

Impact of Covariates in Compositional Models and Simplicial Derivatives

Joanna Morais
Avisia, Bordeaux, France

Christine Thomas-Agnan
Toulouse School of Economics

Abstract

In the framework of Compositional Data Analysis, vectors carrying relative information, also called compositional vectors, can appear in regression models either as dependent or as explanatory variables. In some situations, they can be on both sides of the regression equation. Measuring the marginal impacts of covariates in these types of models is not straightforward since a change in one component of a closed composition automatically affects the rest of the composition.

Previous work by the authors has shown how to measure, compute and interpret these marginal impacts in the case of linear regression models with compositions on both sides of the equation. The resulting natural interpretation is in terms of an elasticity, a quantity commonly used in econometrics and marketing applications. They also demonstrate the link between these elasticities and simplicial derivatives.

The aim of this contribution is to extend these results to other situations, namely when the compositional vector is on a single side of the regression equation. In these cases, the marginal impact is related to a semi-elasticity and also linked to some simplicial derivative. Moreover we consider the possibility that a total variable is used as an explanatory variable, with several possible interpretations of this total and we derive the elasticity formulas in that case.

Keywords: compositional regression model, marginal effects, simplicial derivative, elasticity, semi-elasticity.

1. Introduction and literature review

We consider regression models involving compositional vectors, i.e. vectors carrying relative information. When relative information is the focus, meaningful functions are functions of ratios of the vector's components therefore using traditional regression models in such cases is not correct. Regression models that respect the compositional nature of such data have been proposed in the literature, for example those introduced by [Aitchison \(1986\)](#) based on log-ratio transformations. Theory for inference in these models is developed for example in [Pawlowsky-Glahn and Buccianti \(2011\)](#), [Van Den Boogaart and Tolosana-Delgado \(2013\)](#), [Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado \(2015b\)](#), and [Filzmoser, Hron, and Templ \(2018\)](#).

When the compositional vectors only appear as dependent variable, we will say that the

model is of the ‘Y-compositional’ type (see e.g. Egozcue, Daunis-I-Estadella, Pawlowsky-Glahn, Hron, and Filzmoser (2012)). When they only appear as explanatory variables, we will say that the model is of the ‘X-compositional’ type (see e.g. Hron, Filzmoser, and Thompson (2012)). Finally, when they appear on both sides, we will say that the model is of the ‘YX-compositional’ type, see e.g. Kynclova, Filzmoser, and Hron (2015), Chen, Zhang, and Li (2017), Morais, Thomas-Agnan, and Simioni (2018a) and Morais, Thomas-Agnan, and Simioni (2018b). A simplified version of the YX-compositional type is presented in Wang, Shangguan, Wu, and Guan (2013) and Morais, Thomas-Agnan, and Simioni (2018b) later show that this model is equivalent to the so-called MCI (multiplicative competitive interaction) model introduced earlier in the marketing literature by Nakanishi and Cooper (1982). It may also be relevant to include in the model the total of the different parts involved in the composition and we will consider each of the above models for the case with or without a total variable, see e.g. Coenders, Martín-Fernández, and Ferrer-Rosell (2017) and Coenders, Ferrer-Rosell, Mateu-Figueras, and Pawlowsky-Glahn (2015). Extensions with compositional functional predictors are presented in Sun, Xu, Cong, and Chen (2018), Bui, Loubes, Risser, and Balaesque (2018) and Combettes and Muller (2019). Case studies using some of these models are presented in Hron, Filzmoser, and Thompson (2012), Trinh, Morais, Thomas-Agnan, and Simioni (2018) for the X-compositional type, Morais, Thomas-Agnan, and Simioni (2017) for the YX-compositional type’.

The focus of the present work is on the definition and interpretation of impacts of covariates in these models, question addressed by much fewer papers. Muller, Hron, Fiserova, Smahaj, Cakirpaloglu, and Vancakova (2018) propose an interpretation for models of X-compositional or Y-compositional types based on using a specific type of orthogonal coordinates (called pivot coordinates, see e.g. Filzmoser, Hron, and Templ (2018)). Moreover they promote the replacement of the natural logarithm by the base-2 logarithm for enhancing the interpretability. The first drawback is that the resulting interpretation requires rerunning the model once for each component in the Y-compositional case. Moreover changes in log-ratios correspond to multiplicative increase (of the dependent or independent variables) in terms of relative dominance, i.e. the ratio of one component to the geometric mean of the others (while keeping all other log-ratios constant) which is not a very intuitive notion. This point of view is extended in Coenders and Pawlowsky-Glahn (2020) by considering changes in more general log-ratios leading to changes in any subset of components by a common factor (while reducing the remaining components accordingly).

Morais, Thomas-Agnan, and Simioni (2018b) show that a natural interpretation tool in the YX-compositional model is the notion of elasticity. The notion of elasticity is frequently used in econometrics: it corresponds to the percent increase of the dependent variable induced by a percent increase of the explanatory variable and is natural in a log-log regression model because it coincides with the explanatory variable parameter (see Section 3). Indeed elasticities are commonly computed for the MCI model in the marketing literature (see Nakanishi and Cooper (1982)). Morais, Thomas-Agnan, and Simioni (2018b) relate it to the notion of simplicial derivatives introduced in Egozcue, Jarauta-Bragulat, and Díaz-Barrero (2011) and Barceló-Vidal, Martín-Fernández, and Mateu-Figueras (2011).

With a graphical approach, Nguyen, Laurent, Thomas-Agnan, and Ruiz-Gazen (2018) bring a different light on the evaluation of these impacts by plotting the predicted components as a function of the explanatory variables but this graphing tool is limited to compositional dependent or explanatory variables with three components.

Finally, for the X-compositional model, Coenders and Pawlowsky-Glahn (2020) consider the introduction of the total variable among the explanatory and adapt the resulting interpretations, still in terms of log-ratio changes.

The objective of this paper is to extend Morais, Thomas-Agnan, and Simioni (2018b) to the Y-compositional and the X-compositional models and to allow inclusion of the total variable in the models. In Section 2, we introduce notations and define the different specifications

of the considered models. In Section 3, we demonstrate the equations linking elasticities or semi-elasticities (depending on the considered model) with simplicial derivatives. Section 4 establishes the formulas for the elasticities and semi-elasticities in terms of model parameters in the simplex as well as in coordinate space. Finally, Section 5 provides examples of applications to the X-compositional and to the Y-compositional models and Section 6 concludes.

2. Compositional model specifications

Using the notation \mathbf{x}' for the transpose of a vector \mathbf{x} , let us denote by $\check{\mathbf{X}} = (\check{X}_1, \dots, \check{X}_{D_X})' \in \mathbb{R}_+^{D_X}$ a vector of D_X positive components corresponding to the components of a compositional vector expressed in original units: we call these components *volumes* as opposed to *shares*. For example, in the case studied in [Morais, Thomas-Agnan, and Simioni \(2017\)](#), the volumes are numbers of cars sold during a given month by the different brands of cars whereas the shares represent the corresponding proportion of cars sold during that month by each brand relative to the other brands in the study. The closure of the vector $\check{\mathbf{X}}$ of volumes is the corresponding vector of shares

$$\mathbf{X} = \mathcal{C}(\check{X}_1, \dots, \check{X}_{D_X})' = \left(\frac{\check{X}_1}{\sum_{i=1}^{D_X} \check{X}_i}, \dots, \frac{\check{X}_{D_X}}{\sum_{i=1}^{D_X} \check{X}_i} \right)' = (X_1, \dots, X_{D_X})'$$

and belongs to the simplex space \mathcal{S}^{D_X} of positive vectors in \mathbb{R}^{D_X} with sum equal to 1.

