WORKING PAPERS

N° 17-825

June 2017

"Impact of socioeconomic factors on nutritional diet in Vietnam from 2004 to 2014: new insights using compositional data analysis"

Huong Thi Trinh, Joanna Morais

Toulouse School of Economics

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

Huong Thi Trinh*, Joanna MORAIS†

June 28, 2017

Abstract

This paper contributes to the analysis of 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) 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. The compositional model highlights the important role of food expenditure, size of the household and dwelling region in the determination of diet choices. 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: Macronutrient shares, diet pattern, compositional regression models, expenditure elasticity, Vietnam.

Jel codes: C02, C21, C51, P46

1 Introduction

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.

*Toulouse School of Economics, INRA, University of Toulouse Capitole, Toulouse, France
†BVA, 52 rue Marcel Dassault, Boulogne-Billancourt, France

Many empirical researches focus on the relationship between socioeconomic characteristics of households and their food consumption behavior. Food consumption is measured initially by calorie, i.e food categories in quantity are converted in calorie intake. The relationship between calorie intake and income (or expenditure) have been well studied for many countries in order to implement policies which reduce starvation and redress nutritional deficiencies (see a recent meta–analysis by Ogundari & Abdulai (2013)). Then, economic development and urbanization in developing countries have affected global diet, leading to many empirical researches focusing on food sources, 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 results of rising incomes and urbanization (Global Food Policy Report, IFPRI (2017)). Another concern about food consumption is its composition in terms of macro and micronutrient (such as protein, fat, carbohydrate, vitamin A, zinc). Recently, Santeramo & Shabnam (2015) do a review of estimated income elasticities of calories, macronutrients and micronutrients with a total of 26 studies 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 addition, there are only 5 over 26 empirical researches which focus on all macronutrients, i.e protein, fat and carbohydrate.

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. For example, Liaskos et al. (2003) perform 13 multiple linear regressions, one for each of the 13 nutrients (dependent variable), using household characteristics as explanatory variables (the same variables in all models), in Greece. Similarly, You et al. (2016) fit three specifications of health production functions with the same explanatory variables in China, the outcome of the models being the macronutrients consumptions in protein, fat and carbohydrate. 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. The percentage of losses and wastes varies from 5% to 12% across countries (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 the quantities 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.

Vietnam is a good example of middle-income country that has recorded impressive achievements in economy and population welfare after the launch of economic reforms in 1986. However, this country has also experienced a nutrition transition like many other middle-income countries. Nutrition transition has motivated many empirical works in Vietnam, such as Mishra & Ray (2009), and Thang & Popkin (2004). Particularly, Thang & Popkin (2004)) show a significant evolution of the structure of the diet during the 1990s, which contains less and less starchy staples and more and more proteins and lipids coming from meat, fish, and other protein-rich and higher fat food items. In the 1992–1993 period, the main consumed food items by the Vietnamese people were cereals, potatoes, rice, and other starches, contributing up to 85.9% of total energy intake, while calories coming from other food items were low: only 6.8% of total calories were obtained from meat, fish, tofu, and other protein-rich food items, and 2.4% from fats and oils. In the 1997–1998 period, even though the total amount of calories consumed per capita remained at about the same level as 5 years earlier, there was a remarkable increase in daily proteins and lipids consumption (4.7 points) while the consumption of rice and other starches reduced significantly (5.6 points). Recently, National Institute of Nutrition in Vietnam has defined an indicator of what should be a balanced diet for Vietnamese people, with a vision toward 2030: "the proportion of households with a balanced diet (protein:fat:carbohydrate ratios (P:F:C)– 14:18:68) will reach 50% by 2015 and 75% by 2020." (see National Institute of Nutrition (2012)).

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 approach allows us to take into account the dependence among macronutrients and to avoid the problem of overestimation of total calorie intake when using household survey data. We use compositional data analysis (CODA) in order to model shares. CODA is a well established field of statistics with diverse fields of application (see Pawlowsky-Glahn & Buccianti (2011)). A composition is a vector of shares, where shares are called the components. The advantage of CODA is to deal with compositional vectors since composition provides relative information as opposed to absolute value. In our study, diet components are the proportions of protein, fat and carbohydrate in the average per capita calorie intake. CODA allows to analyze the shift in protein, fat, and carbohydrate shares in diets. 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, such that shares ratios and ternary diagrams, to show the evolution of the three components over the years. Then, we model macronutrients composition as a function of household characteristics, using compositional regression models. We first check the quality of our estimates using various model diagnostics, and then we focus on the impact of food expenditure on the share of each macronutrient in the consumption, measuring elasticities of macronutrient shares relative to food expenditure. We also compare these shares elasticities to elasticities of the volumes of macronutrient, and to the elasticity of the total calorie intake. This study uses six waves of the Vietnam Household Living Standard Survey (VHLSS), from 2004 to 2014.

The rest of this article is structured as follows: section 2 presents the data and provides an analysis of the diet patterns of Vietnam households during a ten–year period. Section 3 discusses the use of CODA when analyzing the impact of food expenditure, urbanization and household characteristics on macronutrient consumption. Food expenditure elasticities of macronutrient shares are presented in a fourth section, and are compared to classical food expenditure elasticities of macronutrient volumes. The last section concludes.

2 The diet pattern of Vietnamese households during a ten-year period

2.1 Data

This study uses data from the Vietnam Household Living Standard Survey, carried out in 2004, 2006, 2008, 2010, 2012 and 2014 by the General Statistics Office of Vietnam in collaboration with World Bank. Each wave sample comprises nearly 9000 households and is nationwide representative for all the 63 Vietnamese provinces. Our analysis makes use of expenditures on food and drink items provided from VHLSS questionnaires[1]. Quantities for 56 food items, including purchased foods and self–subsidies, as well as expenditures for purchased food are recorded.[2]

Conversion factors of grams into calories coming from the food composition table constructed by the Vietnam National Institute of Nutrition in 2007 are used to compute macronutrient consumption amounts (see Table A1 in the appendix). For each household, we compute the total calorie intake (in Kcal), and the protein and fat intakes (in gram) per day. Then, we convert for each household the quantity in gram of protein (resp. fat) into Kcal[3] by multiplying by 4 (resp. 9). Finally, using the recent methodology proposed by Aguiar & Hurst (2013), we calculate a per capita calorie intake (namely $PCCI$), a per capita volume of calories obtained from protein (namely V_P), and a per capita volume of calories obtained from fat (namely V_F), by dividing by an equivalence scale computed for each household (these scales are household specific, for more detail see Trinh et al. (2017)). As the total per capita calorie intake $PCCI$ comes from three types of macronutrients (protein, fat and carbohydrate), the per capita calorie intake obtained from carbohydrate (namely V_C) is calculated as:

$$V_C = PCCI - V_P - V_F$$

The macronutrient shares S_P, S_F and S_C are defined as the percentage of calories coming from protein, fat and carbohydrate:

$$S_P = \frac{V_P}{PCCI}, \quad S_F = \frac{V_F}{PCCI}, \quad S_C = 1 - S_P - S_F$$

We also concentrate on many household socioeconomic characteristics such as food expenditure (Exp)[4], household location ($Urban$, $Area$), household size ($HSize$), the characteristics of the head of the household, including education ($Educ$), gender ($Gender$) and ethnicity ($Ethnic$). These explanatory variables can have a potential impact on macronutrient consumption as shown in Thang & Popkin (2004), and Mishra & Ray (2009). Table 1 provides a description of our data. The food expenditure has changed dramatically from 2004 to 2014. The average food expenditure in 2014 is twice its value

[1] In 2004, 2006, 2008, household food consumption was surveyed using 12–month recall. In 2010, 2012, 2014, household food consumption was surveyed using 30–day recall.

