

MESURES D'IMPACT DES VARIABLES EXPLICATIVES DANS LES MODÈLES DE RÉGRESSION POUR VARIABLES COMPOSITIONNELLES

Joanna Morais ¹ & Christine Thomas-Agnan ²

¹ *Quantmetry, Paris, France et joanna.morais@live.fr*

² *Toulouse School of Economics et christine.thomas@tse-fr.eu*

Résumé. Des vecteurs de composition peuvent intervenir dans un modèle de régression soit en tant que variables explicatives (voir Hron et al. 2012) soit en tant que variables à expliquer (voir Egozcue et al., 2012) et un modèle peut présenter les deux cas (voir Chen et al. 2016 and Morais et al. 2018b). Cependant il n'est pas facile de mesurer l'impact marginal des variables dans ces modèles car un changement sur une des composantes d'une composition affecte toutes les autres composantes.

Morais et al. (2018a) ont montré que le bon outil pour mesurer ces impacts est la notion de dérivée simpliciale, c'est-à-dire dérivée d'une fonction dont soit l'ensemble de départ, soit l'ensemble d'arrivée, soit les deux est un simplexe (Pawlowsky-Glahn, V. and A. Buccianti, 2011). Dans le cas où les deux sont dans le simplexe, cela conduit à des interprétations en terme d'élasticités qui sont fréquemment utilisées en économétrie et dans les applications en marketing.

Nous montrons ici comment ces dérivées simpliciales peuvent être utilisées dans d'autres cadres: cas où seule la variable dépendante est une composition (basé sur le chapitre 12 de Egozcue et al. dans Pawlowsky-Glahn et al. 2011), celui où seule la variable indépendante est une composition (basé sur le chapitre 13 de Barcelo-Vidal et al. dans Pawlowsky-Glahn et al. 2011). Nous envisageons aussi les cas où un total associé à la composition intervient dans la régression en variable explicative. Les interprétations correspondantes sont en terme de semi-élasticités. Enfin, nous montrons comment construire des intervalles de confiance pour ces élasticités ou semi-élasticités qui permettent une meilleure interprétation de ces modèles. Nous illustrons sur un jeu de données réelles.

Mots-clés. Modèles de régression compositionnelle, effets marginaux, dérivée simpliciale, élasticité.

Abstract. Compositions can be used as variables in regression models, either as explanatory variables (see Hron et al. 2012) or as dependent variables (see Egozcue et al. 2012), or both (see Chen et al. 2016 and Morais et al. 2018a). However, it is not straightforward to measure the marginal impacts in these types of models, as the change in one component of a composition may affect the rest of the composition.

Morais et al. 2018a have shown that a tool for measuring these impacts is the notion of simplicial derivative, i.e. the derivative of a function whose domain and/or target set are in a simplex (Pawlowsky-Glahn, V. and A. Buccianti, 2011). In the case both the

dependent and the explanatory are compositional, the resulting interpretation is in terms of elasticity, commonly used in econometrics and marketing applications.

The present contribution aims to present how simplicial derivatives can be computed in other situations: the simplicial derivative of a composition relative to a non-compositional variable (based on Egozcue et al. in Pawlowsky-Glahn et al. 2011, chapter 12), the simplicial derivative of a non-compositional variable relative to a composition (based on Barcelo-Vidal et al. in Pawlowsky-Glahn et al. 2011, chapter 13), in the cases where the total is used or not as an explanatory variable. The resulting interpretations are based on semi-elasticities.

Moreover, we also demonstrate how to compute confidence intervals for these elasticities or semi-elasticities, which significantly improves the interpretability of the compositional regression models. This contribution will be illustrated by real-data applications.

Keywords. compositional regression model, marginal effects, simplicial derivative, elasticity .

Bibliographie

- Chen, J., X. Zhang, and S. Li (2016). Multiple linear regression with compositional response and covariates. *Journal of Applied Statistics*, pp. 1-16.
- Egozcue, J. J., J. Daunis-I-Estadella, V. Pawlowsky-Glahn, K. Hron, and P. Filzmoser (2012). Simplicial regression. The normal model. *Journal of Applied Probability and Statistics* (6), pp. 87-108.
- Hron, K., P. Filzmoser, and K. Thompson (2012). Linear regression with compositional explanatory variables. *Journal of Applied Statistics* 39(5), pp. 1115-1128.
- Morais, J., C. Thomas-Agnan, and M. Simioni (2018a). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, 47(5), pp. 125.
- Morais, J., C. Thomas-Agnan, and M. Simioni (2018b). Using compositional and Dirichlet models for market shares regression. *Journal of Applied Statistics* 45(9), pp. 1670-1689.
- Pawlowsky-Glahn, V. and A. Buccianti (2011). *Compositional data analysis: Theory and applications*. John Wiley & Sons