Supermarkets and their impacts on the relationship between food acquisition patterns and socio-economic and demographic characteristics of households: empirical evidence from Vietnam

Thi Huong Trinh, Dharani Dhar Burra, Michel Simioni, Stef de Haan, Tuyen Thi Thanh Huynh, Tung Van Huynh, Andrew D. Jones

Supermarkets and their impacts on the relationship between food acquisition patterns and socio–economic and demographic characteristics of households: Empirical evidence from Vietnam

Thi Huong TRINH*
International Center for Tropical Agriculture (CIAT)-Asia Office, Hanoi, Vietnam
Regional Office for Asia Agricultural Genetics Institute, Pham Van Dong St., Tu Liem, Hanoi, Vietnam
t.huong@cgiar.org
Phone: + 84 (024) 37 57 69 69

Department of Mathematics and Statistics, Thuongmai University, Hanoi, Vietnam

Dharani DHAR BURRA†
International Center for Tropical Agriculture (CIAT)-Asia Office, Hanoi, Vietnam
D.Burra@cgiar.org

Michel SIMIONI
INRA, UMR-1110 MOISA
2, place Pierre Viala - Bât. 26, 34 060 Montpellier Cedex 2, France
michel.simioni@inra.fr

Stef DE HAAN
International Center for Tropical Agriculture (CIAT)-Asia Office, Hanoi, Vietnam
S.DeHaan@CGIAR.ORG

Tuyen Thi Thanh HUYNH
International Center for Tropical Agriculture (CIAT)-Asia Office, Hanoi, Vietnam
T.Huynh@cgiar.org

Tung Van HUYNH
Can Tho City Institute for Socio-Economic Development, Vietnam
hvtung@cids.org.vn

Andrew D JONES
School of Public Health, University of Michigan, USA
jonesand@umich.edu

June 28, 2019

Acknowledgements: This research was funded by the CGIAR Research Program on Agriculture for Nutrition and Health (CRP-A4NH) through its Flagship Program on Food Systems for Healthy Diets through the Flagship Project on "Food Systems for Healthy Diets", 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" and the INRA-CIRAD GloFoodS meta-program (TAASE project).

*Corresponding author
†Corresponding author

Supermarkets and their impacts on the relationship between food acquisition patterns and socio–economic and demographic characteristics of households: Empirical evidence from Vietnam

Abstract Food environments in developing economies are rapidly evolving, alongside fast-paced changes in the socioeconomic and demographic characteristics of populations. These changes are evident in Vietnam with the widespread emergence of supermarkets, and restructuring in traditional markets that are poised to have profound effects on household diets and patterns of food acquisition. This paper examines the relationship between province level supermarket density, quantity and quality indices of food groups acquired by households within those provinces, between 2010 and 2014. An original approach on the basis of open access mixed data sets (administrative data on the number of supermarkets at provincial level as a proxy for supermarket density, and household living standard survey) is proposed and implemented.We find that the differential presence of supermarkets across provinces in Vietnam is associated with the diversity and macronutrient quality of food groups acquired by households. In addition, households with higher per capita expenditure, and those that purchase a larger proportion of food (relative to food obtained from own production), acquire a higher diversity of food groups. Additionally, diversity of food acquired is associated with higher fat and lower carbohydrate shares, and this is independent of the presence of supermarkets. We observe a significant interplay between low household financial capabilities (i.e., low per capita expenditure and low proportion of income spent on food), large household size, ethnic minority status, and the existence of limited number of supermarkets in the food environment. All of these factors are associated with a limited diversity of food groups acquired, as well as higher carbohydrate and lower fat shares. Our findings highlight potential intervention opportunities that can "rewire" local food environments to address the challenge of double burden of malnutrition in the country.

Keyword: Supermarkets, Diet Diversity Score, Macronutrient shares, Poisson regression, Compositional data analysis, Vietnam Household Living Standard Survey.

1 Introduction

There has been an ongoing evolution in food retail in both developed and developing countries for the past century. This evolution is characterized by the transition from local traditional stores to supermarkets, to e-commerce platforms. The rapid spread of supermarkets has transformed food environments in several countries, including Vietnam (Reardon and Timmer, 2007; Reardon et al., 2012; Lu and Reardon, 2018). Furthermore, supermarkets have also influenced the traditional relationships between household socio-economics, population demographics and food acquisition and consumption patterns (Popkin and Gordon-Larsen, 2004; Chandon and Wansink, 2012; Rischke et al., 2015; HLPE, 2017).

The Doi Moi reforms in Vietnam, launched in late 1980's and early 1990's, have enabled Vietnam to achieve impressive economic growth (Barker and Üngör, 2019). This growth has been associated with the introduction and expansion of supermarkets, supported by increased foreign direct investments (FDI) in the food retail sector (Reardon et al., 2012). Local, traditional traders, such as informal markets, however are still predominant (Vo and Smith, 2016; Maruyama and Trung, 2010; Umberger et al., 2018). Therefore, this complex food environment in Vietnam is characterised by the coexistence of an expanding supermarket sector with traditional markets. According to Euromonitor, sales from modern retail food markets (that include supermarkets, shopping malls and markets) have more than doubled between 2011 and 2018, from 30.9 to 70.9 trillion VND. This steep increase in sales is also associated with an increase in the number of modern supermarkets nationally from 897 in 2012 to 3272 in 2017

There are multiple trade-offs resulting from this increase in supermarkets. For instance, supermarkets offer the opportunity to provide safe food. As emphasized by Nguyen-Viet et al. (2017), food-borne diseases and food poisoning are a concern in Vietnam, and currently local authorities are promoting policies (211/BC-CP, 2017) that provide improved support for the establishment and spread of supermarkets,

and various traceability solutions to ensure delivery and sale of safe food. In contrast, the expansion of supermarkets has also led to increased availability of ready made and processed foods (see Asfaw, 2008; Zhou et al., 2015; Hawkes, 2008). In the case of Vietnam, while the effects of supermarket expansion on crop value chains, farmer incomes, and food have been evaluated (Maruyama and Trung, 2010; Cadilhon et al., 2006; Moustier et al., 2010), large-scale quantitative evidence on transitions in household-level food acquisition patterns as a result of supermarkets is limited, mainly because either there is lack of useful data, or data is collected by different institutions with limited scope for interoperability.

This paper first aims at contributing quantifiable evidence on the transformative impacts of supermarkets, on the relationships between household socioeconomics, demographics and food acquisition patterns, across years and provinces in Vietnam. Existing administrative data has been used in the past to investigate transitions in patterns of food groups acquired among Vietnamese households, however these have not yet been described in the context of supermarkets, across both time and geographies. To our knowledge, this is the first comprehensive attempt to elucidate the link representative at the national level. Secondly, to circumvent data related challenges linked to the analysis, we implemented an analysis strategy that leverages upon two different official data sources: (a) province level numbers of supermarkets, and (b) the Vietnam Household Living Standard Survey (VHLSS), both collected by the General Statistics Office of Vietnam (GSO). This approach combines multiple statistical approaches, which include time-series cluster analysis, poisson regression analysis and compositional data analysis, thus enabling a multidimensional (diversity and quality of acquired food groups) elucidation of the differences in relationship between household food acquisition, and socioeconomic, demographic characteristics of the same households, across differential presence of supermarkets.

