

MODELING MARKET-SHARES OR COMPOSITIONAL DATA: AN APPLICATION TO THE AUTOMOBILE MARKET

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

¹ *BVA, 52 rue Marcel Dassault, 92100 Boulogne-Billancourt et
joanna.morais@ut-capitole.fr*

² *TSE, 21 allée de Brienne, 31000 Toulouse et christine.thomas@tse-fr.eu*

³ *INRA, 1 place Pierre Viala, 34060 Montpellier et michel.simioni@supagro.inra.fr*

Résumé. Pour analyser l'impact des investissements marketing sur les ventes de véhicules, le contexte concurrentiel ne peut être ignoré. C'est pourquoi la modélisation des ventes en parts de marché et/ou des dépenses média en parts de voix semble intéressante. Cette contribution a pour but de présenter les méthodes statistiques adaptées pour la modélisation de parts ou de proportions, qui sont caractérisées par les contraintes de positivité et de somme unitaire. Par définition, les parts sont des données compositionnelles : une composition est un vecteur de parts d'un certain ensemble et qui porte une information relative. Deux principales approches répondent à cette problématique : les modèles de part de marché issus de la littérature économétrique et du marketing quantitatif, et l'analyse compositionnelle (CODA) provenant de la littérature statistique et initialement appliquée à la géologie. Le point commun de ces deux approches est qu'elles utilisent des transformations log-ratio pour prendre en compte les contraintes propres à ce type de données. Les différences entre l'une et l'autre reposent principalement sur les hypothèses faites sur la distribution des données compositionnelles elles-mêmes et sur la distribution des termes d'erreur du modèle. Une application au marché automobile ainsi qu'une comparaison des deux méthodes sont présentées.

Mots-clés. Parts de marché, Données compositionnelles, Automobile

Abstract. To analyze the impact of marketing investment on sales in the automobile market, the competitive context cannot be ignored. Thus the modeling in market-shares for sales and/or in shares-of-voice for media investments is of interest. This contribution aims to present and compare statistical modeling methods adapted for shares or proportions, which are characterized by the following constraints: positivity and sum equal to 1. By definition shares are compositional data: a composition is a vector of parts of some whole which carries relative information. Two major approaches address this question: the market-share models developed in the econometric literature and in the quantitative marketing, and the compositional data analysis (CODA) coming from the statistical literature and initially applied to geology. The common point of these two approaches is to use log-ratio transformations in order to model shares accounting for their constraints. The differences are mostly coming from the assumptions made on the distribution of the

data itself or on the error terms of the models. An application to automobile market and a comparison of these two methods are presented.

Keywords. Market-shares, Compositional data, Automobile

1 Introduction

Cette contribution a pour but de présenter les méthodes statistiques adaptées pour la modélisation de parts ou de proportions, qui sont caractérisées par les contraintes de positivité et de somme unitaire.

Par définition, les parts sont des données compositionnelles : une composition est un vecteur de parts d'un certain ensemble et porte une information relative. Pour une composition de D parts, si $D - 1$ parts sont connues, la D^{eme} est simplement égale à 1 moins la somme des $D - 1$ autres parts. Une composition de taille D peut donc être représentée dans un espace euclidien réel de dimension $D - 1$. Du fait de ces contraintes, les méthodes statistiques usuelles ne peuvent être utilisées directement. Le fameux *ceteris paribus* n'est plus valable pour l'interprétation des modèles contenant des données compositionnelles, car lorsqu'une part bouge, les autres bougent aussi.

Deux principales littératures permettent de tenir compte des contraintes propres à ce type de données : les modèles de parts de marché provenant de la littérature économétrique et du marketing quantitatif, et l'analyse compositionnelle (CODA) issue de la littérature statistique et initialement appliquée à la géologie.

2 Modèles de part de marché

Les modèles de parts de marché ont été développés dans les années 80 en grande partie par Nakanishi et Cooper (1988). Ils ont pour but de modéliser les parts de marchés de D différentes marques en utilisant des variables marketing comme explicatives (prix, publicité).

Ces modèles sont des versions agrégées des modèles de choix discrets tels que le logit multinomial (MNL), où la notion d' "utilité" d'une alternative est remplacée par la notion d' "attractivité" d'une marque. La spécification de l'attractivité de la marque j est une expression multiplicative des variables explicatives décrivant la marque j . La part de marché de la marque j est définie comme son attractivité relative par rapport aux concurrents, c'est-à-dire comme son attractivité divisée par la somme des attractivités de toutes les marques présentes sur le marché. Cela justifie que l'on appelle souvent ces modèles MCI pour Multiplicative Competitive Interaction models.

L'estimation de ces modèles se fait généralement sur la version log-linéarisée des parts, par MCO ou MCG. Cependant, on peut prouver que sous certaines hypothèses, une estimation par maximum de vraisemblance est possible. Cette partie de la littérature ne considère que le cas où seule la variable à expliquer est une variable compositionnelle.

3 Modèles compositionnels

L'analyse des données compositionnelles (CODA) a été initiée par Aitchison (1986). Depuis les années 90, d'importants développements mathématiques ont été menés par un groupe de chercheurs de l'Université de Girona en Espagne (V. Pawlowsky-Glahn, J.J. Egozcue, C. Barcelo, J.A. Martin Fernandez).

