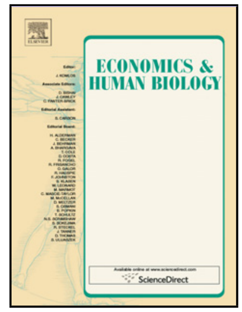
Decomposition of changes in the consumption of macronutrients in Vietnam between 2004 and 2014

Huong Thi Trinh, Michel Simioni, Christine Thomas-Agnan

Accepted Manuscript



Title: Decomposition of Changes in the Consumption of Macronutrients in Vietnam Between 2004 and 2014

Author: Huong Trinh Thi Michel Simioni Christine Thomas-Agnan

PII: S1570-677X(18)30099-6
DOI: https://doi.org/doi:10.1016/j.ehb.2018.09.002
Reference: EHB 742

To appear in: *Economics and Human Biology*

Received date: 3-4-2018
Revised date: 13-8-2018
Accepted date: 16-9-2018

Please cite this article as: Huong Trinh Thi, Michel Simioni, Christine ThomasndashAgnan, Decomposition of Changes in the Consumption of Macronutrients in Vietnam Between 2004 and 2014, *<![CDATA[Economics and Human Biology]]>* (2018), https://doi.org/10.1016/j.ehb.2018.09.002

Decomposition of Changes in the Consumption of Macronutrients in Vietnam Between 2004 and 2014

Huong Trinh Thi[a,b,*], Michel Simioni[c,d], Christine Thomas–Agnan[e]

[a] *Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam*
[b] *International Center for Tropical Agriculture (CIAT), Hanoi, Vietnam*
[c] *MOISA, INRA, University of Montpellier, Montpellier, France*
[d] *IREEDS-VCREME, Hanoi, Vietnam*
[e] *Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France*

Abstract

Vietnam is undergoing a nutritional transition like many middle-income countries. This transition is characterized by an increase in per capita total calorie intake resulting from an increase in the consumption of fat and protein while the carbohydrate consumption decreases. This paper proposes to highlight the sociodemographic drivers of this transition over the period 2004-2014, using Vietnam Household Living Standard Survey data. We implement a method of decomposition of between-year differences in economic outcomes recently proposed in the literature. This method decomposes the between-year change in various indicators related to the outcome distribution (mean, median, quantiles...) into the effect due to between-year change in the conditional distribution of the outcome given sociodemographic characteristics, or "structure effect", and the effect due to the differences in sociodemographic characteristics across years, or "composition effect". In turn, this last effect is decomposed into direct contributions of each sociodemographic characteristics and effects of their interactions. The composition effect, always positive, generally outweighs the structure effect when considering the between-year changes in distributions of per capita calorie intake or calorie intake coming from protein or fat. The effects of changes in the composition of the Vietnamese population thus overcome the effects of changes in preferences of the same population. This finding is reversed in the case of carbohydrates. Food expenditure and household size appear to be the main contributors to the composition effect. The positive effects of these two variables explain well most of the between-year shifts observed in the calorie intake distributions. Urbanization and level of education contribute negatively to the composition effect, with the noticeable exception of fat where the effect of urbanization is positive. But these two variables effects are negligible compared to those of food expenditure and household size.

Keywords: Nutrition transition, macronutrient consumption, decomposition methods in economics, copula, Vietnam.

*Corresponding author
Email address: trinhthihuong@tmu.edu.vn (Huong Trinh Thi)

1. Introduction

Since the launch of economic reforms in 1986, Vietnam has recorded impressive achievements in growth performance and has gone from one of the world's poorest nations to a lower middle-income country. Vietnamese GDP has grown from 383.381 in 1986 to 1834.652 US 2010 dollars in 2017. Poverty[1] was reduced from 35.5% of the population in 1998 to only 2.6% in 2014. Life expectancy has increased from 69.30 years in 1986 up to 76.05 in 2016. Meanwhile, Vietnam has also experienced a nutrition transition like many other middle-income countries in South East Asia. Recently, IFPRI (2017) shows that dietary diversity in South-East Asia and China has considerably increased from 2005 to 2015. The share of cereal demand (in terms of quantity) has decreased by 12% while the share of meat and fish demand and those of dairy and eggs have increased by 8% and 30% respectively, the share of fruits and vegetables staying steady.

As Popkin (1993, 1994) pointed out first, improvements in per-capita incomes are generally accompanied by a major shift in dietary patterns. Total energy intake increases and diet moves from a relatively monotonous and starchy diet with low fat and high fiber intakes, to a more varied diet with more fruits and vegetables. Furthermore, intakes in meat, fat and sugar increase while high fiber intakes decline. Popkin (1993, 1994) coined the expression "nutrition transition" to describe these changes. He also emphasized that two other processes of change occur simultaneously with the nutrition transition (Popkin, 2002). The first process, or epidemiologic transition, is characterized by a transition from a situation where infectious diseases related to chronic malnutrition, periodic starvation, and poor environmental sanitation prevail, to a situation now characterized by the prevalence of chronic and degenerative diseases related to urban-industrial lifestyles. The second process, or demographic transition, involves a major shift in age-specific mortality patterns and a consequent increase in life expectancy. These three concomitant transitions occur more rapidly in lower- and middle-income countries than they have in western countries (Popkin, 2002).

Many studies have been devoted to the evolution of food consumption in both developed and developing countries. Some of them aim to document how the evolution of the socioeconomic status of country's inhabitants has influenced their diets (see Thang and Popkin, 2004; Burggraf et al., 2015, among others). Recently, Mayen et al. (2014) reviewed 33 studies on this issue. These studies show that (1) high socioeconomic status or living in urban areas is associated with higher intakes of calories, protein, total fat, cholesterol, polyunsaturated, saturated, and mono-unsaturated fatty acids, iron, and vitamins A and C and with lower intakes of carbohydrates and fiber, and (2) high socioeconomic status is also associated with higher fruit and/or vegetable consumption, diet quality, and diversity. The improvement of the socio-economic status of populations

[1] Poverty is measured using poverty headcount at 1.90 US dollars a day (2011 Purchasing Power Parity).

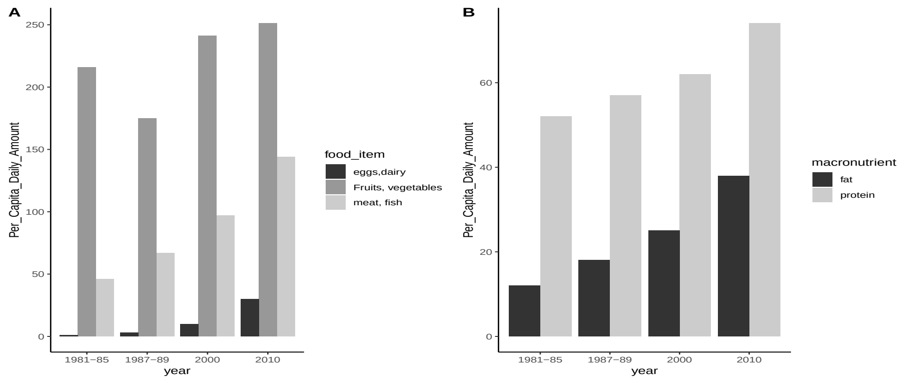
thus seems to lead to a better feeding of human beings. But the other side of the coin is the link between improved diets and noncommunicable diseases as emphasized by Popkin (2006) and Riera-Crichton and Tefft (2014). Thus, both policy makers and citizens are concerned by these concomitant evolutions and the fight against their consequences in terms of nutrition and health. All this requires first of all knowledge of the drivers of these evolutions.

However, most studies mentioned above have important limitations. Some limitations are purely technical. The issue of potential interactions among factors that may explain observed changes in the diet is not addressed. For instance, it is thus conceivable that, in some countries, increasing urbanization of the population is concomitant with the increase in incomes of the same population. Moreover, these studies do not take into account that the link between socio-demographic variables can have changed over the period under study. For instance, the relationship between income and location in an urban area may have become stronger over time. Nevertheless, the major limitation of most studies cited above is that they do not take into account the fact that, in addition to changes in the composition of the population, observed changes in dietary patterns may also be related to changes in consumer behavior (Popkin, 2002). For the same level of education or income, a consumer would buy different amounts of food items at different points of time. Consumer habits may evolve over time (for example, a more pronounced addiction for sweet foods or a growing consumption of dairy products based on the belief that they are good for health, in countries not-used to produce cow's milk) and this can lead to changes in dietary patterns. Thus, the analysis of the drivers of nutritional transition in a given country must consider two potential sources of observed trends: one related to shifts in the composition of the population in that country, and the other to changes in consumer behavior, or behavioral shifts.

Decomposition methods in economics make it possible to evaluate the two aforementioned effects (see Fortin et al., 2011, for an extensive survey). Hereinafter, we use the method recently proposed by Rothe (2015) which, in addition, addresses the two technical problems mentioned above. We apply it to the analysis of the drivers of nutrition transition in Vietnam over the 2004-2014 period. Indeed, thanks to data from Vietnamese Households Living Standard Survey, or VHLSS, we can calculate total energy intakes of Vietnamese households (in Kcal). We convert these intakes into adult equivalent, or per capita, calorie intakes, thus allowing comparison between households. Finally, per capita calorie intakes are decomposed into per capita calorie intakes coming from the three macronutrients: proteins, fat and carbohydrates. VHLSS also contains detailed information on the socio-demographic characteristics of Vietnamese households. Each wave of this survey is, moreover, representative of the Vietnamese population and capture the composition of this population. VHLSS can therefore be used for a comparison of the nutritional status of the Vietnamese population between the two 2004 and 2014 waves.

The remainder of the paper is structured as follows. Section 2 gives a picture of the nutrition transition in Vietnam. Section 3 discusses the pros and cons of usual decomposition methods and introduces the method based on copulas

Figure 1: Trends in per capita daily consumption of foods (panel A) and macronutrients (panel B) (in grams)



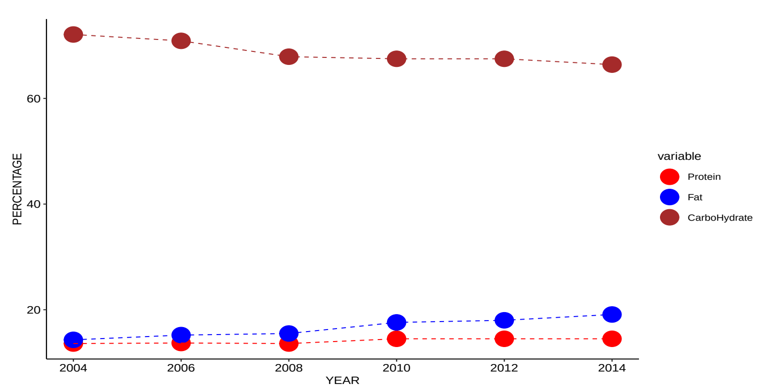
recently proposed by Rothe (2015), we use in this paper. Section 4 gives a description of VHLSS data and details per capita calorie intake computation. Results are presented and commented in section 5. Section 6 draws concluding remarks.

2. Nutrition transition in Vietnam

Recently, Nguyen and Hoang (2018) provide a survey of the literature dealing with the issues of non-communicable diseases, food and nutrition security in Vietnam. They point out that a large part of this literature is gray literature, i.e. reports, presentations . . . They report only 13 published papers including Dien et al. (2004) and Thang and Popkin (2004) who deal with patterns of food consumption at the end of the twentieth century in Vietnam. The detailed analysis of these 13 papers as well as 10 significant reports provides insight into the three transitions discussed in the introduction (epidemiological, demographic, and nutritional). Figure 1, we extracted from Nguyen and Hoang (2018), outlines changes in dietary patterns experienced by Vietnam during the last thirty years. This figure shows an increase in the consumption of meat, fish, eggs, milk and dairy products, and fruits and vegetable, and thus an increase in the consumption of fat and protein between 1981 and 2010. More detailed data covering the period 2004-2014 are presented in Trinh et al. (2018). This paper reveals an increase in the median value of per capita total energy intake, and shows that the share obtained from fat in total calorie intake has increased by 37.5% (resp. 23%) for Vietnamese rural households (resp. urban households), at the expense of calories obtained from carbohydrates, calories obtained from proteins staying quite stable. See Figure 2 using data extracted from Trinh et al. (2018).

