WORKING PAPERS

N° 16-742

December 2016

"A tour of regression models for explaining shares"

Joanna Morais, Christine Thomas-Agnan, and Michel Simioni

Toulouse School of Economics

A tour of regression models for explaining shares

Joanna Morais[ab], **Christine Thomas-Agnan**[a], **Michel Simioni**[c]

[a] Toulouse School of Economics, University of Toulouse Capitole, 21 allee de Brienne, Toulouse, France

[b] BVA, 52 rue Marcel Dassault, Boulogne-Billancourt, France

[c] INRA, UMR 1110 MOISA, 2 Place Pierre Viala, Montpellier, France

joanna.morais@live.fr

December 6, 2016

Abstract

This paper aims to present and compare statistical modeling methods adapted for shares as dependent variables. Shares are characterized by the following constraints: positivity and sum equal to 1. Four types of models satisfy this requirement: multinomial logit models widely used in discrete choice models of the econometric literature, market-share models from the marketing literature, Dirichlet covariate models and compositional regression models from the statistical literature. We highlight the properties, the similarities and the differences between these models which are coming from the assumptions made on the distribution of the data and from the estimation methods. We prove that all these models can be written in an attraction model form, and that they can be interpreted in terms of direct and cross elasticities. An application to the automobile market is presented where we model brand market-shares as a function of media investments in 6 channels in order to measure their impact, controlling for the brands average price and a scrapping incentive dummy variable. We propose a cross-validation method to choose the best model according to different quality measures.

Keywords: Multinomial logit; Market-shares models; Compositional data analysis; Dirichlet regression.

1 Introduction

A large number of fields are concerned by the analysis of share data. Our primary motivation for studying this type of data comes from marketing where there is an interest for modeling market-shares, see for example Cooper and Nakanishi [7]. In political economy, Elff [8] studies voting behaviors and analyzes the relationship between the shares of political parties and their policy positions in different groups of voters. In geology, Solana-Acosta and Dutta

[39] are interested in the lithologic composition of sandstone according to whether it is quartz, feldspar or rock fragments. For environmental planning purposes, land use models focus on what are the proportions of different types of uses (forest, agriculture, urban, etc...) on a given piece of land, see for example Chakir et al. [6].

This paper aims to present statistical modeling methods adapted for share data as dependent variable (e.g. market-shares), which are characterized by the following constraints: they are positive and sum up to 1. By definition shares are "compositional data": a composition is a vector of parts of some whole which carries relative information. For a composition of D parts, if $D-1$ parts are known the D^{th} part is simply 1 minus the sum of the $D-1$ other parts. Indeed, D-compositions lie in a space called the simplex $\mathcal{S}^D$, not in the Euclidean space $\mathbb{R}^D$. Because of these constraints, classical regression models cannot be used directly.

There are four main types of models which incorporate the constraints of this kind of data: multinomial logit models are very frequent in the econometric literature, market-share models in the marketing literature, Dirichlet covariate models and compositional regression models in the statistical literature. We highlight the properties and the differences between these models which are coming from the assumptions made on the distribution of the data and from the estimation methods. We prove that all these models can be written in an attraction model form, and we derive the direct and cross elasticities formula to interpret their parameters. We also show that the market-shares models can be expressed in a compositional way.

An application to the automobile market is presented where we model brands market-shares as a function of media investments in 6 channels (TV, press, radio, outdoor, digital, cinema) in order to measure their impact, controlling for the brands average price and a scrapping incentive dummy variable. We propose a cross-validation method to choose the best model according to different quality measures adapted to shares data.

The present paper is organized as follows: the four models adapted to model shares data are presented in Section 2, and theoretically compared in detail in Section 3. Section 4 presents an application to an automobile market data set, along with an empirical comparison of the models in terms of cross-validated goodness-of-fit measures, and an example of elasticity interpretation. Finally, last section concludes on the findings and on directions to be investigated.

2 Models for explaining shares

2.1 Notations

In order to compare the four different models, the notations are standardized in Table 1 depending on whether the variables are considered in volume or in share, in the left or in the right part of the regression equation, and if they are alternative and/or observation dependent. For example, in the case we use for illustration, the dependent variable is the sales of vehicles observed across time; among the explanatory we have media investments, price (depends on brands and on time) and scrapping incentive (depends only on time). The sales can be considered in volume (number of sales) or in share (market-shares). Similarly, media investments in volume correspond to the amount of euros spent, in share it corresponds to the so-called “shares-of-voice” in marketing.

$\mathcal{C}()$ denotes the closure operation which transforms volumes into shares:

$$\mathcal{C}(y_1,\dots,y_D) = \left(\frac{y_1}{\sum_{j=1}^{D} y_j},\dots,\frac{y_D}{\sum_{j=1}^{D} y_j}\right)$$

Variable	**Volumes** (absolute values)	**Shares** (relative values)
Dependent	N_{jt}	$\mathbf{S_t} = (S_{1t},\dots,S_{Dt}) = \mathcal{C}(N_{1t},\dots,N_{Dt})$
Explanatory (observation and component characteristic)	X_{jt}	$\mathbf{Z_t} = (Z_{1t},\dots,Z_{Dt}) = \mathcal{C}(X_{1t},\dots,X_{Dt})$
Explanatory (observation characteristic only)	W_t	
General notations		
D	Number of components (3 in the application)	
$j,l,m = 1,\dots,D$	Index of components (brands in the application)	
T	Number of observations (123 in the application)	
$t = 1,\dots,T$	Index of observations (time in the application)	
K, K_X, K_W	Number of explanatory variables / of type X / of type W	
$k = 1,\dots,K$	Index of explanatory variables (by default)	
$k = 1,\dots,K_X$	Index of explanatory variables of type X	
$\kappa = 1,\dots,K_W$	Index of explanatory variables of type W	
s_j	Theoretical mean share (expected value of S_j)	
Notations for the application		
C	Number of media channels (6 in the application)	
$c = 1,\dots,C$	Index of media channels	
M_{cjt}	Media investment in channel c at time t for brand j	
P_{jt}	Average price at time t of brand j	
I_t	Scrapping incentive dummy at time t	

Table 1: Notations

2.2 Multinomial logit models

In econometrics, Multinomial logit (MNL) models are widely used to model discrete choices of individuals, i.e. to model the probability that an individual i chooses an alternative j, using individual data. Sometimes this data are aggregated using a group variable (time, space, age-group, etc) and then the counts for each alternative and the covariates are recorded in each group. We are going to describe how an individual-level MNL model can be adapted to aggregated data, provided that the explanatory variables are either describing the alternatives (and are constant for all decision makers in a group) or are group characteristics.

2.2.1 Discrete choice model: a random utility model for individual data

Multinomial logit models (MNL) are usually widely known by statisticians because they are a generalization of the famous binary logistic regression model. MNL is a particular case of discrete choice models, used to explain and predict polytomous, discrete or qualitative, response variable (a finite set of mutually exclusive and collectively exhaustive alternatives) by a set of explanatory variables[1].

In econometrics, random utility models are based on the idea that decision makers are choosing the alternative that maximizes their utility. For an introduction to utility in econometrics, see for example McFadden [27]. Thus, the probability for decision maker i to choose alternative j at choice situation t is defined as:

$$p_{ijt} = \mathbb{P}(Choice_{it} = j) = \mathbb{P}[U_{ijt} \geq U_{ilt}, \quad \forall\, l \neq j] \tag{1}$$

where $Choice_{it}$ is the variable of choice of individual i at choice situation t, and U_{ijt} is the utility associated to alternative j for decision maker i at choice situation t.

Random utility models decompose the utility U_{ijt} as a sum of a deterministic part V_{ijt} and a random part ϵ_{ijt}:

$$U_{ijt} = V_{ijt}(X_t) + \epsilon_{ijt}$$

where X is a set of explanatory variables for the deterministic part of the utility.

If error terms are extreme-value (Gumbel) distributed, the computations of probabilities from (1) (Koppelman and Bhat [20]) have a closed form leading to the Multinomial Logit model (also called random coefficient logit model)

$$p_{ijt} = \frac{\exp(U_{ijt})}{\sum_{l=1}^{D} \exp(U_{ilt})} \tag{2}$$

which can be estimated by maximum likelihood (using the density of the multinomial distribution) on individual-level data.

2.2.2 Conditional logit model: alternative-specific explanatory variables

If explanatory variables only characterize alternatives (and not individuals), MNL is called "conditional logit model". If alternative characteristics do not change across decision makers, the conditional logit model can be expressed in an aggregated way, using count data instead of individual data, which means that only the numbers of individuals who have chosen each alternative are needed instead of the individual choices. This is the case for our illustration data: it allows us to estimate the market-share (probability) of a brand depending on the characteristics of this brand relatively to the characteristics of other brands in competition.

[1] See Koppelman and Bhat [20]

The expected share of alternative j at choice situation t (e.g. market-share of brand j at time t) corresponds actually to the probability of j to be chosen by an individual, and is expressed as

$$s_{jt} = \mathbb{E}(S_{jt}|X_t) = \frac{\exp(a_j + \sum_{k=1}^{K} b_k X_{kjt})}{\sum_{l=1}^{D} \exp(a_l + \sum_{k=1}^{K} b_k X_{klt})} \quad (3)$$

with $a_D = 0$ for identifiability reasons.

2.2.3 Estimation by maximum likelihood for aggregated data

The multinomial distribution is a generalization of the binomial distribution. For N independent individuals who choose exactly one of D alternatives, with each alternative having a given probability to be chosen, the multinomial distribution gives the probability of any particular combination of numbers of choices for the various alternatives: $(N_1, \ldots, N_D) \sim \mathcal{MN}(N; s_1, \ldots, s_D)$, where $N = \sum_{j=1}^{D} N_j$, such that:

$$\mathbb{E}(N_j) = Ns_j \quad ; \quad Var(N_j) = Ns_j(1 - s_j) \quad ; \quad Cov(N_j, N_l) = -Ns_j s_l$$

If the explanatory variables characterizing the alternatives do not change across individuals (for example the price of the vehicle j is the same for all households i), then the utility for alternative j, and thus the probability to choose the alternative j, will be the same for all individuals i. Therefore, the log-likelihood is only function of the counts N_{jt} of individuals for each alternative.

