Wageningen University - Department of Social Sciences

MSc Thesis Chair Group: Development Economics Group

Rural-Urban Migration, Food Consumption Patterns and Trends in the Local Food System: A Case Study for Viet Nam

MSc Thesis by Marieke Meeske
February 2018

MSc Management, Economics and Consumer Studies
Specialisation: Economics and Governance

Supervisor: Marrit van den Berg

DEC80433

WAGENINGEN UNIVERSITY & RESEARCH

Rural-Urban Migration, Food Consumption Patterns and Trends in the Local Food System: A Case Study for Viet Nam

W.M. Meeske (940517553040)
2nd of Feburary 2018

Supervisor:
Dr. Ir. M.M. van den Berg
Associate Professor
Development Economic Group
Wageningen University & Research

Examinator:
Prof. Dr. Ir. E.H. Bulte
Professor
Development Economic Group
Wageningen University & Research

Master Thesis Development Economics Group in partial fulfilment of the degree of Master of Sceince in master program Management Economics and Consumer Studies at Wageningen University & Research, the Netherlands (Course Code: DEC-80433).

Abstract

In the context of increasing population and internal migration rates, together with impressive economic growth, food consumption patterns in Viet Nam are changing. Food consumption patterns have moved gradually from a rice and cereal based diet to a more diverse and healthy bundle of foods, including greater quantities of meat, vegetables, and dairy products. These shifts have been more profound in urban areas, which have experienced more economic growth than rural areas. From total internal migration, rural-urban migration is the largest flow. Given the fact that urban areas are different than rural areas and that they might be at a further stage of the nutritional transition than rural areas, rural-urban migration might affect the diet of migrants. All these processes take place within Viet Nam's broader food system. The objective of this thesis is to explore trends in the local food system in Viet Nam, specifically focusing on food consumption patterns, agricultural production, and international trade. Furthermore, it aims to explore whether rural-urban migration and having rural origins affect the diet of migrants. Data in this thesis comes from literature studies, the World Bank, the Food and Agricultural Organization of the United Nations, and the Vietnam Household Living Standards Survey (VHLSS). It is found that the number of consumed kilocalories per capita is increasing, with calories increasingly coming from proteins and fat instead of carbohydrates. Furthermore production, exports, and imports of nearly all food crops has increased over the years. Especially the production of rice is considerable. The analysis of the effect of rural-urban migration and having rural origins on diets of migrants points to the pattern of migrant households adopting urban diets when moving into urban areas. This has greatest consequences for the consumption of rice and food away from home.

Keywords: Dietary Intake, Food Consumption Patterns, Food System, Nutritional Transition, Rural-Urban Migration, Urbanization, VHLSS, Viet Nam.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

10YFP	10-Year Framework of Programmes on Sustainable Consumption and Production
A4NH	Agriculture for Nutrition and Health
AES	Adult Equivalent Scale
ATE	Average Treatment Effect
ATT	Average Treatment Effect on the Treated
BLUE	Best Linear Unbiased Estimator
CGIAR	Consultative Group for International Agricultural Research
CIA	Conditional Independence Assumption
CIAT	International Center for Tropical Agriculture
CPI	Consumer Price Index
DASH	Dietary Approaches to Stop Hypertension
EAR	Estimated Average Requirements
EUFIC	European Food Information Council
FAFH	Food Away From Home
FAO	Food and Agricultural Organization of the United Nations
FCS	Food Consumption Score
FSHD	Food Systems for Healthier Diets
FVS	Food Variety Score
GDP	Gross Domestic Product
GFDRR	Global Facility of Disaster Reduction and Recovery
Glopan	Global Panel on Agriculture and Food Systems for Nutrition
GSO	General Statistics Office of Viet Nam
HDDS	Household Dietary Diversity Score
HLPE	High Level Panel Experts on Food Security and Nutrition
IFPRI	International Food Policy Research Institute
ITA	Ignorable Treatment Assignment
Kcal	Kilocalorie
LSMS	Living Standards Measurement Study
NIN	National Institute of Nutrition
OECD	Organisation for Economic Cooperation and Development
OLS	Ordinary Least Squares
PSM	Propensity Score Matching
RDA	Recommended Dietary Allowances
RRD	Red River Delta
SFSP	Sustainable Food Systems Programme
SSB	Sugar Sweetened Beverage
UNEP	United Nations Environment Programme
UNFPA	United Nations Population Fund
VAM	Vulnerability Analysis and Mapping Branch

VHLSS	Viet Nam Household Living Standard Surveys
VND	Viet Nam Dong
WDI	World Development Indicators
WECR	Wageningen Economic Research
WFP	World Food Programme
WHO	World Health Organization
WB	World Bank
WUR	Wageningen University & Research

1. Introduction

In recent years, population and migration rates in Viet Nam have rapidly increased. These trends go together with Viet Nam's economic expansion towards a market economy (the *Doi Moi* economic reforms) (Anh et al., 2013; Anh & Sautier, 2011; General Statistics Office of Viet Nam [GSO] & United Nations Population Fund [UNFPA], 2016a and 2016b; International Food Policy Research Institute [IFRPI] & Compact2025 Initiative, 2016; Leaf, 2002; Nguyen et al., 2008; Nguyen and Winters, 2011; Quang, 2015). World Development Indicators (WDIs) show that the total population of Viet Nam increased from around 70.8 million in 1994 to 91.7 million in 2015. Furthermore, in 2015 13.6 percent of the population were internal migrants, which is a percentage that nearly doubled in just over 10 years – it equalled only 6.5 percent in 1994 (GSO & UNFPA, 2016b). Although still 66.4 percent of the population lived in rural areas in 2015, rural to urban migration accounted for the largest flow of internal migration (49.8 percent) (GSO & UNFPA, 2016b), which is a proportion that has increased over time - it equalled 27.1 percent in 1994, 31.4 percent in 1999, and 29.0 percent in 2009 (the percentage slightly dropped due to the impact of the 2008 economic crisis) (GSO & UNFPA, 2016a). The significance of the rural-urban migration flow in Viet Nam might be explained by the attractiveness of relatively well-paid employment and other opportunities in urban areas compared to rural areas, e.g. related to living standards or institutional services (Duc et al., 2012; GSO & UNFPA, 2016a and 2016b; Harris & Todaro, 1970; Loi, 2005; Nguyen & Winters, 2011; Tacoli et al., 2017). Other studies show as well that economic reasons remain the leading reasons for internal migration decisions (Nguyen et al., 2015; Phan & Coxhead, 2010).

Nguyen & Winters (2011) and Nesheim et al. (2015) argue that Viet Nam's recent economic growth has been accompanied by widespread changes in consumption patterns, as income per capita has increased and access to new goods has expanded (e.g. the number of supermarkets in the country increased from 386 in 2008 to 869 in 2016 and the number of street food and restaurants has also increased significantly (GSO, 2016; Moustier et al., 2003)). The National Institute of Nutrition of Viet Nam (NIN) (2010) argues as well that dietary intake has changed (and improved) as a consequence of significant economic growth and changes in socio-economic status. They argue that food consumption patterns in Viet Nam have moved gradually from a rice and cereal based diet to a more diverse and healthy bundle of foods, including greater quantities of meat, vegetables, and dairy products. This gradual shift in food consumption patterns can be seen as a nutritional transition, which in general marks a shift away from relatively monotonous and indigenous diets towards more varied industrialized ones. The latter ones are usually composed of more pre-processed food, more food of animal origin, and more added sugar and fat (Bansal et al., 2010; Bojorquez, 2014; Drewnowski & Popkin, 1997; Kearney, 2010; Popkin, 1997; Popkin et al., 2001; World Health Organization [WHO], 2003). Although several forces influence the nutritional transition (e.g. changes in size and age distribution of populations, climate change, income growth, competition for natural resources, and globalization of diets (Global Panel on Agriculture and Food Systems for Nutrition [Glopan], 2016)), urbanization is increasingly considered as being a crucial determinant of dietary patterns, and thus the nutritional transition (Cockx et al., 2017; Glopan, 2016; IFPRI, 2017; Kearney, 2010). Research on the nutritional transition often relates to the consequences for human health, e.g. due to the increasing prevalence of overweight (Cockx et al., 2017; Hawkes, 2006; Heald et al., 2005; Nguyen & Pham, 2008; Popkin et al., 2012; Renzaho & Burns, 2006; Torun et al., 2002). As urban areas are likely at a further stage of the nutritional transition than rural areas (e.g. due to faster changing diets) (Drewnowski & Popkin, 1997; IFPRI, 2017), the adding of an urban focus becomes increasingly important (Cockx et al., 2017) - especially given the fact that rapidly growing cities now start reshaping food systems by driving change and transformation in agricultural value chains (IFPRI, 2017).

Urban areas in Viet Nam have experienced more economic growth in recent decades than rural areas, and most dynamic economic sectors have been located in urban centres (Nguyen and

Winters, 2011). Therefore large differences exist between rural and urban areas environmentally wise, e.g. differences in available jobs and differences in wages. Furthermore, urban and rural areas might also differ in factors like prices, availability of food items, lifestyles, food consumption patterns, and the exposure to more global eating patterns (Cockx et al., 2017). This means that when people move from rural to urban areas they enter a new food environment, different from the old one in many aspects. As migrants are exposed to this new food environment, rural-urban migration might affect the diet of rural-urban migrants.

Migration and changing diets take place within Viet Nam's broader food system, which takes into account the full set of actors, resources, processes and activities that are related to the domains of food production, processing, distribution, consumption and the disposal of waste and the outputs of these activities (incl. nutrition and health, socioeconomic wellbeing and environmental quality outcomes) (Melesse et al., 2017). As migration potentially influences the food system, both on the supply and demand side (De Brauw, 2010; IFPRI and Compact2025 Initiative, 2016; Kearney, 2010), it is important to shed light on Viet Nam's food system and how it has evolved over time.

This brings us to the objective of this thesis, which is twofold. First of all it aims to explore trends in the local food system in Viet Nam, specifically focusing on food consumption patterns, agricultural production, and international trade. Secondly, it aims to explore whether rural-urban migration affects the diet of rural-urban migrants. I explore both the effect of the actual migration and the effect of having rural origins on dietary intake of rural-urban migrant households. Data in this thesis comes from the World Bank, Food and Agricultural Organization of the United Nations (FAO), and the Vietnam Household Living Standard Survey (VHLSS). This is complemented by literature in the field of food systems in Viet Nam.

This thesis and its objectives are part of a broader research on Food Systems for Healthier Diets (FSHD), which is one of the flagship projects of a special research program on Agriculture for Nutrition and Health (A4NH) by the Consultative Group for International Agricultural Research (CGIAR). The program is led by IFPRI, and aims to *"fill the existing gap between agricultural development and its unfulfilled health and nutritional benefits"* (CGIAR, n.d., c). The FSHD project, which is led by Wageningen University & Research (WUR), aims at understanding how the poor and vulnerable populations can be helped to consume healthier diets using a food system approach (IFPRI, 2017). This thesis contributes to the mentioned project by describing the context of one of its focal countries - Viet Nam, and by analysing the associated trends in the food system that can be seen over time. Furthermore, it potentially provides insights into the diet of an economically vulnerable group in society (Duc et al., 2012; Tacoli et al., 2014; UNFPA, 2007), which directly contributes to the objective of the FSHD project. The results could help with the design of interventions related to dietary intake, nutritional transition, and health effects.

The analysis of parallel trends points to the fact that over the years the number of consumed kilocalories per capita is increasing, where the share of calories from proteins and fat are increasing and the share of calories from carbohydrates is decreasing. Furthermore, production of nearly all food crops has increased over the years, where especially cereal (i.e. rice) production has been considerable. Large quantities of rice are produced for export. We see increasing trends in exports and imports for nearly all food groups. Results from the econometric analysis indicate that rural-urban migrant households adapt their diet to urban diets when they move to urban areas. Compared to rural non-migrant households, the food expenditure share of rice, other staples, meat, eggs, tofu and other is predicted to decrease when migrating, while the share of food away from home is predicted to increase. In terms of calorie intake shares, the share of rice and tofu is predicted to decrease whereas the share of food away from home is predicted to increase. Considering the largest significant effects, migrant households start consuming less rice while eating away from home more often when moving to urban areas. Furthermore, household's economic access to food (HDDS) and household's food variety (FVS) are predicted to increase as well. Compared to urban non-migrant households, we see no significant differences for any of the identified food consumption indicators.

The rest of this thesis is structured as follows: the next chapter contains the theoretical framework, explaining the main concepts at stake and the associated models. Next I discuss the data and methodology. The fourth chapter presents the main results of the analysis of parallel trends, whereas the fifth chapter presents the results from the analysis of the effect of rural-urban migration and having rural origins on diets of migrants. Chapter 6 discusses the limitations associated with the study, while chapter 7 concludes and raises questions for future research.

2. Theoretical Framework

This chapter serves to provide the theoretical framework of this thesis. By defining the main concepts and their corresponding models, it therefore serves as a foundation for the rest of the thesis. The chapter is structured as follows: the first section describes the household consumption model. This is followed by a section on dietary determinants and dietary change. The third section of this chapter is about the food system, and the fourth section identifies preliminary hypothesis on the effect of rural-urban migration and having rural origins on diets of migrants.

2.1 Household Consumption Model

This thesis aims to assess the diet of rural-urban migrants. This is done by analysing the consumption patterns of a rural-urban migrant household; the rural-urban migrant household is the unit of analysis. I follow standard neoclassical theory by assuming that each household has consumer preferences, which it tries to maximize. Or in more economic terms: every household tries to maximize its utility, given its budget and a set of prices (Varian, 2010).

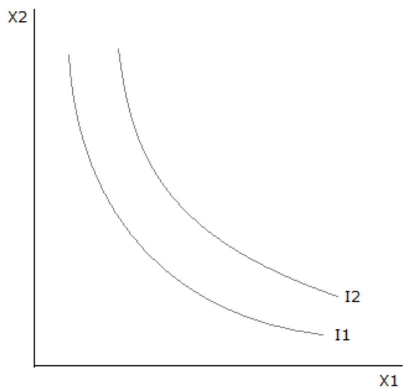
This can be formalized into some equations. Suppose we have an economy with only two food items, $x1$ and $x2$, with corresponding prices $p1$ and $p2$. If we define the amount of money that each household can spend on food as m, the budget constraint of the household can be written as:

$$p1x1 + p2x2 \leq m. \qquad (2.1)$$

Equation 2.1 shows us that each household can buy any combination of food items $x1$ and $x2$ as long as expenditures on those food items is no more than the total amount of money that it can spend on food. This leads us to the household's affordable consumption bundles, namely those combinations of food items that cost no more than its budget set m (Varian, 2010). It follows logically that equation 2.1 can be expanded to include as much food items as are available in the household's food environment.

However, some households have different preferences than others, based on e.g. taste, culture, religion, accessibility, and skills (see also next section); every household has its own consumer preferences. In microeconomic theory consumer preferences are described by indifference curves, which are curves through those consumption bundles to which the household has no preference (Varian, 2010). This can graphically be illustrated as following:

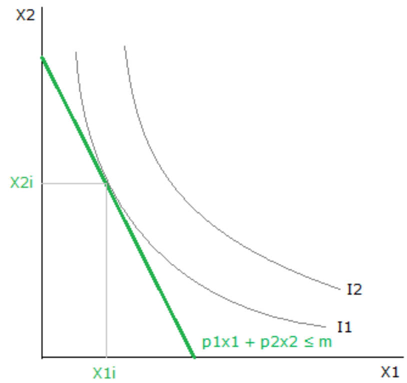
Figure 2.1: Indifference curves



Source: Figure is based on Varian (2010).

Figure 2.2 illustrates two indifference curves, $I1$ and $I2$, with each different combinations of food item $x1$ and $x2$. However, all combinations of $x1$ and $x2$ which lie on the same indifference curve are equally preferred to the household. If we combine the budget constraint defined in equation 2.1 with figure 2.1 we get the following:

Figure 2.2: Utility maximization



Source: Figure is based on Varian (2010).

Each household is assumed to maximize its utility, given current prices $p1$ and $p2$ and its budget m. Therefore the household will consume $x2i$ amount of food item $x2$ and $x1i$ amount of food item $x1$, as utility is maximized (i.e. consumption preferences are optimal) where the indifference curve is tangent to the budget constraint (Varian, 2010).

The household consumption model, which is based on the maximization of utility within the household, lies at the base of this thesis. It is assumed that each rural-urban migrant household tries to optimize its consumption preferences, given the set of prices and its budget. This allows us to derive demand functions (i.e. quantities demanded), which relate to optimal choice, different values of prices and incomes (Varian, 2010).

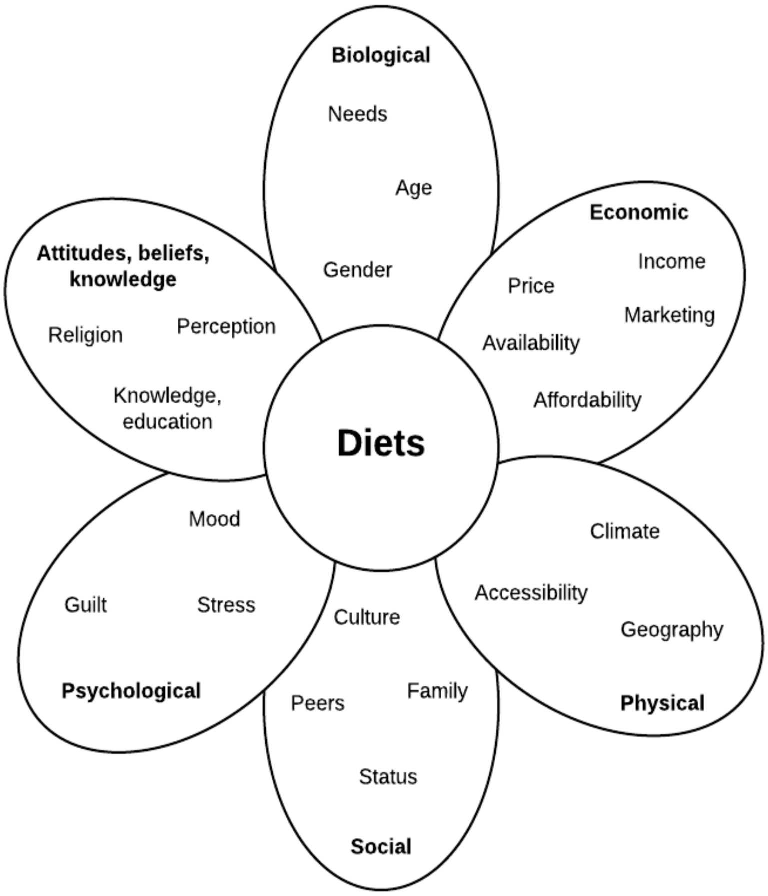
2.2 Determinants of Diets and Dietary Change

2.2.1 Dietary Determinants

The household consumption model assumes that food consumption is shaped by prices, income, and preferences. The determinants of food consumption can, however, be described in a more detailed way. The kinds of food that a household habitually eats can be defined as 'diet' (Oxford Dictionaries, 2017). According to the World Health Organization (WHO) (n.d., a), dietary habits are social - not just individual. Furthermore, individual dietary patterns are influenced by income, food prices (which both affect the availability and affordability of healthy foods), urbanization, globalization, marketing, individual preferences, attitudes and beliefs, cultural traditions, as well as geographical, demographical, environmental, social and economic factors (Cohen and Babey, 2012; Kearney, 2010; WHO, 2015). The National Institute of Nutrition of Viet Nam (NIN) (2010), which argues that production, processing, and distribution of foods, and habits, education, and technological achievements of the country and socio-economic conditions influence individual food consumption, confirms this for the case of Viet Nam. All these dietary determinants, which influence food availability, food accessibility, and food choice, are argued to interact with each other in a complex and interrelated manner (Kearney, 2010; WHO, 2015).

Johnston et al. (2014) have made an attempt to define and link the components of a sustainable diet. They came up with a model, based on the model of Lairon (2012), which illustrates the six interrelated key components of a sustainable diet (i.e. (1) well-being, health, (2) biodiversity, environment, climate, (3) equity, fair trade, (4) eco-friendly, local, seasonal foods, (5) cultural heritage, skills, and (6) food and nutrient needs, food security, accessibility). Each key component has its own factors and processes that influence and determine sustainable diets. Although in this thesis we are not merely interested in the components of a sustainable diet, the model proposed by Johnston et al. (2014) serves as a useful base for a model of dietary determinants. Figure 2.3 below illustrates this model of dietary determinants. We have defined six key determinants that constitute 'food choice', and thus diets. These determinants are based on the determinants of food choice established by the European Food Information Council (EUFIC) (2006). They are: (1) biological, (2) economic, (3) physical, (4) social, (5) psychological, and (6) attitudes, beliefs, and knowledge (an elaboration on each component follows below). Some of these determinants, thus, are determinants related to the environment (i.e. food system) that one lives in, whereas other determinants are merely personal. The determination of diet takes place within the food system, as some of the environmental food choice determinants are directly influenced by the food system and as diets are the outcome of consumption – one of the central food system activities (more on this in the next section). All key determinants, indicated by the 'petals' of the flower, are directly connected to the flower's 'heart' (i.e. diets). This direct connection together with the connection between the petals illustrate that each key determinant relates to and influences one another and the constitution of diets. The different factors and processes within each dietary determinant, which are based on EUFIC (2006), Kearney (2010), and WHO (2015), are listed on the corresponding petal. This illustrates the interdependence and influence that exists across the system. If one or more factors or processes change, this could influence other factors and processes within the same key determinant category and others. Unlike Johnston et al. (2014), we argue that the weight of each determinant varies according to life stage, personal and social characteristics, and the like (EUFIC, 2006).

Figure 2.3: Model of dietary determinants



Source: Adjusted from Johnston et al. (2014) and based on EUFIC (2006).

As can be seen from the figure, the first determinant of food choice is 'biological'. Related factors and processes that are depicted in the figure are gender, age, and needs. For example, it depends on gender and age what the recommendations regarding daily intake of kilocalories are. Often men are heavier in terms of weight than women, which is one of the reasons why they on average need to consume more kilocalories than women. The second determinant is 'economic'. Factors and processes within this component are related to income, price, marketing, availability, and affordability. An example is that richer families are able to afford different (and often higher quality) products relative to poorer families. Furthermore, the factor of availability implies that e.g. in urban areas 'global food products' are more widely available than in rural areas in Viet Nam, which likely influences the relative frequency of consuming these 'global food products' in both urban and rural areas. The third determinant is 'physical', which has the related factors and processes of climate, accessibility, and geography. The determinant therefore relates to the physical environment that one lives in. Food systems in Viet Nam are shaped by a favourable climate for cultivating rice (Anh & Sautier, 2011), inducing that rice is produced in large quantities and is an important component of the Vietnamese diet. Also, if one lives on the outskirt of the city center and markets are only accessible by travelling for at least 1.5 hour, this likely influences the frequency of making use of these markets. 'Social' is the fourth determinant. Related factors and processes are culture, family, peers, status, and meal patterns. EUFIC (2006) argues that the family is significant in food decisions, as the shaping of food choices takes place within the home. Furthermore, in Viet Nam different ethnic groups have different traditions and customs. This likely also influences dietary intake. The fifth determinant is 'psychological', and related factors and processes are mood, guilt, and stress. For example a person with an eating disorder might feel guilty for eating foods rich in calories and fat, which is likely to influence his or her food choices. Also, chronic stress and the consumption of foods rich in fat and sugar are often seen as being correlated (Warne & Dallman, 2007). The last determinant is 'attitudes, beliefs, and knowledge'. Related factors and processes are religion, perception, and knowledge and education. An example is that Muslims will try to avoid the consumption of pork meat in their diet, and that vegetarians avoid the consumption of meat at all. Summarizing, figure 2.3 shows that food choice determinants are widespread and interrelated. An illustrative example applied to the context of this thesis is as follows: Viet Nam has favourable ecological conditions for the production of rice, and therefore rice represents about 60 percent of cultivated areas (Anh & Sautier, 2011). Furthermore, rice has long been a dominant food staple in Viet Nam, and is deeply ingrained in the country's culture, traditions and economy (Anh & Sautier, 2011). Following, rice is an important component of the Vietnamese diet. Therefore, factors and processes of climate, geography, availability, and culture are interrelated in shaping the diet of Vietnamese people. In the case of this example the economic, physical, and social determinants are, thus, interrelated.

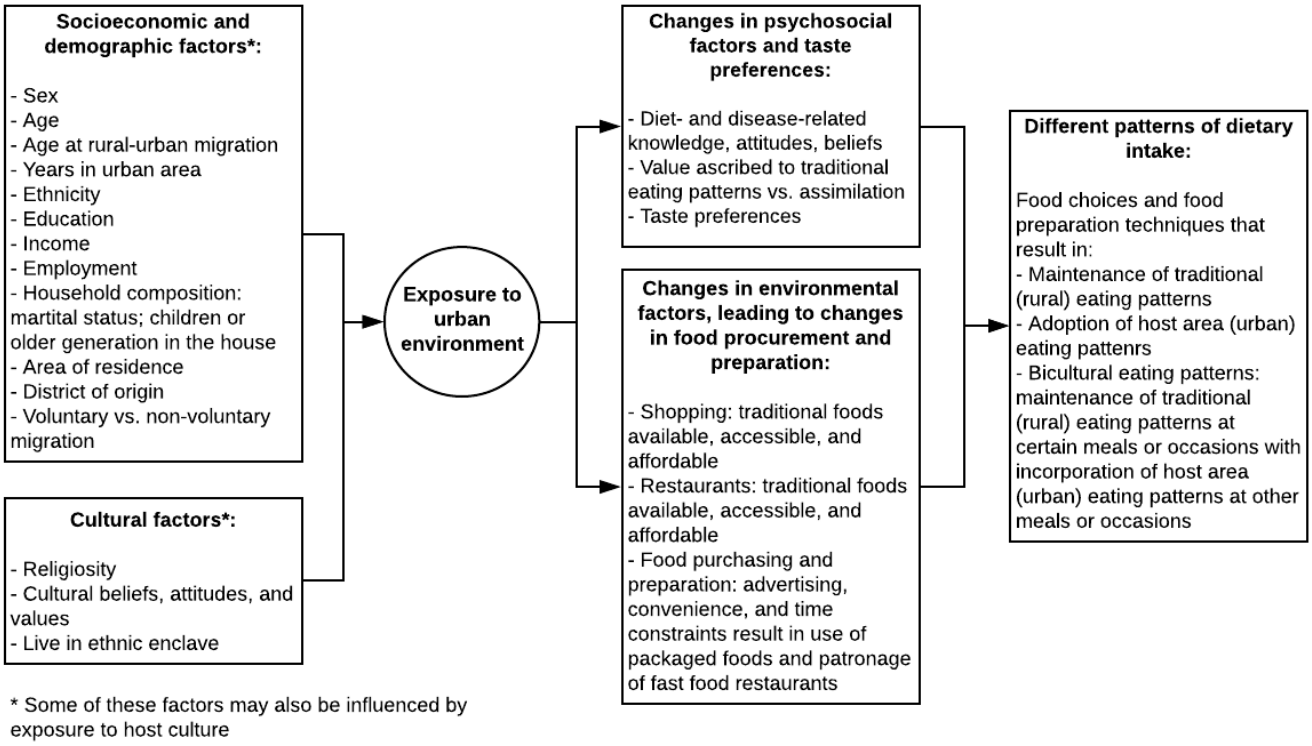
2.2.2 Dietary Change

Now that dietary determinants are defined, one can imagine diets of migrants to be influenced when moving to urban areas, as migrants enter into a new food environment with e.g. different food availability, accessibility, and affordability (Cockx et al., 2017; FAO, 2011). Or said differently, factors and processes within the six defined key dietary determinants might differ for both rural and urban areas. After entering a new sub regional food system, migrants might adopt the dietary practices predominant in urban areas, which is a concept known as 'dietary acculturation' (Satia-Abouta, 2003).

In general acculturation can be described as the process by which individuals or a group, usually a minority group, adopt the attitudes, values, customs, beliefs, and behaviors of a dominant or host group (Abraído-Lanza, 2006; Satia-Abouta, 2003). In this case this would imply rural-urban migrants adopting e.g. the cultural patterns of urban areas. The literature in general proposes two accepted theories that describe the concept of acculturation and its processes: (1) the Park model describes it as being linear and directional (Abraído-Lanza, 2006; Satia-Abouta, 2003), while (2) the Gordon Model takes a dynamic perspective with bidirectional movement between different stages of assimilation (Satia-Abouta, 2003). Both theories assume acculturation occurs both at the micro (individual) and macro (social/group) level (Satia-Abouta, 2003). Satia-Abouta (2003) argues that dietary acculturation can be seen as a reciprocal process, as, in the case of this thesis, the urban community might adopt some of the foods and dietary practices of rural-urban migrants.

In her article Satia-Abouta (2003) has made an attempt to define and model the multidimensional, dynamic, and complex concept of dietary acculturation, which is used in this thesis to theorize dietary changes related to rural-urban migration. This model, which is slightly adjusted in order to better model the context of this thesis, can be found in figure 2.4 below.