This nutrition transition to energy-dense, poor quality diets has led to obesity and diet-related chronic diseases. Using two nationally representative sur-

Figure 2: Trends in macronutrient shares between 2004 and 2014 in Vietnam



veys, Ha et al. (2011) show that the nationwide prevalence of overweight (body mass index $\geq$ 25kg/m^2) and obesity (body mass index $\geq$ 30kg/m^2) was 6.6% and 0.4% respectively in 2005, almost twice the rates of 2000 (3.5% and 0.2%). Using the Asian body mass index cut-off of 23kg/m^2 the overweight prevalence was 16.3% in 2005 and 11.7% in 2000. According to the World Health Organization, the percentage of overweight people in the total population of Vietnam is 21% in 2014, the percentage of obese people being 4%. Moreover, the same study points out that the underweight prevalence (body mass index < 18.5kg/m^2) of 20.9% in 2005 is lower than the rate of 25.0% in 2000. This rate has decreased by half in ten years and is currently 11%. Finally, Ha et al. (2011) also analyze the possible sources of this evolution. They note urban residents were more likely to be overweight and less likely to be underweight compared to rural residents in both years. The shifts from underweight to overweight were clearer among the higher levels of food expenditure.

3. Methodology

We propose to use decomposition methods for assessing the determinants of change in macronutrients consumption in Vietnam using the 2004 and 2014 waves of VHLSS. Decomposition methods were first introduced in order to quantify the contributions of labor, capital, and unexplained factors (productivity) to economic growth (Solow, 1957). They have been extensively used in labor economics, following the seminal papers of Oaxaca (1973) and Blinder (1973). The Oaxaca and Blinder decomposition method is still used in many fields in economics (see Nie et al., 2018, for instance). This method has been refined in several methodological papers and extended to the cases of distributional parameters besides the mean over the last four decades. Fortin et al. (2011) provide a comprehensive overview of these developments. Applications to Vietnam include Nguyen et al. (2007) on urban-rural income inequality, Sakellariou and Fang (2014) on wage inequality and the role of the minimum wage, and recently, Benjamin et al. (2017) and Nguyen et al. (2017) on the sources of inequality in Vietnam. To our knowledge, there is no work using decomposition methods to study the evolution of the nutritional diet and its socio-demographic determinants for Vietnam.

The first objective of decomposition methods is to decompose between-group differences in economic outcomes such as wage or income, into two components: a *composition* effect due to differences in observable covariates across groups, and a *structure* effect due to differences in the relationship that links the covariates, such as the level of education or age, to the considered outcome. The second objective is then to evaluate the individual contributions of these covariates to each of the two effects. Stated differently, let Y denote the outcome of interest (per capita calorie intake in the sequel), with distributions in two different groups denoted by F_Y^t and F_Y^s. Decomposition method deals with decomposing the between-group difference

$$\Delta_Y^\nu = \nu(F_Y^t) - \nu(F_Y^s) \tag{1}$$

where $\nu(.)$ measures a feature of the considered distribution: mean, median, quantile... The question decomposition methods address is thus: how is the between-group difference Δ_Y^ν between the joint distributions in the two groups F_X^t and F_X^s, related to a vector of covariates (denoted by X) ? Implementing the decomposition then requires defining a *counterfactual* distribution. For instance, we can define the counterfactual distribution as what would have been the distribution of outcome in group s if the covariates had been distributed as in group t, or $F_Y^{s|t}$ whose definition is:

$$F_Y^{s|t}(y) = \int_x F_{Y|X}^s(y,x) dF_X^t(x) \tag{2}$$

The between-group difference Δ_Y^ν can then be written as

$$\Delta_Y^\nu = \underbrace{\left(\nu\left(F_Y^t\right) - \nu\left(F_Y^{s|t}\right)\right)}_{\equiv \Delta_S^\nu} + \underbrace{\left(\nu\left(F_Y^{s|t}\right) - \nu\left(F_Y^s\right)\right)}_{\equiv \Delta_X^\nu} \tag{3}$$

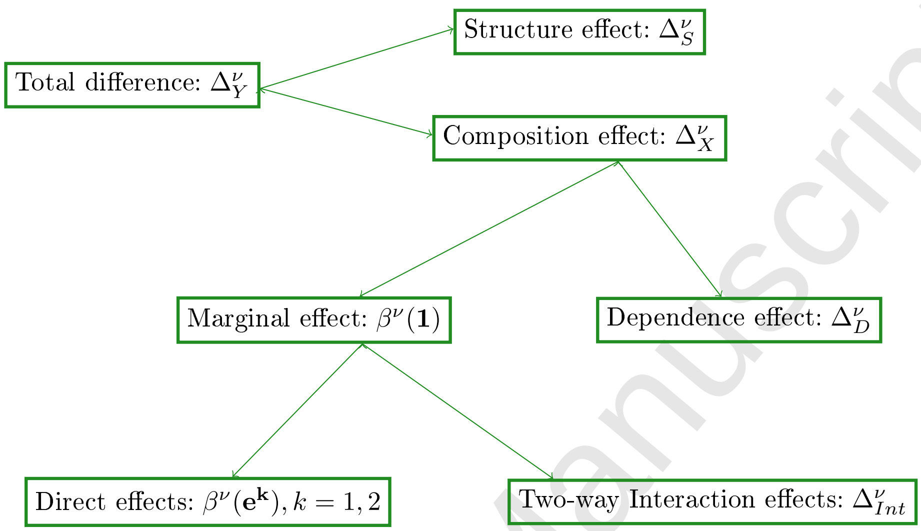
Let us consider the first term in equation (3). The same covariate distribution is used when calculating the two terms in Δ_S^ν, i.e. F_X^t. These two terms only differ by the conditional distribution of the outcome given the covariates used in their computation, i.e. $F_{Y|X}^t$ and $F_{Y|X}^s$. Δ_S^ν captures the between-group difference in the relationship between the outcome and covariates. Δ_S^ν is called the *structure* effect. Consider now the second term in equation (3).The difference between the two terms of Δ_X^ν comes from the marginal covariate distribution used in their computation, i.e. F_X^t and F_X^s. Indeed, the same conditional distribution $F_{Y|X}^s$ is used in their computation. Δ_X^ν is thus defined as the *composition* effect.

The Oaxaca-Blinder decomposition method focuses on between-group difference in mean and is only valid when the data are generated by a linear model. Therefore, Barsky et al. (2002) show that, if the linear model is misspecified, this method leads to misleading classification into structure or composition effects. Moreover, focusing only on the mean introduces a limitation when addressing more general changes in the outcome distribution and it can be more interesting to look at what happens at different quantiles of this distribution. Machado and Mata (2005) and Melly (2005) extended decomposition methods to between-group difference in quantiles. But they kept the linear structure for quantile regression and their methodologies cannot be used to separate the composition effects into the contribution of each covariate. Recently, Chernozhukov et al. (2013) put more theoretical foundations into the construction of the counterfactual distribution needed when decomposing the between group difference Δ_Y^ν into structure and composition effects. But once again, the implementation of the decomposition method is based on a linearity assumption and it is not possible to decompose the composition effect.

The semiparametric reweighting method introduced by DiNardo et al. (1996) can be used to overcome the limitations of the methods discussed so far. The idea of this method is to reweight individual observations depending on whether they are over- or underrepresented when computing the counterfactual distribution. A weighted kernel density routine is used when estimating the counterfactual density of the outcome (see Leibbrandt et al., 2010, for an application). This decomposition method is attractive because it makes no parametric assumptions about the conditional distribution of the outcome given the covariates. Nevertheless, this method is "path dependent" as it does not allow a non-sequential decomposition of the composition effect into the individual covariate effects in the same way they can be decomposed using the conventional Oaxaca-Blinder method.

The Recentered Influence Function, or RIF, regression decomposition method is in line with the Oaxaca-Blinder method, proposing decomposition of the structure and composition effects which is not path-dependent and generalizing it to any distributional feature, not only the mean (Firpo et al., 2007, 2018). In a first stage, the RIF method proposes the decomposition of the distributional statistics of interest into the structure and composition effect using a reweighting approach. The idea of this first stage is very similar to DiNardo et al. (1996) and comes from theoretical results on the identification of distributional statistics besides the mean. In a second stage, the structure and composition effects are decomposed into the contribution of each covariate, using a linear regression-based method. However, unlike the Oaxaca-Blinder method where the dependent variable in the regression is the outcome of interest, the RIF method uses the recentered influence function for the distributional statistics of interest as the dependent variable, following Firpo et al. (2009) work on unconditional quantile regressions. As emphasized by Firpo et al. (2018), "using a linear specification has the advantage of providing a much simpler interpretation of the decomposition, as in Oaxaca-Blinder decomposition." But, they also underline that the linear regression-based procedure only provides a first-order approximation to the composition effect. So, there is an approximation error. This error can be obtained as a by-product of the decomposition of the composition effect and used as a specification test of the regression-based procedure.

Figure 3: **Decomposition framework**



If its size is small, it may be thought that the approach of Firpo et al. (2007, 2018) provides an accurate enough approximation for the problem at hand. No solution is proposed when the approximation error is large. So, for example, should we incorporate interactions between variables? To our knowledge, no response is given in the literature on the RIF method. And although it has been applied in many fields (see Etile, 2014, for instance), few empirical works provide a discussion about the magnitude of the approximation error.

In this paper, we use the copula-based decomposition procedure recently proposed by Rothe (2015). This method also allows decomposition involving any distributional feature, not only the mean. But, unlike most of the methods mentioned above, this method (1) does not rely on any linearity assumption, (2) is not path-dependent, and (3) introduces into this decomposition the effect of the interaction between covariates. Specifically, this decomposition method expands classical methods by adding to the usual decomposition of the composition effect into the *direct contribution* of each covariate due to between-group differences in their respective marginal distributions, several *two way* and *higher order interaction effects* due to the interplay between two or more covariates, and a *dependence effect*, accounting for the between-group difference in the dependence pattern among the covariates. The copula-based decomposition procedure therefore provides a detailed analysis of one of the two sides of the decomposition, the composition effect, but it says nothing about the structure effect. The issue of deriving a decomposition of the structure effect, that is, dividing between-group differences in the structural functions that link the covariates and the outcome variable, into components that can be attributed to individual covariates, still is an open issue.

To get a better understanding of the goals of the decomposition method we use, we will illustrate it by a simple example. Here, we analyze the difference in calorie intake distributions for two years, 2004 and 2014. Our outcome is measured by per capita calorie intake. We are interested in two potential drivers of the difference in per capita calorie intake distributions in 2004 and 2014: (1) evolution of Vietnamese households' food expenditures, and (2) urbanization. For instance, Vietnamese households increased their food spending between 2004 and 2014 and Vietnamese population is more urban in 2004 than in 2014. Moreover urban citizens tend to spend more on food (dependence between these two covariates) hence leading to an extra increase in overall food expenditures. We are interested by decomposing the difference between per capital calorie intake averages in 2014 and 2004. The *structure* effect is the part of this difference that can be explained by the between-year difference in the conditional distributions of per capita calorie intake given food expenditures and location in an urban area. The *composition* effect is the part of the difference that can be explained by the between-year differences in observable characteristics (food expenditures and living in an urban area). The first *direct* contribution is the part of the composition effect that can be attributed to the fact that Vietnamese households have higher food expenditures in 2014 compared to 2004. The second *direct* effect captures the part in the composition effect due to the fact that the Vietnamese population is more urban in 2014 than in 2004. The (only) *interaction* effect measures the additional contribution of the fact that the Vietnamese population at the same time spends more for food and is more urban in 2014. Finally, the *dependence* effect accounts for between-year differences in association patterns among the two covariates, food expenditures and location in an urban area. In other words, the *dependence* effect captures the fact that the relative food expenditure of urban and rural households differs in the two years. Figure 3 summarizes the different stages of the copula-based decomposition method. The notations used for the different elements in the decomposition correspond to those used in appendices A and B where technical details about the copula-based decomposition method and its implementation are given.

