ un vecteur contenant des observations d'une certaine variable réponse y à partir de laquelle est calculée la statistique de domaine d'intérêt dans le domaine i . Soit $\mathbf{X}_i^u \in \mathbb{R}^{n_i \times p^u}$ la matrice du plan d'expérience des effets fixes dans le domaine i contenant des covariables au niveau de l'unité. Soit $\mathbf{X}_i^a = (\mathbf{x}_i^a, \dots, \mathbf{x}_i^a)' \in \mathbb{R}^{n_i \times p^a}$ la matrice du plan d'expérience des effets fixes résultant de la dilatation du vecteur $\mathbf{x}_i^a \in \mathbb{R}^{1 \times p^a}$ contenant des covariables au niveau du domaine. Notons que $(p^u + p^a) > n$ est permis. Soit $\mathbf{Z}_i \in \mathbb{R}^{n_i \times q}$ la matrice du plan d'expérience de l'effet aléatoire dans le domaine i avec $q \leq (p^u + p^a)$. Dans la majorité des modèles sur petits domaines, la structure de l'effet aléatoire est généralement limitée à une ordonnée à l'origine aléatoire propre au domaine. Toutefois, la formulation générale du modèle multiniveau permet un effet aléatoire propre à un domaine sur toutes les covariables, potentiellement. Le modèle multiniveau qui combine les données au niveau de l'unité et du domaine est donné par

$$\mathbf{y}_i = \mathbf{X}_i^u \boldsymbol{\beta}^u + \mathbf{X}_i^a \boldsymbol{\beta}^a + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i \quad \forall i = 1, \dots, m,$$

où, $\beta^u \in \mathbb{R}^{p^u \times 1}$, $\beta^a \in \mathbb{R}^{p^a \times 1}$ sont les vecteurs de coefficient des effets fixes pour chaque niveau et $\mathbf{b}_i \sim MVN(\mathbf{0}, \Psi)$ indique le vecteur de coefficient d'effet aléatoire dans une normalité multivariée avec une matrice de covariance définie positive générale Ψ . $\mathbf{e}_i \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$ est un vecteur d'i.i.d. erreurs aléatoires avec le paramètre de variance du modèle σ^2 . Notons que $\mathbf{b}_1, \dots, \mathbf{b}_m, \mathbf{e}_1, \dots, \mathbf{e}_m$ sont supposés stochastiquement indépendants. Ainsi, le vecteur de réponse a la distribution suivante dans le modèle :

$$\mathbf{y}_i \sim MVN(\mathbf{X}_i^u \beta^u + \mathbf{X}_i^a \beta^a, \mathbf{V}_i(\sigma^2, \boldsymbol{\psi})),$$

avec $\mathbf{V}_i(\sigma^2, \boldsymbol{\psi}) = \sigma^2 \mathbf{I}_{n_i} + \mathbf{Z}_i \Psi \mathbf{Z}_i'$, où la matrice de covariance d'effet aléatoire est paramétrée par un vecteur $\boldsymbol{\psi}$ de dimension $q^* < n$, par exemple à la suite d'une décomposition de Cholesky. La réitération du modèle dans tous les domaines donne

$$\mathbf{y} = \mathbf{X}^u \beta^u + \mathbf{X}^a \beta^a + \mathbf{Z} \mathbf{b} + \mathbf{e},$$

avec $\mathbf{X}^u = (\mathbf{X}_1^{u'}, \dots, \mathbf{X}_m^{u'})'$, $\mathbf{X}^a = (\mathbf{X}_1^{a'}, \dots, \mathbf{X}_m^{a'})'$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$ comme matrices empilées et $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_m)'$, $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_m)'$, $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_m)'$ comme vecteurs empilés. On définit le vecteur de paramètre complet $\boldsymbol{\theta} := (\beta^u, \beta^a, \sigma^2, \boldsymbol{\psi}) \in \mathbb{R}^{p^u + p^a + 1 + q^*}$. La fonction de log-vraisemblance négative est alors

$$-l(\boldsymbol{\theta}) = \frac{1}{2} (n \cdot \log(2\pi) + \log(|\mathbf{V}|) + (\mathbf{y} - \mathbf{X}^u \beta^u - \mathbf{X}^a \beta^a)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}^u \beta^u - \mathbf{X}^a \beta^a)),$$

avec $\mathbf{V} = \text{diag}(\mathbf{V}_1, \dots, \mathbf{V}_m)$ et $|\mathbf{V}|$ désignant le déterminant de \mathbf{V} .

