Macronutrient balances and body mass index: a new insight using compositional data analysis with a total at various quantile orders

LE Danh Tuyen[1], Huong TRINH THI[2,3,*], Christine THOMAS-AGNAN[2],
Michel SIMIONI[4], Ty BEAL[5] and NGUYEN Duy Son[1]

[1]National Institute of Nutrition (NIN), Hanoi, Vietnam

[2]Toulouse School of Economics, University of Toulouse Capitole, France

[3]Department of Mathematics and Statistics, Thuongmai University, Hanoi, Vietnam

[4]MOISA, INRA, University of Montpellier, Montpellier, France

[5]Global Alliance for Improved Nutrition (GAIN), Washington DC, USA

May 2018

*Corresponding author:trinhthihuong@tmu.edu.vn

Macronutrient balances and body mass index: a new insight using compositional data analysis with a total at various quantile orders

Abstract

The impact of food consumption on diseases is complex due to the confounding effects between macronutrients on a diet. We are interested in the impact of both the volume and the proportions of macronutrients on body mass index. We develop a compositional regression model with a total at various quantile orders. Then we compute the elasticities of BMI with respect to each macronutrient. Our methodology is applied to Vietnamese adults from 18 to 60 years of age. The results first reveal significant impacts of some socio–economics factors, such as the total as geometric mean, age, gender, job type, no drinking status and geographical region. All elasticities of BMI with respect to each macronutrient increase as BMI increases until a threshold (BMI=20) and then remain stable.

Keywords: Macronutrient balances, Body mass index (BMI), a compositional regression model with a total, quantile regression, semi-elasticity.

1 Introduction

The Nutrition transition has occurred in both developing countries and developed countries (Popkin (2006)). There is an increase of the double burden of malnutrition characterized by the coexistence of undernutrition along with overweight and obesity, called diet-related noncommunicable diseases (WHO, 2016). The World Health Organization (WHO, 2018) declares that "the fundamental cause of obesity is an energy imbalance between calories consumed and calories expended". Researchers from several disciplines have focused on the relationship between diet composition and disease (see some review in Hooper et al., 2001; Riera-Crichton and Tefft, 2014; Hall et al., 2011; Albar et al., 2014). These findings remain controversial due to the complex associations between total energy intake, physical activity, body size and the prevalence of disease and due to limitations of the datasets.

Total energy intake consists of macronutrient and micronutrient and each specific nutrient is correlated with the total energy intake: i.e each nutrient provides directly a part of the energy intake. A person who has a larger total energy intake also consumes larger volumes of all specific nutrients, on average. In addition, the contribution of each macronutrient in a total energy intake (measured by kcal) may have a different effect. Several empirical studies show the impact of a diet with the same amount of caloric content but different compositions of macronutrients on health (for example, Camacho and Ruppel, 2017). In the US, "Dietary Guidelines for Americans", issued by the US Department of Agriculture and the US Department of Health, Education and Welfare (now the Department of Health and Human Services) in 1980, recommended a reduction in the consumption of the share of total macronutrients attributable to fat and saturated fat, and a reduction in the absolute consumption of cholesterol. To compensate, the guidelines recommended increasing the share (in grams) of carbohydrates in the total consumption of calories because carbohydrates contain less than half the number of calories per ounce than fats (Cohen et al., 2015).

From a mathematical perspective, to control confounding in epidemiologic analysis, Wacholder et al. (1994), Willett et al. (1997), Trichopoulou et al. (2002), and Randi et al. (2007) have discussed various methods of adjustment for total energy intake, such as: nutrient density model, standard multivariate model, nutrient residual (energy-adjusted) model, partition regression. However, "the specific effects of individual macronutrients and the generic effect of energy cannot be disentangled by multivariate analysis" (Wacholder et al., 1994). All of the above regression models still fail to solve the comprehensive effect of total energy from that of each component of energy, i.e protein, fat, and carbohydrate. We recall the compositional nature of the dietary intake in Kcal, i.e total energy = energy from protein + energy from fat + energy from carbohydrate. Thus the four variables: total energy, energy from protein, energy from fat and energy from carbohydrate are perfectly linearly related. Recently, Leite (2016), Dumuid et al. (2017) and Trinh et al. (2018) propose to use a compositional data approach (CoDa) to analyze dietary data and show its advantages over the usual methods. Leite and Prinelli (2017) has applied this approach to analyze the associations between macronutrient balances and diseases. This study, conducted in 1992–1993 from the database of the Italian Bollate Eye study, focuses on adults of between 40 and 70 years of age. The authors discuss a diet which consists of three macronutrients and then go into further detail by widening the composition, including now: saturated versus unsaturated fats.

Our empirical study focuses on Vietnam. This country has experienced a strong economic development after Doi Moi reforms in the 1980s. Now, Vietnam is a lower middle-income country. Due to an increase in income and changes in other socioeconomic characteristics, there is an increase in per capita calorie intake (Trinh et al., 2017). In addition, the Vietnamese diet patterns have also changed with a larger proportion of animal source, fat and protein intake (Nguyen and Popkin, 2004; Trinh et al., 2018). However, Vietnam still faces the double burden of malnutrition as many developed countries. According the the United Nations, Vietnam ranks always among the thirty-six countries with the highest stunting rates in the world. Among Vietnamese 18-65 years old, the prevalence of overweight and obesity increased from 2.0% in 1992 to 5.2% in 2002 using a national survey (Tuan et al., 2008). Similarly, Nguyen and Hoang (2018) show that the prevalence of overweight and obesity increased from 2.3% in 1993 to 15% in 2015 in the same age group. The figures in urban sites are much higher than in rural sites. Cuong et al. (2007) show that 26.2% (resp. 6.4%) of adults living in the urban area of Ho Chi Minh City[1] were already considered as overweight (resp. obese) in 2004. Prevalence of obese among children under 5 has increased much faster than among adults. In the 2000-2010 period, the prevalence of overweight and obesity increased from 0.6% (resp. 0.9%, 0.5%) to 5.6% in the whole country (resp. in urban areas, in rural ones). In 2011, 14% of children (resp. 8.6%, 4.4%) in Vietnam under 5 were still stunted (resp. underweight, thin). In addition, both figures for children under 5 are higher in big cities (Huynh et al., 2007).

This paper contributes to the literature by focusing on the impact of the macronutrient diet and

[1] This is the biggest city in Vietnam

socio-economics characteristics, such as age, gender, job, and living location, on the body mass index (BMI) of 18–65 years old adults by using the 2009 - 2010 wave of the General Nutrition Survey in Vietnam. We contribute to the literature in various ways:

- We apply CoDa regression with a total variable to take into account both the relative importance of each macronutrient in the whole diet and total energy.
- We perform regression both for the average BMI to obtain a general relationship and for the 15% and 90% conditional quantiles of BMI in order to be more precise for vulnerable groups. These limits correspond to underweight and overweight thresholds in the marginal distribution of BMI.
- We adapt semi-elasticity computations in the above two regression models to obtain a direct interpretation of a change of the volume of a given macronutrient on BMI.

2 Descriptive analysis of the nutrition issue of adults aged 18–60 years old in Vietnam using compositional data analysis

We use the General Nutrition Survey 2009 - 2010 in Vietnam which was conducted by the Vietnam National Institute of Nutrition (NIN) (National Institute of Nutrition, 2010). This cross-sectional survey is representative of the Vietnamese population and has been conducted every ten years since 1981. Household dietary intake is based on a 24-hour dietary recall. Food categories in quantities are converted into calorie intake and grams using the Food composition table for Vietnam in 2007. We use the average daily intake of households. In this survey, we only focus on adults between 18 to 60 years of age. Diet intake can be divided according to macronutrient sources. From a macronutrient component perspective in terms of Kcal, we divide the diet intake into three macronutrients: protein (P), fat (F) and carbohydrates (C). From a macronutrient component perspective in term of grams, we divide the diet intake into four macronutrients: protein (P), fat (F), carbohydrate (C)

and fiber[2] (Fi). Table 1 displays some summary descriptive statistics of the Vietnamese diets and their macronutrient intakes. In terms of kcal, the average per capita calorie intake (PCCI) is 1923.9 Kcal: note that this number follows the recommendation of NIN[3]. In terms of grams, per capita per day food intake is around 440 grams. In addition, the volumes of fiber are quite small compared to other macronutrients (6 (g) per person per day and it only accounts for 1.4% of total diet intake). The average total number of fiber grams is lower than in the recommendation but this number is reasonable in Vietnam due to the fact that there is only a small quantity of fiber in ordinary polished rice – the most common rice in Vietnamese meals[4].

