Impact of socioeconomic factors on nutritional diet in Vietnam from 2004 to 2014
New insights using compositional data analysis.

Thi Huong TRINH [1], **Joanna MORAIS**[2]

[1] *TSE, 21 allée de Brienne, 31000 Toulouse et trinhthihuong@tmu.edu.vn*
[2] *TSE, 21 allée de Brienne, 31000 Toulouse et joanna.morais@bva.fr*

Abstract. This contribution aims to analyze the impact of socioeconomic factors, like food expenditure level and urbanization, on diet patterns in Vietnam, from 2004 to 2014. Contrary to the existing literature, we focus on the diet balance in terms of macronutrients consumption (protein, fat and carbohydrate) and we take into account the fact that the volumes of each macronutrient are not independent. In other words, we are interested in the shares of each macronutrient in the total calorie intake. We use the compositional data analysis (CODA), adapted for shares data, to describe the evolution of diet patterns over time, and to model the impact of household characteristics on the macronutrient shares vector. We compute food expenditure elasticities of macronutrient shares, and we compare them to classical elasticities for macronutrient volumes and total calorie intake. Our results are consistent with the rest of the literature, but they have the advantage to highlight the substitution effects between macronutrients in the context of nutrition transition.

Keywords. Compositional regression models, expenditure elasticity, macronutrient shares, diet pattern, Vietnam

Résumé. Cette contribution a pour but d'analyser l'impact des facteurs socioéconomiques, tels que le niveau de dépense alimentaire et l'urbanisation, sur les habitudes alimentaires au Viêt Nam, de 2004 à 2014. Contrairement à la littérature existante, nous nous intéressons à l'équilibre alimentaire en termes de macronutriments (protéines, lipides, glucides), et nous prenons en compte le fait que les volumes consommés de chaque macronutriment ne sont pas indépendants. Autrement dit, nous nous intéressons à la part de chaque macronutriment dans l'apport calorique total. Nous utilisons l'analyse de données compositionnelles (CODA), adaptée pour les données de parts, pour décrire l'évolution des habitudes alimentaires au cours du temps, et pour modéliser l'impact des caractéristiques des ménages sur le vecteur des parts de macronutriments. Nous calculons les élasticités des parts de macronutriments à la dépense alimentaire, et nous les comparons aux élasticités classiques, calculées pour les volumes de macronutriments, et pour l'apport calorique total. Nos résultats sont cohérents avec le reste de la littérature, mais ils ont l'avantage de mettre en lumière les effets de substitution entre les macronutriments dans un contexte de transition nutritionnelle.

Mots-clés. Modèles de régression compositionnels, élasticité à la dépense, parts de macronutriments, habitudes alimentaires, Viêt Nam

Résumé long

Food security and nutrient affordability have become a main concern of governmental and non-profit organizations due to their effects on health and economic development. Economic development and urbanization in developing countries have affected global diet, leading to many empirical researches focusing on food sources leading to categories such as vegetable, staple cereals, meat, etc. Widespread trends include an increase of animal-source foods, sugar, oils processed food and staple cereal refining, as a consequence of rising incomes and urbanization (Global Food Policy Report (2017)). Recently, Santeramo and Shabnam (2015) do a review of estimated income elasticities of calories, macronutrients and micronutrients with a total of 26 studies (but only 5 focus on macronutrients) in many different countries. Through meta-analysis, the authors found that calories intake and proteins intake are more income–inelastic than fat intake and micronutrients intake. In order to assess the relationship between nutrients consumption and socioeconomic characteristics, several regressions are usually performed in parallel with the same explanatory variables and the different nutrients as dependent variables (see Liaskos and Lazaridis (2003), You et al. (2016)). These specifications do not take into account the fact that the three macronutrients constitute the whole diet of each household (or individual) so the volumes of consumed macronutrients are not independent. Moreover, the computation of consumed macronutrient volume can be criticized when using household survey data due to the impossibility to take into account losses and wastes in food preservation, preparation and consumption(see Porkka et al. (2013)). Household survey data have also limitations due to recalled bias and self-reported measures as emphasized by Deaton (1997). Assuming that these two problems affect the computation of all macronutrients in the same way, we can expect the shares of the macronutrients not to be affected by the consecutive biases, contrary to volumes.

The aim of this study is to contribute to this literature by analyzing the evolution of diet patterns in Vietnam, focusing on macronutrient shares in the diet, instead of macronutrient volumes. This study uses six waves of the Vietnam Household Living Standard Survey (VHLSS), from 2004 to 2014. In order to measure the impact of household characteristics on the diet balance, and in particular to calculate the food expenditure elasticity of each macronutrient share, we use compositional regression models, coming from the compositional data analysis (CODA) literature. CODA is a well established field of statistics with diverse fields of application (see Pawlowsky-Glahn and Antonella (2011)). A composition is a vector of shares, where shares are called the components. In our study, diet components are the proportions of protein, fat and carbohydrate in the household calorie intake. As far as we know, our study is the first to use CODA tools to analyze the

evolution of diet patterns.

We first use descriptive tools of CODA, like ternary diagrams, to analyze the evolution of the three components over years, and according to the place of residence (urban or rural, regions). We observe that households tend to increase their proportion of fat at the expense of carbohydrate, while the share of protein stays almost stable.

