WORKING PAPERS

N° 1455

Juillet 2023

“The link between multiplicative competitive interaction models and compositional data regression with a total”

Lukas Dargel and Christine Thomas-Agnan

Toulouse School of Economics

The link between multiplicative competitive interaction models and compositional data regression with a total

WORKING PAPER

Lukas Dargel [1,2] and Christine Thomas-Agnan [1]

[1]Toulouse School of Economics, University of Toulouse Capitole, Toulouse France
[2]BVA Group, France

July 20, 2023

Abstract

This article sheds light on the relationship between compositional data (CoDa) regression models and multiplicative competitive interaction (MCI) models, which are two approaches for modeling shares. We demonstrate that MCI models are special cases of CoDa models and that a reparameterization links both. Recognizing this relation offers mutual benefits for the CoDa and MCI literature, each with its own rich tradition. The CoDa tradition, with its rigorous mathematical foundation, provides additional theoretical guarantees and mathematical tools that we apply to improve the estimation of MCI models. Simultaneously, the MCI model emerged from almost a century-long tradition in marketing research that may enrich the CoDa literature. One aspect is the grounding of the MCI specification in intuitive assumptions on the behavior of individuals. From this basis, the MCI tradition also provides credible justifications for heteroskedastic error structures – an idea we develop further and that is relevant to many CoDa models beyond the marketing context. Additionally, MCI models have always been interpreted in terms of elasticities, a method only recently revealed in CoDa. Regarding this interpretation, the change from the MCI to the CoDa perspective leads to a decomposition of the influence of the explanatory variables into contributions from relative and absolute information. This decomposition also opens the door for testing hypothesis about the importance of each information type.

Keywords: Marketing, MCI, Compositional Data, Regression

1 Introduction

The MCI models of Cooper and Nakanishi (1989) are well-established regression techniques in marketing research, used extensively to analyze market shares. Their long history can be traced back to the law of retail gravitation of Reilly (1931) whose lasting influence is reviewed by Friske and Choi (2013). Reilly's original approach translates Newton's formula for gravitational attraction into a retail context. Huff (1964) apply this idea to model the attraction of stores on individuals living in different districts of a city and further convert these attractions to shares. Nakanishi and Cooper (1974) study a generalized version of Huff's model, which they name MCI model, and also show how its parameters can be estimated by ordinary least squares (OLS). In their textbook Cooper and Nakanishi (1989) introduce several extensions of the original MCI model, which they refer to as the simple effects MCI model. Its two extensions, the differential effects and fully extended MCI models relate the dependent shares in increasingly complex ways to the explanatory variables.

Although the CoDa literature also aims to analyze shares, it developed independently from the MCI approach. The origins of CoDa date back to Aitchison (1982), who noticed the importance of a unified mathematical framework for the analysis of share vectors. Since share vectors must be positive and usually sum to one, most classical statistical methods do not apply, but Aitchison (1982, 1986) popularized using log-ratios to overcome these constraints. Working with log-ratio transformations has since become a principle of all CoDa analysis and also underpins the CoDa regression approach that has been developed by Daunis-i-Estadella et al. (2002), Hron et al. (2012), Egozcue et al. (2012), Chen et al. (2017), Coenders et al. (2017), Chen et al. (2018) and Morais and Thomas-Agnan (2021), among others. The link between CoDa regression and MCI was uncovered by Morais (2017), who identify the simple effects MCI model as an instance of CoDa regression and showed that the attraction formulation of MCI models can also be used to represent CoDa models. While Wang et al. (2013) already use a CoDa model that can be classified as a simple effects MCI model, the authors do not emphasize this connection. In this article, we develop the equivalence between all MCI models and CoDa regression, which has the obvious benefit that all findings in each strand of the literature carry over to the other.

An asset of the CoDa approach is its insights into the geometric structure of log-ratio spaces and their relationship to the simplex, which is the sample space for share vectors. This includes well-understood distance measures and notions for the expectation and variance of simplex-valued variables. These notions allow, for example, to formalize an unbiasedness property for predictions that was already remarked by Boogaart et al. (2021). Additionally, it is common to use isometric log-ratio (ilr) transformations (Egozcue et al. 2003) in CoDa regression. This solves the issue of a singular covariance structure in MCI models and enables us to improve their estimation efficiency.

One intuitively appealing side of the MCI models is their development as extensions of the gravity formula, which can be more accessible than the abstract formalism usually employed in CoDa analysis. Cooper and Nakanishi (1989) also link the MCI models to the random utility framework of McFadden (1974), which can thus be used to derive a microeconomic foundation for the CoDa regression specification. The grounding of MCI models in individual behavior also justifies heteroskedastic error structures whenever the dependent shares are calculated by aggregating individual choices. While Nakanishi and Cooper (1974) already recognize this point, we will simplify their heteroskedasticity structure, making it more practical for MCI as well as general CoDa models.

A rare example of cross-fertilization between the two approaches is Morais and Thomas-Agnan (2021), who developed the interpretation of CoDa models in terms of elasticities that do not depend on any log-ratio transformation. This interpretation is well known for MCI models since Cooper and Nakanishi (1989) and, indeed, commonplace in economics and marketing. In

Dargel and Thomas-Agnan (2023), we further develop the use of elasticities differences, whose advantage is that they are constant over observations. In the present article, we show that the MCI models can be interpreted with the CoDa version of these elasticities, which corresponds to a decomposition of the explanatory variables' influence into contributions from relative and absolute information.

Before entering into the body of the article we want to highlight some differences in the vocabulary used in both strands of the literature. For CoDa regression it is commonplace to refer to observations as the statistical individual of the model and to components or parts for the elements of the dependent CoDa vector. In the MCI approach the statistical individual is a choice situation, which usually refers to specific geographic and temporal window, i.e. a city district in the year 2023. Instead of components the MCI approach refers to options among which consumer in a given choice situations can choose. These options are usually a set of products, brands, or stores, and the dependent CoDa vectors contain the market shares of each option in a all choice situation.

The following two sections briefly review the relevant MCI and CoDa literature. Section 4 demonstrates the link between the two approaches, covering the subjects of model specifications, error structures, estimation, and interpretation. Section 5 presents an illustrative application and a Monte Carlo study, and the final section concludes.

2 Review of MCI models

The use of interaction models in marketing research can be traced back to Reilly (1929, 1931). According to Reilly's law of retail gravitation, a population center in a region (j) attracts individuals living in another region (i) with force proportional to the population mass and negatively proportional to the square of the distance between the two regions

$$A_{ij} \propto \frac{\text{population}_j}{\text{distance}^2_{ij}}. \tag{1}$$

Reilly presents this formula as a rule of thumb to estimate the volume of retail customers that live in location i and chose region j as a destination for their shopping trip. Converse (1949) uses (1) to determine the points where the attraction of two possible shopping destinations is equal, which he called balance points. These balance points can be used to trace the borders of trading areas for multiple retail destinations $(j = 1, ..., D)$ on a map. Customers in location i would then be allocated to the trading area of shop j if its attraction is maximal i.e. $A_{ij} = \max(A_{i1}, \dots, A_{iD})$.

The next important developments are due to Huff (1962, 1963, 1964), who is motivated by the problem of defining trade areas for different stores within a city. He criticizes the trade areas obtained from the balance point method for being deterministic because the customers are supposed to choose the same retail locations as a destination for all their shopping trips. Huff models the attraction that a shop (j) exerts on individuals living in a city district (i) as a function of the distance between them and the store's area: $A_{ij} = \text{area}_j \cdot \text{distance}^b_{ij}$. Unlike his predecessors, he minimized a quadratic loss function over a grid search to estimate b instead of fixing it at -2. Huff further converts the attraction, measured in volumes of customers or revenues, into shares

$$S_{ij} = \frac{A_{ij}}{\sum_{l=1}^{D} A_{il}}. \tag{2}$$

Since these shares indicate for every location i how likely a it is that a customer chooses one of the potential retail destinations, they can be used to define probabilistic trade areas.

Haines et al. (1972) recognized that Huffs model is a direct application of Luce (1959) choice axiom, implying that it is subject to "independence of irrelevant alternatives" (IIA) assumption, whose limitations are famously discussed by McFadden (1974). For Huff's model IIA implies, for example, that the opening of a new store leaves the share ratios between the existing ones unchanged, irrespective of the new store's location. This is unrealistic because we expect that stores that are located closer to the new competitor would lose a higher proportion of their customers to it. An argument in favor of share models with a multiplicative attraction is given by Naert and Bultez (1973), who point out that they are "logically consistent" in the sense that they predict positive shares that sum to one.

Nakanishi and Cooper (1974) introduce the term MCI for an extension of Huff's model to K explanatory variables. However, their main contribution is the first consistent ordinary least squares (OLS) estimation procedure, which they further explain and simplify in Nakanishi and Cooper (1982). Their estimator is defined for a generalized attraction formula that includes a multiplicative log-normal error term

$$A_{ij} = \prod\nolimits_{k=1}^{K} \check{x}_{kij}^{h_k} \check{\epsilon}_{ij}, \tag{3}$$

where we use the check notation $\check{x}_{kij}, \check{\epsilon}_{ij} > 0$ to indicate non-negative variables or volumes. The errors are assumed to follow a log-normal distribution with independence across city districts (i), which are the statistical individuals in our geomarketing context. However, the errors are potentially correlated across the D stores that consumers can choose as a destination for their shopping trips. More formally, with $\check{\epsilon}_{i\bullet} = (\check{\epsilon}_{i1}, ..., \check{\epsilon}_{iD})$ being the vector of errors for one district and all stores, we have $\log(\check{\epsilon}_{i\bullet}) \sim \mathcal{N}(0, \Sigma)$ where Σ $(D \times D)$ is the correlation matrix across stores.

The key step to derive OLS estimates of the model parameters is the "log-centering" operation, which is identical to the clr (centered log-ratio, see Section 3) in the CoDa literature. To properly define this transformation, we first introduce the geometric mean of the components of a positive vector $\mathbf{v} = (v_1, ..., v_D)'$ as $g(\mathbf{v}) = (\prod_{j=1}^{D} v_j)^{1/D}$. The clr is then given by $\text{clr}(\mathbf{v}) = \log(\mathbf{v}/g(\mathbf{v}))$, and from now on we use the star notation $v_j^* = \log(v_j/g(\mathbf{v}))$ to access one element of a clr-transformed vector. Applying the clr to (2) yields $S_{ij}^* = A_{ij}^*$, and the transformed equation becomes linear in the parameters

$$S_{ij}^* = \sum\nolimits_{k=1}^{K} x_{kij}^* h_k + \epsilon_{ij}^*. \tag{4}$$

The errors of the clr-transformed model in (4) and those of the attraction (3) are linked by the equation $\epsilon_{i\bullet}^* = \mathbf{G}_D \log(\check{\epsilon}_{i\bullet})$, with $\mathbf{G}_D = \mathbf{I}_D - \frac{1}{D}\boldsymbol{\iota}_D\boldsymbol{\iota}_D'$ and $\boldsymbol{\iota}_D$ being a D-vector of ones. This implies a multivariate normal $\epsilon_{i\bullet}^* \sim \mathcal{N}(0, \mathbf{G}_D\Sigma\mathbf{G}_D)$ for the clr-errors. With normal errors centered at zero, the OLS estimator of the parameters in equation (4) is unbiased and consistent but not efficient since the errors are not homoskedastic. Additionally, even if Σ has full rank, the covariance matrix $\mathbf{G}_D\Sigma\mathbf{G}_D$ of $\epsilon_{i\bullet}^*$ is degenerate, which complicates the use of more efficient estimators. Although Nakanishi and Cooper (1974) have proposed some solutions to these problems, we will defer their discussion to Section 4.5 after introducing some CoDa tools that allow us to address these issues more efficiently.

After their breakthrough in the estimation methodology, Cooper and Nakanishi (1989) devote a whole textbook to market share models, exploring several generalizations of the MCI model. The first of these is the general MCI (GMCI) framework that relaxes the functional form of the explanatory variables. When $f_k(x) = x$, the original MCI model is recovered, while choosing $f_k(x) = \exp(x_k)$ yields a model whose mean expression coincides with that of the multinomial logit (MNL) model:

$$
\begin{aligned}
\text{GMCI: } A_{ij} &= \prod_{k=1}^{K} f_k(\breve{x}_{kij})^{h_k}\breve{\epsilon}_{ij} \\
\text{MCI: } A_{ij} &= \prod_{k=1}^{K} \breve{x}_{kij}^{h_k}\breve{\epsilon}_{ij} && \implies \forall k = 1, ..., K : f_k(x) = x \\
\text{MNL type: } A_{ij} &= \prod_{k=1}^{K} \exp(\breve{x}_{kij})^{h_k}\breve{\epsilon}_{ij} && \implies \forall k = 1, ..., K : f_k(x) = \exp(x).
\end{aligned} \tag{5}
$$

However, the original MNL model does not include an explicit error term, since it belongs to the family of generalized linear models. To ease the notation for the remainder of the article, we continue with non-transformed variables, but all results would remain valid for the GMCI version as long as the transformations are strictly positive $f_k(\breve{x}_{kij}) > 0$.