In the aggregated case, it is needed to observe several choice situations t in order to estimate the model, that is to have a group variable. It could be a time variable, a space variable or an age group variable for example. In our illustrative application, choice situations will be months of observation. The corresponding log-likelihood (up to a constant) that has to be maximized is:

$$\log L = \sum_{t=1}^{T} \sum_{j=1}^{D} N_{jt} \log(s_{jt}) = \left[\sum_{t=1}^{T} \sum_{j=1}^{D} N_{jt}(X_{jt}b) \right] - \left[\sum_{t=1}^{T} N_t \log \left(\sum_{j=1}^{D} \exp(X_{jt}b) \right) \right]$$

with $X_{jt}b = \sum_{k=1}^{K} b_k X_{kjt}$.

Implementation in R: the package "mclogit" developped by Martin Elff [9] allows to fit conditional logit models with count data, using the Fisher-scoring/IWLS algorithm[2].

2.3 Market-share models

Market-share models were developed in the 80's, mainly by Cooper and Nakanishi [7]. The aim is to model market-shares of D brands using their marketing factors (price, advertising) as explanatory variables, with aggregated data (not individual-level but market-level data).

[2]For details on IWLS algorithm, see for example Green [13].

2.3.1 Attraction models

Market-share models are similar to an aggregated version of the Multinomial Logit model (MNL). The concept of "attractivity" of a brand is central in this literature, and is comparable to the "utility" concept in discrete choice models. The specification of the attractivity of brand j is an expression of the explanatory variables describing brand j. The market-share of brand j is defined as its relative attractivity compared to competitors, i.e. as its attractivity divided by the sum of attractivities of all the brands of the market.

$$0 < \ S_{jt} = \frac{\mathcal{A}_{jt}}{\sum_{l=1}^{D} \mathcal{A}_{lt}} \ < 1$$

with $\mathcal{A}_{jt}$ the attractivity of firm j at observation t such that $\mathcal{A}_{jt} > 0$.

Two main market-share models exist. One is referred to as MNL in this literature (we call it "MNL-type" to distinguish from MNL of section 2.2) because of the similarity between the expected share in equation (3) and the observed share expressed below. The other one is called MCI (Multiplicative Competitive Interaction model). The only difference between them is the functional form of explanatory variables: the log-linearized MCI takes the $\log(X)$ as explanatory, whereas the log-linearized MNL-type takes directly the X. In the subsequent expressions of the models, the attractivities will involve a multiplicative random part.

MNL-type model:

$$\mathcal{A}_{jt} = \exp(a_j + \sum_{k=1}^{K} b_k X_{kjt} + \varepsilon_{jt})$$

$$S_{jt} = \frac{\mathcal{A}_{jt}}{\sum_{l=1}^{D} \mathcal{A}_{lt}} = \frac{\exp(a_j + \sum_{k=1}^{K} b_k X_{kjt} + \varepsilon_{jt})}{\sum_{l=1}^{D} \exp(a_l + \sum_{k=1}^{K} b_k X_{klt} + \varepsilon_{lt})}$$

where $\exp(\varepsilon_{jt})$ is the multiplicative random component.

MCI model:

$$\mathcal{A}_{jt} = \exp(a_j) \prod_{k=1}^{K} X_{kjt}^{b_k} \nu_{jt}$$

$$S_{jt} = \frac{\mathcal{A}_{jt}}{\sum_{l=1}^{D} \mathcal{A}_{lt}} = \frac{\exp(a_j + \sum_{k=1}^{K} b_k \log X_{kjt} + \varepsilon_{jt})}{\sum_{l=1}^{D} \exp(a_l + \sum_{k=1}^{K} b_k \log X_{klt} + \varepsilon_{lt})}$$

where $\nu_j = \exp(\varepsilon_{jt})$ is the multiplicative random component.

With this specification X should be quantitative and strictly positive (marketing actions) in order to respect $\mathcal{A}_{jt} > 0$.

General MCI attraction model (GMCI): introduced by Cooper and Nakanishi, it combines both types of explanatory variables.

$$\mathcal{A}_{jt} = \exp(a_j + \varepsilon_{jt}) \prod_{k=1}^{K} f_k(X_{kjt})^{b_k} \quad \text{and} \quad S_{jt} = \frac{\mathcal{A}_{jt}}{\sum_{l=1}^{D} \mathcal{A}_{lt}}$$

where $\exp(\varepsilon_{jt})$ is the multiplicative random component and with f_k a monotonic transformation of X_k such that $f_k(.) > 0$. If all f_k are the identity function (resp. the exponential function), it corresponds to the MCI model (resp. the MNL-type model).

2.3.2 Estimation by OLS or GLS

The estimation method proposed by Nakanishi and Cooper [30] relies on a log linearization that they call "log-centering transformation" which is actually the log ratio between a share S_{jt} and the geometric mean of all shares at observation t, $\widetilde{\mathbf{S}}_t$, also called CLR (centered log-ratio) transformation in the CODA literature.

MNL-type model

$$\log\left(\frac{S_{jt}}{\widetilde{\mathbf{S}}_t}\right) = a_1 + \sum_{l=2}^{D}(a_j - a_1)d_l + \sum_{k=1}^{K} b_k(X_{kjt} - \overline{\mathbf{X}}_{kt}) + (\varepsilon_{jt} - \overline{\varepsilon}_t)$$

MCI model

$$\log\left(\frac{S_{jt}}{\widetilde{\mathbf{S}}_t}\right) = a_1 + \sum_{l=2}^{D}(a_j - a_1)d_l + \sum_{k=1}^{K} b_k \log\left(\frac{X_{kjt}}{\widetilde{\mathbf{X}}_{kt}}\right) + (\varepsilon_{jt} - \overline{\varepsilon}_t)$$

where $d_l = 1$ if $l = j$, 0 otherwise (brand dummy). $\overline{\mathbf{S}}_t$ is the arithmetic mean of S_{jt}.

It is suggested to use a GLS estimation instead of an OLS estimation due to the potential heteroscedasticity and/or correlation of error terms (if observations are time periods for example). But as Cooper and Nakanishi [7] said, we found that the GLS procedure, which is quite heavy in terms of implementation for this kind of models, does not give empirically better results than the OLS procedure.

Implementation in R: the function lm() allows to fit a linear model on the log-centered model by ordinary least square.

2.4 Dirichlet covariate models

The Dirichlet distribution is the distribution of a composition obtained as the closure of a vector of D independent gamma-distributed variables with the same scale parameter. Thus, it is a distribution adapted for variables lying in the simplex. Campbell and Mosimann [5] developed Dirichlet covariate models to explain a compositional dependent variable, supposed to be Dirichlet distributed, by classical (non-Dirichlet) covariates.

2.4.1 Dirichlet distribution

Let $\mathbf{S} = (S_1, \ldots, S_D) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_D)$ where $S_j > 0$ and $\sum_{j=1}^{D} S_j = 1$, $\alpha_j > 0$ and $\sum_{j=1}^{D} \alpha_j = \alpha_0$. α_0 is called the precision parameter (when this value increases, the concentration around the expected value increases, the variance and covariance decrease). The density function is defined by:

$$f(\mathbf{S}) = \left(\frac{\Gamma(\alpha_0)}{\prod_{j=1}^{D} \Gamma(\alpha_j)} \right) \prod_{j=1}^{D} S_j^{\alpha_j - 1}$$

with Γ the Euler Gamma function.

Moreover,

$$\mathbb{E}(S_j) = \frac{\alpha_j}{\alpha_0} \quad ; \quad Var(S_j) = \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0 + 1)} \quad ; \quad Cov(S_j, S_l) = -\frac{\alpha_j \alpha_l}{\alpha_0^2(\alpha_0 + 1)}$$

An alternative parametrization can be considered, using the parameters $\mu_j = \mathbb{E}(S_j)$ to account for the expected values of the shares, and $\phi = \alpha_0$ to account for the precision. The correspondence between this parametrization and the previous one is based on the fact that $\alpha_j = \mu_j \phi$ and $\alpha_0 = \phi$.

The *negative covariance structure* of the Dirichlet distribution was criticized by Aitchison [1] but actually Campbell and Mosimann [5] show that this is not true for Dirichlet covariate models, contrary to the simple Dirichlet distribution. Thus, each observation indexed by t follows a different Dirichlet distribution. The fact that the negative correlation happens between the shares of a same Dirichlet distribution does not imply that the vectors of shares coming from different Dirichlet distributions are negatively correlated. Indeed the formula of generalized covariance proves that:

$$Cov(X, Y) = \mathbb{E}[Cov(X, Y|Z)] + Cov[\mathbb{E}(X|Z), \mathbb{E}(Y|Z)]$$

Thus, if the covariance between the conditional expected values of two vectors of shares is positive and larger than the negative expected value of the conditional covariance between these two shares, then the unconditional covariance between the two shares can be positive[3].

In addition, Brehm et al. [4] show in a simulation study that the *strong independence between the initial gamma-distributed variables* (before closure) is not a problem: the Dirichlet covariate model successfully fits data with or without strong independence of variables before closure.

[3] The same argument can be used for the Multinomial model.

2.4.2 Dirichlet covariate model

Two parametrizations exist for the Dirichlet regression model. The first one is the *"common parametrization"* and the second one is the *"alternative parametrization"*[4].

Common parametrization Under the common parametrization, the parameters of the Dirichlet distribution, the α_j's, are allowed to depend on the explanatory variables X_k.

$$\log(\alpha_j) = a_j + \sum_{k=1}^{K} b_{kj} X_{kj}$$

Each component is allowed to have different explanatory variables (a different number of explanatory variables and/or explanatory variables which take different values for the different components), but for the sake of simplicity $\mathbf{X}$ denotes explanatory variables for all components. In this parametrization, $\alpha_0 = \sum_{j=1}^{D} \alpha_j$ can be interpreted as a precision parameter.

Alternative parametrization Under the alternative parametrization, the regression is defined by two equations:

$$\log\left(\frac{\mu_j}{1-\mu_j}\right) = a_j + \sum_{k=1}^{K} b_{kj} X_k \quad ; \quad \log(\phi) = \gamma_0 + \sum_{k=1}^{K} \gamma_k Z_k$$

However, the alternative parametrization does not allow to use different explanatory variables for each component, thus the common parametrization is preferred in our illustrative application and this is why we use it.