Figure 2.4: Model of dietary acculturation



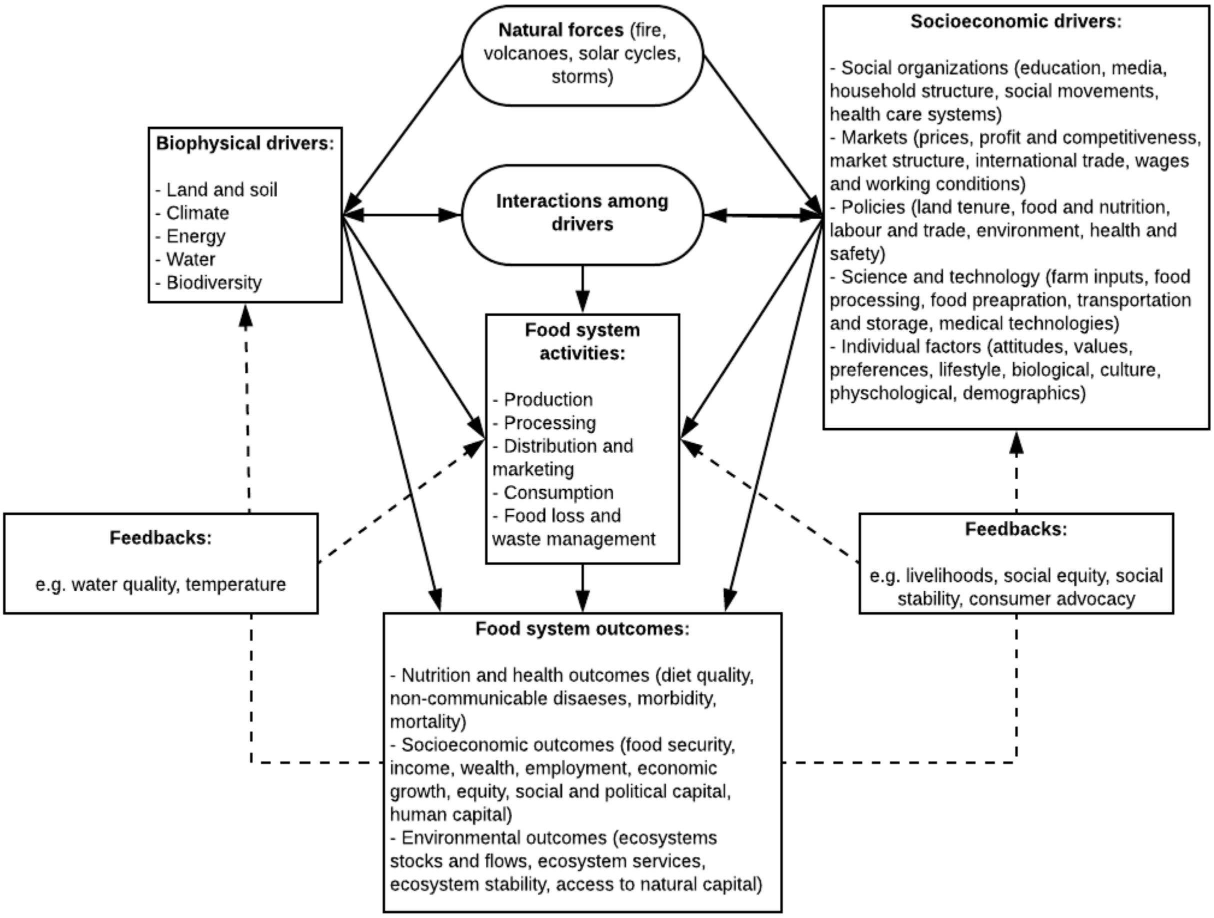
Source: Adjusted from Satia-Abouta (2003).

The model in figure 2.4 above shows the complex and dynamic relationship between socioeconomic, demographic, and cultural factors and exposure to a new environment - in this case the urban environment. The model mainly consists of four parts: (1) characteristics of migrants, both socioeconomic, demographic, and cultural, (2) exposure to the new environment in urban areas, (3) consequences of this exposure to diet-related psychosocial and environmental factors, taste preferences, food procurement and preparation, and (4) consequences of these changes for dietary intake. Exposure to the host environment can happen through for example television, advertisements, friendships, new food supply, different climate, and different prices (i.e. differences in the environmental dietary determinants, see previous subsection), and these different environmental or 'daily life' factors might trigger dietary acculturation (Satia-Abouta, 2003). All these factors are captured in the dietary determinants illustrated in figure 2.3 in the previous subsection. Each determinant, together with its corresponding factors and processes, contributes to the constitution of someone's diet. Therefore, exposure to the host environment in this model of dietary acculturation illustrated in figure 2.4 influences the key determinants in the model of diets illustrated in figure 2.3, which, in the end, may affect the diet of the rural-urban migrant. The model in figure 2.4 shows that the diet of rural-urban migrants may be affected in three different ways: (1) the migrant maintains his or her rural eating patterns, (2) the migrant fully adopts urban eating patterns, and (3) parts of urban eating patterns are added to the rural diet (i.e. biculturalism).

2.3 Food System

Dietary intake (i.e. consumption) takes place within the food system. In 2014, the High Level Panel Experts on Food Security and Nutrition (HLPE) defined the food system as: "*A food system gathers all the elements (environment, people, inputs, processes, infrastructures, institutions, etc.) and activities that relate to the production, processing, distribution, preparation and consumption of food and the outputs of these activities, including socio-economic and environmental outcomes*" (HLPE, 2014 p.29). This definition is used by the FAO-UNEP Sustainable Food Systems Programme (SFSP) of the 10-Year Framework of Programmes on Sustainable Consumption and Production (10YFP) (FAO, 2014), Ericksen (2008), and the Global Panel on Agriculture and Food Systems for Nutrition [Glopan] (2016). However, this definition does not explicitly capture the aspect of food loss and waste management. Therefore, Melesse et al. (2017) define the food system as: *"the full set of actors, resources, processes and activities that encompass the domains of food production, processing, distribution, consumption and the disposal of waste, and the outcomes of these activities, including nutrition and health, socioeconomic wellbeing and environmental quality, as well as the tradeoffs and synergies between the various outcomes* (Melesse et al., 2017 p.3). In this thesis I follow the latter definition. The associated conceptual framework is illustrated in figure 2.5 below.

Figure 2.5: Model of the food system



Source: Melesse et al. (2017).

The activities of the food system lie at the center of the food system, and include production (e.g. land use and tenure, soil management, crop management, and harvesting), processing (e.g. emulsification, pasteurization, and canning), distribution and marketing (e.g. transportation, storage, and packaging), consumption (e.g. preparing, cooking, household decision-making regarding food, and food choices), and food loss and waste management (e.g. redistribution, and prevention and reduction at source). These key activities lead to food system outcomes, which are nutrition and health, socioeconomic wellbeing, and environmental quality (Melesse et al., 2017). Furthermore, within (location-specific) dynamic and interactive food environments with multiple actors, biophysical, socioeconomic and natural factors shape food system activities and outcomes.

As can be seen from the definition and model of food systems above, food systems are multifaceted and complex, with complex interactions and tradeoffs and synergies between various system outcomes (Melesse et al., 2017). Therefore, for the sake of simplicity, I choose to analyse trends in only three activities of the local food system (i.e. the food system in Viet Nam), namely (1) food consumption patterns, (2) agricultural production, and (3) international trade. The first activity falls within the core aspect of consumption, the second activity corresponds with production, and the third activity corresponds with the core aspect of distribution. Food consumption directly relates to dietary intake, which is one of the key interests in this thesis. It, thus, follows that analysing trends in food consumption is important for describing the context of dietary intake and dietary change. Agricultural production is essential in order to provide foods available for consumption, which is essential in order to constitute a diet (Glopan, 2016; Thu & Dinh, 2011). Therefore, the food system cannot exist without production of foods, and the volume and variety of production influence the consumption standard of the population. Because of these reasons I focus on agricultural production when analysing the food system. The rationale behind focussing on international trade comes from first of all Viet Nam's large role in the export of products as diverse as rice, shrimp, coffee, cashews, and pepper (World Bank [WB], 2016a). Additionally, both import and export influence domesic supply and food available for human consumption (FAO, 2001), which in turn influence the availability of food items for consumers. Furthermore, trade in general plays an important role in the access to larger markets, which opens up opportunities for specialization and optimisation in production and economies of scale and scope (Busse & Königer, 2012; FAO, 2003; Krugman et al., 2014). Trade provides access to cheaper supplies and food imports, triggers competition at both the domestic and international market, and might foster the flow of technology and investment (Busse & Königer, 2012; FAO, 2003). Furthermore, as trade is expected to stimulate economic growth, it could also have a considerable impact on household food security (FAO, 2003).

All concepts and key activities described in this chapter take place or are shaped in the location-specific and multifaceted food system. By focussing on Food Systems for Healthier Diets (FSHD), the Consultative Group for International Agricultural Research (CGIAR) aims at understanding how the poor and vulnerable populations can be helped to consume healthier diets using a food system approach (IFPRI, 2017). This approach recognizes the full range of activities and outcomes of food systems, and aims to understand relationships, interactions, system dynamics and dependencies in food systems that define nutrition outcomes (Melesse et al., 2017). The multifacetity and complexity of the food system are central to this approach, and are recognized in this thesis as well.

2.4 Preliminary Hypotheses on Effect of Rural-Urban Migration and Having Rural Origins on Diet

In section 2.2 the 'model of dietary acculturation' was proposed. Rural-urban migrants have certain characteristics (both socioeconomic, demographic, and cultural), and when moving to urban areas they are exposed to the host culture and food environment of their new, urban, place of residence. This might trigger changes in psychosocial factors and taste preferences, as well as changes in food procurement and preparation. In the end there are three possible tracks which the diet of rural-urban migrants can follow: (1) fully maintain the rural eating patterns, (2) fully adopting urban eating patterns, and (3) partly adopting urban eating patterns, while still maintaining parts of the rural diet.

I expect that rural-urban migrant households partly adopt urban eating patterns when entering into the new urban food environment (i.e. diets become more urban). Furthermore, I expect that having rural origins compared to being born in urban areas leads to maintaining parts of the rural diet.

3. Data and Methodology

This chapter serves to describe the data and methods used in the analysis of this thesis. First the data sources and variable identification are described, which is followed by an elaboration on the methodological approach.

3.1 Data Sources and Variable Identification

3.1.1 Data Sources

In this study several data sources are used. The analysis of parallel trends in the food system is built upon amonst other things an elaborative literature review on research about food consumption patterns in Viet Nam. Important papers are Hoang (2009b), Nguyen & Winters (2011), Mishra & Ray (2009), Thang & Popkin (2004), and Trinh & Morais (2017). Data and results from these papers have been compared, analysed, and where possible aggregated. The literature is complemented by statistics of the World Development Indicators and the FAOSTAT database. The World Bank provides free and open access to global development data (i.e. the World Development Indicators). It provides aggregated data on e.g. population growth (both rural and urban), trends in Gross Domestic Product (GDP), and trade and exports of goods and services. Data comes from officially recognized international sources, and include national, regional and global estimates (WB, n.d., a). The Food and Agriculture Organization of the United Nations (FAO) provides free and open access to food and agricultural data. Examples of this data are the share of land devoted to agriculture, and trade indices. Furthermore, the FAOSTAT database provides detailed information on e.g. production, import and export, and domestic supply for a broad range of food products in its Food Balance Sheets (FAO, n.d.). The majority of the data in the Food Balance Sheets comes from direct enquiries or records, estimates of government agencies, marketing authorities and factories, farmer stock surveys, industrial and manufacturing surveys, and cost of production surveys (FAO, 2001). All these statistics together provide essential information on Viet Nam's food system.

The analysis of the effect of rural-urban migration and having rural origins on diets of migrants builds upon data from the Viet Nam Household Living Standard Surveys (VHLSS). The VHLSS is a rotating panel data set administered by the General Statistics Office of Viet Nam within the survey year. Data collection has been done in collaboration with the World Bank. The VHLSS is a comprehensive longitudinal national survey consisting of two main parts: a household survey and a commune-level survey. In the household survey information is gathered which reflects living standards, including income and expenditure, durables, housing and key household facilities, levels of education, employment, and involvement in poverty reduction programs (GSO, 2010). The latest available round of the VHLSS is used, which is the VHLSS of 2014. The dataset for this year consists of 9.399 households, from which 70.42 percent live in rural and 29.59 percent in urban areas, and is argued to be representative for Viet Nam (GSO, 2010). Furthermore I used corresponding kilocalorie intake and price data of Huong Thi Trinh, PhD candidate at the Toulouse School of Economics (TSE), who has been working with the VHLSS of 2004 to 2014 extensively.

As the econometric analysis of the effect of rural-urban migration and having rural origins on diets of migrants is based on the key concepts of migration and food consumption, rural-urban migrant households and food consumption indicators needed to be identified within the VHLSS dataset.

3.1.2 Rural-Urban Migrant Households

De Brauw & Carletto (2012) provide five individual characteristics that can help determining which individuals should be considered migrants in household survey data. These are: (1) place of birth, (2) whether or not the individual resides in the place of birth, (3) household membership, (4) the duration of any stays away from the residence, and (5) a time period of reference. However, as the name of this subsection predicts in this thesis we need to define 'migrant households' (i.e. households that have migrated as one unit) instead of just 'individual migrants', as the VHLSS provides food consumption data only at the household level (more on this in the next subsection). This induces that rural-urban migrants need to be identified in a somewhat more implicit way, as the VHLSS is not too elaborate on questions related to migration.

The VHLSS does provide information on province of birth of each household member. First of all, only those households were selected from which the current place of residence is urban[1]. Within this sample I identified the province of birth of the head of the household, and the province of birth of the spouse. A household was defined as migrant household when both the head and the spouse had a different province of birth than the current province of residence; if the household had no spouse only the province of birth of the head of the household was considered. Unfortunately data on place of birth is available only at the provincial level[2], indicating that migrant households moving within province cannot be identified. The choice is made for focussing on only the head of the household and the spouse as it is assumed that they are the ones who are responsible for migration and food consumption expenditures decisions. Furthermore, by focussing on the head and the spouse only, the identified definition of migration controls for (newly born) children. Next only those households were selected that moved from rural to urban areas, as we are interested in rural-urban migration only. I used a list by the World Bank and Cities Alliances (2011) of provinces and their corresponding classification status in order to identify whether the province of birth of the head and the spouse was rural or urban. The Minister of Construction (2009) identified six categories of urban centres (i.e. special category, type 1, type 2, type 3, type 4, and type 5), which were based on six criteria: (1) urban function, (2) number of urban inhabitants, (3) urban population density, (4) the percentage of non-agricultural labour, (5) the system of urban infrastructure works, and (6) urban architecture and landscape. Household were assumed to come from urban areas when they moved out of a province defined as 'special category' or 'type 1'; all other households were assumed to come from rural areas. There are some limitations to this definition as first of all provinces are heterogeneous and therefore an aggregated definition of 'rural' and 'urban' might not hold for every household. Furthermore, at time of migration the classification status of the province might have been different than the classification status in the list published in 2011. However, data on where the head and spouse come from is only available at the provincial level, indicating that assumptions need to be made. By using this definition of migration, 521 migrant households could be identified. From these households, 194 have a head and spouse that were born in the same province, 176 have a head and spouse that were born in a different province, and 151 have a head and no spouse. There are 44 migrant households that consist only of one person, i.e. the head of the household. As stated in the introduction, the GSO and United Nations Population Fund (UNFPA) (2016b) argue that in 2015 49.8 percent of all internal migrants were rural-urban migrants. Given that 13.6 percent of the population were internal migrants, 6.77 percent of the population were, thus, rural-urban migrants. In our sample 5.54 percent of all households are defined as rural-urban migrant households. The different percentages show that our sample slightly underidentified rural-urban

[1] The VHLSS includes a variable about whether the place of residence is rural or urban. However, although this variable provides us with some very important information, the definition of either rural or urban areas is not as clear cut, with e.g. peri-urban areas of intermediate density lying between rural and urban areas. This is of special importance for this thesis, given that most rural migrants tend to concentrate in the outskirts of large cities (De Brauw & Carletto, 2012). The definition of urban areas varies from country to country (sometimes even within country), and can be based on administrative boundaries, size, level of services, or population density (World Bank, 2008). The GSO, which has administered the VHLSS of 2014, defined a household as 'rural' when it resided in a commune. A household was defined as 'urban' when it resided in either a ward or a township.

[2] Viet Nam is divided into eight regions (i.e. (1) Southeast, (2) Red River Delta, (3) Mekong River Delta, (4) Northeast, (5) Northwest, (6) North Central Coast, (7) South Central Coast, and (8) Central Highlands), which can again be subdivided into 58 provinces (*tỉnh*) and five municipalities (*thành phố trực thuộc trung ương*). These five municipalities are administratively on the same level as provinces, and these units together construct the first tier of Viet Nam's administrative system. On the second tier are the urban district (*quận*), the rural district (*huyện*), and the district-level town (*thị xã*). The third tier consists of wards (*phường*), communes (*xã*), and townships (*thị trấn*) (Cities Alliance, 2011).

migrant households, although the percentages are close enough to each other for arguing that the difference is neglectable.

Unfortunately there is no variable in the VHLSS with information about when the head and spouse moved to the current place of residence. I made an attempt to approximate the timing of migration by looking at the age of the children and their province of birth, but in the whole sample no household had children younger than 16 years old which were born in a different province than the current province of residence. This induces that we do not know how long the migrant household has been living in the current province of residence. This is a limitation of our definition of migration, as no distinction could be made between short-term and long-term migrant households[3]. Nevertheless, one could argue that there is a high likelihood that most of the identified migrant households in our sample are long-term migrant households. This as we look at migrant *households* (with an average household size of 3.65 persons) instead of *individual* migrants. Moving with more persons (i.e. a household) takes more time and effort, induces more administrative difficulties, and, moreover, is more expensive. Therefore I assume that the decision to migrate is less flexible for a household than for an individual, especially when the household has children. This reduces the likelihood of the identified migrant households being short-term migrants. Furthermore, as the sampling frame of the VHLSS 2014 is based on the enumeration area of the 2009 Housing and Population Census, which is again based on official residence records (specifically the granting of official KT1 and KT2 registration status, i.e. permanent registration status), this provides additional support for the statement that it is not likely that the identified migrant households fall into the type of short-term, seasonal, return, and circular migration.

3.1.3 Food Consumption Indicators

Having defined rural-urban migrant households, we turn to the identification of food consumption indicators. These indicators should proxy for the dietary intake of households, which is the key outcome interest in the analysis. The VHLSS includes a section on expenditures at the household level, where amongst other things expenditures are reported for food and drinks. These expenditures, which have a recall period of 30 days, can be subdivided into expenditures on festive occasions and recurrent expenditures. Following Trinh & Morais (2017), only recurrent expenditures on food and drinks are used for constructing outcome variables; expenditures on festive occasions are excluded. This is done in order to give a more reliable picture of an average diet (FAO, 2011b). Furthermore, top 1 percent and bottom 1 percent households by total recurrent food consumption expenditures and total consumed kilocalories are deleted from the sample. This should control for major outliers.

The VHLSS reports the total value (in 1000 VND) and quantity (in kilograms or litres) for each food item (i.e. the sum of consumption of a certain food item). Out of 53 food items I have identified eleven food groups, which are (1) rice, (2) other staples, (3) meat, (4) fish and seafood, (5) eggs, (6) tofu, (7) vegetables, (8) fruits, (9) dairy, (10) food away from home (FAFH), and (11) other. The latter category of 'other' includes (alcoholic) beverages, coffee and tea, sugar and confectionery, salt, additives and seasonings, oils/lards, fish sauce, peanuts, betel leaves, and processed foods. A detailed list of which food item is included in which food group can be found in

[3] The survey does include a question about how long a household has been living in the current province or city for those households which have registration status elsewhere than the current province of residence or which have no registration status at all. This allows us to identify migrant households based on having registration status elsewhere, including making the distinction between short-term and long-term migrant households. However, by implementing this definition only 94 migrant households could be identified, from which 91 households have been living in the current province or city for 12 months or more, and 3 households have been living in the current province for less than or equal to 6 months. The former group falls into the definition of long-term migration (i.e. 12 months or more), while the latter group does not fall into the definition of short-term migration (i.e. between 6 and 12 months) nor in the definition of long-term migration. Furthermore, this definition would impose missing out on the households that did register when they have migrated to a new place of residence and the households having no registration status at all. Because of these reasons and as the sample size of 91 long-term migrant households imposes difficulties for the power of the analysis, I have chosen to stick with the definition of migrant household based on place of birth of the head and the spouse.

Appendix A. The choice for these food groups is based on food groups defined by other papers in the field of food consumption patterns in Viet Nam (i.e. GSO (2012), Hoang (2009b), Mishra & Ray (2009), Nguyen & Winters (2011), and Thang & Popkin (2004)), trends seen in food consumption patterns in Viet Nam over the last years (see next chapter), and the trade-off between the significance of each food group and limiting the total number of food groups. One interesting food group is the category of 'food away from home'. This category includes all food and drinks consumed by household members outside the home, including food and drinks of household members working, studying, or having health treatments away from home. It should be noted, however, that this category is an aggregate of all expenditures outside the home, and therefore does not give any information about the actual consumed types of food and drinks. Furthermore, as a consequence of the FAFH category being constituted of different types of food and drinks, quantity information on FAFH is not reported in the VHLSS. Values, on the contrary, are reported.

Several variables of interest are identified. First of all food expenditure shares ($EXPij$) are calculated, where j (= 1, .., 11) corresponds to the eleven identified food groups ((1) rice, (2) other staples, (3) meat, (4) fish and seafood, (5) eggs, (6) tofu, (7) vegetables, (8) fruits, (9) dairy, (10) food away from home, and (11) other) and i represents the individual households. These shares follow from the household consumption model in section 2.1. In order to generate these variables, an aggregate is created which contains the sum of the total value of all consumed food groups. This aggregate, which thus also includes the value of foods consumed which are self-produced or given as a present, is used as a base to create food consumption expenditure shares for each food group. Food consumption expenditure shares are, therefore, based on the total value of consumption of each food group.

Similar to food expenditure shares, food kilocalorie intake shares ($KCALij$) are calculated, where j (= 1, .., 11) corresponds to the eleven identified food groups ((1) rice, (2) other staples, (3) meat, (4) fish and seafood, (5) eggs, (6) tofu, (7) vegetables, (8) fruits, (9) dairy, (10) food away from home, and (11) other) and i represents the individual households. For the calculation of these shares I used kilocalorie intake data of Huong Thi Trinh, PhD candidate at Toulouse School of Economics (TSE). She has converted consumed quantities of each food item into consumed kilocalories by using the SMILING D3.5-a food composition table for Viet Nam (2013) established by the National Institute of Nutrition (NIN) and Wageningen University. Difficulties arose for the aggregated categories, like 'other vegetables' and 'food away from home', as first of all aggregated categories are not included in the food composition table, and secondly households did not report the consumed quantity of aggregated categories. However, Trinh was able to calculate the associated consumed kilocalories of these food items by using the reported value information of each item (see Trinh (2017) for more extensive calculations). The kilocalories of each food item were summed into the kilocalories of each identified food group. I created an aggregate that contained the sum of the total consumed kilocalories of all consumed food groups. This aggregate is used as a base to create food kilocalorie intake shares for each food group. Food kilocalorie intake shares are, therefore, based on the total consumed kilocalories of each food group[4].

Furthermore, food consumption indicators related to diversity of the diet were created. The Household Dietary Diversity Score (HDDS) is an indicator for the economic ability of the household to access a variety of foods (FAO, 2011b). The indicator is based on twelve food groups proposed by FANTA (Swindale & Bilinisky, 2006), which are (1) cereals, (2) white tubers and roots, (3) vegetables, (4) fruits, (5) meat, (6) eggs, (7) fish and other seafood, (8) legumes, nuts and seeds, (9) milk and milk products, (10) oils and fats, (11) sweets, and (12) spices, condiments and beverages (note: these food groups are different than the ones identified for the calculation of food expenditure and calorie intake shares). As the main purpose of this score is to provide an indication of economic access to food, groups like sweets and spices are included in the score as well, although their nutritional contribution to the diet may not be so important (FAO, 2011b).

[4] Bias could arise due to the fact that food kilocalorie intake shares are based on unprepared quantities; i.e. conversions into kilocalories are *not* based on edible quantities but on the reported quantities.

Interest lies in whether or not a food group is consumed; not in how much or how often. The HDDS basically sums the number of categories consumed, and therefore ranges from 0 to 12:

$$HDDSij = \sum FGij, \quad j = 1, \dots, 12 \qquad (3.1)$$

where $FGij$ represents a dummy variable for food group j for each household i. Normally the HDDS is based on 24-hour recall data (FAO, 2011b). However, given the fact that the VHLSS has a 30 day recall the scores will be inflated, as obviously more food items and therefore more food groups will be consumed the longer the recall period. Yet, the score can be used to compare economic access across household types in Viet Nam.

Another indicator of dietary diversity is created, namely the Food Variety Score (FVS). As is illustrated below this indicator is similar to the HDDS in construction, however it is not based food groups but on food items. The FVS assesses the variety of the diet, by summing the number of food items consumed (i.e. dummy variables for whether or not a food item is consumed) (Hodgson et al. (2010). The VHLSS provides 53 food items, and therefore the FVS ranges from 0 to 53:

$$FVSik = \sum FIik, \quad k = 1, \dots, 53 \qquad (3.2)$$

where $FIij$ represents a dummy variable for food item k for each household i. Normally the FVS is based on 7-day recall data. However, given the fact that the VHLSS has a 30-day recall the scores will be inflated, as obviously more food items will be consumed the longer the recall period. Yet, the score can be used to compare food variety across household types in Viet Nam.

Fruits and vegetables are important sources of vitamins and are recommended to be consumed on a daily basis (NIN, n.d., a). Because of its importance for a diet which improves overall health and reduces the risk of certain non-communicable diseases (Siegel et al., 2014; World Health Organization [WHO], n.d., b), I have created an indicator (FVi) for whether or not a household meets the WHO recommendations for fruits and vegetables consumption (i.e. more than 400 grams of fruits and vegetables per adult per day, WHO (2015)). In order to create this dummy variable I first of all aggregated the reported consumed quantities of fruits and vegetables[5], and converted this quantity into grams. This value was divided by 30 to represent average consumed grams of fruits and vegetables per household per day. However, the WHO recommendations are based on adults, which is why I identified an Adult Equivalent Scale (AES) for each household. The adult equivalent scale takes into account differences in age and economies of scale in consumption (Deaton, 2002; Khandker et al., 2010; Organisation for Economic Cooperation and Development [OECD], n.d.; White & Masset, 2002), by correcting for the size of the household and the age of its members (i.e. adult or child). Coates et al. (2017) argue that energy based adult male equivalence estimates can produce useful proxies of average intake for certain population subgroups. The OECD proposed a scale that assigns a value of 1 to the first household member, of 0.7 to each additional adult and of 0.5 to each child (OECD, n.d.; Khandker et al., 2010). However, as we are interested in the consumed *grams* of one adult the correction for economies of scale is not needed (i.e. one additional adult still needs to consume at least 400 grams of fruits and vegetables per day). Furthermore, given the facts that children in general consume less than adults and that consumption recommendations for children are similar to that for adults (WHO, 2015), I used the following formula to calculate the AES:

$$AESi = 1 * number\ of\ adults + 0.8 * number\ of\ children, \qquad (3.3)$$

[5] Households reported quantities for the consumption of beans, peas, morning glory, kohlrabi, cabbage, tomato, orange, banana, and mango. Unfortunately households did not report their consumed quantities for "other vegetables (gourd, winter melon, cucumber, cabbage, squash, ..)" and "other fruits (rambutan melon, papaya, guava, litchi, grapes, ..)". Therefore figures are deflated, as the estimations are based on only those fruits and vegetables that have quantity information.

where i represents the individual households. The value of 0.8 corrects for the fact that children in general consume less than adults, but recognizes that their dietary recommendations are similar. The variable containing the consumed grams of fruits and vegetables per household per day was divided by the household's AES to represent the adult equivalent consumed grams of fruits and vegetables per day. This value was compared to the WHO's recommendations of consuming at least 400 grams of fruits and vegetables per day. The dummy variable FVi was equal to 1 if the adult equivalent fruit and vegetable consumption was at least 400 grams; 0 otherwise[6].

Other food consumption indicators have been explored as well, but due to data limitations it was not possible to include them into the analysis. A list of these indicators is included in Appendix A.

3.2 Methodology

Overall, the methodology of this thesis has mainly been quantitative. The analysis of parallel trends mainly consists of a descriptive analysis, comparing different trends in both literature and statistics. Trends are represented by national averages from different data sources (see subsection 3.1.1). These national averages are compared over the years, for both rural and urban areas where possible (e.g. in the case of food consumption patterns).

The methodology for the analysis of the effect of rural-urban migration and having rural origins on diet of migrants is somewhat more complicated. However, the approach comes down to comparing rural-urban migrant households to comparable rural non-migrant households for analysing the effect of rural-urban migration, and to comparing rural-urban migrant households to comparable urban non-migrant households for analysing the effect of having rural origins. More details on the methodological design follows in the next subsections.

3.2.1 Endogeneity

An important difficulty associated with the analysis of migration in household surveys is related to endogeneity. This as the decision to migrate from rural to urban areas is not made separately from other decisions within households, and may be influenced by either observable or unobservable characteristics that explain household food consumption expenditures as well (De Brauw & Carletto, 2012). Examples of observable characteristics are the age and level of education of the household head. Examples of unobservable characteristics are level of risk aversion, ambition, attitudes and perception. As these factors are likely to influence both food consumption expenditures and the decision to migrate, migration should be considered endogenous. If one includes migration as an explanatory variable in a regression but neglects the endogeneity of migration, coefficient estimates on migration will be biased. A possible way this bias could occur is via selection bias; households who decide to migrate might be inherently different than households who decide not to, and therefore migration cannot be seen as random (De Brauw & Carletto, 2012). This indicates that migrant households are not directly comparable to non-migrant households, as they are 'selected' into migration (McKenzie, 2012; McKenzie & Yang, 2010). Another way to look at the occurrence of endogeneity bias is omitted variable bias; migrant households may have characteristics that are unobservable to the researcher, but that influenced both their decision to migrate to urban areas and the outcome variables of interest (i.e. food consumption indicators). If these unobservables, or observables that are omitted from the model, differ systematically between migrant households and non-migrant households, this induces the ordinary least squares (OLS) estimates of the impact of migration on food consumption expenditures to be biased (De Brauw & Carletto, 2012). Because of these endogeneity problems concerned with migration, a simple Ordinary Least Squares (OLS) regression cannot be run. A different approach has to be taken instead to deal with the endogeneity of migration.

[6] Calculations have been done using per capita consumption as well. However, given that in general children are assumed to consume less grams of fruits and vegetables per day than adults I argue that using the AES is more accurate.

De Brauw & Carletto (2012) propose four strategies of dealing with endogeneity of migration. In this thesis the third strategy is used, namely the use of matching methods. Propensity Score Matching (PSM) is an econometric approach that enables the matching of migrant households to non-migrant households on observable characteristics. In this way 'treatment' observations (= rural-urban migrant households) are matched with 'control' observations (= non-migrant households) that have similar values of explanatory variables, while it is assumed that both types of observations have equal unobservables. In order to prevent the PSM estimator from being biased, all variables should be included in the matching that might affect both rural-urban migration and food consumption expenditures. An elaborate description of the PSM model is presented in subsection 3.2.2.