Following another direction of generalizations, Cooper and Nakanishi (1989) introduce two model variants with increasingly complex specifications. The original MCI model (3) is classified as simple effects MCI model because it supposes the effect of a variable on the attractions to be constant across the D stores. Hence, in this simple version, the model parameters h_k are only indexed by the variable number k. In the differential effects MCI model, the effects of a variable on the attraction can vary across the D stores, leading to one parameter h_{kj} for each variable k and brand j. The fully extended MCI model further adds cross-effects that capture the influence of the k^{th} characteristic $\breve{x}_{kij}$ of brand j on the attraction A_{il} of the other brands $l \neq j$. Including these cross-effects leads to parameters h_{kjl} with a triple index, where h_{kjj} reflect the direct effects and h_{kjl} for $l \neq j$ the cross-effects. By introducing these cross-effects, the fully extended MCI model also overcomes the restrictions of the IIA assumption. The attraction forms of the three model variants are given by

$$
\begin{aligned}
\text{simple effects MCI: } A_{ij} &= \prod_{k=1}^{K} \breve{x}_{kij}^{h_k}\breve{\epsilon}_{ij} \\
\text{differential effects MCI: } A_{ij} &= \breve{\alpha}_j \prod_{k=1}^{K} \breve{x}_{kij}^{h_{kj}}\breve{\epsilon}_{ij} \\
\text{fully extended MCI: } A_{ij} &= \breve{\alpha}_j \prod_{k=1}^{K} \prod_{l=1}^{D} \breve{x}_{kij}^{h_{kjl}}\breve{\epsilon}_{ij},
\end{aligned} \tag{6}
$$

where the brand-specific constants must be positive $\breve{\alpha}_j > 0$. Cooper and Nakanishi (1989) estimate the two largest MCI models with the explanatory variables transformed by logs instead of clr, which is in contrast to what they do in (4) for the simple MCI. They further caution against "model induced collinearity" in the extended versions unless certain constraints on the parameters are imposed. However, instead of restating their estimation approach at this point we will show in Section 4 that all the MCI models in (6) are special cases of a more general CoDa model, which enables us to use a unified estimation methodology.

Table 1: Elasticities and Cross-Elasticities in the MCI model

Model	**Elasticity** $\frac{\partial S_{ij}/S_{ij}}{\partial \breve{x}_{kij}/\breve{x}_{kij}}$	**Cross-Elasticity** $\frac{\partial S_{ij}/S_{ij}}{\partial \breve{x}_{kil}/\breve{x}_{kil}}, \quad l \neq j$
Simple effects	$h_k(1 - S_{ij})$	$-h_k S_{il}$
Differential effects	$h_{kj}(1 - S_{ij})$	$-h_{kl} S_{il}$
Fully extended	$h_{kjj} - \sum_{j=1}^{D} S_{ij} h_{kjl}$	$h_{kjl} - \sum_{j=1}^{D} S_{ij} h_{kjl}$

To interpret the MCI models (6) Cooper and Nakanishi (1989) suggest to use point elasticities $\frac{\partial S_{ij}/S_{ij}}{\partial \breve{x}_{kil}/\breve{x}_{kil}}$ that measure the relative change of the market share S_{ij} resulting from a relative change in the k^{th} variable $\breve{x}_{kil}$. When the component index of the share j is different from the component index of the variable l they refer to cross-elasticities. Table 1 presents the expression of these elasticities for the three versions of the MCI model.

Another noteworthy distinction between the three models in (6) concerns the data they require for estimations and predictions. To clarify these data requirements let us consider a hypothetical scenario involving two brands competing in market A, two other brands competing in market B, and an undocumented market C. In addition, there exists a fifth brand as a potential new participant in all markets. Table 2 outlines what information sets are available to the modeller of this scenario.

Table 2: Available information for marketing models

		existing brands				new brands
		1	2	3	4	5
existing markets	A	✓	✓			
	B			✓	✓	
new markets	C					

Checked cells (✓) indicate that information about the market share of a brand in a market is available and blank cells indicate the absence of information.

Depending on the modeling objectives some of the extensions of the MCI model become impractical. The fully extended model has the highest data requirements, as it only allows predictions for situations in which the same brand constellation has already been observed. In Table 2, for instance, the fully extended model could be used to predict a scenario where brands 1 and 2 simultaneously enter market C, but not when brand 2 and 3 would do so. It could also not predict what would happen when brand 3 enters market A, because the available information does not suffice to estimate the required cross-effect parameters. In such situations, the differential effects model is much more permissive, as it does not use these cross-effects. Thus, with the differential MCI model, we could predict the impact of the entry of any existing firm in any existing or new market. However, when the objective is to predict the entry of a new firm, we have no choice but resort to the constant parameter hypothesis of the simple MCI model.

3 Review of CoDa models

Compositional data are vectors, also called CoDa vectors, of non-negative components which are often imposed a constant sum constraint of one. The objective of this requirement is to select a unique representer in the class of vectors proportional to a given CoDa vector. Imposing a unit sum is, however, more of a convention than a binding requirement since the objective of CoDa analysis is to concentrate on the role of the relative values of its components, which are the same for any element of this class. The log-ratio approach to the statistical analysis of CoDa vectors was initiated by Aitchison (1982). It is based on using log-ratio transformations before applying classical statistical techniques to the transformed values. Due to the constraints of CoDa vectors, it is natural to work in the simplex of $\mathbb{R}^D$ which is defined, for an integer D, by

$$\mathcal{S}^D = \left\{ \mathbf{x} = (x_1, \ldots, x_D)' : x_j > 0, j = 1, ..., D; \sum_{j=1}^{D} x_j = 1 \right\}. \tag{7}$$

The vector of shares (or parts, or components) are often obtained by the closure of a vector $\check{\mathbf{x}} \in \mathbb{R}_+^D$, whose elements can be called volumes or abundances, where the closure $\mathcal{C}$ is defined by

$$\mathcal{C}(\check{\mathbf{x}}) = \left(\frac{\check{x}_1}{\sum_{j=1}^D \check{x}_j}, \cdots, \frac{\check{x}_D}{\sum_{j=1}^D \check{x}_j} \right). \tag{8}$$

Let us note that market share vectors are constructed in this way. For ease of notation we sometimes denote $\mathbf{x} = \mathcal{C}(\check{\mathbf{x}})$ the representer of a volume vector $\check{\mathbf{x}}$ in the unit simplex.

Aitchison (1986) equipped the simplex with a Hilbert space structure compatible with the objective of the CoDa analysis. He introduced two operations that play for the simplex the role of addition and scalar multiplication in $\mathbb{R}^D$:

1. the perturbation operation, denoted by $\oplus$, plays the role of the addition:

$$\text{for} \quad \mathbf{x}, \mathbf{y} \in \mathcal{S}^D, \mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, \ldots, x_D y_D),$$

2. the power operation, denoted by $\odot$, plays the role of the scalar multiplication:

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

Aitchison (1986) also defined the centered log-ratio (clr) and its inverse of a vector $\mathbf{x} \in \mathcal{S}^D$:

$$\text{clr}(\mathbf{x}) = \mathbf{G}_D \log \mathbf{x} \qquad \text{and} \qquad \text{clr}^{-1}(\mathbf{x}^*) = \mathcal{C}(\exp \mathbf{x}^*), \tag{9}$$

where the logarithm of $\mathbf{x}$ and the exponential of $\mathbf{x}^*$ are understood componentwise, and where $\mathbf{x}^* = \text{clr}(\mathbf{x})$. The clr is therefore a centered version of the vector of log-transformed components.

The inner product of the simplex, nowadays called Aitchison inner product, corresponds to

$$< \mathbf{x}, \mathbf{y} >_A = < \text{clr}(\mathbf{x}), \text{clr}(\mathbf{y}) >_E, \tag{10}$$

where the right hand side uses the standard inner product in $\mathbb{R}^D$. From the above definition, the clr transformation is therefore an isometry between the simplex equipped with the Aitchison inner product and the classical Euclidean space $\mathbb{R}^D$. More complex transformations, called ilr (for isometric log-ratio), are often used instead of clr, because of the singularity of the clr transform. Indeed, the sum of the components of a clr transformed CoDa vector is zero, so that the image of the simplex by clr is a hyperplane of $\mathbb{R}^D$. In contrast, the ilr transformations send the simplex into $\mathbb{R}^{D-1}$ and they are bijective. An ilr transformation is indexed by a so-called contrast matrix $\mathbf{V}$ of dimension $D \times (D-1)$ (Pawlowsky-Glahn et al. 2015, see). The contrast matrix associated to any given orthonormal basis $(\mathbf{e}_1, \cdots, \mathbf{e}_{D-1})$ of $\mathcal{S}^D$, orthonormality being understood with respect to the Aitchison inner product here, is given by $\mathbf{V} = [\text{clr}(\mathbf{e}_1), \cdots, \text{clr}(\mathbf{e}_{D-1})]$. The corresponding isometric log-ratio transformation is defined by:

$$\text{ilr}_V(\mathbf{x}) = \mathbf{V}' \log(\mathbf{x}).$$

The product of a $D \times D$ matrix $\mathbf{B}$ by a vector $\mathbf{x} \in \mathcal{S}^D$ can be defined by $\mathbf{B} \boxdot \mathbf{x} = \text{clr}^{-1}(\mathbf{B}\,\text{clr}(\mathbf{x}))$ for any matrix $\mathbf{B}$ belonging to the set $\mathcal{A}_D$ of $D \times D$ matrices with zero column sums and row sums. The clr and ilr transformation of a matrix $\mathbf{B}$ are defined in Ruiz-Gazen et al. (2023) by

$$\text{ilr}_V(\mathbf{B}) = \mathbf{V}^T \mathbf{B} \mathbf{V} \qquad \text{clr}(\mathbf{B}) = \mathbf{G}_D \mathbf{B} \mathbf{G}_D. \tag{11}$$

It is then easy to see that

$$\mathbf{B} \in \mathcal{A}_D \Leftrightarrow \mathbf{B} = \text{clr}(\mathbf{B}) \tag{12}$$

For a simplex valued random variable $\mathbf{X}$, we use the following definitions for expectation and variance:

$$\begin{aligned} \mathbb{E}^{\oplus}\mathbf{X} &:= \operatorname{clr}^{-1}(\mathbb{E}\operatorname{clr}(\mathbf{X})) && (13)\\ \mathbb{V}\mathrm{ar}^{\oplus}\,\mathbf{X} &:= \mathbb{V}\mathrm{ar}(\operatorname{clr}(\mathbf{X})). && (14) \end{aligned}$$

The normal distribution on the simplex which can be traced back to Aitchison and Shen (1980) is described in a more modern way in Pawlowsky-Glahn et al. (2015). $\mathcal{N}_{\mathcal{S}^D}(\mu, \mathbf{M})$ denotes the normal distribution of a simplex valued random variable $\mathbf{X}$ with $\mu = \mathbb{E}(\operatorname{clr}(\mathbf{X}))$ and $\mathbf{M} = \mathbb{V}\mathrm{ar}(\operatorname{clr}(\mathbf{X}))$.

The statistical analysis of CoDa vectors has first developed in the direction of descriptive and multivariate methods. Regression with CoDa vectors appeared in Daunis-i-Estadella et al. (2002). The case of compositional explanatory variables is discussed for example in Hron et al. (2012) and the case of compositional response for example in Egozcue et al. (2012). The case with CoDa vectors on both sides of the equation appears for example in Chen et al. (2017). Let us write the general equation for a CoDa regression model with a compositional dependent variable $S_{i\bullet} = (S_{i1}, \dots, S_{iD})$ and a mixture of classical and compositional explanatory variables using the simplex operations. Assuming that the compositional explanatory variables have the same dimension as the dependent, the simplex equation writes as follows

$$S_{i\bullet} = \alpha_\bullet \oplus (\bigoplus_{k=1}^{K_X} \mathbf{B}_k \boxdot x_{ki\bullet}) \oplus (\bigoplus_{k=1}^{K_Z} z_{ki} \odot \gamma_{k\bullet}) \oplus \epsilon_{i\bullet} \text{ with } \epsilon_{i\bullet} \sim \mathcal{N}_{\mathcal{S}^D}(\mathcal{C}(\boldsymbol{\iota}_D), \mathbf{M}), \qquad (15)$$

where:

- the parameter $\alpha_\bullet \in \mathcal{S}^D$ corresponds to the vector of component specific constants,
- the matrices of parameters $\mathbf{B}_k \in \mathcal{A}_D$, $k = 1, \dots, K_X$ are associated with compositional variables $x_{ki\bullet} \in \mathcal{S}^D$ observed on unit i,
- the variables $z_{ki} \in \mathbb{R}$ for $k = 1, \dots, K_Z$ are real-valued explanatory variables observed on unit i, associated with the parameters $\gamma_{k\bullet} \in \mathcal{S}^D$,
- the covariance matrix of clr coordinates $\mathbf{M} \in \mathcal{A}_D$ is valid for the normal in the simplex, which means that it is symmetric and that for any contrast $\mathbf{V}$ the quadratic form $\mathbf{V}'\mathbf{M}\mathbf{V}$ is positive definite.

A more recent concept of compositional data analysis removes the constant sum constraint. This approach was studied by Pawlowsky-Glahn et al. (2014), who named it CoDa with a total, and its application to CoDa regression can be found in Ferrer-Rosell et al. (2016), Coenders et al. (2017) and Morais and Thomas-Agnan (2021). The absolute information involved in a so-called total variable $T(\mathbf{X}) > 0$ associated to any positive explanatory vector variable $\mathbf{X}$ can be included as well on the right hand side of the regression equation among the classical variables. A valid total is any function that preserves the absolute information in a vector of volumes $\check{\mathbf{x}} \in \mathbb{R}_+^D$, but the most intuitive choices are the arithmetic total $\sum_{j=1}^D \check{x}_j$ and the geometric total $\prod_{j=1}^D \check{x}_j$, which we denote by $G(\check{\mathbf{x}})$.

4 The link between CoDa and MCI models

This section aims to establish the relationship between MCI and CoDa regression models with a total. Using this approach, it is possible to account for totals in compositional explanatory and response variables, and in principle, all these variables could have a different number of components. However, we only consider totals for the explanatory variables since MCI models are used for explaining market shares whose total is uninformative. We further restrict our attention to the case where the dimension of all compositional variables is equal to the number of components of the dependent variable. This requirement arises naturally when dealing with interaction data whose structure we discuss in the following.