[2] Self–subsidy, gift, donation, and present foods are estimated values.

[3] Protein contains 4 calories per gram and fat contains 9 calories per gram

[4] Expenditures are expressed in 2006 dollars, with 1 dollar being equal to 15,994.25 VNdong in 2006.

Table 1: VHLSS description variables

Variable		Description	2004	2006	2008	2010	2012	2014
N		Nb of observations	8244	8290	8333	8548	8670	8712
V_P		Nb of calories from protein	453.5	461.2	390.1	543.5	537.9	544.3
			(150.0)	(159.5)	(116.5)	(194.4)	(216.7)	(218.6)
V_F		Nb of calories from fat	476.4	510.5	443.8	658.5	664.1	709.1
			(227.5)	(238.6)	(198.7)	(313.5)	(332.8)	(340.8)
V_C		Nb of calories from carbohydrate	2416.5	2383.4	2047.3	2554.1	2516.7	2511.0
			(744.7)	(757.1)	(578.7)	(893.7)	(1005.3)	(1031.2)
S_P		Share of calories from protein	13.6%	13.7%	13.6%	14.5%	14.5%	14.5%
			(1.9%)	(1.9%)	(2.0%)	(2.0%)	(2.0%)	(1.9%)
S_F		Share of calories from fat	14.3%	15.2%	15.5%	17.6%	18.0%	19.1%
			(5.2%)	(4.7%)	(5.5%)	(5.8%)	(6.0%)	(6.5%)
S_C		Share of calories from carbohydrate	72.1%	70.9%	67.9%	67.5%	67.5%	66.4%
			(6.2%)	(5.8%)	(6.6%)	(7.0%)	(6.9%)	(7.4%)
Exp		Food expenditure per year (US$)	598.5	622.8	706.4	966.4	1032.4	1010.2
			(330.8)	(348.1)	(383.8)	(554.1)	(612.4)	(597.9)
$ExpTot$		Total Expenditure per year (US$)	1426.5	1541.2	1763.3	2173.1	2262.4	2303.4
			(947.0)	(1008.5)	(1141.8)	(1398.7)	(1435.5)	(1424.3)
$Engel$		Engel coefficient	46.0%	44.2%	44.0%	49.8%	48.1%	46.0%
			(12.5%)	(12.2%)	(12.4%)	(11.3%)	(11.3%)	(10.9%)
$Urban$	1	Urban	23.34 %	25.28 %	25.86 %	27.56 %	28.54 %	29.61 %
	0	Rural	76.66 %	74.72 %	74.14 %	72.44 %	71.46 %	70.39 %
$HSize$	2	$\leq$ 2 people	11.07 %	12.98 %	14.32 %	16.34 %	18.06 %	19.72 %
	3	3 people	15.74 %	17.13 %	17.58 %	20.12 %	18.92 %	20.02 %
	4	4 people	30.65 %	31.54 %	32.03 %	33.29 %	32.2 %	30.84 %
	5	5 people	21.51 %	20.21 %	19.36 %	16.66 %	17.53 %	16.41 %
	6	$\geq$ 6 people	21.02 %	18.14 %	16.72 %	13.58 %	13.29 %	13.01 %
$Ethnic$	1	Kinh	86.31 %	86.14 %	86.39 %	83.26 %	83.13 %	83.67 %
	0	Minorities	13.69 %	13.86 %	13.61 %	16.74 %	16.87 %	16.33 %
$Gender$	1	Male	76.63 %	75.78 %	75.83 %	75.98 %	75.97 %	75.2 %
	0	Female	23.37 %	24.22 %	24.17 %	24.02 %	24.03 %	24.8 %
$Educ$	1	Below primary	54.25 %	52.06 %	50.76 %	51.1 %	50.68 %	49.15 %
	2	Secondary, High school	41.47 %	43.53 %	44.77 %	42.96 %	43.62 %	44.42 %
	3	University	4.28 %	4.4 %	4.46 %	5.94 %	5.7 %	6.43 %
$Area$	1	Red River Delta	21.57 %	21.79 %	22.13 %	17.57 %	17.26 %	21.54 %
	2	Midlands Northern Mountains	18.63 %	18.23 %	18.13 %	13.35 %	13.01 %	17.3 %
	3	Northern Central Coast	20.44 %	20.53 %	20.05 %	22.18 %	22.16 %	22.08 %
	4	Central Highlands	6.22 %	6.15 %	6.22 %	7.07 %	6.85 %	6.65 %
	5	South East	12.34 %	12.75 %	12.76 %	11.39 %	11.44 %	11.96 %
	6	Mekong River Delta	20.89 %	20.49 %	20.9 %	28.35 %	29.23 %	20.51 %

Average corresponds to arithmetic (resp. closed geometric) mean for volume (resp. share) variables.

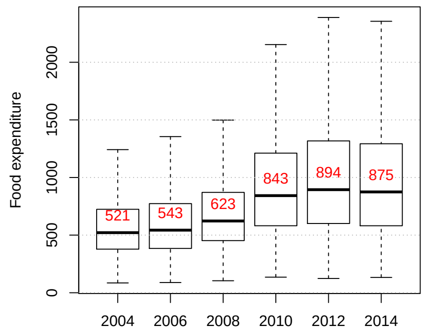
in 2004 (see Table 1 and boxplots in Figure 1 where figures in red are the medians). We also calculate the average Engel coefficient for each year which is the ratio of food expenditure over total expenditure[5]. The average Engel coefficients are quite stable from 2004 to 2014 (around 46%). However, there is an increase of 5.8 points of mean Engel coefficients from 2008 to 2010. The difference is first caused by the 2009 year in the wake of the world crisis (see Cling et al. (2010)). In addition, it may come from the fact that the survey is redesigned between 2008 and 2010 using different population and household census (see Benjamin et al. (2016)).