In some cases, it may be relevant to include in the regression model a variable measuring a total (hence not scale-invariant) which may be $\mathbf{T}(\mathbf{X})$ or $\mathbf{T}(\mathbf{Y})$. [Pawlowsky-Glahn, Egozcue, and Lovell \(2015a\)](#) argue that different formulas can be used for this total, for example one of the following two:

- Arithmetic total: $T_A(\check{\mathbf{X}}) = \sum_{i=1}^D \check{X}_i$
- Geometric total: $T_G(\check{\mathbf{X}}) = (\prod_{i=1}^D \check{X}_i)^{1/\sqrt{D}}$

The general principle of simplicial regression is to use transformations to transport the simplex space \mathcal{S}^D , equipped with the Aitchison geometry, into the Euclidian space \mathbb{R}^{D-1} thus eliminating the simplex constraints problem. It is generally agreed upon to use log-ratio orthonormal coordinates ([Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado \(2015b\)](#)). We recall that to each $D \times (D - 1)$ contrast matrix \mathbf{V} , constructed from an orthonormal basis of \mathcal{S}^D , corresponds an isometric transformation traditionally called *ilr* $_{\mathbf{V}}$. As advocated recently by [Martín-Fernández \(2019\)](#), we will rather use the name *olr* (orthogonal log ratio) for these transformations. We then have $\mathbf{z}^* = \text{olr}_{\mathbf{V}}(\mathbf{z}) = \mathbf{V}' \log(\mathbf{z})$, where the natural logarithm (denoted by \log) is understood componentwise and the inverse transformation is $\text{olr}_{\mathbf{V}}^{-1}(\mathbf{z}^*) = \mathcal{C}(\exp(\mathbf{V}\mathbf{z}^*))$. Note that *olr*-coordinates take the same value regardless if you use shares or volumes. However the inverse transformation always returns shares.

Using the traditional notations for the simplex operations (see [Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado \(2015b\)](#)) and denoting by \langle, \rangle_A the Aitchison scalar product, the first row of Table 1 presents the formulation of the regression models explaining a collection of n i.i.d. random variables (simplex valued or not) by corresponding explanatory variables which may be simplex valued or not. The observations are indexed by t , $t = 1, \dots, n$. Because marginal effects only involve one explanatory at a time, if we had a model explaining a simplex valued variable by both types of explanatory variables, we would use the first and last columns of this table. The second row of Table 1 presents the corresponding model formulations in coordinate space for a given choice of *olr* transformation *olr* $_{\mathbf{V}}$. Parameters \mathbf{a}^* , \mathbf{b}^* or \mathbf{B}^* are then estimated by maximum likelihood in coordinate space where the regression is classical. Formulas to compute the corresponding parameter estimates in the simplex \mathbf{a} , \mathbf{b} or \mathbf{B} are available and it is known that these estimated parameters in the simplex are independent of

the particular choice of $olr_{\mathbf{V}}$, i.e. of the particular choice of contrast matrix. In Table 1 the different formulations may involve a total variable $T(\mathbf{X})$ or $T(\mathbf{Y})$ and it is printed in grey to indicate that it is an option. Finally, we included in the formulations the particular case of the MCI model obtained when $D_X = D_Y$ and the matrix \mathbf{B}^* is a multiple of the identity resulting in $\mathbf{B} \boxtimes \mathbf{X} = b \odot \mathbf{X}$.

Table 1: Specifications of the compositional models and notations

Space	Y-compositional model	X-compositional model	YX-compositional model
			‘CODA’ model:
\mathcal{S}^D	$\mathbf{Y}_t = \mathbf{a} \oplus \check{X}_t \odot \mathbf{b} \oplus \epsilon_t$ $\oplus T(\check{Y})_t \odot \mathbf{c}$	$\check{Y}_t = a + \langle \mathbf{b}, \mathbf{X}_t \rangle_A$ $+ \epsilon_t + dT(\check{X})_t$	$\mathbf{Y}_t = \mathbf{a} \oplus \mathbf{B} \boxtimes \mathbf{X}_t \oplus \epsilon_t$ $\oplus T(\check{X})_t \odot \mathbf{d} + \oplus T(\check{Y})_t \odot \mathbf{c}$
			‘MCI’ model:
			$\mathbf{Y}_t = \mathbf{a} \oplus b \odot \mathbf{X}_t \oplus \epsilon_t$ $\oplus T(\check{X})_t \odot \mathbf{d} + \oplus T(\check{Y})_t \odot \mathbf{c}$
			‘CODA’ model:
\mathbb{R}^{D-1}	$\mathbf{Y}_t^* = \mathbf{a}^* + \mathbf{b}^* \check{X}_t + \epsilon_t^*$ $+ \mathbf{c}^* T(\check{Y})_t$	$\check{Y}_t =$ $a + \sum_{k=1}^{D_X-1} b_k^* X_{t,k}^* + \epsilon_t$ $+ dT(\check{X})_t$	$\mathbf{Y}_t^* = \mathbf{a}^* + \mathbf{B}^* \mathbf{X}_t^* + \epsilon_t^*$ $+ \mathbf{d}^* T(\check{X})_t + \mathbf{c}^* T(\check{Y})_t$
			‘MCI’ model:
			$\mathbf{Y}_t^* = \mathbf{a}^* + b \mathbf{X}_t^* + \epsilon_t^*$ $+ \mathbf{d}^* T(\check{X})_t + \mathbf{c}^* T(\check{Y})_t$
	$\mathbf{Y}_t, \mathbf{a}, \mathbf{b}, \epsilon_t \in \mathcal{S}^{D_Y}, \check{X}_t \in \mathbb{R}$ $\mathbf{Y}_t^*, \mathbf{a}^*, \mathbf{b}^*, \epsilon_t^* \in \mathbb{R}^{D_Y-1}$	$\mathbf{X}_t, \mathbf{b} \in \mathcal{S}^{D_X},$ $\check{Y}_t, \check{X}_t, a, \epsilon_t \in \mathbb{R}$ $\mathbf{X}_t^*, \mathbf{b}^* \in \mathbb{R}^{D_X-1}$	$\mathbf{B} \in \mathbb{R}^{D_Y, D_X}, b \in \mathbb{R}$ $\mathbf{B}^* \in \mathbb{R}^{D_Y-1, D_X-1}$

3. Elasticities, semi-elasticities and simplicial partial derivatives

A marginal impact in a linear regression model is usually understood as the change in the expected value of the dependent variable Y induced by an additive increase of the explanatory of interest X . In nonlinear models, it is rather understood as the infinitesimal equivalent, i.e. the derivative of the expected value of Y with respect to X and it may be non constant throughout the range of X . In some nonlinear models, an elasticity or a semi-elasticity may be more natural. Indeed in a log-log model, if $\mathbb{E}(\log(Y))$ depends linearly on $\log(X)$, then the parameter of $\log(X)$ is equal to the derivative of $\mathbb{E}(\log(Y))$ with respect to $\log(X)$ (also called the logarithmic derivative) and can be interpreted as the percent increase of $\mathbb{E}(Y)$ induced by a one percent increase of X : this quantity is called elasticity of Y with respect to X . Finally, if the model is a semi-log model, the natural quantity is either the partial derivative of $\mathbb{E}(Y)$ with respect to $\log(X)$ (if the logarithm is on the right hand side of the regression equation) or symmetrically the partial derivative of $\mathbb{E}(\log(Y))$ with respect to X in the other case (if the logarithm is on the left hand side of the regression equation): in both cases it is called a semi-elasticity. Because log-log models and semi-log models are frequent in econometrics, elasticities and semi-elasticities are often used to measure the impact of covariates.