Other strategies have also been explored. De Brauw & Carletto (2012) mention the Instrumental Variable (IV) approach as possible strategy to deal with endogeneity. In the IV literature on migration examples of proposed IVs are 'social network of the migrant' (Acosta, 2006; Tu et al., 2008; De Brauw & Carletto, 2012) and 'distance of the household to some fixed point' (De Brauw & Carletto, 2012). However, as the definition of migrant household used in this thesis is based on province of birth, these proposed IVs do not work. Furthermore, as IVs should both be uncorrelated with the error term and highly correlated with migration (Deaton, 1997; Verbeek, 2012), other appropriate IVs are difficult to find. Given that invalid instruments induce inconsistent estimators, and that weak instruments induce large variance and inconsistency as well (Verbeek, 2012), I argue that in the case of this thesis instrumental variables do not provide a good solution to deal with the endogeneity of migration.

As described in the previous section, the VHLSS is a rotating panel data set (i.e. fifty percent rotation of households from implementation of the survey to the next (Phung & Nguyen, 2004)). Since 2002 every other year GSO together with the World Bank gathers the household data. There are several advantages associated with the use of panel data, the first one being that it provides us with baseline data (e.g. demographics like employment, income and expenditures, and dietary intake before migration). Furthermore, panel data allows for controlling for household fixed effects, as it captures any observables and/or unobservables about the household that do not change over time (De Brauw & Carletto, 2012). De Brauw & Carletto (2012) argue that panel data can be particularly effective in controlling for endogeneity bias when it is combined with either matching or instrumental variables. However, migrant households are only traced within the district. This induces that migrant households moving to some place outside the district are no longer part of the panel, which removes the possibility of doing differences-in-differences analysis with fixed effects for those households. Also Deaton (1997) argues that it is not usually possible to match individuals from one census to another. Nevertheless, possibilities have been explored to compare the place of residence in 2012 to the place of residence in 2014 for those households that did participate in the VHLSS of 2012. However, only 8 households could be identified that were surveyed while living in a different place of residence in 2012 than in 2014. Because of these reasons I argue that unfortunately we cannot make use of the favourable characteristics of panel data in this thesis.

Another possibility to deal with the endogeneity of migration it using the Heckman two-step procedure ('Heckit') (Verbeek, 2012). This procedure basically means that you estimate selectivity into migration in a non-linear probit model in the first step. From this you calculate the inverse Mills ratio, which you include in the OLS model of interest in the second step (Verbeek, 2012). Bushway et al. (2007) evaluate the Heckman technique as an approach for dealing with selection bias in criminology. They argue, however, that the Heckman correction is sensitive to distributional assumptions (i.e. normality in the probit model) and furthermore that the Heckman approach can inflate standard errors due to collinearity between the correction term (i.e. the inverse Mills ratio) and the included regressors, especially when exclusion restrictions are not present (Anderson, 2017; Bushway et al., 2007). Exclusion restrictions are similar to instrumental variables in the sense that they are variables that affect, in the case of this thesis, migration but not dietary intake. One favourable characteristic of the Heckman approach is that it does not require exclusion restrictions to be estimated (Bushway et al., 2007); i.e. the same covariates can be included both in the selection model and the OLS model. However, it is argued that using Heckman's two-step procedure to deal with selection bias can lead to highly unstable estimates if used without exclusion restrictions (Bushway et al., 2007; Stolzenberg & Relles, 1990). I already argued in one of the previous paragraphs that in the case of this thesis it is difficult to come up with good instrumental variables; those that satisfy both criteria of no correlation with the error term and high correlation with migration. When no or only weak instruments are available, no or only weak exclusion restrictions are available. As implementing the Heckman procedure regardless of these absent or weak exclusion restrictions will lead to inconsistent estimates (Anderson, 2017; Bushway et al., 2007), I argue that the Heckman approach does not provide the solution of dealing with the selectivity of migration.

The last option De Brauw & Carletto (2012) propose in dealing with endogeneity is to ignore the bias. They argue that if OLS estimates are largely consistent with estimates using other methods, one might choose to ignore the bias or argue that it is of minimal importance. As a robustness check to applying a different weighting, I choose to estimate the effect of migration on diets of migrants by simple OLS as well next to PSM, although estimates will possibly be biased. However, I use the OLS estimates to say something about the sign of the estimates only, rather than the magnitude of the coefficients. This is in line with the recommendations by De Brauw & Carletto (2012) regarding the ignorance of endogeneity bias.

Although propensity score matching also certainly has its limitations, which I turn to in the discussion of this thesis, I argue that given the limited data related to migration at hand it is the best possible approach to estimate the effect of rural-urban migration and having rural origins on diets of migrants.

3.2.2 Propensity Score Matching

The previous subsection described the endogeneity of migration and the choice for propensity score matching (PSM) as the main method to estimate the effect of rural-urban migration and the effect of having rural origins on diets of migrant households. This subsection, which is based on Caliendo & Kopeinig (2005), Ham et al. (2005), Rosenbaum & Rubin (1983), and Sparrow (2017), presents the PSM method, proposed by Rosenbaum & Rubin (1983), and how it can be implemented and used.

Define $Di = 1$ if a household migrates, and $Di = 0$ otherwise. The outcome variable for migrant households is, in turn, defined as $Y1i$ $(Di = 1)$ and the outcome variable for non-migrant households is defined as $Y0i$ $(Di = 0)$. Having identified these variables, the average treatment effect (ATE) can be defined as

$$ATE = E(\Delta i) = E(Y1i) - E(Y0i), \qquad (3.4)$$

where $E(.)$ represents the expectation in the population for each individual household i. The ATE, thus, measures the difference in expectation between outcome variables of migrant households and outcome variables of non-migrant households. Our goal, however, is to estimate the average treatment effect on the treated (ATT) (i.e. the effect of rural-urban migration and the effect of having rural origins on the outcome variables of migrant households). The ATT can be defined as

$$ATT = E(\Delta i | Di = 1) = E(Y i | Di = 1) - E(Y0i | Di = 1). \qquad (3.5)$$

We observe the first term on the right-hand side of equation 3.5. However, the second term on the right-hand side, the counterfactual, is not observed (i.e. the outcome variables of non-migrant households would they have migrated). Saying it more generally: although we would like to know the difference between a household's outcome variable with and without migration, we cannot observe both outcomes (with and without migration) for the same household at the same time. This is where matching comes in: matching with comparable non-migrant households is used to estimate $E(Y0i|\, Di = 1)$, which should work as a solution for the above mentioned fundamental evaluation problem of a missing counterfactual outcome (Caliendo & Kopeinig, 2005). In order for matching to be valid, certain assumptions must hold. The fundamental assumption of matching is the assumption of unconfounded assignment. This assumption is represented by

$$(Y1i, Y0i) \perp Di|\, Xi, \qquad (3.6a)$$

where Xi is a vector of variables (i.e. covariates) that are unaffected by the treatment. It is, thus, assumed that conditional on a set of observed characteristics, treatment (i.e. rural-urban migration) can be considered as random. This assumption is also known as ignorable treatment assignment (ITA). Given that we are estimating the ATT, the condition (3.6a) can be weakened to the following mean independence assumption involving only $Y0i$

$$E(Y0i \mid Xi, Di) = E(Y0i \mid Xi). \qquad (3.6b)$$

Another assumption should hold in order to identify the average treatment effect of the treatment, namely the assumption of common support. This assumption is defined as

$$0 < Pr(Di = 1 \mid Xi) < 1, \qquad (3.7)$$

which requires that at each level of Xi, the probability of either migrating or not migrating is positive. The assumption ensures that there is some overlap between the migrant group and the non-migrant group in the probability of migrating or not.

However, matching on all variables in Xi would become impractical as the number of variables increases. This is known as the curse of dimensionality, which can be overcome by using the propensity score introduced by Rosenbaum & Rubin (1983). The propensity score is a balancing score which measures the probability of being treated (i.e. the probability that a household has migrated), and summarizes Xi into one single indicator. The propensity score is defined as

$$P(Xi) = Pr(Di = 1 \mid Xi). \qquad (3.8)$$

By the use of the propensity score the dimensionality is reduced to only one, inducing matching to be done based on only one variable. Following Ham et al. (2005) we see that if condition 3.6a and 3.8 are satisfied, then

$$(Y0i, Y1i) \perp Di \mid P(Xi) \qquad (3.9a)$$

and

$$0 < Pr(Di = 1 \mid P(Xi)) < 1. \qquad (3.10)$$

If condition 3.9b and 3.10 are satisfied we have

$$E(Y0i \mid P(Xi), Di) = E(Y0i \mid P(Xi)) \qquad (3.9b)$$

and

$$0 < Pr(Di = 1 \mid P(Xi)) < 1. \qquad (3.10)$$

Equations 3.9b and 3.10 show that if the mean independence assumption holds given Xi, it also holds conditional on $P(Xi)$. This allows the usage of the propensity score distribution of households to weight the mean difference in outcomes over the common support.

One important condition that should hold in order to appropriately apply propensity score matching for estimating causal effects, is that both the treatment (migrant) and control (non-migrant) group should be balanced in their covariates Xi. This means that the distribution of Xi should be similar across both groups. If this is the case, the only difference between treatment and control group is participation into migration (given the ITA assumption), which indicates that we have identified a comparable control group which could proxy for the unobserved term of $E(Y0i \mid Di = 1)$ in equation 3.5. The balanced matched sample of comparable non-migrant households allows us to estimate the unbiased average treatment effect on the treated, namely the effect of rural-urban migration and the effect of having rural origins on the outcome variables of migrant households.

Propensity score matching can be implemented in STATA using the command **psmatch2** developed by Leuven and Sianesi (2003). The reported estimates of the standard errors of the ATT, however, do not account for the variance due to the estimation of the propensity score. Therefore I bootstrapped the results and its standard errors. Repeating the bootstrapping *N* times leads to *N* bootstrap samples and *N* estimated average treatment effects on the treated. The distribution of these means approximate the sampling distribution, and thus the standard error, of the population mean (Caliendo & Kopeinig, 2005; Deaton, 1997).

Comparisons have been made across different number of bootstrapping repetitions. Generally speaking, the higher the number of repetitions the better the precision of the inference. However, including a high number of repetitions becomes very time-consuming (Caliendo & Kopeinig, 2005), illustrating the trade-off involved in the choice for the number of bootstrap repetitions. Comparisons have been made between the results of 50, 200, 500 and 1000 bootstrap repetitions. As these were more or less similar for the different numbers of repetitions, the choice has been made to include 200 bootstrap repetitions in the main analysis. This number of repetitions should be sufficiently high to approximate the ATTs and their corresponding standard errors.

Propensity score estimation results can be found in Appendix C, together with the implemented matching algorithms and their associated balance tests.

3.2.3 Sensitivity Analysis – Rosenbaum Bounds

The propensity score matching estimation method assumes that people who look comparable are in fact comparable, as matching is done on observed covariates (Rosenbaum, 2010). However, people might differ in ways that we do not observe, which violates the assumption of equal unobservables in expectation. If migration is not random then there might be reasons that some households migrate and others not and that simultaneously influence dietary intake, but those reasons are not visible to us because the observed covariates provide an incomplete picture of the situation (Rosenbaum, 2010). This is also known as 'hidden bias' (Becker & Caliendo, 2007), which is bias to which matching estimators are not robust. Therefore it is important to do a sensitivity analysis on the estimated model, which allows us to speak about the reliability of the underlying assumptions of the model - or as Rosenbaum (2010, p.76) calls it *"speaking about the degree to which the naïve model is false".* He continues arguing that the sensitivity of an observational study, like the one in this thesis, to bias from unmeasured covariates is represented by the magnitude of the departure from the naïve model that would need to be present to alter inference about treatment effects. Although the assumption of unconfounded assignment is untestable (Becker & Caliendo, 2007), Rosenbaum (2010) proposes a way to calculate bounds which provide evidence on the degree to which any significance results hinge on the assumption of no hidden bias.

The following paragraphs are based on Aakvik (2001), Becker & Caliendo (2007), Caliendo & Kopeinig (2005), and Rosenbaum (2010). I follow the notation in subsection 3.3.2 by assuming that the propensity to migrate is given by:

$$P(Xi) = \Pr(Di = 1 \mid Xi) = F(\beta Xi + \gamma Ui), \qquad (3.11)$$

where $F(.)$ is the logistic distribution, Xi are the observed characteristics for individual i, Ui is the unobserved variable, and γ is the effect of Ui on the migration decision. If the study is free of hidden bias, γ will be zero and the probability to migrate is solely determined by Xi. However, if there is hidden bias, two households with the same observed covariates Xi might both have a different chance to migrate or to have migrated (Becker & Caliendo, 2007). As we assume that the propensity score estimation model has a logistic distribution (Rosenbaum, 2002), the odds that individual household i is a rural-urban migrant household can be written as:

$$(P(Xi) / (1 - P(Xi)) = exp(\beta Xi + \gamma Ui). \qquad (3.12)$$

Suppose we have two matched individual households i and j, then the odds ratio (i.e. relative odds) of migrating can then be written as:

$$(P(Xi) / (1 - P(Xi)) / (P(Xj) / (1 - P(Xj)) = (P(Xi)(1 - P(Xj)) / (P(Xj)(1 - P(Xi)) \qquad (3.13)$$
$$= (exp(\beta Xj + \gamma Uj)) / (exp(\beta Xi + \gamma Ui) = exp[\gamma(Ui - Uj)].$$

The matching procedure implies that both households have identical observed covariates, which is why the Xi vector is cancelled out. However, both households differ in their odds of migrating by a factor that involves the parameter γ and the difference in unobserved covariates γ. If unobserved variables are similar ($Ui = Uj$) or if unobserved variables have no influence on the decision to migrate ($\gamma = 0$) the odds ratio is one, implying that there is no hidden or unobserved selection bias (Becker & Caliendo, 2007). In this case, controlling for observed selection would produce unbiased estimates of the effect of migration and having rural origins on diets of migrants. However, in the sensitivity analysis we evaluate how inference about this effect is altered by changing the values of γ and ($Ui = Uj$).

For the sake of simplicity I follow Aakvik (2001), Becker & Caliendo (2007) and Caliendo & Kopeinig (2005) by assuming that the unobserved variable is a dummy variable with $Ui \epsilon \{0, 1\}$. An example would be whether the household is risk averse ($Ui = 1$) or not ($Ui = 0$). Equation 3.13 can be rewritten into:

$$1 / e^\wedge\gamma \leq (P(Xi)(1 - P(Xj)) / (P(Xj)(1 - P(Xi)) \leq e^\wedge\gamma . \qquad (3.14)$$

Equation 3.14 shows that there are bounds on the odds-ratio that either of the two matched households migrates compared to staying behind or being born in urban areas. Furthermore, the equation shows that both individuals have the same probability of migration only if $e^\wedge\gamma = 1$, which is assumed by the propensity score estimation model (i.e. no unobserved selection bias). If $e^\wedge\gamma = 2$, then two households who look comparable in terms of Xi may in fact differ in their odds of migrating by a factor of two. In this sense, $e^\wedge\gamma$ is a measure of the degree of departure from a study that is free of hidden bias (Becker & Caliendo, 2007; Caliendo & Kopeinig, 2005; Rosenbaum, 2010); if $e^\wedge\gamma$ close to 1 changes the inference about the effect of migration on diets of migrants, then estimated effects are said to be sensitive to unobserved selection bias. However, if a large value of $e^\wedge\gamma$ does not change conclusions of the study, the study is not sensitive to selection bias (Aakvik, 2001). In this way Rosenbaum bounds can be used to determine the end point of the significant test that leads one to accept the null hypothesis of zero effect of rural-urban migration on diets of migrants.

Diprete & Gangl (2004) proposed a way to implement Rosenbaum bounds for the average treatment effects on the treated in the presence of hidden bias in STATA (i.e. **rbounds**). This command takes the difference in the response variable between treatment and control cases as input variable *varname*, and then calculates Wilcoxon sign rank tests that give upper and lower bound estimates of significance levels at given levels of hidden bias. However, this command only works for one to one matching. As in this thesis I have used radius caliper matching for the main analysis (see Appendix C), Rosenbaum bounds could not be applied to the main results. However, ATTs have been estimated using one to one nearest neighbour matching as well, although the balancing exercise was less successful for this particular matching algorithm (see Appendix C). Chapter 5 presents the results produced by both PSM models, and what follows is that both of the models predict more or less similar significant ATTs. Furthermore, one to one nearest neighbour matching predicts significant effects less often than does radius caliper matching – it is more strict in this sense. Rosenbaum bounds have been implemented on the results from the one to one nearest neighbour matching, in order to assess the sensitivity of these results to selection bias. As the one to one nearest neighbour matching model is based on the same observable covariates as the radius caliper model, this should say something about the degree of hidden bias in the analysis in general.

Furthermore, a different procedure to implement Rosenbaum bounds in STATA is proposed by Becker & Caliendo (2007) (i.e. **mhbounds**). They calculate the average treatment effects on the treated in the presence of hidden bias by taking the response variable as input variable *newvar*. The procedure then calculates Mantel-Haenszel tests statistics that give bound estimates of significance levels at given levels of hidden bias under the assumption of either systematic over- or underestimation of treatment effects. The approach is suited for *kth* nearest neighbour matching without replacement, and for stratification matching. Furthermore, this command only works for both binary treatment and response variables. Given that all response variables in this thesis are continuous instead of binary (except for the dummy for whether or not the household meets the WHO's fruit and vegetable guidelines), I cannot use the **mhbounds** command for the implementation of Rosenbaum bounds.

Liu et al. (2013) point to the possibilities of calculating Rosenbaum bounds by hand, but they argue that assessing the sensitivity of studies beyond one to one matching is not easily implemented. Therefore I stick with the implementation of Rosenbaum bounds on one to one nearest neighbour matching, using the same covariates as in the radius caliper matching model[7].

---

[7] Ichino, Mealli and Nannicini (2006) proposed another sensitivity analysis for propensity score matching estimators. Nannicini (2007) presented the implementation of this analysis in STATA with the command **sensatt**, which simulates a potential confounder in order to assess the robustness of the estimates treatment effect with respect to specific deviations from the assumption of unconfounded assignment. However, this command makes use of the PSM estimator proposed by Becker and Ichino (2002) (i.e. **att***) instead of the one proposed by Leuven and Sianesi (2003) (i.e. **psmatch2**), which is used in this thesis. Therefore in this thesis I could not use the **sensatt** command for the implementation of a sensitivity analysis to unobserved confounding.

Other methods than 'Rosenbaum's Approaches' that focus on the statistical significance of the 'true' outcome-exposure association are mentioned in an article by Liu et al. (2013). They mention five studies which aim to obtain the point estimate of the 'true' outcome-exposure association with a 95 percent confidence interval (i.e. Greenland (1996), Harding (2003), Lin et al. (1998), Van der Welee and Arah (2011) and Arah, Chiba & Greenland (2008)) (Liu et al., 2013). The first two studies imagine the data that would be observed if the unobserved confounders were observed; the goal is to specify aspects of so that the data can be re-created. The studies differ, however, in the necessary parameter inputs and in the way the data is re-created. From the re-created data, a weighted logistic regressions is performed to obtain the observed odds ratio of the relationship between the outcome and the exposure from the data, adjusted for the observed confounders, including its confidence interval (Liu et al., 2013). The latter three studies have the same objective, but use a slightly different approach. For an elaborate overview of the proposed methods I refer the reader to Liu et al. (2013). Given that the proposed methods require the specification of quite some parameters and that they should at least partly be calculated manually, I argue that it is beyond the scope of this thesis to implement these proposed sensitivity analysis approaches.

Other methods than 'Rosenbaum's Approaches' that focus on the statistical significance of the 'true' outcome-exposure association are mentioned in an article by Liu et al. (2013). They mention five studies which aim to obtain the point estimate of the 'true' outcome-exposure association with a 95 percent confidence interval (i.e. Greenland (1996), Harding (2003), Lin et al. (1998), Van der Welee and Arah (2011) and Arah, Chiba & Greenland (2008)) (Liu et al., 2013). The first two studies imagine the data that would be observed if the unobserved confounders were observed; the goal is to specify aspects of so that the data can be re-created. The studies differ, however, in the necessary parameter inputs and in the way the data is re-created. From the re-created data, a weighted logistic regressions is performed to obtain the observed odds ratio of the relationship between the outcome and the exposure from the data, adjusted for the observed confounders, including its confidence interval (Liu et al., 2013). The latter three studies have the same objective, but use a slightly different approach. For an elaborate overview of the proposed methods I refer the reader to Liu et al. (2013). Given that the proposed methods require the specification of quite some parameters and that they should at least partly be calculated manually, I argue that it is beyond the scope of this thesis to implement these proposed sensitivity analysis approaches.

3.2.5 Robustness Check – Ordinary Least Squares

I estimate the effect of rural-urban migration and having rural origin on diets of migrants by Ordinary Least Squares (OLS) as well, as a robustness check on the estimated ATTs by Propensity Score Matching (PSM). The results of the OLS model are compared to the results from the PSM model, to assess whether both models yield similar (significant) signs of treatment effects. In this case I follow the suggestions of De Brauw & Carletto (2012) regarding the ignorance of endogeneity bias by only analysing the sign of the estimates rather than the magnitude of the coefficient. Before elaborating on the OLS model it should be noted, however, that in principal OLS and PSM estimation methods are similar to each other in the sense that both models build on the same assumption for causal inference (i.e. the conditional independence assumption). Angrist & Pischke (2008) point to the fact that the OLS estimator is a particular sort of weighted matching estimator; the only difference between OLS and PSM estimators is the weights used to sum the covariate-specific treatment effect into a single treatment effect. Where PSM puts greater weight on those households which are most *similar*, OLS puts greater weight on those households which are most *different* (Angrist & Pischke, 2008).

Similar as the notation in subsection 3.3.2, define $Di = 1$ if a household migrates, and $Di = 0$ otherwise, where i represents the individual household. The outcome variable is defined as Yi. A simple OLS model can then be defined as:

$$Yi = \beta 0 + \beta 1Di + \beta 2Xi + \varepsilon i, \qquad (3.15)$$

where Xi are the household-level covariates, and εi is the error term. The OLS model, which predicts its parameters while minimizing the sum of squared residuals, hinges on the Gauss-Markov Theorem, which is based on five assumptions namely: (1) the model is linear in the parameters, (2) the sample at hand is a random sample from the population which the model applies to, (3) there is no perfect multicollinearity, (4) the expectation of the error term conditional on the independent variables is zero, and (5) the error term is homoscedastic and not autocorrelated (Verbeek, 2012). If these assumptions hold, the OLS estimator is the Best Linear Unbiased Estimator (BLUE). However, as I already explained in subsection 3.2.1 migration is not random, and therefore there likely are unobserved factors (like risk aversion, attitude and perception) influencing both the decision to migrate and the food consumption indicators. This is a violation of both the second and the fourth Gauss-Markov assumption. Therefore I only use the sign of the possibly biased estimates for assessing the robustness of the PSM treatment effects to applying a different weighting.

The OLS model is estimated for all food consumption indicators Yi (i.e. food expenditure shares, calorie intake shares, HDDS, FVS, and FV (= a dummy for whether or not the household meets the WHO's recommendations for fruit and vegetable intake)), and for both the effect of rural-urban migration on diets of migrants and the effect of having rural origins on diets of migrants. Both models have been estimated with and without covariates. In the presence of covariates, the same covariates Xi are included as in the associated propensity score estimation models (see Appendix B), including the covariate of ethnicity. The OLS model has been estimated with inclusion of the option *vce(robust)* to correct the standard errors for heteroskedasticity. Furthermore, 200 bootstrap repetitions are included in the estimation as well. Estimation results are presented in chapter 5.

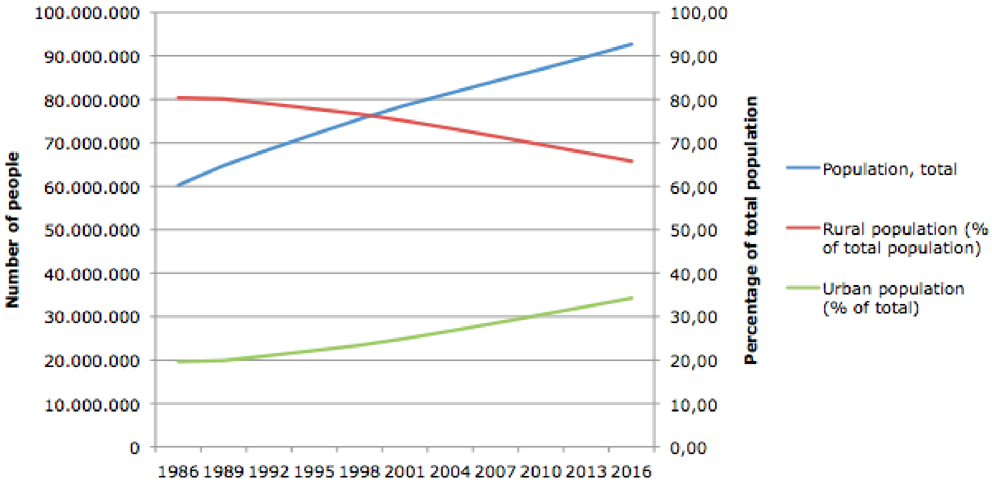
4. Analysis of Parallel Trends

This chapter presents the results of the first research objective and question, namely the analysis of parallel trends in the local food system. As stated in the theoretical framework, the food system can be seen as a multifaceted and complex concept. Therefore, I chose to focus on only food consumption patterns, agricultural production, and international trade for analysing trends in the local food system (i.e. the national food system of Viet Nam). However, in order to appropriately analyse these trends, it is useful to first elaborate on trends in demographics and general statistics of Viet Nam. This because these are socio economic drivers of the abovementioned trends (Global Panel on Agriculture and Food Systems for Nutrition [Glopan], 2016; Melesse et al., 2017), and they therefore shed light on the context in which these trends occur. The chapter is structered as follows: the first section contains trends in demographics and general statistis. This is followed by a section on trends in food consumption patterns. The third section is about trends in agricultural production and international trade, and the chapter finishes with a section combining and summarizing all the analysed trends.

4.1 Demographics and General Statistics

As stated in the introduction, Viet Nam's population rate has rapidly increased over the years. In just 30 years, Viet Nam's population has increased with over 30 million people (World Bank [WB], n.d., b), which is illustrated in figure 4.1 below. Furthermore, Viet Nam has become more and more urban, reaching around 35 percent of total population in 2016. Logically following from the increasing population rate, population density has been increasing over the years as well; it reached nearly 300 people per square kilometres of land area in 2016, while this number was only 185 in 1986 (WB, n.d., b).

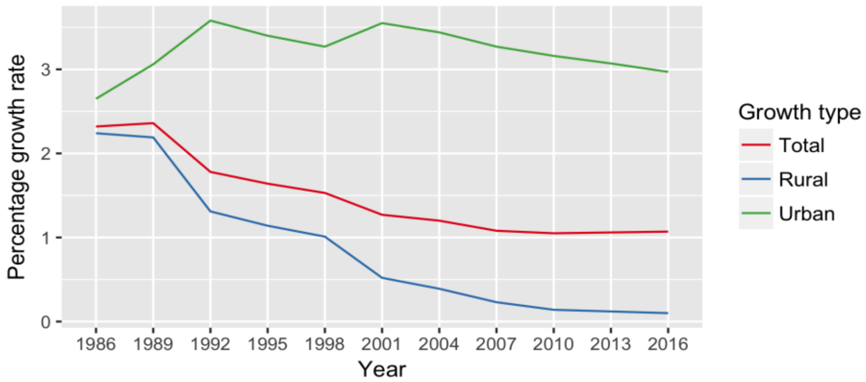
Figure 4.1: Total population by rural and urban



Source: World Development Indicators (WB, n.d., b).

It should be noted, however, that urbanisation and urban growth are two different concepts. Whereas urbanisation is related to the share of the population living in urban areas (see figure 4.1 above), urban growth is related to the absolute number of people living in urban areas (Tacoli et al., 2015). This indicates that urban growth is highly influenced by overall population growth. In fact, Tacoli et al. (2015) argue that the rate of urban population growth is roughly equal to the sum of the national population growth rate and the urbanisation rate. In turn, the contribution of rural-urban migration to urban population growth is approximately equal to the share of the urbanisation rate in the urban population growth rate (Tacoli et al., 2015). Viet Nam's population growth rate, together with the rate for both the rural and urban population, is illustrated in figure 4.2 below. The figure shows that all growth rates are positive for all years, although declining in scope. Furthermore, for all years the urban population growth rate is much higher than both the rural and the national rate. As the urbanisation rate is increasing (see figure 4.1) and the urban population growth rate is positive but declining (see figure 4.2), this would imply that the contribution of rural-urban migration to urban population growth is increasing over the years.

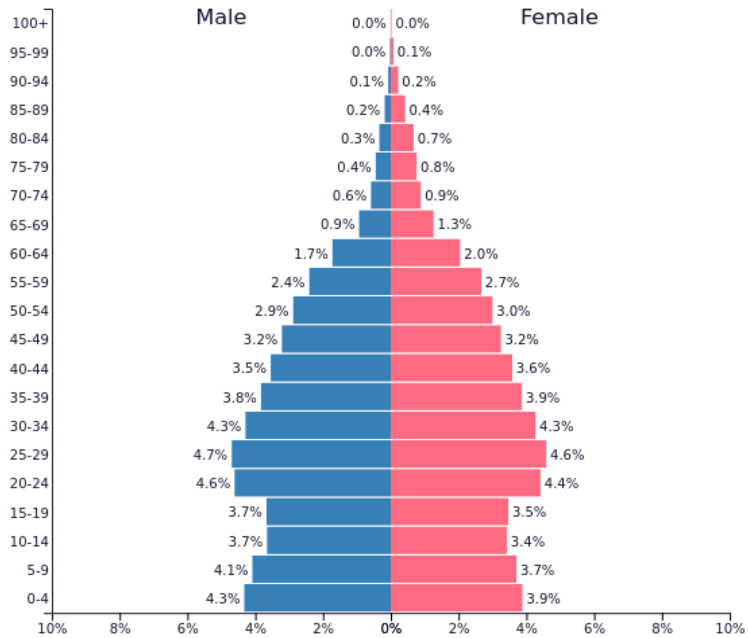
Figure 4.2: Population growth rates



Note: Growth rates are estimated as annual percentages.
Source: World Development Indicators (WB, n.d., b).