Table 1: Descriptive statistics of Vietnamese diets and their macronutrients composition

Variable	Description	Value
N	Number of observations	15035
$PCCI$	Per capita calorie intake (Kcal)	1923.9 (501.8)
$PCCIg$	Per capita per day food intake (gram)	440 (114.6)
V_P	Volume of calorie obtained from protein (Kcal)	318.5 (98.7)
V_F	Volume of calorie obtained from fat (Kcal)	338.4 (177.1)
V_C	Volume of calorie obtained from carbohydrate (Kcal)	1267 (369.7)
S_P	Share of calorie obtained from Protein (%)	16.7 (3.5)
S_F	Share of calorie obtained from Fat (%)	17.4 (7.5)
S_C	Share of calorie obtained from Carbohydrate (%)	65.9 (8.8)
Vg_P	Volume of intake per day from protein (gram)	79.6 (24.7)
Vg_F	Volume of intake per day from fat (gram)	37.6 (19.7)
Vg_C	Volume of intake per day from carbohydrate (gram)	316.8 (92.4)
Vg_{Fi}	Volume of intake per day from fiber (gram)	6 (3.1)
Sg_P	Share of intake per day from protein (gram)	18.3 (4.1)
Sg_F	Share of intake per day from fat (gram)	8.6 (4.2)
Sg_C	Share of intake per day from carbohydrate (gram)	71.7 (7)
Sg_{Fi}	Share of intake per day from fiber (gram)	1.4 (0.6)

Standard errors are in parenthesis

Figure 1 shows the prevalence of obesity and underweight in 2010 in Vietnam, based on the cut-off of BMI classification of World Health Organization[5]. 16% of Vietnamese adults are underweight and about 7% are overweight. These figures are less than in developed countries but they are increasing every year.

Figure 2 reports the ratios of macronutrient intakes expressed in logarithm, or log ratios for both Kcal and grams measurements. The first figure shows log ratios when macronutrient are measured in Kcal. The median of the two boxplots of log ratio $log(\frac{S_P}{S_C})$ and log ratio $log(\frac{S_F}{S_C})$ are negative.

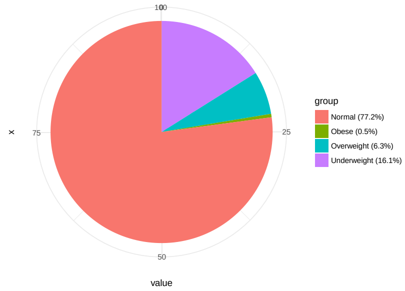
[2]Fiber do not provide any calories.

[3]A household with energy intake below 1800 Kcal will be considered as a low energy intake.

[4]Ordinary polished rice has 0.4g Fiber per 100g.

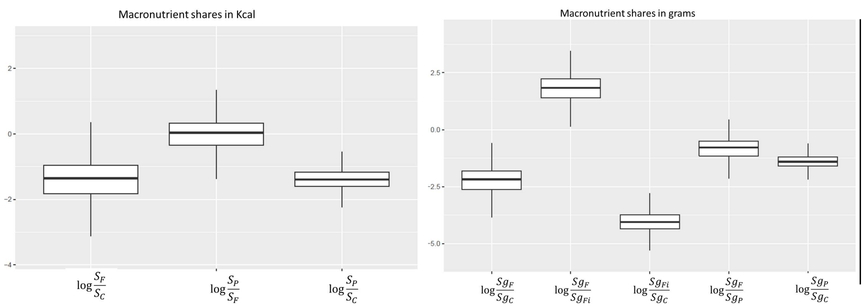
[5]Body Mass Index (BMI)is defined as the weight in kilograms divided by the square of the height in meters (kg/m^2). According to WHO (2004), people with a BMI less than 18.49 are underweight. The normal range of BMI is 18.50 - 24.99. People with a BMI larger than 25 are overweight. In addition, people are obese if BMI is larger than 30

Figure 1: Prevalence of obesity and underweight in Vietnam - 2010.



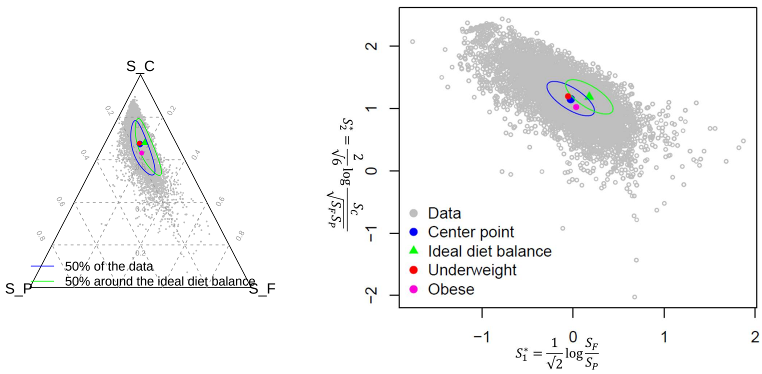
Carbohydrates represent the largest source of calories in the Vietnamese diet. Although having similar median values, the log ratio $log(\frac{S_P}{S_C})$ exhibits more variation than the log ratio $log(\frac{S_F}{S_C})$. The boxplot in the middle shows that the median value of log ratio $log(\frac{S_F}{S_P})$ is close to zero and that its distribution seems symmetric around zero. The right figure shows the log ratios of the four macronutrients when they are measured in grams. In the left figure, the median log ratios between protein, fat versus carbohydrate, i.e $log(\frac{Sg_P}{Sg_C})$ and $log(\frac{Sg_F}{Sg_C})$ are always negative but their absolute values are larger than when we measure macronutrient intakes in Kcal. The shares of fiber are very small compared to other macronutrients shares.

Figure 2: Boxplots of macronutrients log shares ratios.



Recently, Ministry of Health (2012) has issued recommendations on the ideal balanced diet for the Vietnamese population (in Kcal), namely (Protein:Fat:Carbohydrate = $14\% : 18\% : 68\%$). A ternary diagram can be used to plot this ideal diet and compare it with the observed center point of the sample. A ternary diagram is the adequate representation of shares data, incorporating information that these shares sum to one. The left panel of Figure 3 shows the scatterplot of the observed vectors of shares and the three center points for the whole population and for the two vulnerable groups: obese and underweight, respectively. Ellipses are added to show where half of the population is located around these center points in the simplex (Mahalanobis distance level curves). The same is done for the ideal balanced diet. The right panel of Figure 3 is simply a transformation of the previous one using ilr coordinates. Its lecture is easier as data are projected onto a plane (Van den Boogaart, K. G. and Tolosana-Delgado, R., 2013), but the interpretation stays the same whatever representation of the data we use. . . In our data, the center point is not far from the ideal point. The line passing through the two center points for underweight and obese groups is parallel to the edge S_C–S_F of the triangle which means that underweight and obese groups have a similar proportion of protein. However, the diets of the obese group has a larger fat share (similarly, smaller carbohydrate share) than the underweight group.