This supports the idea that, in a simplicial regression model, one should turn attention to simplicial derivatives to evaluate the impacts of explanatory variables. Adapting the definition of derivative to the case where a function is simplex valued or is defined on the simplex stems

from the fact that a change in a vector of shares cannot be just reduced to a change in one of the components since they are linked by their sum constraint: in other words, it is due to the fact that one of the variables lies in a subspace of \mathbb{R}^D .

More precisely, the quantities of interest are

- $\frac{\partial^{\oplus} \mathbb{E}^{\oplus} \mathbf{Y}}{\partial X}$ in the case of the Y-compositional model
- $\frac{\partial \mathbb{E} Y}{\partial^{\oplus} \mathbf{X}}$ in the case of the X-compositional model,
- $\frac{\partial^{\oplus} \mathbb{E}^{\oplus} \mathbf{Y}}{\partial^{\oplus} \mathbf{X}}$ in the case of the YX-compositional model,

where \mathbb{E}^{\oplus} denotes the expectation of a simplex valued random variable (see Pawlowsky-Glahn and Buccianti (2011)) and where the symbol ∂^{\oplus} indicates that the derivative should be understood in the simplicial derivative sense with respect to that variable (see Barceló-Vidal, Martín-Fernández, and Mateu-Figueras (2011) and Egozcue, Jarauta-Bragulat, and Díaz-Barrero (2011)).

For the Y-compositional and X-compositional models, we are first going to express the relevant simplicial derivatives in terms of semi-elasticities.

Indeed, for the case of the X-compositional model, let us consider an homogeneous function of degree zero f defined from \mathbb{R}_+^D to \mathbb{R} inducing a function \underline{f} on \mathcal{S}^D by $\underline{f}(\mathbf{x}) = \underline{f}(\mathcal{C}(\tilde{\mathbf{x}})) = f(\tilde{\mathbf{x}})$. Propositions (13.10) and (13.13) in Barceló-Vidal, Martín-Fernández, and Mateu-Figueras (2011) imply that the part- \mathcal{C} derivatives of \underline{f} , which we denote here by $\frac{\partial \underline{f}(\mathbf{x})}{\partial^{\oplus} \mathbf{x}}$ are given by:

$$\frac{\partial \underline{f}(\mathbf{x})}{\partial^{\oplus} \mathbf{x}} = \frac{\partial f(\tilde{\mathbf{x}})}{\partial \log(\tilde{\mathbf{x}})}$$

Therefore the derivative of a function \underline{f} of a simplex valued variable $\mathbf{x} = \mathcal{C}(\tilde{\mathbf{x}})$ corresponds to the ordinary semi-log derivative of the corresponding homogeneous function f of the volumes $\tilde{\mathbf{x}}$. Applying this result to the function expressing $\mathbb{E}Y$ as a function of the share vector \mathbf{X} , we obtain the link between the simplicial derivative of this function and the semi-elasticity (or semi-log derivative) in the classical sense of the corresponding function of the volume vector $\tilde{\mathbf{X}}$.

Similarly, for the case of the Y-compositional model, for a simplex-valued function \mathbf{f} of a real variable $x \in \mathbb{R}$, Theorem 12.2.6 in Egozcue et al. in Egozcue, Jarauta-Bragulat, and Díaz-Barrero (2011) implies that:

$$\frac{\partial^{\oplus} \mathbf{f}(x)}{\partial x} = \mathcal{C} \left(\exp \left(\frac{\partial \log \mathbf{f}(x)}{\partial x} \right) \right)',$$

where $\partial^{\oplus} \mathbf{f}$ denotes the simplicial derivative of \mathbf{f} at x . This result links the simplicial derivatives of a simplex-valued function \mathbf{f} to the semi-log derivatives (in the ordinary sense) of this function. Applying this result to the function expressing $\mathbb{E}^{\oplus} \mathbf{Y}$ as a function of X , we obtain the link between the simplicial derivative of this function and the semi-elasticity (or semi-log derivative) in the classical sense of $\mathbb{E}^{\oplus} \mathbf{Y}$ as a function of X .

For the YX-compositional model, Morais, Thomas-Agnan, and Simioni (2018a) linked simplicial derivatives to elasticities in the case of a model without a total and in the particular case where the number of components D_Y of the Y composition is the same as that of the X composition (D_X). The limitation $D_Y = D_X$ in Morais, Thomas-Agnan, and Simioni (2018a) was simply due to the particular application framework of this work but there is no additional mathematical difficulty to extend the result to $D_Y \neq D_X$. The corresponding formulas are recalled in Table 2 for completeness.

Finally, considering models including a total, one would need to define infinitesimal paths in the \mathcal{T} -space. Instead we consider three types of infinitesimal variations as described in Section 4.3.

Table 2: Simplicial derivative and (semi-)elasticities

Y-compositional model	X-compositional model	YX-compositional model
$= \mathcal{C} \left(\exp \left(\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial X} \right) \right)'$	$\frac{\partial \mathbb{E} \mathbf{Y}}{\partial^{\oplus} \mathbf{X}} = \frac{\partial \mathbb{E} \mathbf{Y}}{\partial \log \mathbf{X}}$	$= \mathcal{C} \left(\exp \left(\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial \log \mathbf{X}} \right) \right)'$

For upcoming interpretations, it is interesting to consider first order Taylor approximations of such functions (of a simplex variable or simplex valued). For a function f from \mathcal{S}^D to \mathbb{R} , consider as in [Barceló-Vidal, Martín-Fernández, and Mateu-Figueras \(2011\)](#) the generating system $\mathbf{u}_1, \dots, \mathbf{u}_D$ of \mathcal{S}^D defined by

$$\begin{aligned} \mathbf{u}_j &= \left(\frac{D-1}{D} \right) \odot \mu_j \\ \mu_j &= (1, \dots, 1, \exp(1), 1, \dots, 1), \end{aligned}$$

where $\exp(1)$ is at the j^{th} position. From [Barceló-Vidal, Martín-Fernández, and Mateu-Figueras \(2011\)](#), the first order Taylor's approximation is given by

$$\underline{f}(\mathbf{x} \oplus \delta \odot \mathbf{u}_j) \sim \underline{f}(\mathbf{x}) + \delta \frac{\partial f(\check{\mathbf{x}})}{\partial \log(\check{\mathbf{x}}_j)}. \quad (1)$$

This additive (in the simplex sense) increase of $\delta \odot \mathbf{u}_j$ corresponds to a multiplicative increase of the j^{th} component while holding constant all other ratios of remaining components. It is also equivalent in coordinate space, for a proper choice of olr transformation, to increase additively one olr component while keeping all others constant. To summarize, note that the increment is given by the product of δ by the classical semi-elasticity, i.e., a semi-log derivative in the ordinary sense of the corresponding function of the volumes. As we will see in [Section 5](#), δ is proportional to the rate of change of x .