2.2 Estimation pénalisée des paramètres du modèle

Dans le modèle, l'estimation des paramètres du modèle pénalisée est caractérisée par le problème d'optimisation

$$\hat{\boldsymbol{\theta}} = \underset{\beta^u, \beta^a, \sigma^2 > 0, \boldsymbol{\psi} > 0}{\text{argmin}} \{-l(\boldsymbol{\theta}) + \lambda^u P^u(\beta^u) + \lambda^a P^a(\beta^a)\},$$

où $P^u(\beta^u): \mathbb{R}^{p^u} \rightarrow \mathbb{R}$, $P^a(\beta^a): \mathbb{R}^{p^a} \rightarrow \mathbb{R}$ sont des pénalités selon le niveau sur les coefficients des effets fixes et $\lambda^u, \lambda^a > 0$ sont des paramètres de réglage selon le niveau, déterminés à partir de k fois une procédure de validation croisée. Les pénalités suivantes sont considérées ($l \in \{u, a\}$) :

1. Lasso (Tibshirani, 1996) : $P^l(\beta^l) = \|\beta^l\|_1$
2. ridge (Hoerl et Kennard, 1970) : $P^l(\beta^l) = \|\beta^l\|_2^2$
3. filet élastique (Zou et Hastie, 2005) : $P^l(\beta^l) = \alpha^l \|\beta^l\|_1 + (1 - \alpha^l) \|\beta^l\|_2^2$, où $\alpha^l \in [0, 1]$ est un hyperparamètre selon le niveau.

Les pénalités ont des effets différents sur les solutions optimales pour les coefficients des effets fixes. Le Lasso produit une solution parcimonieuse pour $\hat{\beta}^u$ et $\hat{\beta}^a$. La parcimonie selon le niveau est contrôlée par λ^u, λ^a . L'inclusion de la pénalité entraîne une sélection automatique de variables dans le processus d'estimation, car les coefficients qui ne sont pas pertinents pour la description de la variable de réponse sont réglés à zéro. Au contraire, la pénalité ridge produit une solution dense pour $\hat{\beta}^u$ et $\hat{\beta}^a$. Elle lisse les contributions individuelles des coefficients dans le processus d'estimation. Le lissage selon le niveau est contrôlé par λ^u, λ^a . La pénalité ridge n'effectue pas de sélection de variables, mais on sait qu'elle produit des résultats plus stables en présence de multicollinéarité et de structures de regroupement dans les covariables. Le filet élastique est une combinaison linéaire du LASSO et de la pénalité ridge. Il induit une solution parcimonieuse tout en permettant des structures de regroupement dans les covariables. Le poids de chaque pénalité selon le niveau est contrôlé par α^u, α^a . L'intérêt de la pénalisation selon le niveau est qu'on choisit $P^u(\beta^u), P^a(\beta^a), \lambda^u, \lambda^a$ en fonction des caractéristiques de distribution des covariables de chaque niveau pour obtenir des prévisions optimales du modèle.

Afin de résoudre le problème de minimisation selon une pénalisation donnée, on utilise la modification stochastique de l'algorithme de descente de gradient par coordonnée proposée par Tseng et Yun (2009) ainsi que par Schelldorfer et coll. (2011). La minimisation par descente par coordonnée implique que la valeur de la fonction objective est minimisée progressivement en mettant à jour un seul élément $\theta_r \in \boldsymbol{\theta}$ à la fois tout en gardant $\boldsymbol{\theta}_{-r}$ fixe. Les éléments restants $\boldsymbol{\theta}_{-r}$ sont mis à jour en conséquence de façon itérative de sorte qu'il y ait un mouvement cyclique dans toutes les coordonnées de $\boldsymbol{\theta}$. Cette méthode cyclique est particulièrement utile pour le modèle multiniveau proposé, car elle permet de mettre en œuvre facilement la pénalisation selon le niveau dans le processus d'estimation. L'ordre des coordonnées qui correspondent aux coefficients des effets fixes est modifié aléatoirement dans chaque itération afin que soit améliorée la probabilité de convergence de l'algorithme à la lumière du problème de minimisation non convexe. L'ordre général des estimations commun dans les modèles sur

petits domaines (effets fixes conditionnels aux paramètres de variance et vice versa) n'est pas varié. Notons que si $\theta_r \in (\boldsymbol{\beta}^u, \boldsymbol{\beta}^a)$, alors la mise à jour requise pour la minimisation dépend en plus de la pénalité choisie pour son niveau respectif. Cela a des répercussions supplémentaires sur la méthode de calcul de la direction de descente et de détermination d'une longueur d'étape appropriée dans chaque itération. Nous ne traiterons toutefois pas de ces détails ici. Pour de plus amples renseignements sur l'algorithme utilisé aux fins d'estimation des paramètres du modèle, nous renvoyons le lecteur à Burgard et coll. (2019b).