Figure 3: Plotof centers diets of the whole population, of he overweight people and of the obese people compared to the "ideal" diet balance (S_P=14%,S_F=18%,S_C=68%) in a ternary diagram in the simplex and in ILR coordinates.

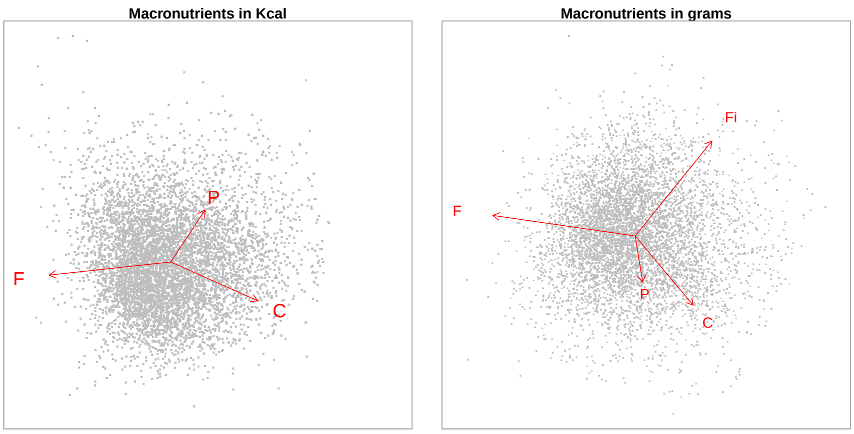


Covariance biplots in Figure 4 show a comprehensive compositional exploratory analysis of the three macronutrient shares (in Kcal) and of the four macronutrient shares (in grams). The left biplot has a 3–part composition, the biplot explains 100% of the variance. The three components protein, fat, and carbohydrate are very long and they point towards different directions (making angles of approximately 90^0 to 120^0). The log-ratio corresponding to the longest link is that of Fat versus

Carbohydrate. The right biplot has a 4–part components, i.e adding share of fiber. Three group links (P, F, Fi) points towards different directions as in the left biplot. In the above descriptive statistics of fiber, we see the small amount of fiber in the diet. But the three group links (P, F, Fi) indicate that the fiber share, although small, is very important in the diet.

The Other protein (OP) and Carbohydrate (C)links appear to be close to each other, thus revealing possibly a collinearity between (OP) and (C). The sets of rays: protein–fat and carbohydrate–fiber appear to be orthogonal, thus revealing two possibly uncorrelated log ratios, i.e $\log(\frac{Sg_P}{Sg_F})$ and $\log(\frac{Sg_{Fi}}{Sg_C})$.

Figure 4: Covariance biplot of a principal component analysis of the macronutrient shares for each year.



3 A compositional data perspective on studying the associations between macronutrient balances and BMI

3.1 A total as geometric mean as an determinant of obesity

As suggested byPawlowsky-Glahn et al. (2015), Coenders et al. (2017), Ferrer-Rosell and Coenders (2017), we use a total variable defined as the geometric mean of macronutrients volumes. This total corresponds to an average value in the space of the logarithm of absolute volumes values. The choice of logarithm in this total has some advantages: (1) it naturally converts an absolute positive value to a value belonging to $\mathbb{R}$, (2) it allows interpreting regression coefficients using the link between coefficients and elasticities (in economics studies) or odd ratios (epidemiologic studies).

We use the following two total variables as geometric means denoted by T (resp. Tg) when macronutrients are measured in Kcal (resp. in grams).

$$\ln T = \frac{1}{3}\big[\ln(V_P) + \ln(V_C) + \ln(V_F)\big]$$

$$\ln Tg = \frac{1}{4}\big[\ln(Vg_P) + \ln(Vg_C) + \ln(Vg_F) + \ln(Vg_{Fi})\big]$$

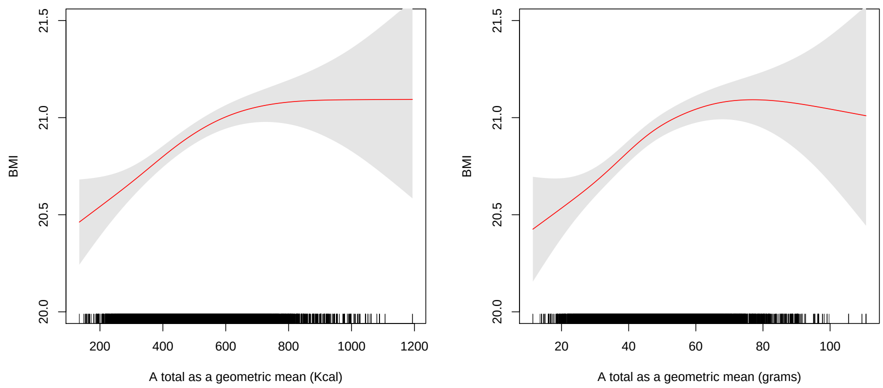
Table 2 shows the descriptive statistics of these two totals. In terms of Kca, the average of this geometric mean is equal to 497.5 Kcal. This number is smaller than one third of PCCI, i.e $\frac{1923.9}{3} = 641.3$ Kcal. The difference between T and $\frac{PCCI}{3}$ is due to the logarithm. Similarly, an average of Tg of macronutrients in grams is 46.8 (g). This number is smaller than one fourth of $PCCIg$, i.e $\frac{440}{4} = 110$ (g).

Table 2: Descriptive statistics for the total variables

Variable	Description	Average value
T	Total in kcal	497.5 (149.2)
Tg	Total in gram	46.8 (14.1)

Figure 5 shows a scatterplot of BMI and the total variable, together with a semi-parametric regression curve (Wood, 2017). There figures show a potential non-linear relationship between BMI and totals. In both figures, at the beginning of the range of totals, BMI indicators increase as totals increase. Then, when totals exceeeds a threshold, say 600Kcal and 55 grams, BMI tends to remain constant.

Figure 5: BMI indicator as a function of total



3.2 Various regression models with compositional predictor and a total

Compositional data describe parts of a whole and, consequently, convey only relative information. A model has been proposed in the so-called CODA (compositional data analysis) literature, which is the standard method of statistic to deal with a positive vector which carries only relative information (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011; Pawlowsky-Glahn et al., 2015). In our approach, we are interested in the BMI indicator, denoted by Y_i, $Y_i \in \mathbb{R}, Y_i > 0$ and several explanatory variables. Among the explanatory variables, we will include the macronutrient shares of a diet. Due to the constant sum of the fitted components (equal to 1 here), classical regression models cannot be used directly. For example, the three macronutrient shares (in Kcal) (S_P, S_F, S_C) have the following constraint $S_P + S_C + S_C = 1$. Each vector of shares (S_P, S_F, S_C) belongs to the simplex S^3. To overcome this difficulty, shares are transformed, using an isometric log-ratio (ILR) transformation (Egozcue and Pawlowsky-Glahn, 2003). We will illustrate our strategy in the case of three macronutrient shares, a similar strategy will be applied in the case of four macronutrient shares (in grams). For three components in the simplex, the Ilr transformation transforms them into two isometric log ratios (Ilr) coordinates $Ilr1$ and $Ilr2$ that vary in $\mathbb{R}$.[6] Importantly, coefficients of compositional regression in simplex are invariant to the choice of sequential binary partition.

$$Ilr1 = \sqrt{\frac{2}{3}} log \frac{S_C}{\sqrt{S_P S_F}}, \quad Ilr2 = \sqrt{\frac{1}{2}} log \frac{S_P}{S_F}$$

Using the ilr coordinates, a linear compositional model can be formulates to estimate the impact of some explanatory variables Z_i and (S_P, S_F, S_C) on the average of the outcome variable Y_i

$$\mathbb{E}(Y_i) = \alpha + \beta Ilr1 + \gamma Ilr2 + a.Z_i \qquad (EC)$$

where $\mathbb{E}(Y_i)$ denotes the expectation of the conditional distribution of Y_i given the covariates. Here, Z includes several explanatory variables described in Table 7. They are total expenditure per week (ExpWeek), age, gender, ethnicity, education levels, job (farmer or non-farmer), plain, drinking beer status, smooking status.