For a function \mathbf{f} from \mathbb{R} to \mathcal{S}^D , [Egozcue, Jarauta-Bragulat, and Díaz-Barrero \(2011\)](#) obtain the following first order Taylor approximation for a small additive increase $\delta > 0$ of $x \in \mathbb{R}$

$$\mathbf{f}(x + \delta) \sim \mathbf{f}(x) \oplus \delta \odot \frac{\partial^{\oplus} \mathbf{f}(x)}{\partial x}$$

As in [Morais \(2017\)](#), let us go one step further in the approximation. Indeed,

$$\mathbf{f}(x) \oplus \delta \odot \frac{\partial^{\oplus} \mathbf{f}(x)}{\partial x} = \mathbf{f}(x) \oplus \exp\left(\delta \frac{\partial \log \mathbf{f}(x)}{\partial x}\right).$$

Combining with a first order approximation of the exponential in a neighborhood of zero $\exp\left(\delta \frac{\partial \log \mathbf{f}(x)}{\partial x}\right) \sim 1 + \delta \frac{\partial \log \mathbf{f}(x)}{\partial x}$, we get the following approximation for the m^{th} component of $\mathbf{f}(x + \delta)$

$$\mathbf{f}_m(x + \delta) \sim \mathbf{f}_m(x) \left(1 + \delta \frac{\partial \log \mathbf{f}_m(x)}{\partial x}\right). \quad (2)$$

Taking the derivative of $\sum_{m=1}^D \mathbf{f}_m(x) = 1$, we get $\sum_{m=1}^D \mathbf{f}_m(x) \frac{\partial \log \mathbf{f}_m(x)}{\partial x} = 0$. Therefore the RHS vector in equation (2) belongs to \mathcal{S}^D . To summarize, note that in this case the percent increase of each component of $\mathbf{f}(x)$ is given by the classical semi-elasticity, i.e., a semi-log derivative in the ordinary sense of the function.

Finally for a function \mathbf{f} from \mathcal{S}_X^D , to \mathcal{S}_Y^D , a similar approximation has been obtained in [Morais \(2017\)](#) for the particular case $D_X = D_Y$. Combining the above two results, we obtain easily that the Taylor approximation of a function \mathbf{f} from \mathcal{S}^{D_X} to \mathcal{S}^{D_Y} is given by

$$\underline{f}_m(\mathbf{x} \oplus \delta \odot \mathbf{u}_j) \sim \underline{f}_m(\mathbf{x}) \left(1 + \delta \frac{\partial \log \underline{f}_m(\tilde{\mathbf{x}})}{\partial \log \tilde{x}_j} \right).$$

showing that a percent increase of the components of \mathbf{x} , proportional to δ , induces a percent increase of each component of $\mathbf{f}(x)$ given by the classical elasticity of the corresponding component $\frac{\partial \log \underline{f}_m(\tilde{\mathbf{x}})}{\partial \log \tilde{x}_j}$.

4. Elasticities and semi-elasticities in terms of model parameters

The aim is now to relate the elasticities/semi-elasticities of the previous section to the model parameters. The results of this section will be based on the following two lemmas which establish the formulas for the semi-log derivatives of an olr transformation and its inverse.

Lemma 4.1 *If \mathbf{z} is a D -composition which is the closure of the vector $\tilde{\mathbf{z}}$ of \mathbb{R}_+^D , and if $\mathbf{z}^* = \text{olr}_V(\mathbf{z}) = \mathbf{V}' \log(\mathbf{z})$ is the olr-transformed vector associated to the contrast matrix \mathbf{V} , then*

$$\frac{\partial \text{olr}_V(\mathbf{z})}{\partial \log \tilde{\mathbf{z}}} = \mathbf{V}'$$

This first lemma just results from the definition of the olr which is linear with respect to $\log \tilde{\mathbf{z}}$, and could be used for any other log-ratio linear transformation.

Lemma 4.2 *If \mathbf{z} is a D -composition which is the closure of the vector $\tilde{\mathbf{z}}$ of \mathbb{R}_+^D , and if $\mathbf{z}^* = \text{olr}_V(\mathbf{z}) = \mathbf{V}' \log(\mathbf{z})$ is the olr-transformed vector associated to the contrast matrix \mathbf{V} , then*

$$\frac{\partial \log(\text{olr}_V^{-1}(\mathbf{z}^*))}{\partial \mathbf{z}^*} = \mathbf{W}_z \mathbf{V},$$

where $\mathbf{z} = \text{olr}_V^{-1}(\mathbf{z}^*)$ and where $\mathbf{W}_z = I_D - \mathbf{1}_D \mathbf{z}'$ with I_D the the $D \times D$ identity matrix and $\mathbf{1}_D$ the $D \times 1$ vector of ones.

Let v_{ij} , ($i = 1, \dots, D$ and $j = 1, \dots, D - 1$), be the general term of the matrix \mathbf{V} . To prove Lemma 4.2, using the formula for the inverse transformation of an olr, one representative of $\log \tilde{\mathbf{z}} = \log(\text{olr}_V^{-1}(\mathbf{z}^*)) = \log \mathcal{C}(\exp(\mathbf{V} \mathbf{z}^*))$ is given by $\mathbf{V} \mathbf{z}^*$ and therefore its derivative with respect to \mathbf{z}^* is \mathbf{V} . Since $\log(\mathbf{z}) = \log(\tilde{\mathbf{z}}) - \log(S) \mathbf{1}_D$, where $S = T_A(\tilde{\mathbf{z}}) = \sum_{i=1}^D \tilde{z}_i$, and since

$$\frac{\partial S}{\partial z_j^*} = \sum_{k=1}^D \frac{\partial \log(\tilde{z}_k)}{\partial z_j^*} \tilde{z}_k = \sum_{k=1}^D v_{kj} \tilde{z}_k,$$

we have

$$\frac{\partial \log(S)}{\partial z_j^*} = \frac{1}{S} \frac{\partial S}{\partial z_j^*} = \sum_{k=1}^D v_{kj} z_i$$

Combining first and second terms yields, for $i = 1, \dots, D$ and $j = 1, \dots, D - 1$

$$\frac{\partial \log(z_i)}{\partial z_j^*} = v_{ij} - \sum_{k=1}^D v_{kj} z_k,$$

and this is the general term of the matrix $\mathbf{W}_z \mathbf{V}$.

If we define $\mathbf{W}_z^* = \mathbf{W}_z \mathbf{V}$, note that $\mathbf{W}_z^* \mathbf{V}' = \mathbf{W}_z$ (will be used later on).