With this study, we aim to advance the evidence base for policy makers to develop policies related to market modernization, with emphasis on their impacts on food environments and food acquisition

patterns, at both national and provincial levels in Vietnam.

The paper is organized as follows. Section 2 introduces the methodology. Technical details about statistical and econometric tools are presented in the appendixes. Section 3 presents and discusses the main results, and section 4 provides final conclusions.

2 Methodology

A data analysis strategy was designed, that enabled interoperability, and facilitated integrated analysis of the two data sets (outlined in Fig. 1). The strategy and the identification of resulting estimations are as follows. First, yearly data (for years 2010, 2011, 2012, 2013 and 2014), on supermarket numbers for each Vietnamese province[1] and was subjected to time series clustering analysis[2] (Aghabozorgi et al., 2015) was applied to classify provinces into three distinct clusters, with significant differences in supermarket numbers between clusters across time. They are: 1) provinces with highest presence of supermarkets (namely, HighSM); 2) provinces with medium presence of supermarkets (namely, MediSM); and 3) provinces with the least presence or absence of supermarkets (namely, LowSM).

We recognize that by using these data we do not take into account the variety of formats that a supermarket can take in general, and more specifically in Vietnam. Not going into the details of these different formats, the data we used capture supermarket as a format with a recognizable and rather uniform product assortment. Thus we cannot address the issue of the diversity in retail assortment, even if it is well known that outcomes on healthy or unhealthy diets may depend on the presence of "safe" fresh foods or ultra-processed foods in supermarkets (See, among others, Monteiro et al., 2013).

[1] Supermarket enumeration data for 63 provinces of Vietnam from 2010 to 2014 were obtained from the General Statistics Office (GSO). This annual data can be obtained from the statistical data section, under the "Trade, Price and Tourism" subsection, on the GSO website (https://www.gso.gov.vn/).

[2] *TSclust* package in *R*, to cluster provinces based on supermarket numbers.

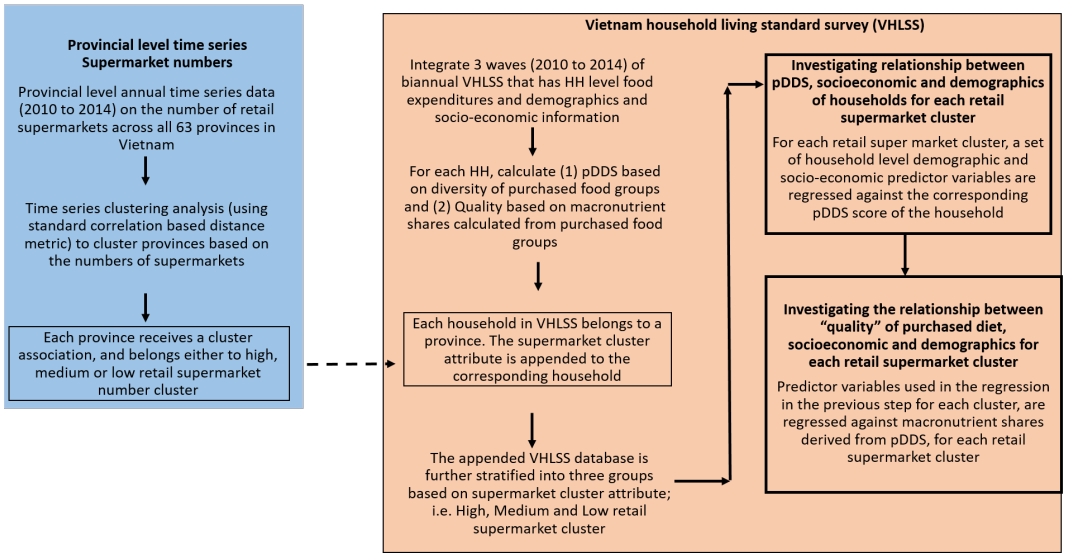

Figure 1: A data analysis strategy that facilitates interoperability of two independent data sets was implemented, which are the province level time series data on supermarket numbers (in the blue colored box), and the Vietnam household living standard survey (in orange colored box). The blue box represents steps taken for time series clustering and statistical analysis of supermarket numbers and the identified clusters, while the orange box describes calculation of the response variables, i.e. aDDS and derived quality indices, and the regression strategies used to estimate relationship between the response variables and socioeconomic, demographic characteristics of the household. The dashed line connecting the two boxes indicates the step at which supermarket cluster information was appended to the VHLSS data set.

Moreover, using that data, we also oversimplify the complexity of Vietnamese food environment in an assumed duality between supermarkets (= modern) and traditional markets. But, here too, the official data that is available, which we use does not allow us to capture the highly complex and dynamic character of retail environment in Vietnam. Nevertheless, clustering analysis of provinces will take into account some interesting characteristics of supermarket presence in Vietnam.

Second, bi-annually (for the years 2010, 2012 and 2014) conducted Vietnam Household Living Standard Survey (VHLSS) was used to obtain household level socioeconomic, demographic and food acquisition data[3]. Each household in the VHLSS survey can be linked to its respective province. This feature of the VHLSS dataset was used, and the supermarket cluster information of the province, obtained as

[3]See Appendix A for a presentation of VHLSS.

a result of time-series clustering was appended to the household data, such that all households received the supermarket cluster membership of the corresponding province. The appended VHLSS data set with supermarket cluster information was then stratified in into subsamples, each representing a supermarket cluster.

Belonging to a given supermarket cluster is a very crude measure of supermarket accessibility for households. Retail density and/or retail proximity variables would have been more meaningful. But, here too, we do not have a sufficiently disaggregated information on spatial coordinates describing the exact location of each supermarket or at least the commune where the supermarket is located. Thus, we can not associate with each household an indicator of its accessibility to supermarkets other than that provided by the clusters.

Household level diversity in relation to acquired food groups over the past 30 days, or aDDS - one of the primary outcome variables - was measured for each household in each subsample. The aDDS was constructed by counting the number of food groups consumed by a household from a list of 9 groups: Starchy Staple foods, Flesh foods, Eggs, Nuts and seeds, Pulses, Vitamin A–rich dark green leafy vegetables, Other vegetables, Other fruits, Dairy product (detailed description of food items belonging to each food group, except OT group, is given in Table A-1). These 9 groups are based on guide to measurement (FAO, 2016) of a representative Vietnamese diet. As emphasized by Verger et al. (2019), aDDS can be viewed as a proxy of access to a variety of food groups in the field of food security research.