4. Data

This study relies on the "Vietnam Household Living Standard Survey", or VHLSS. This survey is conducted by the General Statistics Office of Vietnam, or GSO, with technical assistance of the World Bank, every two years since 2002. Each VHLSS survey contains modules related to household demographics, education, health, employment, income generating activities, including household businesses, and expenditures. The survey is conducted in all of the 64 Vietnamese provinces and data are collected from about 9000 households for each wave. The survey is nationally representative and covers rural and urban areas.[2]

[2]The reader is referred to Trinh Thi et al. (2018) for a detailed description of VHLSS.

Below, we use the two waves of VHLSS conducted in 2004 and 2014 as our groups when implementing the copula-based decomposition method. We do not choose the 2002 wave as our reference group since food consumption was less detailed than in the subsequent waves: only 16 aggregate food items in 2002 instead of 56 in the following waves. Moreover, the demarcation of households is not the same in 2002 and in the following waves. In 2002, VHLSS considered all persons present in the household at the time of the survey as members of the household. Thereafter, household membership has been defined on the basis of physical presence: individuals must eat and live with other members for at least six out of the past twelve months, and contribute to collective income and expenses. Among other things, this means that family members who have moved away to work or school (e.g., migrants) are not considered anymore as household members.

4.1. Sociodemographic variables

Table 1 summarizes the sociodemographic variables we use as covariates when implementing the copula-based decomposition method. Detailed summary statistics on these variables are given in Table 2. These statistics show several interesting developments. First, total food expenditures of Vietnamese households increased over the considered period. Second, the population of these same households is more urbanized in 2014 than in 2004. Third, the average household size has decreased slightly, with about 65% of these households having four or fewer members in 2014 compared to about 55% ten years earlier. Fourth, heads of households are, on average, more educated in 2014 than in 2004. Furthermore, the proportion of heads with more than 12 schooling years (high school level) increased significantly from 2004 to 2014. Finally, the proportions of households with heads belonging to the Kinh ethnicity or living in South Vietnam remained stable.

4.2. Macronutrient intakes

Average annual or monthly food expenditures and quantities about 56 food items are collected for each household surveyed in the two considered VHLSS waves.[3] Conversion factors of grams into calories coming from the food composition table constructed by the Vietnam National Institute of Nutrition in 2013 are used to compute macronutrient consumption amounts (see Table 3). For each household, we compute the total calorie intake (in Kcal), and the protein and fat intakes (in grams) per day. Then, we convert for each household the quantity in grams of protein (resp. fat) into Kcal by multiplying by 4 (resp. 9). Carbohydrate intakes is obtained by subtracting the protein and fat intakes (in Kcal) to total calorie intake (in Kcal). Finally, we compute adult-equivalent calorie intakes to make household comparable. We use the household equivalence scale calculation procedure recently proposed by Aguiar and Hurst (2013),

[3]Only average annual food consumption was recorded in 2004 while monthly average food consumption was surveyed in 2014.

Table 1: **Description of sociodemographic variables**

Variable	Values	Description
$lExp$		Food expenditures per year in US$ (in logarithms) (inflation adjusted)
$Hsize$		Number of household members
$Urban$		Location of the household:
	$= 1$	if household is located in Urban area
	$= 0$	if household is located in rural area
$Ethnic$		Ethnicity of head of household
	$= 1$	if Kinh Ethnicity
	$= 0$	if minority
$Yeducc$		Highest educational level of the head of households (year):
	$= 0$	No schooling
	$= 5$	primary level
	$= 9$	Secondary school level
	$= 12$	High school level
	$= 16$	College degree
	$= 18$	Master degree
	$= 21$	Ph.D level
$South$		Region:
	$= 1$	if Household is located in the South of Vietnam
	$= 0$	otherwise

to calculate the per capita calorie intake (namely $PCCI$), per capita volume of calories obtained from protein (namely VP), per capita volume of calories obtained from fat (namely VF), and per capita volume of calories obtained from carbohydrate (namely VC) for each household. These per capita measures are obtained by dividing the corresponding energy intakes by household specific equivalence scale.[4] Table 2 gives various summary statistics on $PCCI$, VC, VP, and VF. These statistics will be discussed below.

The previous calculations of calorie intake may suffer from measurement errors. Indeed, recall food expenditure data, such as collected in VHLSS, is believed to suffer from considerable measurement error. Diary records should be preferred because they are believed to be more accurate when dealing with calorie intake measurement. But a clear evidence of the superiority of the latter relative to the recall data does not seem to emerge from the literature. For instance, a recent study using a unique data set that collects recall and diary data from the same households show evidence that, while recall data exhibits substantial measurement errors, diary measures are themselves imperfect (Brzozowski et al., 2007). Moreover, as emphasized by Magrini et al. (2017) which use VHLSS data to assess household vulnerability from trade, “it is generally agreed that VHLSS data can be considered to be of high quality and provide legitimate nationally representative household data based on stratified random samples.” But, even if the VHLSS survey meets all the criteria defining a "good" living standard survey (Deaton and Grosh, 2000; Browning et al., 2014), we can-

[4]The computation of household specific equivalence scales using Aguiar and Hurst (2013) is detailed in Trinh Thi et al. (2018).

Table 2: **Descriptive statistics by VHLSS wave**

	Mean	Standard deviation	Quantile 10%	50%	90%
			Wave = 2004		
$PCCI$	3359.746	1015.451	2259.852	3195.859	4609.399
V_C	2415.078	756.170	1565.208	2318.522	3343.795
V_P	457.920	156.403	294.643	428.629	653.904
V_F	486.748	239.576	247.206	433.159	792.876
$lExp$	6.135	0.547	5.461	6.125	6.844
$Urban$	0.235		–	–	–
$Hsize$	4.355	1.636	2	4	6
$Ethnic$	0.893		–	–	–
$Yeducc$	6.222	4.712	0	5	12
$South$	0.345		–	–	–
			Wave = 2014		
$PCCI$	3764.194	1421.362	2313.206	3488.157	5528.041
V_C	2493.419	1032.906	1445.146	2297.777	3764.969
V_P	548.367	219.059	320.181	501.073	830.010
V_F	722.409	343.119	367.404	647.299	1174.950
$lExp$	6.638	0.611	5.843	6.667	7.399
$Urban$	0.311		–	–	–
$Hsize$	3.808	1.526	2	4	6
$Ethnic$	0.869		–	–	–
$Yeducc$	7.097	5.047	0	9	12
$South$	0.339		–	–	–
			Differences 2014 - 2004		
$PCCI$	404.448	405.911	53.354	292.298	918.642
V_C	78.341	276.736	-120.062	-20.745	421.174
V_P	90.447	62.656	25.338	72.444	176.106
V_F	235.661	103.543	120.198	214.140	382.074
$lExp$	0.503	0.064	0.382	0.542	0.555
$Urban$	0.076		–	–	–
$Hsize$	0.447	0.110	0	0	0
$Ethnic$	-0.024		–	–	–
$Yeducc$	0.895	0.335	0	4	0
$South$	0.006		–	–	–

not avoid all possible sources of measurement errors. Indeed, following Nakata et al. (2009), measurement errors can be expected to be related to household size. Consequently, caution should be exercised in interpreting a significant effect of this last variable in the change in caloric intake between 2004 and 2014.

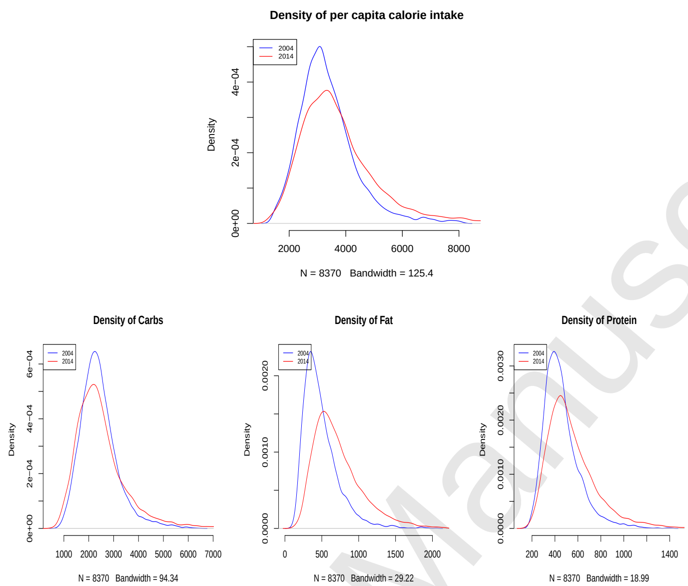
The calculation procedure presented above allows more to compute calorie availability than calorie intake (Bouis and Haddad, 1992). Indeed, VHLSS data do not include losses and waste from food preservation and preparation. Calorie intakes are thus overestimated. Unlike United States for which these losses have been evaluated for each food item (Muth et al., 2011), reliable data on food losses and waste are not yet available for Vietnam, and differences in consumption habits between the two countries prevent us from applying the estimated loss coefficients for the US to Vietnamese data. As in Muth et al. (2011), we assume that there is a systematic bias to overestimation when transforming consumption data into nutrition data. This bias is assumed to be the same regardless of the considered household. Due to lack of data allowing a thorough treatment of this assumption, we maintain it in this paper.

Table 3: **Conversion table**

Food Item	Energy (Kcal)	Protein (gram)	Fat (gram)
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
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	-	-	-
Lard, cooking oil	863.5	0	99.8
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	-	-	-

Note: Amount per 100 grams of food ; protein contains 4 calories per gram and fat contains 9 calories per gram

Figure 4: Density of per capita calorie intake



5. Results

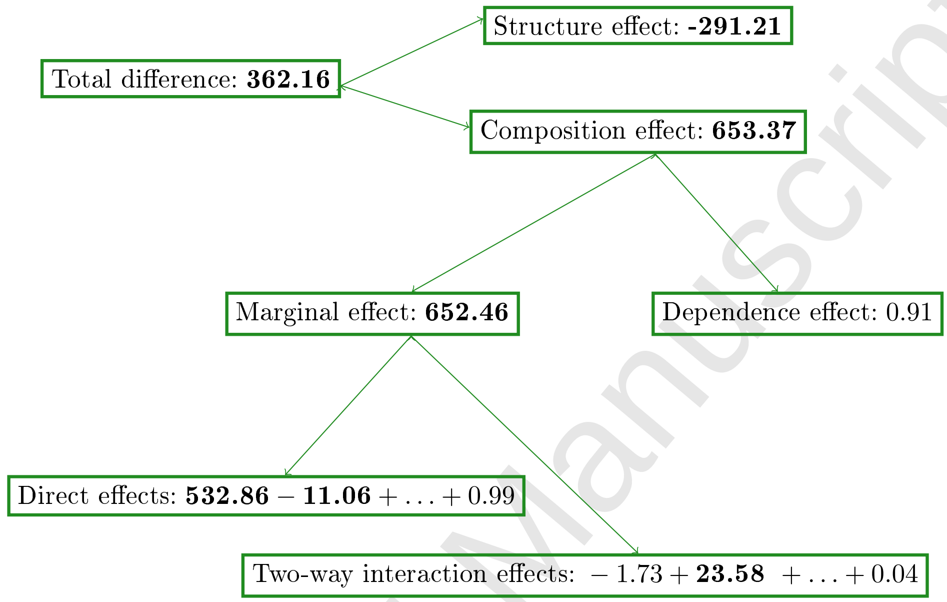
5.1. Changes in per capita calorie intake distributions

Figure 4 reports the weighted kernel estimates[5] of the densities of per capita calorie intake for the two years. There is a shift to the right for the density from 2004 to 2014, indicating an increase in per capita calorie intake over the period, not only on average but also for all quantiles such as those reported in Table 2.