4.1. Semi-elasticities for Y-compositional models and X-compositional models

In the case of Y-compositional and X-compositional models, the natural tool is semi-elasticities. However the formulas differ in the two cases:

- X-compositional case: $se(Y, \check{\mathbf{X}}) = \frac{\partial \mathbb{E}Y}{\partial \log \check{\mathbf{X}}}$
- Y-compositional case: $se(\mathbf{Y}, \check{X}) = \frac{\partial \log \mathbb{E}^\oplus \mathbf{Y}}{\partial \check{X}}$

Let us denote by $\mathbf{V}^{\mathbf{X}}$, respectively $\mathbf{V}^{\mathbf{Y}}$, the contrast matrices used for \mathbf{X} , respectively \mathbf{Y} . The computation in the X-compositional case uses Lemma 4.1. Indeed, for $j = 1, \dots, D_X$

$$\frac{\partial \mathbb{E}Y}{\partial \log \check{X}_j} = \sum_{k=1}^{D_X-1} \frac{\partial \mathbb{E}Y}{\partial X_k^*} \frac{\partial X_k^*}{\partial \log \check{X}_j} = \sum_{k=1}^{D_X-1} b_k^* \mathbf{v}_{jk}^{\mathbf{X}}$$

The result is reported in Table 3 with a matrix formulation

$$\frac{\partial \mathbb{E}(\check{Y})}{\partial \log \check{\mathbf{X}}} = \mathbf{V}^{\mathbf{X}} \mathbf{b}^* = \mathbf{V}^{\mathbf{X}} \mathbf{V}^{\mathbf{X}'} \log \mathbf{b} = clr(\mathbf{b}). \quad (3)$$

The computation in the Y-compositional case uses Lemma 4.2 since $\mathbb{E}^\oplus \mathbf{Y} = olr_V^{-1}(\mathbb{E}\mathbf{Y}^*)$. We have

$$\begin{aligned} \frac{\partial \log \mathbb{E}^\oplus \mathbf{Y}}{\partial \check{X}} &= \frac{\partial \log \mathbb{E}^\oplus \mathbf{Y}}{\partial \mathbb{E}\mathbf{Y}^*} \frac{\partial \mathbb{E}\mathbf{Y}^*}{\partial \check{X}} = \mathbf{W}_{\mathbf{z}}^* \mathbf{b}^* = \mathbf{W}_{\mathbf{z}}^* \mathbf{V}^{\mathbf{Y}'} \log \mathbf{b} = \mathbf{W}_{\mathbf{z}} \mathbf{V}^{\mathbf{Y}} \mathbf{V}^{\mathbf{Y}'} \log \mathbf{b} = \\ &= \mathbf{W}_{\mathbf{z}} clr(\mathbf{b}) \end{aligned} \quad (4)$$

where $\mathbf{z} = olr_V^{-1}(\mathbb{E}(olr_V \mathbf{Y})) = \mathbb{E}^\oplus \mathbf{Y}$.

Expressions (3) and (4) underline the fact that the semi-elasticities are independent of the particular contrast matrix. They are observation dependent in the Y-compositional case through $\mathbf{z} = \mathbb{E}^\oplus \mathbf{Y}$.

4.2. Elasticities for the YX-compositional model

For the YX-compositional model, [Morais, Thomas-Agnan, and Simioni \(2018a\)](#) have obtained the expressions of the elasticities when the dimension of the Y composition is the same as that of the X composition. Let us extend this result to the case $D_X \neq D_Y$ using the above two lemmas.

We can see the relationship between $\log \check{\mathbf{X}}$ and $\log \mathbb{E}^\oplus \mathbf{Y}$ as the composition of three functions (listed from inside to outside)

- the function which maps $\log \check{\mathbf{X}} \in \mathbb{R}^{+D_X}$ to $\mathbf{X}^* \in \mathcal{S}^{D_X}$
- the function which maps $\mathbf{X}^* \in \mathcal{S}^{D_X}$ to $\mathbb{E}\mathbf{Y}^* \in \mathbb{R}^{+D_Y}$
- the function which maps $\mathbb{E}\mathbf{Y}^* \in \mathbb{R}^{+D_Y}$ to $\log \mathbb{E}^\oplus \mathbf{Y} \in \mathcal{S}^{D_Y}$

Using the generalized chain rule for functions of several variables which states that the Jacobian matrix of the composite function is the product of the Jacobian matrices of the composed functions evaluated at appropriate points, we get

$$\frac{\partial \log \mathbb{E}^\oplus \mathbf{Y}}{\partial \log \check{\mathbf{X}}} = \frac{\partial \log \mathbb{E}^\oplus \mathbf{Y}}{\partial \mathbb{E}\mathbf{Y}^*} \frac{\partial \mathbb{E}\mathbf{Y}^*}{\partial \mathbf{X}^*} \frac{\partial \mathbf{X}^*}{\partial \log \check{\mathbf{X}}} \quad (5)$$

The rightmost term on the right hand side of (5) is obtained using Lemma 4.1:

$$\frac{\partial \mathbf{X}^*}{\partial \log \check{\mathbf{X}}} = \mathbf{V}^{\mathbf{X}'}$$

The central term yields the matrix \mathbf{B}^* of parameters in coordinate space since the relationship between $\mathbb{E}\mathbf{Y}^*$ and \mathbf{X}^* is linear. The leftmost term on the right hand side is obtained using Lemma 4.2:

$$\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial \mathbb{E} \mathbf{Y}^*} = \mathbf{W}_z \mathbf{V}^{\mathbf{Y}},$$

where $\mathbf{z} = \mathbb{E}^{\oplus} \mathbf{Y}$. We finally get

$$\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial \log \tilde{\mathbf{X}}} = \mathbf{W}_z \mathbf{V}^{\mathbf{Y}} \mathbf{B}^* \mathbf{V}^{\mathbf{X}'} = \mathbf{W}_z \mathbf{V}^{\mathbf{Y}} \mathbf{V}^{\mathbf{Y}'} \mathbf{B} = \mathbf{W}_z \mathbf{B}, \quad (6)$$

using the relationships between B and B^* (see e.g. Nguyen, Laurent, Thomas-Agnan, and Ruiz-Gazen (2018)), and using the fact that the matrix B satisfies the zero-sum property (sum of rows equal sum of columns equals 0.)

Note that the elasticity is observation dependent through $\mathbf{z} = \mathbb{E}^{\oplus} \mathbf{Y}$.

For the MCI model, we have that $\mathbf{B} = b \mathbf{V}^{\mathbf{Y}} \mathbf{V}^{\mathbf{Y}'}$, and therefore

$$\mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}} \mathbf{B} = b \mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}} \mathbf{V}^{\mathbf{Y}} \mathbf{V}^{\mathbf{Y}'} = b \mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}}$$

Table 3 summarizes the different formulas for semi-elasticities and elasticities for the three types of models as a function of parameters estimates, in the simplex or in coordinate space. Both expressions are important to keep in mind: the expression as a function of the simplex parameters makes it clear that these are intrinsically simplex quantities independent of any transformation. The expression as a function of coordinate space parameters is handy for computations.