En plus de l'estimation des paramètres du modèle, il faut prédire les effets aléatoires. Pour ce faire, nous utilisons une estimation a posteriori maximale, comme le proposent Schelldorfer et coll. (2011). Il s'agit d'une approche bayésienne où la quantité d'intérêt est estimée à partir du mode de la distribution postérieure. Soit f une densité de probabilité normale. Nous avons

$$\begin{aligned}\tilde{\mathbf{b}}_i &= \operatorname{argmax}_{\mathbf{b}_i} \left\{ \frac{f(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}^u, \boldsymbol{\beta}^a, \sigma^2, \boldsymbol{\psi}) \cdot f(\mathbf{b}_i | \boldsymbol{\psi})}{f(\mathbf{y}_i | \boldsymbol{\beta}^u, \boldsymbol{\beta}^a, \sigma^2, \boldsymbol{\psi})} \right\} \\ &= \operatorname{argmin}_{\mathbf{b}_i} \left\{ \frac{1}{\sigma^2} \|\mathbf{y}_i - \mathbf{X}_i^u \boldsymbol{\beta}^u - \mathbf{X}_i^a \boldsymbol{\beta}^a - \mathbf{Z}_i \mathbf{b}_i\|_2^2 + \mathbf{b}_i' \boldsymbol{\Psi}^{-1} \mathbf{b}_i \right\}.\end{aligned}$$

D'après les hypothèses du modèle, on peut conclure que

$$\mathbf{b}_i = (\mathbf{Z}_i' \mathbf{Z}_i + \sigma^2 \boldsymbol{\Psi}^{-1})^{-1} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i^u \boldsymbol{\beta}^u - \mathbf{X}_i^a \boldsymbol{\beta}^a),$$

qui est ensuite prédit par

$$\hat{\mathbf{b}}_i = (\mathbf{Z}_i' \mathbf{Z}_i + \hat{\sigma}^2 \boldsymbol{\Psi}^{-1})^{-1} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i^u \hat{\boldsymbol{\beta}}^u - \mathbf{X}_i^a \hat{\boldsymbol{\beta}}^a)$$

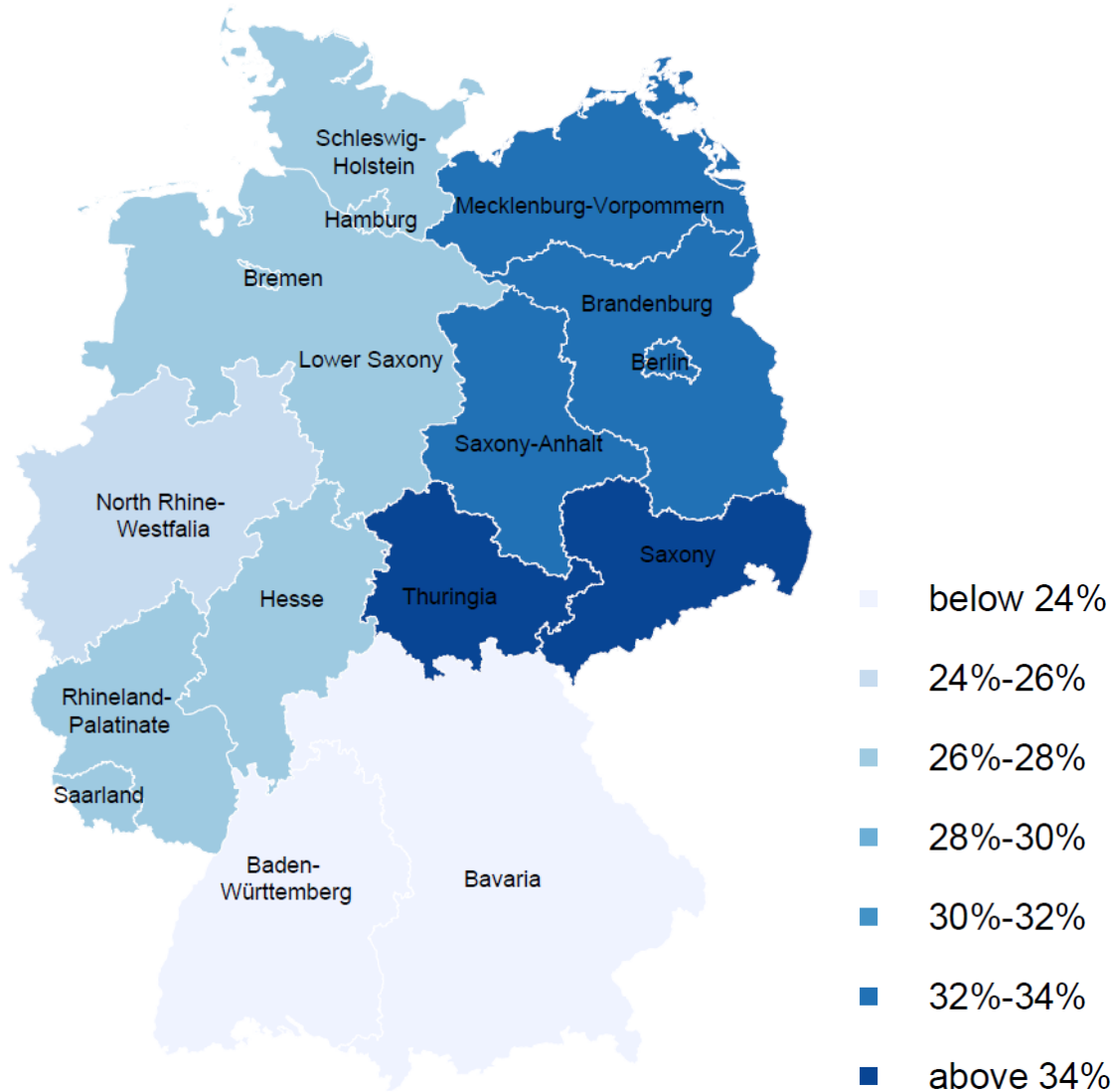
au moyen des estimations des paramètres du modèle obtenues à partir de l'algorithme de descente par coordonnée.