Figure 4 also reports the weighted kernel estimates of the densities of per capita calorie intakes of carbohydrates, fat, and proteins, for the two years. Significant changes appear when comparing the estimated densities for fat and proteins, while the estimated densities for carbohydrates appear to be very similar. There is a significant shift to the right for the estimated densities for fat and protein in 2014. Meanwhile, the estimated density for carbohydrates in 2014 has the same mode as in 2004, but becomes flatter. This visual observation is confirmed by the evolution of average values, standard deviations, and quantiles at 10, 50 and 90% as reported in Table 2. All these values increase significantly for fat and proteins. Average and median values stay quite stable between 2004 and 2014 for carbohydrates while standard deviation increases, 10% quantile decreases, and 90% quantile increases. In other words, per capita calorie intakes from fat and proteins in Vietnamese households have increased

[5] All observations are weighted by their respective VHLSS sample weights.

Figure 5: **Decomposition of the between-year difference in mean, in per capita calorie intake**



over the considered period. Per capita calorie intake from carbohydrates remained stable on average, while this stability hides a contrasted picture with an increase for some households and a decrease for others.

Table 4: **Results of Kolmogorov-Smirnov test of equality of the CDFs in 2004 and 2014**

Variable	KS-statistics	p-value
PCCI	0.14	< 0.0001
Protein	0.20	< 0.0001
fat	0.34	< 0.0001
Carbohydrate	0.06	< 0.0001

Kernel estimates of per capita calorie intake densities and summary statistics show the observed changes in energy intake and macronutrients consumption over the 2004-2014 period. Kolmogorov-Smirnov nonparametric test (Conover, 1971) can be implemented to confirm if these changes are significant or not. Results are presented in Table 4. The null hypothesis of equality of the CDFs in 2004 and 2014 is clearly rejected whatever the considered variable. We can now analyze which covariates have contributed significantly to these changes over time.

5.2. Decomposition results

Significant changes in distributions of per capita intake (total or macronutrients) between 2004 and 2014, have just been noted. What are the main drivers of these changes? To answer this question, we implement the decomposition method proposed in Rothe (2015). This implementation is based on the preliminary estimation of marginal distributions of covariates, dependence parameters in copulae and conditional distributions of per capita calorie intake given covariates. These results, which are not central compared to those of decomposition, are summarized in Appendix C.

Tables 5, 6, 7, and 8 present the results of our decomposition of per capita total calorie intake and calorie intake coming from the three macronutrients, for two measures of location: mean and median, and for the two quantiles at 10% and 90%. Row by row, we report estimates of total change, i.e. Δ_Y^ν, usual structure and composition effects, i.e. Δ_S^ν and Δ_X^ν. Then the composition effect is in turn decomposed into the dependence effect, i.e. Δ_D^ν, and marginal distribution effect, i.e. $\beta^\nu(\mathbf{1})$. Finally, this last effect is decomposed into the direct contribution for each of the six covariates, i.e. the $\beta^\nu(\mathbf{e^k})$, and the "two-way" interaction effects, i.e. the Δ_{Int}^ν. Each estimated value in a decomposition is accompanied by the estimated value of its standard error. Theses standards errors are computed using using nonparametric bootstrap with 300 replications as proposed by Rothe (2015).[6]

Figure 5 explains how tables 5, 6, 7, and 8 should be read. This figure considers the decomposition of the between-year difference in mean, in per capita calorie intake presented in table 5. Bold estimated values are significantly different from zero.

Overall quality of decompositions. Knowledge of the estimated values of total difference and the associated standard errors first allow to have an indication as to whether the chosen modeling of decomposition using parametric restrictions on copulas and conditional distributions, provides a reasonable fit. Indeed, these estimated values of total difference can be compared with the differences that can be directly calculated from the descriptive statistics given in Table 2. It should be noted that, in all cases, the difference computed from the descriptive statistics belongs to the 95% confidence interval that can be constructed from the estimated value of total difference and its estimated standard error. Moreover, the estimated values of total difference for quantiles capture well the observed shifts in empirical quantiles of calorie intake distributions. The chosen model thus provides a reasonable fit to the data.

[6] Rothe (2015) shows the asymptotic convergence of the estimator of each element in a decomposition to a mean zero normal distribution. But, as the asymptotic variance of these estimators takes a fairly complicated form, a practical way to estimate this variance is standard nonparametric bootstrap in which the estimates are recomputed a large number of times on bootstrapped samples $\{\widetilde{Y}_i^t, \widetilde{X}_i^t\}_{i=1}^{n_t}$ drawn with replacement from the original data $\{Y_i^t, X_i^t\}_{i=1}^{n_t}$. The bootstrap variance estimator then coincides with the empirical variance of the bootstrapped estimates.

Table 5: **Estimated decomposition of per capita calorie intake**

	Mean		Quantiles 10%		Quantiles 50%		Quantiles 90%	
	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error
Total difference Δ_Y^ν	362.16	(28.90)	18.62	(26.38)	279.48	(20.81)	830.71	(78.76)
Structure effect Δ_S^ν	-291.21	(52.53)	-283.63	(49.12)	-361.79	(39.69)	-328.38	(213.09)
Composition effect Δ_C^ν	653.37	(44.47)	302.25	(46.16)	641.27	(36.71)	1159.09	(202.78)
Composition effect:								
Dependence effect Δ_D^ν	0.91	(23.08)	-30.6	(22.94)	0.25	(23.26)	-7.97	(135.35)
Marginal effect Δ_M^ν	652.46	(39.97)	332.85	(42.83)	641.02	(33.65)	1167.06	(206.92)
"Direct" contributions to composition effect:								
lexp	532.86	(36.16)	250.05	(35.54)	538.66	(33.05)	900.04	(137.13)
Urban	-11.06	(2.90)	-11.55	(3.28)	-9.56	(3.41)	-9.49	(8.96)
Hsize	131.12	(8.94)	53.62	(8.46)	112.77	(12.26)	246.08	(27.01)
Ethnic	0.69	(1.53)	1.90	(1.33)	0.34	(1.69)	-1.61	(3.95)
Yeduc	-18.16	(7.08)	-3.09	(6.21)	-14.00	(5.72)	-26.18	(12.53)
South	0.99	(1.06)	0.88	(0.96)	0.83	(1.30)	1.11	(1.30)
"Two-way" interaction effects:								
lexp:Urban	-1.73	(5.08)	7.41	(6.83)	-9.83	(9.02)	3.38	(23.43)
lexp:Hsize	23.58	(10.38)	50.01	(24.08)	30.04	(20.97)	34.87	(129.97)
lexp:Ethnic	0.61	(2.70)	3.89	(3.58)	2.36	(4.12)	2.36	(14.9)
lexp:Yeduc	-6.14	(6.01)	-7.56	(10.87)	-11.06	(11.26)	-14.48	(32.18)
lexp:South	0.44	(0.70)	0.03	(0.80)	0.21	(1.28)	0.29	(3.96)
Urban:Hsize	0.14	(1.19)	2.62	(2.73)	-6.45	(4.10)	1.47	(6.32)
Urban:Ethnic	-0.45	(0.29)	-0.54	(0.47)	-0.26	(0.54)	-0.39	(1.48)
Urban:Yeduc	0.41	(0.81)	0.17	(1.22)	-2.37	(2.39)	-1.43	(3.73)
Urban:South	-0.20	(0.22)	-0.01	(0.19)	-0.05	(0.31)	-0.64	(0.70)
Hsize:Ethnic	0.84	(0.48)	0.90	(1.09)	1.25	(1.37)	0.73	(2.68)
Hsize:Yeduc	-2.38	(2.05)	-1.76	(3.85)	-14.84	(6.15)	-5.78	(10.89)
Hsize:South	-0.06	(0.15)	-0.43	(0.46)	0.63	(0.72)	-0.22	(0.74)
Ethnic:Yeduc	-0.32	(0.40)	-0.61	(0.50)	0.29	(0.62)	-0.17	(1.63)
Ethnic:South	0.03	(0.05)	0.05	(0.09)	-0.01	(0.14)	-0.10	(0.14)
Yeduc:South	0.04	(0.07)	-0.01	(0.14)	0.04	(0.38)	-0.20	(0.44)

Note: Bootstrapped standard errors are computed using 300 replications.

Table 6: **Estimated decomposition of calorie intake from fat**

	Mean		Quantiles					
			10%		50%		90%	
	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error
Total difference Δ_Y^ν	221.51	(8.68)	119.61	(6.13)	200.73	(7.06)	364.92	(28.06)
Structure effect Δ_S^ν	-17.63	(13.93)	-1.92	(6.8)	-15.85	(10.51)	-5.33	(55.00)
Composition effect Δ_C^ν	239.14	(12.39)	121.53	(8.13)	216.57	(9.58)	370.25	(55.94)
Composition effect:								
Dependence effect Δ_D^ν	-0.77	(8.01)	2.34	(5.94)	-0.67	(5.01)	6.33	(46.76)
Marginal effect Δ_M^ν	239.91	(11.11)	119.19	(5.95)	217.24	(8.49)	363.92	(54.53)
"Direct" contributions to composition effect:								
lexp	178.97	(9.34)	80.78	(6.57)	173.74	(7.12)	296.02	(36.63)
Urban	2.51	(0.68)	0.77	(0.3)	2.39	(0.75)	2.91	(2.11)
Hsize	47.64	(2.68)	26.16	(2.55)	44.33	(2.74)	71.02	(7.29)
Ethnic	0.24	(0.44)	-0.18	(0.21)	0.14	(0.33)	-0.65	(1.17)
Yeduc	-0.98	(0.92)	-0.04	(1.19)	0.75	(1.27)	-6.75	(2.79)
South	0.28	(0.34)	0.10	(0.15)	0.30	(0.37)	0.67	(0.93)
"Two-way" interaction effects:								
lexp:Urban	0.31	(1.37)	2.62	(1.28)	-1.40	(1.24)	-3.23	(7.07)
lexp:Hsize	10.13	(3.30)	9.36	(5.19)	-1.56	(5.72)	25.03	(32.16)
lexp:Ethnic	0.54	(0.91)	0.48	(0.47)	-0.22	(0.75)	1.85	(3.51)
lexp:Yeduc	-0.56	(1.26)	1.83	(2.46)	-2.07	(2.02)	2.70	(7.42)
lexp:South	0.19	(0.23)	0.00	(0.16)	0.13	(0.38)	-0.14	(0.69)
Urban:Hsize	0.54	(0.33)	0.62	(0.32)	-0.74	(0.67)	1.63	(1.90)
Urban:Ethnic	-0.02	(0.13)	-0.01	(0.03)	0.04	(0.11)	0.07	(0.31)
Urban:Yeduc	-0.05	(0.13)	-0.01	(0.10)	-0.32	(0.25)	0.20	(0.86)
Urban:South	-0.04	(0.05)	0.00	(0.02)	-0.03	(0.06)	0.07	(0.21)
Hsize:Ethnic	0.04	(0.16)	0.14	(0.16)	0.35	(0.27)	0.53	(0.76)
Hsize:Yeduc	-0.33	(0.32)	0.77	(1.03)	-0.83	(0.90)	-0.88	(2.71)
Hsize:South	0.04	(0.05)	0.05	(0.06)	0.01	(0.14)	0.01	(0.46)
Ethnic:Yeduc	-0.08	(0.10)	-0.04	(0.06)	-0.06	(0.11)	-0.34	(0.47)
Ethnic:South	0.00	(0.02)	0.00	(0.01)	0.00	(0.02)	0.03	(0.10)
Yeduc:South	0.02	(0.02)	0.00	(0.03)	0.01	(0.06)	0.01	(0.29)

Note: Bootstrapped standard errors are computed using 300 replications.