2.4.3 Estimation by maximum likelihood

As explained in Hijazi and Jernigan [16], "a different Dirichlet distribution is modeled for every value of the explanatory variables, resulting in a conditional Dirichlet distribution". The conditional distributions $\mathbf{S}_t|\mathbf{X}_t$ are mutually independent: $\mathbf{S}_t|\mathbf{X}_t \sim \mathcal{D}(\alpha_1(\mathbf{X}_t), \ldots, \alpha_D(\mathbf{X}_t))$ with unknown parameters.
Thus, the log-likelihood to maximize is:

$$\log L(\mathbf{S}|\alpha(\mathbf{X})) = \sum_{t=1}^{T} \left[\log\Gamma\left(\sum_{j=1}^{D} \alpha_j(\mathbf{X}_t)\right) - \sum_{j=1}^{D} \log\Gamma(\alpha_j(\mathbf{X}_t)) + \sum_{j=1}^{D} (\alpha_j(\mathbf{X}_t) - 1) \log S_{jt} \right]$$

Implementation in R: the package "DirichReg" created by Maier [24] allows to fit Dirichlet model based on the common or alternative parametrization, by maximum likelihood.

[4] See Hijazi and Jernigan [16].

2.5 Compositional models

Compositional data analysis was developed in the 80's by John Aitchison [1]. Since the 90's, a group of researchers is particularly active in the domain and has developed a large mathematical framework for this literature (see for example V. Pawlowsky-Glahn, J.J. Egozcue, J.A. Martin Fernandez, K. Hron).

First applications were made on geological data, with the objective to analyze the composition of a rock sample in terms of the relative presence of different chemical elements. More generally, CODA aims to analyze relative information between the components (parts) of a composition where the total of the components is not relevant or is not of interest.

Let us remind that a composition $\mathbf{S}$ is a vector of D shares S_j potentially coming from the closure of D positive numbers N_j and belonging to the simplex $\mathcal{S}^D$:

$$\mathbf{S} = (S_1, \ldots, S_D) = \mathcal{C}(N_1, \ldots, N_D) \in \mathcal{S}^D \quad \text{with} \quad S_j > 0 \quad \text{and} \quad \sum_{j=1}^{D} S_j = 1$$

2.5.1 The log-ratio transformation approach

Compositional data analysis is based on the log-ratio transformation of compositions in order to obtain coordinates which can be represented in a $\mathbb{R}^{D-1}$ Euclidean space. Then, classical methods suited for data in the Euclidean space can be used on coordinates.

Three main transformations are developed: the ALR (additive log-ratio), the CLR (centered log-ratio) and the ILR (isometric log-ratio) transformations, each of them having specific advantages. We will need to use the simplex inner product (associated to a norm and a distance) called the Aitchison inner product and given by

$$< S_1, S_2 >= \frac{1}{D} \sum_{j=1}^{D} \sum_{l=1}^{D} \log \frac{S_{1j}}{S_{1l}} \log \frac{S_{2j}}{S_{2l}}.$$

ALR Additive log-ratio transformation It is the first transformation proposed by Aitchison in 1986.

$$alr(\mathbf{S}) = \left(\log \frac{S_1}{S_D}, \ldots, \log \frac{S_{D-1}}{S_D} \right) = \mathbf{y} = (y^{(1)}, \ldots, y^{(D-1)})$$

Its inverse transformation is given by: $S_l = alr^{-1}(y^{(l)}) = \frac{\exp(y^{(l)})}{1+\sum_{l'=1}^{D-1} \exp(y^{(l')})}$ for $l = 1, \ldots, D-1$ and $S_D = \frac{1}{1+\sum_{l'=1}^{D-1} \exp(y^{(l')})}$.

ALR is isomorphic but not isometric from the simplex endowed with the Aitchison geometry to the Euclidean space $\mathbb{R}^{D-1}$ with the canonical inner product. This means that the Aitchison distance between two compositions is not equal to Euclidean distance between the corresponding points in $\mathbb{R}^{D-1}$.

CLR Centered log-ratio transformation

$$clr(\mathbf{S}) = \left(\log \frac{S_1}{\widetilde{\mathbf{S}}}, \ldots, \log \frac{S_D}{\widetilde{\mathbf{S}}}\right) = \mathbf{y} = (y^{(1)}, \ldots, y^{(D)})$$

with $\widetilde{\mathbf{S}}$ the geometric mean of the D components.
Its inverse transformation is given by: $\mathbf{S} = clr^{-1}(\mathbf{y}) = \mathcal{C}(\exp(y^{(1)}), \ldots, \exp(y^{(D)}))$, that is $S_l = clr^{-1}(y^{(l)}) = \frac{\exp(y^{(l)})}{\sum_{l'=1}^{D} \exp(y^{(l')})}$.

CLR is isometric but the covariance matrix of the CLR data is singular.

ILR Isometric log-ratio transformation It consists in a projection of components in an orthonormal basis of $\mathcal{S}^D$ in order to obtain $D-1$ orthonormal coordinates.

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_{D-1}\}$ be an arbitrary orthonormal basis in $\mathbb{R}^{D-1}$, then $\mathbf{e}_l = clr^{-1}(\mathbf{v}_l),\ l = 1, \ldots D-1$ represent an orthonormal basis in the simplex $\mathcal{S}^D$. Considering the $D \times (D-1)$ matrix $\mathbf{V}$ with columns $\mathbf{v}_l = clr(\mathbf{e}_l),\ l = 1, \ldots D-1$, ILR coordinates are defined as:

$$ilr(\mathbf{S}) = clr(\mathbf{S})\mathbf{V} = \log(\mathbf{S})\mathbf{V} = \mathbf{y} = (y^{(1)}, \ldots, y^{(D-1)})$$

Its inverse transformation is given by: $\mathbf{S} = ilr^{-1}(\mathbf{y}) = \mathcal{C}(\exp(\mathbf{y}\mathbf{V}^T))$.
ILR is isometric with full rank covariance matrix, but the interpretability of coordinates can be lost if the chosen ILR transformation does not have meaningful log-ratios[5].

Example A particular ILR transformation that could be used is the following:

$$y^{(l)} = \sqrt{\frac{D-l}{D-l+1}} \log \frac{S_l}{\left(\prod_{l'=l+1}^{D} S_{l'}\right)^{\frac{1}{D-l}}}, \quad l = 1, \ldots, D-1$$

$y^{(1)}$ contains all the relative information of part S_1 to the parts $S_2, \ldots, S_D$.
If $D = 3$ for example, it leads to $y^{(1)} = \sqrt{\frac{2}{3}} \log \frac{S_1}{\sqrt{S_2 S_3}} = \sqrt{\frac{2}{3}} \log S_1 - \frac{1}{\sqrt{6}}(\log S_2 + \log S_3)$ and $y^{(2)} = \sqrt{\frac{1}{2}} \log \frac{S_2}{S_3} = \frac{1}{\sqrt{2}}(\log S_2 - \log S_3)$.
Thus, $\mathbf{V} = \begin{bmatrix} \sqrt{\frac{2}{3}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$.

2.5.2 CODA regression models

CODA regroups different tools for analyzing compositions: graphical tools (ternary diagrams, biplots), analytic tools (compositional PCA) and compositional regression models.

[5] The Sequential Binary Partition is a way to obtain meaningful log-ratios. See Pawlowsky-Glahn and Buccianti [32].

Compositional regression models are of different types depending on whether the response variable and/or the explanatory variables are compositional. In our application case, the response variable is a composition (market-shares of D brands) and the explanatory variables describe each brand (e.g. media investments for each brand). The corresponding CODA regression model considers explanatory variables as compositions too (e.g. market-shares are explained by relative media investments).

CODA models can be expressed either in terms of the initial compositional observations in the simplex or alternatively in terms of the corresponding transformations: the coordinates in Euclidean space. The second presentation has the advantage to look like a classical linear model but its connection with the original data is obscured by the transformation. On the other side, the first presentation in terms of the original share data is obscured by the simplex operations involved in the model equation.

Linear CODA model in the simplex (expressed in compositions):

$$\mathbf{S}_t = \mathbf{a} \bigoplus_{k=1}^{K} \mathbf{Z}_{\mathbf{k}t} \boxdot \mathbf{B}_{\mathbf{k}} \oplus \boldsymbol{\varepsilon}_t \tag{4}$$

with $\mathbf{S}, \mathbf{a}, \mathbf{Z}_{\mathbf{k}}, \boldsymbol{\varepsilon} \in \mathcal{S}^{D}$[6] and $\mathbf{B}_{\mathbf{k}} \in \mathbb{R}_{D\times D}$ such that row and column sums are equal to zero[7].

$\oplus$ is the *perturbation operation*, equivalent to the addition operation in the simplex:

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1y_1, \ldots, x_Dy_D) \qquad \text{with } \mathbf{x}, \mathbf{y} \in \mathcal{S}^D$$

$\odot$ is the *power transformation*, equivalent to the multiplication operation in the simplex:

$$\lambda \odot \mathbf{x} = \mathcal{C}(x_1^{\lambda}, \ldots, x_D^{\lambda}) \qquad \text{with } \lambda \in \mathbb{R}, \mathbf{x} \in \mathcal{S}^D$$

$\boxdot$ is the *compositional matrix product*, equivalent to the matrix product in the simplex:

$$\mathbf{x} \boxdot \mathbf{B} = \mathcal{C}\big(\textstyle\prod_{j=1}^{D} x_j^{b_{j1}}, \ldots, \prod_{j=1}^{D} x_j^{b_{jD}}\big) \qquad \text{with } \mathbf{B} \in \mathbb{R}_{D\times D}, \mathbf{x} \in \mathcal{S}^D$$

Linear CODA model in the Euclidean space (expressed in ilr coordinates):

$$ilr(\mathbf{S}_t)^{(l)} = y^{(l)} = \alpha^{(l)} + \sum_{k=1}^{K} \sum_{l'=1}^{D-1} \beta_{kl'}^{(l)} \times ilr(\mathbf{Z}_{kt})^{(l')} + \epsilon_t^{(l)} \quad \forall\, l \in 1, \ldots, D-1$$

$\epsilon^{(l)} \sim \mathcal{N}(0, \sigma^2)$; l: index of $\mathbf{S}$'s ILR coordinates ; l': index of $\mathbf{Z}$'s ILR coordinates.

[6] Note that the dependent composition and the explanatory compositions can be of different dimensions: $\mathbf{S} \in \mathcal{S}^D$, $\mathbf{Z}_{\mathbf{k}} \in \mathcal{S}^d$, resulting in a matrix of parameters $\mathbf{B}_{\mathbf{k}} \in \mathbb{R}_{D\times d}$

[7] Under these conditions, $\mathbf{B} \boxdot \mathbf{Z}$ is a linear transformation with respect to Aitchison geometry and an endomorphism on the simplex $\mathcal{S}^D$ (See Kynclova et al. [22]). Thus the model (4) is a linear model in the simplex.

