Impact des variables explicatives dans les modèles compositionnels: effet des investissements publicitaires sur le marché automobile français

Christine Thomas-Agnan [1] & Joanna Morais [2] & Michel Simioni [3]

[1] *Toulouse School of Economics, University of Toulouse 1 Capitole, 21 allee de Brienne, Toulouse, France, christine.thomas@tse-fr.eu*
[2] *BVA, 52 rue Marcel Dassault, Boulogne-Billancourt, France, joanna.morais@bva-group.com*
[3] *INRA, UMR 1110 MOISA, 2 Place Pierre Viala, Montpellier, France, michel.simioni@inra.fr*

Résumé. On considère des modèles de régression comportant des variables compositionnelles. Un vecteur de parts est appelé une composition. Les parts sont les composants et la composition appartient au simplexe. Nous nous intéressons au cas où la variable dépendante ainsi que certaines variables explicatives sont compositionnelles (des variables explicatives classiques peuvent aussi être incluses). Plusieurs modèles sont disponibles pour ce type de données dans la littérature et nous nous concentrons sur la famille des modèles de régression compositionnels (CODA). Dans le but d'améliorer l'interprétabilité de ces modèles, nous montrons que les élasticités sont un outil adapté à la mesure de l'impact d'une variable explicative sur un composant. Cependant leur definition et leur calcul nécéssite l'utilisation de la notion de dérivée d'une fonction à valeurs dans le simplexe. De plus lorsqu'il s'agit de l'impact d'une variable compositionnelle, un soin particulier doit être apporté au calcul. Nous présentons une application au marché automobile où le but est d'étudier l'impact des investissements publicitaires sur les parts de marché des différents constructeurs sur le segment B de ce marché.

Mots-clés. Modèles compositionnels, Effets marginaux, Elasticités, Rapports des côtes, parts de marché

Abstract. We consider regression models involving compositional data. A vector of shares is called a composition. The shares are the components, and the composition lies in a space called the simplex. We are interested in the case where the dependent variable as well as some explanatory variables are compositional (classical explanatories can also be included). Several models are available in the literature for such data and we focus on the family of compositional (CODA) models. In order to enhance the interpretability of these models, we show that elasticities are useful to isolate the impact of an explanatory variable on a particular share. However their definition and computation requires using the notion of derivative of a simplex valued function. Moreover we show that special care has to be taken when the impact of a compositional variable is concerned. An application to an automobile market data set is presented, where the aim is to study the impact of brand media investments on brand market shares in the B segment of this market.

Keywords. Compositional models, Marginal effects, Elasticities, Odds ratio, market shares

1 Résumé long

Les méthodes de régression classiques ne s'appliquent pas lorsque la variable dépendante d'un modèle de régression est un vecteur de parts, par exemple des parts de marché de différentes marques, des proportions de votes pour les différents partis lors d'une élection. Divers modèles adaptés à ces données ont été développés dans la littérature (voir Morais et al. (2016) pour une revue sur ce sujet). On distingue notamment les modèles CODA (pour compositional data analysis, voir Van den Boogaart et Tolosana-Delgado (2013)), mais aussi les modèles de Dirichlet et les modèles de la littérature marketing (logit multinomial et MCI pour "multiplicative competitive interaction"). Nous nous concentrons dans ce travail sur les modèles CODA dans lesquels la variable dépendante mais également une ou plusieurs variables explicatives sont de type compositionnel. Nous distinguons deux modèles: un modèle A où un unique coefficient est associé à chaque variable compositionnelle et un modèle plus complexe B qui a des coefficients spécifiques aux différents composants et qui présente des effets croisés.

Une composition de dimension $D \in \mathbb{N}$ est un vecteur de D parts qui appartient au simplexe de $\mathbb{R}^D$:

$$\mathcal{S}^D = \left\{ \mathbf{s} = (s_1, s_2, \ldots, s_D)' : s_j > 0,\ j = 1, \ldots, D;\ \sum_{j=1}^{D} s_j = 1 \right\},$$

c'est à dire que ses composants sont positives et de somme un. Les opérations suivantes sont utiles pour travailler dans cet espace:

- $\mathcal{C}()$ est l'opération de *clotûre* qui tranforme les volumes $x_j > 0$ en parts s_j: $\mathcal{C}(x_1, \ldots, x_D)' = \left(\frac{x_1}{\sum_{j=1}^{D} x_j}, \ldots, \frac{x_D}{\sum_{j=1}^{D} x_j} \right)'$
- $\oplus$ est l' operateur de *perturbation* , qui correspond à l'addition dans le simplexe: $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, \ldots, x_D y_D)'$ avec $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$
- $\odot$ est la *transformation puissance*, correspondant à la multiplication dans le simplexe: $\mathbf{x} \odot \lambda = \mathcal{C}(x_1^\lambda, \ldots, x_D^\lambda)'$ avec $\lambda \in \mathbb{R}, \mathbf{x} \in \mathcal{S}^D$
- $\boxdot$ est le *produit matriciel compositionnel*, correspondant au produit de matrices dans le simplex: $\mathbf{B} \boxdot \mathbf{x} = \mathcal{C}\big(\prod_{j=1}^{D} x_j^{b_{1j}}, \ldots, \prod_{j=1}^{D} x_j^{b_{Dj}} \big)'$ avec $\mathbf{B} \in \mathbb{R}_{D \times D}, \mathbf{x} \in \mathcal{S}^D$