Like all interaction models, MCI models are based on information at three different levels. In a geomarketing context, there are the characteristics of the different stores, which we refer to more generally as type C data (C for components). Additionally, there is type I data that relates to the statistical individual, which is typically a city district in geomarketing. Finally, type IC data refers to information on the pairs of individuals and components. Huff's model, for example, uses the area of a shop as type C information and the distance between a region and the shop as type IC information but does not include type I information. The omission of type I information is indeed a shortcoming of all MCI models in (6), and a benefit of the CoDa methods is the seamless integration of this information into a regression model. On the other hand, type C information is not used in general CoDa models due to collinearity problems. This issue does not occur in the simple effects MCI model. When translating interaction models with this data structure into the CoDa framework (15) we will treat type I information as real-valued variables and data of type IC and C as compositions. More precisely, they are D-compositions, indicating that, like the market shares, they have D components. In principle, the CoDa approach would allow us to handle compositions of dimensions $J \neq D$ on all information levels (type I, C, IC), but given the scope of this article, we do not require this level of generality.

From the previous discussion, it should be clear that the most general model we consider for this article is a CoDa model with a total of the explanatory compositional variables and with a restriction on their dimension. In the following subsections, we demonstrate that this model nests all versions of the MCI model as special cases. The precise nesting structure of all models we consider is illustrated in Figure 1, where the symbol $\Leftrightarrow$ indicates that two models are equivalent up to a reparametrization, and the symbol $\equiv$ when the parametrization is also identical. The four models on the left use a total and those on the right are classical CoDa models that do not include absolute information. Below the model names, we show the model complexity in terms of the number of parameters. The expression of this complexity indicator is obviously simplified by the fact that we only consider D-compositions. The constants K_I, K_{IC}, and K_C indicate the number of variables of each type, where $K_C \geq D$ would lead to collinearity problems in the simple effects MCI model.

The most general model uses a general total $T(\mathbf{X})$, but with a geometric total $G(\mathbf{X})$ the CoDa model becomes equivalent to a fully extended MCI, though with a non-standard parametrization. The differential and simple effects MCI models are special cases with additional restrictions on the parameters. In the differential effects MCI model, the parameter matrix is constrained to be diagonal. Using this constraint we derive two additional models, all called Diagonal CoDa models, but with different influences of the total. The simple effects MCI model, adds the additional constraint that the parameters for each variable are constant across components. All models that have a diagonal parameter matrix are subject to the IIA assumption.

The different models presented in Figure 1 can be taken as a guideline for the following subsections. We will first focus on adjusting the classical CoDa model, i.e. the CoDa model without a total, to interaction data. In the next subsections, we discuss the three MCI models For each of these models, we explore vectorized data representations, as it makes the relationship between them more explicit and allows us to treat their estimation in a unified framework. The problem of parameter estimation is discussed in a separate subsection, where we also touch upon alternative error structures that are relevant to all MCI models and CoDa models in general. In the last subsection, we discuss the model interpretation in terms of elasticities.

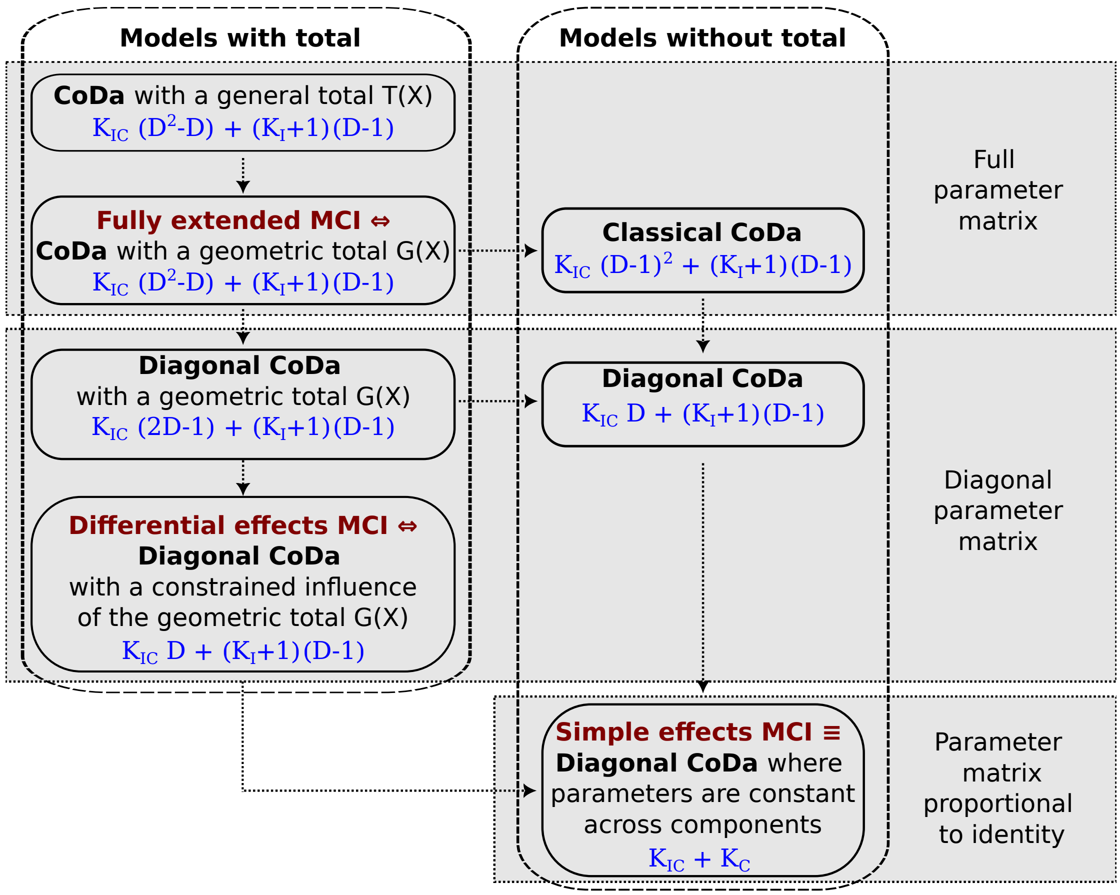

Figure 1: Nesting of MCI models under CoDa models with a general total $T(X)$

4.1 Classical CoDa regression for interaction data

To facilitate the comparison of the CoDa and MCI approaches, we begin by examining the relationship between the simplex representation used for the CoDa model in (15) and the attraction representation employed for the MCI models (6). Comparing these representations reveals that most of the differences in the MCI and the classical CoDa approaches originate in alternative ways of addressing the non-uniqueness of the attraction formulation of the model. We start by adapting the CoDa model to the previously introduced interaction data types:

$$S_{i\bullet} = \alpha_\bullet \oplus \bigoplus_{k=1}^{K_{IC}} (B_k \boxdot x_{ki\bullet}) \oplus \bigoplus_{k=1}^{K_I} (z_{ki} \odot \gamma_{k\bullet}) \oplus \epsilon_{i\bullet} \text{ with } \epsilon_{i\bullet} \sim \mathcal{N}_{\mathcal{S}^D}(\mathcal{C}(\boldsymbol{\iota}_D), \mathbf{G}_D \Sigma \mathbf{G}_D), \qquad (16)$$

where $S_{i\bullet}, \alpha_\bullet, x_{ki\bullet}, \gamma_{k\bullet}, \epsilon_{i\bullet} \in \mathcal{S}^D$, $z_{ki} \in \mathbb{R}$, and $B \in \mathcal{A}_D$. The attraction form of the same model is based on the model expression obtained by replacing in (16) the simplex operations by their definitions in terms of the market share components before closure. Therefore we can express A_{ij} in terms of the volumes $\breve{\alpha}_\bullet, \breve{x}_{ki\bullet}, \breve{\gamma}_{j\bullet}, \breve{\epsilon}_{i\bullet} > 0$ instead of their simplex-valued counterparts:

$$A_{ij} = \breve{\alpha}_j \prod_{k=1}^{K_{IC}} \prod_{l=1}^{D} \breve{x}_{kij}^{b_{kjl}} \prod_{k=1}^{K_I} \breve{\gamma}_{kj}^{z_{ki}} \breve{\epsilon}_{ij} \text{ with } \log(\breve{\epsilon}_{i\bullet}) \sim \mathcal{N}(0, \Sigma). \qquad (17)$$

The above model resembles the fully extended MCI model in (6), with additional type I variables. Let us already note that the distributional assumption on the errors in the MCI models is coherent with CoDa models. The log-normal errors in the attraction model (17) and the simplex-normal errors in (16) are related by $\epsilon_{ij} = \check{\epsilon}_{ij} / \sum_{j=1}^{D} \check{\epsilon}_{ij}$, and after clr-transformation, both versions are equal $\mathrm{clr}(\check{\epsilon}_{i\bullet}) = \mathrm{clr}(\epsilon_{i\bullet})$. This identity is sufficient to conclude that $\log(\check{\epsilon}_{i\bullet}) \sim \mathcal{N}(0, \Sigma) \implies \epsilon_{i\bullet} \sim \mathcal{N}_{\mathcal{S}^D}(\mathcal{C}(\mathbf{1}_D), \mathbf{G}_D \Sigma \mathbf{G}_D)$ because the normal distribution in the simplex is defined by the inverse image of a normal in log-ratio space. It is, however, worth noting that this distribution is invariant to adding row-wise or column-wise constant terms to the correlation matrix yielding $\Sigma + r\boldsymbol{\iota}_D' + \boldsymbol{\iota}_D c'$, where $r, c \in \mathbb{R}^D$. In the usual CoDa approach, this ambiguity is overcome by defining the errors in an ilr space or by directly assuming $\mathbb{V}\mathrm{ar}^{\oplus}(\epsilon_{i\bullet}) \in \mathcal{A}_D$.

The lack of uniqueness also shows up in the deterministic part of the attraction model in (17) and is, in fact, a direct consequence of the closure operation. This becomes obvious when noting that the shares S_{ij} are invariant to scaling the attractions by any factor c_i that varies with the individual i but is constant over components j

$$S_{ij} = \frac{A_{ij} \times c_i}{\sum_{l=1}^{D} A_{il} \times c_i}. \tag{18}$$

In this light, we should understand the parameter restrictions in model (16) not as actual constraints on the parameter space, but as a convention to identify the parameters uniquely. This makes the CoDa literature somewhat more rigorous than most presentations of MCI models, whose attraction specification is usually ambiguous. However, the MCI tradition solves this ambiguity at the estimation step by imposing constraints on the parameters in the clr transformed model. Considering these constraints when specifying the attraction model would only reflect an alternative convention for identifying the parameters.

For the component-specific parameters α_j, the CoDa approach imposes a unit sum $1 = \sum_j \alpha_j$, while the MCI approach leads to a unit product $1 = \prod_j \alpha_j$. This could be represented by different choices of c_i in (18), where the CoDa approach uses $c_i^{-1} \propto \sum_j \alpha_j$, while $c_i^{-1} \propto g(\alpha_\bullet)$ in the MCI approach. Thus, we can link the constrained parameter vectors under both conventions by a reparametrization with the scaling factor $g(\check{\alpha}_\bullet)/\sum_j \check{\alpha}_j$, which is always identified since it does not depend on the scale of $\check{\alpha}_\bullet$.

The identification condition for the parameters $\gamma_{k\bullet}$ associated with the k^{th} type I variable is similar. In the CoDa approach, there is an arithmetic constraint: $1 = \sum_j \gamma_{kj}$, which implies $c_i^{-1} \propto (\sum_j \gamma_{kj})^{z_{ki}}$. Although the original MCI models do not use type I variables, the geometric constraint $1 = \prod_j \gamma_{kj}$ would be equally valid.

When it comes to the IC-type variables, the identification condition on the parameter matrix B_k can be more difficult to understand because it is interdependent with the normalization of the compositional variables $\check{x}_{ki\bullet}$. To simplify the discussion, let us consider a model with a single type IC variable and omit the index k. We note that the condition $\mathbf{B} \in \mathcal{A}_D$ is equivalent to $\mathbf{B}$ having zero as row-means $0 = \bar{b}_{j\bullet}$ and as column-means $0 = \bar{b}_{\bullet l}$, for $l, j = 1, ..., D$. Using these notations, we may develop

$$S_{ij} \propto \frac{\Pi_{l=1}^{D} \check{x}_{il}^{b_{jl}}}{\sum_{p=1}^{D} \Pi_{l=1}^{D} \check{x}_{il}^{b_{pl}}} = \frac{\Pi_{l=1}^{D} \check{x}_{il}^{b_{jl}}}{\sum_{p=1}^{D} \Pi_{l=1}^{D} \check{x}_{il}^{b_{pl}}} \times \frac{\Pi_{l=1}^{D} \check{x}_{il}^{-\bar{b}_{\bullet l}}}{\Pi_{l=1}^{D} \check{x}_{il}^{-\bar{b}_{\bullet l}}} = \frac{\Pi_{l=1}^{D} \check{x}_{il}^{b_{jl} - \bar{b}_{\bullet l}}}{\sum_{p=1}^{D} \Pi_{l=1}^{D} \check{x}_{il}^{b_{pl} - \bar{b}_{\bullet l}}}. \tag{19}$$

Equation (19) show that the zero column sums constraint on $\mathbf{B}$ appears as an identification constraint. In contrast, assuming zero row-sums for $\mathbf{B}$ is a restriction that cannot be derived from the ambiguity of the attraction form. The implication of this restriction becomes clear when we scale the variables $\check{x}_{ij}$ in the attraction by a term r_i that is constant across the D

brands

$$A_{ij} \propto \prod\nolimits_{l=1}^{D} \breve{x}_{il}^{b_{jl}} = \prod\nolimits_{l=1}^{D} \left(\frac{\breve{x}_{il}}{t_i}\right)^{b_{jl}} \cdot r_i^{\sum_{l=1}^{D} b_{jl}}. \tag{20}$$

When $\mathbf{B}$ has zero row-sums, the scaling factor r_i becomes irrelevant as $r_i^{\sum_{l=1}^{D} b_{jl}} = 1$. In particular, if we use the arithmetic total $r_i = \sum_j \breve{x}_{ij}$, equation (20) implies the equivalence between a model based on type IC variables represented as volumes or as shares. Thus, assuming zero row-sums for the matrix $\mathbf{B}$ reflects the fundamental concept of classical CoDa models, namely that only relative information in compositional variables are relevant. It is precisely this assumption that makes the classical CoDa regression inconsistent with the fully extended MCI model, where absolute information has an impact on the market shares.