A quality indicator of acquired food groups, in terms of relative macronutrient share of fat, carbohydrates and proteins, was also calculated, and used as an additional outcome variable, following Trinh et al. (2018a).[4] Household food acquisition is converted into total kilocalories (namely THCC), protein

[4]Other quality indicators have been used in literature such as "problem food" (Kelly et al., 2014), shares of eight food groups in total calorie availability (Asfaw, 2008), and Diet Quality Index International (Tessier et al., 2008). Most of these indicators are not computable from data such as VHLSS because they require information not available in this kind of survey

(namely, THPC) and fat (namely THFC)[5] using a food composition table (NIN, 2007). As total calorie intakes comes from three types of macronutrients: protein, fat and carbohydrate, the macronutrient shares, denoted by S_P, S_F and S_C, are defined as the proportion (in percentage) of calories coming from protein, fat and carbohydrate, i.e.

$$S_P = \frac{THPC}{THCC},\ S_F = \frac{THFC}{THCC} \text{ and } S_C = 1 - S_P - S_F. \tag{1}$$

Two different regression approaches were used to model the relationships between outcome variables, i.e. acquired diet diversity, and diet quality, and various household-level demographic and socio-economic predictor variables for each supermarket cluster separately, i.e.

$$Y_{i,s} = \alpha_{0,s} + \sum_{j=1}^{J} \alpha_{j,s} X_{i,s}^{j} + \epsilon_{i,s} \tag{2}$$

where $Y_{i,s}$ is diet indicator of household i in cluster s, and $X_{i,s}$ is a vector of characteristics of households, including per capita expenditure, proportion of food consumed from acquired food items, location (urban/rural sites and socio-economic regions), ethnicity, household size and year. Model 2 covers both two indicators:

i. Household diet diversity score, or aDDS: Estimates of parameters were recovered using Poisson regression techniques (see more details in Appendices B).

ii. Household macronutrient shares of diet, or S_P, S_F, S_C: Estimates of parameters were used by

[5]The quantity of protein in grams (resp. fat) is converted into kilocalories by multiplying by 4 (resp. 9), namely THPC and THFC, respectively.

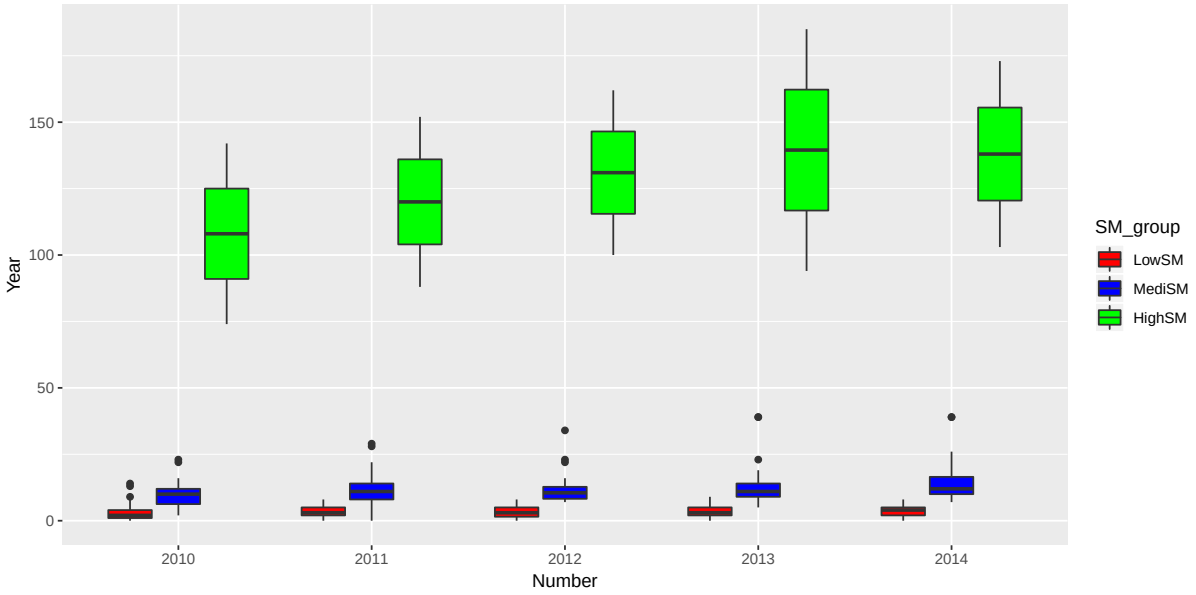

Figure 2: Distributions of the number of supermarkets according to clusters and over time

Compositional data analysis (CoDa). There three shares are transformed in two coordinates:

$$Ilr_1 = \frac{2}{\sqrt{6}} \ln \frac{S_C}{\sqrt{S_F S_P}} \text{ and } Ilr_2 = \frac{1}{\sqrt{2}} \ln \frac{S_F}{S_P}$$

where Ilr_1 is thus the ratio of Carbohydrate share over the geometric mean of other shares while Ilr_2 corresponds to the ratio of Fat share over Protein share (more details are given in Appendices C).

3 Results

3.1 Cluster analysis

Time-series cluster analysis identified three distinct clusters. The distributions of the number of supermarkets in each cluster are depicted in Fig. 2. These three clusters classify Vietnamese provinces on a

Table 1: Number of supermarkets per 100,000 households

	2010	2012	2014
LowSM	1.40	1.35	1.61
MediSM	3.27	4.15	5.11
HighSM	7.21	8.84	9.29

continuum of supermarket presence, with each cluster representing a food environment with differential presence of supermarkets. More specifically, provinces are classified into: i) provinces with highest number of supermarkets (cluster denoted by HighSM); ii) provinces with medium number of supermarkets (cluster MedSM); and, iii) provinces with the lowest number of supermarkets (cluster LowSM). As expected, provinces largely dominated by urban areas like Hanoi and Ho Chi Minh City were associated with the HighSM cluster (126 supermarkets per province, on average), while medium provinces like Tuyen Quang, Hai Phong, and Quang Binh had lower supermarket numbers than those in the largest Vietnamese cities, and hence were clustered in MedSM cluster (13 supermarkets per province on average). Poor provinces such as Ha Giang, Ha Tinh, and Quang Nam had significantly lower number of supermarkets than the provinces in HighSM and MedSM clusters (3 supermarkets per province, on average). Fig. 3 describes the spatial distribution of the three clusters.

Fig. 2 also shows the increase in the number of supermarkets in the HighSM provinces, with a slowdown in 2013 and 2014. This increase also appears for provinces in the MedSM cluster but is less pronounced, while it is very slow for LowSM cluster provinces. Table 1 shows the number supermarket per 100,000 households by year and by cluster. These numbers continue showing the increase of supermarket per 100,000 households in MediSM and HighSM over years.

Table 2 shows the average characteristics of households living in the three clusters. As expected, households living in HighSM cluster provinces spend more proportion of their income on purchase of food items, they belong to urban areas, and more educated than those living in the other two clusters.