2.5.3 Estimation of CODA model

After log-ratio transformation, the estimation is usually made with the OLS method, separately on the $D-1$ CODA models expressed in coordinates[8].

Then, the estimated model can be expressed in the simplex using the inverse transformation, to transform $\boldsymbol{\alpha}$ into $\mathbf{a}$, $\boldsymbol{\beta}$ into $\mathbf{b}$, $ilr(\mathbf{S})$ into $\mathbf{S}$ and $ilr(\mathbf{Z})$ into $\mathbf{Z}$:

$$\begin{aligned}
\widehat{\mathbf{a}} &= ilr^{-1}(\widehat{\alpha}^{(1)}, \ldots, \widehat{\alpha}^{(D-1)}) = \mathcal{C}(\exp(\widehat{\boldsymbol{\alpha}}\mathbf{V}^T)) \\
\widehat{\mathbf{B}}_{D\times D} &= \mathbf{V}_{D\times(D-1)} \cdot \widehat{\boldsymbol{\beta}}_{(D-1)\times(D-1)} \cdot \mathbf{V}^T_{(D-1)\times D} \\
\widehat{\mathbf{S}} &= ilr^{-1}(\widehat{y}^{(1)}, \ldots, \widehat{y}^{(D-1)}) = \mathcal{C}(\exp(\widehat{\boldsymbol{y}}\mathbf{V}^T))
\end{aligned}$$

with $\boldsymbol{\beta} = [\beta_{l',l}] = \begin{bmatrix} \beta_1^{(1)} & \ldots & \beta_1^{(D-1)} \\ \ldots & \beta_{l'}^{(l)} & \ldots \\ \beta_{D-1}^{(1)} & \ldots & \beta_{D-1}^{(D-1)} \end{bmatrix}$, and $\mathbf{B} = [b_{m',m}] = \begin{bmatrix} b_{1,1} & \ldots & b_{1,D} \\ \ldots & b_{m',m} & \ldots \\ b_{D,1} & \ldots & b_{D,D} \end{bmatrix}$ where $b_{m',m}$ is the parameter corresponding to the impact of $Z_{m'}$ on S_m.

Implementation in R: the packages “compositions” [41] and “robCompositions” [40] allow to transform compositional data, to fit the compositional model by OLS on the coordinates and to back transform the results in compositions. Van den Boogaart and Tolosana-Delgado [42] wrote a book for analysing compositional data with R.

2.6 Alternative models

In the literature, some articles mix compositional data analysis and aggregated choice models. In Bechtel [2] and Fry and Chong [11], the shares are specified according to a nested multinomial logit model which does not embody the IIA property (see section 3.2). They use an additive log-ratio transformation of their model (ALR) as can be found in the compositional analysis, in order to be able to estimate the model by OLS or GLS.

Some authors propose to transform compositional data to directional data by the square root transformation mapping the simplex into the unit hypersphere. Wang et al. [44] further use classical regression models for the polar coordinates, whereas Scealy and Welsh [36] use the additive Kent regression model.

3 Theoretical comparison of share models

This section aims to highlight the similarities and differences of the four presented models from a theoretical perspective. Because these models are deeply linked with the type of applications they have been proposed for, the following comparison refers not only to statistical

[8] The orthonormality of coordinates allows us to estimate the $D-1$ models separately.

properties, but also to econometric and marketing properties.

First of all, Table 2 summarizes the distributional assumptions, the estimation methods, the properties and the complexity of each model. These items are subsequently discussed in detail. Then, we derive the expressions of direct and cross elasticities for the four models, and we highlight the fact that GMCI can be expressed in a CODA way. Finally, the problem of zero observations is addressed.

Name	Model	Distribution	Estimation	Properties	Nb param.
MNL GLM type	$\mathbb{E}(S_{jt}\|X_t,W_t) = \frac{\exp(a_j+\sum_{k=1}^{K_X} b_k X_{kjt}+\sum_{\kappa=1}^{K_W} b_{\kappa j} W_{\kappa t})}{\sum_{l=1}^{D}\exp(a_l+\sum_{k=1}^{K_X} b_k X_{klt}+\sum_{\kappa=1}^{K_W} b_{\kappa l} W_{\kappa t})}$ with $a_1 = 0$ for identifiability reasons.	$(N_{1t},\dots N_{Dt}) \sim \mathcal{MN}(N_t, s_{1t},\dots,s_{Dt})$ Indep. distributed over t	Maximum Likelihood	Permutation invariant, Scale invariant, Random utility model, IIA	$(D-1)(1+K_W)+K_X$
GMCI TRM type	Share: $S_{jt} = \frac{\exp(a_j+\sum_{k=1}^{K_X} b_k X_{kjt}+\sum_{\kappa=1}^{K_W} b_{\kappa j} W_{\kappa t}+\varepsilon_{jt})}{\sum_{l=1}^{D}\exp(a_l+\sum_{k=1}^{K_X} b_k X_{klt}+\sum_{\kappa=1}^{K_W} b_{\kappa l} W_{\kappa t}+\varepsilon_{lt})}$ Equivalently in terms of CLR coordinate: $\log\left(\frac{S_{jt}}{\tilde{S}_t}\right) = a_1 + \sum_{j'=2}^{D} a'_{j'} d_{j'} + \sum_{k=1}^{K_X} b_k (X_{kjt} - \overline{X}_{kt}) + \sum_{\kappa=1}^{K_W}\left(b_{\kappa 1} W_{\kappa t} + \sum_{j'=2}^{D} b'_{\kappa j'} W_{\kappa t} d'_j\right) + (\epsilon_{jt} - \overline{\epsilon}_t)$	$clr(S_{jt}) \sim \mathcal{N}(\mu_{jt},\sigma^2)$ Indep. distributed over t	OLS on coordinates	Permutation invariant, Scale invariant, Perturbation invariant, IIA	$D(1+K_W)+K_X$
DIR GLM type	With the common parametrization: $\mathbb{E}(S_{jt}\|X_t,W_t) = \frac{\exp(a_j+\sum_{k=1}^{K_X} b_k X_{kjt}+\sum_{\kappa=1}^{K_W} b_{\kappa j} W_{\kappa t})}{\sum_{l=1}^{D}\exp(a_l+\sum_{k=1}^{K_X} b_k X_{klt}+\sum_{\kappa=1}^{K_W} b_{\kappa l} W_{\kappa t})}$ $\log \alpha_{jt} = a_j + \sum_{k=1}^{K_X} b_k X_{kjt} + \sum_{\kappa=1}^{K_W} b_{\kappa j} W_{\kappa t}$	$(S_{1t},\dots S_{Dt}) \sim \mathcal{D}(\alpha_{1t},\dots,\alpha_{Dt})$ Indep. distributed over t	Maximum likelihood	Permutation invariant, Scale invariant, IIA	$(1+K_X+K_W)\times D$
CODA TRM type	Composition in the simplex: $\mathbf{S}_t = \mathbf{a} \bigoplus_{k=1}^{K_X} \mathbf{X_{kt}} \boxdot \mathbf{B_k} \bigoplus_{\kappa=1}^{K_W} W_{\kappa t} \odot \mathbf{b}_\kappa \oplus \varepsilon_t$ Equivalently in terms of share in the simplex: $S_{jt} = \frac{a_j \cdot \prod_{l=1}^{D}\prod_{k=1}^{K_X} X_{klt}^{b_{klj}} \cdot \prod_{\kappa=1}^{K_W} b_{\kappa j}^{W_{\kappa t}} \cdot \varepsilon_{jt}}{\sum_{m=1}^{D} a_m \cdot \prod_{l=1}^{D}\prod_{k=1}^{K_X} X_{klt}^{b_{klm}} \cdot \prod_{\kappa=1}^{K_W} b_{\kappa m}^{W_{\kappa t}} \cdot \varepsilon_{mt}}$ Equivalently in terms of ILR coordinates: $ilrS_t^{(l)} = \alpha^{(l)} + \sum_{k=1}^{K_X}\sum_{l'=1}^{D-1} \beta_{kl'}^{(l)} ilrX_{kt}^{(l')} + \sum_{\kappa=1}^{K_W} \beta_\kappa^{(l)} W_{\kappa t} + \epsilon_t^{(l)}$	$\mathbf{S}_t \sim \mathcal{N}_{\mathcal{S}}(\mu_t, \mathbf{\Sigma})$ with $\mathcal{N}_{\mathcal{S}}$ the normal distribution on the simplex, μ a mean vector, $\mathbf{\Sigma}$ a variance matrix. $ilr(\mathbf{S}_t) \sim \mathcal{N}_{D-1}(\mu_t,\sigma^2)$ with $\mathcal{N}_{D-1}$ the multivariate normal distribution.	OLS on coordinates	Sub-compositional coherence, Permutation invariant, Scale invariant, Perturbation invariant	$((D-1)\times K_X + K_W + 1) \times(D-1)$

Table 2: Benchmark of models for explaining shares
(GLM: Generalized linear model ; TRM: Transformation model)

3.1 Distributional assumptions

The distribution of the dependent variable is different across models as one can see in Table 2. Indeed, in the MNL model the dependent variable is a vector of positive numbers N_j which follow a multinomial distribution. In the other three models the dependent variable is directly the vector of shares S_j which are Dirichlet distributed in the case of DIR and Gaussian in the simplex distributed for GMCI and CODA (the coordinates are Gaussian in the transformed space). Note that the MNL model of section 2.2 differs from the MNL-type model of section 2.3 by its underlying distributional assumptions.

MNL and Dirichlet models belong to the family of GLM (Generalized linear models): see Peyhardi et al. [34] for MNL and Maier [24] for Dirichlet. GMCI and CODA models belong to the family of transformation models (TRM) in which a classical linear model is postulated in the transformed space.

The "intercept only" model: If the models are defined with only intercepts as explanatory variables, the fitted shares are not the same across the four models. CODA and GMCI models yield the center of the compositional data, that is the closed vector of geometric means of each component, while MNL and DIR models yield the arithmetic means of components (weighted in the case of MNL). However, the geometric mean which is coherent with the simplex geometry is more adapted than the arithmetic mean to summarize shares data, as shown in Figure 1. This is an argument in favor of CODA and GMCI models.

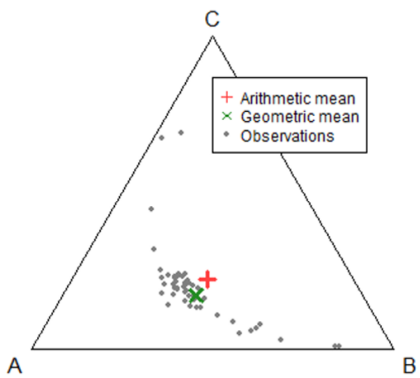

Figure 1: Arithmetic and geometric means of a compositional data set in a ternary diagram

3.2 Properties

We discuss here whether the properties that are clearly established for a given model are valid for the other models.

