Estimation de la densité de coefficients aléatoires quand les regresseurs sont bornés

Christophe Gaillac [1] & Eric Gautier [2]

[1] *Centre de Recherche en Economie et Statistique, 15 Boulevard Gabriel Péri, 92245 Malakoff, France*
christophe.gaillac@ensae.fr
[2] *Toulouse School of Economics, 21 allée de Brienne, 31 000 Toulouse, France*
eric.gautier@tse-fr.eu

Résumé. Nous considérons l'estimation de la densité de coefficients aléatoires dans un modèle linéaire à coefficients aléatoires où l'intercept et la pente aléatoires sont indépendants de p régresseurs continus. Ces derniers sont supposés continus mais leur support est restreint à être un strict sous-ensemble de $\mathbb{R}^p$. Nous donnons de nouvelles conditions sous lesquelles le modèle est identifié, et qui n'imposent pas de conditions sur la distribution de l'intercept. Ces conditions sont formulées en termes d'observables. Puis nous présentons des bornes inférieures pour le risque minimax basé sur l'erreur quadratique moyenne ainsi qu'un estimateur adaptatif. Selon la classe de régularité considérée, les vitesses de convergence vont du logarithmique en la taille de l'echantillon, au paramétrique à un terme en logarithme près. Après avoir illustré les bonnes performances de l'estimateur sur des simulations, nous estimons l'hétérogénéité des élasticités prix et revenu des ménages anglais dans un modèle de demande linéaire à coefficients aléatoires.

Mots-clés. Adaptation, Problèmes inverses mal posés, Minimax, Coefficients aléatoires.

Abstract. We consider the estimation of the density of random coefficients in a linear random coefficients model where the random intercept and slopes are independent from p continuously distributed regressors. We address the case where the support of the regressors is a strict subset of $\mathbb{R}^p$. We provide new identification conditions which impose no restriction on the random intercept. These conditions involve the observables, namely the conditional distribution of the outcome given the regressors. We present lower bounds on the minimax risk based on the mean integrated squared error and an adaptive estimator. Rates of convergence range from logarithmic in the sample size to parametric up to a log factor depending on the smoothness class. Finally, we estimate the heterogeneity in price and income elasticities of British households in a linear demand model with random coefficients.

Keywords. Adaptation, Ill-posed Inverse Problem, Minimax, Random Coefficients.

1 Résumé long

1.1 Introduction

Les modèles à coefficients aléatoires permettent de décrire l'hétérogénéité inobservée au sein d'une population, et ainsi de prendre en compte celle-ci dans la modélisation de l'impact d'une mesure de politique économique. Les applications possibles sont nombreuses : hétérogénéité de l'effet de la taille de la classe sur la réussite scolaire, estimation de fonctions de production, de demande... Nous considérons un modèle à coefficients aléatoire linéaire

$$Y = \alpha + \beta^T X \tag{1}$$

où X et β sont des vecteurs aléatoires de dimension $p \times 1$, α est une variable aléatoire, et (α, β^T) et X sont indépendants. Le chercheur dispose d'un échantillon de n observations i.i.d. (y_i, x_i^T) de (Y, X^T) mais n'observe pas les réalisations i.i.d. (α_i, β_i^T) de (α, β^T) pour $i = 1, \ldots, n$. La linéarité est une limitation qui peut être relachée comme dans [5].

Suivant les applications, réduire l'hétérogénéité inobservée à une variable scalaire peut conduire à limiter les éffets de substitutions (modèle logit de demande), ou à imposer des restrictions sur le comportement des agents (monotonicité dans le modèle de Roy). Ce modèle, étant totalement nonparamétrique, limite aussi le risque de mis-spécification.