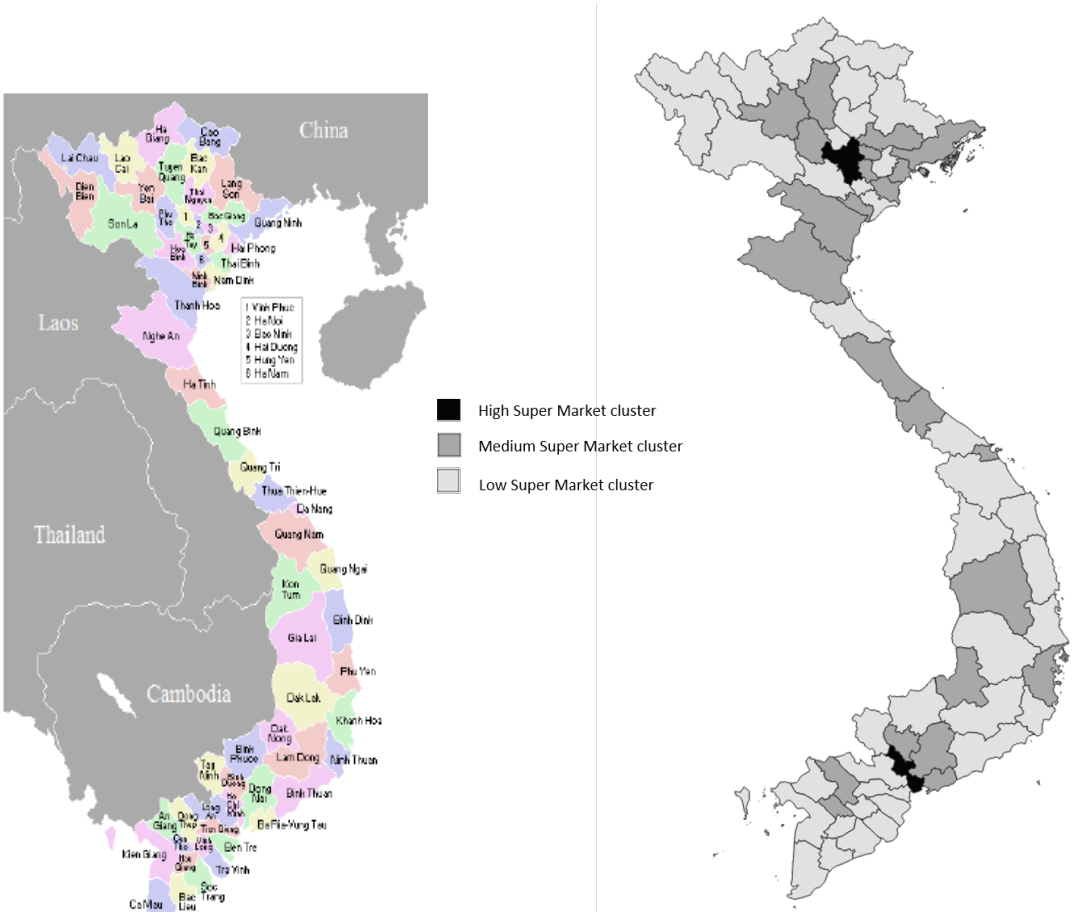

Figure 3: Spatial distribution of clusters

More women are household head in HighSM cluster provinces, and most households in this cluster belong to the dominant ethnicity in Vietnam, i.e. Kinh. Moreover, HighSM cluster provinces households are consuming more purchased food items than households in the other two clusters. The key differences between households in the MedSM and LowSM clusters are as follows. Households in MedSM cluster provinces are spending more in food items than Households in LowSM cluster provinces, even if the proportion of food consumed from purchases are the same. They are also more educated. Lastly, ethnic minorities are more prevalent among households in the LowSM cluster provinces.

Table 2: Descriptive statistics of social, economic and demographic variables by supermarket clusters

Variable	HighSM	MedSM	LowSM
Per capita expenditure (US$)	883.4 (417.2)	609.8 (334)	547.1 (307.8)
Proportion of food consumed from purchases	90%	80%	80%
Rural residence	42.37%	70.16%	76.74%
Urban residence	57.63%	29.84%	23.26%
Household with 2 members	17.42%	19.98%	15.31%
Household with 3 members	21.26%	18.78%	19.73%
Household with 4 members	33.05%	32.49%	32.52%
Household with 5 members	14.74%	16.47%	17.97%
Household with 6 members	13.53%	12.29%	14.47%
Minority ethnic	4.32%	10.94%	23.28%
Kinh ethnic	95.68%	89.06%	76.72%
Male head of household	63.26%	76.40%	77.77%
Female head of household	36.74%	23.60%	22.23%
Head of household obtained Below primary	34.05%	42.65%	58.11%
Head of household obtained Secondary, High school	50.74%	51.71%	37.58%
Head of household obtained University	15.21%	5.65%	4.31%
Year 2010	32.42 %	33.67%	33.69%
Year 2012	33.01%	32.65%	32.42%
Year 2014	34.58%	33.67%	33.89%

aDDS outcome ranges from 1 to 9 which corresponding to households who consume 1 to 9 over total of nine food groups. Figure 4 shows number of household who consumes 9, 8, 7 6, and less 6 food groups, in percentage. Over three years, the percentage of households who acquired 9 and 8 food groups are much higher in HighSM than two other clusters. The percentage of households acquired each number of food groups in LowSM are slightly higher than those of MediSM.

Table 3 shows the evolution of macronutrient shares by supermarket clusters between 2010 and 2014.

Figure 4: aDDS Distributions by cluster and year (in percentage)

MedSM and LowSM clusters tend to have similar mean profiles, even if in 2010 carbohydrate share was higher in the LowSM cluster. The HighSM cluster differs markedly form the other two, with higher average fat and protein levels, and consequently lower average carbohydrate share.

Table 3: Evolution of macronutrient shares by supermarket clusters and over years.

	2010			2014		
Cluster	S_P	S_F	S_C	S_P	S_F	S_C
LowSM	14.6	15.4	70.0	14.9	16.9	68.3
MediSM	14.8	16.7	68.5	14.7	17.9	67.4
HighSM	16.5	21.4	62.1	16.3	21.6	62.1

The "recommended" (or "ideal") composition of macronutrients in a Vietnamese diet, according to Vietnamese Ministry of Health, consists of 14% for protein, 18% for fat, and 68%, for carbohydrate (Ministry of Health, 2010). A comparison of average macronutrient shares in supermarket clusters to "ideal" macronutrient shares is given in Fig. 5 using a ternary diagram[6]. Both MedSM and LowSM clusters have, on average, reached the target, while the High SM cluster exceeded it, with higher protein and fat shares, and, consequently, smaller carbohydrate share. Similar observation about urban Vietnamese households when compared to rural ones can be found in Trinh et al. (2018).

[6]A ternary diagram is a triangular coordinate system where the edges of the triangle are the axes. Each axe is associated with a component with a variation from 0 to 100%. So, for example, the axis from the top titled "Carbohydrate" to the top "Protein" is associated with the variation in protein component. We added the location of the vector with $S_p = 10\%$, $S_F = 20\%$ and $S_C = 70\%$, to make this diagram reading easier.

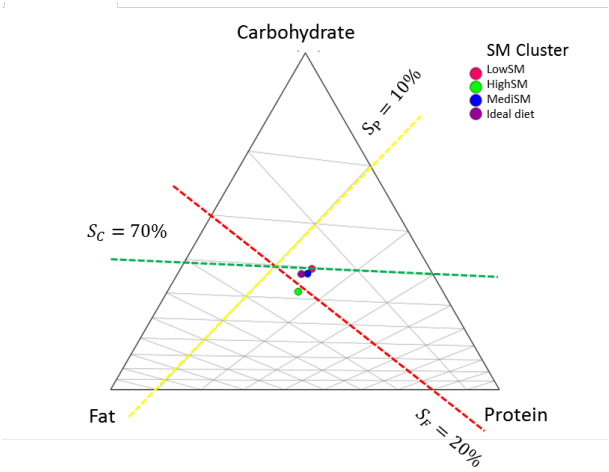

Figure 5: Ternary diagram of average macronutrient shares for each supermarket cluster.