Table 7: **Estimated decomposition of calorie intake from protein**

	Mean		Quantiles					
			10%		50%		90%	
	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error
Total difference Δ_Y^ν	85.74	(4.54)	23.76	(4.37)	70.93	(3.97)	163.34	(11.17)
Structure effect Δ_S^ν	-52.32	(9.08)	-43.97	(7.05)	-61.23	(7.35)	-41.46	(31.46)
Composition effect Δ_C^ν	138.06	(8.09)	67.73	(7.06)	132.16	(7.18)	204.8	(30.44)
Composition effect:								
Dependence effect Δ_D^ν	2.94	(6.06)	2.80	(4.98)	2.66	(4.74)	-5.93	(25.79)
Marginal effect Δ_M^ν	135.12	(7.13)	64.93	(6.78)	129.5	(6.73)	210.73	(26.85)
"Direct" contributions to composition effect:								
lexp	108.11	(6.08)	49.37	(5.10)	108.17	(5.54)	169.77	(19.08)
Urban	-0.57	(0.40)	-0.82	(0.32)	-0.91	(0.45)	0.14	(1.06)
Hsize	26.89	(1.43)	15.33	(1.63)	23.34	(1.46)	43.44	(6.05)
Ethnic	-0.21	(0.19)	-0.04	(0.15)	-0.27	(0.26)	-0.37	(0.54)
Yeduc	-1.84	(0.82)	-0.29	(0.94)	-2.44	(0.86)	-4.16	(1.51)
South	-0.06	(0.09)	0.00	(0.03)	-0.01	(0.04)	-0.08	(0.17)
"Two-way" interaction effects:								
lexp:Urban	-0.23	(0.98)	1.28	(0.99)	-2.89	(1.42)	4.36	(3.93)
lexp:Hsize	2.74	(2.10)	0.80	(4.47)	6.84	(4.70)	-7.37	(17.44)
lexp:Ethnic	-0.56	(0.44)	-0.11	(0.53)	-0.29	(0.60)	-1.60	(2.56)
lexp:Yeduc	0.14	(0.86)	-1.41	(1.97)	-1.76	(1.77)	-1.35	(3.93)
lexp:South	0.14	(0.21)	0.06	(0.10)	0.21	(0.28)	0.08	(0.32)
Urban:Hsize	0.28	(0.19)	-0.10	(0.28)	0.25	(0.37)	1.51	(1.49)
Urban:Ethnic	-0.04	(0.04)	0.00	(0.04)	-0.06	(0.08)	0.03	(0.12)
Urban:Yeduc	-0.03	(0.09)	-0.10	(0.16)	-0.01	(0.30)	0.12	(0.31)
Urban:South	-0.03	(0.04)	0.00	(0.01)	-0.01	(0.02)	-0.08	(0.10)
Hsize:Ethnic	0.04	(0.06)	0.19	(0.12)	0.19	(0.15)	-0.19	(0.66)
Hsize:Yeduc	-0.15	(0.28)	-0.07	(0.78)	-0.18	(0.74)	-1.24	(1.85)
Hsize:South	-0.02	(0.02)	-0.01	(0.03)	0.01	(0.03)	-0.19	(0.23)
Ethnic:Yeduc	-0.02	(0.04)	0.03	(0.05)	-0.02	(0.12)	-0.09	(0.20)
Ethnic:South	0.00	(0.05)	0.00	(0.00)	0.00	(0.00)	0.01	(0.04)
Yeduc:South	0.00	(0.01)	0.00	(0.01)	0.01	(0.02)	-0.01	(0.04)

Note: Bootstrapped standard errors are computed using 300 replications.

Table 8: **Estimated decomposition of calorie intake from carbohydrates**

	Mean		Quantiles 10%		50%		90%	
	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error	Estimated value	Standard error
Total difference Δ_Y^ν	45.36	(22.04)	-120.24	(21.14)	-27.88	(18.25)	372.23	(50.91)
Structure effect Δ_S^ν	-244.93	(36.73)	-238.99	(30.56)	-252.19	(24.66)	-300.78	(134.48)
Composition effect Δ_C^ν	290.29	(32.66)	118.75	(29.32)	224.31	(22.52)	673.01	(128.18)
Composition effect:								
Dependence effect Δ_D^ν	0.68	(17.57)	-6.03	(15.75)	5.50	(10.70)	18.43	(74.42)
Marginal effect Δ_M^ν	289.61	(29.3)	124.78	(24.67)	218.81	(19.74)	654.58	(110.40)
"Direct" contributions to composition effect:								
lexp	253.01	(25.7)	135.13	(20.79)	207.47	(19.2)	528.18	(88.29)
Urban	-12.63	(2.34)	-16.17	(3.88)	-13.01	(2.78)	-11.54	(6.03)
Hsize	59.84	(5.07)	19.53	(6.60)	43.53	(5.54)	140.86	(22.94)
Ethnic	0.60	(1.35)	1.29	(1.95)	0.93	(1.45)	-0.50	(2.55)
Yeduc	-15.53	(5.07)	-6.82	(6.62)	-19.57	(4.18)	-18.59	(7.90)
South	0.75	(0.78)	0.93	(1.00)	0.83	(0.94)	0.47	(0.69)
"Two-way" interaction effects:								
lexp:Urban	-2.20	(4.04)	8.73	(5.57)	-0.93	(5.67)	-14.17	(17.3)
lexp:Hsize	13.24	(6.80)	-7.25	(8.55)	-1.78	(10.89)	14.15	(59.55)
lexp:Ethnic	0.83	(1.92)	4.82	(2.70)	-0.38	(1.78)	3.43	(8.75)
lexp:Yeduc	-6.16	(4.28)	-9.55	(8.78)	-2.97	(7.21)	0.29	(16.59)
lexp:South	0.08	(0.48)	-0.52	(0.76)	-0.25	(0.56)	1.08	(2.45)
Urban:Hsize	-0.73	(0.79)	0.20	(3.37)	0.08	(3.28)	3.76	(5.80)
Urban:Ethnic	-0.41	(0.25)	-0.52	(0.61)	-0.67	(0.54)	-0.39	(0.78)
Urban:Yeduc	0.52	(0.52)	1.80	(2.12)	3.10	(2.39)	-0.73	(1.91)
Urban:South	-0.14	(0.16)	-0.29	(0.26)	-0.09	(0.27)	-0.32	(0.35)
Hsize:Ethnic	0.80	(0.34)	1.69	(1.02)	0.68	(0.91)	1.22	(2.63)
Hsize:Yeduc	-1.93	(1.30)	-4.10	(2.70)	-1.57	(4.60)	-12.03	(7.96)
Hsize:South	-0.09	(0.14)	0.00	(0.37)	-0.04	(0.33)	-0.07	(0.52)
Ethnic:Yeduc	-0.28	(0.29)	-0.68	(0.74)	-0.28	(0.67)	-0.27	(0.67)
Ethnic:South	0.03	(0.04)	0.07	(0.13)	0.03	(0.09)	0.03	(0.06)
Yeduc:South	0.03	(0.05)	-0.08	(0.22)	-0.07	(0.32)	0.00	(0.15)

Note: Bootstrapped standard errors are computed using 300 replications.

Per capita total energy intake. Let us now look more closely at each of the tables. Table 5 presents the estimated values of the various elements in the decomposition of differences in mean, median and quantiles at 10% and 90% between the two years for per capita calorie intake. The decomposition of total difference in structure effect and composition effect reveals two effects that play in opposite directions. A strong positive composition effect then appears while the structure effect is negative and quite stable among quantiles. The composition effect is only counterbalanced by the structural effect in the case of the quantile at 10%. Moreover, the composition effect increases with the quantile.

In other words, the between-year change in the conditional distributions of per capita calorie intake given the sociodemographic characteristics, or in the relationship between per capita calorie intake and these covariates, caused a significant decrease in per capita calorie intake on average as well as on the three considered quantiles. Meanwhile, the change in the composition of the sample of households between the two years led to a significant increase in per capita calorie intake. This increase was larger than the decrease due to changes in the relationship between per capita calorie intake and sociodemographic variables, except for the 10% quantile where the two compensate.

The dependence effect that captures the contribution of between-year differences in the covariates' copula functions plays no role in the decomposition of composition effect. The dependence effect is never significantly different from zero. The composition effect is almost always equal to the marginal effect resulting from differences in the marginal covariate distributions across the two years.

Consider now the decomposition of the marginal effect into direct effects of each covariate and "two-way" interactions effects. This decomposition shows the importance of the contribution of food expenditures and household size to total marginal distribution effect, i.e., here, the composition effect. These contributions are indeed positive, large, and significantly different from zero. It should be noted that these contributions increase according to the considered quantile order. Food expenditures and household size play a more and more important role in the increase of per capita calorie intake when moving from the 10% quantile to the 90% quantile. The effects of these two variables are barely offset by the negative and significantly different from zero effects of urbanization and years of education of the head of the household. Moreover, almost all "two-way" interaction effects are negligible.

Macronutrients. Similar comments can be made regarding decompositions for consumption in terms of calories from fat and protein (see Tables 6 and 7) Thus, the estimated values of the total difference for the different quantiles closely trace the observed uniform changes in these distributions towards higher consumption of the two macronutrients. Again, the main source of change comes from the composition effect that the structural effect only partially compensates for. It should be noted that the structural effect is never significantly different from zero in the case of fat. The dependence effect is negligible, and the main contributors to the composition effect are still food expenditures and household size. The estimated values of the impacts of these two covariates on changes in consumed calories from fat and protein increase when moving from the 10% quantile to the 90% quantile. Now, the number of years of education of the head of household still impacts negatively on changes, the effects being sometimes not significantly different from zero. The effect of urbanization is negligible in the case of proteins, whereas it becomes positive in the case of fat. Nevertheless, although significantly different from zero for most of the considered statistics, the effect of urbanization is negligible when compared to those of food expenditure or household size.

The results obtained in the case of carbohydrates are more contrasted than the previous ones (see Table 8). Here again, the estimated values of the total difference trace well what is observed for the empirical distributions of calories consumed from carbohydrates, whether in terms of location or spread statistics. Thus, total differences for mean and median are not significantly different from zero at the 10% and 5% threshold respectively, while total differences for 10% and 90% quantiles are significantly different from zero, the first being negative while the second is positive. The results capture well the flattening of the distribution between 2004 and 2014. But now, the structure effect compensates the composition effect in the cases of the mean and median, or even exceeds it for the 10% quantile when decomposing total difference. As for the decomposition of the composition effect, it gives rise to similar comments to those made above for per capita calorie intake: negligible dependence effect, and strong positive contributions of food expenditures and household size compensated in part by negative contributions of urbanization and level of education of head of household.

6. Conclusion and discussions

The aim of this paper is to document the evolution of Vietnamese household consumption in terms of total calorie intake and consumption of macronutrients over the period 2004-2014. The availability of VHLSS surveys makes it possible to have detailed data on these consumptions. The descriptive analysis of the data reveals an increase in per capita calorie intake over the period not only on average but also at all the quantiles of the corresponding distribution. The same evolution is observed for the consumption of proteins and that of fat. The distribution of carbohydrate consumption, on the other hand, flattens, showing an increase in low and high consumption between the two years while staying stable on average.

The characterization of the drivers of these evolutions is based on the use of a decomposition method recently proposed by Rothe (2015). In addition to the classical decomposition of between-year changes in terms of structure and decomposition effects, this method allows to compute the direct contributions of various socio-demographic variables and the effects of their interactions in these between-year changes. Moreover, this method allows to identify the impact of between-year changes in the dependence structure among covariates.