Table 3: (Semi-)elasticities without total

Y-compositional model	X-compositional model	YX-compositional model
		‘CODA’ Model
$\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial X} = \mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}}^* \mathbf{b}^*$ $= \mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}} \text{clr}(\mathbf{b})$	$\frac{\partial \mathbb{E}(\check{Y})}{\partial \log \tilde{\mathbf{X}}} = \mathbf{V}^{\mathbf{X}} \mathbf{b}^*$ $= \text{clr}(\mathbf{b})$	$\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial \log \tilde{\mathbf{X}}} = \mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}}^* \mathbf{B}^* \mathbf{V}^{\mathbf{X}'} =$ $\mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}} \mathbf{B}$
		‘MCI’ Model
		$\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial \log \tilde{\mathbf{X}}} = b \mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}}$

Notations $\mathbf{W}_z = I_D - \mathbf{1}_D \mathbf{z}'$, $\mathbf{W}_z^* = \mathbf{W}_z \mathbf{V}^{\mathbf{Y}}$

4.3. Models including a total

The presence of the total variable has to be taken into account in the partial impact measure computations. We consider including among the explanatory variables

- a total of \mathbf{Y} in the Y-compositional model (model A)
- a total of \mathbf{X} in the X-compositional model (model B)
- a total of \mathbf{X} and/or a total of \mathbf{Y} in the YX-compositional model (model C)

The right hand side of model equations from Table 1 are modified as follows

- model A: add $\oplus T(\mathbf{Y}_t) \odot \mathbf{c}$, where $\mathbf{c} \in \mathcal{S}^{D_Y}$ is the parameter corresponding to the total effect of \mathbf{Y}

- model B: add $+dT(\mathbf{X}_t)$, where $d \in \mathbb{R}$ is the parameter corresponding to the total effect of \mathbf{X}
- model C: add $\oplus T(\mathbf{Y}_t) \odot \mathbf{c} \oplus T(\mathbf{X}_t) \odot \mathbf{d}$, where $c \in \mathcal{S}^{D_Y}$ and $d \in \mathcal{S}^{D_X}$ are the parameters corresponding to the two total effects.

In the presence of a total, as mentioned in Section 3, we need to distinguish three types of infinitesimal variations for a compositional variable \mathbf{X} . The three types are as follows

- Type 1: the total $T(\mathbf{X})$ remains constant and we look at infinitesimal variations of the composition \mathbf{X} . Such variations correspond to considering derivatives in the direction of one of the unitary vectors of an orthonormal basis of \mathcal{S}^{D_X} . With a proper choice of basis and of contrast matrix as in [Hron, Filzmoser, and Thompson \(2012\)](#), this corresponds to an infinitesimal change in one component, along a linear path in the simplex, keeping all but the first olr coordinate constant.
- Type 2: the composition \mathbf{X} remains constant while the total is subject to an infinitesimal variation. Such variations correspond to considering ordinary derivatives with respect to the total $T(\mathbf{X})$.
- Type 3: one of the components of \mathbf{X} varies together with the total $T(\mathbf{X})$.

In model A, the impact of additively increasing the total of \mathbf{Y} is the same question as the impact of a non-compositional variable and therefore the formula of Table 3 can be applied with c instead of b .

Type 1 variations of \mathbf{X} In model B, the impact of a type 1 variation of \mathbf{X} with fixed total can be computed as in the X-compositional model in Table 3. The impact of a type 1 variation of \mathbf{X} in model C can be computed as in the YX-compositional model in Table 3.

Type 2 variations of \mathbf{X} Type 2 variations of \mathbf{X} correspond to ordinary derivatives with respect to the total of \mathbf{X} .

For model B, a type 2 variation of \mathbf{X} results in an ordinary derivative

$$\frac{\partial \mathbb{E}Y}{\partial T} = d$$

In model C, a type 2 variation of \mathbf{X} can be computed as in a Y-compositional model treating the total $T(\mathbf{X})$ as an ordinary variable and formula from Table 3 can be applied with \mathbf{d} instead of \mathbf{b} .

Type 3 variations of \mathbf{X} In this case, evaluating the effect of the variation of \mathbf{X} or of $T(\mathbf{X})$ is equivalent since they are linked together, therefore one of the two formulas is enough. For type 3 variations of \mathbf{X} , since both total and composition vary, the easiest way out is to express the dependent as a function of the volumes and use ordinary derivatives of the ensuing function of the volumes.

In model B, for computing the effect of a type 3 variation of \mathbf{X} , we need to adapt equation (3) adding an extra term taking into account the fact that the total depends upon the volumes and we get

$$\frac{\partial \mathbb{E}Y}{\partial \log \check{\mathbf{X}}} = \mathbf{V}^{\mathbf{X}} \mathbf{b}^* + \frac{\partial \log \mathbb{E}Y}{\partial T} \frac{\partial T}{\partial \log \check{\mathbf{X}}} = \mathbf{V}^{\mathbf{X}} \mathbf{b}^* + d \frac{\partial T}{\partial \log \check{\mathbf{X}}}$$

This result shows that the derivatives of the total with respect to the volumes play a role in the final expression of this semi-elasticity (hence we get a different formula for an arithmetic or a geometric total).

In model C, for a type 3 variation of \mathbf{X} , the derivative with respect to \mathbf{X} of the first term $\mathbf{B} \boxplus \mathbf{X}$ is obtained as in the YX-compositional model without total and and the derivative of the second term $T(\mathbf{X}) \odot \mathbf{d}$ is obtained as in the X-compositional model with a $T(\mathbf{X})$ total (equation (6)) yielding overall

$$\frac{\partial \log \mathbb{E}^{\oplus} \mathbf{Y}}{\partial \log \check{\mathbf{X}}} = \mathbf{W}_{\mathbb{E}^{\oplus} \mathbf{Y}} (\mathbf{B} + \log(\mathbf{d})) \frac{\partial T}{\partial \log \check{\mathbf{X}}}$$

Once again, the result involves the derivatives of the total with respect to the volumes.

No additional complexity is introduced if we consider models involving a total of \mathbf{Y} as an additional dependent variable, which makes sense in the Y-compositional or YX-compositional cases. In that case consequently, there would be no total of \mathbf{Y} among the explanatory variables. The impact of variations of a classical (resp: compositional) explanatory variable on the compositional part of \mathbf{Y} would be studied as in the Y-compositional model (resp: YX-compositional) and on the total of \mathbf{Y} as in an ordinary linear model (resp: X-compositional model).

5. Illustration

Let us give two toy examples of interpretation to illustrate our approach. We focus on the X-compositional and the Y-compositional models since the case of the YX-compositional model was already illustrated in [Morais, Thomas-Agnan, and Simioni \(2018a\)](#). Both examples involve time series data and would justify a compositional time series model. However since the focus is just on illustrating the impacts evaluation, we will ignore the time series aspect and pretend the observations are i.i.d.. The subsequent computations were done using the R package *compositions*.

5.1. Economic context and automobile market: Y-compositional model

In [Morais and Thomas-Agnan \(2020\)](#), the relationship between the socioeconomic context on the demand of new cars by segments is investigated with a data set coming from the French Renault company for market shares and from publicly available data bases. The data coming from Renault has been blurred with a small noise for confidentiality reasons. The automobile market is divided into five segments, from the smallest vehicles (A segment) to the largest vehicles (E segment). The available explanatory variables are consumption expenditure, an economic sentiment indicator, Gross Fixed Capital Formation of household, Gross Domestic Product, diesel price and short term interest rate. The data is recorded monthly from 2003 to 2015 (167 observations). The model explaining the market shares of each segment by the above explanatory is therefore a Y-compositional model in our terminology. We use the following sequential binary partition: B versus A, C versus A and B, D versus A, B and C, and E versus A, B, C and D to construct an orthonormal basis of the simplex and an associated olr transformation. Figure 1 displays the observed and predicted segments shares over time and we can see that the compositional model catches the general tendency and smoothes the jiggly patterns without suffering from overfitting, but not all the variance of this data which differs across shares. The quality of fit of this model can be assessed by the multivariate adjusted coefficient of determination based on the proportion of metric variance explained by the model and which is equal to 0.86. Table 4 contains the average semi-elasticities of segments shares with respect to GDP.