Independence of irrelevant alternatives and subcompositional coherence
In the econometric literature, an important question which is often discussed is whether or not a model satisfies the IIA (Independence from Irrelevant Alternatives) property. IIA is

an axiom of decision theory which means that the ratio of shares of an alternative j with respect to an alternative l only depends on characteristics of j and l and is not affected by the presence or absence of irrelevant alternatives. Actually, there is no difference whether the third alternative is irrelevant or not: even if the third alternative can indeed change the choice situation, it will not impact in the model the ratio of shares of the first two alternatives. This property allows to simplify the models but it is not always realistic (see the famous red bus - blue bus example of McFadden [27]). Without cross-effects, MNL, GMCI and Dirichlet models satisfy IIA.

In the CODA literature, the subcompositional coherence property (see Pawlowsky-Glahn [33]) means that the results of an analysis made on a subcomposition (i.e. remove some alternatives) should not contradict the results of the analysis made on the whole composition. This is coming from the fact that the Compositional Data Analysis (CODA) is based on the use of log-ratios. However, if we look at the forthcoming equation (6), we can see that the market-share of brand j is determined by the explanatory variables of all the brands.

In the econometrics literature, it is considered that IIA can be a severe limitation of MNL models and a lot of models (for example nested logit, GEV) have been developed in the framework of individual and agregated choice models to overcome this limitation (Bechtel [2], Fry et Chong [11], Hossain [17], Koppelman [19]).

Scale Invariance Scale invariance is mentioned in the CODA literature as a desirable property and CODA regression satisfies it (Pawlowsky-Glahn [33]). Scale invariance means that if the count data is multiplied by a constant, it does not affect the estimation results. It is known that scale-invariant functions of a composition are necessarily functions of log-ratios $\log(x_j/x_l)$ (reference) and reversely. All models that we have described satisfy this property.

Permutation invariance Permutation invariance corresponds to invariance through a permutation of the components of a composition and is a desirable property. It is clearly satisfied by all the described models.

Perturbation invariance Perturbation invariance corresponds to coherence when performing a change of units possibly different for each component of a composition: coherence corresponds here to the fact that the inverse perturbation of the results obtained on the perturbed data should correspond to the results obtained on the original data. For example, we can model brand market-shares in terms of sales volumes or in terms of sales values (that is sales volumes perturbed by the vector of prices). The estimated market-shares and parameters from the “volume” model should be equal to those of the “value” model after perturbation by the vector of prices. This property is satisfied by CODA regression and GMCI (as it uses the CLR transformation which is coherent with CODA principles). We can show empirically that it is not satisfied by MNL and Dirichlet.

3.3 Attraction and cross effects

GMCI models are presented using the notion of attraction (or attractivity). This presentation is somehow comparable to the utility presentation of random utility models in individual choice models of Section 2.2: what corresponds to attraction in the MNL model would be the numerator of (2), that is the exponential of the utility function.

We show here that a similar attraction presentation can be done for Dirichlet and CODA, even when the application background does not lead to an intuitive notion of "attraction".

Attraction form of the Dirichlet model (under the common parametrization)

$$\mathbb{E}(S_{jt}|X_t) = \frac{\alpha_{jt}}{\sum_{l=1}^{D}\alpha_{lt}} = \frac{\exp(a_j + \sum_{k=1}^{K} b_{kj}X_{kjt})}{\sum_{l=1}^{D}\exp(a_l + \sum_{k=1}^{K} b_{kl}X_{klt})} = \frac{\mathcal{A}_{jt}}{\sum_{l=1}^{D}\mathcal{A}_{lt}}$$

$$\Leftrightarrow \mathcal{A}_{jt}^{DIR} = \exp\left(a_j + \sum_{k=1}^{K} b_{kj}X_{kjt}\right) \quad (5)$$

Attraction form of the CODA model First, we express the expression of the market-share of brand j in the CODA model[9]:

$$\mathbf{S}_t = \mathbf{a}_t \bigoplus_{k=1}^{K} \mathbf{Z_k}_t \boxdot \mathbf{B_k} \oplus \boldsymbol{\varepsilon}_t = \mathcal{C}\left(a_1 \cdot \prod_{k=1}^{K}\prod_{l=1}^{D} X_{klt}^{b_{kl1}} \cdot \varepsilon_{1t}, \ldots, a_D \cdot \prod_{k=1}^{K}\prod_{l=1}^{D} X_{klt}^{b_{klD}} \cdot \varepsilon_{Dt}\right)$$

Thus,

$$S_{jt} = \frac{a_j \cdot \prod_{k=1}^{K}\prod_{l=1}^{D} X_{klt}^{b_{klj}} \cdot \varepsilon_{jt}}{\sum_{m=1}^{D} a_m \cdot \prod_{k=1}^{K}\prod_{l=1}^{D} X_{klt}^{b_{klm}} \cdot \varepsilon_{mt}} = \frac{\mathcal{A}_{jt}}{\sum_{m=1}^{D}\mathcal{A}_{mt}} \quad (6)$$

$$= \frac{\exp\left(\log(a_j) + \sum_{k=1}^{K}\sum_{l=1}^{D} b_{klj}\log(X_{klt}) + \log(\varepsilon_{jt})\right)}{\sum_{m=1}^{D}\exp\left(\log(a_m) + \sum_{k=1}^{K}\sum_{l=1}^{D} b_{klm}\log(X_{klt}) + \log(\varepsilon_{mt})\right)} \quad (7)$$

$$\Leftrightarrow \mathcal{A}_{jt}^{CODA} = a_j \cdot \prod_{k=1}^{K}\prod_{l=1}^{D} X_{klt}^{b_{klj}} \cdot \varepsilon_{jt} = \exp\left(\log(a_j) + \sum_{k=1}^{K}\sum_{l=1}^{D} b_{klj}\log(X_{klt}) + \log(\varepsilon_{jt})\right) \quad (8)$$

Note that taking the composition of X as explanatory variable in (6) actually corresponds to take the $\log(X)$ in the CODA attraction model under the exponential form (7).

Cross effects & number of parameters The dependence of the attractivity on other alternative characteristics corresponds to the existence or not of cross effects. Note

[9] Here the market-share S_{jt} is expressed as a function of X_{klt} directly and not as a function of Z_{klt} because S_{jt} is obtained by a closure operation (dividing by the denominator), thus it can be shown that the explanatory variables can be used in "volumes" as they are closed at the end.

that in the usual MNL, GMCI and Dirichlet models, the attractivity A_{jt} is a function of the explanatory variables characterizing alternative j only. In the CODA model, the attractivity depends automatically on all alternative characteristics. This is why CODA is the most complex model with the higher number of parameters.

It is not possible to estimate all the cross effects in the MNL model (see So and Kuhfeld [38]). Cross effects can be estimated in the GMCI model (see Cooper and Nakanishi [7]) and in the Dirichlet models but the number of parameters dramatically increases. CODA is relatively parsimonious in the sense that it allows to estimate all cross effects with a relatively lower number of parameters than the other models ($(D-1)\times(D-1)$ versus $D\times D$ for others), thanks to the constraints on the B matrix of parameters.

It is interesting to see that using the same dependent and explanatory variables, the complexity is totally different from a model to another. For example (as in our application, see Section 4), if the number of components (shares) of the dependent variable is $D=3$, explained by $K_X=7$ compositions of size $D=3$ and $K_W=1$ time-dependent variable, the numbers of estimated parameters are the following: 11 for MNL, 13 for GMCI, 27 for DIR and 32 for CODA. With 32 parameters, the CODA model reflects all the cross-effects between shares whereas the DIR and the GMCI models would require 69 parameters ($D(1+D\times K_X+K_W)$). Note also that the number of parameters increases dramatically with the number of components (brands), especially in the CODA model. For example if D becomes equal to 5 (with K_X and K_W fixed), the numbers of parameters become 15, 17, 45, and 120. This is a serious limitation for the CODA model.

3.4 Elasticities

In econometrics, explanatory variable impact is often measured through elasticities. The elasticity of a dependent variable Y to an explanatory variable X is the infinitesimal impact on Y, in percentage of Y, of an infinitesimal change of X, in percentage of X. Mathematically, it can be seen as the logarithmic derivative of Y with respect to the logarithm of X. The elasticity of the market-share of brand j to the explanatory variable k of brand j is:

$$e_{S_j}^{X_{kj}} = \frac{\partial S_j/S_j}{\partial X_{kj}/X_{kj}} = \frac{\partial \log S_j}{\partial \log X_{kj}}$$

Elasticities in MNL, GMCI and DIR models Expressions for elasticities are developed for the MNL and GMCI models in the literature (Cooper and Nakanishi [7]).

The log-linearized MCI, which takes the log of X as explanatory, leads to decreasing elasticity of market-share S_j with respect to X_{kj}, and is particularly well adapted for price in our example whereas log-linearized MNL-type takes directly the X and leads to increasing elasticity until a certain saturation level and decreasing elasticity after that level, particularly well adapted for advertising effects (see Cooper and Nakanishi [7]).

The expressions of elasticities in the GMCI model and in the MNL-type model are respectively $e_{S_j}^{X_{kj}} = b_k(1-S_j)$ and $e_{S_j}^{X_{kj}} = b_k(1-S_j)X_{kj}$. We claim that, due to the similar attraction presentation of the Dirichlet model (5) and the MNL model, the elasticities in these models can be derived from the ones of the MNL-type model.

Elasticities in the CODA model In the CODA literature people usually interpret parameters estimates in the transformed model and it turns out to be complex: Hron et al. [18] choose particular transformations so as to simplify this interpretation. We show that it is possible to interpret the CODA parameters in terms of elasticities. Actually, the CODA model has the same shape than the fully extended GMCI model (with all cross-effects, see Cooper and Nakanishi [7]), the only difference relies on the assumptions made on the data and the estimation of the parameters (constraints in the CODA framework) which allow to take into account the relative information of the explanatory variables. Thus, the elasticity in CODA can be derived from equation (6):

$$\mathbb{E}_{\mathcal{S}} S_{jt} = \frac{a_j \cdot \prod_{k=1}^{K} \prod_{l=1}^{D} X_{klt}^{b_{klj}}}{\sum_{m=1}^{D} a_m \cdot \prod_{k=1}^{K} \prod_{l=1}^{D} X_{klt}^{b_{klm}}} := \frac{NUM}{DEN}$$

where $\mathbb{E}_{\mathcal{S}} = ilr^{-1}(\mathbb{E}(ilr(X))$ is the expectation in the simplex.