3.2 Regression analysis

Table 4 summarizes the results of estimation of the different regression models. For simplicity, we report the signs of the impacts of variables whose impacts on the different outcomes are significantly different from zero, and that for each supermarket cluster. Detailed estimation results are given in Tables A-2 and A-3.

Differential effects of per capita expenditure and proportion of food consumed from purchases on acquired diet diversity and quality across supermarket clusters

Strong growth of Vietnam economy has improved financial capabilities of households in the past years, with increase in per capita expenditure and proportional increase in total acquired food groups from purchases. From 2010 to 2014, households in LowSM and MedSM have shown an increase of 3.5% in total proportion of acquired food from purchases, in contrast to households in HighSM cluster, that have shown a smaller increase of 2.4% on average. Our results clearly indicate a positive impact of per capita expenditure and proportion of food consumed form purchases on diet diversity, i.e aDDS. This has also been shown in previous studies (Drescher et al., 2009; Chandon and Wansink, 2012). Interestingly

Table 4: Summary of estimation results: directions of significant impacts

Predictor variables:	HighSM			MedSM			LowSM		
Name	aDDS	Ilr1	Ilr2	aDDS	Ilr1	Ilr2	aDDS	Ilr1	Ilr2
Per capita expenditure (US$)	+	-	-	+	-	+	+	-	+
Proportion of food consumed from purchases	+	-	+	+	-	+	+	-	+
Household status	||								
(Reference: Rural)									
Urban residence				+	-			-	
Household size									
(Reference: 2 members)									
Household with 3 members	+	+	+	+	+		+	+	
Household with 4 members	+	+	+	+	+		+	+	
Household with 5 members	+	+	+	+	+	-	+	+	-
Household with 6 members	+	+	+	+	+		+	+	-
Household Ethnicity									
(Reference: Kinh)	||								
Minority ethnicity				-	-		-	+	-
Area									
(Reference: Red River Delta)				||				
Midlands Northern Mountains							-	+	+
Northern Central Coast	||						-	+	-
Central Highlands							-	+	-
South East	-		-				-	+	-
Mekong River Delta							-	+	-
Year	||								
(Reference: 2010)									
Year 2012									
Year 2014			+	-	-	+	-	-	+

+ and − means positive and negative impact
Double vertical bars mean that these explanatory variables were not chosen by the best performing Poisson regression model
HighSM, LowSM and MedSM correspond to clusters with high, low and medium number of supermarkets.

though, presence of supermarkets in the food environment do not impact this positive relationship, as this trend is observed across all supermarket clusters.

In terms of magnitude, the effect of an increase of one point in per capita expenditure is always smaller than that of an increase of one point in the proportion of food consumed from purchases (see Table A-2). Indeed, the first effect is between 9% (HighSM) and 12% (LowSM) while the second effect goes from 12% (MedSM) to 23% (HighSM).

These result seem to translate firstly a wealth effect. It is indeed possible to consider that the more a household is able to spend and the higher its share of purchased food items in its food consumption, the

more its income is large. Thus, households are able to acquire more diverse food groups, hence leading to high aDDS. In addition, food supply in Vietnam has diversified in recent years, allowing diversification in food purchases.

The impacts of per capita expenditure on diet quality are more nuanced. Consider first HighSM cluster households. For these households, an increase of per capita expenditure results in a decrease of the Ilr_1 variable. A negative sign of Ilr_1 means a decrease in carbohydrate share and a subsequent increase in the geometric mean of protein and fat shares. Similarly, the effect of per capita expenditure on Ilr_2 is also negative, i.e an increase in per capita expenditure results in a decrease in fat share and an increase in protein share. To sum up, an increase in per capita expenditure results in a substitution of protein macronutrient to the other two macronutrients for HighSM cluster households. Interestingly, a different trend is observed among households in MedSM and LowSM cluster. Unlike HighSM cluster households, Ilr_2 increases when per capita expenditure increases, meaning a substitution of fat to protein. The analysis of the impact of proportion of food consumed from purchases on diet quality, for its part, clearly demonstrates a substitution effect of fat macronutrient to carbohydrate macronutrient, whatever the considered supermarket cluster. In addition, in HighSM cluster, an increase of proportion of food consumed from purchases leads to an increase of fat share with the expense of protein share.

The interpretation of effect magnitude here is much more difficult than in the case of Poisson model. Nevertheless, it should be noted that, although significantly different from zero, the effects of per capita expenditure are almost insignificant when compared to those of the proportion of food consumed from purchases (see Table A-3).

To sum up, households with better financial means (i.e. higher per capita expenditure, and higher proportion of income to purchase food groups), acquire much more diverse food groups than those with lower financial means, and this is independent of supermarket clusters. However analysis of the quality

of acquired food groups, reveals that households with higher per capita expenditure, acquire food groups with larger fat and protein shares, in HighSM households, whereas households with higher per capita expenditure in LowSM and MedSM cluster acquire food groups with larger carbohydrate shares.

Other socio-demographic variables effects. Significant impacts of living in an urban area on aDDS and macronutrient shares only appear when considering MedSM and LowSM clusters. These results are in line with the argument of larger presence and improved access of households to supermarkets, i.e leading to increase in diversity of acquired food groups. This argument is further strengthened by the observation that, urban households, with higher per capita expenditure, with increased purchase of total acquired food (instead of self-production), especially in the MedSM cluster, have higher aDDS, than that of rural households in the same cluster. Moreover, results show a substitution effect of fat and protein to carbohydrate in MedSM and LowSM clusters.

Larger size households across all clusters display larger food diversity, compared to households with at most two members. The magnitude of incidence rate ratios in Table A-2 confirm this increasing effect. It is easy to see the need to increase food diversity to feed more members in household. Moreover, an increase in household size clearly also seems to imply a substitution of carbohydrate macronutrient to the other two macronutrients in all clusters, and a substitution of fat to protein for all clusters. It should be noted here that, in terms of magnitude, substitution of carbohydrate macronutrient to the other two macronutrients patterns are very similar for HighSM and MedSM clusters, while much smaller for LowSM cluster (see Table A-3). All of this supports the argument that larger households acquire food items with higher carbohydrate content. Moreover, this relationship is more complex in food environments with high number of supermarkets, wherein large sized households also acquire food groups with higher shares of both carbohydrate and fat macronutrients, while especially in lowSM cluster, households acquire higher carbohydrates shares, but also significantly lower shares of fats.

A substitution effect in quality of purchased food groups is also observed in ethnic minority households of LowSM and MedSM clusters. In both low SM and MedSM, ethnic minority households have a lower aDDS score in comparison to Kinh ethnic households. Interestingly, quality of food groups acquired by ethnic minority households, differ in terms of carbohydrate shares between the two clusters. This results may further reveal the traditional diet of minority people with smaller aDDS.

Limited access to supermarkets in food environments is also reflected when assessing the impact of household location. Interestingly the impact of area variable is only observed for LowSM cluster. Within the LowSM cluster, households in the low lying, fertile, with diverse agricultural production systems, Red River Delta (Kurosawa et al., 2004), have higher aDDS in comparison to households in other areas in the same cluster. This additional diversity, even in the absence of supermarkets, can be potentially explained by other food sourcing channels (such as traditional markets) in the food environment of the Red River Delta.