Viet Nam's population is more or less equally divided across males and females; in 2016 for every 100 women the country has 97.3 men (General Statistics Office of Viet Nam [GSO], 2016). The population pyramid in figure 4.3 shows that the largest contributors to the overall population are people within the age group 25-29. Furthermore, the working-age population (age 15 - 64) comprises around 84.3 percent of the total population, which is a relatively large share compared to other countries (Organisation for Economic Cooperation and Development [OECD], n.d., a). The shape of the population pyramid tells us that in 2016 Viet Nam's population was more or less stationary, although the size of the youngest groups indicate that the birth rate is increasing.

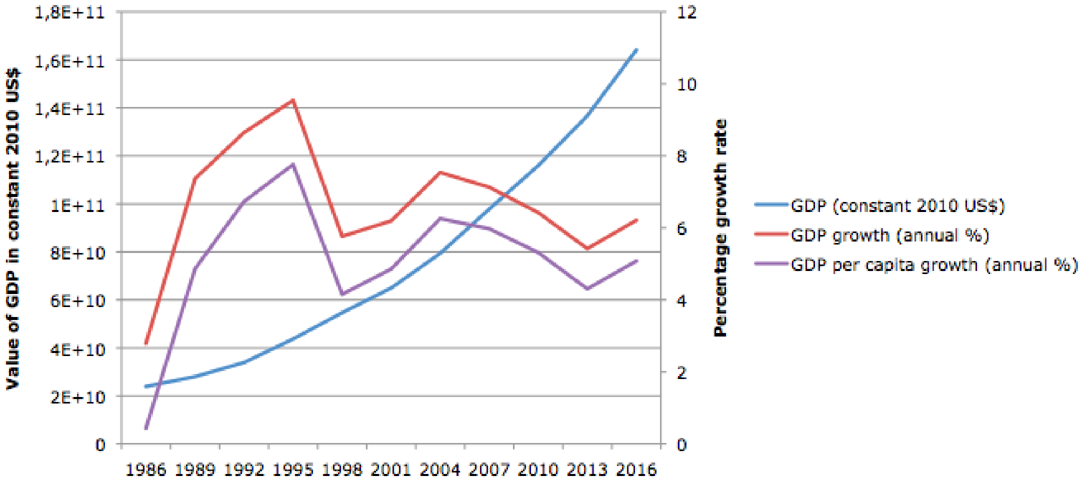
Figure 4.3: Population pyramid



Notes: Figure illustrates the population pyramid for Viet Nam for 2016. In 2016 the total population was estimated to equal 94.444.200 people (Population Pyramid Net, 2017).
Source: Population Pyramid Net (2017).

Viet Nam's GDP also shows an increasing trend (see figure 4.4 below). In 1986 economic and political reforms were launched in Viet Nam, known as the *Doi Moi* reforms. These reforms induced the industrialization of the country, which led to a restructuring of the economy and impressive economic growth (Nguyen & Winters, 2011). Therefore, the main economic driver of this incredible increase in GDP during the 1990s has been increased productivity (WB, n.d., b). As can be seen in the figure, during the 1990s and into the 2000s the average growth rates reached around 8 percent. This induces Viet Nam's GDP growth to be among the fastest in the world (WB, n.d., b). The figure shows that the trend in GDP per capita growth is following the same trend as the overall GDP growth rate, although the per capita growth rate is on average 1.57 percent point lower. The World Bank (n.d., b) argues that GDP growth has been equitable, given that over the years extreme poverty has fallen dramatically and social outcomes have improved. However, following from the GINI index, inequality within Viet Nam has increased over the years; the index equalled 35.65 in 1992, and increased to 42.68 in 2010 (WB, n.d., b). The increased GINI index might be explained by the increased rural-urban income gap, which multiplied by four times in just ten years (WB, 2016a). Also Anh et al. (2012) argue that the benefits of *Doi Moi* have been unequally distributed among regions, with cities as Hanoi and Ho Chi Minh City as well as surrounding provinces receiving large levels of industrial capital, and other rural areas lagging behind.

Figure 4.4: GDP and growth rates



Notes: Annual percentage growth rates of GDP (per capita) are based on constant local currency.
Source: World Development Indicators (WB, n.d., b).

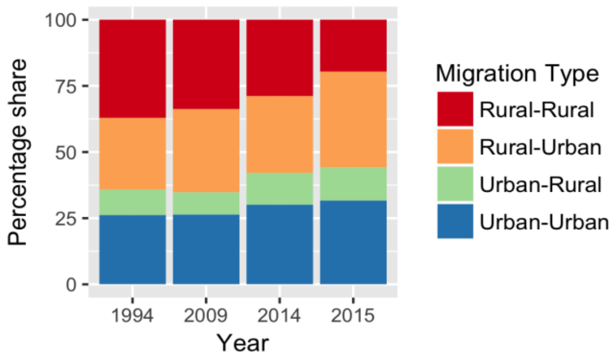
Over the years total employment in agriculture has been decreasing, whereas the percentage of people working in industry and services has been increasing (see figure 4.5 below). This is in line with the decreasing (increasing) trend we saw in rural (urban) population, as most jobs in the sectors of industry and services are located in or near urban areas. Furthermore, Agergaard & Thao (2011) and Anh et al. (2012) argue that the majority of rural-urban migrants find work in the informal sector, which is, however, often associated with relatively high vulnerability.

Figure 4.5: Employment shares

Source: World Development Indicators (WB, n.d., b).

As stressed in the introduction of this thesis, internal migration rates within Viet Nam have rapidly increased over the years. Where 6.5 percent of the total population were internal migrants in 1994, this percentage more than doubled towards 13.6 in 2015 (GSO & United Nations Population Fund [UNFPA], 2016b). Internal migration can be divided into four different types: (1) rural to rural, (2) rural to urban, (3) urban to rural, and (4) urban to urban. Over the years the share of rural-rural migration has decreased, whereas rural-urban migration and urban-urban migration have increased in share (the rural-urban share equalled 36.2 percent in 2013 (GSO & UNFPA, 2016b)). These trends are in line with the trends we saw in decreasing (increasing) rural (urban) population, and they are illustrated in figure 4.6 below. The attractiveness of urban areas has been described by Lewis (1979) who proposed a dual economy model in which labour surplus in rural areas moves to urban areas in order to supplement the urban labour shortage. Also Harris and Todaro (1970) propose a theory on the attractiveness of expected higher wages in urban areas, which might trigger rural-urban migration. As stressed above the rural-urban income gap rapidly increased over the years, and the Vietnamese internal migration survey held in 2015 indeed shows that economic reasons, like employment and income opportunities, is the largest group of reasons for migration (34.7 percent) (GSO & UNFPA, 2016b). This group of reasons is followed by family (25.5 percent) and education (23.4 percent).

Figure 4.6: Structure of internal migration flows



Source: GSO & UNFPA (2016a and 2016b).

When disaggregating internal migration flows into in and out migration rates by province, we see that in 2016 the region with the largest in-migration rate was the South East (GSO, 2016); this region also had the largest net-migration rate. The Mekong River Delta had the highest out-migration rate, and the lowest net-migration rate (GSO, 2016).

4.2 Food Consumption Patterns

Since the beginning of the 21th century, quite some studies have been done on food consumption in Viet Nam (Hoang, 2009; Le et al., 2003; Mergenthaler et al., 2009; Mishra & Ray, 2009; Molini, 2007; Nguyen & Winters, 2011; Thang & Popkin, 2004; Thi et al., 2005; Trinh & Morais, 2017). In this thesis report, these studies are used for analysing the trends in food consumption patterns in both rural and urban areas in Viet Nam for roughly the period 1992 to 2012. It is worth noting that often these studies are based on Viet Nam Household Living Standard Surveys (VHLSS). These surveys consist of a household level survey together with a commune-level survey, and provide fruitful information about food consumption expenditures and quantities.

To begin with, the average total number of kilocalories (kcal) consumed by the Vietnamese population has increased over the years, which goes together with narrowing of the food deficit (WB, n.d., b). The positivity of this trend holds for both rural and urban areas, although on average it was found that rural households consume slightly more kcal than urban households. This might be explained by the fact that rural households have in general more strenuous occupations, which correspond with higher energy demand (Molini, 2007; Thang & Popkin, 2004).

An overview of total per capita calorie consumption estimates by study can be found in table 4.1 below. As can be seen differences exist between the calorie estimates of different studies, although all studies are based on the VHLSS. This might be explained by the fact that different calorie conversion tables are used across studies (the Food and Agricultural Organisation of the United Nations [FAO] food consumption table for international use versus the calorie conversion table constructed by Viet Nam's National Institute of Nutrition [NIN]), and that some studies have applied an adult equivalence scale while others have not. Furthermore, Hoang (2009b) points to differences in dealing with food items that do not have quantity information, and with the food item of food away from home. However, in general the estimates point to increasing total per capita consumption of kilocalories, although we see a drop in 2008 due to a difficult climatic year and increased food prices (Trinh et al., 2017).

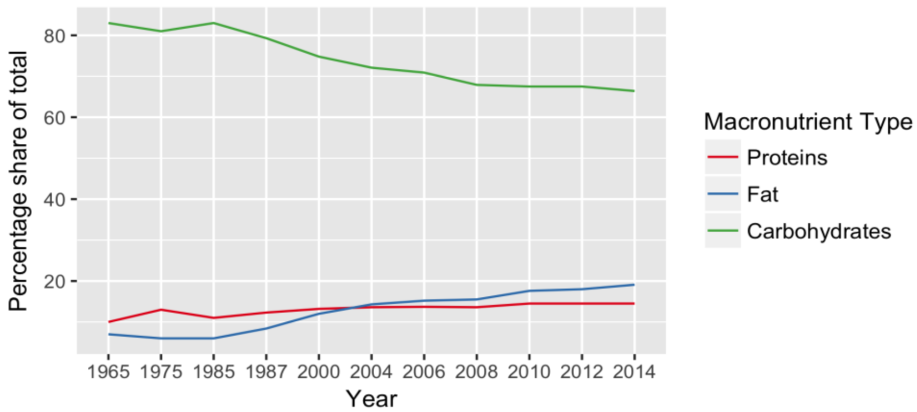
Table 4.1: Estimates of total consumed kilocalories per capita per day by study

	1992-93			1997-98			2004			2006			2008	2010	2012	2014
	All	Rural	Urban	All	Rural	Urban	All	Rural	Urban	All	Rural	Urban	All	All	All	All
Molini (2007)	2053	2060	2021	2267	2281	2218										
Thang & Popkin (2004)	2129	2173	1893	2111	2158	1783										
Mishra & Ray (2009)		2571	2165		2553	2039		3206	2824							
Nguyen & Winters (2011)							3144	3212	2949	3074	3163	2829				
Hoang (2009a)										2348	2376	2265				
Trinh et al. (2017)							3291			3272			2818	3632	3611	3651

Note: All studies have used the VHLSS to calculate their results. Estimates of Thang & Popkin (2004), Nguyen & Winters (2011), and Trinh et al. (2017) are based on adult equivalence scale. The rest is based on per capita. Molini (2007) and Mishra & Ray (2009) have used the FAO's food consumption table for international use to obtain calorie consumption, whereas Thang & Popkin (2004), Nguyen & Winters (2011), Hoang (2009a), and Trinh et al. (2017) have used the calorie conversion table constructed by Viet Nam's National Institute of Nutrition. Units in this table are total consumed kilocalories per capita per day.

The total number of kilocalories roughly come from three macronutrient categories. These are proteins, fat, and carbohydrates. Over the years the share of calories from proteins and fat have gradually increased, whereas the share of calories from carbohydrates has decreased (see figure 4.7 below). This means that macronutrient diet shares have moved closer to the 'ideal' balanced diet (Sp= 14%, Sf= 18%, Sc= 68%) established by the National Institute of Nutrition in Viet Nam (NIN, 2011b). The urge of this move towards the 'ideal' balanced diet follows from Le et al. (2003), who argue that in 1985 around 15 percent of the population suffered inadequate energy intake with low protein intake. They argue that much of the proteins came from rice, and that the consumption of meats, beans and fish was negligible. The mentioned trends in macronutrient diet shares hold for both rural and urban areas (Thang & Popkin, 2004; Trinh & Morais, 2017). However, in all analysed years the share of calories from protein and fat is higher fur urban households than for rural households. On the contrary, the share of calories from carbohydrate is higher for rural households.

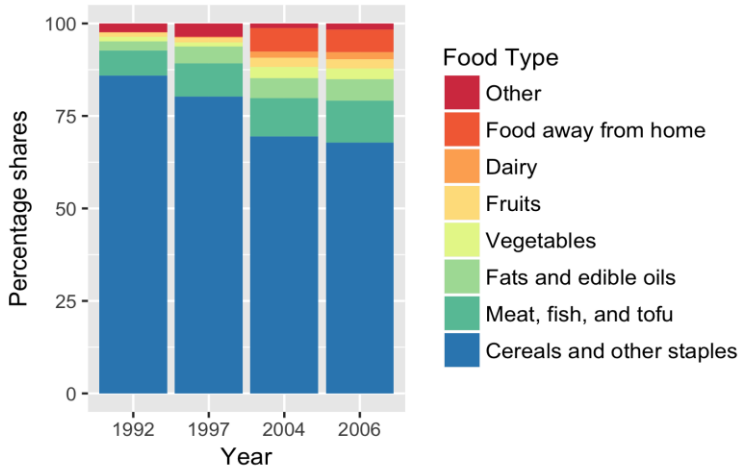
Figure 4.7: Macronutrient calorie intake share



Source: Figure is based on data from a study by Le et al. (2003) for the years 1965 to 2000, and Trinh & Morais (2017) for the years 2004 to 2014. Both studies have used the VHLSS to calculate their results.

Over the years, the Vietnamese diet has become more diverse (Nhan & Tuong, 2006). Generally speaking, a diet can be divided into different food categories. When looking at each food category and its calorie intake share, illustrated in figure 4.8 below, it can be seen that the largest proportion of calories consumed come from cereals and other starches. These food items fall into the macronutrient category of carbohydrates, which is, as mentioned above, the main macronutrient category in the Vietnamese diet. However, Mishra & Ray (2009) argue that the Vietnamese diet, with its great dependency on rice consumption even in the new millennium, faces a lack of diversification. They argue that this is especially the case for poor households. However, as can be seen from figure 4.8 the above mentioned calorie share has rapidly decreased over the years. The second largest category of calorie intake share is meat, fish, shellfish and tofu, which shows increases in calories consumed as well. Furthermore, the shares of fats and edible oils, vegetables, fruits, and milk and other dairy products have also increased, although with less percentage points than the other two categories. All mentioned changes are in line with the trends in macronutrient diet shares analysed above.

Figure 4.8: Calorie intake share per category

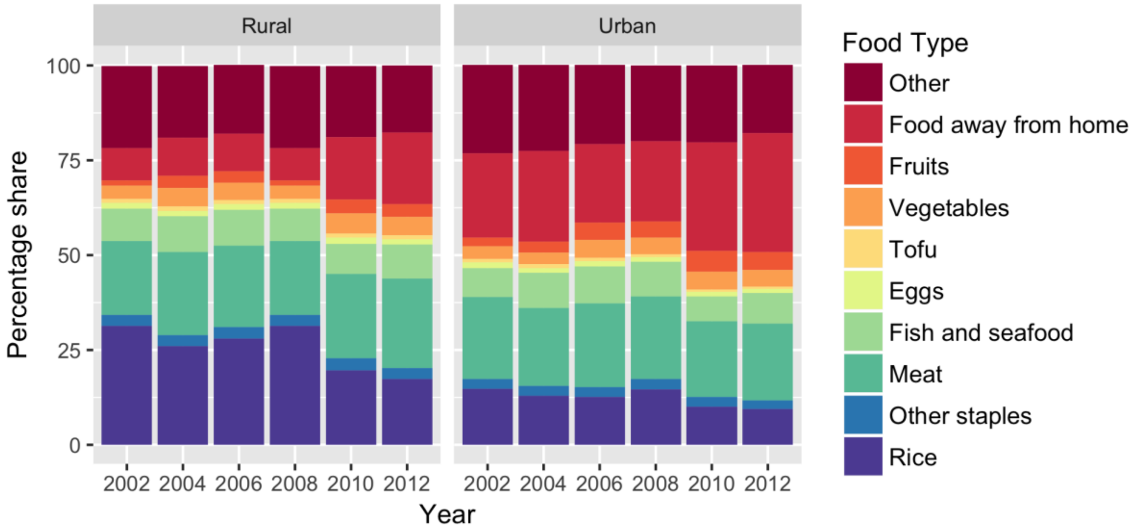


Notes: The estimations for the years 1992 and 1997 do not include the category of 'food away from home'. The category of 'food away from home' is an aggregate, and therefore does not provide information about which types of food are consumed away from home. This should be kept in mind when analysing the proportion of each share.
Source: Data for the years 1992 and 1997 come from a study by Thang & Popkin (2004), whereas data for the years 2004 and 2006 come from a study by Nguyen & Winters (2011). Both studies have used the VHLSS to calculate their results.

The Engel coefficient is an estimate of the ratio of food expenditures over total expenditures (Deaton, 1997), which can easily be derived from the data provided in the VHLSS. The coefficient is estimated to be quite stable over the years for Vietnamese households, averaging around 46 percent in the period 2004 to 2014 (Trinh & Morais, 2017). It should be noted, however, Trinh & Morais (2017) have taken into account only regular food expenditures in their calculations (they omitted food expenditures on festive occasions). Hoang (2009b), on the contrary, used both regular food expenditures and food expenditures on festive occasions to calculate the average Engel coefficient for 2006. Therefore in his study the estimation for the coefficient turned out to be much higher, equalling 53.5 percent. Furthermore, rural households were estimated to spend more of their budget on food than urban households (55.0 percent compared to 48.2 percent respectively) (Hoang, 2009b). Although the Engel coefficient has been quite stable, total consumption expenditures per capita on food have increased for both rural and urban households (GSO, 2012; Trinh & Morais, 2017). Moreover, the average price of 1000 kcal has decreased in the period 2004 to 2014 for all food categories (Trinh, 2017), indicating that households on average have increased their calorie intake, which is in line with what we saw in table 4.1 above.

Like calorie intake, food consumption expenditures can be divided into different categories. The shares are illustrated in figure 4.9 below; it should be kept in mind that the category of food away from home is an aggregate of different types of foods, which can, thus, contain food types mentioned in the other categories - as long as they are consumed away from home. In all years the share of rice is larger for rural households than for urban. However, Mishra & Ray (2009) show that in 1992 and 1997 this was the opposite; rice expenditures were 8.37 and 10.47 percent points higher for urban households than for rural respectively. Moreover, in both years rice expenditures were much higher than in the beginning of the 21st century. It follows that the expenditure share for rice has been decreasing considerably over the years for both rural and urban areas. The fact that from 2002 onwards rural households spend a higher share of their food budget on rice than urban households might be explained by urban areas in Viet Nam having experienced more economic growth in recent decades than rural areas, and most dynamic economic sectors being located in urban centres (Nguyen & Winters, 2011). This indicates that economic growth might has induced changes in food consumption patterns more rapidly in urban areas than rural. This is confirmed by Mishra & Ray (2009), who argue that non-poor households, especially in urban areas, have diversified their diet more profoundly than poor households. Next to the decrease in rice expenditures, a large increase can be seen in the share of food eaten outside the home for both rural and urban households. This might be alarming given that home-cooked meals are in general considered more healthy and nutritious (NIN, n.d., b). From 1992 to 2004 the share of the categories of fish and meat show considerable increases for both rural and urban households (Mishra & Ray, 2009), although those shares are quite stable after 2004 (see figure 4.9). The most important category of meat consumption is pork (FAO, n.d.; Hoang, 2009b), accounting for around 60 percent of total meat expenditures in 2006 for both rural and urban areas (Hoang, 2009b). Furthermore, urban households are estimated to spend a larger share of their food budget on fruits than rural households, although the share for vegetables is more or less similar for both types of households. However, for all households both shares (i.e. fruits and vegetables) are more or less stable in the period 2002 to 2012. This is striking given the profound changes in other food categories, the fact that fruits and vegetables are important components of a healthy diet (Glopan, 2016; World Health Organization [WHO], 2015), and the fact that Viet Nam in general has a favourable climate and ground for fruits and vegetables production (Vietnam Trade Promotion Agency, 2008). In the period 1990 to 2013 South-East Asia as a region saw a positive percentage change for the consumption of fruits, and a negative percentage change for the consumption of vegetables (Glopan, 2016). Especially the percentage changes for the consumption of whole grains, seafood, nuts and seeds, and processed meat were large and positive for the region (Glopan, 2016). The changes in food consumption expenditure shares per category illustrated in figure 4.9 are in line with the changes in food calorie intake shares per category and the changes in macronutrient calorie intake shares that we saw above.

Figure 4.9: Food consumption expenditures per category



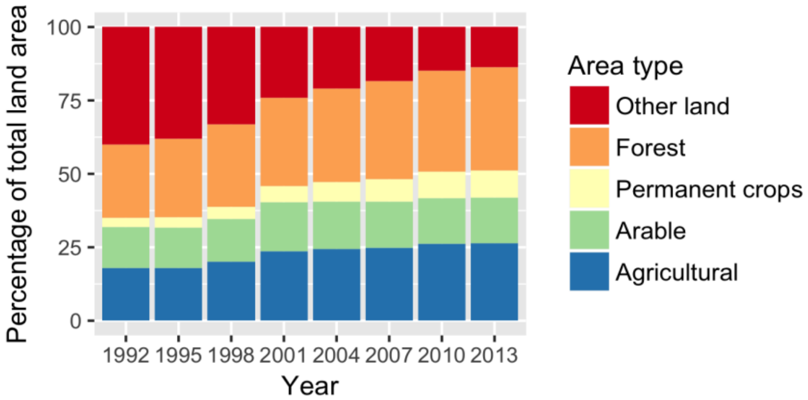
Notes: The category of 'food away from home' is an aggregate, and therefore does not provide information about which types of food are consumed away from home. This should be kept in mind when analysing the proportion of each share.
Source: (GSO, 2012).

It is worth emphasizing that the average price of 1000 kcal has decreased in the period 2004 to 2014 for all food categories. For example, where in 2004 the average price of 1000 kcal of plain rice equaled 2.08 thousand VND, it decreased to 1.10 thousand VND in 2014 (Trinh, 2017). Furthermore, the average price of 1000 kcal of pork meat equaled 19.87 thousand VND in 2004, and decreased to 9.79 thousand VND in 2014 (Trinh, 2017). However, these price calculations are based on national averages, and therefore do not control for e.g. region, taste, quality of the product, and availability of products. The GSO (n.d.) provides data on the Consumer Price Index (CPI) in Viet Nam for both rural and urban areas. In 2006, compared to 2005, the CPI for food equaled 100.2 for both rural and urban areas. In 2014, compared to 2013, the CPI for food equaled 101.52 for rural areas and 101.02 for urban areas. These CPIs show that relative price changes were more or less similar across rural and urban areas. Unfortunately no data is available on absolute price levels in both rural and urban areas.

4.3 Agricultural Production and International Trade

To begin with, the share of total land area devoted to agriculture, arable land, permanent crops, and forest has increased over the years (see figure 4.10 below); since the year 2000 the total area under agriculture has grown nearly 15 percent (WB, 2016a). This goes together with an increase in agriculture value added of around 15 billion US dollars in the period 1986 to 2016 (W WB, n.d., b). However, the percentage share of agriculture contributing to GDP has decreased (21.02 percent in 2010 compared to 18.14 percent in 2016), which can mainly be described by the rapid increase in GDP shown in the first section of this chapter (WB, n.d., b). Different types of production units are responsible for Viet Nam's agricultural production. These are agricultural households, farms, agricultural enterprises, and other entities. About 90 percent of all agricultural land falls under the former two categories, where the agricultural households segment is dominated by very small farms (WB, 2016a). However, this also means that agricultural land is fragmented as small farms often hold three or four, or many more tiny plots of land, due to the past allocation of land in an egalitarian way. This has implications for the efficiency of small farms, especially because plots are sometimes separated by considerable distance (WB, 2016a). The total number of farms has increased from 22.655 in 2012 to 33.488 in 2016 (GSO, 2016). Furthermore, the majority of economic activities at farms is related to livestock (62.32 percent), which is followed by cultivation (27.52 percent).

Figure 4.10: Land area by types



Notes: Other land is the land not classified as agricultural land and forest area. It includes built-up and related land, barren land, other wooded land, etc. (FAO, 2001).
Source: FAOSTAT database (n.d.).

Over the years, the production of food, livestock, and crops has rapidly increased (see table 4.2 below). This is likely the consequence of the increased productivity that happened after the *Doi Moi* reforms in the 1990s (WB, n.d., b), which led to expanded and more intensive use of land and other natural resources, and relatively heavy use of fertilizer and other agro-chemicals (WB, 2016a). The largest increase in production can be seen in the production of livestock, having the lowest production index in 1986 and the highest in 2013. This increase is mainly occurring in medium- and large-size farms (WB, 2016a).

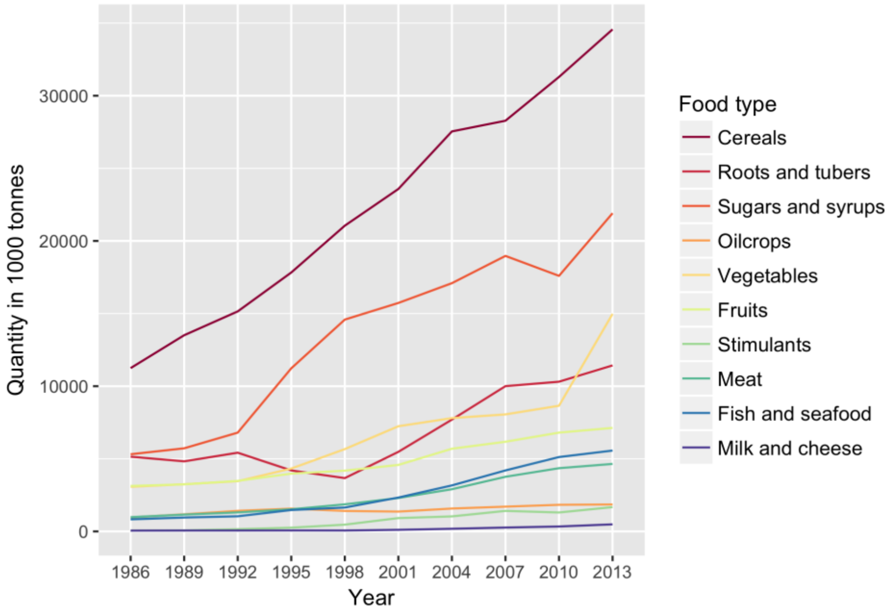
Table 4.2: Production indices

	1986	1989	1992	1995	1998	2001	2004	2007	2010	2013
Food production index	39,74	45,27	51,42	59,83	69,28	80,25	96,25	107,49	118,64	131,60
Livestock production index	32,48	37,55	42,48	49,47	58,96	72,83	90,14	117,85	137,36	147,45
Crop production index	39,55	44,73	50,79	59,78	69,88	82,33	97,74	106,87	115,07	130,06

Notes: Base year in this table is 2004-2006 (2004-2006 = 100).
Source: World Development Indicators (WB, n.d., b).

In the section on food consumption patterns we saw that the consumption of cereals, including rice, has decreased over the past years. However, an increasing trend can be found in the production of cereals (see figure 4.11 below), which mainly comprises of the production of rice (FAO, n.d.). Furthermore, for all years cereals is the food group with the largest production quantities, which is followed by the group of sugars and syrups. Sugarcane is the main crop responsible for this latter high level of production (FAO, n.d.). Considerable increases over the years can be seen in the production of sugars and syrups, vegetables, and roots and tubers as well, supporting the claim of increased agricultural productivity in Viet Nam over the years. It is argued that one of the major drivers of agricultural production has been exports (WB, 2016a).

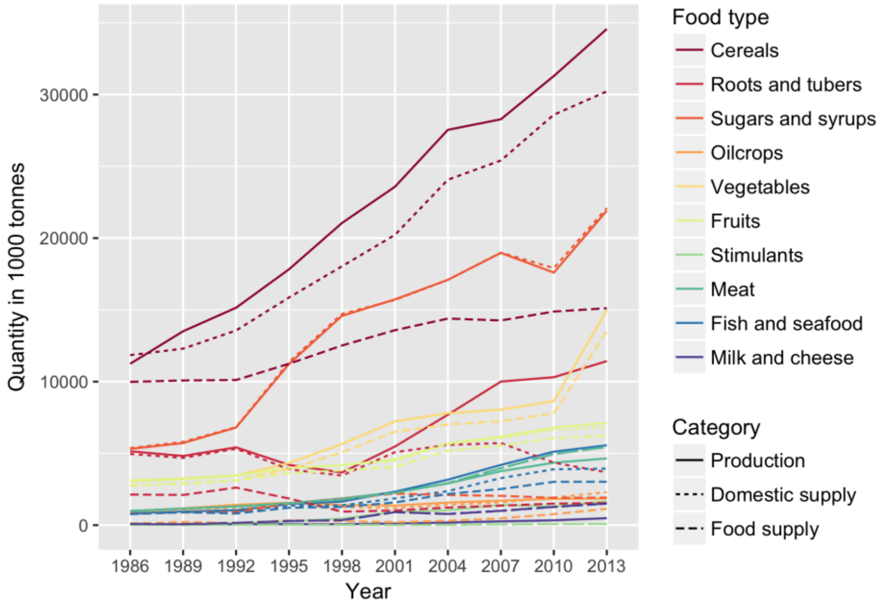
Figure 4.11: Production quantity



Notes: Production relates to the total domestic production whether inside or outside the agricultural sector, i.e. including non-commercial production and production in kitchen gardens (FAO, 2001). Commodities are classified based on major food groups defined by the FAO (2001).
Source: FAOSTAT database (n.d.).