Assuming the fitted model is correct, let us interpret for example the effect of a small increase of GDP on the small cars (A segment) market shares. From formula (2), a small additive increase $\delta = 1$ billion euros (this amount representing 0.6% of the average monthly GDP) results on average in a multiplicative increase of 0.0028 % of the A segment market share. Instead of focusing on average elasticities, we could concentrate on a given point in time and

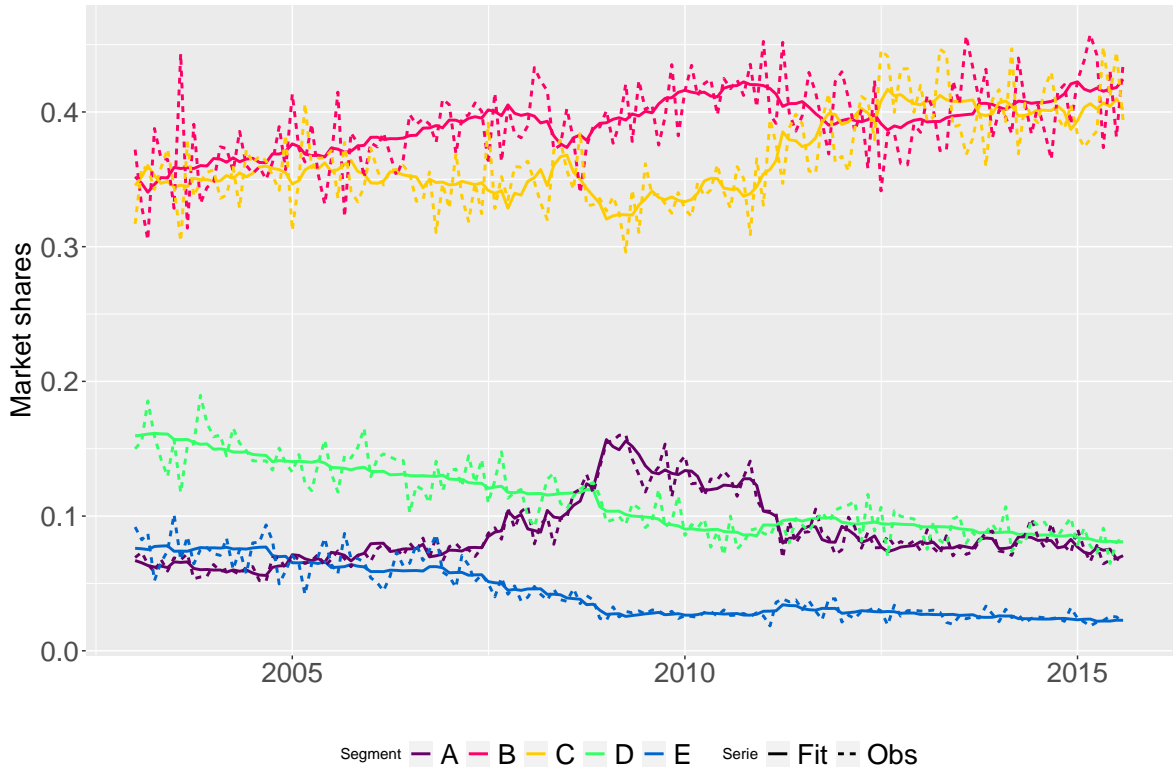


Figure 1: Observed (in dotted line) and predicted (in solid line) segments shares over time

Table 4: Average semi-elasticities of segments shares with respect to GDP

	$\overline{se}(S_t, GDP_t)$
A	2.88e-05
B	-0.17e-05
C	-0.96e-05
D	0.99e-05
E	1.18e-05

compute the impact on the whole share vector of such a small increase in GDP. We could then check easily that the new shares vector is indeed in the simplex.

5.2. French GDP and job market: X-compositional model

In this second illustration, we are interested in the impact on French GDP of the structure (composition) and the volume (total) of the French job market in the three main sectors of activity: Agriculture (primary), Industry (secondary), and Services (tertiary). GDP is expressed in million euros (current price) and total employment in thousands of people. The data is collected quarterly from 2004 to 2018¹. We use the *olr* transformation corresponding to the sequential binary partition: Agriculture versus Industry and Services, and Industry versus Services. We consider the model explaining the GDP as a function of total employment and the two *olr* coordinates associated to the above *olr* transformation. It is therefore an X-compositional model including a total, in this case the simple arithmetic total employment. The adjusted R square for this model is of 0.92. Table 5 reports the semi-elasticities of GDP with respect to the three sectors at the mean value of the sector composition corresponding to 788, 9196 and 19385 thousand employees for respectively Agriculture, Industry and Services. To

¹<https://data.oecd.org/emp/employment-by-activity.htm>

apply formula (1) in the neighborhood of the mean sector composition in volumes, we consider a small $\delta > 0$ and a variation of $\oplus\delta \ominus \mathbf{u}_j$ of \mathbf{x} , where \mathbf{u}_j is the unit vector in the direction of the component Services. This variation of \mathbf{x} is equivalent, when δ is small, to a relative variation of $\sqrt{3/2\delta}$ (i.e. multiplying \mathbf{x} by $1 + \sqrt{3/2\delta}$.) The factor $\sqrt{3/2}$ is $\sqrt{D_X/D_X - 1}$ in the general case, corresponding to $\log(u_j)$ in the Taylor expansion in [Barceló-Vidal, Martín-Fernández, and Mateu-Figueras \(2011\)](#). Taking $\delta = 0.01\%$ results in an increase of around $\sqrt{3/2} * 19385 * 0.0001 = 2450$ people of the Services employment while the ratio between Agriculture and Industry employments remain constant, and, assuming the fit is correct, we can see in [Table 5](#) that the model predicts that GDP should increase by 84 million euros. The marginal effect of the size of the job market, assuming that its composition stays the same is obtained by the parameter estimate of total employment in the model, which is equal to 26.52. When total employment increases by 1000 people, the predicted GDP increase is 26.5 millions.

Table 5: Semi-elasticities of GDP with respect to employment sectors composition

	$\overline{se}(GDP, \mathbf{Emp} \mathbf{Sect})$
AGR	-10157.26
INDU	-51706.00
SERV	841030.75

Note that using a base 2 logarithm as in [Muller, Hron, Fiserova, Smahaj, Cakirpaloglu, and Vancakova \(2018\)](#) is not useful in our approach and would rather introduce an unnecessary constant.

6. Conclusion

This contribution highlights the fact that elasticities or semi-elasticities are well-adapted to interpret the impacts of explanatory variables in all types of compositional regression models. It also links these elasticities or semi-elasticities to the simplicial derivatives of the expected response with respect to the considered explanatory variable. The models may contain compositional variables on the right hand side and/or on the left hand side of the regression equation, and may contain or not total variables (relative to the dependent or the explanatory variables). Further work should be done about confidence intervals for (semi-)elasticities which can be computed by the Delta method, or simply using a bootstrap approach. An extension to time series compositional model as well as to spatial compositional model involves more complex elasticities computations which take into account the time-lag and spatial-lag operators as can be seen in [Thomas-Agnan, Laurent, Ruiz-Gazen, Nguyen, Chakir, and Lungarska \(2020\)](#) for the spatial case.

