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

Huong TRINH THI
Toulouse School of Economics, INRA, University of Toulouse Capitole
21 Allée de Brienne, 31015 Toulouse Cedex 6, France
trinhthihuong@tmu.edu.vn
Michel SIMIONI
INRA, UMR-1110 MOISA
2, place Pierre Viala - Bât. 26, 34 060 Montpellier Cedex 2, France
michel.simioni@inra.fr
and
Christine THOMAS-AGNAN
Toulouse School of Economics, University of Toulouse Capitole
21 Allée de Brienne, 31015 Toulouse Cedex 6, France
christine.thomas@tse-fr.eu

April 26, 2018

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


**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:** Macronutrient consumption, Nutritional transition, Decomposition method, Copula, Vietnam.

# 1 Introduction

Since the launch of economic reforms in 1986, Vietnam has recorded impressive achievements in growth performance and, at the same time, has also experienced a nutrition transition like many other middle-income countries in South East Asia. Dietary diversity from 2005 to 2015 in South-East Asia and China has considerably increased: 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 (IFPRI, 2017). On one hand, this nutrition transition to energy-dense, poor quality diets has led to obesity and diet-related chronic diseases. Using two nationally representative surveys, Ha et al. (2011) show that the nationwide prevalence of overweight (body mass index $\geq 25\text{kg/m}^2$) and obesity (body mass index $\geq 30\text{kg/m}^2$) 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 $23\text{kg/m}^2$ 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%. On the other hand, Ha et al. (2011) point out that the underweight prevalence (body mass index $< 18.5\text{kg/m}^2$) 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%. Ha et al. (2011) also analyze the possible sources of this evolution and note that women were more likely to be both underweight and overweight compared to men in both 2000 and 2005. 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.

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 ( Thang and Popkin (2004), Burggraf et al. (2015)). 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 thus leads to a better feeding of human beings. But the other side of the coin is the link between improved diets and noncommunicable disease 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 malnutrition or over-food consumption. All this requires first of all knowledge of the drivers of these evolutions.

In this paper we document shifts in consumption of macronutrients in Vietnam over the period 2004 to 2014. Thanks to data from Vietnamese households living standard survey, we can calculate total calorie intakes of Vietnamese households, convert them into an adult equivalent, or per capita, calorie intakes (thus allowing comparison between households), and their decomposition into the three macronutrients : proteins, fat and carbohydrates (Trinh Thi et al., 2018). This survey also contains detailed information on the socio-demographic characteristics of Vietnamese households. Each wave of this survey is, moreover, representative of the Vietnamese population. This survey can therefore be used for a comparison of the nutritional status of the Vietnamese population between two waves.

We propose the use of decomposition methods to assess 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). Fortin et al. (2011) provide a comprehensive overview of decomposition methods that have been developed since then. This method is recently wide used in the health sector, among them: Nie et al. (2018), (Anderson, 2018). The common 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 to the considered outcome. Applications to Vietnamese economy 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, very recently, Benjamin et al. (2017) on income inequality. 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 Oaxaca-Blinder decomposition method has been refined in a large number of methodological papers and extended to the cases of distributional parameters besides the mean over the last four decades. Among all these methodological developments, we use the decomposition procedure recently proposed by Rothe (2015) which can be applied to mean, quantiles, or other parameters characterizing the distribution of the considered outcome (in our application, per capita calorie intake or calorie intakes coming from the three macronutrients). 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, and several *two way* and *higher order interaction effects* due to the interplay between two or more covariates, a third effect, or *dependence effect*, accounting for the between-group difference in the dependence pattern among the covariates. 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 explanatory) 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 Vietnamese population is more urban in 2014 than in 2004. The (only) *interaction* effect measures the additional contribution of the fact that Vietnamese population at the same time spends more for food and is more urban in 2014. Finally, the *dependence* effect accounts for between-year difference 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.

The remainder of the paper is structured as follows. Section 2 describes the decomposition method based on copulas and its practical implementation. Section 3 gives a description of the data we use in this study. Results are presented and commented in section 4. Section 5 concludes.

# 2 Decomposition method

## 2.1 Decomposing the decomposition effect

This section 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 $\alpha$–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.