We can derive from the previous equation the elasticity of the market-share of brand j to the explanatory variable k of brand l,X_{kl}, making the distinction between two cases:

- If the composition of X_k is used as an explanatory variable:

$$e_{\mathbb{E}_{\mathcal{S}} S_j}^{X_{kl}} = \frac{\partial \log \mathbb{E}_{\mathcal{S}} S_j}{\partial \log X_{kl}} = b_{klj} - \frac{\sum_{m=1}^{D} b_{klm} \times NUM}{DEN} = b_{klj} - \sum_{m=1}^{D} b_{klm} \mathbb{E}_{\mathcal{S}} S_m$$

- If the composition of $\log(X_k)$ is used as an explanatory variable:

$$e_{\mathbb{E}_{\mathcal{S}} S_j}^{X_{kl}} = \frac{\partial \log \mathbb{E}_{\mathcal{S}} S_j}{\partial \log X_{kl}} = \left(b_{klj} - \sum_{m=1}^{D} b_{klm} \mathbb{E}_{\mathcal{S}} S_m \right) / \log(X_{kl})$$

The elasticity of S_j to X_{kl} depends on the parameter corresponding to the effect of X_{kl} on S_j (b_{klj}), but also on parameters corresponding to the effects of X_{kl} on the other market-shares (b_{klm}) and on all market-shares S_m, $\forall\, m = 1, \ldots, D$.

Table 3 summarizes direct and cross estimated elasticities for the four considered models. A distinction is made whether the concerned explanatory variable is used directly as X or as $\log X$ (but in both cases, the elasticity is computed with respect to X). Note that the elasticity depends on S and/or X, which vary among observations $t = 1, \ldots, T$. Thus we can be interested in the elasticity for a precise observation (e.g. at time period t) or in the average elasticity over all observations.

3.5 Compositional form of GMCI

The use of a log-ratio transformation for estimating purpose is similar to the GMCI approach but it does not correspond exactly to the CODA model that we present in this paper.

Model	Direct elasticities	Cross elasticities
MNL, GMCI, DIR for non-logged var. (X)	$\widehat{e}_{S_j}^{X_{kj}} = \widehat{b}_k(1 - S_j)X_{kj}$	$\widehat{e}_{S_j}^{X_{kl}} = -\widehat{b}_k S_l X_{kl}$
MNL, GMCI, DIR for logged var. ($\log X$)	$\widehat{e}_{S_j}^{X_{kj}} = \widehat{b}_k(1 - S_j)$	$\widehat{e}_{S_j}^{X_{kl}} = -\widehat{b}_k S_l$
CODA for non-logged var. (X)	$\widehat{e}_{S_j}^{X_{kj}} = \widehat{b}_{kjj} - \sum_{m=1}^{D} S_m \widehat{b}_{kjm}$	$\widehat{e}_{S_j}^{X_{kl}} = \widehat{b}_{klj} - \sum_{m=1}^{D} S_m \widehat{b}_{klm}$
CODA for logged var. ($\log X$)	$\widehat{e}_{S_j}^{X_{kj}} = \frac{\widehat{b}_{kjj} - \sum_{m=1}^{D} S_m \widehat{b}_{kjm}}{\log(X_{kj})}$	$\widehat{e}_{S_j}^{X_{kl}} = \frac{\widehat{b}_{klj} - \sum_{m=1}^{D} S_m \widehat{b}_{klm}}{\log(X_{kl})}$

In this table, S are the observed shares.

Table 3: Direct and cross estimated elasticities

Wang et al. [43] propose a CODA regression model for the case when both dependent and explanatory variables are compositional which is simpler than the one we present: instead of having a matrix of parameters for each compositional explanatory variable, the model has a unique real parameter for all components of the explanatory composition. This model does not include cross effects between components contrary to our CODA model.

Actually Wang et al.'s model is exactly similar to the MCI model proposed by Cooper and Nakanishi in 1988 [7], except that Wang et al. use ILR coordinates while CLR coordinates are used in the MCI model.

From this correspondence we derive a compositional form for the GMCI model:

$$\mathbf{S}_t = \mathbf{a} \bigoplus_{k=1}^{K} b_k \odot \mathbf{Z_k}_t \oplus \boldsymbol{\varepsilon}_t \tag{9}$$

$$\Leftrightarrow S_{jt} = \frac{a_j \cdot \prod_{k=1}^{K} X_{kjt}^{b_k} \cdot \varepsilon_{jt}}{\sum_{l=1}^{D} a_l \cdot \prod_{k=1}^{K} X_{klt}^{b_k} \cdot \varepsilon_{lt}} = \frac{\exp(\log a_j + \sum_{k=1}^{K} b_k \log X_{kjt} + \log \varepsilon_{jt})}{\sum_{l=1}^{D} \exp(\log a_l + \sum_{k=1}^{K} b_k \log X_{klt} + \log \varepsilon_{lt})}$$

Equation (9) highlights the similarities and differences between GMCI and CODA models: in place of the B_k matrix in Equation (4) of the CODA model, we now have a single b_k parameter in the GMCI model.

3.6 Treatment of zero observations

Zeros are often an issue with share data. For GMCI, Dirichlet and CODA models, zeros cannot be tolerated because of the presence of the log transformation of shares in the likelihood. Many solutions to this problem have been considered, depending on the nature of zeros. Among the main ones, let us mention amalgamation of components (Pawlowsky-Glahn et al. [33]), ratio-preserving zero replacement (Martin-Fernadez et al. [25]) and conditional modelling for the CODA literature. Several transformations have been proposed for this problem, for example (Smithson and Verkuilen [37]) in the Dirichlet model . Wang et al. [43] and Scealy and Welsh [36] use a square root transformation together with models on the hypersphere. Fry et al. [10] compare their performance for the case of economic micro-data.

4 Empirical comparison of share models

In the above sections, we have presented four models adapted to model shares dependent variables which are based on different assumptions. In this section we propose a way to determine which one of these four models is the most adapted for a particular case study. After presenting the application and the data of our illustrative example, a cross-validation process is proposed in order to compute quality measures adapted for shares models on the four types of models. Finally, we compare the interpretation of the parameters of the four models in terms of elasticities.

4.1 Application and data

The main objective of this application is to understand the impact of media investments (relative or absolute values) on market-shares (the response variable is compositional) controlling for other factors like price and scrapping incentive. In each model specification the interest is on the marginal impact of each canal of media on relative sales, that is on the elasticities of market-shares to media investments by channel.

We focus here on the B segment[10] of the automobile market, which represents half of the sales in France in terms of volume. More precisely, following the subcompositional coherence property of CODA, we focus on 3 brands of this segment: Renault, Nissan and Dacia ($D = 3$).

The studied period runs from June 2005 to August 2015. This period is characterized by the birth of Dacia on the French automobile market, a low-cost brand belonging to Renault, at the beginning of 2005. It is also characterized by the economic crisis which has hurt the French automobile market a lot from 2008 to 2012 (at least). The French government tried to help this market setting up a scrapping incentive[11] which has "artificially" boosted the sales during 2009 and 2010. Note that Dacia increased a lot its market-share during the crisis thanks to its low price. These facts have to be kept in mind in order to understand the evolution of market-shares, and it justifies the use of a scrapping incentive dummy as control variable.

The four methods are applied to an automobile market data set coming from Renault containing for each brand of the B segment the sales volume in units N_{jt}, the catalog price in euros P_{jt}, the media investments by canal in euros M_{cjt} (TV, press, radio, outdoor, digital, cinema), and the periods of scrapping incentive I_t (dummy variable), monthly from June 2005 to August 2015 ($T = 123$ periods of observation).

The ternary diagram allows to represent compositions of 3 components in the simplex. Figure 2 represents for example the annual market-shares of Dacia, Nissan and Renault from 2005 to 2014. We can see easily that Dacia increases its market-share easily at the expense

[10] Segments of the automobile market are determined according to the size of the chassis. Segment B corresponds to small mainstream vehicles like the Renault Clio which is the most famous of this segment in France.

[11] A scrapping incentive is an incentive given by a government to promote the replacement of old vehicles with modern vehicles.

of Renault from 2005 to 2010.

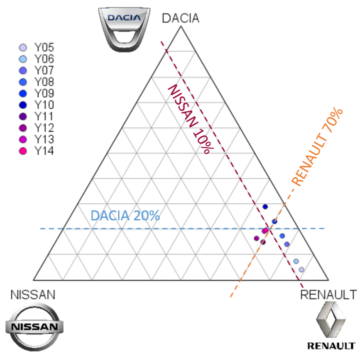

Figure 2: Ternary diagram of annual market-shares of Dacia, Nissan and Renault

According to the marketing literature, it is preferable to use the *logarithm of price* instead of the raw price[12]. Indeed, for our four models, using the log of price instead of the price gives best in-sample fits. The media investments have to be considered with a lag with respect to sales. Statistically in this application, a lag of 4 months gave the best results on the 4 models[13]. When media investments at time t are equal to zero, we replace them by 1 euro, which is a very small amount compare to the non-zero investments.

Table 4 summarizes the four models which are fitted in order to model brands market-shares S_{jt} as a function of media investments of brands M_{cjt} $(C = 6)$, the log of the average brand price $\log P_{jt}$, and the scrapping incentive dummy I_t (during years 2009 and 2010).

4.2 A cross-validation comparison

A cross-validation process is used to compute out-of-sample goodness-of-fit measures (presented in the next subsection) on the four considered models, in order to avoid overfitting effect due to the fact that the considered models do not have the same number of parameters.

1. Randomly draw a sub-sample of 100 observations among 123, resulting in 81% (100) in-sample observations and 19% (23) out-of-sample observations
2. Fit the 4 models to the sub-sample, store the fitted parameters
3. Apply the 4 models to the out-of-sample observations, store the fitted values of the shares
4. Compute the quality measures using the out-of-sample predicted share values

[12]The reason of that is linked to the shape of the elasticity of market-shares to the price (see 3.4). Moreover, to keep the market shares equal, the logged variables have to increase in the same proportion while the non-logged variables have to increase by the same amount

[13]Later on, we consider using an adstock function, which is a cumulative function of actual and past investments.