Le principe des modèles de régression compositionnelles est d'utiliser une transformation dite en log-ratios pour transporter le simplexe $\mathcal{S}^D$ dans l'espace euclidien $\mathbb{R}^{D-1}$. On peut alors écrire un modèle de régression sur les coordonnées (images des composants) et utiliser la transformation inverse pour revenir aux composants lorsque le modèle est estimé. Plusieurs transformations classiques sont utilisées (appelées alr, clr et ilr) mais la transformation ilr est en général préférable. Dans le simplexe, si l'on désigne par $\mathbf{S}_t$ le vecteur des parts (de dimension D_S) pour l'observation t, et par $\mathbf{X}_t$ de dimension D_X (respectivement Z_t) les variables explicatives compositionnelles (respectivement non compositionnelles), le modèle A s'écrit

$$\mathbb{E}\mathbf{S}_t = \mathbf{a} \oplus \mathbf{X}_t \odot b \oplus \mathbf{c} \odot Z_t \tag{1}$$

et le modèle B s'écrit

$$\mathbb{E}\mathbf{S}_t = \mathbf{a} \oplus \mathbf{B} \boxdot \mathbf{X}_t \oplus \mathbf{c} \odot Z_t \tag{2}$$

où $\mathbf{B}$ est une matrice de paramètres des dimension $D_S \times D_X$. Morais et al. (2016) montrent que ces modèles ont aussi une écriture sous forme d'attractivité comme les modèles logits multinomiaux et MCI de la littérature marketing. Pour le modèle A, cela s'écrit

$$\mathbb{E}S_{jt} = \frac{a_j X_{jt}^b c_j^{Z_t}}{\sum_{m=1}^D a_m X_{mt}^b c_m^{Z_t}} \tag{3}$$

et pour le modèle B

$$\mathbb{E}S_{jt} = \frac{a_j \prod_{l=1}^D X_{lt}^{b_{jl}} c_j^{Z_t}}{\sum_{m=1}^D a_m \prod_{l=1}^D X_{lt}^{b_{ml}} c_m^{Z_t}}. \tag{4}$$

Cette écriture permet de calculer facilement les élasticités, telles que définies dans la littérature marketing

$$e(S_{jt}, X_{lt}) = \frac{\frac{\partial S_{jt}}{S_{jt}}}{\frac{\partial X_{lt}}{X_{lt}}} = \frac{\partial \log S_{jt}}{\partial \log X_{lt}}. \tag{5}$$

Nous montrons cependant que cette notion ne convient pas en raison de la nature compositionnelle du vecteur de parts : en effet lorsqu'on les utilise, une variation de la variable explicative peut faire sortir le vecteur de parts du simplexe. Nous définissons donc la notion d'élasticité dans le simplexe et calculons ces élasticités pour les modèles A et B. Le résultat montre que ces quantités sont bien adaptées à ces modèles. Nous calculons également les élasticités des ratios de parts S_j/S_j' par rapport à un composant X_l de la variable compositonnelle X, ainsi que les élasticités des ratios de parts S_j/S_j' par rapport à un ratio de parts X_l/X_l' de la variable compositionnelle. Finalement, nous envisageons aussi l'utilisation de rapports des côtes, comme dans les modèles logits et montrons leurs liens avec les élasticités.

Dans l'application qui a motivé ce travail, nous expliquons les parts de marché des marques leaders Renault, Peugeot et Citroen sur le segment B du marché automobile, les autres constructeurs étant regroupés dans la catégorie “Autres”. Parmi les variables explicatives figurent les parts d’investissements publicitaires de chaque marque (variable compositionnelle), les prix (en compositions) et l’indicatrice de prime à la casse. Les données portent sur 152 mois. Les résultats montrent que Renault a l’élasticité directe la plus forte par rapport aux dépenses publicitaires.

Bibliographie

[1] Van den Boogaart, K.G. et Tolosana-Delgado, R. (2013), *Analyzing Compositional Data with R*, Springer.
[2] Morais, J. Thomas-Agnan, C. et Simioni M. (2016), A tour of regression models for explaining shares, *TSE Working papers*, 16-742 (12 2016).
[3] Chen, J., Zhang X. et Li, S. (2016), Multiple linear regression with compositional response and covariates, *Journal of Applied Statistics*, 1–16.
[4] Egozcue, J. J., Jarauta-Bragulat, E., and Díaz-Barrero, J. L. (2011), Calculus of simplex-valued functions. in : Pawlowsky-Glahn, V., and Buccianti, A. Compositional data analysis: Theory and applications. John Wiley and Sons.