The coefficients of model (EC) are estimated using ordinary least squares. Examples of applications of these models in social sciences can be found in Muller et al. (2016), Leite (2016), Leite and Prinelli (2017).

The above compositional model ignores the information about total abundance of all components while focusing only on relative information between shares. In this epidemiologic study, the totals, i.e per capita calorie intake or per capita per day food intake, are also important due to their impact on BMI. Then, we adapt a compositional model including these totals, initially proposed by Pawlowsky-Glahn et al. (2015) and Coenders et al. (2017), called the T–space model. In this T–space model, the total is defined as in the previous subsection such that its logarithm equals to the geometric mean of the volumes.

We can then formulate a compositional model with the two ilr coordinates together with a total

[6] The Ilr coordinates we are using here are based on a sequential binary partition: Carbohydrate vs protein and fat, fat vs protein. We can apply alternative sequential binary partitions, such as protein vs fat and carbohydrate, fat vs carbohydrate

by

$$\mathbb{E}(Y_i) = \alpha + \beta Ilr1 + \gamma Ilr2 + T_i.\delta + a.Z_i \qquad (EF)$$

In addition, the classical linear model explaining Y with the total only is nested in model (EF) and can be used to estimate the impact of the total on the outcome variable Y. Thus, our "total only" regression model will be

$$\mathbb{E}(Y_i) = \alpha + T_i.\delta + a.Z_i \qquad (ET)$$

Finally, model (EC), (EF) and (ET) can be extended to the quantile regression framework, as in Koenker and Hallock (2001). Here, we are interested in the estimation of the impact of explanatory variables Z_i and (S_P, S_F, S_C) on the τ^{th} conditional quantile of the outcome variable Y_i, so that we write

$$\mathbb{Q}_\tau(Y_i) = \alpha_\tau + \beta_\tau CIlr1 + \gamma_\tau Cilr2 + a.Z_i \qquad (QC)$$

$$\mathbb{Q}_\tau(Y_i) = \alpha_\tau + \beta_\tau CIlr1 + \gamma_\tau Cilr2 + T_i.\delta + a.Z_i \qquad (QF)$$

$$\mathbb{Q}_\tau(Y_i) = \alpha_\tau + T_i.\delta + a.Z_i \qquad (QT)$$

where $\mathbb{Q}_\tau$ denotes a $\tau-$ quantile level of Y_i given the explanatory variables. The interpretation of the coefficients in the three quantile models are similar to that in the classical regression models (EC), (EF) and (ET).

To obtain a comprehensive and complex impact of diet pattern on BMI, the above models are also applied to the case of four shares $Sg_F, Sg_P, Sg_C, Sg_{Fi}$.

Figure 8 shows the density of $log(BMI)$ which has a shape similar to a normal density. This figure supports our choice of using $log(BMI)$ as an outcome variable and shows that its distribution is approximately gaussian. To decide which quantile order to focus on, we use the cut-off for underweight (BMI is less than 18.5) and for overweight (BMI is larger than 25) which are based on BMI Asian populations (WHO, 2004). We then fit quantile regression at 15% quantile and 90% quantile levels of the marginal distribution of BMI. Table 3 shows all potential regression models at various quantile orders. To choose among these various models, we use an analysis of variance table (resp. an analysis of deviance table) comparing conditional mean linear models (resp. quantile regression (Koenker and Bassett, 1982). Table 4 shows the corresponding F–value and significance levels of the tests. For all these various models defined at mean or or for quantiles, results show that the full model is always preferred. This strategy is similar to Coenders et al. (2017) when choosing among alternative compositional models.

Muller et al. (2016), Leite (2016), Leite and Prinelli (2017) gives interpretation of the coefficients of the Ilr coordinates. We will rather adopt the same kind of interpretation as Morais et al. (2018) and adapting it to our case, i.e. using semi-elasticities in the next subsection to have direct interpretation of the impact of each macronutrient. Table 5 shows the coefficients of (traditional, non-compositional) explanatory variables[7] for both two kinds of food intake measures.

[7]These coefficients are dependent on the choice of Ilr coordinates.

Table 3: Strategy to study the associations between macronutrient balances and BMI

Shares	**Macronutrient measured in Kcal** S_P, S_F, S_C	**Macronutrient measured in gram** $Sg_P, Sg_F, Sg_C, Sg_{Fi}$
A total	$\ln T = \frac{1}{3}\left[\ln(V_P) + \ln(V_C) + \ln(V_F)\right]$	$\ln Tg = \frac{1}{5}\left[\ln(Vg_P) + \ln(Vg_C) + \ln(Vg_F) + \ln(Vg_{Fi})\right]$
Ilr coordinates	$Ilr1, Ilr2$	$Ilrg1, Ilrg2, Ilrg3$
Models **Conditional mean**	$\mathbb{E}(Y_i) = \alpha + T_i.\delta + a.Z_i \quad (ET)$ $\mathbb{E}(Y_i) = \alpha + \beta Ilr1 + \gamma Ilr2 + a.Z_i \quad (EC)$ $\mathbb{E}(Y_i) = \alpha + \beta Ilr1 + \gamma Ilr2 + T_i.\delta + a.Z_i \quad (EF)$	$\mathbb{E}(Y_i) = \alpha + Tg_i.\delta + a.Z_i \quad (ET)$ $\mathbb{E}(Y_i) = \alpha + \beta Ilrg1 + \gamma Ilrg2 + \nu Ilrg3 + a.Z_i \quad (EC)$ $\mathbb{E}(Y_i) = \alpha + \beta Ilrg1 + \gamma Ilrg2 + \nu Ilrg3 + Tg_i.\delta + a.Z_i \quad (EF)$
τ quantile	$\mathbb{Q}_\tau(Y_i) = \alpha_\tau + T_i.\delta + a.Z_i \quad (QT)$ $\mathbb{Q}_\tau(Y_i) = \alpha + \beta Ilr1 + \gamma Ilr2 + a.Z_i \quad (QC)$ $\mathbb{Q}_\tau(Y_i) = \alpha + \beta Ilr1 + \gamma Ilr2 + T_i.\delta + a.Z_i \quad (QF)$	$\mathbb{Q}_\tau(Y_i) = \alpha_\tau + Tg_i.\delta + a.Z_i \quad (QT)$ $\mathbb{Q}_\tau(Y_i) = \alpha + \beta Ilrg1 + \gamma Ilrg2 + \nu Ilrg3 + a.Z_i \quad (QC)$ $\mathbb{Q}_\tau(Y_i) = \alpha + \beta Ilrg1 + \gamma Ilrg2 + \nu Ilrg3 + Tg_i.\delta + a.Z_i \quad (QF)$

Table 4: Analysis of Variance table for alternative models

Models	Macronutrient in Kcal		Macronutrient in Kcal	
	Full vs Total	Full vs Composition	Full vs Total	Full vs Composition
Conditional mean	12.76***	25.31***	9.35***	25.84***
$\tau = 0.15$ quantile	2.91.	20.16***	2.49.	25.78***
$\tau = 0.9$ quantile	5.03**	9.61**	4.23**	8.49**

Note: ., *, **, and *** mean significant at 10%, 5%, 1% and 0.1%, respectively

Table 5: Multiple linear regression analysis of the relationship between the first ilr coordinate and the total as geometric mean and BMI.