In addition to Red River Delta, the Midlands Northern Mountainous area also supports diverse production systems, however in similarity to the other areas (excepting the Red River Delta), this region has relatively lower infrastructure development (reflected by its presence in the LowSM cluster). The presence of diverse production systems, and their access through existing food sourcing channels (such as traditional markets), in the Midlands Northern Mountain area, can be attributed to increased acquisition of food groups with both high fat and carbohydrate shares by households in this area.

Traditional markets are still the most predominant channel for sourcing food for households in Vietnam (Maruyama and Trung, 2007; Vo and Smith, 2016). This analysis also points to the observation, that traditional markets are a potential source of food groups with larger carbohydrate share and low fat share, unlike the supermarkets, as observed from the quality of acquired food by ethnic minority households, large sized families both in low and medium supermarket clusters, or households in Central Highlands,

South East, Northern Central Coast areas, that acquire major proportion of their food from traditional market. This is in contrast households in HighSM cluster or households living in cities in MedSM cluster, for which more food acquired is associated with high fat and low carbohydrate shares.

4 Conclusion

Existing administrative data (such as the VHLSS and others) have been used in the past to investigate transitions in patterns of food groups acquired among Vietnamese households (see, among others, Trinh et al., 2018a). However, these data have not yet been used to describe associations between changes in supermarkets and consumption patterns across both time and geographies. In addition, the use of relatively new statistical approaches, such as compositional data regression analysis to capture nuanced associations of quality of food group acquired in relation to supermarkets have also not been performed previously.

A data analysis strategy that leverages the quality of information across two disparate administrative datasets (i.e. province level supermarket numbers and household levels in VHLSS), and a combination of statistical approaches, enabled this multidimensional (diversity and quality of acquired food groups) elucidation of the differences in relationship between household food group acquisition, and socioeconomic, demographic characteristics of households, across differential presence of supermarkets (as signified by the three supermarket clusters, i.e. HighSM, MedSM and LowSM).

Overall, analysis of existing data reveals that supermarkets have had, and continue to have complex and profound influence on the food environment surrounding Vietnamese households, that is further leading to an evolved relationship between socioeconomics, demographics on diversity and quality of food groups acquired. There are also specific relationships that have not yet been influenced by supermarkets.

This relationship is in line with the comprehensive framework and critical actions on promote food environment in other countries (Turner et al., 2018).

Vietnam like other developing economies is starting to fall "prey" to the problem of double burden of malnutrition, with co-existence of both undernutrition and over nutrition related chronic disorders (Khan and Khoi, 2008; Nguyen et al., 2018). Results from this piece of work provide potential intervention options to address this issue, and use the expansion of supermarkets, as an opportunity to "rewire" local food environments, to address challenges related to the double burden of malnutrition. There is a need to catalyze equitable infrastructure development, which will lead to more supermarkets in food environments surrounding ethnic minority households, or in areas with income and infrastructure disparities. This will help households in such environments acquire a larger diversity of food groups, possibly bringing in more diversity to their diets, thereby contributing to malnutrition reduction in these communities. Increased acquisition of diverse food groups, does not imply healthy diets. Results from this work provide ample evidence of this, such as increased proportion of fats and proteins in relation to carbohydrates in food groups purchased in food environments with higher presence of supermarkets. Food environments need to be designed in a manner, that facilitate the ability to acquire diverse food groups, but also balance the macronutrient shares closer to the recommended diet. Our results show that it is possible to achieve this, through a mix of improved ability to spend (increased per capita expenditure) by a household, and by increasing the number of supermarkets in food environments of relatively poor households. However, the growth and expansion of supermarkets needs to be regulated, as there is a strong possibility of facing the negative-effects of development, such as increased incidence of diabetes, obesity etc. as observed in economically well developed countries. Supermarkets are bound to expand in the coming future, however "intelligent" food environments with planned, nutritionally sensitive introduction and expansion of supermarkets, can indeed rewire the food environment, and better tackle the problem of double burden

of malnutrition.

References

211/BC-CP (2017). The situation of policy implementation on food safety management in the period of 2011-2016 in vietnam.

Aghabozorgi, S., A. S. Shirkhorshidi, and T. Y. Wah (2015). Time-series clustering – a decade review. *Information Systems 53*, 16–38.

Asfaw, A. (2008). Does supermarket purchase affect the dietary practices of households? some empirical evidence from guatemala. *Development Policy Review 26*(2), 227–243.

Barker, T. and M. Üngör (2019). Vietnam: The next asian tiger? *The North American Journal of Economics and Finance 47*, 96–118.

Cadilhon, J.-J., P. Moustier, N. D. Poole, P. T. G. Tam, and A. P. Fearne (2006). Traditional vs. modern food systems? insights from vegetable supply chains to ho chi minh city (vietnam). *Development Policy Review 24*(1), 31–49.

Cameron, A. C. and P. K. Trivedi (2013). *Regression analysis of ccount data*. Cambridge: Cambridge University Press.

Chandon, P. and B. Wansink (2012). Does food marketing need to make us fat? a review and solutions. *Nutrition reviews 70*(10), 571–593.

Drescher, L. S. ansd Thiele, S., J. Roosen, and G. B. Mensink (2009). Consumer demand for healthy eating considering diversity âĂŞ an economic approach for german individuals. *International Journal of Consumer Studies 33*, 684–696.

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

FAO, F. (2016). Minimum dietary diversity for women: a guide for measurement. *Rome: FAO*.

Hawkes, C. (2008). Dietary implications of supermarket development: a global perspective. *Development Policy Review 26*(6), 657–692.

HLPE (2017). Nutrition and food systems. a report by the high level panel of experts on food security and nutrition of the committee on world food security. *FAO-Rome*.

Kelly, M., S.-a. Seubsman, C. Banwell, J. Dixon, and A. Sleigh (2014). Thailand's food retail transition: supermarket and fresh market effects on diet quality and health. *British Food Journal 116*(7), 1180–1193.

Khan, N. C. and H. H. Khoi (2008). Double burden of malnutrition: the vietnamese perspective. *Asia Pacific journal of clinical nutrition 17*(S1), 116–118.

Kilic, T., U. Serajuddin, H. Uematsu, and N. Yoshida (2017). *Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity*. The World Bank.

Kurosawa, K., D. Nguyen Hai, N. Huu Thanh, H. Thi Lam Tra, N. Tat Canh, and K. Egashira (2004). Monitoring of inorganic nitrogen levels in the surface and ground water of the red river delta, northern vietnam. *Communications in soil science and plant analysis 35*(11-12), 1645–1662.

Lu, L. and T. Reardon (2018). An economic model of the evolution of food retail and supply chains from traditional shops to supermarkets to e-commerce. *American Journal of Agricultural Economics 100*(5), 1320–1335.

Maruyama, M. and L. V. Trung (2007). Supermarkets in vietnam: Opportunities and obstacles. *Asian Economic Journal 21*(1), 19–46.

Maruyama, M. and L. V. Trung (2010). The nature of informal food bazaars: Empirical results for Urban Hanoi, Vietnam. *Journal of Retailing and Consumer Services 17*(1), 1–9.

Ministry of Health (2010). National plan of action for nutrition to 2020. *Hanoi, Vietnam: Medical Publishing House*.