2.2 The diet pattern of Vietnamese households during 2004-2014

The diet pattern of Vietnamese households has changed dramatically from 2004 to 2014. The volume and the share of macronutrient consumption along time are presented in Figure 2. The median volume of per capita calorie intake (in red color) has increased

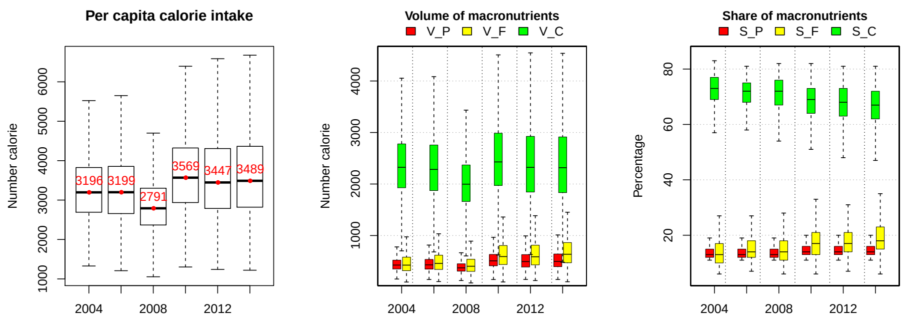
[5]Expenditure are regular consumptions which include education expenditures, healthcare expenditures, food and drink consumption on festive occasions, regular food and drink consumption, daily consumption of non-food items, annual consumption of non-food items, expenditures on durables over the past 12 months, recurrent expenditures on housing, electricity, water, and daily-life waste. We do not add the costs of production and business.

Figure 1: Food expenditure in US$



from 2004 to 2014, except there is a strong fall of PCCI in 2008 due to a difficult climatic year and a very significant increase in food prices (double-digit inflation). With respect to the volume of macronutrient consumption, calories obtained from carbohydrate are quite stable across the six years (except a decrease in 2008) while calories obtained from protein and fat have increased gradually. Shares of macronutrient consumption show an interesting trend during the years.

Figure 2: Volume and share of macronutrient consumption over year



Broadly speaking, during this ten-year period, the average protein share and the average fat share are between 10% and 20%, and the average carbohydrate share is between 60% and 80% (see Table 1). 3 represents the ternary diagrams of the share of macronutrients for the rural and urban sites. The arrows indicate the evolution over the years. Particularly, households in both sites tend to decrease their proportion of carbohydrate and increase their proportion of fat. The evolution of macronutrient consumption in rural and urban sites are going in the same direction. However, the starting points (in 2004) in terms of diet balance are different between rural and urban sites (see Table 2). Moving from (S_P =13.3%, S_F =12.8%, S_C =73.9%) in 2004 to

(14.2%, 17.6%, 68.2%) in 2014, Vietnamese rural households have increased of around 4.8 points the part of calories obtained from fat at the expense of calories obtained from carbohydrate while the calories obtained from protein are quite stable (increase of 0.9 point). In contrast, starting from (14.5%, 16.5%, 69.0%) in 2004 to (15.4%, 20.3%, 64.3%) in 2014, urban households have increased by 3.8 points the part of calories obtained from fat at the expense of calories obtained from carbohydrate, while there is a small change in the proportion of protein (0.9 point).

Regions in Vietnam are different in terms of socio-economic characteristics, and in terms of diet patterns. The map in Figure 4 shows the geometric average of macronutrient shares (S_P, S_F, S_C) and the arithmetic average of food expenditure (Exp), by region ($Area$) in 2014. Red River Delta and South East areas have the highest averages in food expenditure. They also have the largest shares of fat and protein. On the contrary, Midlands Northern Mountains and Mekong River Delta areas have the smallest values for average food expenditure. In the same line, Midlands Northern Mountains has the smallest protein share (13.4%) and Mekong River Delta has the lowest fat share (15.6%). These average macronutrient shares are similar to the results in the General Nutrition Survey 2009-2010 which was conducted by the National Institute of Nutrition (2010). Red River Delta and South East are the two regions who have the highest food consumption of animal-based foods, eggs and milk (in kilograms of food). The General Nutrition Survey also reveals a high proportion of vegetable, such as leafy vegetables and edible flowers and tuberous vegetables for Mekong River Delta and Midlands Northern Mountains. Both our results and the General Nutrition Survey show an average proportion for macronutrient intake and food group consumption for the other regions.

Figure 3: Ternary diagrams of average macronutrient shares in urban and rural sites

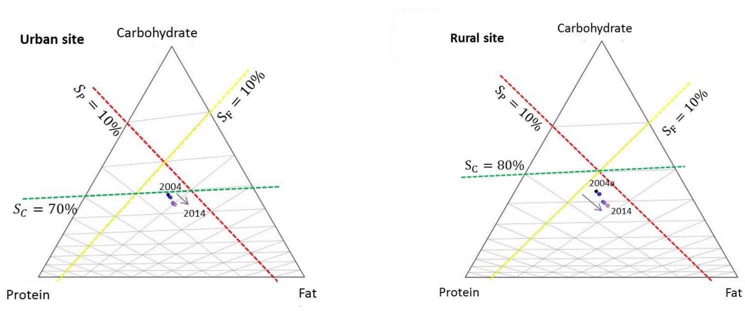
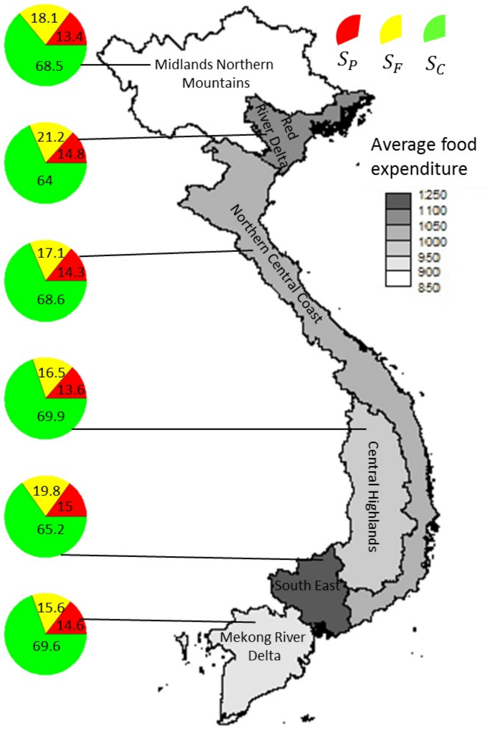


Table 2: Closed geometric mean of macronutrient shares in urban and rural sites

	Urban site			Rural site		
Year	S_P	S_F	S_C	S_P	S_F	S_C
2004	14.5%	16.5%	69.0%	13.3%	12.8%	73.9%
2014	15.4%	20.3%	64.3%	14.2%	17.6%	68.2%