4.2 The fully extended MCI model from a CoDa perspective

To accommodate the fully extended MCI model within the CoDa framework we have to consider CoDa models with a total. The main distinction between the CoDA and MCI models is the representation of compositional explanatory variables. The MCI approach works directly on non-negative volumes $\breve{x}_{i\bullet} \in \mathbb{R}_+^D$, while the CoDa approach separates the influence of the relative information form that of a total using the so-called $\mathcal{T}$-space, introduced and analyzed by Pawlowsky-Glahn et al. (2014). In essence, the $\mathcal{T}$-space is a product space for a simplex valued vector and a positive total: $(\mathcal{C}(\breve{x}_{i\bullet}), T(\breve{x}_{i\bullet})) \in \mathcal{T} = \mathcal{S}^D \times \mathbb{R}_+$. In principle, the functional form of this total is quite flexible, but to obtain an equivalence between the MCI and CoDa approaches, the geometric total $G(\breve{x}_{i\bullet}) = \prod_{j=1}^{D} \breve{x}_{ij}$ turns out as a natural choice. This observation is coherent with Pawlowsky-Glahn et al. (2014), who showed that using geometric total in $\mathcal{T}$ preserved geometric properties of $\mathbb{R}_+^D$, while this link is broken for more general totals. Our approach only differs from Coenders et al. (2017) in that the total emerges with an alternative parametrization if we derive the CoDa model from the MCI representation.

To illustrate the relation between the fully extended MCI model and CoDa regressions we focus on a model with a single explanatory variable of type IC. In attraction form of this MCI model is given by $A_{ij} = \prod_{l=1}^{D} \breve{x}_{il}^{h_{jl}} \breve{\epsilon}_{ij}$, and its simplex form is $S_{i\bullet} = \mathbf{H} \boxdot \breve{x}_{i\bullet} \oplus \breve{\epsilon}_{i\bullet}$. This model does obviously not satisfy the usual CoDa assumptions, as it associates the volumes in $\breve{x}_{i\bullet}$ to the unconstrained parameter matrix $\mathbf{H}$, and because the errors are not in the simplex. The first step in the right direction is to remove the column means of $\mathbf{H}$, using (19). This leads to the new parameter matrix $\mathbf{G}_D\mathbf{H}$, which is the one actually used by Cooper and Nakanishi (1989) for the fully extended MCI model if we consider the constraints they impose on the parameters. We may additionally normalize the errors $\epsilon_{ij} = \breve{\epsilon}_{ij} / \sum_j \breve{\epsilon}_{ij}$, leading to the following equivalence

$$A_{ij} = \prod\nolimits_{l=1}^{D} \breve{x}_{il}^{h_{jl}} \breve{\epsilon}_{ij} \Longleftrightarrow A_{ij} = \prod\nolimits_{l=1}^{D} \breve{x}_{il}^{h_{jl} - \bar{h}_{\bullet l}} \epsilon_{ij}. \tag{21}$$

The remaining gap from the right hand side equation in (21) to a CoDa regression model is that the variables should be simplex valued and that the parameter matrix should satisfy the zero sum property. To bridge this gap, we use the clr of the matrix $\mathbf{H}$ which allows relating the elements of the clr transformed matrix $\mathbf{H}^* = \text{clr}(\mathbf{H})$ to those of $\mathbf{H}$ by $h_{jl}^* = h_{jl} - \bar{h}_{\bullet l} - \bar{h}_{j\bullet} + \bar{h}_{\bullet\bullet}$, where $\bar{h}_{\bullet l}$ represents the column-means, $\bar{h}_{j\bullet}$ the row-means, and $\bar{h}_{\bullet\bullet}$ the overall mean of $\mathbf{H}$. With this in mind, we may further develop the right hand side of (21):

$$\begin{aligned} A_{ij} = \prod_{l=1}^{D} \check{x}_{il}^{h_{jl}-\bar{h}_{\bullet l}} \epsilon_{ij} &= \prod_{l=1}^{D} \check{x}_{il}^{h_{jl}-\bar{h}_{\bullet l}-\bar{h}_{j\bullet}+\bar{h}_{\bullet\bullet}} \check{x}_{il}^{\bar{h}_{j\bullet}-\bar{h}_{\bullet\bullet}} \epsilon_{ij} \\ &= \left(\prod_{l=1}^{D} \check{x}_{il}^{h^*_{jl}}\right) \left(\prod_{l=1}^{D} \check{x}_{il}\right)^{\bar{h}_{j\bullet}-\bar{h}_{\bullet\bullet}} \epsilon_{ij} \\ &= \left(\prod_{l=1}^{D} x_{il}^{h^*_{jl}}\right) \exp(\bar{h}_{j\bullet} - \bar{h}_{\bullet\bullet})^{\log G(\check{x}_{i\bullet})} \epsilon_{ij}. \end{aligned} \tag{22}$$

In the second line of (22) the geometric total appears naturally because the associated parameters are not indexed by l. Going from the second to the third line we replace the elements of the volume vector $\check{x}_{i\bullet}$ by those of the composition $x_{i\bullet}$, which is made possible by $\sum_j h^*_{jl} = 0$ in conjunction with (20). Additionally, we use a representation where the total appears as the exponent of a parameter that we denote $\tau_j = \exp(\bar{h}_{j\bullet} - \bar{h}_{\bullet\bullet})$. This allows rewriting (22) in terms of simplex operators as

$$S_{i\bullet} = \mathbf{H}^* \boxdot x_{i\bullet} \oplus \tau_\bullet \odot \log G(\check{x}_{i\bullet}) \oplus \epsilon_{i\bullet}, \tag{23}$$

where almost all the usual requirements: $x_{i\bullet} \in \mathcal{S}^D$, $\mathbf{H}^* \in \mathcal{A}$, and $\epsilon_{i\bullet}$ being normal in the simplex, are satisfied. The only conventional restriction of CoDa models that is violated is $\tau_\bullet \in \mathcal{S}^D$. However, all parameters are still uniquely identified since $\prod_j \tau_j = 1$. Changing this geometric normalization into the conventional unit sum constraint would only lead to a different parametrization for the same model. However, for our purpose, we believe that the geometric constraint makes the link between the MCI and the CoDa parameters clearer. At this stage, we already want to point out that the parametrization is only relevant if we want to interpret the model directly in the simplex. These differences disappear when the model is interpreted in the clr or any ilr space.

After clr-transformation of (22), the correspondence between the MCI and the CoDa representations is given by

$$\begin{aligned} S^*_{ij} &= \sum_{l=1}^{D} (h_{jl} - \bar{h}_{\bullet l}) \log \check{x}_{ij} + \epsilon^*_{ij} \\ &= \sum_{l=1}^{D} h^*_{jl} x^*_{ij} + (\bar{h}_{j\bullet} - \bar{h}_{\bullet\bullet}) \log G(\check{x}_{i\bullet}) + \epsilon^*_{ij}. \end{aligned} \tag{24}$$

The first line in (24) mirrors the representation of the fully extended MCI model from which Cooper and Nakanishi (1989) estimate the parameters, and the second line represents the the corresponding CoDa model in clr-space. Using $x^*_{ij} = \log \check{x}_{ij} - \log G(\check{x}_{i\bullet})/D$ and $\sum_j x^*_{ij} = 0$, we can easily go from the first to the second line. One convenient aspect of keeping a geometric normalization for $\tau_\bullet$ is that we have $\log(\tau_j) = \tau_j* = \bar{h}_{j\bullet} - \bar{h}_{\bullet\bullet}$.

Table 3 shows a vectorized representation of the data for the clr-transformed model in the second line of (24). Vectorizing the data involves stacking the share vectors $S^*_{i\bullet}$ for all individuals into a single column. The individual and component indexes of these shares are shown in the first two columns of the table.

The explanatory variable of type IC is spread over D^2 columns that we grouped into D sets, according to the component index l of x^*_{il}. In the fully extended MCI model, each component of $x^*_{i\bullet}$ influences each component of $S^*_{i\bullet}$ with a distinct parameter. In the last three rows, we recall these parameters h^*_{jl} and the associated indexes. The transformed total is used to explain all components of $S^*_{i\bullet}$ with a dedicated parameter, which is why the same value is repeated D times for each individual. The constraints $0 = \sum_j \tau^*_j = \sum_j h^*_{jl} = \sum_l h^*_{jl}$ on the parameters also seem natural when we consider that, for each individual, the transformed share vector $S^*_{i\bullet}$ and the transformed type IC variables $x^*_{i\bullet}$ are centered at zero. If we included additional information of

Table 3: Vectorized data for the fully extended MCI model in clr space

Indiv. (i)	S-Comp. (j)	Shares (S^*_{ij})	Type IC variable (x^*_{il} is spread over D groups of D columns)														Total $T(\check{x}_{i\bullet})$			
1	1	S^*_{11}	x^*_{11}				x^*_{12}				...	x^*_{1D}					t_1			
1	2	S^*_{12}		x^*_{11}				x^*_{12}			...		x^*_{1D}					t_1		
...	...	...			...				...		...			...					...	
1	D	S^*_{1D}				x^*_{11}				x^*_{12}	...				x^*_{1D}					t_1
2	1	S^*_{21}	x^*_{21}				x^*_{22}				...	x^*_{2D}					t_2			
2	2	S^*_{22}		x^*_{21}				x^*_{22}			...		x^*_{2D}					t_2		
...	...	...			...				...		...			...					...	
2	D	S^*_{2D}				x^*_{21}				x^*_{22}	...				x^*_{2D}					t_2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		...	...	...	...
N	1	S^*_{N1}	x^*_{N1}				x^*_{N2}				...	x^*_{ND}					t_D			
N	2	S^*_{N2}		x^*_{N1}				x^*_{N2}			...		x^*_{ND}					t_D		
...	...	...			...				...		...			...					...	
N	D	S^*_{ND}				x^*_{N1}				x^*_{N2}	...				x^*_{ND}					t_D
Parameters			h^*_{11}	h^*_{21}	...	h^*_{D1}	h^*_{12}	h^*_{22}	...	h^*_{D2}	...	h^*_{1D}	h^*_{2D}	...	h^*_{DD}		τ^*_1	τ^*_2	...	τ^*_D
S-Comp. (j)			1	2	...	D	1	2	...	D	...	1	2	...	D		1	2	...	D
X-Comp. (l)			1	1	...	1	2	2	...	2	...	D	D	...	D					

Blank cells correspond to zeros. When the values in the total column are computed as $t_i = \sum_j \log \check{x}_{ij}$, we obtain a CoDa consistent parametrization of the fully extended MCI model.

type I in the model, this data should have the same format as the totals and requires replacing t_i with z_{ki} and τ^*_j with γ^*_j. The constant terms α^*_j would be associated to data in the same format with the difference that $t_i = 1$.

Although it would be computationally inefficient to estimate the fully extended MCI model from this representation it helps to clarify how this model uses the available information, particularly when compared to the simple and differential effects MCI models. In the differential effects model, all data except for type IC data, could be presented in the same format. When it comes to type IC data, this model only requires the information in the columns that are shaded in gray, where the component index of S^*_{ij} and x^*_{il} coincide. In the simple effects MCI model we could only use Type IC variables, and each of these should be represented like the shares as a single ND vector. In the context of Table 3 this vector would correspond to the sum of the gray columns and should be associated with a single parameter.

4.3 The simple effects MCI model from a CoDa perspective

The simple effects MCI model can be regarded as a special case of the fully extended version, in which we assume the influence of all explanatory variables to be constant across the D brands. Therefore it is also a special case of CoDa regression models with a total. However, since the constraints on the parameters imply zero influence of this total on the shares, the simple MCI model is also consistent with classical CoDa regression, as already acknowledged by Morais (2017). Considering the earlier discussion on data types, we may restate the simple effects MCI model as

$$A_{ij} = \prod_{k=1}^{K_{IC}} \check{x}_{kij}^{h_k} \prod_{k=1+K_{IC}}^{K_{IC}+K_C} \check{x}_{kj}^{h_k} \check{\epsilon}_{ij}, \tag{25}$$

where the number of type C variables $K_C \leq D-1$ is limited to avoid collineraty issues. Type C variables cannot be used in the two largest MCI models in (6) because the associated parameters cannot be identified when the model includes component-specific constants α_j.

To demonstrate that model (25) is consistent with the standard CoDa assumptions, we may simply repeat the steps in (22). Since these developments are identical for all type C and type IC variables, we may again consider a model with a single type IC variable $\check{x}_{i\bullet}$ associated with the parameter matrix $\mathbf{H}$. The sole difference is that $\mathbf{H}$ is a full matrix in the fully extended MCI model, while it is proportional to the identity $\mathbf{H} = h * \mathbf{I}_D$ in the simple MCI model. Consequently, the overall mean $\mathbf{H}$ and the row-wise means are identical $\bar{h}_{\bullet\bullet} = \bar{h}_{j\bullet} = h/D$ and therefore the total $G(\check{x}_{i\bullet})$ disappears in (22) when considering the simple MCI model.

Another interesting fact is that the simple MCI model is directly equivalent to its CoDa version, while reparametrizations must be considered to achieve an equivalence between a CoDa model and the two larger MCI versions. Let us note that $\mathbf{H} = h\mathbf{I}_D$ implies $\mathbf{H}^* = h\mathbf{G}_D$. Using this CoDa consistent parameter matrix the attraction becomes $A_{ij} = x_{ij}^h (\prod_{l=1}^D x_{il}^{-h/D})\epsilon_{ij}$, where the constant $-h/D$ disappears after the clr transformation:

$$S_{ij}^* = hx_{kij}^* - \frac{h}{D} \underbrace{(x_{i1}^* + x_{i2}^* + ... x_{iD}^*)}_{=0} + \epsilon_{ij}^*.$$

Thus, under any log-ratio transformation the CoDa version and the original MCI model based on $A_{ij} = \check{x}_{ij}^h \check{\epsilon}_{ij}$ become identical.