Monteiro, C. A., J.-C. Moubarac, G. Cannon, S. W. Ng, and B. Popkin (2013). UltraâĂŘprocessed products are becoming dominant in the global food system. *Obesity Reviews 14*(52), 21–28.

Morais, J., C. Thomas-Agnan, and M. Simioni (2018a). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics 47*(5), 1–25.

Morais, J., C. Thomas-Agnan, and M. Simioni (2018b). Using compositional and dirichlet models for market share regression. *Journal of Applied Statistics 45*(9), 1670–1689.

Moustier, P., P. T. G. Tam, D. T. Anh, V. T. Binh, and N. T. T. Loc (2010). The role of farmer organizations in supplying supermarkets with quality food in vietnam. *Food policy 35*(1), 69–78.

Müller, I., K. Hron, E. Fiserova, J. Smahaj, P. Cakirpaloglu, and J. Vancakova (2018). Interpretation of compositional regression with application to time budget analysis. *Austrian Journal of Statistics 47*(2), 3–19.

Nguyen, T. T., M. V. Hoang, et al. (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.

Nguyen-Viet, H., T. T. Tuyet-Hanh, F. Unger, S. Dang-Xuan, and D. Grace (2017). Food safety in vietnam: where we are at and what we can learn from international experiences. *Infectious diseases of poverty 6*(1), 39.

NIN (2007). Vietnamese food composition table. *Hanoi, Vietnam: Medical Publishing House*.

Popkin, B. M. and P. Gordon-Larsen (2004). The nutrition transition: worldwide obesity dynamics and their determinants. *International journal of obesity 28*(S3), S2.

Reardon, T. and C. P. Timmer (2007). Transformation of markets for agricultural output in developing countries since 1950: How has thinking changed? *Handbook of agricultural economics 3*, 2807–2855.

Reardon, T., C. P. Timmer, and B. Minten (2012). Supermarket revolution in asia and emerging development strategies to include small farmers. *Proceedings of the National Academy of Sciences 109*(31), 12332–12337.

Rischke, R., S. C. Kimenju, S. Klasen, and M. Qaim (2015). Supermarkets and food consumption patterns: the case of small towns in kenya. *Food Policy 52*, 9–21.

Solans, M., G. Coenders, R. Marcos-Gragera, A. Castelló, E. Gràcia-Lavedan, Y. Benavente, V. Moreno, B. PÃľrez-GÃşmez, P. Amiano, T. FernÃąndez-Villa, M. Guevara, I. GÃşmez-Acebo, G. FernÃąndez-TardÃşn, M. Vanaclocha-Espi, M. Chirlaque, R. Capelo, R. Barrios, N. AragonÃľs, A. Molinuevo, F. Vitelli-Storelli, J. Castilla, T. Dierssen-Sotos, G. CastaÃśo-Vinyals, M. Kogevinas, M. PollÃąn, and M. Saez (2018). Compositional analysis of dietary patterns. *Statistical Methods in Medical Research*.

Tessier, S., P. Traissac, B. Maire, N. Bricas, S. Eymard-Duvernay, J. El Ati, and F. Delpeuch (2008). Regular Users of Supermarkets in Greater Tunis Have a Slightly Improved Diet Quality. *Journal of Nutrition 138*(8), 768–775.

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

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

Trinh, H. T., M. Simioni, and C. Thomas-Agnan (2018b). Decomposition of changes in the consumption of macronutrients in vietnam between 2004 and 2014. *Economics and Human Biology 31*, 259 – 275.

Turner, C., A. Aggarwal, H. Walls, A. Herforth, A. Drewnowski, J. Coates, S. Kalamatianou, and S. Kadiyala (2018). Concepts and critical perspectives for food environment research: A global framework with implications for action in low-and middle-income countries. *Global food security 18*, 93–101.

Umberger, W., D. Zeng, J. Rupa, N. Dumbrell, and A. D. Nguyen (2018). The vietnam urban food consumption and expenditure study. factsheet 4: Where do consumers shop? wet markets still dominate.

Verger, E. O., T. J. Ballard, M. C. Dop, and Y. Martin-Prevel (2019). Systematic review of use and interpretation of dietary diversity indicators in nutrition-sensitive agriculture literature. *Global Food Security 20*, 156–169.

Vo, K. and G. Smith (2016). Vietnam retail foods sector report 2016. *Report Number: VM6081 (Global Agricultural Information Network. USDAA Foreign Agricultural Service)*.

Zezza, A., C. Carletto, J. L. Fiedler, P. Gennari, and D. Jolliffe (2017). Food counts. measuring food consumption and expenditures in household consumption and expenditure surveys (hces). introduction to the special issue. *Food Policy 72*, 1–6.

Zhou, Y., S. Du, C. Su, B. Zhang, H. Wang, and B. M. Popkin (2015). The food retail revolution in china and its association with diet and health. *Food Policy 55*, 92–100.

Appendices

A Vietnam Household Living Standard Survey (VHLSS)

VHLSS belongs to the family of Household and Expenditure Surveys (HCES) which are conducted regularly in several countries over several years (Kilic et al., 2017). HCES have also been used previously to study food consumption and nutrition at national levels, with the objective of supporting the Sustainable Development Goals (SDG) (Zezza et al., 2017). Household level socio-economic, demographic and expenditure on food groups were obtained from three waves: 2010, 2012 and 2014, of the nationally representative VHLSS surveys, comprising about 9000 households in each wave. These variables have been previously used to study the impact of socioeconomic characteristics on Vietnamese diets previously: see Trinh et al. (2018a), Trinh et al. (2018b), and Trinh et al. (2018).

Household level food acquisition patterns are based on regular food acquisition, including purchased goods, self-supplied food (home production) and received in kind in kg, over a recall period of 30 days for a wide range of food items. Table A-1 shows detail food categories in each food group. The number of food groups acquired by a household over a recall period of thirty days, during each wave of VHLSS, is used to calculate the acquired Dietary diversity score (aDDS) for each household.

B Poisson regression analysis

Poisson regression analysis (Cameron and Trivedi, 2013) was performed to estimate the relationship between socio-economic demographic characteristics and year on aDDS, as response variable. For each supermarket cluster, a full model with all explanatory variables and a null model with only the intercept as explanatory variable were constructed. A step wise variable selection and regression approach was employed. Residuals from both forward and backward step wise selection approach were used to assess the residual deviance, and a goodness of fit test was performed for each model, obtained from forward and backward selection. The model that was significantly different from the null model, and not different from a full model, and the one that obtained a relatively lower AIC (Akaike Information Criterion) score, in comparison to the full model, was selected and further described. *Sandwich* and *lmtest* packages in *R* were used to calculate the standard errors of the final model using a robust variance estimator to correct for over-dispersion in the residuals. The *msm* package was used to calculate incidence rate ratios. Each incidence rate ratio measures by which factor expected aDDS will increase (when the ratio is greater than one) or decrease (when the ratio is smaller than one) when the corresponding explanatory variable increases by one point, while holding all other variables in the model constant.