Notre but est d'estimer la densité jointe du vecteur (α, β). Estimer cette loi jointe est indispensable pour de nombreuses applications nécessitant plus que la connaissance de moyenne des coefficients (α_i, β_i). Par exemple, [1] considère la construction d'un intervale de prédiction pour Y, sachant les valeurs du regresseur $X = x$ en dehors du domaine observé. Comme nous observons $(\alpha_i, \beta_i)_{i=1}^n$ seulement au travers de d'observations i.i.d de $(Y_i, X_i)_{i=1}^n$, ceci constitue un problème inverse. Ce problème a été considéré par [2] et [8] en utilisant la transformée de Radon, qui a été étudiée dans le contexte de la tomographie (voir [3],[10],[9]). Dans le modèle (1), contrairement à la tomographie, la distribution des regresseurs est aléatoire et inconnue. L'identification du modèle (1) nécéssite soit que le support de X soit l'espace entier $\mathbb{R}^p$, soit de faire des hypothèses sur les marginales $\{\beta_j\}_{j\in\{1,\ldots,p\}}$. Dans cette présentation nous plaçons dans ce second cas, en supposant que les régresseurs X sont continus et bornés.

1.2 Identification de $\mathbb{P}_{\alpha,\beta}$

Si l'on suppose (α, β) independent de X, et le support de X est d'intérieur non vide, [1] montre que $\mathbb{P}_{\alpha,\beta}$ est identifié si le support de (α, β) est compact. De manière plus générale, nous montrons que $\mathbb{P}_{\alpha,\beta}$ est identifié sans faire d'hypothèses sur $\mathbb{P}_\alpha$, et en relachant l'hypothèse de support compact en β. Cela permet de considérer les distributions des termes d'erreurs habituels qui sont non bornées.

Dénotons $\mathbb{S}_X$ le support de X. Notre résultat se base sur le fait que la distribution des observables $\mathbb{P}_{Y,X}$ est charactérisée par $\mathbb{P}_X$ et $\mathbb{P}_{Y|X=x}$ pour x dans $\mathbb{S}_X$, soit par $\mathbb{P}_X$ et la fonction $h : \ (t,x) \in \mathbb{R} \times \mathbb{S}_X \to \mathbb{E}\left[e^{itY} \middle| X = x\right]$. En utilisant l'indépendance entre (α, β) et X on lie la fonction h à la distribution des inobservables $\mathbb{P}_{\alpha,\beta}$ par la relation $h(t,x) = \mathcal{F}\left[\mathbb{P}_{\alpha,\beta}\right](t,tx)$ pour tout $(t,x) \in \mathbb{R} \times \mathbb{S}_X$. Nous donnons alors une condition suffisante formulée à partir des moments des marginales de β sous $\mathbb{P}_{\alpha,\beta}$

$$\forall j = 1, \ldots, p, \ \sum_{k \in \mathbb{N}_0} \left(M_k^{\beta_j}\right)^{-1/k} = \infty \ , \tag{2}$$

où $M_k^{\beta_j} = \mathbb{E}\left[|\beta_j|^k\right]$ sont les moments absolus de β_j, qui permet de prolonger la transformée de Fourier de la distribution $\mathbb{P}_{\alpha,\beta}$ du cône $\{(t,tx) \ : \ t \in \mathbb{R}, \ x \in \mathbb{S}_X\}$, où elle est observée, à l'espace entier $\mathbb{R}^{p+1}$, permettant d'identifier $\mathbb{P}_{\alpha,\beta}$. Lorsque cette condition n'est pas satisfaite, nous exhibons un contre-exemple adapté de [11] en constuisant deux probabilités distinctes conduisant aux mêmes observables. Enfin, nous donnons une formulation de cette condition basée sur des observables.

1.3 Estimation

Dans un second temps, nous supposons que (α, β) admet une densité et, sans perte de généralité, que les régresseurs varient dans $[-\underline{x}, \underline{x}]^p$ avec $\underline{x} > 0$. Sous ces conditions, nous proposons un estimateur non-paramétrique de la densité jointe $f_{\alpha,\beta}$. Nous nous plaçons dans un espace de Hilbert, et décrivons l'opérateur associé au problème inverse considéré. Nous montrons que cet opérateur est linéaire, compact et injectif, et qu'il admet donc une décomposition en valeurs singulières (SVD) qui forme une base orthonormale de l'espace de départ. Celle-ci étant facilement calculable, nous proposons un estimateur à partir de cette SVD, avec deux paramètres de régularisation. En introduisant des espaces de régularité adaptés à la géométrie du problème, nous donnons une bonne inférieure pour le risque en moyenne quadratique. Nous montrons que notre estimateur adaptatif, basés sur deux critéres pour le choix des paramètres combinant les aspects de [6] et de [7] conduit à les vitesses de convergence qui vont du logaritmique au paramétrique à un terme en logarithme près suivant la classe de régularité condiérée. Dans le cas de régularité le moins favorable, la vitesse obtenue est minimax. Nous illustrons sur des simulations que même avec de faibles valeurs de $\underline{x}$, et une taille de l'échantillon allant de 1000 à 5000, notre estimateur parvient à estimer distinctement les deux composantes lorsque la loi de (α, β) est prise sous forme d'un mélange de deux gaussiennes.