Figure 4: Macronutrient shares and food expenditure averages by area in 2014

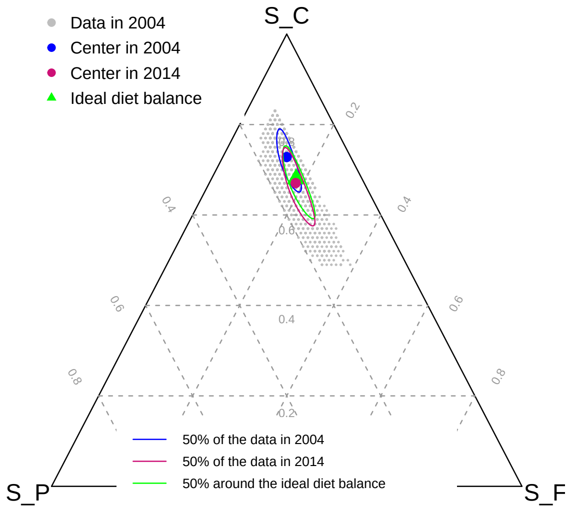


Beyond analyzing the center of the data, it is also interesting to look at the dispersion of data around this center. Figure 5 represents in a ternary diagram the centers of data in 2004 and 2014, along with ellipses delimiting half of the population around these points. The "ideal" balanced diet according to the National Institute of Nutrition in Vietnam (S_P=14%,S_F=18%,S_C=68%) is represented with a triangle. This ternary diagram shows that half of the population in 2014 have a diet balance very close to the ideal one, closer than in 2004. In addition, we can check that the centers of the "very poor" and "very rich"[6] are very far from each other. In 2004, the center of the "very poor" (S_P =13.0%, S_F =12.1%, S_C =74.9%) is far from the ideal diet point while the center of the "very rich" (15.4%, 17.8%, 66.8%) is close to the ideal diet balance. In 2014, the center of the "very poor" and "very rich" are (13%, 16.8%, 69.2%) and (15.9%, 22.1%, 61.9%). Thus, the "very poor" households in 2014 still do not consume enough protein and fat, while the "very rich" households consume too much fat.

Note that the information carried by a vector of three shares can be summarized in two ratios of shares, thanks to the summing up to one constraint. We define the following two shares ratios in our case: $R_{PC} = \frac{S_P}{S_C}$ and $R_{FC} = \frac{S_F}{S_C}$. Figure 6 represents

[6]Households who have food expenditure less than 5% (217.7$) and higher than 95% quantile 1247.1$ in 2004 (resp. 304.8$ and 2165.6$ in 2014)

Figure 5: Ternary diagram of centers in 2004 and 2014 compared to the “ideal” diet balance

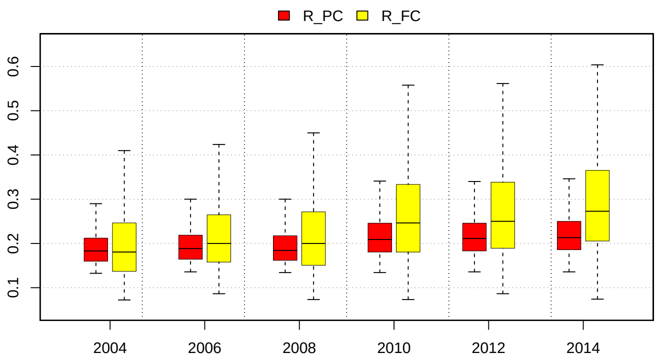


the boxplots of the shares ratios over the years. Except for some outliers from 2010, they are all less than 0.5 (i.e the proportion of carbohydrate is more than twice the proportions of protein and fat). However, these ratios have increased over time. Moreover, in 2004, the median values for both R_{PC} and R_{FC} are quite similar, but in 2014 the median value of R_{FC} is much higher than the value of R_{PC}. The evolution shows an increase of the consumption of fat and protein at the expense of carbohydrate, and this increase is more pronounced for fat than for protein. The evolution of Vietnamese diet pattern is consistent with the global change in diets consisting in an increase in consumption of animal-source foods, fats and oils at the expense of grains and cereals (see IFPRI (2017)).

3 Compositional data analysis approach to describe and explain macronutrient consumption

In the literature, different types of models are available for shares regression (see Morais et al. (2016) for a comparison). In the case where the dependent variable is a vector of shares (e.g. the composition of macronutrients) and explanatory variables are classical variables which depend only on the observations (e.g. household characteristics), a model has been proposed in the so-called CODA (compositional data analysis) literature. This model is very simple to implement and is based on a log-ratio transformation of

Figure 6: Boxplots of ratios of protein and fat shares over carbohydrate share



shares. A composition $\mathbf{S}$ of D shares belongs to the simplex space $\mathcal{S}^D$:

$$\mathcal{S}^D = \left\{ \mathbf{S} = (S_1, S_2, \dots, S_D)' : S_j > 0,\ j = 1, \dots, D;\ \sum_{j=1}^{D} S_j = 1 \right\}$$

Compositions are subject to the following constraints: the components are positive and sum up to 1. Because of these constraints, classical regression models cannot be used directly. Thus, shares are transformed, using an isometric log-ratio (ILR) transformation for example, in $D-1$ coordinates which belong to the classical Euclidean space so that linear regression models can be used separately on the $D-1$ coordinates. The ILR coordinates are defined as:

$$ilr(\mathbf{S}) = \mathbf{W}' \log(\mathbf{S}) = \mathbf{S}^* = (S_1^*, \dots, S_{D-1}^*)'$$

where the $D \times (D-1)$ matrix $\mathbf{W}$ allows the projection of shares on an orthonormal basis of $\mathcal{S}^D$. For example, for $D = 3$, the following $\mathbf{W}$ matrix can be used:

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

Finally, inverse transformation of results allows to go back to the simplex in order to interpret the model on shares. The inverse transformation is given by: $\mathbf{S} = ilr^{-1}(\mathbf{S}^*) = \mathcal{C}(\exp(\mathbf{W}\mathbf{S}^*))'$, where $\mathcal{C}(.)$ is the closure operation allowing to go from a vector of volumes $\mathbf{V}$ to a vector of shares $\mathbf{S}$: $\mathcal{C}(V_1, \dots, V_D)' = \left(\frac{V_1}{\sum_{j=1}^{D} V_j}, \dots, \frac{V_D}{\sum_{j=1}^{D} V_j} \right)' = (S_1, \dots, S_D)'$. Let us introduce the following operators used in the simplex: the operators $\oplus$ and $\odot$

are called perturbation operation and power transformation, and play a similar role as operators $+$ and $\times$. They are defined as follows:

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, \ldots, x_D y_D)' \qquad \text{with } \mathbf{x}, \mathbf{y} \in \mathcal{S}^D$$
$$\lambda \odot \mathbf{x} = \mathcal{C}(x_1^\lambda, \ldots, x_D^\lambda)' \qquad \text{with } \lambda \in \mathbb{R}, \mathbf{x} \in \mathcal{S}^D$$

3.1 Compositional model for macronutrient shares

We are interested in the impact of Vietnamese household characteristics on their macronutrient composition, and the evolution of this impact across time, from 2004 to 2014. An adapted 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 $\mathbf{S} = (S_P, S_F, S_C)'$, and the index i denotes the i^{th} household. $\mathbf{S}, \mathbf{a}, \mathbf{b_k}, \boldsymbol{\epsilon} \in \mathcal{S}^D$ are compositional and X_k are classical explanatory variables (Exp is a positive continuous variable, used in logarithm, and others are categorical variables).