We implement Rothe (2015) method on VHLSS data to characterize the different effects on between-year mean, median, and 10% and 90% quantiles changes in per capita calorie intake and macronutrient consumptions in Vietnam. The main results we obtained can be summarized as follows (see also Figures 6 and 7 that summarize most of the results presented in Tables 5, 6, 7, and 8):

1. Decompositions using parametric restrictions on copulas and conditional distributions provide a reasonable fit. The estimated values of the between-year total differences clearly reflect the observed differences, either on average or for the considered quantiles.
2. The structure and composition effects play in an opposite direction, whatever the considered decomposition. Structure effects, which come from between-year differences in the relationship that links the covariates to the considered outcome, are always negative, while composition effects, which are due to differences in the distributions of observable covariates across years, are always positive.
3. The composition effect often outweighs the structure effect when considering the between-year changes in distributions of per capita calorie intake or calorie intake coming from protein or fat. The effects of changes in the composition of the Vietnamese population thus overcome the effects of changes in preferences of the same population. This finding is particularly striking in the case of calorie intake from fat where structure effects are never distinguishable from zero. In the case of carbohydrates, this finding is reversed, with the exception of the 90% quantile.
4. Food expenditure and household size appear to be the main contributors to the composition effect, regardless of the considered decomposition. The positive effects of these two variables explain well most of the between-year shifts observed in the calorie intake distributions. Urbanization and level of education contribute negatively to the composition effect, with the noticeable exception of fat where the effect of urbanization is positive. In all cases, the effects of the latter two variables are negligible compared to those of food expenditure and household size.
5. Dependence effects and two-way interaction effects appear to be negligible or insignificant.

Food expenditure and household size appear as the main drivers of the composition effect, thus corroborating the results of many previous studies (Mayen et al., 2014). While urbanization is often advocated as an important driver in dietary pattern changes (Popkin, 2002), its between-year changes do not seem to matter in explaining the composition effect in Vietnam.

As noted above, the structure effect captures between-year behavioral changes in the population under study. Specifically, this effect captures between-year changes in consumer preferences and habits. Thus, for example, the impact of mass media or advertising can change consumer choices from traditional choices linked to the country's food culture towards choices more oriented to Western culture. For the same level of education or income, a consumer would then buy different amounts of food items at different points of time, faced to a different mass media or advertising exposure. The preceding analysis shows the importance of taking this effect into account when studying nutrition transition in a country such as Vietnam, a country with a strong food culture. It is thus remarkable to see that between-year behavioral changes have practically no impact on the observed changes in fat and protein consumption, whereas they have an important effect in the case of carbohydrates. In the latter case, the structural effect may even offset the effect of between-year changes in the composition of the population, effect on which most published papers on the drivers of nutrition transition focus (Mayen et al., 2014). Nevertheless, we are then faced with a limitation of the approach proposed by Rothe (2015) which does not make it possible to decompose the structure effect and to analyze its main drivers.

Acknowledgements

The authors would like to thank the Editor of the review and two referees for their thoughtful comments and suggestions on the earlier draft of this paper. Comments by participants at the Vietnamese Economists Annual Meeting, Ho Chi Minh city, August 2017, 10th Asia Pacific Conference on Clinical Nutrition (APCCN 2017), Adelaide, November 2017, and International Center for Tropical Agriculture (CIAT) – Regional Office for Asia, November 2017, are fully acknowledged. We would like to thank Brian McCaig (Wilfrid Laurier University) and Binh T. Nguyen (Asian Development Bank, Philippines) for sharing with us data and program files. We are grateful to Thibault Laurent for technical assistance in R. This research was funded by the INRA-CIRAD GloFoodS meta-program (TAASE project), the Vietnam International Education Department Fellowships (911 Project), the CGIAR Program on Agriculture for Nutrition and Health (CRP-A4NH) through the Flagship Project on "Food Systems for Healthy Diets", and the Graham Institute of Sustainability through the project "Leveraging existing data and insights into the policy process to accelerate progress toward achieving sustainable diets in the Global South".

Figure 6: **Total differences, composition and structure effects**

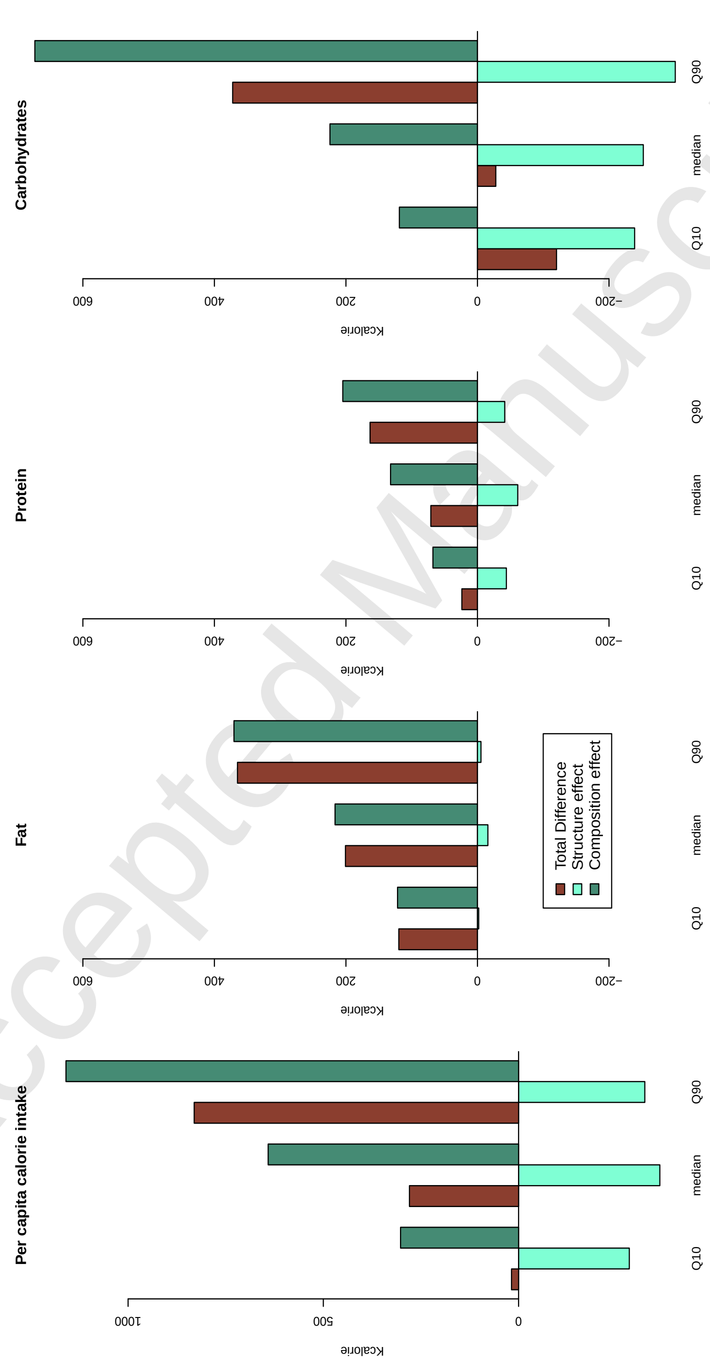
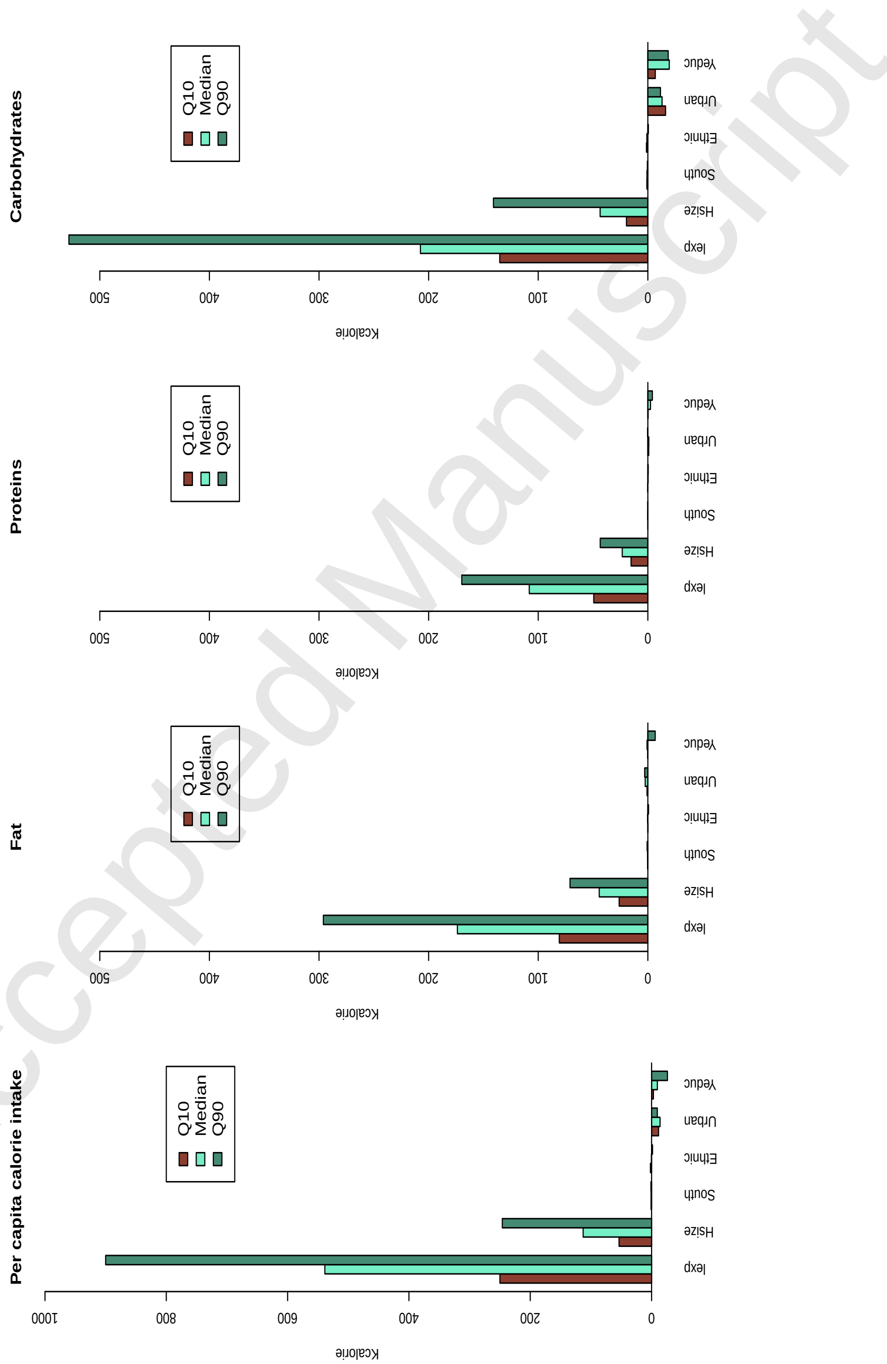


Figure 7: **Direct contribution to the composition effects**



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

Barsky, R., J. Bound, K. C. Kerwin, and J. P. Lupton (2002). Accounting for the black-white wealth gap: A nonparametric approach. *Journal of the American Statistical Association 97*, 663–673.

Benjamin, D., L. Brandt, and B. McCaig (2017). Growth with equity: income inequality in Vietnam, 2002-14. *Journal of Income Inequality 8*, 436–455.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources 15*, 25–46.

Bouis, H. E. and L. J. Haddad (1992). Are estimates of calorie-income elasticities too high? A recalibration of the plausible range. *Journal of Development Economics 39*, 333–364.

Browning, M., T. F. Crossley, and J. Winter (2014). The measurement of household consumption expenditures. *Annual Review of Economics 6*(1), 475–501.

Brzozowski, M., T. F. Crossley, and J. K. Winter (2007). A comparison of recall and diary food expenditure data. *Food Policy 72*, 53–61.

Burggraf, C., R. Teuber, S. Brosig, and T. Glauben (2015). Economic growth and the demand for dietary quality: Evidence from russia during transition. *Economics & Human Biology 19*, 184–203.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Conover, W. J. (1971). *Practical Nonparametric Statistics*. New-York: John Wiley & Sons.