As large quantities of produced goods are exported, this indicates that not all goods produced end up with Vietnamese consumers. Furthermore, part of production is used as feed for animals, used as inputs for other products, or for other non-human consumption purposes. Therefore the FAO has distinguished the concepts of 'domestic supply quantity' and 'food quantity', which are also influenced by imports and changes in stocks. The first concept refers to production plus imports minus exports plus changes in stocks. The second concept refers to food available for human consumption. In figure 4.12 below one can find the production, domestic supply, and food quantity for ten major food groups. First of all nearly all trends in the figure are increasing over the years. Large differences can be found in the production, domestic supply, and food availability of cereals. This indicates that large quantities of cereals are produced for export, which is also shown later in this section. Furthermore maize, which is part of this food group, is produced and supplied for domestic utilization in large quantities, but on the contrary human food consumption turns out to be relatively low (FAO, n.d.). This is because maize is no important component of the Vietnamese diet, and is mainly produced for animal feed instead. From around the year 2000 the levels of production and domestic supply of roots and tubers start widening, with sharp increases in production and slight increases followed by decreases in domestic supply. This widening is mainly due to large increases in the production and export of cassava (FAO, n.d.). Another interesting finding is the more or less similar production and domestic supply quantity for sugars and syrups, while food available for human consumption is much lower. This might be due to the fact that large quantities of sugar are used as raw material in the food processing industry (Vu, 2014), but could also indicate that Viet Nam mainly produces sugars and syrups for non-human consumption activities. Furthermore, after 1992 the production of stimulants (e.g. coffee) starts increasing, whereas domestic supply and food supply stay more or less similar. This indicates that the export of stimulants increases significantly after the year 1992 (more on this below). Lastly, from the year 2004 onwards the domestic supply of meat is slightly higher than the production of meat, indicating the import of meat after 2004 (more on this below).

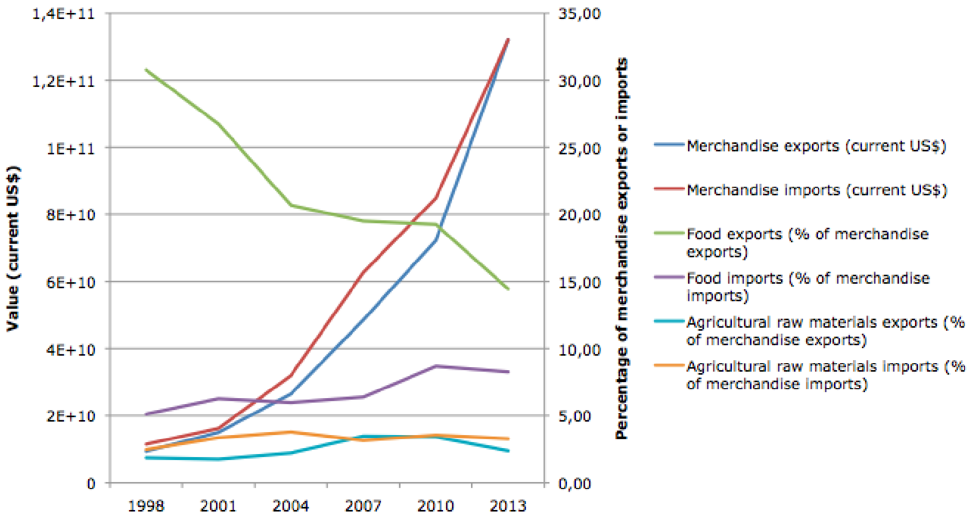
Figure 4.12: Production, domestic supply and food quantity



Notes: Production relates to the total domestic production whether inside or outside the agricultural sector, i.e. including non-commercial production and production in kitchen gardens. Supply for domestic utilization is equal to production plus imports minus exports plus changes in stocks (decrease or increase). Food comprises the amounts of the commodity in question and of any commodities derived therefrom not further pursued in the food balance sheet that are available for human consumption during the reference period. It is important to note that the quantities of food available for human consumption reflect only the quantities reaching the consumer. The actual amount of food consumed may be lower than the quantity shown in the graph, depending on the degree of losses of edible food and nutrients in the household (FAO, 2001). Commodities are classified based on major food groups defined by the FAO (2001).
Source: FAOSTAT database (n.d.).

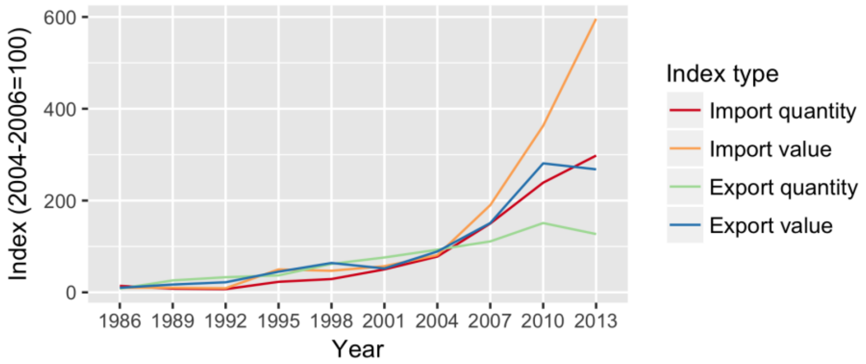
Total exports and imports have rapidly increased over the years (see figure 4.13 below). However, the share of food exports out of all merchandise exports has decreased, whereas the share of food imports has slightly increased. On the contrary, the share of export and import of agricultural raw materials has been quite stable over the years. Figure 4.14 shows that for the period 1986 to 2004 both the quantity and value of imports and exports follow a more or less similar increase. After the year 2004, both exports and imports keep increasing, although imports exceed exports for both quantity and value. Furthermore, value exceeds quantity for both imports and exports, indicating increased prices after 2004.

Figure 4.13: Merchandise exports and imports



Source: World Development Indicators (WB, n.d., b).

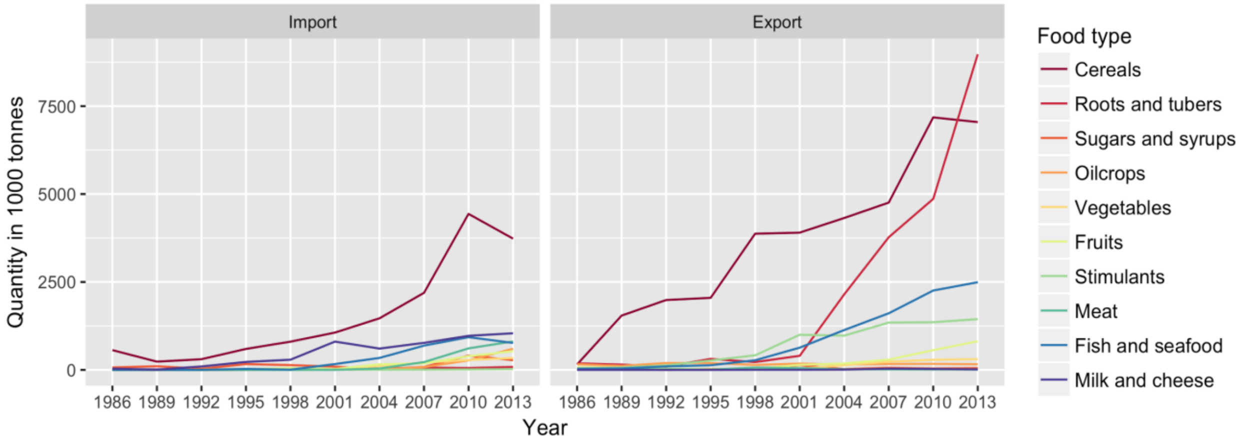
Figure 4.14: Trade indices



Source: FAOSTAT database (n.d.).

When looking at the figures of import and export quantity by major food groups (see figure 4.15 below), we see that both the import and export quantity of nearly all food groups increases over the years, especially after the year 2001. However, as both the import and export data include the imports and exports of foods for non-human consumption (FAO, 2001), this should be taken into account when interpreting the data. First of all cereals is both the largest imported and exported food group. This can be explained by the different commodities constituting this major food group. Cereals that are imported in large quantities are wheat and maize, whereas cereal export mainly consists of rice (FAO, n.d.). The large import of maize is mainly the consequence of a high demand for animal feed components, as maize is not widely consumed by the Vietnamese population itself (which we also saw in figure 4.12 above). After the year 2004, the import of meat increases rapidly, where poultry and bovine meat are the largest contributors (FAO, n.d.). These imports complement domestic livestock production, which mainly comprises of pig meat. Furthermore the import quantity of milk and cheese is considerable, and the same can be said for the import of fish and seafood products (i.e. marine fish) and oil crops. However, an example of oil crops is soybeans, which is widely used for animal feed as well. Figure 4.15 shows that next to cereals, roots and tubers have increased rapidly in export quantity from the year 2001 onwards. The food group even exceeds cereal (i.e. rice) exports in the year 2013. This increase in export of roots and tubers comes from a rapid increase in the export of cassava. Also the export of fish and seafood and stimulants has rapidly increased over the years. The former increase is mainly due to exports in freshwater fish, marine fish, and crustaceans, whereas the increase in stimulants exports is mainly constituted of the increase in exports of coffee (FAO, n.d.). All these findings are in line with findings of the World Bank (2016a), which argues that Viet Nam now ranks among the top five global exporters in products as diverse as shrimp, coffee, cashews, rice, and pepper.

Figure 4.15: Import and export quantity



Notes: Imports, measured in 1000 tonnes, cover all movements of the commodity in question into the country as well as of commodities derived therefrom and not separately included in the food balance sheet. Exports, measured in 1000 tonnes, cover all movements of the commodity in question out of the country during the reference period. Imports and exports both include imports and exports for non-human consumption, like animal feed (FAO, 2001). Commodities are classified based on major food groups defined by the FAO (2001).
Source: FAOSTAT database (n.d.).

4.4 Summary of Trends

We have seen several trends in demographics, food consumption patterns, agricultural production and international trade. This section serves to highlight and summarize the most important trends. Viet Nam's population has rapidly increased over the period 1986 to 2016. The proportion of the population living in urban areas is increasing, whereas the rural population declines. The fact that more and more people are employed in the sectors of industry and services instead of agriculture perfectly corresponds with the above mentioned increasing trend in urbanisation. Furthermore, after the economic and political reforms known as the *Doi Moi* reforms, Viet Nam became increasingly industrialized, inducing impressive economic growth. Viet Nam's GDP growth rate even was among the fastest in the world. The main economic driver of this incredible increase in GDP has been increased productivity. Over the years, internal migration has increased, and its structure has shifted towards rural-urban migration being the largest share of total internal migration. In turn, rural-urban migration is contributing to the increasing urban population.

Trends can be seen in food consumption patterns as well. In the period 1992 to 2014 total consumed kilocalories per capita have increased. The average share of calories from fats and proteins has increased, whereas the contrary can be seen for the share from carbohydrates. These trends hold for both rural and urban areas, although urban households are predicted to consume a higher share of total kilocalories from fat and proteins than rural households. When we disaggregate these trends into different food groups, we see that they still hold. For all years rice is an important component of the Vietnamese diet, especially for rural households, although its importance is declining. Meat is another food group with a considerable expenditure share; from 2002 to 2008 it is the second largest food expenditure group for both rural and urban households and its share is quite stable over time. The most frequently consumed type of meat is pork. The food category of food away from home has become increasingly important for both types of households over the years, although especially for urban households; from 2010 onwards this food group has the largest expenditure share. The expenditure shares of fruits and vegetables are quite stable for both rural and urban households. The share of fruits for urban households has been slightly larger than the one for rural households for most years. In general all trends in food groups are going in the same direction for both rural and urban areas, although on average the patterns are more profound in urban areas. The trends we see in shifting food consumption patterns correspond with the definition of a nutritional transition, introduced in the introduction of this thesis.

Considering Viet Nam's agricultural production and international trade, we see increasing trends in land devoted to agriculture, arable land, permanent crops and forestry. Agricultural production has increased rapidly over the years, which is mainly due to increased productivity after the *Doi Moi* reforms and increasing export demand. Increasing production can be seen for both crops and livestock, where especially the latter increase has been impressive. The largest production quantities can be seen for cereals, where especially the production of rice contributes to the scope of this group. The second largest food group in production quantity is sugars and syrups. Increases are seen in the production of nearly all food groups, although the increases in cereals, sugars and syrups, vegetables, and roots and tubers for the period 1986 to 2013 have been largest. However, the Vietnamese population does not consume all foods produced. Large quantities are also exported or used as animal feed, which is why trends have been analysed in domestic supply and foods consumed also. These trends showed that the production of cereals largely exceeded the domestic supply and food consumption of cereals. This as rice has been exported in large quantities and as maize is mainly produced (and imported) for animal feed. Furthermore, after 2002 the production and export of cassava has rapidly increased, inducing a gap between the production and domestic supply of roots and tubers. Large differences can be seen in the domestic supply and food consumption of sugars and syrups as well, indicating that large quantities of this category are used for non-human consumption. After 1992 the production of stimulants increases rapidly, corresponding with the increasing trend in export that we see for stimulants (i.e. coffee). Imports for meat increase after 2004, which is mainly due to imports in bovine and poultry meat. Overall, total imports and exports have increased, although the share of food in total merchandise exports (imports) has decreased (increased). The main imported and exported food group is, again, cereals, where cereal imports mainly exist of wheat and maize, and cereal exports of rice. The imported quantities of milk and cheese, marine fish, and soybeans have increased over the years, whereas Viet Nam has become one of the top five global exporters of shrimp, coffee, cashews, rice, and pepper.

Summarizing we see increasing trends for nearly all analysed concepts. Where total number of kilocalories consumed increase, total quantities of production and imports and exports increase as well. Trends in macronutrients and food consumption expenditures point to Vietnamese diets following the definition of a nutritional transition.

5. Effect of Rural-Urban Migration and Having Rural Origins on Diet

This chapter presents the results of the second research objective and question, namely the analysis of the effect of rural-urban migration and having rural origins on diet of migrants. The chapter is structured as following: first an elaboration is given on characteristics of rural, urban and rural-urban migrant households. This is followed by a section on diets in 2014, as this provides useful background information before jumping into the analysis of the effect of rural-urban migration (section 5.3) and having rural origions (section 5.4) on diets of migrants. The chapter concludes with a section on robustness checks and sensitivity analyses on the estimated results.

5.1 Characteristics of Rural, Urban and Rural-Urban Migrant Households

In order to provide some more context around the analysed effects, the main characteristics of different types of households are presented in table 5.1 below. The numbers are produced using one-way ANOVA, and pairwise comparisons of means are executed using Tukey tests. The results from these Tukey tests can be found in table D.1 in Appendix D. First of all, interesting is the fact that from all rural-urban migrant households only 1.15 percent belongs to an ethnic minority group (i.e. *not* to the 'Kinh' ethnic group). This percentage is much lower (and significantly different at the 1 percent level) than the percentage for rural and urban households (21.79 and 8.62 respectively). The average household head in the sample is 50.52 years old, which is 46.16 for the average spouse. The age of the head and spouse is quite similar across household types, although the spouse is estimated to be significantly older in urban and migrant households than in rural households (p= 0.000 and p= 0.007 respectively). This holds for the head of urban compared to rural households as well (p= 0.001). Furthermore, the percentage of married household heads is significantly different across all types of households; it is highest for rural households (82.45 percent) and lowest for rural-urban migrant households (74.47 percent). It is estimated that the head and spouse of rural-urban migrant households have obtained higher levels of education than the head and spouse of both rural and urban households; all pairwise comparisons of means are significant at the 1 percent level. On average the obtained level of education of the head and spouse of migrant households ranges between higher secondary and college, university, MA/MSc, or PhD qualification, whereas this is between lower and higher secondary education for urban households, and between primary and lower secondary qualification for rural households. The percentage of households having registration status in the current residential area is significantly smaller for migrant households (79.46 percent) compared to both rural and urban households (98.67 and 95.97 percent respectively). Although all types of households have on average a household size between 3 and 4 members, rural-urban migrants households are predicted to have a significantly smaller household size than both rural and urban households (p= 0.000 and 0.008 respectively). The number of children in the household differs significantly across all types of households, although on average it is estimated to equal 1. In 2014 13.19 percent of rural households were classified as 'poor' in the commune or ward by local authorities, which is a significantly higher percentage than the one for both urban and rural-urban migrant households (3.52 and 1.54 percent respectively). Similar patterns hold for the percentages of having insufficient food and foodstuff to meet the needs of the household; these percentages are significantly higher at the 1 percent level for rural households than for both urban and rural-urban migrant households. More than 50 percent of rural-urban migrant households reported to have savings at a bank (e.g. via a bank account or savings passbook), which is much higher (and significantly different) than this percentage for rural and urban households (12.22 and 35.33 percent respectively). Furthermore, both urban and migrant households are estimated to have significantly larger residential area and value of accommodation than rural households. Moreover, migrant households are estimated to have significantly larger value of accommodation than urban households as well (p= 0.000). The average total food consumption expenditures on both regular and festive occassions are estimated to be significantly different from each other in all cases, and they are estimated to be highest for rural-urban migrant households and lowest for rural households (the predicted difference is 257 and 110 million VND respectively). Average total calorie intake, on the contary, is estimated to be similar across households; no significant differences are found. Lastly, total revenue and income are estimated to be significantly larger for both urban and migrant households than for rural households. Moreover, total income of migrant households is estimated to be significantly larger than for urban households as well (p= 0.000).

Table 5.1: Main characteristics of different household types

Characteristics	Total	Rural	Urban	RU Migrants
Ethnic minority (%)	17.49	21.79	8.62	1.15
Age of household head	50.52	50.18	51.39	51.07
Age of spouse	46.16	45.69	47.23	47.71
Male household head (%)	75.15	79.47	65.55	61.61
Married household head (%)	81.26	82.45	79.37	74.47
Education of household head (highest obtained level)	1.98	1.58	2.75	3.48
Education of spouse (highest obtained level)	1.97	1.56	2.87	3.26
Registration status in resident area* (%)	96.93	98.67	95.97	79.46
Household size	3.84	3.88	3.82	3.60
Number of children	0.99	1.04	0.91	0.78
Classified as 'poor' in 2014 (%)	10.23	13.19	3.52	1.54
Insufficient food* (%)	1.91	2.31	1.11	0.39
Insufficient foodstuff* (%)	5.21	6.52	2.23	1.54
Savings at bank (%)	19.99	12.22	35.33	52.21
Residential area (m2)	80.55	76.17	91.61	88.67
Value of accommodation (1000 VND)	616916.40	368491.00	1160979.00	1419922.00
Total food consumption expenditure, regular (1000 VND)**	4371.25	3739.18	5783.83	6305.59
Total food consumption expenditure, festive (1000 VND)**	3222.39	2959.52	3801.27	4064.28
Total calorie intake, regular (kcal)**	262773.30	262911.90	263122.00	259622.00
Total revenue (1000 VND)**	165542.70	147922.00	206019.20	214927.20
Total income (1000 VND)**	108040.30	91073.38	143865.40	168631.30

*Notes: Numbers have been produced using one-way ANOVA. The sample size for 'Total' equals 9.093, 6.415 for 'Rural', 2.157 for 'Urban' (excl. RU Migrants), and 521 for 'RU Migrants'. *The percentage of households having registration status in the resident area (in commune or ward) is based on the registration status of the head of the household. **Estimated over the past 30 days. Pairwise comparisons of means have been executed using Tukey tests, for which the results can be found in Appendix D.*
Source: Authors calculations based on the VHLSS of 2014.

Summarizing the results described above, significant differences can be found between rural and urban households for all characteristics except for household size and total calorie intake.

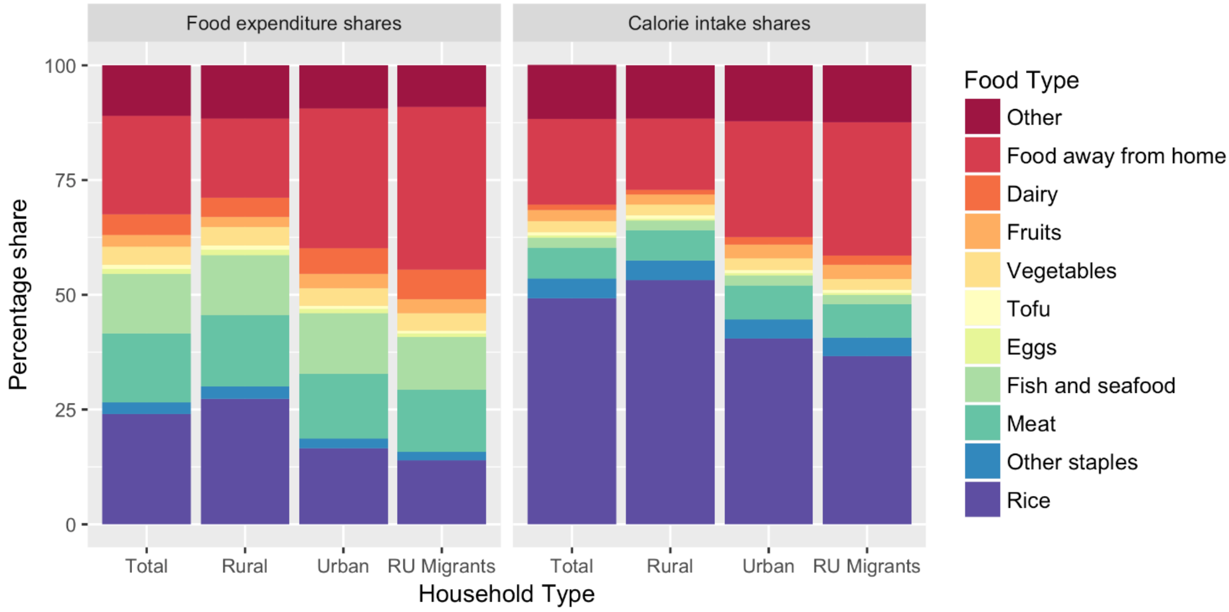
Compared to rural households a smaller proportion of urban households belongs to ethnic minority groups, more urban households are female headed, head and spouse of urban households have obtained higher levels of education, a smaller percentage of urban households is classified as poor in 2014, and urban households have higher food consumption expenditures and revenue and income. Similar figures hold for rural-urban migrant households compared to rural households as well.

The characteristics in table 5.1 show that rural-urban migrant households look most similar like urban households. Compared to urban households, however, largest differences can be seen for the percentage belonging to an ethnic minority group (smaller for migrant households), obtained level of education of head and spouse (larger for migrant households), percentage of households which have savings at a bank (larger for migrant households), total food consumption expenditures and total income (all larger for migrant households).

5.2 Descriptive Analysis of Diets in 2014

This section provides a cross-sectional analysis of diets of different types of households in Viet Nam in 2014. Food consumption expenditure shares and food calorie intake shares are illustrated in figure 5.1 below (as a reference, the average price of 1000 kcal for each food item is included in Appendix E). Again, it should be noted that the category of food away from home is an aggregated food group, and thus no information is available about which types of food are consumed away from home. Furthermore, recall from table 5.1 that total regular food expenditures in 1000 VND were estimated to be largest for rural-urban migrant households, and lowest for rural households. Total calorie intake was estimated to be more or less similar across household types. When looking at food expenditure shares in the left part of the figure, we see that for all types of households the shares of rice, meat, fish and seafood, and food away from home are considerable. Interesting is the estimated small difference in size of expenditure share for the categories of meat and fish and seafood for the total population (the shares were on average 15.12 and 12.96 percent respectively), given that the estimated difference was much larger in 2012 (the shares were on average 22.2 and 8.6 percent respectively, see figure 4.9) (GSO, 2012). This might be explained by the fact that the GSO has used both regular and festive food consumption expenditures to calculate its results, while I only focused on regular expenditures. As the share of meat consumption is much larger on festive occasions (38.5 percent), this might at least partly explains the narrowing difference between the shares of both categories. The right part of figure 5.1 illustrates the calorie intake shares per category; we see that on average for the total population nearly 50 percent of all consumed calories come from rice consumption. Furthermore, the share of food away from home is considerably large as well - 18.62 percent on average.

Figure 5.1: Food expenditure and calorie intake shares per category



Notes: The sample size for 'Total' equals 9.093, 6.415 for 'Rural', 2.157 for 'Urban' (excl. RU Migrants), and 521 for 'RU Migrants'. The category of 'food away from home' is an aggregate, and therefore does not provide information about which types of food are consumed away from home. This should be kept in mind when analysing the proportion of each share.
Source: Author's calculations based on the VHLSS of 2014.

Tukey tests for pairwise comparisons of means show that the food consumption expenditures shares for rural and urban households are all significantly different from each other at either the 1 or 5 percent level, except for the category of 'fish and seafood' (p= 0.594). The largest differences, however, can be seen in the categories of 'rice' and 'food away from home'. It is estimated that rural households spend 27.34 percent of their total food budget on rice, compared to 16.63 for urban. For the category of 'food away from home' the shares are 17.25 for rural households and 40.48 for urban households. It should be noted, however, that the differences in percentage shares do not say anything about the absolute level of expenditures on each of the identified categories. When turning to calorie intake shares of rural and urban households, all shares are significantly different from each other at either the 1 or 5 percent level, except for the category of 'other staples' (p= 0.493) and 'tofu' (p= 0.099). The largest differences can again be found in the categories of 'rice' (12.76 percent) and 'food away from home' (9.75 percent).

When we turn to the food consumption expenditure shares of rural-urban migrant households, however, Tukey tests for pairwise comparisons of means show that the shares of all food categories are significantly different at the 1 percent level from the shares of rural households, except for the category of 'vegetables' (p= 0.103). The largest difference in percentage points exists for the categories of 'rice' (- 13.46 percent points) and 'food away from home' (+ 18.20 percent points). When turning to calorie intake shares, the only categories that are not significantly different between rural-urban migrant households and rural households are 'other staples' (p= 0.607), 'fish and seafood' (p= 0.170), and 'vegetables' (p= 0.833). Again the largest differences are found for the categories of 'rice' (-16.64 percent points) and 'food away from home' (+ 13.64 percent points). Figure 5.1 shows that both the expenditure shares and calorie intake shares of rural-urban migrant households are quite similar to the shares of urban households. In fact, the only food expenditure shares that are significantly different from each other for both groups are the shares of 'rice', 'fish and seafood', and 'food away from home' (p= 0.000 in all cases). The cross-sectional comparison predicts that rural-urban migrant household spend 2.75 percent points less of their food budget on rice, 1.78 percent points less on fish and seafood, and 4.97 percent points more on food away from home than urban households. For calorie intake shares significant differences can be found for the categories of 'rice' (p= 0.000), 'fish and seafood' (p= 0.010), 'dairy' (p= 0.004), and 'food away from home' (p= 0.000). The comparison predicts that on average rural-urban migrant households have a 3.88 percent points lower calorie intake from rice, 0.22 percent points lower calorie intake from fish and seafood, 0.29 percent points higher calorie intake from dairy products, and 3.89 percent point higher calorie intake from food away from home than urban households.

In the definition of migration used above we aggregated all households from which the head and spouse (if the household has a spouse) come from a different province than the current province of residence. However, within this group we can distinguish between households from which the head and spouse were born in either the same province or a different province, and households which only have a head and no spouse. This distinction is made as head and spouse who are born in the same province might have migrated together, and head and spouse born in a different province might have migrated separately but met in the current province of residence. Another possibility in the latter case would be that the spouse (77.03 percent of them are female) would move first to the province of residence of the head after marriage, which is the traditional custom in Viet Nam, and afterwards they would move together to the current province of residence. There are a lot of different possibilities, which might indicate heterogeneity across households. This might induce differences in dietary patterns, which is the reason for making this distinction. The food consumption expenditure shares and calorie intake shares of all these groups have been estimated and can be found in table 5.2 below. Also Tukey tests for pairwise comparisons of means have been executed. The results of these tests, which can be found in table D.2 in Appendix D, show that all shares are very similar across groups; only a few significant differences can be found. Furthermore, even these statistically significant estimated differences are still very small in absolute terms. Therefore, in the analysis about the effect of rural-urban migration on the diet of migrant households (see next section) I choose to focus on the 'aggregated migrant group' only. This, moreover, increased the sample size used in the analysis.

Table 5.2: Food expenditure and calorie intake shares per category and per type of rural-urban migration

		RU Migrants *Aggregate*	RU Migrants *H&S born in same province*	RU Migrants *H&S born in different province*	RU Migrants *No spouse*
Rice	**Food expenditure share**	13.88 (0.39)	14.68 (0.66)	13.42 (0.68)	13.39 (0.71)
	Calorie intake share	36.60 (0.67)	38.96 (1.09)	37.60 (1.13)	32.40 (1.26)
Other staples	**Food expenditure share**	1.93 (0.07)	1.97 (0.11)	1.82 (0.12)	2.00 (0.13)
	Calorie intake share	4.05 (0.14)	4.11 (0.23)	3.97 (0.22)	4.07 (0.28)
Meat	**Food expenditure share**	13.52 (0.35)	13.85 (0.45)	13.55 (0.53)	13.04 (0.74)
	Calorie intake share	7.33 (0.19)	7.41 (0.29)	7.62 (0.32)	6.88 (0.37)
Fish and seafood	**Food expenditure share**	11.45 (0.33)	11.83 (0.54)	11.21 (0.53)	11.25 (0.67)
	Calorie intake share	1.95 (0.07)	2.02 (0.12)	1.93 (0.10)	1.89 (0.17)
Eggs	**Food expenditure share**	0.83 (0.03)	0.81 (0.07)	0.74 (0.04)	0.95 (0.06)

	Calorie intake share	0.48 (0.01)	0.46 (0.02)	0.46 (0.02)	0.54 (0.04)
Tofu	**Food expenditure share**	0.57 (0.03)	0.55 (0.04)	0.52 (0.04)	0.67 (0.08)
	Calorie intake share	0.65 (0.03)	0.65 (0.04)	0.62 (0.04)	0.68 (0.07)
Vegetables	**Food expenditure share**	3.82 (0.11)	3.66 (0.15)	3.93 (0.22)	3.90 (0.21)
	Calorie intake share	2.23 (0.07)	2.19 (0.10)	2.19 (0.10)	2.71 (0.18)
Fruits	**Food expenditure share**	3.04 (0.01)	2.89 (0.15)	3.12 (0.17)	3.14 (0.18)
	Calorie intake share	3.12 (0.14)	2.99 (0.21)	3.22 (0.27)	3.15 (0.24)
Dairy	**Food expenditure share**	6.47 (0.40)	6.14 (0.56)	8.19 (0.83)	4.88 (0.60)
	Calorie intake share	1.91 (0.10)	1.84 (0.15)	2.25 (0.20)	1.60 (0.18)
Food away from home	**Food expenditure share**	35.45 (0.96)	34.22 (1.47)	35.09 (1.67)	37.46 (1.89)
	Calorie intake share	29.16 (0.93)	26.73 (1.43)	28.47 (1.46)	33.10 (1.95)
Other	**Food expenditure share**	9.04 (0.23)	9.40 (0.35)	8.41 (0.32)	9.31 (0.53)
	Calorie intake share	12.41 (0.32)	12.64 (0.55)	11.68 (0.42)	12.97 (0.68)

Notes: Standard errors in parentheses. The definition of rural-urban migrants is based on the province of birth of the head of the household and the spouse. A distinction is made between (1) all identified rural-urban migrant households (N= 520), (2) households from which the head and spouse are born in the same province (N= 194), (3) households from which the head and spouse are born in a different province (N= 176), and (4) households which do not have a spouse (N= 151).
Source: Author's calculations based on the VHLSS of 2014.