Regressors	Macronutrient in Kcal			Macronutrient in grams		
	* $\tau = 0.15$	Mean	$\tau = 0.9$	$\tau = 0.15$	Mean	$\tau = 0.9$
(Intercept)	1.19***	1.502***	2.000***	1.154***	1.512***	2.057***
A total as geometric mean						
T	0.01***	0.1***	0.01**	0.1***	0.1***	0.1**
Age						
log(Age)	0.935***	0.797***	0.58***	0.954***	0.797***	0.551***
$\log^2$(Age)	-0.125***	-0.102***	-0.068***	-0.128***	-0.102***	-0.064***
Expenditure per week						
log(EXP)	10^{-5}	0.001*	0.002	0.0002	0.002*	0.002.
Gender						
Female	-0.019***	-0.017***	-0.011*	-0.019***	-0.018***	-0.011*
Ethnicity						
Kinh	-0.009*	-0.016***	-0.023***	-0.008.	-0.015***	-0.024***
Education levels						
Secondary school	0.003	-0.005*	-0.01*	0.003	-0.005*	-0.011*
High school	0.006	0	-0.005	0.006	0	-0.006
Univeristy	0.006	-0.002	-0.009	0.006	-0.003	-0.01
Job						
Non–Farmer	0.015***	0.019***	0.023***	0.015***	0.019***	0.023***
Smoking status						
Non smoker	0.015***	0.018***	0.017**	0.015***	0.018***	0.016***
Drinking beer status						
1–4 times per months	-0.005	0.002	0.004	-0.004	0.002	0.003
(no drinking)	-0.019***	-0.009**	-0.003	-0.019***	-0.009**	-0.003
Coastline	0.005	0.005	0	0.006	0.006.	0.001
Geographical region						
Midlands-mountainous	0.005	-0.015**	-0.034*	0.006	-0.015**	-0.031*
Low mountains	0.015***	-0.002	-0.029***	0.015***	-0.002	-0.029***
High mountains	0.023***	-0.002	-0.033***	0.023***	-0.003	-0.034***

The coefficients of T are multiplied by 1000.

Note: ., *, **, and *** mean significant at 10%, 5%, 1% and 0.1%, respectively

The interpretation of Table 5 is as follows

- For both measures (in Kcal and grams), the coefficients of totals as geometric mean are significant positive in all regression models.
- The logarithm of age is significant and positive and the square of the logarithm of age is significant and negative, i.e BMI increases as age increases, then after a given threshold, BMI tends to decrease.[8] For example, in terms of macronutrients in Kcal, the thresholds at 15% quantile, mean and 90% quantile are 42.1, 49.7 and 71.1 years old. It is quite interesting that the threshold of the obese group is the highest number.
- The coefficients of the logarithm of expenditure are all positive but significant only for the conditional mean regression.
- When gender takes the female level, its coefficient is significant and negative. It means that women tend to have a lower BMI than men conditionally on other characteristics.

[8]When we interpret the coefficients of Age, we assume that all other variables remain constant. Then, the peak of Age is equal to $\exp(\frac{a1}{-2*a2})$ where $a1$ and $a2$ are coefficients of $log(Age)$ and $log^2(Age)$.

- In comparison to minority level of ethnicity, the coefficient of the Kinh ethnicity level is significant and negative. It means that the Kinh people have a higher BMI indicator than minority people on average conditional on other characteristics.
- The coefficients of secondary school levels are significant and negative at the mean and at the 90% quantile level. All coefficients of other education levels are insignificant.
- The coefficients of the non-farmers job level are significant and positive. Then, on average, non–farmers tend to have higher BMI than farmers at all regression levels conditional on other characteristics. These results are reasonable since in this study, job type, i.e farmers or non–farmers, plays the role of activities levels. People who have more intensive activities will consume more energy.
- About the drinking beer status, it is interesting that the coefficient of the non-drinking beer people is significant and negative. Then, on average, non–drinking people have smaller BMI than drinking people conditional on other characteristics.
- The coefficients of geographical regions have various signs (positive and negative, significant or insignificant). These mixed effects are due to the fact that there is a confounding effect between the impact of the regions and that of the other characteristics.

3.3 Elasticities computation in these compositional models

In order to interpret the share regression models, Morais et al. (2018) suggest to use elasticities to overcome complex interpretations of the parameters in ILR coordinates regressions. These elaticities are similar to odds ratios which are popular in medical research. The elasticity quantifies the relative variation of an outcome variable due to the relative variation of an explanatory variable, measured in percentage. We adapt the elasticity calculation of Morais et al. (2018) to the case of our preferred model, i.e a the compositional model with a total. In our case, since the dependent variable is not a composition, the adapted tool is a semi-elasticity but since our ourcome is the log of BMI, the semi elasticity of the outcome corresponds to the elasticity of BMI. The mathematical computation of the semi-elasticities are given in the Appendix 5. We also prove that these elasticity formulas are invariant to the choice of ilr coordinates. In the case of three macronutrient shares (in Kcal), the elastiscity of BMI with respect to the volumes of macronutrients are given by:[9]

$$\frac{\partial Y}{\partial lnV_C} = \beta\sqrt{\frac{2}{3}} + \frac{\delta T}{3}, \qquad \frac{\partial Y}{\partial lnV_P} = \frac{-\beta}{\sqrt{6}} + \frac{\gamma}{\sqrt{6}} + \frac{\delta T}{3}, \qquad \frac{\partial Y}{\partial lnV_F} = \frac{-\beta}{\sqrt{6}} - \frac{\gamma}{\sqrt{6}} + \frac{\delta T}{3}$$

and

$$\frac{\partial Y}{\partial \ln T} = \delta$$

Table 6 displays the average elasticities of BMI with respect to macronutrients at various quantile orders and for both units: Kcal and grams. Results indicate that

- The elasticities of BMI with respect to carbohydrate are always negative.
- Positive semi-elasticities are associated to fat, protein and fiber.
- Generally, BMI is more elastic to protein.

Table 6: Average elasticities of BMI with respect to macronutrients at various quantile orders

Macronutrient	Macronutrient in Kcal			Macronutrient in Kcal		
	$\tau = 0.15$	Mean	$\tau = 0.9$	$\tau = 0.15$	Mean	$\tau = 0.9$
Carbohydrate	-0.003	-0.011	-0.012	-0.006	-0.014	-0.016
Fat	0.005	0.005	0.003	0.004	0.004	0.004
Protein	0.020	0.024	0.030	0.020	0.021	0.025
Fiber				0.006	0.008	0.006

Figure 6 and 7 show the average elasticities as a function of BMI for all macronutrients in the two units obtained by smoothing the scatterplot of elasticity as a function of BMI. Average elasticities of BMI with respect to all macronutrients in both units increase rapidly with BMI and then become stable after a threshold around BMI $= 20$, meaning that individuals with a low BMI are more affected by the composition of their nutrition.

[9]These formulas are based on sequential binary partitions: Carbohydrate vs protein and fat, protein vs fat.

Figure 6: Elasticity (in Kcal) as function of BMI.