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

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

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

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 $\Delta_Y^\nu$ into

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

where $\Delta_S^\nu$ is a *structure effect*, solely due to differences in the conditional distribution of the outcome given values of covariates between the two years, and $\Delta_X^\nu$ is a *composition effect*, solely due to differences in the distribution of the covariates between the two years.

The different elements of the decomposition (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

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

following Sklar's Theorem (Sklar, 1959). $C^t(.)$ is a copula function, i.e., a bivariate CDF with standard uniformly distributed marginals, and $F^t_{X_j}(.)$ is the marginal distribution of the $j$th 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. (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 $l$th covariate is equal to that in group $\mathbf{k}$ by

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

with

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

For instance, the counterfactual distribution $F_Y^{2004|2014}$ in Eq. (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 $\Delta_X^\nu$ as

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

The first term of the decomposition in Eq. (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. $\Delta_D^\nu$ 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

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

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 $\Delta_M^\nu(\mathbf{1}) \equiv \beta^\nu(\mathbf{1}) - \beta^\nu(\mathbf{e^1}) - \beta^\nu(\mathbf{e^2})$. $\Delta_M^\nu(\mathbf{1})$ can then interpreted as a "pure" *interaction effect*.

To sum up, the composition effect can be written as

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

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.

## 2.2 Practical implementation

Consider now the general case where the vector of the covariates has $d$ elements, and suppose 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 above requires the estimation of various functions or parameters.

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

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

**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

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

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 $\theta$. A strategy for estimating the parameters characterizing the copula then consists in choosing the minimum distance estimator defined as (Weiß, 2011)

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

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

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

where $\Phi_{\Sigma}^{d}(.)$ denotes the CDF of a $d$-variate standard normal distribution with correlation matrix $\Sigma$, 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 $\Sigma$ affects the pairwise dependence between the covariates $X_a^t$ and $X_b^b$. So minimum distance estimation (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

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

where $\Phi_{\Sigma_{a,b}}^{2}(.)$ 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 (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 (5) and the conditional c.d.f $F^t_{Y|X}(y, x)$ at time $t$ estimated by equation (11), we can construct an estimation of any counterfactual distribution of the outcome.

# 3 Data

This study relies on the survey "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. In this study, we use the two waves of VHLSS conducted in 2004 and 2014.

## 3.1 Macronutrient intakes

Average annual or monthly food expenditures and quantities about 56 food items are collected for each household surveyed in each VHLSS wave.[1] The observed kilograms can then be converted into kilocalories using the conversion coefficients given in the Vietnamese Food Composition Table constructed by the Vietnam National Institute

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

of Nutrition in 2007. Table 1 shows the coefficients that have been applied to perform these conversions into calorie intakes and amounts of proteins and fats, expressed as calorie intakes. Calorie intakes from carbohydrates are then obtained by difference. These annual calorie intakes, which are computed at the household level, are then converted into daily intakes and adjusted in the form of per capita calorie intakes to be comparable between households. This adjustment makes use of the household equivalence scale calculation procedure recently proposed by Aguiar and Hurst (2013).[2]

Figure 1 reports the kernel weighted estimates 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 4.

Figure 1: Density of per capita calorie intake

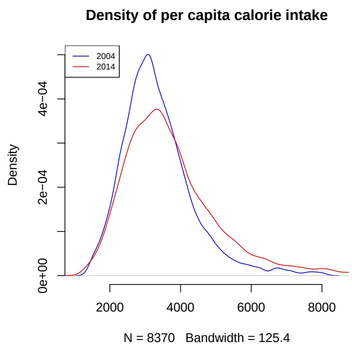
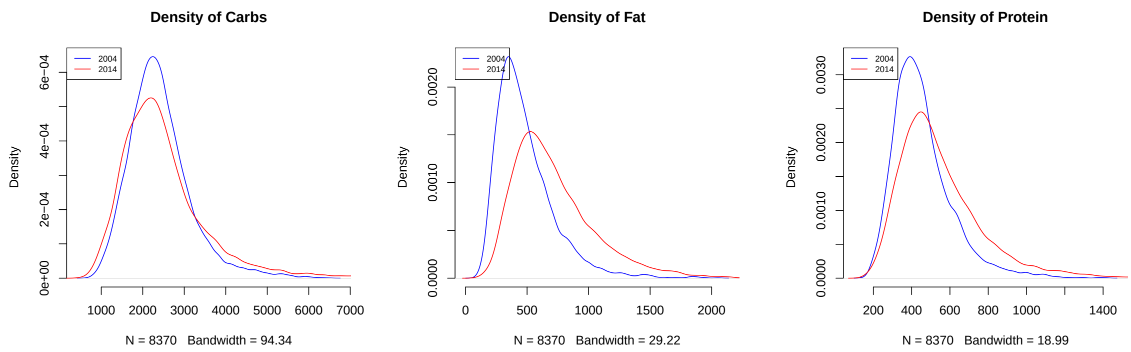