3. Estimation de la prévalence régionale de l'hypertension

La méthodologie est appliquée à la mesure de la santé en Allemagne. L'objectif est d'estimer la prévalence de l'hypertension chez la population de 18 ans ou plus à l'échelle régionale. La définition du profil de la maladie est une adaptation de celle du Robert Koch Institute (2012). À cette fin, nous combinons deux sources de données différentes. La première source de données est l'enquête allemande sur la santé *Gesundheit in Deutschland aktuell (GEDA)* de 2010. Elle contient des renseignements médicaux et relatifs à la santé détaillés sur environ 20 000 participants âgés d'au moins 18 ans. Les observations de cette enquête sont utilisées comme sources de données au niveau de l'unité. La deuxième source de données provient de dossiers agrégés du microrecensement allemand de 2010. Le microrecensement est une enquête à grande échelle couvrant un échantillon de 1 % de la population allemande. Il apporte (entre autres) des renseignements sociodémographiques et économiques que nous utilisons sous forme agrégée à des niveaux régionaux pour maximiser le pouvoir explicatif de l'estimation de la prévalence de l'hypertension.

La pénalité en filet élastique avec les hyperparamètres $\alpha^u = \alpha^a = 0.5$ est utilisée aux fins de l'estimation pénalisée par la méthode du maximum de vraisemblance des paramètres du modèle. Les paramètres de réglage λ^u, λ^a sont déterminés par k fois une validation croisée. Comme la pénalité en filet élastique est une pénalité entraînant la parcimonie, une sélection automatique des variables est effectuée dans le processus d'estimation. Au niveau de l'unité, les variables relatives à la démographie, à la comorbidité et au mode de vie sont sélectionnées. Au niveau de la région, les variables sélectionnées concernent principalement les conditions socioéconomiques et le marché du travail. Cette méthodologie permet d'obtenir les résultats suivants.

Figure 3-1
Prévalence estimée de l'hypertension pour 2010, États fédéraux



La figure 3-1 est une carte de l'Allemagne indiquant les estimations de la prévalence de l'hypertension par État fédéral. La prévalence nationale de l'hypertension est de 26,8 %. Ce chiffre concorde avec les résultats du Robert Koch Institute (2012), qui a calculé un intervalle de confiance de 95 % fondé sur des enquêtes de [25,9 % ; 27,6 %]. En examinant les estimations de l'État fédéral, on constate que la prévalence la plus faible se trouve dans le sud du pays, qui comprend les États fédéraux de Bade-Württemberg et de Bavière. La prévalence la plus élevée se trouve à l'est du pays, soit l'ancien territoire de la République démocratique allemande. La répartition régionale estimée est plausible, car des études antérieures ont constaté des distributions semblables de maladies étroitement liées, comme le diabète sucré de type 2 (voir, p. ex., Schipf et coll., 2014).