1.4 Application

Nous appliquons notre estimateur à la quantification de l'hétérogénité de la demande dans un modèle LA/AIDS ("Linear Approximate Almost Ideal Demand System", voir [4]).

L'estimation de l'hétérogénéité dans ce modèle peut amener à amender les conclusions en terme de bien-être pour le consommateur lors de l'évaluation de l'impact d'une taxe à partir de modèles à coefficients non aléatoires. Nous considerons le système de G équations pour les dépenses des ménages en G biens de consommations : nourriture consommée à la maison, essence, habillement, alcool, livres et magazines, et autres biens non durables, qui pour $g \in \{1, \dots, G\}$ prend la forme

$$W_{g,i} = \alpha_{g,i} + \beta_{g,i} \ln\left(\frac{M_i}{P_{t(i)}}\right) + \gamma_{g,i} \ln\left(P_{g,t(i)}\right) + \sum_{\substack{\tilde{g}=1 \\ \tilde{g}\neq g}}^{G} \gamma_{\tilde{g}} \ln\left(P_{\tilde{g},t(i)}\right) + \delta^T C_i + \nu^T T_{t(i)} \ . \quad (3)$$

Dans l'équation (3), M_i est la dépense totale, $(P_{\tilde{g},t(i)})_{\tilde{g}=1}^{G}$ sont les indices de prix mensuels des biens à la date $t(i)$ quand le ménage i est échantillonné, $P_{t(i)} = \sum_{\tilde{g}=1}^{G} \overline{w}_{\tilde{g},t(i)} \ \ln(P_{\tilde{g},t(i)})$ où $\overline{w}_{\tilde{g},t(i)}$ est la part de dépenses moyenne du bien $\tilde{g}$ dans les dépenses totales pour les ménages échantillonnés à la date $t(i)$ (*i.e* l'indice de Stone (1954)), C_i est un vecteur de variables démographiques, et $T_{t(i)}$ contient une tendance linéaire et trois indicatrices saisonnières. Dans ce modèle nous décrivons comment estimer la densité jointe de $(\beta_{g,i}, \gamma_{g,i})$ au moyen de notre estimateur.

References

[1] Rudolf Beran, *Prediction in random coefficient regression*, Journal of Statistical Planning and Inference **43** (1995), 205–213.

[2] Rudolf Beran, Andrey Feuerverger, and Peter Hall, *On nonparametric estimation of intercept and slope distributions in random coefficient regression*, Annals of Statistics **24** (1996), 2569–2592.

[3] Laurent Cavalier, *Efficient estimation of a density in a problem of tomography*, Annals of Statistics (2000), 630–647.

[4] Angus Deaton and John Muellbauer, *An almost ideal demand system*, The American economic review **70** (1980), 312–326.

[5] Eric Gautier and Stefan Hoderlein, *A triangular treatment effect model with random coefficients in the selection equation*, preprint arXiv:1109.0362 (2011).

[6] Eric Gautier and Erwan Le Pennec, *Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding*, preprint arXiv:1106.3503 (2011).

[7] Yu Golubev, *The principle of penalized empirical risk in severely ill-posed problems*, Probability Theory and Related Fields **130** (2004), 18–38.

[8] Stefan Hoderlein, Jussi Klemelä, and Enno Mammen, *Analyzing the random coefficient model nonparametrically*, Econometric Theory **26** (2010), 804–837.

[9] Iain M Johnstone and Bernard W Silverman, *Speed of estimation in positron emission tomography and related inverse problems*, Annals of Statistics (1990), 251–280.

[10] Aleksandr P. Korostelev and Alexandre B Tsybakov, *Minimax theory of image reconstruction*, Vol. 82, Springer Science and Business Media, 2012.

[11] Walter Rudin, *Real and complex analysis*, Tata McGraw-Hill Education, 1987.