Deaton, A. and M. Grosh (2000). Consumption. In M. Grosh and P. Glewwe (Eds.), *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience*. Washington, DC: World Bank.

Dien, L. N., N. M. Thang, and M. E. Bentley (2004). Food consumption patterns in the economic transition in Vietnam. *Asia Pacific Journal of Clinical Nutrition 13*, 40–47.

DiNardo, J., N. Fortin, and T. Lemieux (1996). Labor market institutions and the distributions of wage, 1973-1992: a semiparametric approach. *Econometrica 81*, 1001–1044.

Etile, F. (2014). Education policies and health inequalities: Evidence from changes in the distribution of Body Mass Index in France, 1981-2003. *Economics & Human Biology 13*, 46–65.

Firpo, S. P., N. M. Fortin, and T. Lemieux (2007). Decomposing wage distributions using recentered influence function regressions. Mimeo, Vancouver: University of British Columbia.

Firpo, S. P., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*, 953–973.

Firpo, S. P., N. M. Fortin, and T. Lemieux (2018). Decomposing Wage Distributions Using Recentered Influence Function Regressions. *Econometrics 6*, doi:10.3390/econometrics6020028.

Foresi, S. and F. Peracchi (1995). The conditional distribution of excess returns: an empirical analysis. *Journal of the American Statistical Association 90*, 451–466.

Fortin, N. M., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, pp. 1–102. Amsterdam: North Holland.

Ha, D. T. P., E. J. M. Feskens, P. Deurenberg, B. M. Le, C. K. Nguyen, and F. J. Kok (2011). Nationwide shifts in the double burden of overweight and underweight in Vietnamese adults in 2000 and 2005: two national nutrition surveys. *BMC Public Health 62*(11).

IFPRI (2017). *2017 Global Food Policy Report.* Washington, DC: International Food Policy Research Institute.

Jiryaie, F., N. Withanage, B. Wu, and de Leon A. R. (2016). Gaussian copula distributions for mixed data, with application in discrimination. *Journal of Statistical Computation & Simulation 86*(9), 1643–1659.

Leibbrandt, M., J. A. Levinsohn, and J. McCrary (2010). Incomes in South Africa after the fall of Apartheid. *Journal of Globalization & Development 1*(1), Article 2.

Machado, J. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics 20*(4), 445–465.

Magrini, E., P. Montalbano, and L. A. Winters (2017). Vulnerability from trade in Vietnam. Working paper 20-2017, Department of Economics, University of Sussex.

Mayen, A.-L., P. Marques-Vidal, F. Paccaud, P. Bovet, and S. Stringhini (2014). Socioeconomic determinants of dietary patterns in low- and middle-income countries: a systematic review. *American Journal of Clinical Nutrition 100*, 1520–1531.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics 12*(4), 577–590.

Muth, M. K., S. A. Karns, S. J. Nielsen, J. C. Buzby, and H. F. Wells (2011). Consumer-level food loss estimates and their use in the ERS loss-adjusted food availablity data. *USDA-ESR, Technical Bulletin Number 1927.*

Nakata, H., Y. Sawada, and M. Tanaka (2009). Asking retrospective questions in household surveys: evidence from vietnam. Discussion paper series 10-e-008, Tokyo: RIETI.

Nguyen, B. T., J. W. Albrecht, S. B. Vroman, and M. D. Westbrook (2007). A quantile regression decomposition of urban–rural inequality in Vietnam. *Journal of Development Economics 83*(2), 466–490.

Nguyen, H.-T.-M., T. Kompas, T. Breusch, and M. B. Ward (2017). Language, mixed communes, and infrastructure: Sources of inequality and ethnic minorities in vietnam. *World Development 96*, 145–162.

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.

Nie, P., A. A. Leon, M. E. D. Sánchez, and A. Sousa-Poza (2018). The rise in obesity in Cuba from 2001 to 2010: An analysis of National Survey on Risk Factors and Chronic Diseases data. *Economics & Human Biology 28*, 1–13.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*, 693–709.

Popkin, B. (1993). Nutritional patterns and transitions. *Population and Development Review 19*(1), 138–157.

Popkin, B. (1994). The nutrition transition in low-income countries: an emerging crisis. *Nutrition Review 52*(9), 285–298.

Popkin, B. (2002). The shift in stages of the nutrition transition in the developing world differs from past experience! *Public Health Nutrition 5*(1A), 205–214.

Popkin, B. (2006). Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. *American Journal of Clinical Nutrition 84*(2), 289–298.

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

Rothe, C. (2015). Decomposing the composition effect: the role of covariates in determining between-group differences in economic outcomes. *Journal of Business & Economic Statistics 33*(3), 323–337.

Sakellariou, C. and Z. Fang (2014). The Vietnam reforms, change in wage inequality and the role of the minimum wage. *Economics of Transition 22*(2), 313–340.

Silverman, B. (1986). Density estimation for statistics and data analysis. *Chapman & Hall*.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris 8*, 229–231.

Solow, R. (1957). Technical change and the agrregate production function. *Review of Economics & Statistics 39*, 312–320.

Thang, N. M. 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*, 145–153.

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*, DOI:10.1177/0962280218770223.

Trinh Thi, H., M. Simioni, and C. Thomas-Agnan (2018). Assessing the nonlinearity of the calorie-income relationship: an estimation strategy - With new insights on nutritional transition in Vietnam. *World Development 110*, 192–204.

Trivedi, P. K. and D. M. Zimmer (2007). Copula modeling: an introduction for practitioners. *Foundations & Trends in Econometrics 1*(1), 1–111.

Weiß, G. (2011). Copula parameter estimation by maximum likelihood and minimum-distance estimators: a simulation study. *Computational Statistics 26*, 31–54.

Appendices

A. Decomposing the composition effect

This appendix introduces through an example the methodology subsequently used, and draws heavily on Rothe (2015).

In the remainder of this article, we will focus on the evolution of certain characteristics of the distribution of the quantities of macronutrients consumed in Vietnam: average values and quantiles, between 2004 and 2014. Let us concentrate, below, on the number of calories obtained from the consumption of carbohydrates per day and per individual. The same reasoning will apply to the number of calories obtained from the consumption of protein or fat. For any household i in year 2004 and any household h in year 2014, we observe an outcome variable: the per capita and per day amount of calories obtained from the consumption of carbohydrates, denoted by Y_i^{2004} and Y_h^{2014}, respectively. These observations are the realizations of two random variables, denoted by Y^{2004} and Y^{2014}, whose marginal cumulative distribution functions, or CDFs, are F_Y^{2004} and F_Y^{2014}, respectively. Our object of interest is a distribution feature, denoted by $\nu(F)$, where $\nu(.)$ is a function from the space of all one-dimensional distribution functions to the real line. The main features we are interested in include the mean, i.e. $\nu : F \to \int y dF(y)$, and the α–quantiles, i.e. $\nu : F \to F^{-1}(\alpha) = \inf\{t : F(t) \geq \alpha\}$ for a given value of $\alpha \in [0, 1]$.

Suppose, for ease of presentation, that we have observed two covariates for each individual in the sample of a given year: for example, food expenditures and location in either urban or rural areas. Of course, the presentation given below can be easily generalized to more than two covariates. We denote the vectors of the two covariates by $X^{2004} = (X_1^{2004}, X_2^{2004})$ and $X^{2014} = (X_1^{2014}, X_2^{2014})$, and their joint CDFs by F_X^{2004} and F_X^{2014}, respectively. The decomposition method aims at understanding how the observed difference between the distribution feature $\nu(F_Y^{2014})$ and $\nu(F_Y^{2004})$, i.e.

$$\Delta_Y^{\nu} = \nu(F_Y^{2014}) - \nu(F_Y^{2004}) \tag{A-1}$$

is related to differences between the distributions F_X^{2004} and F_X^{2014}. For this, we can define the counterfactual outcome distribution $F_Y^{2004|2014}$ that combines the conditional distribution in year 2004 with the distribution of covariates in year 2014, as

$$F_Y^{2004|2014}(y) = \int F_{Y|X}^{2004}(y, x)\, dF_X^{2014}(x) \tag{A-2}$$

where $F_{Y|X}^{2004}(y, x)$ denotes the conditional distribution of outcome given values of the covariates in year 2004. In our example, we can interpret $F_Y^{2004|2014}(y)$ as the distribution of per day and per capita carbohydrates consumption after a counterfactual experiment in which the joint distribution of the two covariates is changed from year

2004 to year 2014, but the conditional distribution of per day and per capita carbohydrates consumption given these characteristics remains that of 2004. One can then decompose the observed between-year difference Δ_Y^ν into

$$\begin{aligned}\Delta_Y^\nu &= \left(\nu\left(F_Y^{2014}\right) - \nu\left(F_Y^{2004|2014}\right)\right) + \left(\nu\left(F_Y^{2004|2014}\right) - \nu\left(F_Y^{2004}\right)\right) \\ &= \Delta_S^\nu + \Delta_X^\nu \end{aligned} \tag{A-3}$$

where Δ_S^ν is a *structure effect*, solely due to differences in the conditional distribution of the outcome given values of covariates between the two years, and Δ_X^ν is a *composition effect*, solely due to differences in the distribution of the covariates between the two years.

The different elements of the decomposition (A-3) can be easily estimated using nonparametric estimates of CDFs. One such strategy, focusing on densities instead of CDFs, is applied in DiNardo et al. (1996) or Leibbrandt et al. (2010). But the application of such a strategy soon encounters the problem of the curse of dimensionality. For a fixed sample size, the precision of the nonparametric estimators deteriorates very rapidly when the number of covariates increases, even if these estimators are free from any specification error (Silverman, 1986). In addition, it is also interesting to break down the composition effect for the different covariates. This can be easily done using the Oaxaca (1973) and Blinder (1973) approach when focusing on the between-year difference of average outcomes. But the possibility of disentangling the impact of each of the covariates in the composition effect rests on the very restrictive assumption that the data are generated from a linear model. As pointed out by Rothe (2015), in the general case, it is difficult to express the composition effect as a sum of terms which depend on the marginal distribution of a single covariate only. Instead, an explicit decomposition of the composition effect in terms of the respective marginal covariate distributions typically contains "interaction terms" resulting from the interplay of two or more covariates, and also "dependence terms" resulting from between-year difference in the dependence pattern among the covariates.

Rothe (2015) proposes to use results from copula theory in order to disentangle the covariates' marginal distributions from the dependence structure among them. Indeed, the CDF of X^t can always be written as

$$F_X^t(x) = C^t(F_{X_1}^t(x_1), F_{X_2}^t(x_2)) \quad \text{for} \quad t \in \{2004, 2014\} \tag{A-4}$$

following Sklar's Theorem (Sklar, 1959). $C^t(.)$ is a copula function, i.e., a bivariate CDF with standard uniformly distributed marginals, and $F_{X_j}^t(.)$ is the marginal distribution of the jth component of X^t (Trivedi and Zimmer, 2007). The copula describes the joint distribution of individuals' ranks in the two components of X^t. The copula accounts for the dependence between the covariates in a way that is separate from and independent of their marginal specifications. This result holds for continuous covariates. When some of them are discrete, some identifiability issues may arise, that can be solved by making parametric restrictions on the functional form of the copula.