4. Conclusion et perspectives

On a proposé un modèle multiniveau afin d'utiliser conjointement des données au niveau de l'unité et de la région. Le modèle permet de combiner des données à plusieurs niveaux provenant de différentes sources de données afin d'optimiser l'estimation sur petits domaines fondée sur un modèle en maximisant le pouvoir explicatif du modèle de régression sous-jacent. On a résolu les problèmes méthodologiques associés à la combinaison de niveaux par une pénalisation selon le niveau au moyen de Lasso, de la pénalité ridge et de la méthode en filet élastique. Le réglage des paramètres de régularisation se fait par k-fois une validation croisée. L'estimation des paramètres du modèle est effectuée par un algorithme de descente de gradient stochastique. Aux fins de la prédiction de l'effet aléatoire, on adopte une approche utilisant le maximum a posteriori.

Les recherches futures pourraient porter sur l'estimation de l'erreur quadratique moyenne des estimations des statistiques de domaine avec pénalisation selon le niveau. D'une part, les estimations des paramètres du modèle avec pénalisation n'ont pas de solution sous forme fermée. D'autre part, la méthode d'estimation par la méthode du maximum de vraisemblance pénalisée introduit un certain biais dans les estimations des paramètres du modèle qu'il est difficile de quantifier. Burgard et coll. (2019a) proposent une méthode du jackknife modifiée pour estimer l'EQM d'un modèle de Fay-Herriot pénalisé. Bien que la procédure générale s'applique à notre approche, il faudrait peut-être d'autres modifications pour inclure la pénalisation selon le niveau dans le processus d'estimation.

Bibliographie

- Battese, G. E., R. M. Harter, et W. A. Fuller (1988), « An error-components model for prediction of county crop areas using survey and satellite data », *Journal of the American Statistical Association*, 83(401), p. 28-36.
- Bertsimas, D., et M. S. Copenhaver (2018), « Characterization of the equivalence of robustification and regularization in linear and matrix regression », *European Journal of Operational Research*, 270, p. 931-942.
- Burgard, J. P., J. Krause, et D. Kreber (2019a), « Regularized area-level modelling for robust small area estimation in the presence of unknown covariate measurement errors », Trier University Working Paper Series.
- Burgard, J. P., J. Krause, et R. Münnich (2019b), « Regularized Small Area Models for the Combination of Unit- and Area-Level Data », Trier University Working Paper Series.
- Clark, W. A. V., et K. L. Avery (1976), « The effects of data aggregation in statistical analysis », *Geographical Analysis*, 8(4), p. 428-438.
- Fay, R. E., et R. A. Herriot (1979), « Estimates of income for small places: An application of James-Stein procedures to census data », *Journal of the American Statistical Association*, 74(366), p. 269-277.
- Hoerl, A. E., et R. W. Kennard (1970), « Ridge regression: Biased estimation for nonorthogonal problems », *Techometrics*, 12(1), p. 55-67.
- Lohr, S., et L. Ybarra (2008), « Small area estimation when auxiliary information is measured with error », *Biometrika*, 95, p. 919-931.
- Münnich, R. T., J. P. Burgard, S. Gabler, M. Ganninger, et J. P. Kolb (2016), « Small area estimation in the German census 2011 », *Statistics in Transition and Survey Methodology*, 17(1), p. 25-40.
- Rao, J. N. K., et I. Molina (2015), *Small area estimation (2 ed.)*, Hoboken: Wiley.
- Robert Koch Institute (2012), *Daten und Fakten: Ergebnisse der Studie Gesundheit in Deutschland aktuell 2010*, Berlin: RKI.
- Schelldorfer, J., P. Bühlmann, et S. van de Geer (2011), « Estimation for high-dimensional linear mixed-effects models using l1-penalization », *Scandinavian Journal of Statistics*, 38, p. 197-214.
- Schipf, S., T. Ittermann, T. Tamayo, R. Holle, M. Schunk, W. Maier, C. Meisinger, B. Thorand, A. Kluttig, K. H. Greiser, K. Berger, G. Müller, S. Moebus, U. Slomiany, W. Rathmann, et H. Völzke (2014), « Regional differences in the incidence of self-reported type 2 diabetes in Germany: Results from five population-based studies in Germany (DIAB-CORE Consortium) », *Journal of Epidemiology and Community Health*, 68, p. 1088-1095.
- Tibshirani, R. (1996), « Regression shrinkage and selection via the lasso », *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), p. 267-288.
- Tseng, P., et S. Yun (2009), « A coordinate gradient descent method for nonsmooth separable minimization », *Mathematical Programming*, 117(1-2), p. 387-402.
- Zou, H., et T. Hastie (2005), « Regularization and variable selection via the elastic net », *Journal of the Royal Statistical Society: Series B (Methodological)*, 67(2), p. 301-320.