After ILR transformation, model (1) can be written in $D-1=2$ 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 (because $D-1=2$) separately for the 6 years of observation, using OLS and the assumption that $\boldsymbol{\epsilon}^*$ follows a Gaussian distribution, $\boldsymbol{\epsilon}$ follows a Gaussian distribution that is in the simplex.
As explained before, 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). For example, $\widehat{\mathbf{b_1}} = \mathcal{C}(\exp(\mathbf{W}\widehat{\mathbf{b_1^*}}))'$, with $\widehat{\mathbf{b_1^*}} = (\widehat{b}_{1,1}^*, \ldots, \widehat{b}_{D-1,1}^*)'$.

3.2 Diagnostic model-checking

In order to determine if the above presented compositional model is reliable to explain macronutrient shares, we have to check several items.

Significance of explanatory variables According to the analysis of the variance of our compositional models, all household characteristics used in the model are very significant (at 1%), at all observation periods[7].

Quality measure The quality of compositional models can be assessed by a measure adapted to share data, called R^2 based on the total variance, denoted R^2_T (see Van den Boogaart & Tolosana-Delgado (2013)). Table 3 shows that our models explain around 30% of the total variability of the compositional data, but the quality of models tends to decrease over time, meaning that recently other factors than those considered explain the household diet balance.

Table 3: Adjusted R^2_T for macronutrient shares modeling

	2004	2006	2008	2010	2012	2014
R^2_T	0.31	0.33	0.28	0.29	0.23	0.22

Inspection of residuals Figure A1 in the appendix represents boxplots of absolute values of share residuals by component. This figure shows that the fitted error for the share of protein in consumption is very low. Errors happen mainly in the fitting of fat and carbohydrate shares, and these two shares are more and more difficult to estimate across time. Our compositional model is based on the assumption that error terms $\boldsymbol{\epsilon}$ in (1) follow a Gaussian distribution in the simplex, which is equivalent to say that error terms ϵ_j^* in (2) or log ratios of error terms in $\boldsymbol{\epsilon}$ follow a Gaussian distribution. Then, we check the normality of residuals, using QQ-plots (one by log ratio of residuals). They show that the residuals are close to follow a Gaussian distribution although there is a heavy tailed distribution (see Figure A2 in the appendix for year 2010). Moreover, the residuals are symmetric according to the residuals log ratios boxplots (see Figure A3 in the appendix for year 2010). Then we conclude that our compositional model is relevant and reliable to explain the diet balance between calories intakes from protein, fat and carbohydrates.

3.3 Regression results

Table 4 summarizes the coefficients of the compositional model (1) over the years. Highlighted numbers correspond to higher and smaller coefficients. They have to be compared to the "compositional zero": $\mathbf{0} = (1/3, 1/3, 1/3)'$. If the coefficient associated to variable X is larger (resp. smaller) than 0.33 for the macronutrient j, it means that an increase of the X variable results in an increase (resp. decrease) of the share of macronutrient j in the consumption. If the coefficient is equal to 0.33, X has no significant impact on j's share.

In our case, we realize 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. This is consistent with the fact that larger households live in rural

[7]Full results available upon request.

sites[8]. 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, except that households living in South East and Mekong River Delta tend to consume more protein calories. These conclusions are quite stable across time.

Table 4: Coefficients of the compositional regression model in the simplex

		2004			2006			2008			2010			2012			2014		
Regressors		S_P	S_F	S_C	S_P	S_F	S_C	S_P	S_F	S_C	S_P	S_F	S_C	S_P	S_F	S_C	S_P	S_F	S_C
(Intercept)		0.05	0.02	0.93	0.06	0.03	0.92	0.05	0.03	0.91	0.07	0.04	0.90	0.07	0.04	0.89	0.07	0.06	0.87
log(Exp)		0.33	0.41	0.26	0.33	0.39	0.27	0.34	0.38	0.27	0.33	0.39	0.28	0.34	0.38	0.28	0.34	0.37	0.29
Urban	0	0.33	0.32	0.35	0.33	0.32	0.35	0.33	0.31	0.36	0.33	0.31	0.35	0.33	0.32	0.35	0.33	0.33	0.34
Gender	0	0.33	0.34	0.33	0.33	0.34	0.33	0.33	0.34	0.33	0.33	0.34	0.33	0.33	0.34	0.32	0.33	0.34	0.33
Educ	2	0.33	0.34	0.33	0.33	0.34	0.33	0.33	0.34	0.33	0.33	0.35	0.32	0.33	0.35	0.33	0.33	0.34	0.33
	3	0.33	0.36	0.31	0.33	0.35	0.32	0.33	0.35	0.32	0.33	0.35	0.32	0.33	0.36	0.32	0.33	0.34	0.32
Hsize	3	0.33	0.29	0.38	0.33	0.30	0.37	0.33	0.30	0.38	0.33	0.30	0.37	0.33	0.31	0.37	0.32	0.31	0.37
	4	0.33	0.27	0.40	0.33	0.28	0.39	0.32	0.28	0.40	0.32	0.29	0.39	0.32	0.29	0.39	0.32	0.30	0.38
	5	0.32	0.25	0.42	0.32	0.26	0.41	0.32	0.26	0.42	0.32	0.27	0.41	0.32	0.28	0.40	0.32	0.29	0.39
	6	0.32	0.23	0.45	0.32	0.25	0.44	0.32	0.24	0.44	0.32	0.25	0.43	0.32	0.26	0.42	0.31	0.28	0.41
Ethnic	0	0.33	0.32	0.35	0.33	0.32	0.35	0.33	0.32	0.35	0.32	0.33	0.35	0.32	0.33	0.35	0.33	0.32	0.35
Area	2	0.33	0.34	0.33	0.33	0.33	0.34	0.33	0.33	0.34	0.33	0.32	0.34	0.33	0.32	0.34	0.33	0.33	0.35
	3	0.34	0.32	0.34	0.34	0.31	0.35	0.34	0.32	0.34	0.34	0.31	0.35	0.34	0.30	0.36	0.34	0.29	0.37
	4	0.33	0.33	0.34	0.34	0.32	0.35	0.33	0.33	0.34	0.34	0.30	0.36	0.34	0.30	0.36	0.34	0.30	0.37
	5	0.35	0.32	0.33	0.35	0.32	0.34	0.34	0.33	0.33	0.34	0.32	0.34	0.34	0.32	0.34	0.34	0.31	0.35
	6	0.36	0.30	0.34	0.36	0.29	0.34	0.36	0.30	0.35	0.35	0.30	0.35	0.35	0.30	0.35	0.35	0.28	0.37