An alternative but more complex tool used in [Wang, Shangguan, Wu, and Guan \(2013\)](#) and in [Morais, Thomas-Agnan, and Simioni \(2018a\)](#) is the elasticity of a ratio of shares. In the framework of an MCI model, it would directly correspond to a parameter of the model, which is attractive, but relates to a change rate of a ratio of components and not of a single component and therefore is more difficult to vulgarize.

Acknowledgements

We acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

References

- Aitchison J (1986). *The Statistical Analysis of Compositional Data*. Monographs on statistics and applied probability. Chapman and Hall, Reprinted in 2003 with additional material by Blackburn Press.
- Barceló-Vidal C, Martín-Fernández JA, Mateu-Figueras G (2011). “Compositional Differential Calculus on the Simplex.” *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons.
- Bui TTT, Loubes JM, Risser L, Balaesque P (2018). “Distribution Regression Model with a Reproducing Kernel Hilbert Space Approach.” *arXiv preprint arXiv:1806.10493*.
- Chen J, Zhang X, Li S (2017). “Multiple Linear Regression with Compositional Response and Covariates.” *Journal of Applied Statistics*, **44**(12), 2270–2285.
- Coenders G, Ferrer-Rosell B, Mateu-Figueras G, Pawlowsky-Glahn V (2015). “MANOVA of Compositional Data with a Total.” *CODAWORK2015*.
- Coenders G, Martín-Fernández JA, Ferrer-Rosell B (2017). “When Relative and Absolute Information Matter: Compositional Predictor with a Total in Generalized Linear Models.” *Statistical Modelling*, **17**(6), 494–512.
- Coenders G, Pawlowsky-Glahn V (2020). “On Interpretations of Tests and Effect Sizes in Regression Models with a Compositional Predictor.” *SORT*, **44**(1).
- Combettes PL, Muller CL (2019). “Regression Models for Compositional Data: General Log-contrast Formulations, Proximal Optimization, and Microbiome Data Applications.” URL <https://arxiv.org/abs/1903.01050>.
- Egozcue JJ, Daunis-I-Estadella J, Pawlowsky-Glahn V, Hron K, Filzmoser P (2012). “Simplicial Regression. The Normal Model.” *Journal of Applied Probability and Statistics*.
- Egozcue JJ, Jarauta-Bragulat E, Díaz-Barrero J (2011). “Calculus of Simplex-valued Functions.” *Compositional Data Analysis: Theory and Applications*.
- Filzmoser P, Hron K, Templ M (2018). *Applied Compositional Data Analysis, With Worked Examples in R*. Springer series in Statistics. Springer.
- Hron K, Filzmoser P, Thompson K (2012). “Linear Regression with Compositional Explanatory Variables.” *Journal of Applied Statistics*, **39**(5), 1115–1128.
- Kynclova P, Filzmoser P, Hron K (2015). “Modeling Compositional Time Series with Vector Autoregressive Models.” *Journal of Forecasting*, **34**(4), 303–314.
- Martín-Fernández JA (2019). “Comments on: Compositional Data: The Sample Space and Its Structure.” *TEST*, **28**, 653–657. URL <https://doi.org/10.1007/s11749-019-00672-4>.
- Morais J (2017). *Impact of Media Investments on Brands’ Market Shares: A Compositional Data Analysis Approach*. Ph.D. thesis, Toulouse School of Economics (TSE).
- Morais J, Thomas-Agnan C (2020). “Impact of the Economic Context on the Automobile Market Segment Shares: A Compositional Approach.” *Preprint*.
- Morais J, Thomas-Agnan C, Simioni M (2017). “Impact of Advertising on Brand’s Market-shares in the Automobile Market: A Multi-channel Attraction Model with Competition and Carryover Effects.” URL <https://hal.archives-ouvertes.fr/hal-01666853/>.
- Morais J, Thomas-Agnan C, Simioni M (2018a). “Interpretation of Explanatory Variables Impacts in Compositional Regression Models.” *Austrian Journal of Statistics*, **47**(5), 1–25.

- Morais J, Thomas-Agnan C, Simioni M (2018b). “Using Compositional and Dirichlet Models for Market Share Regression.” *Journal of Applied Statistics*, **45**(9), 1670–1689.
- Muller I, Hron K, Fiserova E, Smahaj J, Cakirpaloglu P, Vancakova J (2018). “Interpretation of Compositional Regression with Application to Time Budget Analysis.” *Austrian Journal of Statistics*, **47**(2), 3–19. doi:10.17713/ajs.v47i2.652. URL <https://www.ajs.or.at/index.php/ajs/article/view/vol47-2-1>.
- Nakanishi M, Cooper LG (1982). “Simplified Estimation Procedures for MCI Models.” *Marketing Science*, **1**(3), pp. 314–322. ISSN 07322399. URL <http://www.jstor.org/stable/183931>.
- Nguyen THA, Laurent T, Thomas-Agnan C, Ruiz-Gazen A (2018). “Analyzing the Impacts of Socio-economic Factors on French Departmental Elections with CODA Methods.” *TSE Working paper 18-961*.
- Pawlowsky-Glahn V, Buccianti A (2011). *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons.
- Pawlowsky-Glahn V, Egozcue JJ, Lovell D (2015a). “Tools for Compositional Data with a Total.” *Statistical Modelling*, **15**(2), 175–190.
- Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015b). *Modeling and Analysis of Compositional Data*. John Wiley & Sons.
- Sun Z, Xu W, Cong X, Chen K (2018). “Log-Contrast Regression with Functional Compositional Predictors: Linking Preterm Infant’s Gut Microbiome Trajectories in Early Postnatal Period to Neurobehavioral Outcome.” *arXiv preprint arXiv:1808.02403*.
- Thomas-Agnan C, Laurent T, Ruiz-Gazen A, Nguyen T, Chakir R, Lungarska A (2020). “Spatial Simultaneous Autoregressive Models for Compositional Data: Application to Land Use.” *TSE Working Paper*, **20**(1098).
- Trinh HT, Morais J, Thomas-Agnan C, Simioni M (2018). “Relations between Socio-economic Factors and Nutritional Diet in Vietnam from 2004 to 2014: New Insights Using Compositional Data Analysis.” *Statistical Methods in Medical Research*, p. 0962280218770223.
- Van Den Boogaart KG, Tolosana-Delgado R (2013). *Analyzing Compositional Data with R*. Springer.
- Wang H, Shangquan L, Wu J, Guan R (2013). “Multiple Linear Regression Modeling for Compositional Data.” *Neurocomputing*, **122**, 490–500.

Affiliation:

Christine Thomas-Agnan
Toulouse School of Economics
Esplanade de l’université
31080 Toulouse Cedex 06 France
E-mail: christine.thomas@tse-fr.eu
URL: <https://www.tse-fr.eu/fr/people/christine-thomas-agnan>

Austrian Journal of Statistics
published by the Austrian Society of Statistics

<http://www.ajs.or.at/>
<http://www.osg.or.at/>

Volume 50
January 2021

Submitted: 2019-12-02
Accepted: 2020-07-18