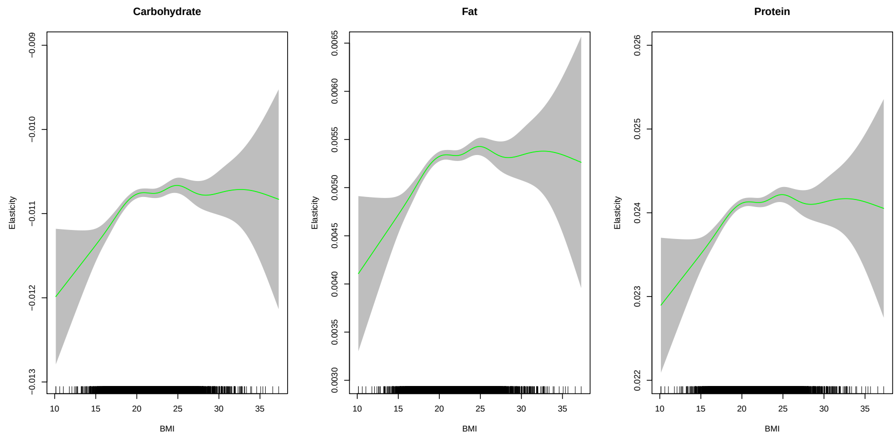
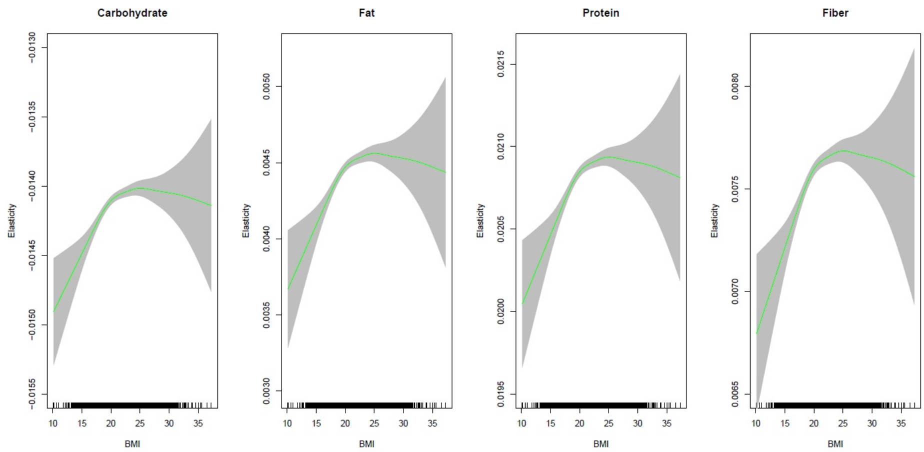


Figure 7: Elasticity (in grams) as function of BMI.



4 Conclusion

This article focuses on the relationship between food consumption and the BMI indicator. This is an important issue since there is a close link between eating habits and the occurrence of chronic diseases. These topics are currently analyzed by many multi-disciplinary researchers but the findings remain controversial due to the complex associations between total energy intake, physical activity and the metabolism of each person.

There are many different approaches to find the relationship between the BMI indicator and the diet intake in epidemiologic analysis. However, the current models in the literature fail to disentangle the comprehensive effect of total energy from that of each component of energy. Our proposal is based on the compositional data approach (CoDa). There are only few empirical epidemiologic studies using the CoDa approach, such as Leite (2016), Dumuid et al. (2017) and Trinh et al. (2018). This advanced methodology has much to bring to epidemiology.

We propose to estimate various regression models: compositional models, total only models and compositional models with a total. We use geometric mean of macronutrient shares as total, as a determinant variable of the BMI indicator. These models are estimated at the conditional mean and two quantile orders: 15% quantile regression (corresponding to the underweight cut–off), and 90% quantile regression (corresponding to the overweight cut–off). Macronutrients are measured in two different units: Kcal and grams. From an analysis of variance comparing alternative models, we conclude that the full model, i.e compositional model with a total, is preferred whatever the conditional mean or quantile regressions.

Average elasticity values increase as BMI increases until a threshold ($BMI = 20$). Some of these results could be due to confounding effects. Protein could be acting as a proxy for unhealthy behaviors: individuals who consume higher amounts of protein may be wealthier, less active, smoke more, and consume more processed foods, sugar-sweetened beverages, and total energy than individuals who consume lower amounts of protein. It is impossible to account for all of the potential confounders (i.e., unhealthy behaviors) using the available dataset.

References

Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman and Hall London.

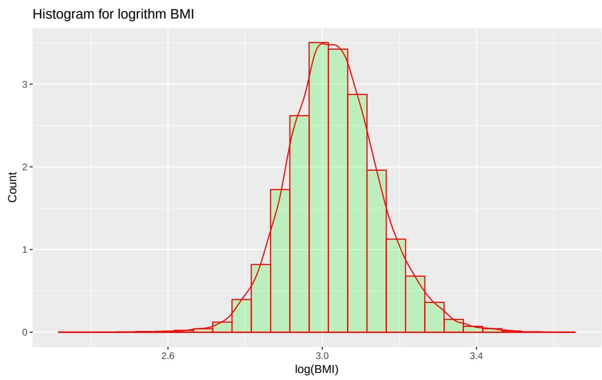
Albar, S. A., N. A. Alwan, C. E. Evans, and J. E. Cade (2014). Is there an association between food portion size and bmi among british adolescents? *British Journal of Nutrition 112*(5), 841–851.

Camacho, S. and A. Ruppel (2017). Is the calorie concept a real solution to the obesity epidemic? *Global health action 10*(1), 1289650.

Table 7: General Nutrition Survey for 2009-2010 description variables

Variable	Description	Value
N	Number of observation	15035
BMI	Body mass index (BMI)	20.9 (2.6)
ExpWeek	Total expenditure per week (thousand vnd)	3416.4 (3211)
Age	Year olds	36.4 (11.4)
Gender	Male	51.25 %
	Female	48.75 %
Ethnic	Minority	15.32 %
	Kinh	84.68 %
Educ	Below primary	35.1 %
	Secondary school	34.04 %
	High school	19.92 %
	Univeristy	10.95 %
Job	Farmer	43.25 %
	Non-Farmer	56.75 %
Plain	Delta	53.47 %
	Coastline	8.92 %
	Midlands-mountainous	3.15 %
	Low mountains	21.31 %
	High mountains	13.15 %
Drinking beer status	Bear1 (more than 1 time per week)	16.72 %
	Bear2 (1-4 times per months)	15.46 %
	Bear0 (no drinking)	67.82 %
Smooking status	Smook	25.3 %
	Nonsmook	74.7 %

Figure 8: Density of log(BMI).



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. *Statistical Modelling 17*(6), 494–512.

Cohen, E., M. Cragg, A. Hite, M. Rosenberg, B. Zhou, et al. (2015). Statistical review of us macronutrient consumption data, 1965–2011: Americans have been following dietary guidelines, coincident with the rise in obesity. *Nutrition 31*(5), 727–732.

Cuong, T. Q., M. J. Dibley, S. Bowe, T. T. M. Hanh, and T. T. H. Loan (2007). Obesity in adults: an emerging problem in urban areas of ho chi minh city, vietnam. *European journal of clinical nutrition 61*(5), 673–681.

Dumuid, D., T. E. Stanford, J. Martin-Fernández, Ž. Pedišić, C. A. Maher, L. K. Lewis, K. Hron, P. T. Katzmarzyk, J. P. Chaput, M. Fogelholm, et al. (2017). Compositional data analysis for physical activity, sedentary time and sleep research. *Statistical methods in medical research*, 0962280217710835.

Egozcue, J. and V. Pawlowsky-Glahn (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology 35*(3), 279–300.

Ferrer-Rosell, B. and G. Coenders (2017). Airline type and tourist expenditure: Are full service and low cost carriers converging or diverging? *Journal of Air Transport Management 63*, 119–125.

Hall, K. D., G. Sacks, D. Chandramohan, C. C. Chow, Y. C. Wang, S. L. Gortmaker, and B. A. Swinburn (2011). Quantification of the effect of energy imbalance on bodyweight. *The Lancet 378*(9793), 826–837.

Hooper, L., C. D. Summerbell, J. P. Higgins, R. L. Thompson, N. E. Capps, G. D. Smith, R. A. Riemersma, and S. Ebrahim (2001). Dietary fat intake and prevention of cardiovascular disease: systematic review. *Bmj 322*(7289), 757–763.

Huynh, T. T. D., M. J. Dibley, D. Sibbritt, and T. M. H. Tran (2007). Prevalence of overweight and obesity in preschool children and associated socio-demographic factors in ho chi minh city, vietnam. *Pediatric Obesity 2*(1), 40–50.

Koenker, R. and G. Bassett (1982). Tests of linear hypotheses and l_1 estimation. *Econometrica: Journal of the Econometric Society*, 1577–1583.