4 Food expenditure elasticity of macronutrient consumption shares and volumes

4.1 Elasticities computation in compositional models

In order to interpret share models, the elasticity is often an adapted measure because it is a measure of the relative impact of an explanatory variable on a share, after a relative change of this explanatory variable. For example, it will allow us to measure the percentage increase of the share of fat in the consumption of a Vietnamese household, when the food expenditure of the household increases by 1%. This elasticity corresponds actually to the following formula:

$$Elast(S_{j,i}, Exp_i) = \frac{\frac{\partial S_{j,i}}{S_{j,i}}}{\frac{\partial Exp_i}{Exp_i}} = \frac{\partial \log S_{j,i}}{\partial \log Exp_i} \tag{3}$$

[8]It was especially true at the beginning of the period: in 2004, 80% of household made of 5 people and more were living in rural sites, whereas in 2014 it was 73% (77% in average on the period).

Morais et al. (2017) prove that model (1) can be written in a similar fashion as the classical attraction models, used in the marketing literature (see Nakanishi and Cooper):

$$S_{j,i} = \frac{a_j \prod_{k=1}^{K} b_{j,k}^{X_{ki}} \epsilon_{j,i}}{\sum_{m=1}^{D} a_m \prod_{k=1}^{K} b_{m,k}^{X_{ki}} \epsilon_{m,i}} = \frac{\exp(\log a_j + \sum_{k=1}^{K} X_{ki} \log b_{j,k} + \log \epsilon_{j,i})}{\sum_{m=1}^{D} \exp(\log a_m + \sum_{k=1}^{K} X_{ki} \log b_{m,k} + \log \epsilon_{m,i})} \quad (4)$$

From equation (4), we can compute the previous elasticity relative to $\log Exp_i$:

$$Elast(S_{j,i}, Exp_i) = \log b_{j,1} - \sum_{m=1}^{D} S_{m,i} \log b_{m,1} \quad (5)$$

where $b_{j,1}$ are the coefficients associated to $\log(Exp)$ for each macronutrient S_j.

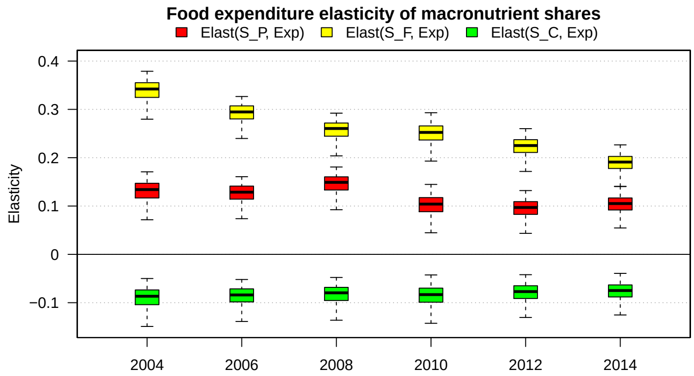
4.2 Elasticity of macronutrient shares

Elasticities of macronutrient shares relative to the household food expenditure are presented in the boxplots in Figure 7, and are summarized in Table 5, for all observation periods. We can see that the fat share is the most elastic macronutrient with respect to the food expenditure: in 2004, the food expenditure was quite low compared to the rest of the periods, and at this time, an increase of 1% of the food expenditure tends to increase the share of fat in the total caloric intake by 0.34%, whereas it tends to increase the share of protein by 0.13% and to decrease the share of carbohydrate by 0.09%. Let us notice that carbohydrate elasticities are negative at all periods, meaning that when households increase their food expenditure, they tend to substitute fat and protein to carbohydrate.
For example, let us consider a household having a diet balance of (14.0%, 20.0%, 66.0%) for protein, fat and carbohydrate, and a food budget of US$1000 in 2014. The corresponding elasticities are (0.1031, 0.1890, -0.0769), thus if the yearly food expenditure of this household increases by US$50 (increase of 5%), its diet balance would change to (14.07%, 20.19%, 65.75%).
Note that the elasticity of the share of fat decreases across time, whereas we know that the food expenditure tends to progress (on average from US$599 in 2004 to US$1010 in 2014). This means that for low food budget households, an increase in food expenditure tends to benefit much more to fat consumption than for high food budget households.

4.3 Elasticity of macronutrient volumes

In order to compare these results with the existing literature, we also perform the usual double-log regression models explaining the consumption volume of each macronutrient and of the total calorie intake ($PCCI$) by the same household characteristics than in model (1) (one model by macronutrient and one for the total, estimated separately by

Figure 7: Boxplot of food expenditure elasticities of macronutrient consumption shares



OLS):

$$\begin{aligned} \log(V_{j,i}) &= \alpha_j + \beta_{1,j}\log(Exp_i) + \sum_{k=2}^{K}\beta_{k,j}X_{ki} + \varepsilon_{j,i} \qquad \text{for } j=1,2,3 \\ \log(PCCI_i) &= \alpha + \beta_1\log(Exp_i) + \sum_{k=2}^{K}\beta_k X_{ki} + \varepsilon_i \end{aligned} \tag{6}$$

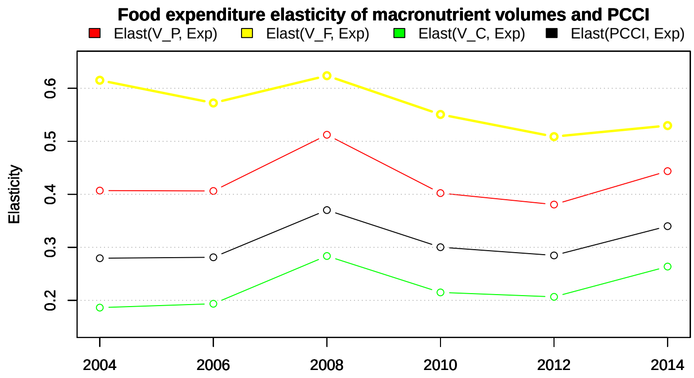
Then, the elasticities of macronutrient volumes relative to food expenditure are equal to:

$$Elast(V_{j,i}, Exp_i) = \frac{\frac{\partial V_{j,i}}{V_{j,i}}}{\frac{\partial Exp_i}{Exp_i}} = \frac{\partial \log V_{j,i}}{\partial \log Exp_i} = \beta_{1,j}$$

and the elasticity of the total calorie intake relative to food expenditure is equal to β_1. Note that for double-log regression models, the elasticity is a constant term which does not depend on the considered household i, whereas the elasticity of the macronutrient share S_j for household i depends on all $S_{m,i}, m = 1, \dots, D$ (on the full composition of macronutrient shares), that is on the diet balance of household i.