Figure 2 reports the kernel weighted estimates of the densities of per capita calorie intakes of carbohydrates, fat, and proteins, for the two years. Significant changes appear when comparing the estimated densities for fat and proteins, while the estimated densities for carbohydrates appear to be very close. 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 4. All these values

[2]More details are given in Trinh Thi et al. (2018).

Figure 2: Density of per capita calorie intake by macronutrient




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 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.

## 3.2 Sociodemographic variables

Table 3 summarizes the sociodemographic variables we use in this paper, and detailed descriptive statistics on these variables are given in Table 4. 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.

—— Insert Tables 3 and 4 ——

# 4 Results

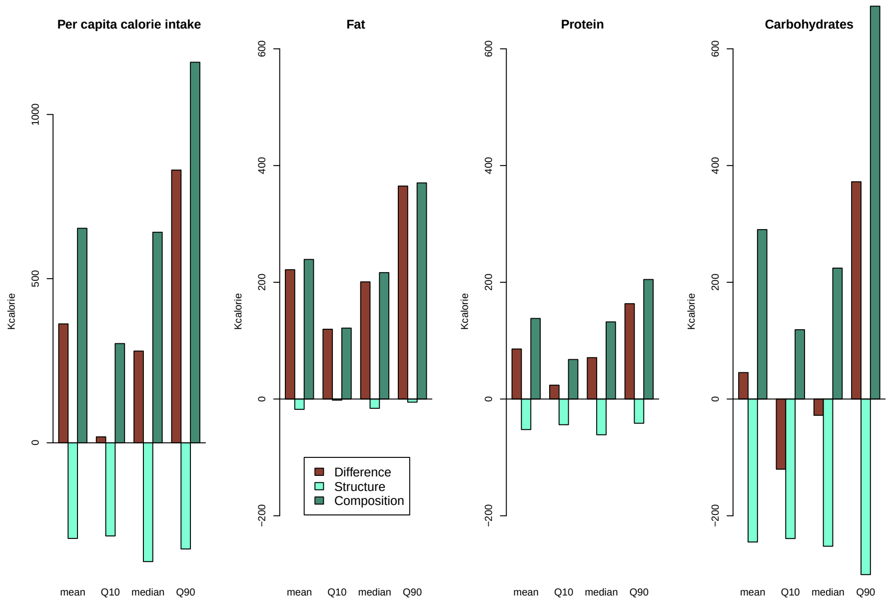
To estimate the various elements of the decomposition of the composition effect, we proceed as described in section 2. Copulas are thereby modeled by a Gaussian copula and the joint CDF of each pair of covariates estimated using marginal empirical CDF estimators and copula estimators. Table 2 reports 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.

—— Insert Table 2 ——

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.