Koenker, R. and K. F. Hallock (2001). Quantile regression. *Journal of economic perspectives 15*(4), 143–156.

Leite, M. and F. Prinelli (2017). A compositional data perspective on studying the associations between macronutrient balances and diseases. *European journal of clinical nutrition 71*(12), 1365.

Leite, M. L. (2016). Applying compositional data methodology to nutritional epidemiology. *Statistical methods in medical research 25*(6), 3057–65.

Ministry of Health (2012). *National Nutrition Strategy for 2011-2020, with a Vision Towards 2030*. Hanoi: Medical Publishing House.

Morais, J., C. Thomas-Agnan, and M. Simioni (2018). Interpreting the impact of explanatory variables in compositional models. *To appear in Austrian Journal of Statistics*.

Muller, I., K. Hron, and E. Fiserova (2016). Interpretation of compositional regression with application to time budget analysis. *arXiv* (1609.07887).

National Institute of Nutrition (2010). *General Nutrition Survey 2009 - 2010*. Medical Publishing House.

Nguyen, M. T. and B. M. Popkin (2004). Patterns of food consumption in vietnam: effects on socioeconomic groups during an era of economic growth. *European journal of clinical nutrition 58*(1), 145.

Nguyen, T. T. and M. V. Hoang (2018). Non-communicable diseases, food and nutrition in Vietnam from 1975 to 2015: The burden and national response. *Asia Pacific Journal of Clinical Nutrition 27*(1), 19–28.

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

Pawlowsky-Glahn, V., J. Egozcue, and R. Tolosana-Delgado (2015). *Modeling and analysis of compositional data*. John Wiley & Sons.

Pawlowsky-Glahn, V., J. J. Egozcue, and D. Lovell (2015). Tools for compositional data with a total. *Statistical Modelling 15*(2), 175–190.

Popkin, B. M. (2006). Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases (NCDs). *American Journal of Clinical Nutrition 84*, 289–298.

Randi, G., C. Pelucchi, S. Gallus, M. Parpinel, L. Dal Maso, R. Talamini, L. S. Augustin, A. Giacosa, M. Montella, S. Franceschi, et al. (2007). Lipid, protein and carbohydrate intake in relation to body mass index: an italian study. *Public health nutrition 10*(3), 306–310.

Riera-Crichton, D. and N. Tefft (2014). Macronutrients and obesity: revisiting the calories in, calories out framework. *Economics & Human Biology 14*, 33–49.

Trichopoulou, A., C. Gnardellis, V. Benetou, P. Lagiou, C. Bamia, and D. Trichopoulos (2002). Lipid, protein and carbohydrate intake in relation to body mass index. *European journal of clinical nutrition 56*(1), 37.

Trinh, H. T., J. Morais, C. Thomas-Agnan, and M. Simioni (2018). Relations between socio-economic factors and nutritional diet in vietnam from 2004 to 2014: New insights using compositional data analysis. *Statistical methods in medical research*, 0962280218770223.

Trinh, T. H., M. Simioni, and C. Thomas-Agnan (2017). A fresh look at the nutrition transitionin vietnam using semiparametric modeling. *TSE working paper Sept*(17-842).

Tuan, N., P. Tuong, and B. Popkin (2008). Body mass index (bmi) dynamics in vietnam. *European journal of clinical nutrition 62*(1), 78.

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

Wacholder, S., A. Schatzkin, L. S. Freedman, V. Kipnis, A. Hartman, and C. C. Brown (1994). Can energy adjustment separate the effects of energy from those of specific macronutrients? *American journal of epidemiology 140*(9), 848–855.

WHO (2016). The double burden of malnutrition: policy brief.

WHO, E. C. (2004). Appropriate body-mass index for asian populations and its implications for policy and intervention strategies. *Lancet (London, England) 363*(9403), 157.

Willett, W. C., G. R. Howe, and L. H. Kushi (1997). Adjustment for total energy intake in epidemiologic studies. *The American journal of clinical nutrition 65*(4), 1220S–1228S.

Wood, S. N. (2017). *Generalized additive models: an introduction with R.* 2nd Edition, Chapman and Hall/CRC.

5 Appendix: Marginal effect and elasticity calculus on ILR

We are going to demonstrate how to compute the semi–elasticities of the dependent variable Y relative to an explanatory variable X_j, using compositional models. The demonstration is made for a CODA model with a compositional explanatory variable and a real valued dependent variable. These semi-elasticities calculations are valid for both linear regression and quantile regression.

Consider for $D = 3$, the ILR transformation defined by the transformation matrix:

$$\mathbf{W} = \begin{bmatrix} \sqrt{\frac{2}{3}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \tag{1}$$

Let us remind that $X^* = ilr(X) = V' \ln(X)$, i.e

$$X_1^* = \sqrt{\frac{2}{3}} \ln X_1 - \frac{1}{\sqrt{6}} \ln X_2 - \frac{1}{\sqrt{6}} \ln X_3$$

$$X_2^* = \frac{1}{\sqrt{2}} \ln X_2 - \frac{1}{\sqrt{2}} \ln X_3$$

We define the total as

$$T = exp(\frac{1}{3}\big[\ln(V_1) + \ln(V_2) + \ln(V_3)\big]) \tag{2}$$

i.e

$$\ln T = \frac{1}{3}\big[\ln(V_1) + \ln(V_2) + \ln(V_3)\big]$$

where V_1, V_2, V_3 are the three volumes of macronutrients. Then,

$$\frac{\partial T}{\partial V_1} = \frac{\partial T}{\partial \ln T}.\frac{\partial \ln T}{\partial V_1} = \frac{T}{3V_1}, \qquad \frac{\partial T}{\partial V_2} = \frac{T}{3V_2}, \qquad \frac{\partial T}{\partial V_3} = \frac{T}{3V_3}$$

We define the following transformations

$$F_T : (V_1, V_2, V_3)' \to (ilr_1, ilr_2, T)'$$

$$M : (ilr_1, ilr_2, T)' \to Y = M(ilr_1, ilr_2, T) = \alpha + \beta ilr_1 + \gamma ilr_2 + \delta T,$$

whether $M = \mathbb{E}(ilr_1, ilr_2, T)$ is a mean level or $M = \mathbb{Q}_\tau(ilr_1, ilr_2, T)$, τ is a quantile level. We are going to use the following property of Jacobian matrices: $J = J_M J_{F_T}$.

$$J_{F_T} = \begin{bmatrix} \sqrt{\frac{2}{3}}\frac{1}{V_1} & -\frac{1}{\sqrt{6}}\frac{1}{V_2} & -\frac{1}{\sqrt{6}}\frac{1}{V_3} \\ 0 & \frac{1}{\sqrt{2}}\frac{1}{V_2} & -\frac{1}{\sqrt{2}}\frac{1}{V_3} \\ \frac{T}{3V_1} & \frac{T}{3V_2} & \frac{T}{3V_3} \end{bmatrix} \tag{3}$$

and

$$J_M = \begin{bmatrix} \frac{\partial Y}{\partial V_1^*} & \frac{\partial Y}{\partial V_2^*} & \frac{\partial Y}{T} \end{bmatrix} = \begin{bmatrix} \beta & \gamma & \delta \end{bmatrix}. \tag{4}$$