Table 4: Vectorized data for the simple MCI model, and clr transformation

Indiv. (i)	S-Comp. (j)	Shares S_{ij}	Type C $\check{x}_j$	Type IC $\check{x}_{ij}$	Type I z_i
1	1	$S_{11} \to \log(S_{11}/g(S_{1\bullet}))$	$\check{x}_1 \to \log(\check{x}_1/g(\check{x}_\bullet))$	$\check{x}_{11} \to \log(\check{x}_{11}/g(\check{x}_{1\bullet}))$	$z_1 \to 0$
1	2	$S_{12} \to \log(S_{12}/g(S_{1\bullet}))$	$\check{x}_2 \to \log(\check{x}_2/g(\check{x}_\bullet))$	$\check{x}_{12} \to \log(\check{x}_{12}/g(\check{x}_{1\bullet}))$	$z_1 \to 0$
...	...	...	...	...	...
1	D	$S_{1D} \to \log(S_{1D}/g(S_{1\bullet}))$	$\check{x}_D \to \log(\check{x}_D/g(\check{x}_\bullet))$	$\check{x}_{1D} \to \log(\check{x}_{1D}/g(\check{x}_{1\bullet}))$	$z_1 \to 0$
2	1	$S_{21} \to \log(S_{21}/g(S_{2\bullet}))$	$\check{x}_1 \to \log(\check{x}_1/g(\check{x}_\bullet))$	$\check{x}_{21} \to \log(\check{x}_{21}/g(\check{x}_{2\bullet}))$	$z_2 \to 0$
2	2	$S_{22} \to \log(S_{22}/g(S_{2\bullet}))$	$\check{x}_2 \to \log(\check{x}_2/g(\check{x}_\bullet))$	$\check{x}_{22} \to \log(\check{x}_{22}/g(\check{x}_{2\bullet}))$	$z_2 \to 0$
...	...	...	...	...	...
2	D	$S_{2D} \to \log(S_{2D}/g(S_{2\bullet}))$	$\check{x}_D \to \log(\check{x}_D/g(\check{x}_\bullet))$	$\check{x}_{2D} \to \log(\check{x}_{2D}/g(\check{x}_{2\bullet}))$	$z_2 \to 0$
...	...	...	...	...	...
N	1	$S_{N1} \to \log(S_{N1}/g(S_{N\bullet}))$	$\check{x}_1 \to \log(\check{x}_1/g(\check{x}_\bullet))$	$\check{x}_{N1} \to \log(\check{x}_{N1}/g(\check{x}_{N\bullet}))$	$z_N \to 0$
N	2	$S_{N2} \to \log(S_{N2}/g(S_{N\bullet}))$	$\check{x}_2 \to \log(\check{x}_2/g(\check{x}_\bullet))$	$\check{x}_{N2} \to \log(\check{x}_{N2}/g(\check{x}_{N\bullet}))$	$z_N \to 0$
...	...	...	...	...	...
N	D	$S_{ND} \to \log(S_{ND}/g(S_{N\bullet}))$	$\check{x}_D \to \log(\check{x}_D/g(\check{x}_\bullet))$	$\check{x}_{ND} \to \log(\check{x}_{ND}/g(\check{x}_{N\bullet}))$	$z_N \to 0$

Table 4 provides a vectorized representation of the data for the simple effects MCI model, and demonstrates the impact of the clr-transformation on the different data types. For this model, vectorizing actually serves a practical purpose, as it enables us to estimate the parameters using unrestricted OLS. For type C variables, the same D values are repeated for all N individuals. In the last column, we illustrate that type I variables would be mapped to a column of zeros by the clr transformation, which could potentially lead to collinearity issues if they were inadvertently included in a simple MCI model.

4.4 The differential effects MCI model from a CoDa perspective

Adopting a CoDa perspective for the differential effects MCI model is arguably more complex than for the other two MCI models because the influence of absolute information and relative information are interdependent. We will show that this interdependence is reflected by rather intricate constraints on the parameters, making estimation and interpretation of this model difficult. For this reason, we will consider three variants of the differential effects MCI model, which we name diagonal CoDa models because they use a diagonal parameter matrix to capture the influence of type IC variables. The three model variants only differ in how they include absolute information. The most general diagonal CoDa model includes a total without constraints. When this total is removed, we obtain a diagonal version of the classical CoDa model, and under certain constraints on the influence of the total we recover the differential effects MCI model.

We begin by deriving the CoDa consistent parametrization of a differential effects MCI model, where we focus on the attraction with a single Type IC variable: $A_{ij} = \check{x}_{ij}^{h_j} \check{\epsilon_{ij}}$. This attraction also corresponds to a fully extended MCI model under the additional constraint that all of-diagonal elements of the parameter matrix $\mathbf{H}$ are zero. Consequently, we can simply repeat the steps in (22) to derive the CoDa consistent parametrization of the differential MCI model. For this model the structure of $\mathbf{H} = \text{diag}(h_1, ..., h_D)$ allows us to simplify the expression of the centered row-means as $\bar{h}_{j\bullet} - \bar{h}_{\bullet\bullet} = \frac{1}{D}(h_j - \bar{h}_\bullet)$, and the elements of $\mathbf{H}^* = \text{clr}(\mathbf{H})$ become

$$h_{jl}^* = \begin{cases} \frac{D-2}{D} h_j + \frac{1}{D} \bar{h}_\bullet & \text{if } j = l \\ \frac{-1}{D}(h_j + h_l) + \frac{1}{D} \bar{h}_\bullet & \text{if } j \neq l \end{cases}. \tag{26}$$

Using these simplifications, the equivalence between the original and the CoDa version of the differential effects MCI model is

$$A_{ij} = \check{x}_{ij}^{h_j} \check{\epsilon_{ij}} \iff A_{ij} = \left(\prod_{l=1}^{D} x_{il}^{h_{jl}^*} \right) \exp(h_j - \bar{h}_\bullet)^{\log g(\check{x}_{i\bullet})} \epsilon_{ij}. \tag{27}$$

In the above equation, the absolute information in the vector $\check{x}_{i\bullet}$ is represented by its geometric mean, while the fully extended version uses the geometric total. Since these two measures are linked by $\log g(\check{x}_{i\bullet}) = \log G(\check{x}_{i\bullet})/D$, the difference is only a matter of absorbing the constant $1/D$ in the data, instead of the parameters. The interpretation of model (27) is complex because the parameters associated with the geometric mean are linked to the elements of $\mathbf{H}^*$ in (26).

To better understand how model (27) uses relative information, let us first consider the diagonal CoDa model as an alternative. In Figure 1, this model is presented as a special case of classical CoDa regression because it does not use absolute information. The attraction form of the diagonal CoDa model is given by

$$A_{ij} = \left(\prod_{l=1}^{D} x_{il}^{h_{jl}^*} \right) \epsilon_{ij}, \tag{28}$$

with h_{jl}^* as in (26). Like the differential MCI model, it has D parameters associated with each variable of type IC and also requires the same information sets to estimate these parameters from observed data (see Table 2). The main difference is that the diagonal CoDa model breaks with the assumption of zero cross-effects that Cooper and Nakanishi (1989) use in the differential MCI model. While differentiated direct effects are maintained, zero cross-effects are replaced with the assumption that the cross-effect of a characteristic x_{ij} of brand j on the market share

S_{il} of brand l is negatively proportional to the sum of the two direct effects. This assumption follows from the form of $\mathbf{H}^*$ in (26), where we can ignore the term $\frac{1}{D}\bar{h}_\bullet$ as it would cancel out when computing the shares. Such an assumption regarding how a variable impacts the market shares may initially seem purely technical. However, in a marketing context, it has some logic, in particular from the perspective of brand differentiation. For instance, if we think about advertisements for soft drinks, it is plausible that a consumer who sees the advertisement of a particular brand might have an increased desire for soft drinks in general and not only those of the brand whose advertisement he saw. In this context, the assumption of the diagonal CoDa model without total is consistent with the idea that the more a brand differentiates itself from its competitors, the less the other brands benefit from its advertisement, while at the same time, the brand itself also benefits less from the advertisement of its competitors.

In the differential effects MCI model, we can use the concept of brand differentiation to interpret the influence of relative information in the type IC variables. However, we also need to account for absolute information that enters (27) through the geometric mean. Here the link between the parameters $\tau_j = \exp(h_j - \bar{h}_\bullet)$ and $\mathbf{H}^*$ postulates that the differences in how the market share of each brand reacts to changes in absolute information must perfectly offset the cross-effects due to relative information, which is a rather unnatural assumption.

A straightforward generalization of the previous two models is to consider the influence of the total as independent from the influence of the shares. Starting from (27) this simply requires considering the parameter associated with $g(\check{x}_{i\bullet})$ as independent from $\mathbf{H}^*$, leading to the model

$$A_{ij} = \prod\nolimits_{l=1}^{D} x_{il}^{h_{jl}^*} \tau_j^{\log g(\check{x}_{i\bullet})} \epsilon_{ij}, \tag{29}$$

where $\prod_j \tau_j = 1$. The above model recovers the diagonal CoDa without total when $\tau_j = 1$, for $j = 1, \ldots, D$ and in the case of $\tau_j = \exp(h_j - \bar{h}_\bullet)$ it is equivalent to the differential MCI model.

One issue with all diagonal CoDa models is that the definition of $\mathbf{H}^*$ as the clr of a diagonal matrix leads to relatively complex constraints on the parameters. Enforcing these constraints within the multivariate estimation approach typically used for CoDa regression is difficult because they are not included in standard software packages. However, if we adopt a vectorized data representation, these constraints become much simpler, and by further rearranging the vectorized data, it is even possible to estimate the model parameters by unconstrained OLS. This arrangement of the data is already used by Cooper and Nakanishi (1989), who presented the clr-transformed differential MCI model as

$$S_{ij}^* = \frac{D-1}{D} \log \check{x}_{ij} h_j - \sum\nolimits_{l \neq j}^{D} \frac{1}{D} \log \check{x}_{il} h_l + \epsilon_{ij}^*. \tag{30}$$

Rewriting this equation for the whole share vector yields $S_{i\bullet}^* = \mathbf{G}_D \operatorname{diag}(\log \check{x}_{i1}, \ldots, \log \check{x}_{iD}) + \epsilon_{i\bullet}^*$, which is easily converted into any ilr representation by pre-multiplication with $\mathbf{V}'$ the transpose of the corresponding contrast matrix. We can use a similar representation for the diagonal CoDa model, which only requires replacing $\log \check{x}_{i1}$ by x_{i1}^* in (30). Adding $\log g(\check{x}_{i\bullet})$ with the associated parameters leads to the equation for the diagonal CoDa model with a geometric total that nests the other two diagonal CoDa models as special cases.

Table 5 illustrates the vectorized data representation for the diagonal CoDa model with a total for the clr and the ilr-based formulations, where rows with component index D would not exist in the ilr case. When the data of type IC is represented in this form, both transformations directly estimate the parameters h_j that also appear on the main diagonal of $\mathbf{H}$ in the original attraction form of the differential effects MCI model. When including type I variables in the model, they should have the same structure as the variables in the total columns, and the associated parameters are estimated in log-ratio coordinates, as is normal for CoDa models. As explained previously, we could adjust this data representation to all three diagonal CoDa model

Table 5: Vectorized data for the diagonal CoDa model with a total in clr and ilr spaces

Indiv. (i)	S-Comp. (j)	clr Shares S^*_{ij}	clr Type IC x^*_{ij}	clr Total $T(\check{x}_{i\bullet})$	ilr Shares S^{*V}_{ij}	ilr Type IC x^*_{ij}	ilr Total $T(\check{x}_{i\bullet})$
1	1	S^*_{11}			S^{*V}_{11}		
1	2	S^*_{12}	$\underbrace{\mathbf{G}_D \operatorname{diag}(x^*_{1\bullet})}_{D\times D}$	$\mathbf{I}_D \cdot t_1$	S^{*V}_{12}	$\underbrace{V' \operatorname{diag}(x^*_{1\bullet})}_{(D-1)\times D}$	$\mathbf{I}_{D-1} \cdot t_1$
...	...	...			...		
1	$D-1$	$S^*_{1(D-1)}$			$S^{*V}_{1(D-1)}$		
1	D	S^*_{1D}					
2	1	S^*_{21}			S^{*V}_{21}		
2	2	S^*_{22}			S^{*V}_{22}	$V' \operatorname{diag}(x^*_{2\bullet})$	$\mathbf{I}_{D-1} \cdot t_2$
...	...	...	$\mathbf{G}_D \operatorname{diag}(x^*_{2\bullet})$	$\mathbf{I}_D \cdot t_2$	...		
2	$D-1$	$S^*_{2(D-1)}$			$S^{*V}_{2(D-1)}$		
2	D	S^*_{2D}					
...	...	...	...	...	...	...	...
N	1	S^*_{N1}			S^{*V}_{V1}		
N	2	S^*_{N2}			S^{*V}_{N2}	$V' \operatorname{diag}(x^*_{N\bullet})$	$\mathbf{I}_{D-1} \cdot t_N$
...	...	...	$\mathbf{G}_D \operatorname{diag}(x^*_{N\bullet})$	$\mathbf{I}_D \cdot t_N$	...		
N	$D-1$	$S^*_{(D-1)}$			$S^{*V}_{N(D-1)}$		
N	D	S^*_{ND}					

When the values in the total columns are computed as $t_i = \frac{1}{D}\sum_j \log \check{x}_{ij}$, the model nests the differential effects MCI model as a special case. The notation S^{*V}_{ij} is a short form for the j^{th} element of the vector $\text{ilr}_V(S_{i\bullet})$.

variants. It is thus fairly easy to estimate the two models nested below the diagonal CoDa model with a total, which means that the problem of model selection could be approached using standard testing procedures for nested multivariate linear models.