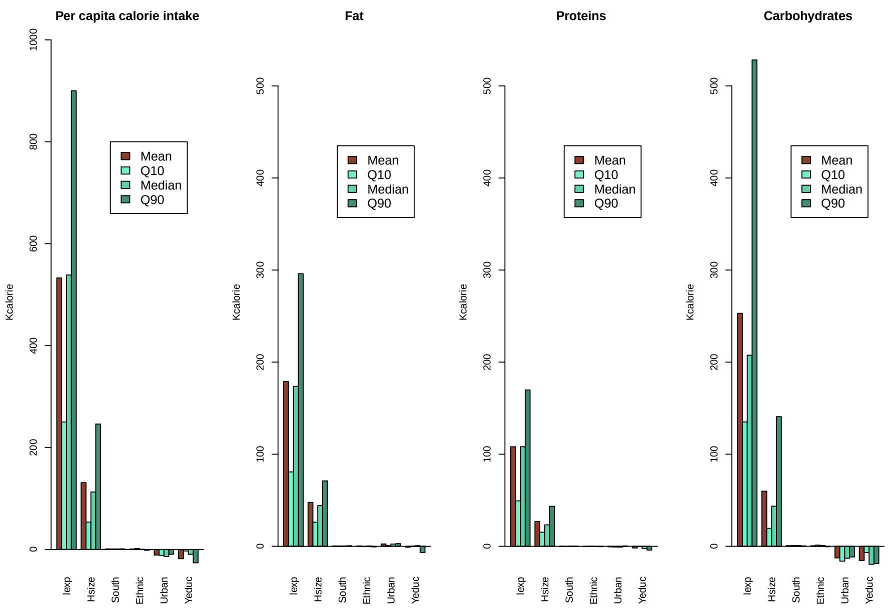
Tables 5, 6, 7, and 8 then present the results of our decomposition of per capita calorie intake and calorie intake coming from the three macronutrients, for two measures of location: mean and median, and for the two quantiles at 10% and 90% allowing to construct a measure of dispersion. Row by row, we report estimates of total change, i.e. $\Delta_Y^\nu$, usual structure and composition effects, i.e. $\Delta_S^\nu$ and $\Delta_X^\nu$. Then the composition effect is in turn decomposed into the dependence effect, i.e. $\Delta_D^\nu$, 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 $\beta^\nu(\mathbf{e}^\mathbf{l})$, and the "two-way" interaction effects, i.e. the $\Delta_M^\nu$. Figures 3 and 4 summarize these same results in the form of barplots.

Figure 3: Total differences, composition and structure effects




Each estimated value in a decomposition is accompanied by the estimated value of its standard error. 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 the use of a 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.

Figure 4: Direct contributions to the composition effects




Here, estimated standard errors are calculated using nonparametric bootstrap with 300 replications.

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 4. 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.

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 means, 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 change in the conditional distributions of per capita calorie intake givenf the sociodemographic characteristics, i.e. in the relationship between per capita calorie intake and these covariates, between the two years 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 total marginal distribution effect resulting from differences in the marginal covariate distributions across the two years.

Consider now the decomposition of the total marginal distribution 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.

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 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 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.

# 5 Conclusion

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 capital 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 us to compute the direct contributions of various socio-demographic variables and the effects of their interactions in these between-year changes. We implement this 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 have obtained can be summarized as follows (see Figures 3 and 4). First, 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. Second, 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. Third, 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. Fourth, 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 compositional 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. Finally, dependence effects and two-way interaction effects appear to be negligible or insignificant.

The decomposition method we use in this paper focuses on the decomposition of the composition effect into its main drivers: the direct effects of covariates or the effects of their interactions. It therefore allows a detailed analysis of one of the two sides of the decomposition, the composition effect, but it says nothing about the structure effect. Our application shows that the latter effect can play an important role. The related issue of deriving a decomposition of the structure effect, that is, dividing between-year 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.

# Acknowledgements

This paper has also benefited from valuable discussions with participants at the 10th Asia Pacific Conference on Clinical Nutrition (APCCN 2017), Adelaide, Nov 2017, Vietnamese Economists Annual Meeting, Ho Chi Minh city, August 2017, the International Center for Tropical Agriculture (CIAT) – Regional Office for Asia, Nov 2017. We would like to thank Dr. Brian McCaig (Wilfrid Laurier University) and Dr. 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.

**Funding** This work was supported by the INRA-CIRAD GloFoodS meta-program (TAASE project) and the Vietnam International Education Department Fellowships (911 Project).

# References

# References

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

Anderson, L. R. (2018). Adolescent mental health and behavioural problems, and intergenerational social mobility: A decomposition of health selection effects. *Social Science & Medicine 197*, 153–160.

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.

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.

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.

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., 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).

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 and 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 and Development 1*(1), Article 2.