Then

$$J = J_M J_{F_T} = \begin{bmatrix} \frac{\partial Y}{\partial V_1} \\ \frac{\partial Y}{\partial V_2} \\ \frac{\partial Y}{\partial V_3} \end{bmatrix} = \begin{bmatrix} \beta & \gamma & \delta \end{bmatrix} \begin{bmatrix} \sqrt{\frac{2}{3}}\frac{1}{V_1} & -\frac{1}{\sqrt{6}}\frac{1}{V_2} & -\frac{1}{\sqrt{6}}\frac{1}{V_3} \\ 0 & \frac{1}{\sqrt{2}}\frac{1}{V_2} & -\frac{1}{\sqrt{2}}\frac{1}{V_3} \\ \frac{T}{3V_1} & \frac{T}{3V_2} & \frac{T}{3V_3} \end{bmatrix}$$

$$= \begin{bmatrix} \beta\sqrt{\frac{2}{3}}\frac{1}{V_1} + \frac{\delta T}{3V_1} \\ \frac{-\beta}{\sqrt{6}}\frac{1}{V_2} + \frac{\gamma}{\sqrt{2}}\frac{1}{V_2} + \frac{\delta T}{3V_2} \\ \frac{-\beta}{\sqrt{6}}\frac{1}{V_3} - \frac{\gamma}{\sqrt{2}}\frac{1}{V_3} + \frac{\delta T}{3V_3} \end{bmatrix}$$

Then,

$$\frac{\partial Y}{\partial lnV_1} = \beta\sqrt{\frac{2}{3}} + \frac{\delta T}{3}, \quad \frac{\partial Y}{\partial lnV_2} = \frac{-\beta}{\sqrt{6}} + \frac{\gamma}{\sqrt{2}} + \frac{\delta T}{3}, \quad \frac{\partial Y}{\partial lnV_3} = \frac{-\beta}{\sqrt{6}} - \frac{\gamma}{\sqrt{2}} + \frac{\delta T}{3}$$

We are now going to demonstrate that the semi–elasticities are invariant to the choices of the transformation matrix. In addition, the demonstration is made in a general case, i.e with D components. Assume V has D components, i.e $V = (V_1, V_2, ..., V_D)$. Assume there are two tranformation matrices $\mathbb{V}^{\mathbb{A}}$ and $\mathbb{V}^{\mathbb{B}}$, the corresponding Ilr coordinates are

$$Ilr^A = (Ilr_1^A, ..., Ilr_{D-1}^A) = \left[\mathbb{V}^{\mathbb{A}}\right]_{(D-1)\times D}[\ln V]_{D\times 1} = \left[\mathbb{V}^{\mathbb{A}}\right]_{(D-1)\times D} \begin{bmatrix} \ln V_1 \\ \ln V_2 \\ ... \\ \ln V_D \end{bmatrix}$$

and

$$Ilr^B = (Ilr_1^B, ..., Ilr_{D-1}^B) = \left[\mathbb{V}^{\mathbb{B}}\right]_{(D-1)\times D}[\ln V]_{D\times 1} = \left[\mathbb{V}^{\mathbb{B}}\right]_{(D-1)\times D} \begin{bmatrix} \ln V_1 \\ \ln V_2 \\ ... \\ \ln V_D \end{bmatrix}$$

A total as geometric mean of V is

$$\ln T = \frac{1}{D}(\ln V_1 + \ln V_2 + ... + \ln V_D).$$

Then, given that there are two ways to construct Ilr coordinates, there are two regression models:

$$Y = \alpha^A + \sum_{j=1}^{D-1} \beta_j^A Ilr^A + \delta^A T + \epsilon^A = \alpha^A + [\beta_1^A ... \beta_{D-1}^A]\left[\mathbb{V}^{\mathbb{A}}\right]_{(D-1)\times D}[\ln V]_{D\times 1} + \delta^A T + \epsilon^A \tag{5}$$

$$Y = \alpha^B + \sum_{j=1}^{D-1} \beta_j^B Ilr^B + \delta^B T + \epsilon^B = \alpha^B + [\beta_1^B ... \beta_{D-1}^B]\left[\mathbb{V}^{\mathbb{B}}\right]_{(D-1)\times D}[\ln V]_{D\times 1} + \delta^B T + \epsilon^B \tag{6}$$

Models (5) and (6) are estimated by the ordinary least squares method. They have the same dependent variable, i.e Y and the same explanatory variables, i.e $\ln V_1, \ln V_2, ..., \ln V_D$ and T. Then, the corresponding coefficients estimated from the two models must be equal. Thus we have

$$\delta^A = \delta^B = \delta \quad \text{and} \quad [\beta_1^A \quad ... \quad \beta_{D-1}^A]\left[\mathbb{V}^{\mathbb{A}}\right]_{(D-1)\times D} = [\beta_1^B \quad ... \quad \beta_{D-1}^B]\left[\mathbb{V}^{\mathbb{B}}\right]_{(D-1)\times D} \tag{7}$$

In addition, we define the following transformations

$$F_T^A : (V_1, ..., V_D)' \to (ilr_1^A, ..., ilr_{D-1}^A, T)'$$

$$M^A : (ilr_1^A, ..., ilr_{D-1}^A, T)' \to Y = M^A(ilr_1^A, ..., ilr_{D-1}^A, T)' = \alpha^A + \sum_{j=1}^{D-1} \beta_j^A Ilr^A + \delta^A T,$$

and

$$F_T^B : (V_1, ..., V_D)' \to (ilr_1^B, ..., ilr_{D-1}^B, T)'$$

$$M^B : (ilr_1^B, ..., ilr_{D-1}^B, T)' \to Y = M^B(ilr_1^B, ..., ilr_{D-1}^B, T)' = \alpha^B + \sum_{j=1}^{D-1} \beta_j^B Ilr^B + \delta^B T,$$

then, we have

$$J^A = J_{M^A} J_{F_T^A} \qquad J^B = J_{M^B} J_{F_T^B}$$

In detail

$$J_{M^A} = \begin{bmatrix} \beta_1^A & ... & \beta_{D-1}^A & \delta \end{bmatrix} \qquad J_{M^B} = \begin{bmatrix} \beta_1^B & ... & \beta_{D-1}^B & \delta \end{bmatrix}$$

$$J_{F_T^A} = \begin{bmatrix} \mathbb{V^A}_{(D-1)\times D} \\ \left[\frac{T}{D}\right]_{1\times D} \end{bmatrix} \left[\frac{1}{V}\right]_{D\times 1} \qquad J_{F_T^B} = \begin{bmatrix} \mathbb{V^B}_{(D-1)\times D} \\ \left[\frac{T}{D}\right]_{1\times D} \end{bmatrix} \left[\frac{1}{V}\right]_{D\times 1}$$

Then,

$$J^A = \begin{bmatrix} \beta_1^A & ... & \beta_{D-1}^A & \delta \end{bmatrix} \begin{bmatrix} \mathbb{V^A}_{(D-1)\times D} \\ \left[\frac{T}{D}\right]_{1\times D} \end{bmatrix} \left[\frac{1}{V}\right]_{D\times 1}$$

and

$$J^B = \begin{bmatrix} \beta_1^B & ... & \beta_{D-1}^B & \delta \end{bmatrix} \begin{bmatrix} \mathbb{V^B}_{(D-1)\times D} \\ \left[\frac{T}{D}\right]_{1\times D} \end{bmatrix} \left[\frac{1}{V}\right]_{D\times 1}$$

The semi-elasticity computed from the two different sets of Ilr coordinates are

$$\left[\frac{\partial Y}{\partial lnV}\right]_{D\times 1} = \begin{bmatrix} \beta_1^A & ... & \beta_{D-1}^A & \delta \end{bmatrix} \begin{bmatrix} \mathbb{V^A}_{(D-1)\times D} \\ \left[\frac{T}{D}\right]_{1\times D} \end{bmatrix}$$

$$\left[\frac{\partial Y}{\partial lnV}\right]_{D\times 1} = \begin{bmatrix} \beta_1^B & ... & \beta_{D-1}^B & \delta \end{bmatrix} \begin{bmatrix} \mathbb{V^B}_{(D-1)\times D} \\ \left[\frac{T}{D}\right]_{1\times D} \end{bmatrix}$$

Applying the results of equation (7), we infer that the calculation of the semi-elasticity is invariant to the choices of transformation matrix $\mathbb{V^A}_{(D-1)\times D}$ and $\mathbb{V^B}_{(D-1)\times D}$.