Two other indicators estimated in this thesis are the Household Dietary Diversity Score (HDDS) and Food Variety Score (FVS); the scores range from 0-12 and 0-53 respectively. These indicators are estimated as a more diversified and varied diet is argued to improve nutrient adequacy (FAO, 2011b; Hodgson et al., 2010; Kuyper et al., 2017; NIN, 2012; Swindale & Bilinsky, 2006). Furthermore, the HDDS is often used as an indicator for food access (FAO, 2011b). The estimates of the HDDS and FVS, together with the percentage of households meeting the fruit and vegetable intake recommendations of the World Health Organization (WHO) (FV), are shown in table 5.3 below for different types of households. It should be noted, however, that in general the computation of the HDDS is based on a 24-hour recall (FAO, 2011b; Swindale & Bilinsky, 2006), and the FVS on a seven days recall (Hodgson et al., 2010). As our data is based on a 30 day recall this imposes some challenges for the interpretation of the indicators and for comparing the estimated HDDS and FVS across countries, as obviously more food groups and food types will be consumed the longer the recall period (i.e. the figures will inflate). Nevertheless, we can use the estimates to compare across different types of households. Table 5.3 shows that the HDDS is more or less similar across groups; it ranges from 10 food groups for rural households to 11 for urban and rural-urban migrant households. The only significant differences in HDDS are found for rural households compared to urban and migrant households (p= 0.000 in both cases). The least frequently consumed food group is white roots and tubers (see figure 5.2 below); only 35.51

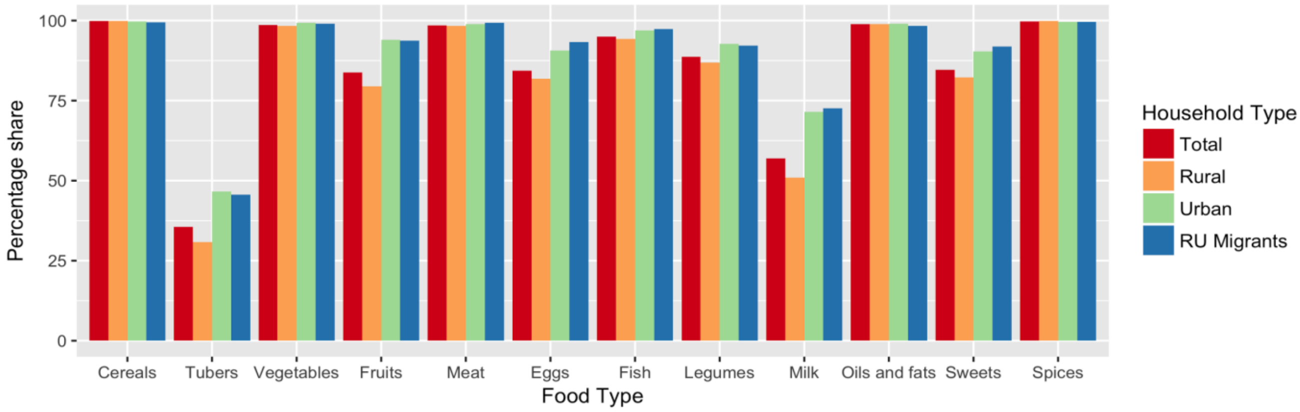
percent of all households reported to have consumed any food falling into this category. Furthermore, only 56.88 percent of all households reported to have consumed milk and milk products. The percentage of households consuming these two food groups is larger for urban and rural-urban migrant households than for rural households. As the HDDS is an indicator for economic access the findings show that in general urban markets provide a higher variety of foods. It can thus be argued that urban food markets are more adequately supplied and easily accessible than rural food markets (FAO, 2011b), although the differences in HDDS between rural and urban households are not too large. This is in line with findings from the regional workshop on food consumption, urbanisation and rural transformations held in Hanoi (2015) that urban markets now constitute about 60 to 70 percent of the food consumed in Asia.

Table 5.3: HDDS, FVS, and FV per type of household

Indicator	Total	Rural	Urban	RU Migrants
HDDS	10.24 (1.51)	10.01 (1.52)	10.78 (1.31)	10.82 (1.34)
FVS	30.01 (8.44)	28.41 (7.92)	33.69 (8.35)	34.44 (8.62)
FV (%)	1.11 (10.48)	1.00 (9.94)	1.44 (11.90)	1.15 (10.68)

Notes: Standard deviation in parentheses. HDDS means Household Dietary Diversity Score, FVS means Food Variety Score, and FV (based on adult equivalence scale) is an indicator for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day. The sample size for 'Total' equals 9.093, 6.415 for 'Rural', 2.157 for 'Urban' (excl. RU Migrants), and 521 for 'RU Migrants'.
Source: Author's calculations based on the VHLSS of 2014.

Figure 5.2: Percentage of households consuming HDDS food groups



Notes: Food types are abbreviations of food groups corresponding with Household Dietary Diversity Score (HDDS). 'Tubers' capture white tubers and roots, 'Fish' captures fish and other seafood, 'Legumes' correspond with legumes, nuts and seeds, 'Milk' corresponds with milk and milk products, and 'Spices' correspond with spices, condiments and beverages. The sample size for 'Total' equals 9.093, 6.415 for 'Rural', 2.157 for 'Urban' (excl. RU Migrants), and 521 for 'RU Migrants'.
Source: Author's calculations based on the VHLSS of 2014.

Table 5.3 shows that the FVS ranges from 28 food types for rural households, to 34 for rural-urban migrant households. Significant differences are found, again, for rural households compared to urban and migrant households (p= 0.000 in both cases). It is predicted that rural households consume a diet of lower variety than urban and rural-urban migrant households.

Vegetables and fruits are important sources of vitamins and are recommended to be consumed on a daily basis (NIN, n.d., a). On average 98.56 percent of the Vietnamese population is estimated to consume vegetables, which is 83.72 percent for fruits. The percentage of rural, urban and rural-urban migrant households consuming vegetables is more or less similar, although we do see differences in shares for the consumption of fruits (see figure 5.2 above); a smaller share of rural households is estimated to consume fruits (79.42 percent) compared to urban households (94.03 percent) and rural-urban migrant households (93.67 percent). These percentages show, thus, that on average a high proportion of Vietnamese households is predicted to have had expenditures on vegetables and fruits over the last 30 days at timing of survey. However, they do not indicate whether these households meet the daily intake recommendations of the National Institute of Nutrition (NIN, n.d., a). The World Health Organization recommends an average adult to consume more than 400 grams of fruits and vegetables a day (WHO, 2015), and the percentage of households meeting these recommendations is shown in table 5.3 above (these percentages are based on adult equivalence scales, see chapter 3). We see that this percentage is extremely small for all households; it is estimated to be at largest 1.44 percent for urban households. The average number of consumed grams of fruits and vegetables per adult per day is estimated to range from 101.13 grams for rural households to 125.99 grams for rural-urban migrant households. The estimations show that, even though most of Vietnamese households have expenditures on fruits and vegetables, only a small proportion of those households consumes a quantity of fruits and vegetables sufficiently large to meet the dietary recommendations of the WHO[8].

5.3 The Effect of Rural-Urban Migration on Diet

In the theoretical framework a model of dietary determinants and dietary acculturation were proposed. The former model, which was based on the European Food Information Council (EUFIC, 2006), showed the importance of physical and economic factors in determining food choice. Examples of factors are prices, availability, marketing, accessibility, geography, and climate. Furthermore, the model of dietary acculturation stressed the influence of environmental factors on food procurement and preparation, which in turn influence dietary intake. Both models assume that the food environment in which one lives influences the diet that one consumes. This assumption triggers the analysis of the effect of rural-urban migration on diets of migrants, as migrants enter a new food environment after migration. The previous section showed the cross-sectional analysis of diets across different groups in the population. Although we saw some significant differences between the food expenditure shares and calorie intake shares of rural-urban migrant households and rural and urban households, this does not allow us to make causal inferences about the effect of migration on diet. As an econometric approach propensity score matching has been executed.

Rural-urban migrant households were matched with comparable rural non-migrant households (based on observable characteristics) by means of radius caliper matching for estimating the average treatment effect on the treated (ATT). In order to be able to implement Rosenbaum Bounds as a sensitivity analysis to hidden bias, ATTs have been estimated using 1x1 nearest neighbour matching as well. Therefore estimation results using both matching algorithms are presented in this section, although the former matching procedure has been much more successful in achieving balance and therefore in the analysis I focus on the estimation results of radius caliper matching only. As described in Appendix C, propensity score matching has been implemented on a subsample of the population, namely the Red River Delta (RRD) (194 migrant households were identified as coming from the RRD).

Different food consumption indicators have been identified, and the estimation results on food expenditure shares and calorie intake shares are shown in table 5.4 below. First of all, both radius caliper and 1x1 nearest neighbour matching yield similar (significant) signs of predicted coefficients. The estimation results show that migration is predicted to negatively and significantly (at the 1 percent level) influence both the food expenditure and calorie intake share of rice. This is

[8] Even when we slightly adapt the calculation of the adult equivalence scale by multiplying the number of children by 0.5 instead of 0.8, the percentage of households meeting the WHO recommendations is estimated to be at largest 1.57 percent for urban households. The average number of consumed grams of fruits and vegetables per adult per day is estimated to range from 109.93 grams for rural households to 134.68 grams for rural-urban migrant households in this case.

not too surprising, given the cross-sectional analysis we saw in section 5.2 (i.e. rural households had much larger expenditure and calorie intake shares for rice than urban and rural-urban migrant households). Furthermore, also the ATT on the expenditure share of other staples (like potatoes and noodles) and meat is negative and significant (at the 5 and 1 percent respectively). Although it is predicted that there is no significant impact on the expenditure share of fish and seafood, we do see a positive and significant effect on the number of calories from fish and seafood consumed. Furthermore, it is predicted that households reduce their expenditure share on eggs when they move to urban areas. Also it is predicted that both the food expenditure and calorie intake share of tofu decrease as a consequence of rural-urban migration. The number of calories from dairy products is predicted to increase, as well as both the food away from home (FAFH) expenditure share and calorie intake share. The expenditure share on the category of 'other' is negatively and significantly associated with rural-urban migration.

Due to limitations associated with the data at hand (e.g. 30 days recall) and the implemented estimation strategy, we should be careful when interpreting the magnitude of the significant coefficients. Nevertheless it is interesting to look at the predicted largest significant effects, which are for the categories of rice, and FAFH (for both food expenditure and calorie intake shares). It is predicted that rural-urban migration decreases the expenditure share of rice with 6.34 percent points, and increases the expenditure share of FAFH with 10.37 percent points. For calorie intake shares the effects are -7.44 and +6.29 percent points for rice and FAFH respectively. Also worth mentioning is the ATT on the expenditure share of meat: it is predicted that the meat expenditure share decreases with 3.88 percent points when households move to urban areas. The magnitude of all other significant ATTs is relatively small. Summarizing, the PSM model predicts that rural-urban migration has the largest consequences for the consumption of rice and FAFH.

Table 5.4: ATT of rural- urban migration on food expenditure and calorie intake shares

Share	Food expenditure shares		Calorie intake shares	
	Radius caliper matching	1x1 Nearest neighbour matching	Radius caliper matching	1x1 Nearest neighbour matching
Rice	-6.34*** (1.66)	-6.49*** (2.03)	-7.44*** (2.51)	-7.29** (3.02)
Other staples	-0.54** (0.24)	-0.42 (0.33)	-0.04 (0.47)	-0.11 (0.63)
Meat	-3.88*** (1.44)	-4.00** (1.76)	-1.00 (0.0075)	-1.27 (0.92)
Fish and seafood	1.40 (1.08)	2.62** (1.29)	0.46** (0.23)	0.61** (0.28)
Eggs	-0.47*** (0.12)	-0.31* (0.17)	-0.06 (0.05)	-0.02 (0.07)
Tofu	-0.65*** (0.12)	-0.59*** (0.17)	-0.55*** (0.14)	-0.46** (0.19)
Vegetables	0.32 (0.37)	0.76 (0.47)	-0.06 (0.23)	0.12 (0.29)
Fruits	-0.04 (0.32)	-0.03 (0.40)	0.16 (0.36)	0.17 (0.45)
Dairy	2.31* (1.37)	2.09 (1.62)	1.00*** (0.33)	0.98*** (0.36)
Food away from home	10.37*** (3.01)	8.39** (3.87)	6.29** (2.56)	5.77* (3.33)

Other	-2.48*** (0.73)	-2.11** (0.91)	1.24 (1.00)	1.48 (1.11)

Notes: For the estimation of the ATTs rural-urban migrant households from which the head of the household is born in the Red River Delta are matched with rural non-migrant households living in the Red River Delta. In the case of radius caliper matching, 96 rural-urban migrant households were matched with 1.041 rural non-migrant households. In the case of 1x1 nearest neighbour matching, 96 rural-urban migrant households were matched with 96 rural non-migrant households.

*** p<0.01, ** p<0.05, * p<0.1
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200).

Interesting is the fact that no significant effects are found on the shares for vegetables and fruits (see table 5.4 above). On average for the whole country the expenditure share equals 4.01 and 2.48 percent for vegetables and fruits respectively, whereas the calorie intake share equals 2.34 and 2.51 percent respectively. These shares are, thus, quite low. Furthermore, recall from section 4.2 that the expenditure share for vegetables and fruits has been quite stable over the period 2002 to 2012. Given that only a very small proportion of the population meets the daily fruits and vegetables intake recommendation of the WHO (see table 5.3), the finding that rural-urban migration does neither significantly influence these shares positively nor negatively provides some further insights into the possibly alarming situation of the Vietnamese population in general consuming not enough vegetables and fruits.

Furthermore, the ATT has been estimated for the Household Dietary Diversity Score (HDDS), Food Variety Score (FVS), and a dummy for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day (FV). The estimation results for both radius caliper and nearest neighbour matching are shown in table 5.5 below. The signs of significant predicted coefficients are similar for both matching algorithms, although nearest neighbour matching does not find any significant effects whereas radius caliper matching does. However, as described above, I focus on the estimation results of radius caliper matching. The table shows that moving from a rural to an urban area is positively and significantly associated with the household's economic access to food at the 5 percent level. The coefficient predicts that the score increases with 0.44 food groups as a consequence of migration. This is in line with the argument made in the previous section, namely that urban markets in general are more adequately supplied and easily accessible than rural food markets (FAO, 2011b). Furthermore, migration is associated with a positive and significant increase in the Food Variety Score as well; it is predicted that rural-urban migrant households consume 2.91 more food items than their comparable rural non-migrant counterparts. No significant impact is found on whether or not the household meets the fruit and vegetable recommendations of the WHO. The bottom line of table 5.5 is that rural-urban migration is predicted to increase the household's economic access to food, and the variety of food items consumed. As in this thesis the calculation of the HDDS and FVS is based on 30 days recall instead of 24 hours and 7 days respectively, we should be cautious with interpreting the exact magnitude of the coefficients. Rather it is interesting to emphasize the positive sign of both ATTs.

Table 5.5: ATT of rural-urban migration on HDDS, FVS, and FV

Indicator	Radius caliper matching	1x1 Nearest neighbour matching
HDDS	0.44** (0.17)	0.24 (0.24)
FVS	2.91** (1.31)	2.19 (1.54)
FV	-0.71 (0.95)	0.00 (1.74)

Notes: For the estimation of the ATTs rural-urban migrant households from which the head of the household is born in the Red River Delta are matched with rural non-migrant households living in the Red River Delta. HDDS means Household Dietary Diversity Score, FVS means Food Variety Score, and FV *(based on adult equivalence scale) is an indicator for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day. In the case of radius caliper matching, 96 rural-urban migrant households were matched with 1.041 rural non-migrant households. In the case of 1x1 nearest neighbour matching, 96 rural-urban migrant households were matched with 96 rural non-migrant households.*

*** p<0.01, ** p<0.05, * p<0.1
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200).

I conclude this section arguing that rural-urban migration is predicted to increase the food expenditure share of food away from home, whereas it is predicted to decrease the food expenditure share of rice, other staples, meat, eggs, tofu, and other. Furthermore, the calorie intake share of dairy and food away from home is predicted to increase, whereas this share is predicted to decrease for the categories of rice and tofu. Largest effects are seen for the shares of rice and food away from home. It is predicted that household's economic access to food increases when migrating from rural to urban areas, as well as that households start consuming a larger variety of food items. Whether or not a household moves from rural to urban areas is predicted to not significantly influence the consumption of vegetables and fruits.

5.4 The Effect of Having Rural Origins on Diet

The previous section analysed the effect of rural-urban migration on diets of migrants, as it was assumed that moving into a different food environment influences dietary intake. However, the model of dietary determinants in the theoretical framework (chapter 2) also includes food determinants related to someone's personal background - factors like taste, culture, religion, and knowledge and education. This triggered us to test the effect of having rural origins on dietary intake; i.e. does the fact that you have a different background, namely having migrated from a rural to an urban area, influence the diet that you consume? This hypothesis was tested by matching rural-urban migrant households to urban non-migrant households living in (a) the Red River Delta and (b) the South Eastern Area (SEA) of Viet Nam (see Appendix C for details about the matching exercises). For the first subsample 66 migrant households could be identified, whereas this was 203 for the latter subsample (N.B.: from all 194 migrant households coming from the RRD, 42 moved to urban areas within the RRD, whereas 50 moved to urban areas in the SEA). Balancing tests in Appendix C show that the matching exercise has been much more successful for the SEA compared to the RRD in achieving balance across groups, which is why I mainly focus on the ATTs of migrant households living in the SEA. Similar to the analysis in the previous section, ATT have been estimated using both radius caliper matching and 1x1 nearest neighbour matching. However, I focus on the estimation results of radius caliper matching, as again this matching algorithm yields much better balance across covariates.

The ATTs for food expenditure shares and calorie intake shares are shown in table 5.6 below. First of all, both radius caliper and 1x1 nearest neighbour matching yield similar (significant) signs of predicted coefficients. Estimations for the RRD show that, at the 5 percent level, the expenditure share and calorie intake share of fish and seafood is predicted to be lower for rural-urban migrant households than for comparable urban non-migrant households. For the SEA, however, no significant effects are found for all categories for both food expenditure shares and calorie intake shares. As the balancing tests for the RRD were not too convincing in arguing that balance has been achieved and as the migrant sample within the RRD is quite small (see Appendix C), we should be cautious when interpreting the significant expenditure and calorie intake share of fish and seafood in the RRD.

Table 5.6: ATT of having rural origins on food expenditure and calorie intake shares

	Food expenditure shares				Calorie intake shares			
	Radius caliper matching		1x1 Nearest neighbour matching		Radius caliper matching		1x1 Nearest neighbour matching	
	RRD	SEA	RRD	SEA	RRD	SEA	RRD	SEA
Rice	-4.93* (2.64)	-0.71 (1.83)	-5.74* (2.97)	-0.66 (2.06)	-6.67* (3.92)	0.18 (3.24)	-5.15 (4.68)	-1.31 (3.48)
Other staples	-0.80* (0.48)	-0.19 (0.41)	-0.85 (0.59)	-0.21 (0.41)	-0.88 (1.13)	-0.05 (0.68)	-0.39 (1.44)	-0.18 (0.74)
Meat	-2.24 (2.13)	0.97 (1.69)	-0.82 (2.36)	-0.22 (1.77)	-0.27 (1.43)	0.64 (1.04)	0.56 (1.41)	-0.20 (0.99)
Fish and seafood	-4.75** (2.18)	0.91 (1.65)	-3.52 (2.60)	-0.30 (1.76)	-0.66** (0.31)	0.26 (0.41)	-0.49 (0.46)	0.05 (0.40)
Eggs	-0.40* (0.23)	-0.08 (0.12)	-0.43 (0.27)	-0.09 (0.13)	-0.14 (0.0010)	-0.04 (0.08)	-0.17 (0.11)	-0.07 (0.08)
Tofu	-0.14 (0.18)	-0.10 (0.17)	-0.24 (0.22)	-0.20 (0.18)	-0.10 (0.26)	-0.15 (0.13)	-0.05 (0.30)	-0.22 (0.13)
Vegetables	-0.21 (0.62)	-0.08 (0.66)	-0.13 (0.72)	-0.33 (0.74)	-0.35 (0.45)	-0.01 (0.36)	-0.78 (0.55)	-0.04 (0.35)
Fruits	-0.44 (0.79)	0.21 (0.54)	-0.23 (0.89)	-0.05 (0.54)	0.12 (0.74)	-0.23 (0.85)	0.31 (0.75)	-0.14 (0.96)
Dairy	3.37 (2.80)	-0.57 (2.36)	3.33 (3.40)	1.48 (2.30)	0.52 (0.73)	-0.02 (0.56)	0.43 (0.87)	0.33 (0.61)
Food away from home	10.84* (6.12)	-0.31 (4.79)	9.03 (7.35)	2.19 (5.87)	8.07 (5.86)	0.73 (4.82)	5.78 (6.29)	5.68 (4.98)
Other	-0.31 (1.62)	-0.04 (1.45)	-0.40 (1.80)	-1.62 (1.40)	0.13 (1.80)	-1.30 (2.04)	-0.05 (1.20)	-3.88 (2.08)

Notes: For the estimation of the ATTs rural-urban migrant households are matched with urban non-migrant households living in (a) the Red River Delta (RRD) and (b) South Eastern Area (SEA). In the case of radius caliper matching in the RRD, 28 rural-urban migrant households were matched with 390 urban non-migrant households. In the case of radius caliper matching in SEA, 80 rural-urban migrant households were matched with 251 urban non-migrant households. In the case of 1x1 nearest neighbour matching in the RRD, 28 rural-urban migrant households were matched with 28 urban non-migrant households. In the case of 1x1 nearest neighbour matching in SEA, 80 rural-urban migrant households were matched with 80 urban non-migrant households.

*** p<0.01, ** p<0.05, * p<0.1
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200).

The ATT has been estimated for the HDDS, FVS, and a dummy for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day (FV), for which the estimation results of both radius caliper and 1x1 nearest neighbour matching are shown in table 5.7 below. No significant effects are found for rural-urban migrant households living either the RRD or the SEA. It is thus predicted that having a rural-urban migrant background compared to being born in an urban area does not significantly influence economic access to food (HDDS), which was to be expected given that living in a similar food environment was controlled for. Furthermore for both regions it is predicted that there are no significant effects for the variety of consumed food items (FVS), and whether or not a household meets the WHO's fruit and vegetable recommendations (FV).

Table 5.7: ATT of having rural origins on HDDS, FVS and FV

	Radius caliper matching		1x1 Nearest neighbour matching	
	RRD	SEA	RRD	SEA
HDDS	0.36 (0.32)	0.10 (0.24)	0.21 (0.38)	0.17) (0.25)
FVS	5.16 (3.30)	1.80 (1.93)	4.79 (3.70)	1.38 (1.91)
FV	-0.71 (1.87)	-1.25 (1.58)	-3.57 (2.5)	-1.41 (2.21)

Notes: For the estimation of the ATTs rural-urban migrant households are matched with urban non-migrant households living in (a) the Red River Delta (RRD) and (b) South Eastern Area (SEA). HDDS means Household Dietary Diversity Score, FVS means Food Variety Score, and FV (based on adult equivalence scale) is an indicator for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day. In the case of radius caliper matching in the RRD, 32 rural-urban migrant households were matched with 498 urban non-migrant households. In the case of radius caliper matching in the RRD, 28 rural-urban migrant households were matched with 390 urban non-migrant households. In the case of radius caliper matching in SEA, 80 rural-urban migrant households were matched with 251 urban non-migrant households. In the case of 1x1 nearest neighbour matching in the RRD, 28 rural-urban migrant households were matched with 32 urban non-migrant households. In the case of 1x1 nearest neighbour matching in SEA, 80 rural-urban migrant households were matched with 80 urban non-migrant households.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200).

Concluding this section I argue that no significant effects are found for the effect of having rural origins compared to being born in urban areas on any of the food consumption indicators of rural-urban migrant households. The only significant effects that were found were for the expenditure and calorie intake share of fish and seafood for migrants living in the RRD, but as the migrant sample in the RRD is considerably small and as the matching exercise was not too successful in creating a balanced sample, I argue that the mentioned significant effects are neglectable. It is therefore predicted that rural-urban migrant households have similar diets as comparable urban non-migrant households living in the same food environment.

5.5 Robustness Checks and Sensitivity Anaylsis

In chapter 3 I argued that although Ordinary Least Squares (OLS) estimates will probably be biased when not controlling for the endogeneity of migration, I run the OLS model for checking the robustness of the estimated PSM treatment effects. Here I specifically look at the (significant) sign of the OLS estimates, rather than the magnitude of the coefficients. Recall that the OLS estimator is a particular sort of weighted matching estimator (Angrist & Pischke, 2008).

5.5.1 OLS Estimation of the Effect of Rural-Urban Migration on Diet

Table 5.8 below shows the estimated OLS regression results for the effect of rural-urban migration on food expenditure shares and calorie intake shares. Both models have been estimated with and without control variables. The models with control variables are similar to the PSM model in the sense that they control for the same covariates and build on the same Conditional Independence Assumption (CIA), but apply a different weighting (Angrist & Pischke, 2008). The models without control variables basically are two-sample t-tests for rural-urban migrant households versus rural non-migrant households. For all estimations the R-squared for those models without covariates is much smaller than the R-squared for those models with control variables. However it should be noted that even when control variables are included, the variance in dependent variable explained by the variance in independent variables remains rather small (this holds for table 5.8 – 5.11).

When looking at the estimates for food expenditure shares, the sign of all significant estimates is similar as the sign of the estimates of the PSM model for both radius caliper and 1x1 nearest neighbour matching (see table 5.4), regardless of whether control variables are included or not. Furthermore, the OLS models (both with and without covariates) predict significant effects at at least the 5 percent level for the shares of rice (-), other staples (-), meat (-), eggs (-), tofu (-), dairy (+), FAFH (+), and other (-). This is again in line with the prediction results of the radius caliper PSM model, although the share of dairy was only significant at the 10 percent level in the PSM model (see table 5.4). Although the interpretation of the sign of the OLS estimates is the interest rather than the interpretation of their magnitude, the OLS model, like the PSM model, predicts largest significant effects for the food expenditure shares of rice and food away from home.

If we turn to calorie intake shares we see that the sign of all significant estimates is similar as the sign of the estimates of the radius caliper PSM model (see table 5.4). Two-sample t-tests show that the calorie intake share of rice, fish and seafood, tofu, fruits, dairy, FAFH and other differs significantly for rural-urban migrant households and rural non-migrant households. However, when looking at the OLS model with controls, significant effects at the 5 percent level were found for the categories of rice (-), meat (-), eggs (-), tofu (-), dairy (+) and FAFH (+). As the PSM model predicts significant effects for the same categories, minus the categories of meat and eggs and plus the category of fish and seafood, I argue that the significant ATTs produced by the PSM model are all robust to putting higher weights on people who are different – except for the ATT of fish and seafood, which is only significant at the 10 percent level in the OLS model. Similar to the PSM model, compared to all other categories the OLS model predicts largest significant effects for the calorie intake shares of rice and food away from home.

Table 5.8: OLS regression results of rural-urban migration on food expenditure and calorie intake shares as dependent variable

		Food expenditure shares		Calorie intake shares	
Dependent variable	**Controls**	**RU Migration**	**R^2**	**RU Migration**	**R^2**
Rice	NO	-9.48*** (0.76)	0.0676	-1.83*** (1.17)	0.0505
	YES	-6.98*** (0.94)	0.1293	-7.49*** (1.39)	0.0962
Other staples	NO	-0.68*** (0.13)	0.0120	0.05 (0.23)	0.0000
	YES	-0.45*** (0.15)	0.0291	-0.03 (0.28)	0.0225
Meat	NO	-3.36*** (0.75)	0.0170	-0.44 (0.37)	0.0011
	YES	-3.54*** (0.80)	0.0411	-0.93** (0.41)	0.0237
Fish and seafood	NO	1.01* (0.62)	0.0022	0.33*** (0.12)	0.0075
	YES	0.56 (0.67)	0.0282	0.27* (0.16)	0.0299
Eggs	NO	-0.53***	0.0202	-0.05*	0.0016

		(0.08)		(0.03)	
	YES	-0.49*** (0.07)	0.0774	-0.08** (0.03)	0.0560
Tofu	NO	-0.73*** (0.06)	0.0422	-0.58*** (0.06)	0.0258
	YES	-0.71*** (0.07)	0.0850	-0.70*** (0.09)	0.0537
Vegetables	NO	0.13 (0.18)	0.0003	-0.07 (0.14)	0.0002
	YES	0.33 (0.25)	0.0655	-0.14 (0.15)	0.0567
Fruits	NO	-0.31* (0.18)	0.0019	0.51** (0.24)	0.0031
	YES	-0.03 (0.21)	0.0353	0.38 (0.27)	0.0243
Dairy	NO	2.14*** (0.69)	0.0071	1.06*** (0.18)	0.0429
	YES	1.69** (0.76)	0.1330	0.78*** (0.19)	0.1813
Food away from home	NO	13.25*** (1.66)	0.0446	8.55*** (1.55)	0.0235
	YES	11.71*** (1.84)	0.0964	6.91*** (1.63)	0.0613
Other	NO	-2.06*** (0.47)	0.0168	1.48*** (0.51)	0.0079
	YES	-2.10*** (0.49)	0.0557	1.03 (0.71)	0.0274

Notes: Included covariates are ethnicity, gender of the household head, age of the household head and spouse, age squared of the household head and spouse, age cubed of the household head and spouse, and four dummy variables for obtained level of education of the household head and spouse. N= 1.539 in the case without covariates (including 194 rural-urban migrant households); N= 1.182 in the case with covariates (including 141 rural-urban migrant households).