4.5 Alternative error structures and estimators

In this subsection, we revisit the discussion on estimation strategies, variance structures, and potential heteroskedasticity in MCI models, tracing back to Nakanishi and Cooper (1974). First, we address the inefficiency of the OLS estimator in simple and differential effects MCI models without considering heteroskedasticity. For this issue, the CoDa methods enable us to improve the existing estimation approaches of MCI models. Next, we integrate various heteroskedasticity forms into the MCI and CoDa models. On this point, the CoDa regression approach benefits from the insights in the MCI literature, although we finally argue for a simpler error structure than the one considered by Nakanishi and Cooper (1974). Our discussion of efficient estimation methods is closely related to the literature on seemingly unrelated regression equations (SURE) initiated by Zellner (1962), and the heteroskedastic error structures we consider resemble the generalizations of SURE proposed by Fiebig et al. (1991). We also want to point to Chen et al. (2018), who compare several estimators of the error covariance in CoDa regression models that are robust to unknown forms of heteroskedasticity. Our case is simpler because we derive the heteroskedasticity structure from the microeconomic foundation of the MCI model, which means that we can treat it as known and correct it directly.

The connection between SURE and classical CoDa models becomes evident when representing the latter in ilr space. This stems from the fact that an ilr-transformed CoDa model can be represented as a multivariate linear model:

$$\underset{N\times(D-1)}{\mathbf{Y}} = \mathbf{X}\Theta + \underset{N\times(D-1)}{\mathbf{E}} \tag{31}$$

where the rows of $\mathbf{Y}$ represent the ilr-transformed share vectors, denoted by $\text{ilr}_V(S_{i\bullet})'$. The matrix $\mathbf{E}$ employs the same representation for the error vectors, each being an independent realization of a multivariate normal distribution $\text{ilr}_V(\epsilon_{i\bullet}) \sim \mathcal{N}_{D-1}(0, \mathbf{V}'\Sigma\mathbf{V})$. The matrix $\mathbf{X}$ contains a constant and all explanatory variables, where the compositional variables are transformed by the same ilr_V as the dependent shares. The parameter matrix Θ associated with $\mathbf{X}$ consists of the ilr-transformed parameters, where the transformation is different for the vector and matrix-valued parameters. For the vector of constants and the parameter vectors associated with classical variables, we use the usual ilr, but matrix-valued parameters should be transformed by (11). In this format, a CoDa model expressed by (31) corresponds to a set of unrelated linear equations except for potential correlations in the Gaussian error terms, which is the core assumption of SURE.

When Zellner (1962) first studied SURE, he identified two alternative conditions under which the OLS estimator $\hat{\Theta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the best linear unbiased estimator (BLUE). The first condition requires the errors to be uncorrelated across equations, which means that $\mathbf{V}'\Sigma\mathbf{V}$ must be diagonal. This condition is unrealistic for MCI models and probably CoDa models in general because it only holds independently of $\mathbf{V}$ when the correlation matrix of the log-normal attraction errors is proportional to an identity $\Sigma = \sigma^2\mathbf{I}_D$.

The second condition under which OLS is BLUE for SURE is that all equations have identical sets of explanatory variables. This is obviously the case for all models that can be written in the form of an unrestricted multivariate linear model as in (31), which is possible for the fully extended MCI model, and classical CoDa models, with and without a total. In contrast, the simple and differential MCI and diagonal CoDa models do not admit such a representation. Therefore, we will consider the following vectorized formulation of the ilr transformed models

$$\text{Vec}(\mathbf{Y}') = \mathbf{Z}\theta + \text{Vec}(\mathbf{E}'), \tag{32}$$

where the Vec-operator stacks the columns of a matrix, and $\mathbf{Z}$ contains the vectorized representation of explanatory variables. All models shown in Figure 1 can be represented by (32), where each model would use its distinct set of explanatory variables in the columns of $\mathbf{Z}$ with their associated parameters θ. For $y = \text{Vec}(\mathbf{Y}')$ the OLS of the parameter θ in model (32) is

$$\hat{\theta}_{OLS} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'y, \tag{33}$$

and it is generally inefficient for the models considered here unless the overall covariance matrix $\Omega = \mathbb{V}\text{ar}(y)$ is diagonal. To obtain an efficient estimator Zellner (1962) proposed to use the generalized least squares (GLS) formula when Ω is assumed to be known:

$$\hat{\theta}_{GLS} = (\mathbf{Z}'\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\Omega^{-1}y. \tag{34}$$

OLS remains efficient for models that admit a multivariate form because, for these models, OLS and GLS coincide. To better understand this relation, let us first note that model (31) is equivalent to model (32), when $\theta = \text{Vec}(\hat{\Theta}')$ and $\mathbf{Z} = \mathbf{X} \otimes \mathbf{I}_{D-1}$, where $\otimes$ denotes the Kronecker product. Using this operation, the structure of $\mathbf{Z}$ would mirror one of the explanatory variables in Table 3, with the difference that Table 3 uses clr instead of ilr transformations. When $\mathbf{Z}$

admits such a Kronecker product structure, the OLS and GLS formulas are equivalent:

$$\begin{aligned}
\mathrm{Vec}(\hat{\Theta}'_{GLS}) &= \left[\mathbf{Z}'\Omega^{-1}\mathbf{Z}\right]^{-1}\mathbf{Z}'\Omega^{-1}y \\
&= \left[(\mathbf{X}\otimes\mathbf{I}_{D-1})'(\mathbf{I}_N\otimes\mathbf{V}'\Sigma\mathbf{V})^{-1}(\mathbf{X}\otimes\mathbf{I}_{D-1})\right]^{-1}(\mathbf{X}\otimes\mathbf{I}_{D-1})'(\mathbf{I}_N\otimes\mathbf{V}'\Sigma\mathbf{V})^{-1}y \\
&= \left[(\mathbf{X}'\mathbf{X})\otimes(\mathbf{V}'\Sigma\mathbf{V})^{-1}\right]^{-1}(\mathbf{X}'\otimes(\mathbf{V}'\Sigma\mathbf{V})^{-1})\,\mathrm{Vec}(\mathbf{Y}') \\
&= ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_{D-1})\,\mathrm{Vec}(\mathbf{Y}') \\
&= \mathrm{Vec}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}') \\
&= \mathrm{Vec}(\hat{\Theta}'_{OLS}).
\end{aligned} \tag{35}$$

Thus, while OLS yields the minimum variance estimators in the three models in the first block of Figure 1 (Full parameter matrix), we need to turn to GLS to obtain efficient estimators for the other models. This problem was already recognized by Nakanishi and Cooper (1974), but they could not directly apply the GLS formula in (34) because they estimate the model in clr-space where the covariance matrix Ω would be singular. This issue is easily solved by estimating the model in an ilr-space.

In their article, Nakanishi and Cooper (1974) further argue that MCI models are likely to have heteroskedastic errors because the true market shares of the brands are usually unavailable and must be derived from the sales to a sample of customers. Therefore, market shares computed from fewer customers are likely associated with higher volatility than those based on a larger sample. In their original approach, Nakanishi and Cooper (1974) model this heteroskedasticity using a composite error term that is a sum of a specification error and a sampling error, which they assume to be independent. Their specification error coincides with the errors we used so far, and their sampling error is derived from independent log-normal distributions. Heteroskedasticity is only present in this sampling error since their variances are inversely proportional to the customer volumes $\breve{v}_{ij}$: $\log\breve{\epsilon}_{ij}\sim\mathcal{N}(0,\sigma^2 v_{ij}^{-1})$. However, we do not adopt this composite variance structure here, mainly because it is not frequently used in statistical models, and to our knowledge, no closed-form solution for an efficient estimator exists in this case.

In the following, we consider an alternative form of heteroskedasticity, where the differences in volatility act as a scaling factor on the usual errors. This approach leads to a similar error structure as the one of Fiebig et al. (1991) but still allows us to model the customer-volume effects like the one considered by Nakanishi and Cooper (1974). To formalize this heteroskedasticity assumption, we can start from the most general error structure that maintains independence across individuals [1]. This case corresponds to $\log\breve{\epsilon}_{i\bullet}\sim\mathcal{N}(0,\Sigma_i)$ with different correlation matrices for each statistical individual. The associated error covariances in the simplex and in clr space are $\mathbf{G}_D\Sigma_i\mathbf{G}_D$, and $\mathbf{V}'\Sigma_i\mathbf{V}$ for the ilr space with contrast matrix $\mathbf{V}$. However, without imposing further structure on Σ_i we could not identify all covariance terms since their number is larger than the ND observations. To solve this issue, we use the same correlation matrix Σ for all individuals. This shared correlation matrix is then adjusted through a scaling factor defined at the individual-component pair level, leading to $\Sigma_i=\Lambda_i\Sigma\Lambda_i$, with $\Lambda_i=\mathrm{diag}(\sqrt{\lambda_{i1}},...,\sqrt{\lambda_{iD}})$. With this structure, we can link the volatility of the errors to the customer volumes using, for example, $\lambda_{ij}=1/\breve{v}_{ij}$. Since the scaling factors λ_{ij} can be considered Type IC information, we refer to Type IC heteroskedasticity. Imposing further constraints on Λ_i allows us to define Type I and Type C heteroskedasticity and recover the usual homoskedastic case, as summarized by Table 6. According to this classification, the estimators of Chen et al. (2018) correct for unknown Type I heteroskedasticity.

[1] One way to extend MCI models to include potential correlation across individuals is offered by spatial econometric methods. While such an extension is out of the scope of this article, we refer to Lee and Pace (2005) for a spatial model of retail sales and to Nguyen et al. (2021) for a spatial CoDa model

Table 6: Heteroskedasticity structures for MCI and CoDa models

Type	Correlation: Σ_i	Scaling factors: Λ_i
Type IC Heteroskedasticity	$\Sigma_i = \Lambda_i \Sigma \Lambda_i$	$\Lambda_i = \text{diag}(\sqrt{\lambda_{i1}}, ..., \sqrt{\lambda_{iD}})$
Type I Heteroskedasticity	$\Sigma_i = \lambda_i^2 \Sigma$	$\Lambda_i = \lambda_i \mathbf{I}_D$
Type C Heteroskedasticity	$\Sigma_i = \Lambda \Sigma \Lambda$	$\Lambda_i = \Lambda = \text{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_D})$
Homoscedasticity	$\Sigma_i = \Sigma$	$\Lambda_i = \mathbf{I}_D$

A practical advantage of type I heteroskedasticity is that the scaling factors λ_i are not affected by any ilr transformation, as for example in $\text{ilr}_V(\lambda_i^2 \Sigma) = \lambda_i^2 \mathbf{V}' \Sigma \mathbf{V}$. For this reason, we could simply use weighted versions of the OLS or GLS estimators in (33) and (34) to correct for this form of heteroskedasticity. Unfortunately, this simplification does not arise with type IC and type C heteroskedasticity since the scaling matrix Λ_i would also be transformed by the operation $\text{ilr}_V(\Lambda_i \Sigma \Lambda_i)$. However, even when there is type C heteroskedasticity, the variance structures are identical for all individuals, which means that an unweighted GLS estimation remains efficient. Correcting for type IC heteroskedasticity is more complicated because it requires weighted estimators, where the weights must be considered in the log-ratio transformations. Although Hron et al. (2022) explain how to work with weighted log-ratios, we leave their application to CoDa regression models for future research.

It is worth noting that the previously outlined arguments for heteroskedasticity are not exclusive to a marketing context and may be useful in many other applications of CoDa regression, in particular when the dependent shares are computed by aggregating choices of individuals. For example, when modeling election results at a regional level, where the vote shares of parties or candidates are expected to be more stable for regions with a larger population.

4.6 Fitted shares and estimated elasticities

In this section, we develop the equivalence between the elasticity expressions of Cooper and Nakanishi (1989) and the elasticities for CoDa models with a total developed in Morais and Thomas-Agnan (2021). In line with our previous results, we find that adopting a CoDa perspective for MCI models leads to a decomposition of these elasticities into contributions from relative and absolute information. For completeness, we also treat the semi-elasticities that Morais and Thomas-Agnan (2021) proposed to interpret the influence of classical variables in CoDa models. In contrast to some of the previous work on this topic, we define the elasticities in terms of the fitted values, as opposed to the observed shares used by Cooper and Nakanishi (1989), or their expectations in the simplex-sense used by Morais and Thomas-Agnan (2021).

Before entering into the details of the elasticities, we need to define the fitted values. As in any regression model, they are computed by evaluating a prediction formula for the observations used to fit the model. Since the CoDa model with a total subsumes all other models as special cases, we can focus on its prediction formula:

$$\hat{S}_{i\bullet} = \hat{\alpha}_{\bullet} \oplus \bigoplus_{k=1}^{K_{IC}} \left(\hat{\mathbf{H}}_k^* \boxdot x_{ki\bullet} \oplus f_k(T(\check{x}_{ki\bullet})) \odot \hat{\tau}_{k\bullet} \right) \oplus \bigoplus_{k=1}^{K_I} (z_{ki} \odot \hat{\gamma}_{k\bullet}), \tag{36}$$

where choosing a log-transformed geometric total $f_k(T(\check{x}_{ki\bullet})) = \log G(\check{x}_{ki\bullet})$ recovers the prediction formula of a fully extended MCI model with additional variables of type I.