In this application, estimated coefficients $\hat{\beta}_{1,j}$ and $\hat{\beta}_1$ are all significantly different from zero at 0.1%, at all periods, meaning that the food budget has a real impact on the consumption of macronutrients and on the total calorie intake. Figure 8 represents the volume elasticities relative to the food expenditure across time. Table 5 compares elasticities obtained from the share model (1) and the volume model (6). All elasticities are positive for macronutrient volumes, meaning that when the food budget of an household increases, all types of caloric intakes increase: the household consumes more of all macronutrients. This is consistent with the fact that $PCCI$ food expenditure elasticities are positive and significant too. However, similarly to the study of macronutrient shares, we conclude that fat is the more elastic and carbohydrate the less elastic to food budget. If the food expenditure of a household increases by 1%, the calories coming from fat tend to increase by 0.62% in 2004 and by 0.53% in 2014.

Figure 8: Food expenditure elasticities of macronutrient volumes



Our results are consistent with those of previous studies, like Liaskos et al. (2003) for example.

Table 5: Food expenditure elasticities of macronutrients shares and volumes

	Protein		Fat		Carbohydrates		PCCI
Year	Share	Volume	Share	Volume	Share	Volume	Volume
2004	0.1296	0.4071	0.3377	0.6152	-0.0911	0.1863	0.2795
2006	0.1261	0.4063	0.2921	0.5723	-0.0866	0.1936	0.2813
2008	0.1450	0.5123	0.2564	0.6237	-0.0836	0.2837	0.3703
2010	0.1011	0.4023	0.2494	0.5507	-0.0862	0.2150	0.3003
2012	0.0946	0.3807	0.2227	0.5088	-0.0795	0.2067	0.2848
2014	0.1031	0.4437	0.1890	0.5296	-0.0769	0.2637	0.3400

* Average in the case of shares

Note that the log of food expenditure is very significant (P-value ¡ 2e-16) for all macronutrients and all periods. The quality measures (R^2) of models relative to the volumes of macronutrient consumption in Table 6 indicate that the volume of carbohydrate is the most complicated to estimate using household characteristics. In contrast, fat and protein consumptions are well determined by household characteristics we are using.

5 Conclusion and discussion

This paper analyzes the evolution of diet patterns in terms of macronutrients (protein, fat and carbohydrate) and the impact of socioeconomic factors on diet balance in Vietnam, using six waves of the VHLSS data, from 2004 to 2014.

Table 6: Adjusted R^2 for macronutrient volume models

	2004	2006	2008	2010	2012	2014
Protein	0.36	0.32	0.52	0.31	0.30	0.39
Fat	0.46	0.41	0.48	0.39	0.38	0.42
Carbohydrate	0.10	0.09	0.20	0.11	0.09	0.14
PCCI	0.19	0.17	0.33	0.19	0.18	0.25

In the existing literature, food consumption is usually analyzed in terms of nutrient volumes, leading to biases due to the overdeclaration of households in survey data, to the failure to account for waste, and to ignoring the dependence between the different macronutrients consumption. In order to avoid these problems, we propose to focus on the diet balance in terms of macronutrient shares in the total consumption. We use the compositional data analysis (CODA) tools and regression models to highlight the nutrition transition and to explain it according to household characteristics.

The compositional analysis reveals that the share of fat, which was almost equal to the share of protein at the beginning of the period (around 14%), increases a lot at the expense of the carbohydrate share. The compositional model highlights the important role of food expenditure, size of the household and dwelling region in the determination of diet choices. For example, the larger the household is, the lower the fat share tends to be. Concerning the role of food expenditure, elasticities of macronutrient shares have been computed and compared to classical elasticities for macronutrient volumes and total calorie intake. Our results are consistent with the existing literature: the fat is the most elastic macronutrient (in a positive way) to food expenditure, but this elasticity tends to slowly decrease over time (from 0.3 to 0.2 in average from 2004 to 2014). The carbohydrate share is negatively elastic to food expenditure (between -0.09 and -0.08). This reflect the substitution effects in a context of nutrition transition. Moreover, the positive elasticities of the three macronutrients volumes capture the positive impact of food expenditure on the total calorie intake of households.

This research contributes to important findings in the literature about the evolution of diets at country level. As nutrition transition is well–known to be correlated with the rise of non-communicable diseases, such as obesity and heart disease (see Bloom et al. (2012)), national policies are needed to encourage Vietnamese people to improve their diet balance in terms of macronutrients. Indeed, policies should be targeted toward different groups. For example, they should tend to encourage “very poor” households to consume a higher share of fat and protein, and “very rich” households to stabilized their fat share in order to limit the risk of obesity. A limitation of our study comes from the fact that our data does not allow to distinguish between different types of fat.

In further research, similar empirical studies about macronutrients shares in the diet can be done for other countries in order to design a whole picture about food consumption composition. Moreover, it could be interesting to focus on the relationship between macronutrients shares and non-communicable diseases at country level.

Acknowledgements

We would like to acknowledge our thesis advisors, Christine THOMAS-AGNAN and Michel SIMIONI, for their expert comments that greatly improved our manuscript, and their kind supports for the authors collaboration. We would like to thank John GALLUP and NGUYEN Ngoc Hieu who re-calculated expenditure data used in this article. We are grateful to Thibault LAURENT for technical assistance in R. We thank the market research agency BVA and the TAASE project of INRA-CIRAD GloFoodS meta-program for their support.

References

References

Aguiar, M. & Hurst, E. (2013), 'Deconstructing life cycle expenditure', *Journal of Political Economy* **121**(3), 437–492.

Benjamin, D., Brandt, L. & McCaig, B. (2016), 'Growth with equity: Income inequality in vietnam, 2002–14', *The Journal of Economic Inequality* **15**, 25–46.

Bloom, D., Cafiero, E., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L., Fathima, S., Feigl, A., Gaziano, T., Hamandi, A., Mowafi, M. et al. (2012), 'The global economic burden of noncommunicable diseases', *Program on the Global Demography of Aging* .

Cling, J.-P., Razafindrakoto, M., Roubaud, F. et al. (2010), 'Assessing the potential impact of the global crisis on the labour market and the informal sector in vietnam', *Journal of Economics & Development* **38**, 16–25.

Deaton, A. (1997), *The analysis of household surveys: a microeconometric approach to development policy*, World Bank Publications.

IFPRI (2017), '2017 global food policy report', *Washington, DC: International Food Policy Research Institute Pub.* .

Liaskos, G., Lazaridis, P. et al. (2003), 'The demand for selected food nutrients in greece: The role of socioeconomic factors', *Agricultural Economics Review* **4**(2).