C Compositional data analysis (CODA)

To estimate the relationship between socioeconomic, demographic and time variable impact on diet quality indices, CODA approach was performed separately for each supermarket cluster. Compositional data are those which contain only relative information and are parts of some whole. In most cases they are recorded as closed data, i.e. data summing to a constant, such as 100% - macronutrient shares in diet being a good example. Compositional data are always positive and range only from 0 to 100, or any

other constant, when given in closed form. If one component increases, others must, perforce, decrease, whether or not there is a genetic link between these components. This means that the results of standard statistical analysis of the relationships between raw components or parts in a compositional data set are clouded by spurious effects. Several recent papers (Trinh et al., 2018; Solans et al., 2018) have applied compositional analysis to food consumption patterns.

A vector of macronutrient shares, or (SC, SF, SP) is a composition of three positive components for which only the relative information is of interest as $SC+SF+SP = 1$. As three components are involved in the composition, these composition lies in an equilateral triangle surface in $\mathbf{R}^3$, and it is common to represent the data in a ternary diagram such as Figure A-4, which is an equivalent representation. Compositional models are "transformation models" in the sense that they assume a distribution (generally Gaussian) for a log-ratio transformation of shares. Transformation models have several advantages compared to other share models: they are easy to estimate (usually by OLS on coordinates) and flexible in terms of explanatory variables (they can be compositional or non-compositional variables, with or without component-specific parameters) (Morais et al., 2018b). For instance, in our application, the three macronutrient shares are summarized in two Isometric Log-Ratios, or Ilr, coordinates.

It has been shown that ILR transformation of shares into real coordinates is very flexible as it preserves all the metric properties that characterize shares in the simplex (Egozcue et al., 2003).

The interpretation of regression parameters of above compositional regression models are complex. In our analysis, we follow the interpretation proposed in Müller et al. (2018). Alternative interpretation is elasticity of shares, as in Morais et al. (2018a). The entire regression analysis and processing of the compositional data was conducted in *R* statistical computing environment, using the *compositions* and *robCompositions* packages.

D Additional Tables

Table A-1: Description of food groups

Name	Description	Included food items
SSF	Starchy staple foods	Rice, Maize, Cassava, Potato Wheat grains, bread, wheat powder Floor noodle, instant rice noodle Pooridge, rice noodle, Vermicelli
FF	Flesh foods	Pork, Beef Buffalo meat, Chicken meat, other poultry meat Other types of meat (goats, dogs, etc... Processed meat Lard, cooking oil, dresh shrimp Fish, dried and processed shrimps Fish, Other aquatic products and seafoods
Eggs	Eggs	Eggs of chicken, ducks, Muscovy ducks, geese
ND	Nuts and seeds	Tofu, Peanuts, sesame
Pulses	Pulses	Beans of various kinds
VDV	Vitamin A–rich dark green leafy vegetables	Morning glory vegetables
OV	Other vegetables	Fresh peas, Kohlrabi, Cabbage, Tomato Other vegetables
OF	Other fruits	Other fruits
Dairy	Dairy	Condensed milk, milk powder Ice cream, yogurt, Fresh milk
OT	Other foods	All other foods and food away from home

Table A-2: Incidence rate ratios obtained from Poisson regressions

	Supermarket clusters		
Predictor variables:	High SM	MedSM	LowSM
Per capita expenditure (US$)	1.09***	1.10***	1.12***
Proportion of food consumed from purchases	1.23*	1.12***	1.17***
Urban residence (Reference level: Rural residence)		1.01*	1.01
Minority ethnic (Reference level: Kinh Ethnic)		0.90***	0.93***
Household with 3 members (Reference level: Household with 2 members)	1.05*	1.05***	1.05***
Household with 4 members (Reference level: Household with 2 members)	1.09***	1.08***	1.10***
Household with 5 members (Reference level: Household with 2 members)	1.11***	1.12***	1.13***
Household with 6 members (Reference level: Household with 2 members)	1.12***	1.15***	1.14***
Midlands Northern Mountains (Reference level: Red River Delta)			0.93***
Northern Central Coast (Reference level: Red River Delta)			0.97*
Central Highlands (Reference level: Red River Delta)			0.96*
South East (Reference level: Red River Delta)	-0.97		0.95*
Mekong River Delta (Reference level: Red River Delta)			0.92***
Year 2012 (Reference level: Year 2010)		0.97**	0.98**
Year 2014 (Reference level: Year 2010)		0.98	0.98**

Results from Poisson regressions with aDDS as the response variable and household level social, economic and demographic variables obtained from VHLSS as predictor variables.
Poisson regressions are performed for each supermarket cluster. Reported figures are incidence rate ratios, wherein a value above 1 (resp. below 1)indicates positive (resp. negative) relationship between aDDS and the corresponding predictor variable.
Asterisks correspond to significance at p-value < 0.05 for *, p-value < 0.01 for **, and p-value < 0.001 for ***.

Table A-3: Results for the compositional data regression model estimation

	Supermarket clusters					
Predictor variables	HighSM		MediSM		LowSM	
Per capita expenditure (US$)	-1e-04 ***	-1e-04 ***	-2e-04 ***	1e-04 ***	-3e-04 ***	1e-04 ***
Proportion of food consumed from purchases	-0.4206 ***	0.1764 ***	-0.1758 ***	0.00777	-0.2549 ***	0.1273 ***
Urban residence			-0.0247 ***	-0.0049	-0.0285 ***	0.0052
(Reference level: Rural residence)						
Minority ethnic (Reference level: Kinh Ethnic)			0.0409 ***	-0.0015	-1.4886 ***	0.0353 ***
Household with 3 members (Reference level: Household with 2 members)	1.5298 ***	0.0587 *	1.5217 ***	-0.0007	0.0371 ***	-0.0009
Household with 4 members (Reference level: Household with 2 members)	1.5029 ***	0.0615 *	1.5403 ***	-0.006	0.0475 ***	-0.0046
Household with 5 members (Reference level: Household with 2 members)	1.5171 ***	0.0719 *	1.571 ***	-0.0216 *	0.0721 ***	-0.0085
Household with 6 members (Reference level: Household with 2 members)	1.5169 ***	0.0523 .	1.5908 ***	-0.026	0.0922 ***	-0.0279 ***
Midlands Northern Mountains (Reference level: Red River Delta)					0.0211 *	0.0231 ***
Northern Central Coast (Reference level: Red River Delta)					0.1129 ***	-0.171 ***
Central Highlands (Reference level: Red River Delta)					0.0897 ***	-0.0928 ***
South East (Reference level: Red River Delta)	1.61E-2	-0.0878 ***			0.1025 ***	-0.0945 ***
Mekong River Delta (Reference level: Red River Delta)					0.101 ***	-0.2044 ***
Year 2012 (Reference level: Year 2010)			-0.0365 ***	0.0725 ***	-0.0345 ***	0.0692 ***
Year 2014 (Reference level: Year 2010)			-0.0209 **	0.0485 ***	-0.0366 ***	0.0457 ***
R_2 adjust	0.657	0.894	0.324	0.958	0.294	0.953

Results from composition regression analysis with ilr_1 and ilr_2 as the outcomes and household level social, economic and demographic variables obtained from VHLSS as predictor variables.
Each regression corresponds to a supermarket cluster. Numbers indicate regression coefficients, wherein values positive values signify positive and negative values signify negative relationship between the outcome variables and the corresponding predictor variable from VHLSS.
Asterisks correspond to significance at p-value < 0.05 for *, p-value < 0.01 for ** and p-value < 0.001 for ***.