*** p<0.01, ** p<0.05, * p<0.1
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200)

The OLS estimation results for the effect of rural-urban migration on the HDDS, FVS, and a dummy for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day (FV) are presented in table 5.9 below. The OLS model predicts positive and significant effects for the HDDS and FVS (both for the model with and without covariates), and negative significant effects for the dummy variable FV. This is partly in line with the PSM estimation results: the radius caliper PSM model has predicted positive and significant effects for the HDDS and FVS, but insignificant effects for FV (see table 5.5). However, the sign of all significant and insignificant estimates are similar for both the PSM and OLS model. I argue that the significant ATTs of the radius caliper PSM model are robust to applying the OLS weighting.

Table 5.9: OLS regression results of rural-urban migration on HDDS, FVS, and FV as dependent variable

Dependent variable	Controls	RU Migration	R^2
HDDS	NO	0.64*** (0.10)	0.0256
	YES	0.54*** (0.10)	0.0989
FVS	NO	5.10*** (0.64)	0.0435
	YES	4.13*** (0.76)	0.1208
FV	NO	-1.34** (0.65)	0.0012
	YES	-0.90*** (0.35)	0.0135

Notes: Included covariates are ethnicity, gender of the household head, age of the household head and spouse, age squared of the household head and spouse, age cubed of the household head and spouse, and four dummy variables for obtained level of education of the household head and spouse. HDDS means Household Dietary Diversity Score, FVS means Food Variety Score, and FV (based on adult equivalence scale) is an indicator for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day. N= 1.539 in the case without covariates (including 194 rural-urban migrant households); N= 1.182 in the case with covariates (including 141 rural-urban migrant households).

*** p<0.01, ** p<0.05, * p<0.1
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200).

I conclude this subsection by arguing that both the PSM and OLS models predict similar signs for the significant effects of rural-urban migration on food expenditure shares, calorie intake shares, the HDDS, FVS, and whether or not households meet the recommendations of the WHO for fruit and vegetable consumption. Furthermore, all estimated significant ATTs of the radius caliper PSM model are robust to putting higher weights on people who are different instead of on people who are similar.

Altonji et al. (2005) and Bellows and Miguel (2009) have developed an approach to assess the importance of selection bias in OLS models. The approach, and its implementation on the results from this subsection, can be found in Appendix F.

5.5.1 OLS Estimation of the Effect of Having Rural Origins on Diet

The OLS model has been run as well for estimating the effect of having rural origins on food expenditure shares and calorie intake shares. Like the estimations using PSM, estimations have been done for two sub-samples: (a) the Red River Delta (RRD) and (b) South Eastern Area (SEA) of Viet Nam. The estimation results are shown in table 5.10 below. Recall from section 5.4 that no significant ATTs were found for any of the food consumption indicators of rural-urban migrant households living in the SEA. Significant negative ATTs were predicted to exist for the food expenditure and calorie intake share of fish and seafood for rural-urban migrant households living in the RRD. When looking at the estimates for food expenditure shares, the sign of all significant estimates is similar as the sign of the estimates of the PSM model for both radius caliper and 1x1 nearest neighbour matching (see table 5.6), regardless of whether control variables are included or not. Furthermore, two-sample t-tests (i.e. the OLS models without covariates) estimate that the food expenditure shares of rice, other staples, meat, eggs, tofu, dairy, FAFH, and other differ significantly for rural-urban migrant households and urban non-migrant households living in the RRD. For households living in the SEA this is only the case for the food expenditure share of rice and FAFH. Furthermore, the OLS model predicts significant effects at at least the 5 percent level for the expenditure shares of rice (-), other staples (-), meat (-), eggs (-), tofu (-), dairy (+), FAFH (+), and other (-) for being rural-urban migrant compared to being born in urban areas in the RRD. No significant effects on food expenditure shares were found for households living in the SEA. Therefore the OLS and PSM model produce similar results for the SEA. For the RRD, however, although signs of significant estimates are similar across both models, both models estimate significant effects for different categories.

When turning to calorie intake shares we see, again, similar signs of significant effects for the OLS and PSM estimation methods. The OLS estimation predicts significant effects for the shares of rice (-), dairy (+), and FAFH (+) for being rural-urban migrant households compared to being born in urban areas in the RRD; no significant effects are found for the SEA. Therefore the OLS and PSM model produce similar results for the SEA. For the RRD, however, although signs of significant estimates are similar across both models, both models estimate significant effects for different categories.

Table 5.10: OLS regression results having rural origins on food expenditure and calorie intake shares as dependent variable

		RRD				SEA			
		Food expenditure shares		Calorie intake shares		Food expenditure shares		Calorie intake shares	
Dependent variable	Controls	RU Migration	R^2	RU Migration	R^2	RU Migration	R^2	RU Migration	R^2
Rice	NO	-5.16*** (0.89)	0.0329	-6.45*** (1.52)	0.0239	-1.68*** (0.64)	0.0103	-1.84 (1.26)	0.0037
	YES	-4.13*** (1.22)	0.1655	-4.61** (1.82)	0.1648	-0.57 (0.82)	0.1567	0.57 (1.49)	0.1517
Other staples	NO	-0.63*** (0.14)	0.0143	-0.09 (0.37)	0.0001	-0.09 (0.15)	0.0006	0.13 (0.28)	0.0004
	YES	-0.63*** (0.20)	0.1104	-0.39 (0.49)	0.0528	-0.15 (0.16)	0.0947	-0.04 (0.31)	0.0624
Meat	NO	-2.76*** (0.81)	0.0165	-0.19 (0.52)	0.0002	-1.00* (0.57)	0.0049	-0.29 (0.35)	0.0012
	YES	-2.98*** (0.88)	0.0805	-0.41 (0.64)	0.0428	0.09 (0.71)	0.0454	0.26 (0.49)	0.0454
Fish and seafood	NO	-1.02 (0.89)	0.0020	-0.34*** (0.10)	0.0074	-0.14 (0.56)	0.0001	0.05 (0.13)	0.0003
	YES	-1.40 (0.97)	0.0655	-0.29* (0.15)	0.0403	0.48 (0.70)	0.0377	0.20 (0.18)	0.0487
Eggs	NO	-0.23*** (0.08)	0.0075	0.02 (0.05)	0.0002	-0.07 (0.05)	0.0039	-0.04 (0.03)	0.0021
	YES	-0.27*** (0.09)	0.0896	-0.05 (0.04)	0.0451	-0.04 (0.05)	0.1108	-0.05 (0.03)	0.0466
Tofu	NO	-0.34*** (0.07)	0.0170	-0.12 (0.10)	0.0017	-0.07 (0.07)	0.0018	-0.07* (0.04)	0.0040

	YES	-0.25*** (0.07)	0.1435	-0.12 (0.11)	0.0919	-0.04 (0.07)	0.0673	-0.08 (0.05)	0.0314
Vegetables	NO	-0.23 (0.27)	0.0013	-0.14 (0.16)	0.0007	0.04 (0.25)	0.0001	-0.06 (0.14)	0.0003
	YES	-0.12 (0.28)	0.0580	-0.27 (0.20)	0.0611	0.22 (0.29)	0.0553	-0.04 (0.16)	0.0175
Fruits	NO	0.01 (0.25)	0.0000	0.16 (0.26)	0.0003	-0.35* (0.20)	0.0048	-0.42 (0.29)	0.0035
	YES	-0.20 (0.33)	0.0692	0.43 (0.31)	0.0611	-0.19 (0.24)	0.0219	-0.27 (0.37)	0.0475
Dairy	NO	4.37*** (1.36)	0.0262	1.43*** (0.32)	0.0492	-0.62 (0.76)	0.0010	-0.03 (0.18)	0.0000
	YES	3.67*** (1.39)	0.1554	0.90** (0.38)	0.1373	-0.55 (1.05)	0.1272	-0.18 (0.26)	0.1362
Food away from home	NO	7.38*** (2.22)	0.0155	6.11*** (1.81)	0.0135	4.05** (1.81)	0.0080	2.60 (1.81)	0.0034
	YES	7.64*** (2.41)	0.1500	5.48** (2.21)	0.1117	0.91 (2.18)	0.0732	0.53 (2.24)	0.0717
Other	NO	-1.37*** (0.49)	0.0095	-0.38 (0.60)	0.0005	-0.07 (0.45)	0.0000	-0.04 (0.71)	0.0000
	YES	-1.33** (0.55)	0.0746	-0.67 (0.70)	0.0670	-0.14 (0.61)	0.0640	-0.90 (0.95)	0.0389

Notes: Included covariates are ethnicity, gender of the household head, age of the household head and spouse, age squared of the household head and spouse, and four dummy variables for obtained level of education of the household head and spouse. OLS models have been run with the inclusion of province of origin dummies as well; results do not significantly change. For the subsample of the RRD: N= 564 in the case without covariates (including 66 rural-urban migrant households); N= 436 in the case with covariates (including 46 rural-urban migrant households). For the subsample of the SEA: N= 565 in the case without covariates (including 203 rural-urban migrant households); N= 385 in the case with covariates (including 134 rural-urban migrant households).

*** p<0.01, ** p<0.05, * p<0.1
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200).

The OLS model has been estimated for analysing the effect of having rural origins on the HDDS, FVS, and whether or not the household meets the WHO's recommendations on fruit and vegetable intake, for which the estimation results are shown in table 5.11 below. Table 5.11 shows that only significant effects are found for the subsample of the RRD (for both the models with and without control variables); no significant effects are found for the SEA. The PSM models, on the contrary, did not predict any significant effects for the HDDS, FVS, and FV for both subsamples (see table 5.7). However, the sign of all significant estimates predicted by the OLS model is similar to the sign of the insignificant PSM results.

Table 5.11: OLS regression results of having rural origins on HDDS, FVS, and FV as dependent variable

		RRD		SEA	
Dependent variable	**Controls**	**RU Migration**	**R^2**	**RU Migration**	**R^2**
HDDS	NO	0.50*** (0.10)	0.0216	-0.15 (0.13)	0.0024
	YES	0.42*** (0.13)	0.0945	-0.04 (0.11)	0.0348

FVS	NO	5.22*** (1.26)	0.0309	-0.31 (0.64)	0.0004
	YES	4.32*** (1.35)	0.1591	0.33 (0.79)	0.0630
FV	NO	-1.41** (0.57)	0.0017	-0.40 (0.98)	0.0003
	YES	-0.32 (0.36)	0.0148	-0.27 (0.45)	0.0551

Notes: Included covariates are ethnicity, gender of the household head, age of the household head and spouse, age squared of the household head and spouse, and four dummy variables for obtained level of education of the household head and spouse. OLS models have been run with the inclusion of province of origin dummies as well; results do not significantly change. HDDS means Household Dietary Diversity Score, FVS means Food Variety Score, and FV (based on adult equivalence scale) is an indicator for whether or not the household meets the WHO's recommendation of consuming at least 400 grams of fruits and vegetables per adult per day. For the subsample of the RRD: N= 564 in the case without covariates (including 66 rural-urban migrant households); N= 436 in the case with covariates (including 46 rural-urban migrant households). For the subsample of the SEA: N= 565 in the case without covariates (including 203 rural-urban migrant households); N= 385 in the case with covariates (including 134 rural-urban migrant households).

*** p<0.01, ** p<0.05, * p<0.1
Bootstrapped standard error in parentheses (nr. of bootstrap repetitions= 200).

Concluding this subsection I argue that both the PSM and OLS model predict similar signs for the significant effects of being rural-urban migrant households on food expenditure shares, calorie intake shares, the HDDS, FVS, and whether or not households meet the recommendations of the WHO for fruit and vegetable consumption. Furthermore, in case of the SEA both the PSM and OLS do not predict any significant effects, which induces that the estimated ATTs are all still insignificant when applying different weights. In the case of the RRD the PSM and OLS model predict significant effects for different food expenditure and calorie intake shares. Furthermore, unlike the PSM model the OLS model does predicts significant effects for the HDDS, FVS, and FV. However, we should be cautious when interpreting the estimates for the subsample of the RRD (both the estimates of PSM and OLS), as the matching exercise was not too convincing in terms of balance achievement. Furthermore, the number of migrant households within the RRD is considerably smaller than the number of migrant households in the SEA.

Altonji et al. (2005) and Bellows and Miguel (2009) have developed an approach to assess the importance of selection bias in OLS models. The approach, and its implementation on the results from this subsection, can be found in Appendix F.

5.5.3 Rosenbaum Bounds on the Effect of Rural-Urban Migration on Diet

The analysis in this thesis focuses on the effect of rural-urban migration on diets of migrants using the matching algorithm of radius caliper matching. However, Rosenbaum bounds for assessing the sensitivity of estimated average treatment effects on the treated (ATTs) to hidden bias (using the STATA command **rbounds**) cannot be applied to an analysis based on this algorithm (Diprete & Gangl, 2004), which is why they have been applied to the results of nearest neighbour matching instead. Section 5.3 presented the estimation results of the effect of rural-urban migration on diets of migrants. For food expenditure shares nearest neighbour matching predicted significant effects for similar categories as radius caliper matching – although nearest neighbour matching did not predict significant effects for the category of other staples, whereas it did for fish and seafood. Both matching methods predicted significant effects for similar calorie intake shares.

Nearest neighbour matching predicted that, at at least the five percent level, the food expenditure share of rice, meat, tofu, and other is predicted to decrease, whereas the share of fish and seafood and FAFH is predicted to increase. The estimated Rosenbaum bounds for these ATTs are shown in table 5.12 below. The food expenditure share of rice is still significant at the five percent level when gamma equals 1.75, meaning that even if two comparable households in terms of □□ differ in their odds of migrating by a factor of 1.75 (due to unobservables), the ATT of a decreasing rice expenditure share is still significant. For the share of meat this is the case for when gamma equals 1.25. The effect of increased share of fish and seafood after migrating to urban areas is removed, however, when gamma increases above 1.00. Furthermore, the share of tofu is still significant at the five percent level when gamma equals 2.00. Like the share of fish and seafood, the effect of increased share of FAFH after migration is no longer significant when gamma increases above 1.00. Lastly, the share for the category of other is no longer significant at the five percent level when gamma equals 1.50. The estimated ATTs for food expenditure shares are thus sensitive to selection bias to a different degree. The gamma levels indicate that the ATT associated with the expenditure share of rice and tofu are insensitive to selection bias. The ATTs associated with meat, fish and seafood, FAFH and other, however, are sensitive to unobserved selection bias.

Significant ATTs at at least the five percent level were found for the calorie intake shares of rice (-), fish and seafood (+), tofu (-), and dairy (+). The Rosenbaum bounds in table 5.12 below show that the effect of decreased share of calories from rice and increased share of calories from dairy is still significant when gamma equals 1.50. Similar to food expenditure shares, the effect of increased share of calories from fish and seafood is removed when gamma increases above 1.00. Furthermore, the ATT of decreased calorie intake share of tofu is still significant at the five percent level when gamma equals 2.00. Again, all estimated ATTs are sensitive to selection bias to a different degree. For the shares of rice, tofu and dairy ATTs are (quite) insensitive to selection bias, whereas the ATT associated with fish and seafood is sensitive to unobserved selection bias.

Table 5.12: Rosenbaum bounds on the ATT of rural-urban migration on food expenditure and calorie intake shares

		Food expenditure shares		Calorie intake shares	
	gamma	sig+	sig-	sig+	sig-
Rice	**1.00**	0.00004	0.00004	0.0003	0.0003
	1.25	4.4e-07	0.001279	6.0e-06	0.0066
	1.50	4.2e-09	0.011075	9.3e-08	0.0410
	1.75	3.8e-11	0.045822	1.3e-09	0.1296
	2.00	3.2e-13	0.120113	1.8e-11	0.2727
Other staples	**1.00**	0.009486	0.009486	0.4347	0.4347
	1.25	0.000463	0.079337	0.1321	0.7840
	1.50	0.000018	0.255297	0.0284	0.9416
	1.75	6.2e-07	0.489256	0.0049	0.9875
	2.00	1.9e-08	0.698253	0.0007	0.9977
Meat	**1.00**	0.001632	0.001632	0.0440	0.0440
	1.25	0.000046	0.022273	0.0038	0.2218
	1.50	1.1e-06	0.102784	0.0003	0.4979

	1.75	2.3e-08	0.259121	0.0000	0.7382
	2.00	4.5e-10	0.455252	7.3e-07	0.8845
Fish and seafood	**1.00**	0.008096	0.008096	0.0094	0.0094
	1.25	0.071001	0.000375	0.0788	0.0005
	1.50	0.236506	0.000014	0.2541	0.0000
	1.75	0.465055	4.5e-07	0.4877	6.1e-07
	2.00	0.676296	1.4e-08	0.6969	1.9e-08
Eggs	**1.00**	0.000924	0.000924	0.3808	0.3808
	1.25	0.000022	0.014518	0.1047	0.7404
	1.50	4.5e-07	0.074452	0.0205	0.9229
	1.75	8.3e-09	0.203919	0.0032	0.9818
	2.00	1.4e-10	0.382677	0.0004	0.9963
Tofu	**1.00**	1.9e-08	1.9e-08	3.3e-06	3.3e-06
	1.25	4.3e-11	2.3e-06	2.1e-08	0.00017
	1.50	9.7e-14	0.000052	1.2e-10	0.002104
	1.75	2.2e-16	0.000469	6.5e-13	0.011551
	2.00	0	0.002331	3.4e-15	0.038253
Vegetables	**1.00**	0.024648	0.024648	0.475232	0.475232
	1.25	0.15218	0.001698	0.812898	0.155375
	1.50	0.39357	0.000091	0.952849	0.035935
	1.75	0.643597	4.2e-06	0.990537	0.00664
	2.00	0.821862	1.8e-07	0.998369	0.001057
Fruits	**1.00**	0.619172	0.619172	0.18629	0.18629
	1.25	0.259159	0.895503	0.521505	0.032276
	1.50	0.076794	0.979626	0.795738	0.004066
	1.75	0.018027	0.996803	0.9312	0.000421
	2.00	0.003615	0.999565	0.980397	0.000039
Dairy	**1.00**	0.037995	0.037995	0.000449	0.000449
	1.25	0.200874	0.003143	0.00833	9.0e-06
	1.50	0.467298	0.0002	0.048462	1.5e-07
	1.75	0.711137	0.000011	0.146809	2.4e-09
	2.00	0.867055	5.4e-07	0.299023	3.6e-11
Food away from home	**1.00**	0.009768	0.009768	0.018459	0.018459

	1.25	0.080791	0.000484	0.125128	0.001146
	1.50	0.258177	0.000019	0.346291	0.000056
	1.75	0.492561	6.6e-07	0.594697	2.4e-06
	2.00	0.700954	2.1e-08	0.785446	9.1e08
Other	**1.00**	0.00065	0.00065	0.145462	0.145462
	1.25	0.000014	0.011165	0.455595	0.022015
	1.50	2.6e-07	0.061104	0.744943	0.002443
	1.75	4.4e-09	0.175945	0.905554	0.000225
	2.00	6.9e-11	0.343436	0.97046	0.000018

Notes: Gamma means the log odds of differential assignment due to unobserved factors. Sig+ represents the upper bound of the significance level, whereas Sig- represents the lower bound of the significance level. Rosenbaum Bounds are implemented on 1x1 nearest neighbour matching; 96 rural-urban migrant households from the Red River Delta (RRD) were matched to 96 rural non-migrant households living in the RRD.

Nearest neighbour matching predicted no significant effects for neither HDDS and FVS, nor whether or not the households meets the WHO's recommendations of fruit and vegetable intake. Therefore the null hypothesis of zero effect of rural-urban migration on HDDS, FVS, and FV is accepted, regardless of the degree of hidden bias. Implications are that Rosenbaum bounds need not to be calculated, as determining the end point of the significanct test that leads one to accept the null hypothesis of zero effect of rural-urban migration on diets of migrants does not make sense if the null hypothesis is already accepted in the case of no hidden bias.

5.5.4 Rosenbaum Bounds on the Effect of Having Rural Origins on Diet

In this thesis the analysis focused on the effect of having rural origins on diets of migrants predicted using the matching algorithm of radius caliper matching. However, Rosenbaum bounds (using the STATA command **rbounds**) cannot be applied to an analysis based on this algorithm (Diprete & Gangl, 2004), which is why they have been applied to the results of nearest neighbour matching instead. Subsection 5.4 presented the estimation results of the effect of having rural origins on diets of migrants. Nearest neighbour matching predicted no significant effects for any of the food consumption indicators in both the subsample of the RRD as well as SEA, although radius caliper matching predicted negative effects on the food expenditure and calorie intake shares of fish and seafood of rural-urban migrant households living in the RRD. As the null hypothesis of zero effect of having rural origins on any of these food consumption indicators is accepted, regardless the degree of hidden bias, Rosenbaum bounds need not to be calculated.

6. Discussion

Several data limitations apply to this thesis. As described in chapter 2, the main analysis was based on household food expenditure data of the Vietnam Household Living Standard Survey (VHLSS), collected by the General Statistics Office of Viet Nam (GSO) in collaboration with the World Bank. However, this data was based on a 30-days recall. Although this provides benefits for capturing 'usual behavioural patterns' and limiting survey collection costs, one could argue whether respondents are able to accurately remember all food items (and their associated quantities and values) purchased and consumed over the last month (Food and Agricultural Organisation of the United Nations [FAO], 2011b; Zezza et al., 2017). Generally speaking the longer the recall period, the higher the imprecision of the data. Zezza et al. (2017) argue that there is enough evidence about how a 'usual month' recall period is detrimental to data quality, while not providing sufficient benefits in other domains that could justify that loss of accuracy or precision. This is especially the case when one is interested in nutritional outcomes. The VHLSS is collected every other year, and I would recommend the GSO to rethink the recall period of its survey. Although there are limitations associated with shorter recall periods - like 24 hours of 7 days - as well, I argue that a recall period of say one week would be better able to provide high quality data while at the same time still capture 'usual consumption behaviour'. Implementing a 7 days recall compared to 30 days would, moreover, not add to the costs of survey implementation.

In the VHLSS data is collected on what is purchased or exchanged, self-produced, donated or given as a gift or present. In this way it tries to capture all food that was acquired with the intention to be consumed and not only what was purchased (Zezza et al., 2017). However, this does not mean that all reported quantities are actually eaten. For example if one reports to have consumed one kilogram of bananas, one reports the weight of bananas including its peel. Furthermore, if some bananas get rotten and thrown away or if some parts get wasted, they are counted as consumed while they are actually not. The food consumption indicators identified in this thesis are not corrected for edible portions, and thus do not represent actual food consumption correctly. However, if we assume that both rural-urban migrant households and non-migrant households are exposed to food waste and degree of losses of edible food and nutrients randomly and in similar scope, we can neglect the fact that we did not control for edible portions. However, especially when making inferences about nutrient intake and reaching Estimated Average Requirements (EAR) on nutrient level, one should be aware of the fact that the reported quantities in the VHLSS should be converted to edible portions.

Furthermore, the VHLSS asks questions about 53 food items (54 food items if you also count 'cigarettes' as food item). In comparison to other Living Standards Measurement Studies (LSMSs) – like the LSMS of Bangladesh – this number is quite small. A smaller number of food items included in the survey induces data to be gathered at a higher aggregated level. Although this is beneficial for survey length, it increases the chance of missing the reporting of parts of consumption. Furthermore, a smaller number of food groups requires the creation of aggregated groups like 'other vegetables' and 'other fruits' which are quite large in scope. Especially when you want to use the VHLSS for nutritional purposes, this imposes difficulties that are not too easy to solve accurately. An example is the aggregated category for 'beverages', which includes both the consumption of sweetened soft drinks as well as the consumption of pure water. The aggregated level of this food item induced that I could not accurately analyse the consumption of soft drinks, which is negatively associated with a healthy diet (Malik et al., 2006). In general I would recommend the GSO to reconsider the currently aggregated food items, as disaggregating some of these categories might enhance the quality of the recall data while at the same time increase the accurate usage possibilities.

This brings us to one special aggregated category that has become increasingly important over time: food away from home (FAFH). FAFH represents an increasing share of food consumption, and as we saw in chapter 5 in 2014 accounted for nearly 20 percent of total kilocalories consumed.

Ha et al. (n.d.) point to the importance of ready-to-eat foods and street foods in the Vietnamese diet, which can, in the current survey format, not be specified. Given the increasing importance of FAFH in the Vietnamese diet, I would recommend to disaggregate this food item somehow, specifically incorporating disaggregated consumption at school-based and other widespread feeding programs as well (Zezza et al., 2017).

For analysing trends in agricultural production, domestic supply, food available for human consumption, and international trade the food balance sheets of the FAO are used (see chapter 4). The FAO (2001) itself argues that one should be aware of the fact that the basic data may possibly be incomplete and or inaccurate given that a variety of sources is used, ranging from direct enquiries or records to estimates of government agencies to cost and production surveys. As the food balance sheets depend on the reliability of the underlying basic statistics, the FAO executes quality assurance to ensure that all collected and published information is reliable. However, no guarantee on quality can be made (FAO, 2001), which induces that no guarantee can be made on the figures based on FAOSTAT data in this thesis as well.

Migrant households were defined based on the province of birth of the head of the household and the spouse. Unfortunately no information was available about how long someone had been living in the current province, except for the situations where someone had registration status in a different province than the province of residence and where someone had never had registered household status. As internal migration has become more and more prevalent in Viet Nam (GSO & United Nations Population Fund [UNFPA], 2016b), I argue that it would be helpful if (different types of) migrant households can be defined more accurately in the VHLSS. One step forward would be to ask the question about how long someone has been living in the current province (i.e. Muc01 1A, Q14) to everyone. Asking this question to everyone is easy to implement, while it does not add too much to total survey length. Furthermore it would be helpful if an additional question is asked about which province someone lived in previously, if the answer to Q14 would be less than or equal to say 36 months[9]. In this way both short-term and long-term migrant households could be identified, and province of birth, previous province of residence, and current province of residence could be compared. This allows us to make more precise assumptions about internal migration flows.

De Brauw & Carletto (2012) point to the fact that migration is a rare event, which has implications for the sample size and therefore the power of making statistically significant inferences. In this thesis I could identify 521 rural-urban migrant households. However, the main average treatment effects on the treated, estimated by the PSM model on a subsample of the population, were based on only 96 and 80 migrant households for the effect of migration and having rural origins on diet respectively. This should be kept in mind when making inferences about the estimated results.

As data on food consumption was at the household level, migrant *households* needed to be identified instead of *individual* migrants. Section 5.1 described the characteristics of rural-urban migrant households, where we amongst other things saw that most households belong to the 'Kinh' ethnic group (98.85 percent), and that the head and spouse of the average migrant household had obtained a relatively high level of education compared to the average Vietnamese household. Furthermore, these households had higher levels of income and higher levels of food consumption expenditures than both rural and urban households. Anh et al. (2012) argue that also a high proportion of *individual* rural-urban migrants had obtained high school education or higher (51.4 percent), and that 95.2 percent belonged to the 'Kinh' ethnic group. However, Anh et al. (2003) found that a large propotion of individual migrants is female and unmarried, and younger than 25 years. An example of female migrants follows from a case study by Agergaard et al. (2011) who analyse 'shoulder-pole women porters' at Hanoi's Long Bien Market. The GSO & UNFPA (2016b) argue that migrants are more likely to work in the industrial and construction sector than

[9] I say 36 months as this allows the researcher to identify the previous province of residence for both short-term migrant households (i.e. between 6 and 12 months) and long-term migrant households (i.e. between or equal to 12 and 36 months).

non-migrants, often living in dormitories together with other migrants. This shows that living conditions are likely to be different for individual and household migrants. Furthermore, Tacoli et al. (2014) argue that individual migrants often lack registration status, leading to disadvantages in case of e.g. housing and public services (World bank [WB], 2016b)[10]. However in our sample 79.46 percent of migrant households did have registration status in the residential area, indicating that the majority of households was able to meet the registration requirements and thus were not restricted formally in access to public services. All these findings could imply that we have identified a special subgroup of migrant households. If this is the case migrant *households* might not be representative for *individual* migrants, inducing that the results of our study might not be generalizable to the whole rural-urban migrant population in Viet Nam.

Furthermore, the propensity score estimation induced removing all migrant households from the migrant sample that have no spouse (as households with no spouse have missing variables for several of the matching variables). Therefore ATTs have been estimated only on migrant households that have both a head and spouse. The descriptive analysis of diets in section 5.2 showed that only a few significant differences exist between different types of migrant households (i.e. with or without spouse). I argue, thus, that the exclusion of households that only have a household head is not too big of a problem for the generalizability of the results to other rural-urban migrant households in Viet Nam.

In this thesis we were dealing with observational data, which is why propensity score matching has been implemented as a way of dealing with selection bias. Chapter 3 already explained some problems related to the implementation of Propensity Score Matching (PSM). Li (2012) argues that the effectiveness of PSM depends on the data quality at hand. Furthermore, De Brauw & Carletto (2012) argue in favour of a large set of covariates. Given the fact that the included covariates should be correlated with both migration and food consumption, but unaffected by migration, the number of covariates included in the propensity score estimation was quite low. This induced the excluded covariates to enter into the unobservables, which are assumed to be equal across treatment and control group (Becker & Caliendo, 2005). However, the more covariates are excluded from the propensity score estimation, the higher the likelihood that the unobservables are not similar across groups. Therefore Rosenbaum (2010) proposed the implementation of bounds to check the degree to which significant results hinge on the assumption of no hidden bias. However, these bounds could only be applied on one to one nearest neighbour matching, for which the matching exercise was less successful than for radius caliper matching (i.e. the matching method producing the main results). Nevertheless, given that both matching methods produced more or less similar significant results, the Rosenbaum bounds on the average treatment effects on the treated (ATTs) of one to one nearest matching should give an indication to which degree the results in general are sensitive to selection bias.