Mishra, V. & Ray, R. (2009), 'Dietary diversity, food security and undernourishment: the vietnamese evidence', *Asian Economic Journal* **23**(2), 225–247.

Morais, J., Thomas-Agnan, C. & Simioni, M. (2016), 'A tour of regression models for explaining shares', *TSE Working Paper* **16-742**.

Morais, J., Thomas-Agnan, C. & Simioni, M. (2017), 'Interpreting the impact of explanatory variables in compositional models', *In progress* .

National Institute of Nutrition (2010), ‘General nutrition survey 2009 - 2010’, *Medical Publishing House* .

National Institute of Nutrition (2012), ‘National nutrition strategy for 2011 - 2020, with a vision toward 2030’, *Medical Publishing House* .

Ogundari, K. & Abdulai, A. (2013), ‘Examining the heterogeneity in calorie–income elasticities: A meta-analysis’, *Food Policy* **40**, 119–128.

Pawlowsky-Glahn, V. & 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.

Santeramo, F. G. & Shabnam, N. (2015), ‘The income-elasticity of calories, macro- and micro-nutrients: What is the literature telling us?’, *Food Research International* **76**, 932–937.

Thang, N. M. & Popkin, B. (2004), ‘Patterns of food consumption in vietnam: effects on socioeconomic groups during an era of economic growth’, *European journal of clinical nutrition* **58**(1), 145–153.

Trinh, T. H., Thomas-Agnan, C. & Simioni, M. (2017), ‘A new perspective on the relationship between calorie intake and income in china and vietnam using semiparametric modeling’, *In progress* .

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

You, J., Imai, K. S. & Gaiha, R. (2016), ‘Declining nutrient intake in a growing china: Does household heterogeneity matter?’, *World Development* **77**, 171–191.

Appendix

Table A1: Conversion table Calories for Vietnam

Food	Energy Kcal	protein gr	fat gr	Food	Calorie Kcal	protein gr	fat gr
Plain rice	344.5	8.5	1.55	Sticky rice	347	8.3	1.6
Maize	354	8.3	4	Cassava	146	0.8	0.2
Potato of various kinds	106	1.4	0.15	Wheat grains, bread, wheat powder	313.7	10.2	1.1
Floor noodle, instant rice noodle, porridge	349	11	0.9	Fresh rice noodle, dried rice noodle	143	3.2	0.2
Vermicelli	110	1.7	0	Pork	26016.5	21.5	
Beef	142.5	20.3	7.15	Buffalo meat	122	22.8	3.3
Chicken meat	199	20.3	13.1	Duck and other poultry meat	275	18.5	22.4
Other types of meat	-	-	-	Processed meat	-	-	-
Fresh shrimp, fish	83	17.75	1.2	Dried and processed shrimps, fish	361	49.16	14.6
Other aquatic products and seafoods	-	-	-	Eggs of chicken, ducks, Muscovy ducks, geese	103.74	8.34	7.74
Tofu	95	10.9	5.4	Peanuts, sesame	570.5	23.8	45.5
Beans of various kinds	73	5	0	Fresh peas of various kinds	596	0.4	
Morning glory vegetables	25	3	0	Kohlrabi	36	2.8	0
Cabbage	29	1.8	0.1	Tomato	20	0.6	0.2
Other vegetables	-	-	-	Orange	37	0.9	0
Banana	81.5	1.2	0.2	Mango	69	0.6	0.3
Other fruits	-	-	-	Fish sauce	60	12.55	0
Salt	0	0	0	MSG	0	0	0
Glutamate	0	0	0	Sugars, molasses	390	0.55	0
Confectionery	412.2	8.9	10.7	Condensed milk, milk powder	395.7	23.4	11.9
Ice cream, yoghurt	-	-	-	Fresh milk	61	3.9	4.4
Alcohol of various kinds	47	4	0	Beer of various kinds	11	0.5	0
Bottled, canned, boxed beverages	47	0.5	0	Instant coffee	353	12	0.5
Coffee powder	0	0	0	Instant tea powder	0	0	0
Other dried tea	0	0	0	Cigarettes, waterpipe tobacco	0	0	0
Betel leaves, areca nuts, lime, betel pieces	0	0	0	Outdoor meals and drinks	-	-	-
Other foods and drinks	-	-	-	Lard, cooking oil	863.5	0	99.8

Amount per 100gr food ; protein contains 4 calories per gram and fat contains 9 calories per gram

Figure A1: Boxplots of absolute values of residuals by component and year

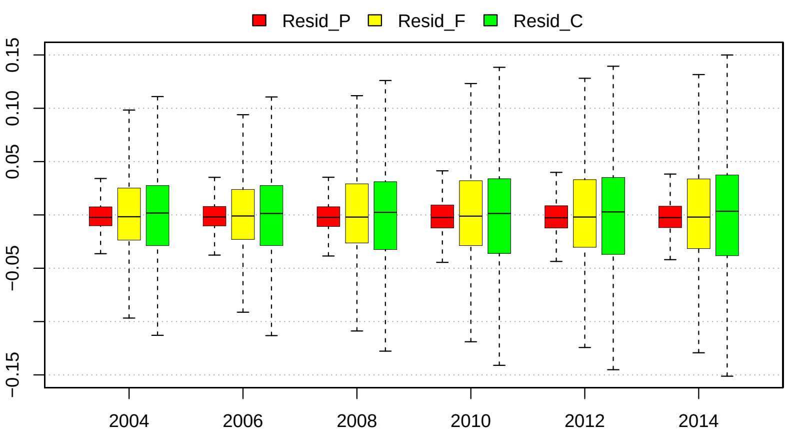


Figure A2: QQ-plot of residuals in 2010

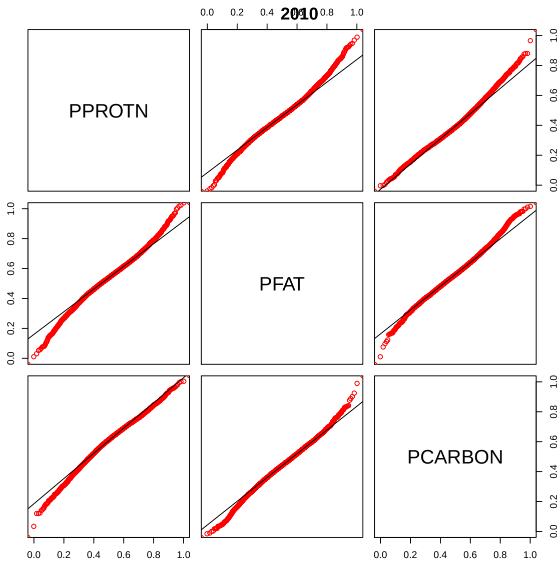


Figure A3: Log ratios of residuals in 2010