One issue for all MCI and CoDa models is that the prediction in (36) is biased when considering the classical sense of expectation: $\mathbb{E}(S_{ij}) \neq \mathbb{E}(\hat{S}_{ij})$. However, the usual definition of the expectation $\mathbb{E}(x) = \int x f(x) dx$ is tailored to unbounded real-valued variables. For simplex-valued variables, the CoDa approach uses an alternative expectation $\mathbb{E}^{\oplus}(x) = \int x f(x) d_A x$ that is based on the Aitchison measure d_A, instead of the more familiar Lebesgue measure d. In terms of the simplex expectation, the predictor (36) is indeed unbiased $\mathbb{E}^{\oplus}(S_{ij}) = \mathbb{E}^{\oplus}(\hat{S}_{ij})$, and this property also implies "logical consistency" in the sense of Naert and Bultez (1973).

Let $\mathcal{D}_i$ denote the values of all explanatory variables for individual i. The elasticities of the fitted shares with respect to the k^{th} compositional variable for individual i are then naturally given by the logarithmic derivative (evaluated at the observed data $\mathcal{D}_i$) of the fitted shares, i.e. $\frac{\partial \log \hat{S}_{ij}}{\partial \log \check{x}_{kil}}(\mathcal{D}_i) = \frac{\check{x}_{kil}}{\hat{S}_{ij}} \frac{\partial \hat{S}_{ij}}{\partial \check{x}_{kil}}(\mathcal{D}_i)$. Similarly the semi-elasticities for individual i with respect to the k^{th} scalar variable are given by the semi-logarithmic derivative (evaluated at the observed data $\mathcal{D}_i$) of the fitted shares, i.e. $\frac{\partial \log \hat{S}_{ij}}{\partial z_{ki}}(\mathcal{D}_i) = \frac{1}{\hat{S}_{ij}} \frac{\partial \hat{S}_{ij}}{\partial z_{ki}}(\mathcal{D}_i)$. Since the point of evaluation $\mathcal{D}_i$ is implicit in definition (36) of the fitted values we omit it for the remainder of the article. Note that the elasticities include the scaling factor $\check{x}_{kij}/\hat{S}_{ij}$ that justifies the usual interpretation of an elasticity as a relative change in $\hat{S}_{ij}$ induced by a relative change in $\check{x}_{kil}$. For the semi-elasticity the scaling factor is $1/\hat{S}_{ij}$ and the according interpretation of a semi-elasticity is a relative change in $\hat{S}_{ij}$ induced by a unit change in z_{kj}.

Semi-elasticities play no role in classical MCI models because these do not use type I variables. However, we could easily include type I variables using CoDa techniques. Simple differentiation rules then allow to derive

$$\frac{\partial \log \hat{S}_{ij}}{\partial z_{ki}} = \log(\hat{\gamma}_{kj}) - \sum\nolimits_{j=1}^{D} \log(\hat{\gamma}_{kj}) \hat{S}_{ij}, \tag{37}$$

which is observation dependent through $\hat{S}_{ij}$.

As seen in Section 4.2, the CoDa perspective of the MCI model employs an equivalent $\mathcal{T}$-space representation $(x_{ki\bullet}, G(\check{x}_{ki\bullet})) \in \mathcal{T}$ for compositional explanatory variables. Not surprisingly, this change of perspective leads to a decomposition of the elasticity into the corresponding contributions from relative and absolute information. This result is also coherent with the elasticity expressions of Morais and Thomas-Agnan (2021), where they correspond to the case of a YX-compositional model in which the total and the particular component of X vary together (type III variation). Here, we adjust these expressions to the prediction formula in (36), which allows transformed totals. It is convenient to work directly with the elasticity matrix that is defined in terms of $W_{\hat{S}_{i\bullet}} = (\mathbf{I}_D - \boldsymbol{\iota}_D \hat{S}'_{i\bullet})$:

$$\frac{\partial \log \hat{S}_{i\bullet}}{\partial \log \check{x}_{ki\bullet}} = W_{\hat{S}_{i\bullet}} \hat{\mathbf{H}}^*_k + W_{\hat{S}_{i\bullet}} \hat{\tau}^*_{k\bullet} \frac{\partial f_k(T(\check{x}_{ki\bullet}))}{\partial \log \check{x}_{ki\bullet}}. \tag{38}$$

The derivative on the right hand side of (38) equals one for $f_k(T(\check{x}_{ki\bullet})) = \log G(\check{x}_{ki\bullet})$. This allows to derive the MCI elasticities and their corresponding CoDa versions as

$$\begin{aligned} \frac{\partial \log \hat{S}_{i\bullet}}{\partial \log \check{x}_{ki\bullet}} &= W_{\hat{S}_{i\bullet}} \hat{\mathbf{H}}_k \\ &= W_{\hat{S}_{i\bullet}} \hat{\mathbf{H}}^*_k + W_{\hat{S}_{i\bullet}} \hat{\tau}^*_{k\bullet} \boldsymbol{\iota}'_D, \end{aligned} \tag{39}$$

which apply to all three versions of the MCI model with their corresponding parameter matrices $\hat{\mathbf{H}}_k$. The only difference to the expressions in Table 1 is a replacement of the parameters by their estimators and the original shares by the fitted shares. In both cases these elasticities are observation dependent.

Dargel and Thomas-Agnan (2023) show that differences of elasticities (and similarly of semi-elasticities) between two shares are observation independent and how the parameters $\hat{\mathbf{H}}^*_k$ and $\hat{\tau}^*_{k\bullet}$ of (39) can be directly interpreted.

5 Application and simulations

This section illustrates the methods outlined in the article using a fictitious but realistic geomarketing example. In the first subsection, we present this example in detail and use it to demonstrate an application of the MCI model. The second subsection presents a Monte Carlo simulation to evaluate the merits of the OLS and GLS estimators of the simple MCI model, under varying scenarios of heteroskedasticity.

5.1 Illustrative application

In our application, we model the market shares of nine commercial centers around the city of Toulouse in southern France. This model is based on the actual location of these commercial centers and local population statistics for the wider metropolitan area with 1.16 million inhabitants. Our study area is constituted by 428 so-called IRIS districts, which are geographic entities defined by the French national statistics institute (INSEE). Figure 2 shows a heat map for the population density in these districts alongside the location and the name of the commercial centers we are interested in.

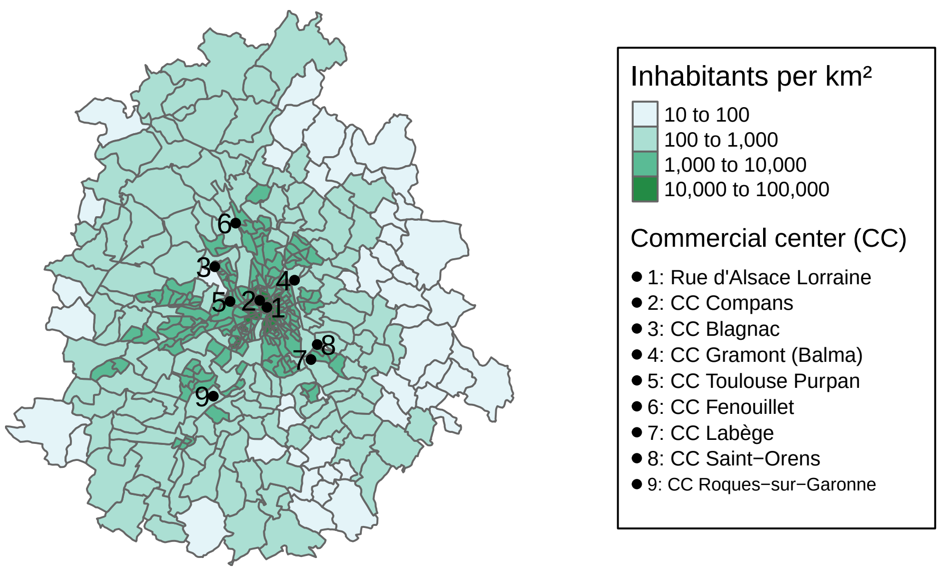

Figure 2: Heat map of the population density for 428 IRIS districts around Toulouse

The market shares of the $D = 9$ commercial centers are simulated from the attraction model:

$$A_{ij} = \alpha_j \cdot (\prod_{l=1}^{D} DIST_{ij} h_{jl}) \cdot \gamma_j^{POP_i} \cdot \delta_j^{INC_i} \cdot \check{\epsilon}_{ij}, \tag{40}$$

where $DIST_{ij}$ denotes the distance between commercial center j and the center of IRIS i. The population count (POP_i) and median income (INC_i) is taken from the French census provided by the INSEE website [2]. The values for the parameters $\alpha_\bullet$, $\gamma_\bullet$, $\delta_\bullet \in \mathcal{S}^D$, and the error distribution $\log \check{\epsilon_{i\bullet}} \sim \mathcal{N}(0, \Sigma)$ are chosen to resemble the estimation results from a model based on actual and confidential data for commercial centers in a comparable city in France. Figure 3 represents the market shares derived from our simulation.

[2]https://www.insee.fr

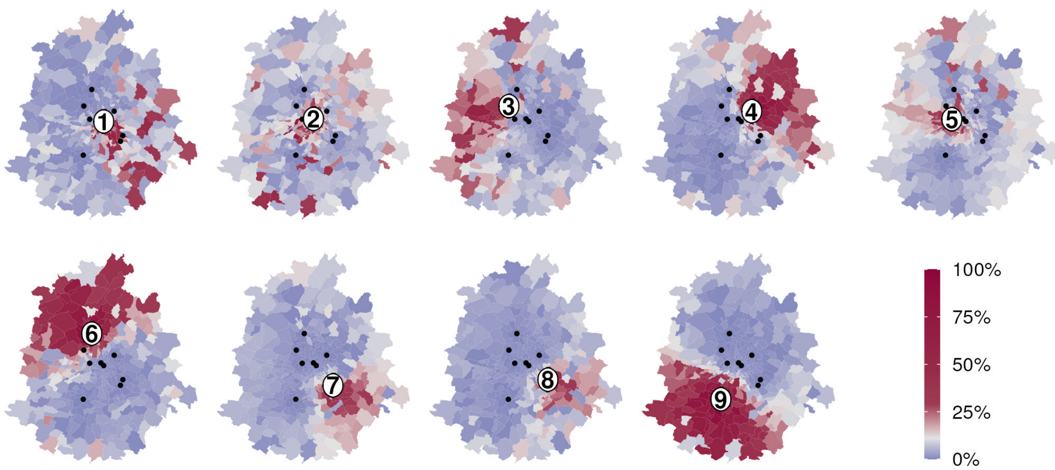

Figure 3: Market shares of the nine commercial centers in all regions (The shift from red to blue colors occurs at one ninth, where the market shares could be balanced. Since the data is heavily skewed towards zero, with a median share at around 4%, most of the color variation is placed in the lower percentages.)

We then derive a simple MCI model only based on distance and a differential effects version with $\mathbf{H} = \mathrm{diag}(h_1, ..., h_D)$, from the fully extended model in (40). The parameters of all three versions of the MCI model are then estimated by OLS in ilr space. Table 7 provides some global statistics for these estimations: In the second column, we recall the number of free parameters K that is 1 for the simple MCI, 33 for the differential effects model, and 88 for the fully extended MCI models (not counting the free elements in the covariance matrix Σ, which are identical in the three MCI models). The third column reports the compositional determination coefficient described in Van den Boogaart and Tolosana-Delgado (2013), which we named R^2_A because it can be derived from the Aitchison norm of the residuals. From the R^2_A, we may conclude that the simple effects MCI model explains about 64.9% of the compositional variation in the data. This figure raises to 70.1% and 78.2% in the differential and fully extended versions. However, since the number of parameters is different in the three models, we cannot rely on direct goodness of fit measures like the R^2_A for model selection. For this purpose, we additionally report the Akaike information criterion (AIC), which is maximized by the fully extended MCI model. This is expected since the data generating process (DGP) in (40) also corresponds to a fully extended MCI model.

Table 7: Model comparison

MCI Model	K	R^2_A	AIC
Simple effects	1	0.649	-7303
Differential effects	33	0.701	-4788
Fully extended	88	0.782	-4076

5.2 Simulation: alternative estimators under heteroskedasticity

Finally, we conduct a Monte Carlo study to understand how heteroskedasticity affects the performance of the OLS and GLS estimators described in Section 4.5. Our simulations are based on a simple effects MCI model with distance as the only explanatory variable, where we include error terms using the different structures of Table 6. As was already suggested by Nakanishi and Cooper (1974) we link the error structure to the number of customers trips $\breve{v}_{ij}$ originating in the region i and having shop j as the destination (numerator of the empirical market shares).

For our fictive example, we derive $\breve{v}_{ij}$ from the population of the IRIS and simple assumptions on the shopping behavior of an average individual.