We then use a compositional regression model (Pawlowsky-Glahn and Buccianti, 2011) where the dependent variable is a vector of shares (the composition of macronutrients) and explanatory variables are classical variables which depend only on the observations (household characteristics: food expenditure, urban/rural and region of residence, size of the household, education level, ethnicity and gender of the head of the household). The corresponding compositional regression model is the following (one by period):

$$\begin{aligned}\mathbf{S}_i &= \mathbf{a} \bigoplus_{k=1}^{K} X_{ki} \odot \mathbf{b_k} \oplus \boldsymbol{\epsilon}_i \\ &= \mathbf{a} \oplus \log(Exp)_i \odot \mathbf{b_1} \oplus Urban_i \odot \mathbf{b_2} \oplus HSize_i \odot \mathbf{b_3} \oplus Educ_i \odot \mathbf{b_4} \\ &\oplus Ethnic_i \odot \mathbf{b_5} \oplus Gender_i \odot \mathbf{b_6} \oplus Area_i \odot \mathbf{b_7} \oplus \boldsymbol{\epsilon}_i\end{aligned} \tag{1}$$

where $\mathbf{S} = (S_P, S_F, S_C)'$ is the vector of protein, fat, and carbohydrate shares, and the index i denotes the i^{th} household. $\mathbf{S}, \mathbf{a}, \mathbf{b_k}, \boldsymbol{\epsilon} \in \mathcal{S}^3$ are compositional (they belong to the simplex of dimension 3, $\mathcal{S}^3$) and X_k are classical explanatory variables.

Due to the fact that the components $(S_P, S_F, S_C)'$ are subject to some constraints (the components are positive and sum up to 1), classical regression models cannot be used directly. Thus, shares are transformed, using an isometric log–ratio (ILR) transformation for example. After ILR transformation, Model (1) can be written in $3 - 1 = 2$ equations (for each period):

$$\begin{aligned}S^*_{j,i} &= a^*_j + \sum_{k=1}^{K} b^*_{j,k} X_{ki} + \epsilon^*_{j,i} \qquad \text{for } j = 1, 2 \\ &= a^*_j + b^*_1 \log(Exp)_i + b^*_2 Urban_i + b^*_3 HSize_i + b^*_4 Educ_i \\ &+ b^*_5 Ethnic_i + b^*_6 Gender_i + b^*_7 Area_i + \epsilon^*_{j,i}\end{aligned} \tag{2}$$

where $S^*_j, a^*_j, b^*_{j,k}, \epsilon^*_j$ are the j^{th} ILR coordinates of $\mathbf{S}, \mathbf{a}, \mathbf{b_k}, \boldsymbol{\epsilon}$. We perform this transformed model made up of two equations separately for the 6 years of observation, using OLS and the assumption that $\boldsymbol{\epsilon}$ follows a Gaussian distribution in the simplex (that is ϵ^* follows a Gaussian distribution). Note that, the estimation of the coefficients of the model in the simplex (1) can be obtained by inverse transformation from the estimated coefficients of the transformed model (2).

In order to determine if our compositional model is relevant and reliable to explain the shares of protein, fat and carbohydrate, we checked several items such as the significance of explanatory variables, the R^2 of the model, and the good properties of residuals.

From the coefficients of the compositional model (1) over years, we highlight the fact that the size of the household ($HSize$) impacts a lot the diet balance: the larger the household is, the larger the carbohydrate share is and the smaller the fat share is. In contrast, the larger the food budget (Exp) is, the smaller the carbohydrate share is and the larger the fat share is. The share of protein is not really affected by the explanatory variables, excepted that households living in the South East and Mekong River Delta tend to consume more protein calories.

Finally, in order to interpret share models, the elasticity is an appropriate tool because it measures the relative impact of an explanatory variable on a share, after a relative change of this explanatory variable. We calculate food expenditure elasticities of macronutrient shares (see Morais et al. (2017) for formula). We conclude that the fat share is the most elastic macronutrient with respect to food expenditure. However, this elasticity tends to decrease across time. The food expenditure elasticity of carbohydrate share is negative, and the one of protein share is positive and stable across time. In order to compare our elasticities with the existing literature, we perform the usual double-log regression models explaining the consumption volume of each macronutrient and of the per capita calorie intake, using the same explanatory variables than in model 1). We obtain positive food expenditure elasticities for all macronutrients volumes, meaning that a higher food budget tends to increase the consumed quantity of food, for all macronutrients. Results in volumes are consistent with results in shares in the sense that fat is the most elastic macronutrient and carbohydrate is the less elastic. The shares elasticities reflect the substitution effect between fat and carbohydrate.

This study can help policymakers to understand Vietnamese behavior on food choices. This is an important topic due to links between the nutrition transition and the rise of non-communicable diseases, such as obesity and heart disease.

bibliography

Deaton, A., (1997), "The analysis of household surveys: a microeconometric approach to development policy," *World Bank Publications*.
International Food Policy Research Institute, (2017), "2017 Global Food Policy Report", *Washington, DC: International Food Policy Research Institute*, doi.org.10.2499/9780896292529.
Liaskos, G. and Lazaridis, P. (2003), "The demand for selected food nutrients in Greece: The role of socioeconomic factors", *Agricultural Economics Review*, 4(2).
Morais, J. and Thomas-Agnan, C. and Simioni, M. (2017), "Interpreting the impact of explanatory variables in compositional models", *TSE Working Paper*.

Pawlowsky-Glahn, V. and Buccianti, A. (2011), "Compositional data analysis: Theory and applications", *John Wiley & Sons*.
Porkka, M., Kummu, M., Siebert, S., Varis, O. (2013), "From Food Insufficiency towards Trade Dependency: A Historical Analysis of Global Food Availability", *PLoS ONE* 8(12): e82714. doi:10.1371/journal.pone.0082714
Santeramo, F. G., and Shabnam, N. (2015), "The income-elasticity of calories, macro- and micro-nutrients: What is the literature telling us?", *Food Research International*, 76, 932-937.
You, J., and Imai, K. S., et Gaiha, R. (2016), "Declining nutrient intake in a growing China: Does household heterogeneity matter?", *World Development*, 77, 171-191.