In this context, the decomposition given by Eq. (A-3) can then be generalized as follows. Let $\mathbf{k}$ denote an element of the 2-dimensional product set $\{2004, 2014\}^2$, i.e. $\mathbf{k} = (k_1, k_2)$ where k_1 (resp. k_2) is equal to either 2004 or 2014. We can define the distribution of the outcome in a counterfactual setting where the conditional distribution is as in year t, the covariate distribution has the copula function of year s, and the marginal distribution of the lth covariate is equal to that in group $\mathbf{k}$ by

$$F_Y^{t|s,\mathbf{k}} = \int F_{Y|X}^t(y, x) dF_X^{s,\mathbf{k}}(x) \tag{A-5}$$

with

$$F_X^{s,\mathbf{k}}(x) = C^s(F_{X_1}^{k_1}(x_1), F_{X_2}^{k_2}(x_2)). \tag{A-6}$$

For instance, the counterfactual distribution $F_Y^{2004|2014}$ in Eq. (A-3) can be written as $F_Y^{2004|2014,\mathbf{1}}$ where $\mathbf{1} = (2014, 2014)$. In other words, the computation of the counterfactual distribution $F_Y^{2004|2014}$ uses the conditional distribution of the outcome given the covariates in year 2004, the dependence structure of year 2014 , and the marginal distributions of the covariates in year 2014. Similarly, we can get $F_Y^{2004} = F_Y^{2004|2004,\mathbf{0}}$ where $\mathbf{0} = (2004, 2004)$.

Now we can write the composition effect Δ_X^ν as

$$\begin{aligned}
\Delta_X^\nu &= \nu\left(F_Y^{2004|2014}\right) - \nu\left(F_Y^{2004}\right) \\
&= \nu\left(F_Y^{2004|2014,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{0}}\right) \\
&= \left(\nu\left(F_Y^{2004|2014,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{1}}\right)\right) + \left(\nu\left(F_Y^{2004|2004,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{0}}\right)\right) \\
&= \Delta_D^\nu + \beta^\nu(\mathbf{1})
\end{aligned} \tag{A-7}$$

The first term of the decomposition in Eq. (A-7), or

$$\Delta_D^\nu = \nu\left(F_Y^{2004|2014,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{1}}\right),$$

captures the contribution of the between-year difference of the covariates' copula functions. Δ_D^ν is thus a *dependence effect*. The second term, or

$$\beta^\nu(\mathbf{1}) = \nu\left(F_Y^{2004|2004,\mathbf{1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{0}}\right)$$

measures the joint contribution of between-year differences in the marginal covariate distributions.

Let now $\mathbf{e^1} = (2014, 2004)$ and $\mathbf{e^2} = (2004, 2014)$. $\beta^\nu(\mathbf{1})$ can in turn be decomposed as

$$\beta^\nu(\mathbf{1}) = \left(\beta^\nu(\mathbf{1}) - \beta^\nu(\mathbf{e^1}) - \beta^\nu(\mathbf{e^2})\right) + \beta^\nu(\mathbf{e^1}) + \beta^\nu(\mathbf{e^2}) \tag{A-8}$$

with

$$\begin{aligned}
\beta^\nu(\mathbf{e^1}) &= \nu\left(F_Y^{2004|2004,\mathbf{e^1}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{0}}\right) \\
&\text{and} \\
\beta^\nu(\mathbf{e^2}) &= \nu\left(F_Y^{2004|2004,\mathbf{e^2}}\right) - \nu\left(F_Y^{2004|2004,\mathbf{0}}\right)
\end{aligned}$$

In other words, $\beta^\nu(\mathbf{e^1})$ and $\beta^\nu(\mathbf{e^2})$ measure the respective direct contributions of the first and second covariate. Let $\Delta_{Int}^\nu(\mathbf{1}) \equiv \beta^\nu(\mathbf{1}) - \beta^\nu(\mathbf{e^1}) - \beta^\nu(\mathbf{e^2})$. $\Delta_{Int}^\nu(\mathbf{1})$ can then interpreted as a "pure" *interaction effect*.

To sum up, the composition effect can be written as

$$\Delta_X^\nu = \beta^\nu(\mathbf{e^1}) + \beta^\nu(\mathbf{e^2}) + \Delta_{Int}^\nu(\mathbf{1}) + \Delta_D^\nu, \tag{A-9}$$

i.e, as the sum of the respective contributions of each covariate, a term measuring the pure effect of their interaction, and a term measuring the contribution due to the between-year variation of the dependence between covariates. This decomposition can easily be generalized in the case of more than two covariates and focus either on individual effect of each of them and the pure effect of their interaction as shown above, or on the effect of groups of variables and of the interaction among these groups.

B. Practical implementation

We consider the general case where the vector of the covariates has d elements, and we have two iid samples $\left\{(Y_i^t, X_i^t)\right\}_{i=1}^{n_t}$ of size n_t from the distribution of (Y^t, X^t) for $t = 2004, 2014$. The practical implementation of the decomposition procedure presented in Appendix A requires the estimation of various functions or parameters.

Univariate CDFs.. Univariate CDFs are estimated nonparametrically using the classical empirical CDF, i.e.

$$\widehat{F}_{X_j}^t(x_j) = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbb{I}(X_{ji}^t \leq x_j) \tag{A-10}$$

where $\mathbb{I}(A) = 1$ if A verified, and $= 0$ if not.

Conditional CDF of $Y^t|X^t$.. The conditional CDF of $Y^t|X^t$ is a multivariate function whose dimension depends on the number of covariates. A nonparametric estimate of this function can be quite imprecise when the number of covariates is large, due to the so-called curse of dimensionality. Flexible parametric specifications can be used to overcome this drawback of nonparametric estimators (see Fortin et al. (2011)). As in Rothe (2015), conditional CDFs $F_{Y|X}^t$ are estimated using the distributional regression approach of Foresi and Peracchi (1995). The distributional regression model assumes that

$$F_{Y|X}^t(y, x) \equiv \Phi(x'\delta^t(y)), \tag{A-11}$$

where $\Phi(.)$ is the standard normal CDF. The finite-dimensional parameter $\delta^t(y)$ is estimated by the maximum likelihood estimate $\widehat{\delta}^t(y)$ in a Probit model that relates the indicator variable $\mathbb{I}(Y^t \leq y)$ to the covariates X^t.

Copula choice.. The last function necessary for the implementation of the decomposition procedure of Rothe (2015) is the copula function. Let us take a copula contained in a parametric class indexed by a k-dimensional parameter θ. A strategy for estimating the parameters characterizing the copula then consists in choosing the minimum distance estimator defined as (Weiß, 2011)

$$\widehat{\theta}^t = \arg\min_{\theta} \sum_{i=1}^{n_t} \left(\widehat{F}_X^t(X_{1i}^t, \ldots, X_{di}^t) - C_\theta(\widehat{F}_{X_1}^t(X_{1i}^t), \ldots, \widehat{F}_{X_d}^t(X_{di}^t))\right) \tag{A-12}$$

Different parametric copula functions can be used (Trivedi and Zimmer, 2007). But, here too, we must keep in mind when choosing this function to select a function that is sufficiently flexible for generating all possible types of dependence. Moreover, we are confronted here with the fact that our variables are a mixture of continuous and discrete variables. To address these issues, we choose the Gaussian copula model

$$C_\Sigma(u) = \Phi_\Sigma^d\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right) \tag{A-13}$$

where $\Phi_\Sigma^d(.)$ denotes the CDF of a d-variate standard normal distribution with correlation matrix Σ, and $\Phi^{-1}(.)$ is the inverse function of the standard normal distribution function $\Phi(.)$. The parameters $\theta \equiv \Sigma$ determine the dependence pattern among the covariates.

The flexibility and the analytical tractability of Gaussian copulas make them a handy tool in applications as emphasized by Jiryaie et al. (2016). First, This specification has a computational advantage, namely, that only the (a, b) element of Σ affects the pairwise dependence between the covariates X_a^t and X_b^b. So minimum distance estimation (A-12) can be performed for each pair of covariates, not by taking all the covariates together simultaneously.

Second, as noted above, the copula function describes the joint distribution of individuals' ranks in the various components of X^t, and, here, the dependence between two components can be measured using a correlation coefficient as we are working with Gaussian copula. Indeed, in the bivariate case, we get

$$C_{\Sigma_{a,b}}(F_{X_a}(X_{ai}), F_{X_b}(X_{bi})) = \Phi^2_{\Sigma_{a,b}}\left(\Phi^{-1}(F_{X_a}(X_{ai})), \Phi^{-1}(F_{X_b}(X_{bi}))\right) \quad \text{(A-14)}$$

where $\Phi^2_{\Sigma_{a,b}}(.)$ denotes the CDF of the bivariate normal distribution with covariance matrix $\Sigma_{a,b}$, and $\Phi^{-1}(F_{X_a}(X_{ai}))$ (resp.$\Phi^{-1}(F_{X_b}(X_{bi}))$ can be interpreted as the quantile of the univariate marginal distribution associated to the observation X_{ai} (resp. X_{bi}).

Third, Gaussian copulas make it possible to have both continuous and discrete variables in the vector of covariates. We only have to assume that each discrete covariate X_j^t can be represented as $X_j^t = t_j(\widetilde{X}_j^t)$ for some continuously distributed latent variable $\widetilde{X}_j^t$ and a function $t_j(.)$ that is weakly increasing in its argument. For instance, if X_j^t is a binary, we could have $X_j^t = \mathbb{I}(\widetilde{X}_j^t > c_j)$ for some constant c_j. Details on the computation of the joint distribution of a vector of continuous and discrete variables using Gaussian copula can be found in Jiryaie et al. (2016).

Counterfactual distributions.. After estimating the copula and the marginal distributions for each time period, we can construct the joint c.d.f. of the explanatory variables given by (A-6) in any counterfactual experiment where the copula is as in time s and the marginals as in time k_1 and k_2. Given this joint c.d.f, using equation (A-5) and the conditional c.d.f $F^t_{Y|X}(y, x)$ at time t estimated by equation (A-11), we can construct an estimation of any counterfactual distribution of the outcome.

C. Preliminary results

To estimate the various elements of the decomposition of the composition effect, we proceed as described in Appendix B. Copulas are thereby modeled by a Gaussian copula and the joint CDF of each pair of covariates estimated using marginal empirical CDF estimators and copula estimators. Table A-1 reports the estimated values of the copula parameters from the 2004 and 2014 VHLSS waves. Estimated copula parameters show positive and significant association between food expenditures and location in an urban area as well as food expenditures and household size. The first association decreased between 2004 and 2014 while the second remained fairly stable. The association between location in an urban area and ethnicity is negative and significant whatever the considered waves, as expected, and increases over the period. The association between location in an urban area and years of education is positive but becomes significant only in 2014. A stable positive and significant association is also shown for location in an urban area and living in South Vietnam. We also notice a negative association between household size and ethnicity in 2004, which disappears completely in 2014. As recently pointed out by Benjamin et al. (2017), the share of minorities in the rural population has risen over time, from below 15% in 2002 to over 18% in 2014. This is a consequence of a higher fertility among minorities, combined with rising urbanization among the Kinh. Finally, the association between the number of years of education and living in South Vietnam is negative and significant but decreasing between 2004 and 2014.

Table A-1: **Estimated copula parameters**

	Urban		Hsize		Ethnic		Yeduc		South	
	2004	2014	2004	2014	2004	2014	2004	2014	2004	2014
lExp	0.502	0.410	0.542	0.582	-0.167	0.206	0.160	0.239	0.364	0.186
	(0.124)	(0.149)	(0.107)	(0.138)	(0.159)	(0.119)	(0.253)	(0.345)	(0.272)	(0.289)
Urban			-0.034	0.007	-0.288	-0.624	0.113	0.278	0.377	0.349
			(0.269)	(0.204)	(0.084)	(0.093)	(0.092)	(0.097)	(0.066)	(0.077)
Hsize					-0.544	-0.033	-0.009	-0.005	0.006	-0.081
					(0.245)	(0.199)	(0.118)	(0.129)	(0.129)	(0.108)
Ethnic							-0.191	-0.131	-0.247	-0.384
							(0.077)	(0.086)	(0.289)	(0.360)
Yeduc									-0.529	-0.320
									(0.127)	(0.122)

Note: Bootstrapped standard errors, based on 300 replications, are in parenthesis.

Conditional CDFs $F^t_{Y|X}$ are modeled by a distributional regression model with a Gaussian link function. We do not report the results as they are not very helpful in the discussions that follow. Nevertheless, they are available from the authors.