Name	Model
MNL	Estimation by ML using the N_{jt} $$\mathbb{E}(S_{jt}\|M_t, P_t, I_t) = \frac{\exp(a_j + \sum_{c=1}^{C} b_c M_{cjt} + b_P \log P_{jt} + b_{Ij} I_t)}{\sum_{l=1}^{D} \exp(a_l + \sum_{c=1}^{C} b_c M_{clt} + b_P \log P_{lt} + b_{Il} I_t)}$$ with $a_1 = b_{I1} = 0$ for identifiability reasons
GMCI	Estimation by OLS on the CLR coordinates $$S_{jt} = \frac{\exp(a_j + \sum_{c=1}^{C} b_c M_{cjt} + b_P \log P_{jt} + b_{Ij} I_t + \varepsilon_{jt})}{\sum_{l=1}^{D} \exp(a_l + \sum_{c=1}^{C} b_c M_{clt} + b_P log P_{lt} + b_{Il} I_t + \varepsilon_{lt})}$$ CLR coordinates: $$\log\left(\frac{S_{jt}}{\overline{S_t}}\right) = a_1 + \sum_{j'=2}^{D} a'_{j'} d_{j'} + \sum_{c=1}^{C} b_c (M_{cjt} - \overline{M_{ct}})$$ $$+ b_P \log\left(\frac{P_{jt}}{\overline{P_t}}\right) + b_{I1} I_t + \sum_{l=2}^{D} b'_{Il} I_t d_l + (\epsilon_{jt} - \overline{\epsilon_t})$$
DIR	Estimation by ML using the S_{jt} (common parametrization) $$\mathbb{E}(S_{jt}\|M_t, P_t, I_t) = \frac{\exp(a_j + \sum_{c=1}^{C} b_{cj} M_{cjt} + b_{Pj} \log P_{jt} + b_{Ij} I_t)}{\sum_{l=1}^{D} \exp(a_l + \sum_{c=1}^{C} b_{cl} M_{clt} + b_{Pl} \log P_{lt} + b_{Il} I_t)}$$ $$\log \alpha_{jt} = a_j + \sum_{c=1}^{C} b_{cj} M_{cjt} + b_{Pj} \log P_{jt} + b_{Ij} I_t$$
CODA	Estimation by OLS on the $(D-1)$ ILR coordinates separately $$\mathbf{S}_t = \mathbf{a} \bigoplus_{c=1}^{C} \mathbf{M_{ct}} \boxdot \mathbf{B_c} \oplus \log \mathbf{P}_t \boxdot \mathbf{b_P} \oplus I_t \odot \mathbf{b}_I \oplus \varepsilon_t$$ $$\Leftrightarrow S_{jt} = \frac{a_j \cdot \prod_{l=1}^{D} \prod_{c=1}^{C} M_{cjt}^{b_{clj}} \cdot \log P_{jt}^{b_{Plj}} \cdot b_{Ij}^{I_t} \cdot \varepsilon_{jt}}{\sum_{m=1}^{D} a_m \cdot \prod_{l=1}^{D} \prod_{c=1}^{C} M_{cmt}^{b_{clm}} \cdot \log P_{mt}^{b_{Plm}} \cdot b_{Im}^{I_t} \cdot \varepsilon_{mt}}$$ ILR coordinates: $$ilr(S_t)^{(l)} = \alpha^{(l)} + \sum_{c=1}^{C} \sum_{l'=1}^{D-1} \beta_{cl'}^{(l)} ilr(M_{ct})^{(l')}$$ $$+ \sum_{l'=1}^{D-1} \beta_{Pl'}^{(l)} ilr(\log P_t)^{(l')} + \beta_I^{(l)} I_t + \epsilon_t^{(l)}$$

Table 4: Fitted models to explain brand market-shares

5. Iterate 100 times steps 1 to 4

6. Compute the average quality measures using the out-of-sample predicted share values over the 100 iterations

N.B.: Here we want to have an efficient model all along the studied period, the aim is not to have a good predictive model for the future. Moreover the presented models are not taking into account the potential auto-correlation of error terms. That is the reason why the cross-validation is made on randomly drawn samples and not on a split of the studied period according to time.

4.3 Quality measures

The out-of-sample accuracy of the 4 models is compared according to a list of different indicators adapted to shares that we found in the literature. Two categories of measures are detailed: the R^2-type measures which are based on the notion of explained variability, and the distance-type measures which evaluate how far are the fitted values from the true values.

R^2 based on total variability (R2T) The CODA literature proposes a R^2 directly adapted to compositional data (see Hijazi [15], Monti et al. [28]). It uses the measure of the total variability of a set of compositions, based on the variance of log-ratios. In terms of interpretation, it is similar to the classical R^2: it measures the proportion of the total variation explained by the model.

$$R_T^2 = \frac{totvar(\widehat{\mathbf{S}})}{totvar(\mathbf{S})}$$

with $totvar(\mathbf{S}) = \frac{1}{2D}\sum_{j=1}^{D}\sum_{l=1}^{D} var(\log \frac{S_j}{S_l})$. This measure is always positive but is not guaranteed to be lower than 1. Note that for the “intercept only” model, R_T^2 equals zero for all models because there is no variability in $\widehat{\mathbf{S}}$.

R^2 based on Aitchison distance (R2A) Another R^2 measure can be found in the CODA literature, based on the Aitchison distance between the observed compositions and the fitted compositions on one hand, and on the Aitchison distance between the observed compositions and the center of the data (closed geometric means of components) on the other hand (see Hijazi [15], Monti et al. [28]).

$$R_A^2 = 1 - \frac{CSSE}{CSST}$$

with $CSST = \sum_{t=1}^{T} d_A^2(\mathbf{S_t}, \mathbf{g})$; $CSSE = \sum_{t=1}^{T} d_A^2(\mathbf{S_t}, \widehat{\mathbf{S}}_\mathbf{t})$. $\mathbf{g}$ is the closed vector of geometric means of each component over observations t, and

$$d_A(\mathbf{S_t}, \widehat{\mathbf{S}}_\mathbf{t}) = \sqrt{\sum_{j=1}^{D} \left(\log \frac{S_{jt}}{g(S_j)} - \log \frac{\widehat{S}_{jt}}{g(\widehat{S}_j)} \right)^2} = \sqrt{\frac{1}{D} \sum_{j=1}^{D} \sum_{l>j}^{D} \left(\log \frac{S_{jt}}{S_{lt}} - \log \frac{\widehat{S}_{jt}}{\widehat{S}_{lt}} \right)^2}$$

However, this R^2 measure can be misleading because it has a large variability and it can take negative values. Note that for the "intercept only" model, R_A^2 equals zero for CODA and GMCI models because $\widehat{\mathbf{S}} = g(\mathbf{S})$.

Kullback-Leibler divergence (KL) The Kullback-Leibler divergence is used as a goodness-of-fit measure or as a prediction accuracy measure (see Haaf et al. [14]). It is a sum of the log-ratios between the observed values and the fitted values of the shares, weighted by the observed value. The log-ratio allows to take into account the relative error, and the weight emphasizes the importance of large errors in large shares.

$$KL(\mathbf{S}, \widehat{\mathbf{S}}) = \sum_{t=1}^{T} \sum_{j=1}^{D} \log \left(\frac{S_{jt}}{\widehat{S}_{jt}} \right) S_{jt}$$

A compositional version of this measure is defined as follows (Martin-Fernandez et al. [26], Palarea et al. [31]):

$$KLC(\mathbf{S}, \widehat{\mathbf{S}}) = \frac{D}{2} \left(KL(\mathbf{0_D}, \mathbf{S} \ominus \widehat{\mathbf{S}}) + KL(\mathbf{0_D}, \widehat{\mathbf{S}} \ominus \mathbf{S}) \right) = \frac{D}{2} \sum_{t=1}^{T} \log \left(\overline{(S_t/\widehat{S}_t)} \cdot \overline{(\widehat{S}_t/S_t)} \right)$$

where $\mathbf{0_D} = (1/D, \ldots, 1/D)$ the compositional zero (center of the simplex $\mathcal{S}^D$), and $\overline{(S_t/\widehat{S}_t)}$ the arithmetic mean of shares ratios $\left(\frac{S_{1t}}{\widehat{S}_{1t}}, \cdots, \frac{S_{Dt}}{\widehat{S}_{Dt}} \right)$ for observation t.

The KLC measure is indeed well adapted to shares data because for the "intercept only" model, this measure of divergence is lower for models which predict the geometric means of the shares (CODA and GMCI models) than for models which predict the arithmetic means (MNL and DIR models), and it is well known that the geometric mean is more adapted to summarize compositional data than the arithmetic mean.

Other quality measures can be used for share data. See for example Kumar [21], Quagrainie [35], Leeflang and Reuyl [23], Naert and Weverbergh [29], Ghosh et al. [12].

Table 5 presents the out-of-sample average quality measures for our four models. For each measure, the best model is in bold. The out-of-sample average quality measures suggest that DIR seems to be the most adapted model to fit our data (27 parameters). However, according to the R^2 based on total variability (R2T), CODA (32 parameters) is better than the Dirichlet model. The GMCI model and the MNL model without cross-effects are almost systematically the worst models, certainly due to their simplicity and low number of parameters.

	MNL		GMCI		DIR		CODA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
R2T	0.425	0.164	0.462	0.179	0.622	0.224	**0.647**	0.227
R2A	0.196	0.270	0.155	0.325	**0.373**	0.235	0.084	0.433
KL	0.465	0.101	0.480	0.121	**0.392**	0.166	0.501	0.168
KLC	0.139	0.034	0.137	0.032	**0.117**	0.071	0.134	0.034
RMSE	0.059	0.008	0.061	0.009	**0.052**	0.008	0.063	0.011

Table 5: Out-of-sample quality measures

4.4 Interpretation of parameters

The fitted parameters of the presented models can be interpreted in terms of elasticities. As an example, the direct elasticities of market-shares of the 3 considered brands are computed for the TV channel, for the 123 observed periods, and the average is presented in the Table 6. It corresponds to the average relative impact on the market-share of brand j, S_j, when the investment in TV of brand j increases by 1%.

	MNL	GMCI	DIR	CODA
DACIA	0.0019	0.0028	-0.0068	-0.0046
NISSAN	**0.0101**	**0.0152**	**0.0389**	**-0.0022**
RENAULT	0.0058	0.0088	0.0145	-0.0038

Table 6: Average direct elasticities for TV investments

We observe that elasticities are not the same across models, and can even be of opposite sign. For example, the DIR model concludes that, on average over the period 2005-2015, if Nissan increases its TV investment by 10%, it will increase its market-share by 0.4%, whereas in CODA, it will have a slightly negative impact. The CODA model, which includes all cross effects, suggests that the impacts of TV investments of Dacia, Nissan and Renault tend to cancel each other. However, all models agree on the fact that Nissan has the highest TV's elasticities (in bold in the table).