Following the discussion in 4.5 we define different forms of heteroskedasticity from the general assumption $\mathbb{V}\mathrm{ar}(\log(\breve{\epsilon_{i\bullet}})) = \Lambda_i \Sigma \Lambda_i$, with $\mathrm{diag}(\sqrt{\lambda_{i1}}, ..., \sqrt{\lambda_{iD}})$. By linking λ_{ij} in different ways to the volumes $\breve{v}_{ij}$ we derive four different error structures:

$$\begin{aligned}
&(1)\ \text{Homoskedasticity}: && \lambda_{ij} = 1 \\
&(2)\ \text{Type C heteroskedasticity}: && \lambda_{ij} = \lambda_j = 1/(\textstyle\sum_{i=1}^{N} v_{ij}) \\
&(3)\ \text{Type I heteroskedasticity}: && \lambda_{ij} = \lambda_i = 1/(\textstyle\sum_{j=1}^{D} v_{ij}) \\
&(4)\ \text{Type IC heteroskedasticity}: && \lambda_{ij} = 1/v_{ij}.
\end{aligned} \tag{41}$$

Using all variance structures, we simulate 1000 models according to the attraction $A_{ij} = DIST_{ij}^{h} \breve{\epsilon_{ij}}$. In each simulation, we scale the four errors to impose the same total variance $\sum_{ij} (\breve{\epsilon_{i\bullet ij}})^2$. This allows attributing potential differences in estimation accuracy solely to the heteroskedasticity structure and not to changing signal-to-noise ratios. The parameter h is estimated from the ilr transformed model using the OLS and GLS estimators in (33) and (34), both implemented in a weighted and a non-weighted version. The four estimators can written using the GLS formula $\hat{h} = (Z'\Omega^{-1}Z)^{-1}(Z'\Omega^{-1}y)$, with the different specifications of the covariance matrix Ω. In all cases $\Omega = \mathrm{diag}(\Omega_i, ..., \Omega_N)$ is a block diagonal matrix with diagonal blocks Ω_i of dimension $(D-1) \times (D-1)$. The individual covariances Ω_i leading to the four estimators are given by

$$\begin{aligned}
&(1) \quad \text{OLS}: \Omega_i = \mathbf{I}_{D-1} \\
&(2)\ \text{WOLS}: \Omega_i = \lambda_i \mathbf{I}_{D-1} \\
&(3) \quad \text{GLS}: \Omega_i = \mathbf{V}'\Sigma\mathbf{V} \\
&(4)\ \text{WGLS}: \Omega_i = \lambda_i \mathbf{V}'\Sigma\mathbf{V}.
\end{aligned} \tag{42}$$

In practice, the GLS estimator is derived using a two-step procedure that is sometimes called feasible GLS (FGLS). For our application, we first estimate the covariance matrix $\widehat{\mathbf{V}'\Sigma\mathbf{V}}$ from the residuals of the OLS regression and then use this matrix in the GLS formula in the second step. Likewise, the WGLS uses the residuals of WOLS in the first step. Since our implementation of WOLS and WGLS is based on weights at the individual level, only Type I heteroskedasticity is adequately addressed. For the other heteroskedasticity types, this also allows assessing the effect of incorrectly specified weights on the estimators' performances.

Crossing the four estimators with the four error structures, we get 16 estimates $\hat{h}$ of the parameters, whose bias is given by $\hat{h} - h$. The distributions of these biases over all simulations are shown in Figure 4, where the four panels correspond to the error structures in (41), and the four estimators are shown on the x-axis of each panel. Since all distributions are centered around zero we may conclude that all estimators are indeed unbiased. However, the concentration of the bias distributions around zero differs substantially, indicating large differences in relative efficiency among the estimators. The dispersion of the bias distribution over the $B = 1000$ simulation can be summarized in the root mean square error, computed as $\mathrm{RMSE} = (\sum_{b=1}^{B} (\hat{h} - h)^2)^{0.5}/B$. To

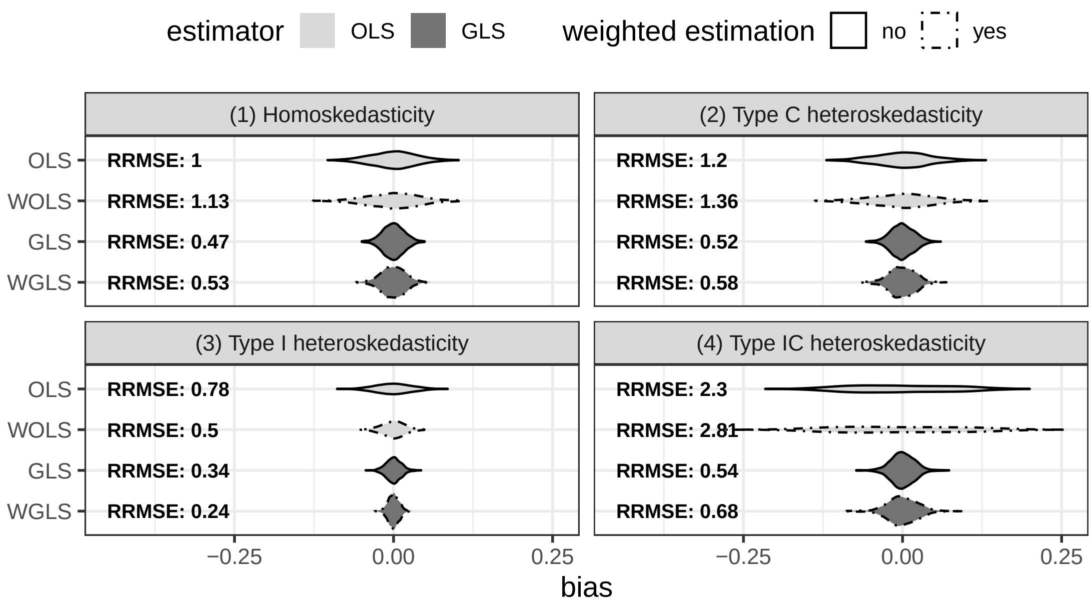

Figure 4: Accuracy of OLS and GLS estimation under four different error structures

facilitate the comparison among estimators, the graphic reports relative RMSE (RRMSE) figures, where the unweighted OLS in the case of homoskedastic errors serves as a reference. Comparing these RRMSE values reveals using (W)GLS instead of (W)OLS always at least doubles the estimation accuracy. In the case of type IC heteroskedasticity, the (W)GLS is more than four times more precise than (W)OLS. Another important comparison is between the weighted and non-weighted versions of the two estimators (OLS vs. WOLS and GLS vs. WGLS). Here we see that using weights only improves the estimation accuracy in the case of type I heteroskedasticity and leads to accuracy loss for the other error structures. In general, the efficiency gains (losses) of using correct (misspecified) weights appear small compared to the efficiency gains associated with using GLS instead of OLS.

6 Conclusion

This article demonstrates that all MCI models introduced by Cooper and Nakanishi (1989) are instances of CoDa regression. More specifically, we show that the fully extended MCI model, which nests the simple and differential effects MCI models as special cases, corresponds to a CoDa model with a log-transformed geometric total under an alternative parametrization. Recognizing this connection offers mutual benefits for both methodologies.

MCI models benefit from mathematical tools and theoretical guarantees developed in the CoDa literature. Using CoDa techniques, it is straightforward to introduce characteristics of the statistical individual as explanatory variables in the two largest MCI models. Additionally, we can solve the problem of singular covariance matrix encountered in the estimation of some MCI models by using an ilr instead of the clr transformation. This allows us to directly apply the GLS estimation to recover efficient estimates in the simple and differential effects MCI models. We confirm the potential efficiency gains of this estimation approach in Monte Carlo simulations. Studying the differential MCI model with CoDa tools also reveals that it is based on an unrealistic assumption of how the dependent shares are influenced by relative and absolute information contained in the compositional explanatory variables. To relax this assumption, we introduce diagonal CoDa models as a bridge between the simple MCI and general CoDa models. Moreover, the reparametrization that converts an MCI into a CoDa model separates the influence of relative and absolute information in the explanatory variables. This enables testing hypotheses regarding the influence of each information type. The separation of the influence of relative absolute information is mirrored in the elasticity expressions of in the MCI models to their CoDa counterparts.

CoDa regression also benefits in multiple ways from the MCI approach, which derives three model specifications as straightforward extensions of the intuitive gravity formula. These three models have varying degrees of complexity in terms of the number of parameters, where only fully extended MCI models match the complexity of CoDa regression models. The differential and simple effects MCI models use fewer parameters and also require less information for model estimation, which can be an advantage when data is scant. The simple effects MCI model is the most parsimonious with only one parameter per variable. It can also incorporate component-level information and make predictions for unobserved components, which is impossible with general CoDa models. Another interesting point of MCI models is their microeconomic foundation in random utility models, which means that the model specification is grounded in behavioral assumption of individual economic agents. This microeconomic foundation, in turn, provides justifications for heteroskedastic error structures, that carry over to the CoDa framework whenever the dependent shares are computed by aggregating individual choices. To account for heteroskedasticity we adjust the error structures of the traditional MCI model, making them more practical because they can be corrected by using a weighted least squares approach.

In summary, our findings show that the CoDa and marketing literates can realize mutual benefits by taking note of the other and we hope that this study contributes to the growing awareness that CoDa methods may be fruitfully applied across all quantitative social sciences.

References

Aitchison, J. and S. M. Shen (1980). "Logistic-Normal Distributions: Some Properties and Uses". In: *Biometrika* 67.2, pp. 261–272.

Aitchison, J. (1982). "The statistical analysis of compositional data". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 44.2, pp. 139–160.

— (1986). *The statistical analysis of compositional data*. eng. Monographs on statistics and applied probability. London New York: Chapman and Hall.

Boogaart, K. G. van den et al. (July 1, 2021). "Classical and Robust Regression Analysis with Compositional Data". In: *Mathematical Geosciences* 53.5. ZSCC: 0000016, pp. 823–858.

Chen, J., X. Zhang, and S. Li (2017). "Multiple linear regression with compositional response and covariates". In: *Journal of Applied Statistics* 12, pp. 2270–2285.

— (2018). "Heteroskedastic linear regression model with compositional response and covariates". In: *Journal of Applied Statistics* 12, pp. 2164–2181.

Coenders, G., J. A. Martín-Fernández, and B. Ferrer-Rosell (2017). "When relative and absolute information matter: compositional predictor with a total in generalized linear models". In: *Statistical Modelling* 17.6, pp. 494–512.

Converse, P. D. (1949). "New Laws of Retail Gravitation". In: *Journal of Marketing* 3, pp. 379–384.

Cooper, L. G. and M. Nakanishi (1989). *Market-share analysis: Evaluating competitive marketing effectiveness*. Springer Science & Business Media.

Dargel, L. and C. Thomas-Agnan (2023). “Share-ratio interpretations of compositional regression models”. In: *TSE Working Paper N°1456*.

Daunis-i-Estadella, J., J. J. Egozcue, and V. Pawlowsky-Glahn (2002). “Least squares regression in the simplex”. In: *Proceedings of IAMG*, pp. 411–416.

Egozcue, J. J. et al. (2003). “Isometric Logratio Transformations for Compositional Data Analysis”. In: *Mathematical Geology* 3, pp. 279–300.

Egozcue, J. J. et al. (2012). “Simplicial regression. The normal model”. In: *Journal of Applied Probability and Statistics* 6.1-2, pp. 87–108.

Ferrer-Rosell, B. et al. (2016). “Understanding low-cost airline users’ expenditure patterns and volume”. In: *Tourism Economics* 22.2, pp. 269–291.

Fiebig, D. G., R. Bartels, and D. J. Aigner (1991). “A random coefficient approach to the estimation of residential end-use load profiles”. In: *Journal of Econometrics* 3, pp. 297–327.

Friske, W. and S. Choi (2013). “Another look at retail gravitation theory: History, analysis, and future considerations”. In: *Academy of Business Disciplines Journal* 1, pp. 88–106.

Haines, G. H., L. S. Simon, and M. Alexis (1972). “Maximum Likelihood Estimation of Central-City Food Trading Areas”. In: *Journal of Marketing Research* 2, pp. 154–159.

Hron, K., P. Filzmoser, and K. Thompson (2012). “Linear regression with compositional explanatory variables”. In: *Journal of Applied Statistics* 39.5, pp. 1115–1128.

Hron, K. et al. (2022). “Weighting of Parts in Compositional Data Analysis: Advances and Applications”. In: *Mathematical Geosciences* 1, pp. 71–93.

Huff, D. L. (1962). “Determination of intra-urban retail trade areas”. In: *publication of Retail Estate Research Program*.

— (1963). “A Probabilistic Analysis of Shopping Center Trade Areas”. In: *Land Economics* 1, pp. 81–90.

— (1964). “Defining and Estimating a Trading Area”. In: *Journal of Marketing* 3, pp. 34–38.

Lee, M.-L. and R. K. Pace (2005). “Spatial Distribution of Retail Sales”. In: *The Journal of Real Estate Finance and Economics* 1, pp. 53–69.

Luce R. Duncan, 1.-2. (1959). *Individual choice behavior: A theoretical analysis*. New York: John Wiley & Sons.

McFadden, D. (1974). “Conditional logit analysis of qualitative choice behavior”. In: *in Paul Zarembka (editor), Frontiers in Econometrics New York: Academic Press*, pp. 105–42.

Morais, J. (2017). “Impact of media investments on brands’ market shares: a compositional data analysis approach”. PhD thesis. Toulouse School of Economics (TSE).

Morais, J. and C. Thomas-Agnan (2021). “Impact of Covariates in Compositional Models and Simplicial Derivatives”. In: *Austrian Journal of Statistics* 2, pp. 1–15.

Naert, P. A. and A. Bultez (1973). “Logically Consistent Market Share Models”. In: *Journal of Marketing Research* 3, pp. 334–340.

Nakanishi, M. and L. G. Cooper (1974). “Parameter Estimation for a Multiplicative Competitive Interaction Model: Least Squares Approach”. In: *Journal of Marketing Research* 3, pp. 303–311.

— (1982). “Simplified Estimation Procedures for MCI Models”. In: *Marketing Science* 3, pp. 314–322.

Nguyen, T. H. A. et al. (2021). “A simultaneous spatial autoregressive model for compositional data”. In: *Spatial Economic Analysis* 2, pp. 161–175.

Pawlowsky-Glahn, V., J. J. Egozcue, and D. Lovell (2014). “Tools for compositional data with a total:” in: *Statistical Modelling*.

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

Reilly, W. J. (1929). *Methods for the Study of Retail Relationships*. W.J. Reilly.

— (1931). *The law of retail gravitation*. W.J. Reilly.

Ruiz-Gazen, A. et al. (2023). "Detecting Outliers in Compositional Data Using Invariant Coordinate Selection Check for updates". In: *Robust and Multivariate Statistical Methods: Festschrift in Honor of David E. Tyler*, p. 197.

Van den Boogaart, K. G. and R. Tolosana-Delgado (2013). *Analyzing compositional data with R*. Vol. 122. Springer.

Wang, H. et al. (2013). "Multiple linear regression modeling for compositional data". In: *Neurocomputing*, pp. 490–500.

Zellner, A. (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias". In: *Journal of the American Statistical Association* 298, pp. 348–368.

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, from the French National Association of Research and Technology (ANRT) and from the market research agency BVA.