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.

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.

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. (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 and 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 and 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 and 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 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. *TSE Working Paper*.

Trivedi, P. K. and D. M. Zimmer (2007). Copula modeling: an introduction for practitioners. *Foundations and 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.

# Appendix: Tables and Figures

Table 1: Conversion table Calories for Vietnam.

| Food | Energy 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 |
| 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 | - | - | - |

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

Table 2: 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.

Table 3: Description of sociodemographic variables

| Variable | Values | Description |
|---|---|---|
| $lExp$ | | Food expenditures per year in US$ (in logarithms) |
| $Hsize$ | | 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 |

Table 4: Descriptive statistics in VHLSS 2004 and 2014

| | Mean | SD | Q10 | Q50 | Q90 |
|---|---|---|---|---|---|
| **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 | | – | – | – |
| **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 | | – | – | – |

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

| | Mean | | Q10 | | Median | | Q90 | |
|---|---|---|---|---|---|---|---|---|
| Total difference | 362.16 | (28.90) | 18.62 | (26.38) | 279.48 | (20.81) | 830.71 | (78.76) |
| Structure effect | -291.21 | (52.53) | -283.63 | (49.12) | -361.79 | (39.69) | -328.38 | (213.09) |
| Composition effect | 653.37 | (44.47) | 302.25 | (46.16) | 641.27 | (36.71) | 1159.09 | (202.78) |
| *Composition effect*: | | | | | | | | |
| Dependence effect | 0.91 | (23.08) | -30.6 | (22.94) | 0.25 | (23.26) | -7.97 | (135.35) |
| Marginal effect | 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, based on 300 replications, are in parenthesis.

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

| | Mean | | Q10 | | Median | | Q90 | |
|---|---|---|---|---|---|---|---|---|
| Total difference | 221.51 | (8.68) | 119.61 | (6.13) | 200.73 | (7.06) | 364.92 | (28.06) |
| Structure effect | -17.63 | (13.93) | -1.92 | (6.8) | -15.85 | (10.51) | -5.33 | (55.00) |
| Composition effect | 239.14 | (12.39) | 121.53 | (8.13) | 216.57 | (9.58) | 370.25 | (55.94) |
| *Composition effect*: | | | | | | | | |
| Dependence effect | -0.77 | (8.01) | 2.34 | (5.94) | -0.67 | (5.01) | 6.33 | (46.76) |
| Marginal effect | 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, based on 300 replications, are in parenthesis.

Table 7: Estimated decomposition of calorie intake from protein

| | Mean | | Q10 | | Median | | Q90 | |
|---|---|---|---|---|---|---|---|---|
| Total difference | 85.74 | (4.54) | 23.76 | (4.37) | 70.93 | (3.97) | 163.34 | (11.17) |
| Structure effect | -52.32 | (9.08) | -43.97 | (7.05) | -61.23 | (7.35) | -41.46 | (31.46) |
| Composition effect | 138.06 | (8.09) | 67.73 | (7.06) | 132.16 | (7.18) | 204.8 | (30.44) |
| *Composition effect*: | | | | | | | | |
| Dependence effect | 2.94 | (6.06) | 2.80 | (4.98) | 2.66 | (4.74) | -5.93 | (25.79) |
| Marginal effect | 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, based on 300 replications, are in parenthesis.

Table 8: Estimated decomposition of calorie intake from carbohydrates

| | Mean | | Q10 | | Median | | Q90 | |
|---|---|---|---|---|---|---|---|---|
| Total difference | 45.36 | (22.04) | -120.24 | (21.14) | -27.88 | (18.25) | 372.23 | (50.91) |
| Structure effect | -244.93 | (36.73) | -238.99 | (30.56) | -252.19 | (24.66) | -300.78 | (134.48) |
| Composition effect | 290.29 | (32.66) | 118.75 | (29.32) | 224.31 | (22.52) | 673.01 | (128.18) |
| *Composition effect*: | | | | | | | | |
| Dependence effect | 0.68 | (17.57) | -6.03 | (15.75) | 5.50 | (10.70) | 18.43 | (74.42) |
| Marginal effect | 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, based on 300 replications, are in parenthesis.