Les premières applications concernaient la géologie avec l'objectif, par exemple, d'analyser la composition d'une roche en fonction des éléments chimiques qui la composent. De façon plus générale, l'analyse CODA s'intéresse à l'information relative entre les composants d'une composition, pour laquelle le total n'a pas de sens ou n'est pas l'objectif de l'analyse.

Une composition de D parts appartient au simplexe \mathcal{S}^D . La géométrie adaptée pour les compositions est la géométrie d'Aitchison, et non la géométrie euclidienne. Les compositions peuvent être transformées en coordonnées via des transformations log-ratio de façon à être représentées dans l'espace euclidien \mathbb{R}^{D-1} . Les méthodes statistiques classiques sont ensuite applicables sur ces coordonnées. Il existe trois principales transformations utilisées en CODA : la transformation ALR (additive log-ratio), la transformation CLR (centered log-ratio) et la transformation ILR (isometric log-ratio), qui ont chacune leurs avantages et leurs inconvénients.

CODA regroupe différents outils pour analyser des compositions : des outils graphiques (diagrammes ternaires, biplot de coordonnées), des outils analytiques (ACP compositionnelle) ainsi que des modèles de régression compositionnels. Les modèles de régression compositionnels sont adaptés aux cas où la variable à expliquer et/ou les variables explicatives sont compositionnelles. L'estimation se fait sur les coordonnées appartenant à l'espace euclidien, par MCO généralement.

4 Application

Ces méthodes sont appliquées à un jeu de données sur le marché automobile français dans le but d'expliquer l'impact des investissements média sur les parts de marchés des véhicules (la variable à expliquer est compositionnelle). Nous avons à notre disposition pour chaque véhicule et par mois le volume des ventes, le prix catalogue, les investissements media par canal (TV, Presse, Radio, Digital, Affichage extérieur, Cinéma), et un

ensemble de caractéristiques du véhicule.

Deux autres modèles sont également testés. Le premier considère la composition des investissements média par canal d'un véhicule comme variable explicative du volume des ventes de ce véhicule. Le deuxième considère la composition des investissements média des véhicules (c'est-à-dire les parts de voix des différents véhicules) comme variable explicative de la composition des ventes, c'est-à-dire des parts de marchés des différents véhicules.

Pour chaque spécification du modèle, on s'intéresse à l'effet marginal de chaque canal de communication et aux interactions entre eux. L'objectif final est de pouvoir faire des recommandations au constructeur automobile concernant le volume d'investissement media et sa répartition entre les différents canaux.

5 Comparaison et résultats

Une comparaison des deux familles de modèles est faite en fonction des propriétés des estimations, des hypothèses sur les distributions des termes d'erreur, et de l'interprétation des coefficients estimés. Il en ressort que les deux méthodes ont des points positifs et négatifs et peuvent s'enrichir l'une l'autre.

En particulier, nous avons prouvé que la propriété d'indépendance des alternatives non pertinentes (IIA) des modèles de choix discrets (que l'on retrouve dans les modèles de parts de marché) est équivalente à la propriété de cohérence des sous-compositions des modèles CODA. Or, tandis que la propriété IIA a souvent été critiquée pour son manque de réalisme, menant au développement de modèles de choix discrets et de parts de marché plus flexibles (nested logit, GEV), la propriété de cohérence des sous-compositions ne semble jamais avoir été remise en cause dans la littérature CODA.

Par ailleurs, CODA propose la transformation ILR qui est une projection des données compositionnelles dans l'espace euclidien en utilisant une base orthonormale dans le simplexe. Cette transformation a de bonnes propriétés mathématiques et permet d'utiliser un grand nombre d'outils statistiques, contrairement aux transformations ALR et CLR. Or, on peut montrer que les transformations utilisées dans les modèles de parts de marché correspondent aux transformations ALR et CLR, mais ILR n'est jamais appliquée dans cette littérature.

De plus, les hypothèses faites sur la structure de covariance entre les composants dans la spécification MNL sont moins flexibles (covariance négative imposée) que dans le cas de CODA où la covariance positive entre deux composants d'une D -composition avec $D > 2$ n'est pas impossible.

Enfin, concernant les résultats de la modélisation, les deux méthodes mènent à des interprétations différentes et apportent des informations complémentaires en fonction de la transformation log-ratio utilisée.

Bibliographie

- [1] Nakanishi, M. et Cooper, L.G. (1982), Simplified Estimation Procedures for MCI Models, *Marketing Science*, 1, 314–322.
- [2] Cooper, L.G. et Nakanishi, M. (1988), *Market-Share Analysis: Evaluating Competitive Marketing Effectiveness*, Springer.
- [3] Haaf, C.G., Michalek, J., Morrow, W.R. et Liu, Y. (2014), Sensitivity of Vehicle Market Share Predictions to Discrete Choice Model Specification, *Journal of Mechanical Design*, 136.
- [4] Aitchison, J. (1986), *The statistical analysis of compositional data*, Chapman and Hall.
- [5] Van Den Boogaart, K.G. and Tolosana-Delgado, R. (2013), *Analysing Compositional Data with R*, Springer.
- [6] Pawlowsky-Glahn, V., Egozcue, J.J. and Tolosana-Delgado, R. (2015), *Modeling and Analysis of Compositional Data*, John Wiley & Sons.