5 Conclusion

Because of the constraints of shares data, classical regression models cannot be used directly. In this paper we present in detail four models adapted to model a composition of shares as dependent variable, using explanatory variables which are characteristics of each component or of observations: the aggregated multinomial logit model (MNL), the generalized multiplicative competitive interaction model (GMCI), the Dirichlet model (DIR) and the linear compositional regression model (CODA). These four models are coming from different literatures and application fields, and use different estimating methods. We express all models in attraction model form and derive the direct and cross elasticities. We also prove that GMCI can be written in a compositional way. In our case, we use this kind of models

to understand the impact of media investments by channel on brand market-shares in the automobile market, controlling for price and scrapping incentive.
We highlight the similarities and the differences of these models. The MNL model requires the volume data whereas the others only need the shares data. MNL and DIR are GLM estimated by maximum likelihood and centered on the arithmetic means, whereas GMCI and CODA are transformation models estimated by OLS, centered on the geometric means. MNL and GMCI models without cross-effects are very simple and parsimonious models but fail to capture the variability of the data in our application. The CODA model is the most complex model but it manages to estimate all cross effects with a relative parsimony, compared to other models thanks to constraints on parameters, resulting in a good fitting quality. The DIR model is very flexible and it successfully fits the data with less parameters than the CODA model. All these models are implemented in R, and can be interpreted in terms of elasticities.
Finally, we propose a method based cross-validation in order to determine the best model for a particular case study. We compare the models in terms of out-of-sample average quality measures adapted for shares data. In our application, the Dirichlet model gives the best out-of-sample results, followed by the CODA model.

We intend to focus in further work on the interpretability of the different models. More precisely, direct and cross elasticities have to be deeply interpreted in order to check that the models make sense for the considered application, and to be able to use them to help decision making in practice. Concerning our particular application, the observations are over time. Thus, the potential auto-correlation of error terms should be tested and taken into account if necessary. Moreover, as we measure the impact of media investments on market-shares, considering "adstock function" of media investments instead of punctual media investments might be more relevant. Adstock functions are often used in the marketing literature, they are cumulative value of past and present advertising expenditures, corresponding to the "carry-over effect" over time. Furthermore, the introduction of random coefficients can be discussed. Such models are considered by Berry, Levinsohn and Pakes [3] in the aggregated MNL framework in econometrics. It seems easy to implement for GMCI and CODA models, but we did not find it in the Dirichlet framework.

Acknowledgment: We would like to thank BVA and the Marketing Direction of Renault, especially Philippe Devailly, Matthias Guegan and Herve Tranger, for sharing valuable data with us, and for their support during the elaboration of the data base and the model specification. We would also like to thank Vera Pawlowsky-Glahn, Marina Vives-Mestres, Juan Jose Egozcue, Josep Antoni Martin Fernandez and Karel Hron for their help to get familiar with the compositional data analysis.

Funding: This work was supported by the market research agency BVA and the ANRT.

References

[1] Aitchison, J. *The statistical analysis of compositional data*. Monographs on statistics and applied probability. Chapman and Hall, 1986.

[2] Bechtel, G. G. Share-ratio estimation of the nested multinomial logit model. *Journal of Marketing Research 27*, 2 (1990), pp. 232–237.

[3] Berry, S., Levinsohn, J., and Pakes, A. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* (1995), 841–890.

[4] Brehm, J., Gates, S., and Gomez, B. A monte carlo comparison of methods for compositional data analysis. In *1998 annual meeting of the Society for Political Methodology* (1998).

[5] Campbell, G., and Mosimann, J. E. Multivariate analysis of size and shape: modelling with the dirichlet distribution. In *ASA Proceedings of Section on Statistical Graphics* (1987), pp. 93–101.

[6] Chakir, R., Laurent, T., Ruiz-Gazen, A., Thomas-Agnan, C., and Vignes, C. Spatial scale in land use models: application to the teruti-lucas survey. *Spatial Statistics* (2016).

[7] Cooper, L., and Nakanishi, M. *Market-Share Analysis: Evaluating Competitive Marketing Effectiveness*. International Series in Quantitative Marketing. Springer, 1988.

[8] Elff, M. Social divisions, party positions, and electoral behaviour. *Electoral Studies 28*, 2 (2009), 297–308.

[9] Elff, M. *mclogit: Mixed Conditional Logit*, 2014. R package version 0.3-1.

[10] Fry, J. M., Fry, T., and McLaren, K. Compositional data analysis and zeros in micro data. Centre of policy studies/impact centre working papers, Victoria University, Centre of Policy Studies/IMPACT Centre, 1996.

[11] Fry, T. R. L., and Chong, D. A tale of two logit, compositional data analysis and zero observations.

[12] Ghosh, A., Neslin, S., and Shoemaker, R. A comparison of market share models and estimation procedures. *Journal of Marketing Research 21*, 2 (1984), pp. 202–210.

[13] Green, P. J. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)* (1984), 149–192.

[14] Haaf, C. G., Michalek, J., Morrow, W. R., and Liu, Y. Sensitivity of vehicle market share predictions to discrete choice model specification. *Journal of Mechanical Design 136* (DECEMBER 2014).

[15] Hijazi, R. H. Residuals and diagnostics in dirichlet regression. *ASA Proceedings of the General Methodology Section* (2006), 1190–1196.

[16] Hijazi, R. H., and Jernigan, R. W. Modelling compositional data using dirichlet regression models. *Journal of Applied Probability and Statistics 4*, 1 (2009), 77–91.

[17] Hossain, M. Z. A review on some alternative specifications of the logit model. *Journal of Business & Economics Research (JBER) 7*, 12 (2011).

[18] Hron, K., Filzmoser, P., and Thompson, K. Linear regression with compositional explanatory variables. *Journal of Applied Statistics 39*, 5 (2012), 1115–1128.

[19] Koppelman, F. S. *Closed Form Discrete Choice Models.* 2007, ch. 13, pp. 257–277.

[20] Koppelman, F. S., and Bhat, C. *A self instructing course in mode choice modeling: multinomial and nested logit models.* FTA, 2006.

[21] Kumar, V. Forecasting with market response models forecasting performance of market share models: an assessment, additional insights, and guidelines. *International Journal of Forecasting 10*, 2 (1994), 295 – 312.

[22] Kynclova, P., Filzmoser, P., and Hron, K. Modeling compositional time series with vector autoregressive models. *Journal of Forecasting 34*, 4 (2015), 303–314.

[23] Leeflang, P. S. H., and Reuyl, J. C. On the predictive power of market share attraction models. *Journal of Marketing Research 21*, 2 (1984), pp. 211–215.

[24] Maier, M. J. Dirichletreg: Dirichlet regression for compositional data in r. Research Report Series/Department of Statistics and Mathematics 125, WU Vienna University of Economics and Business, Vienna, January 2014.

[25] Martin-Fernandez, J., Barcelo-Vidal, C., and Pawlowsky-Glahn, V. Zero replacement in compositional data sets. In *Data Analysis, Classification, and Related Methods.* Springer, 2000, pp. 155–160.

[26] Martin-Fernandez, J. A., Bren, M., Barcelo-Vidal, C., and Pawlowsky-Glahn, V. A measure of difference for compositional data based on measures of divergence. In *Proceedings of IAMG* (1999), vol. 99, pp. 211–216.

[27] McFadden, D. L. Econometric analysis of qualitative response models. *Handbook of econometrics 2* (1984), 1395–1457.

[28] Monti, G., Mateu-Figueras, G., Pawlowsky-Glahn, V., and Egozcue, J. Shifted-dirichlet regression vs simplicial regression: a comparison. *Welcome to CoDawork 2015* (2015).

[29] Naert, P., and Weverbergh, M. On the prediction power of market share attraction models. *Journal of Marketing Research 18*, 2 (1981), pp. 146–153.

[30] Nakanishi, M., and Cooper, L. G. Simplified estimation procedures for mci models. *Marketing Science 1*, 3 (1982), pp. 314–322.

[31] Palarea-Albaladejo, J., Martín-Fernández, J. A., and Soto, J. A. Dealing with distances and transformations for fuzzy c-means clustering of compositional data. *Journal of classification 29*, 2 (2012), 144–169.

[32] Pawlowsky-Glahn, V., and Buccianti, A. *Compositional data analysis: Theory and applications.* John Wiley & Sons, 2011.

[33] Pawlowsky-Glahn, V., Egozcue, J. J., and Tolosana-Delgado, R. *Modeling and Analysis of Compositional Data.* John Wiley & Sons, 2015.

[34] Peyhardi, J., Trottier, C., and Guédon, Y. A new specification of generalized linear models for categorical data. *arXiv preprint arXiv:1404.7331* (2014).

[35] Quagrainie, K. K. Analysis of u.s. catfish fillet market share using a flexible logistic model. *Marine Resource Economics 21*, 1 (2006), pp. 33–45.

[36] Scealy, J., and Welsh, A. Regression for compositional data by using distributions defined on the hypersphere. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*, 3 (2011), 351–375.

[37] Smithson, M., and Verkuilen, J. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods 11*, 1 (2006), 54.

[38] So, Y., and Kuhfeld, W. F. Multinomial logit models. In *in SUGI 20 Conference Proceedings, Cary, NC: SAS Institute Inc* (1995).

[39] Solano-Acosta, W., and Dutta, P. K. Unexpected trend in the compositional maturity of second-cycle sand. *Sedimentary Geology 178*, 3 (2005), 275–283.

[40] Templ, M., Hron, K., and Filzmoser, P. *robCompositions: an R-package for robust statistical analysis of compositional data*, 2011.

[41] van den Boogaart, K. G., Tolosana, R., and Bren, M. *compositions: Compositional Data Analysis*, 2014. R package version 1.40-1.

[42] Van Den Boogaart, K. G., and Tolosana-Delgado, R. *Analysing Compositional Data with R*. Springer, 2013.

[43] Wang, H., Liu, Q., Mok, H. M., Fu, L., and Tse, W. M. A hyperspherical transformation forecasting model for compositional data. *European journal of operational research 179*, 2 (2007), 459–468.

[44] Wang, H., Shangguan, L., Wu, J., and Guan, R. Multiple linear regression modeling for compositional data. *Neurocomputing 122* (2013), 490–500.