Furthermore, Ordinary Least Squares (OLS) estimation has been executed to assess the robustness of the estimated ATTs to putting higher weight on households that are different instead of which are similar. Both the PSM and OLS model are based on the same covariates, assuming there is no omitted variable bias. Although the validity of this assumption was tested by implementing Rosenbaum bounds, Altonji et al. (2005), Bellows and Miguel (2009), and Oster (2016) propose additional methods which one could apply on OLS estimation for assessing the

[10] One important factor influencing internal migration in Viet Nam is the household registration system, known as *ho khau*. This system, which has been implemented in 1964 as an instrument of public security, economic planning, and control of migration, basically means that every citizen can be registered only at the place of permanent residence, and migration can take place only with the permission of authorities (World Bank [WB], 2016b). In practice it induced the household registration of mainly permanent and official migrants, while not including spontaneous and undocumented migrants (Anh et al., 2003). However, the force of the system has diminished since the launch of *Doi Moi* in 1986, inducing that having permission to move is no longer necessary. Although this release of restrictions, local authorities are now allowed by law to set their own policies regarding registration, leading to even tighter requirements (e.g. official and document) for obtaining permanent status in some cities (WB, 2016b). Furthermore, evidence shows that having temporary registration instead of permanent limits the access to some public services, like education and health care, and public employment (Anh, 2017; Anh et al., 2003; Le et al., 2010; WB, 2016b).

degree of selection bias. The coefficient stability ratio proposed by Bellows and Miguel (2009) has been implemented in this thesis (Appendix F), although I argue that we should be cautious with the interpretation of the ratio. This as only a limited number of covariates could be included in the OLS model, and their corresponding explanatory power was considerably low. Given that the main analysis focuses on the results produced by the PSM model, I argue that the results from the coefficient stability approach should not be emphasized on too much.

In general, even after implementing sensitivity analyses and robustness checks, criticism remains related to PSM. McKenzie & Tang (2010) and McKenzie (2012) argue that, even if the approach conditions on a wide array of observable characteristics, comparisons of migrants and non-migrants are unlikely to give convincing estimates of the impact of migration. Because of criticism related to PSM I aimed to focus merely on the sign of significant effects, rather than the exact magnitude. Furthermore, I argue that dealing with the observational data that we have, PSM was one of the best possible approaches for estimating the effect of rural-urban migration and having rural origins on diets of migrants. In the conclusion in the next chapter I raise some suggestions, based on McKenzie & Yang (2010) and McKenzie (2012), for further research related to migration and methodological design.

7. Conclusion

In the context of increasing population and migration rates, impressive economic growth and shifting food consumption patterns, the objective of this thesis was twofold: (1) analysing trends in the local food system in Viet Nam, specifically focussing on food consumption patterns, agricultural production, and international trade, and (2) analysing the effect of rural-migration and having rural origins on diets of rural-urban migrants.

Over the period 1992 to 2014, total consumed kilocalories per capita have been increasing. The average share of these calories coming from fats and proteins has been increasing, whereas the contrary can be seen for the share coming from carbohydrates. This is the case for both rural and urban areas, although the shifts in size of shares are more profound for urban than for rural households. When we disaggregate these trends into different food groups, we see that the importance of rice in diet is declining, although its importance is still considerable. Furthermore, the consumption of meat is considerable as well; from 2002 to 2008 it is the second largest food expenditure group for both rural and urban households, and its share is quite stable over time. Food away from home has become increasingly important over the years - especially for urban households. Expenditure shares for fruits and vegetables, on the contrary, have been quite stable. This might be alarming, given the fact that in 2014 only 1.10 percent of the population was able to meet the adult equivalent recommendations of the WHO of consuming 400 grams of fruits and vegetables per day.

After the *Doi Moi* economic reforms, Viet Nam has become increasingly industrialized inducing impressive economic growth. The reforms also triggered increased productivity, which fostered agricultural production within the country. Increased production can be seen for crops and livestock as well, where especially the latter increase has been impressive. When disaggregating production, we see largest quantities for cereals – i.e. rice production. Also the production of sugars and syrups is considerable, which is mainly due to production of sugarcane. Next to the large increase in production of these two food groups, increases can also be seen for nearly all other food groups. Worth mentioning is the increase in production of vegetables and roots and tubers, where the latter increase mainly consists of increased cassava production. However, not all of these foods produced are consumed by the Vietnamese population; large quantities are exported or used as animal feed. Viet Nam being among the top five global exporters of shrimp, coffee, cashews, rice and pepper, most of rice and cassava produced is produced for export. On the contrary, Viet Nam imports large quantities of maize, mainly used for animal feed. A gap exists between the domestic supply and food consumption of sugars and syrups, indicating that large quantities of this category are used for either non-human consumption or as input for the processing industry. After 1992 the production of stimulants increases rapidly, corresponding with the increasing trend in export that we see for stimulants (i.e. coffee). Imports for meat increase after 2004, which is mainly due to imports in bovine and poultry meat. Overall, total imports and exports have increased, although the share of food in total merchandise exports (imports) has decreased (increased). The main imported and exported food group is, again, cereals, where cereal imports mainly exist of wheat and maize, and cereal exports of rice. Furthermore, the imported quantities of milk and cheese, marine fish, and soybeans have increased over the years.

Several estimation methods have been executed for analysing the effect of rural-urban migration and having rural origins on diets of migrants, although Propensitity Score Matching (PSM) has been de main methodological approach. The predicted average treatment effects on the treated (ATTs) have been assessed by implementing Rosenbaum Bounds on significant ATTs, and by estimating OLS models. Starting with the effect of migration, both the PSM and OLS models predicted significant negative effects for the food expenditure share of rice, other staples, meat, eggs, tofu, and other. Furthermore, a positive significant effect was found for the food expenditure share of food away from home for both models. The calorie intake share of rice and tofu was predicted to decrease, whereas the share of food away from home was predicted to increase. For both food expenditure and calorie intake shares, the effect for the categories of rice and food away from home is predicted to be largest. Furthermore, households are expected to have increased economic access to food when migrating from rural to urban areas. Also the consumed food variety is predicted to increase when moving to urban areas. Furthermore, Rosenbaum Bounds predict that the negative ATTs of the expenditure share of rice and tofu are insensitive to selection bias. Furthermore, the ATTs on the calorie intake shares of rice, tofu and dairy are predicted to be insensitive to selection bias as well. Estimation results for the effect of having rural origins on diets of migrants show zero effect. All estimated models predict that the food consumption indicators of rural-urban migrant households do not significantly differ from urban non-migrant households. Therefore, our estimation results suggest that rural-urban migrant households adapt their diet to urban diets when they move to urban areas. When considering the largest estimated effects, this implies that they start consuming less rice while eating away from home more often.

The results of this thesis point to the fact that the diet of rural-urban migrant households becomes (more) urban when moving into urban areas. Practical implications of these results could therefore point to the implementation of interventions related to nutritional transition and health effects at the urban lifestyle and supply environment. Furthermore, given the fact that rural-urban migration is a widespread phenomenon in Viet Nam, the results could imply a possible acceleration of the nutritional transition in Viet Nam in general. This likely influences the broader food system as well, given that it could trigger changes in socio-economic drivers.

Given the fact that we have observational data and that migration is not random, propensity score matching has been executed. Future research, however, could focus on the use of experiments in research on migration as this allows for considering migration as random. Randomness of migration removes selection bias related to migration, as both treatment and control group have equal possibilities of being treated into migration. McKenzie & Yang (2010) and McKenzie (2012) mention the use of policy experiments, natural experiments and researcher-led field experiments for estimating the determinants or impacts of migration. Therefore, future research on the effects of rural-urban migration and having rural origins could explore the design and implementation of (field) experiments for dealing with self-selection into migration. Implementing these experiments will add to the validity of estimated effects related to migration. Furthermore, future research could include food choice experiments into the analysis, in order to analyse consumer's preferences and beliefs about certain products. These experiments could be conducted both in rural and in urban sites, and for both rural-urban migrants as well as non-migrants. Conducting these experiments could provide valuable information which complements the analysis based on household survey data.

Currently there are no specific standard diet recommendations available for the Vietnamese population. The National Institute of Nutrition (NIN) (n.d., a) provides ten tips on proper nutrition, but these recommendations do not meet details like recommended grams per day. Furthermore, Nguyen & Pham (2008) have revised the Recommended Dietary Allowances (RDA) for the Vietnamese population, but these RDA are at the micronutrient level. This is useful for nutritionists, but it first needs to be translated in order to be valuable for the general population as well. Therefore in order to educate people about healthy diets and to steer consumption towards healthy diets, it would be useful if a description of a standard healthy diet would be developed, specified to the Vietnamese context. This would be useful for both the Vietnamese population and for researchers on nutrition in Viet Nam. Therefore future research could focus on developing standard healthy diet recommendations.

We saw that only a very small proportion of the Vietnamese population meets the adult equivalent daily recommended fruits and vegetables intake of the WHO. Furthermore, the expenditure share on fruits and vegetables has been quite stable over the years, and in 2014 the associated expenditure and calorie intake shares were still not too large. Therefore the results point to the relevance of future research related to fruit and vegetable intake in Viet Nam, as fruits and vegetables are argued to be important elements of a healthy diet.

Lastly, this thesis has been about the effect of rural-urban migration and having rural origins on diets of migrants. These effects are analysed within the broader context of the nutritional transition. Future research could focus on the effects of rural-urban migration on human health as well. The analysis of these effects is relevant for countries like Viet Nam in particular, as urbanization is increasing, urban diets are rapidly changing, and, moreover, urban pollution is becoming more and more problematic.

References

Aakvik, A. (2001). Bounding a matching estimator: the case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics, 63*(1): 115-132.

Abraído-Lanza, A., Armbrister, A.N., Flórez, K.R., & Aguirre, A.N. (2006). Toward a Theory-Driven Model of Acculturation in Public Health Research. *American Journal of Public Health, 96*(8).

Acosta, P. (2006). Labor supply, school attendance, and remittances from international migration: the case for El Salvador. *World Bank Policy Research Working Paper,* 3903. Washington, D.C.: The World Bank.

Agergaard, J., & Thao, V.T. (2011). Mobile, Flexible, and Adaptable: Female Migrants in Hanoi's Informal Sector. *Population, Space and Place, 17*: 407-420.

Altonji, J.G., Elder, T.E., & Taber, C.R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy, 113*(1): 151-184.

Anders, S., & Schroeter, C. (2010). The Impact of Nutritional Supplement Choices on Diet Behavior and Obesity Outcomes. *PLoS ONE, 12*(10): e0185258.

Anderson, B. (2017). Selection models and weak instruments [Blog post]. Retrieved from https://info.umkc.edu/drbanderson/selection-models-and-weak-instruments/

Angrist, J.D., & Pischke, J.S. (2008). *Mostly harmless econometrics: an empiricist's companion*. New Jersey, USA: Princeton University Press.

Anh, D.N. (2017). Rural-to-urban migration in Vietnam: Trend and institutions. *VNU Journal of Social Sciences and Humanities, 3*(2).

Anh, D.N., Tacoli, C., & Thanh, H.X. (2003). Migration in Viet Nam: A review of information on current trends and patterns, and their policy implications. *IIED,* G00479. London, England: Department For International Development.

Anh, D.T., & Sautier, D. (2011). Local Food Systems in Viet Nam: Strengths and Opportunities. *Extension Bulletin - ASPAC. Food and Fertilizer Technology Center*, (640).

Anh, L.T.K, Vu, L.H., Bonfoh, B., & Schelling, E. (2012). An analysis of interprovincial migration in Viet Nam from 1989 to 2009. *Global Health Action, 5*.

Arah, O.A., Chiba, Y., & Greenland, S. (2008). Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Annals of Epidemiology, 18*: 637-646.

Austin, P.C. (2013). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine, 33*: 1057-1069.

Bansal, D., Satija, A., Khandpur, N., Bowen, L., Kinra, S., Prabhakaran, D., Reddy, K.S., & Ebrahim, S. (2010). Effects of migration on food consumption patterns in a sample of Indian workers and their families. *Public Health Nutrition, 13*(12): 1982-1989.

Becker, S., & Caliendo, M. (2007). Sensitivity analysis for average treatment effect. *The Stata Journal, 7*(1): 71-83.

Becker, S., & Ichino, A. (2002). Estimation of Average Treatment Effects based on Propensity Scores. *The Stata Journal, 2*(4): 358-377.

Bellows, J., & Miguel, E. (2009). War and local collective action in Sierra Leone. *Journal of Public Economics, 93*: 1144-1157.

Best, H. & Wolf, C. (2014). *The Sage Handbook of Regression Analysis and Causal Inference*. Sage Publications Ltd.

Bojorquez, I., Rentería, D., & Unikel, C. (2014). Trajectories of dietary change and the social context of migration: a qualitative study. *Appetite, 81*: 93-101.

Bowen, L., Ebrahim, S., Stavola, B., Ness, A., Kinra, S., Bharathi, A.V., Dorairaj, P., & Reddy, K.S. (2011). Dietary Intake and Rural-Urban Migration in India: A Cross-Sectional Study. *PloS One, 6*(6): e14822.

Bushway, S., Jonhson, B.D., & Slocum, L.A. (2007). Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology. *Journal of Quantitative Criminology, 23*(2): 151-178.

Busse, M., & Königer, J. (2012). Trade and Economic Growth: A Re-examination of the Empirical Evidence. *HWWI Research Paper,* 123. Hamburg, Germany: Hamburg Institute of International Economics (HWWI).

Caliendo, M., & Kopeining, S. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. *IZA Discussion Paper,* 1588. Bonn, Germany: Institute of Labor Economics (IZA).

Cities Alliance (2011). Viet Nam Urbanization Review. *Technical Assistance Report,* P122324. Washington, D.C.: The World Bank.

Coates, J., Rogers, B.L., Blau, A., Lauer, J., & Roba, A. (2017). Filling a dietary data gap? Validation of the adult male equivalent method of estimating individual nutrient intakes from household-level data in Ethiopia and Bangladesh. *Food Policy, 72*: 27-42.

Cockx, L., Colen, L., & Weerdt, J. (2017). From Corn to Popcorn? Urbanization and Food Consumption in sub-Saharan Africa: Evidence From Rural-Urban Migrants in Tanzania. *LICOS Discussion Paper Series*, 390/2017.

Cohen, D.A., & Babey, S.H. (2012). Contextual influences on eating behaviours: heuristic processing and dietary choices. *Obesity reviews, 13*(9): 766-779.

Combs, G.F., Welch, R.M., Duxbury, J.M., Uphoff, N.T., & Mesheim, M.C. (1996). *Food-based approaches to prevent micronutrient malnutrition: an international research agenda* [Summary report of an international workshop]. Ithaca, NY, USA: Cornell International Institute for Food, Agriculture and Development (CIIFAD).

Consultative Group on International Agricultural Research (n.d., a). About A4NH. Retrieved from http://a4nh.cgiar.org

Consultative Group on International Agricultural Research (n.d., b). About Us. Retrieved from http://www.cgiar.org

Consultative Group on International Agricultural Research (n.d., c). CGIAR Research Program on Agriculture for Nutrition and Health. Retrieved from http://www.cgiar.org

Coxhead, I., Cuong, N.V., & Vu, L.H. (2015). Migration in Viet Nam: New Evidence from Recent Surveys. *Viet Nam Development Economics Discussion Paper,* 2. Washington, D.C.: World Bank Group.

Dang, A., Goldstein, D., & McNally J. (1997). Internal Migration and Development in Viet Nam. *The International Migration Review, 31*(2): 312-337.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy.* Washington, D.C.: The World Bank.

Deaton, A., & Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates. *Living standards measurement study (LSMS) working paper*, LSM 135. Washington, D.C.: The World Bank.

De Brauw, A. (2010). Seasonal Migration and Agricultural Production in Viet Nam. *Journal of Development Studies, 46*(1): 114-139.

De Brauw, A., Carletto, C. (2012). Improving the measurement and policy relevance of migration information in multi-topic household surveys. *World Bank Working Paper*. Washington, D.C.: The World Bank.

De Brauw, A., & Harigaya, T. (2007). Seasonal Migration and Improving Living Standards in Viet Nam. *American Agricultural Economics Association, 89*(2): 430-447.

Deheija, R.H., & Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics, 84*(1): 151-161.

Deshingkar, P. (2006). Internal Migration, Poverty and Development in Asia. *Asia 2015 Conference - Session 3: Realising the Potential for Poverty Reduction.*

DiPrete, T.A., & Gangl, M. (2004). Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology, 34*(1): 271-310.

Drewnowski, A., & Popkin, B.M. (1997). The Nutrition Transition: New Trends in the Global Diet. *Nutrition Reviews, 55*(2).

Dreyfus, F., Plencovich, C., Petit, M., Akca, H., Dogheim, S., Ishii-Eiteman, M., Jiggings, J., Kiers, T., & Kingamkono, R. (2008). Food Systems Management. In S. Biggs and G. Castillo (Eds.) *Historical Analysis of the Effectiveness of AKST Systems in Promoting Innovation*. Retrieved from http://www.weltagrarbericht.de

Duc, L.N., Raabe, K., & Grote, U. (2012). Rural-Urban Migration in Vietnam: Do Households and Migrants Get Better Off? Hannover, Germany: Institute for Environmental Economics and World Trade, Leibniz University Hannoved.

Ericksen, P.J. (2008). Conceptualizing food systems for global environmental change research. *Global Environmental Change, 18*: 234-245.

European Food Information Council (2006). The determinants of food choice. Retrieved from http://www.eufic.org/en/

Food and Agricultural Organization (1997). The food system and factors affecting household food security and nutrition. In *Agriculture food and nutrition for Africa - A resource book for teachers of agriculture.* Rome, Italy: Food and Agricultural Organization of the United Nations, Information Division. Retrieved from http://www.fao.org

Food and Agricultural Organization of the United Nations (2001). Food Balance Sheets: A Handbook. Rome, Italy: Food and Agricultural Organization of the United Nations. Retrieved from http://www.fao.org

Food and Agricultural Organization of the United Nations (2003). Trade and food security: the importance of agriculture and agricultural trade in developing countries. In *The State of Food Insecurity in the World, 2003*. Rome, Italy: Food and Agricultural Organization of the United Nations, Information Division. Retrieved from http://www.fao.org

Food and Agricultural Organization of the United Nations (2011a). Food, Agriculture and Cities: Challenges of food and nutrition security, agriculture and ecosystem management in an urbanizing world. *FAO Food for the Cities multi-disciplinary initiative position paper.* Retrieved from http://www.fao.org

Food and Agricultural Organization of the United Nations (2011b). *Guidelines for measuring household and individual dietary diversity*. Rome, Italy: Food and Agricultural Organization of the United Nations.

Food and Agricultural Organization of the United Nations (2014). Towards the Development of the Programme on Sustainable Food Systems (SFSP) of the 10-Year Framework of Programmes on Sustainable Consumption and Production (10YFP). Retrieved from http://www.fao.org

Food and Agricultural Organization of the United Nations (n.d.). FAOSTAT Database [Database]. Retrieved from http://www.fao.org

Global Facility of Disaster Reduction and Recovery (2011). Vulnerability, Risk Reduction, and Adaptation to Climate Change: Viet Nam. *Climate Risk and Adaptation Country Profile*.

Global Facility of Disaster Reduction and Recovery (n.d.). Viet Nam. Retrieved from https://www.gfdrr.org/en

Global Panel on Agriculture and Food Systems for Nutrition (2016). Food systems and diets: Facing the challenges of the 21st century. London, England: Global Panel on Agriculture and Food Systems for Nutrition.

Greenland, S. (1996). Basic Methods for Sensitivity Analysis of Biases. *International Journal of Epidemiology, 25*(6): 1107-1116.

Groth, A., Mueller, S., & Wilke, T. (2017). Which PSM Method to Use? Association Between Chosen Propensity Score Method and Outcomes of Retrospective Real-World Treatment Comparisons: Evaluation Of 18 Different PSM Methods. *The Journal of The International Society for Pharmacoeconomics and Outcomes Research, 20*(9): A771-A772.

General Statistics Office of Viet Nam (2010). On the Viet Nam Household Living Standard Survey (VHLSS) 2010. Retrieved from http://www.gso.gov.vn/Default_en.aspx?tabid=491

General Statistics Office of Viet Nam (2012). Data results of the Viet Nam Household Living Standard Survey 2012: Consumption Expenditure. Retrieved from http://www.gso.gov.vn/Default_en.aspx?tabid=491

General Statistics Office of Viet Nam (2016). Statistical Yearbook of Viet Nam 2016. Retrieved from http://www.gso.gov.vn/Default_en.aspx?tabid=491

General Statistics Office of Viet Nam (n.d.). Consumer price index. Retrieved from http://www.gso.gov.vn/Default_en.aspx?tabid=491

General Statistics Office of Viet Nam, & United Nations Population Fund (2016a). The 2014 Viet Nam Intercensal Population and Household Survey: Migration and Urbanization in Viet Nam. Hanoi, Viet Nam: Viet Nam News Agency Publishing House.

General Statistics Office of Viet Nam, & United Nations Population Fund (2016b). The 2015 National Internal Migration Survey: Major Findings. Hanoi, Viet Nam: Viet Nam News Agency Publishing House.

Ha, T.A.D., Le, H.D., & Nguyen, T.D. (n.d.). Nutritional values of a number of traditional Vietnamese foods. Retrieved from http://viendinhduong.vn/research/en/66/28/thanh-phan-dinh-duong-mot-so-thuc-an-truyen-thong-viet-nam.aspx

Ham, J.C., Li, X., & Reagan, P.B. (2005). Propensity Score Matching, a Distance-Based Measure of Migration, and the Wage Growth of Young Men. *Federal Reserve Bank of New York Staff Report,* 212. New York, USA: Federal Reserve Bank of New York.

Harding, D.J. Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy. *The American Journal of Sociology, 109*: 676-719.

Harris, J.R., & Todaro, M.P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, *60*(1): 126-142.

Hawkes, C. (2006). Uneven dietary development: linking the policies and processes of globalization with the nutrition transition, obesity and diet-related chronic diseases. *Globalization and Health, 2*(4).

Heald, A.H., Sharma, R., Anderson, S.G., Vyas, A., Siddals, K., Patel, J., Bhatnagar, D., Prabharkaran, D., Rudenski, A. Hughes, E., Durrington, P., Gibson, J.M., & Cruickshank, J.K. (2005). Dietary intake and the insulin-like growth factor system: effects of migration in two related populations in India and Britain with markedly different dietary intake. *Public Health Nutrition, 8*(6): 620-627.

Heinrich, C., Maffioli, A., & Vázquez, G. (2010). A Primer for Applying Propensity-Score Matching. *Impact-Evaluation Guidelines - Technical Notes*, IDB-TN-161. Inter-American Development Bank.

Heller, M. (2017). What is the DASH Diet? Retrieved from http://dashdiet.org/what_is_the_dash_diet.asp

High Level Panel of Experts (2014). Food losses and waste in the context of sustainable food systems. *HLPE Report,* 8. Retrieved from http://www.fao.org

Himmelgreen, D.A., Cantor, A., Arias, S., & Daza N.R. (2014). Using a biocultural approach to examine migration/globalization, diet quality and energy balance. *Physiology & Behavior, 123*: 76-85.

Hoang, L.V. (2009a). Analysis of Calorie and Macronutrient Consumption in Viet Nam. *Depocen Working Paper Series*, 2009/14.

Hoang, L.V. (2009b). Estimation of Food Demand from Household Survey Data in Viet Nam. *Depocen Working Paper Series*, 2009/12.

Hodgson, J.M., Hsu-Hage, B.H.H., & Wahlqvist, M.L. (2010). Food variety as a quantitative descriptor of food intake. *Ecology of Food and Nutrition, 32*(3-4).

Ichino, A., Mealli, F., & Nannicini, T. (2006). From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity? *IZA Discussion Paper,* 2149. Bonn, Germany: Institute of Labor Economics (IZA).

Ichino, A., Mealli, F., & Nannicini, T. (2008). From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity? *Journal of Applied Econometrics, 23*(3): 305-327.

International Food Policy Research Institute (2017). A4NH flagship 1: Food systems for healthier diets. A4NH Note 1. Washington, D.C.: International Food Policy Research Institute (IFPRI).

International Food Policy Research Institute, & Compact2025 Initiative (2016). Food System Transformations: Brazil, Rwanda and Viet Nam. *World Economic Forum (WEF) Global Agenda Council (GAC) on Food and Nutrition Security.*

Johnston, J.L., Fanzo, J.C., & Cogill, B. (2014). Understanding Sustainable Diets: A Descriptive Analysis of the Determinants and Processes That Influence Diets and Their Impact on Health, Food Security, and Environmental Sustainability. *American Society for Nutrition, 5*: 418-429.

Kearney, J. (2010). Food consumption trends and drivers. *Philosophical Transactions of the Royal Society, 365*: 2793-2807.

Khandker, S.R., Koolwal, G.B., & Samad, H.A. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices.* Retrieved from http://www.worldbank.org

Krugman, P., Obstfeld, M., & Melitz, M. (2014). *International Economics: Theory and Policy, Global Edition* (10th edition). Pearson Education Limited.

Kuyper, E.M., Engle-Stone,R., Arsenault, J.E., Arimond, M., Adams, K.P., & Dewey, K.G. (2017). Dietary gap assessment: an approach for evaluating whether a country's food supply can support healthy diets at the population level. *Public Health Nutrition, 20*(13): 2277-2288.

Lairon, D. (2012). Biodiversity and sustainable nutrition with a food-based approach. In B. Burlingame, S. Dernin, Nutrition and Consumer Protection Division, Food and Agricultural Organization of the United Nations (Eds.), *Sustainable diets and biodiversity: directions and solutions for policy, research and action* (pp. 30-35). Rome, Italy: Food and Agricultural Organization of the United Nations.

Le, B.D., Tran, G.L., & Nguyen, T.P.T. (2010). Social protection for rural-urban migrants in Vietnam: current situation, challenges and opportunities. *SPA Working Papers.* Hanoi, Viet Nam: Institute for Social Development Studies.

Le, T.H., Le, B.M., & Nguyen, C.K. (2003). Trends in Food Production and Food Consumption in Viet Nam During the Period 1980-2000. *Malaysian Journal of Nutrition 9*(1): 1-5.

Leaf, M. (2002). A Tale of Two Villages: Globalization and Peri-Urban Change in China and Viet Nam. *Cities, 19*(1): 23-31.

Leuven, E., & Sianesi. B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.

Lewis, W.A. (1979). The dual economy revisited. *The Manchester School,* 47(3): 211-229.

Li, M. (2012). Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide. *Organizational Research Methods, 1*(39).

Lin, D.Y., Psaty, B.M., & Kronmal, R.A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics, 54*: 948-963.

Liu, W., Kuramoto, S.J., & Stuart, E.A. (2013). An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research. *Prevention Science, 14*(6): 570-580.

Loi, C.C. (2005). Rural to urban migration in Vietnam. *ASEDP,* 71: 115-143.

Lucas, R.E.B. (1997). Internal Migration in Developing Countries. In *Handbook of Population and Family Economics*, 721-798*.* Boston, USA: Boston University.

Malik, V.S., Schulze, M.B., & Hu, F.B. (2006). Intake of sugar-sweetened beverages and weigh gain: a systematic review. *The American Journal of Clinical Nutrition, 84*(2): 274-288.

McKenzie, D. (2012). Learning about migration through experiments. *CDP Discussion Paper Series,* 07/12. London, England: Centre for Research and Analysis of Migration (CReAM).

McKenzie, D., & Yang, D. (2010). Experimental Approaches in Migration Studies. *IZA Discussion Paper,* 5125. Bonn, Germany: Institute of Labor Economics (IZA).

Melesse, M.B., Berg, M., Béné, C., Brauw, A., Brouwer, I.D., & Remans, R. (2017). Improving Diets through Food Systems: Metrics and Methods for Analysis. *Unpublished Article.*

Mergenthaler, M., Weinberger, K., & Qaim, M. (2009). The food system transformation in developing countries: A disaggregate demand analysis for fruits and vegetables in Viet Nam. *Food Policy, 34*: 426-436.

Minister of Construction (2009). Decree on the classification of urban areas. Retrieved from https://thuvienphapluat.vn/van-ban/Xay-dung-Do-thi/Nghi-dinh-42-2009-ND-CP-phan-loai-do-thi-87949.aspx

Mishra, V., & Ray, R. (2009). Dietary Diversity, Food Security and Undernourishment: The Vietnamese Evidence. *Asian Economic Journal, 23*(2): 225-247.

Molini, V. (2007). Food Security in Viet Nam during the 1990s: the empirical evidence. *Food Security: Indicators, Measurement, and the Impact of Trade Openness*: 129-149*.*

Moustier, P., Anh, D.T., & Figuié, M. (2003). Food markets and agricultural development in Vietnam. Hanoi, Viet Nam: MALICA (CIRAD - IOS - RIFAV - VASI).

Nannicini, T. (2007). Simulation-Based Sensitivity Analysis for Matching Estimators. *The Stata Journal, 7*(3): 334-350.

National Institute of Nutrition (2010). General Nutrition Survey 2009-2010. Hanoi, Viet Nam: Medical Publishing House.

National Institute of Nutrition (2011a). Part one national nutrition strategy for 2011-2012, with a vision toward 2030. Retrieved from http://viendinhduong.vn/home/vi/Default.aspx

National Institute of Nutrition (2011b). Part two national nutrition strategy for 2011-2012, with a vision toward 2030. Retrieved from http://viendinhduong.vn/home/vi/Default.aspx

National Institute of Nutrition (n.d., a). Food-based dietary guidelines - Viet Nam. Retrieved from http://viendinhduong.vn/home/vi/Default.aspx

National Institute of Nutrition (n.d., b). Preparing Healthy Meals at Home. Retrieved from http://viendinhduong.vn/home/vi/Default.aspx

Nesheim, M.C., Oria, M., & Yih, P.T. (2015). *A Framework for Assessing Effects of the Food System.* Wasington, DC: The National Academies Press.

Nguyen, C.K., & Pham, V.H. (2008). Vietnam Recommended Dietary Allowances 2007. *Asia Pacific Journal of Clinical Nutrition, 17*(S2): 409-415.

Nguyen, C.V., Berg, M., & Lensink, R. (2011). The impact of work and non-work migration on household welfare, poverty and inequality. *Economics of Transition, 19*(4): 771-799.

Nguyen, L.D., Raabe, K., & Grote, U. (2015). Rural-Urban Migration, Household Vulnerability and Welfare in Viet Nam. *World Development, 71*: 79-93.

Nguyen, M.C., & Winters, P. (2011). The impact of migration on food consumption patterns: The case of Viet Nam. *Food Policy, 36*: 71-87.

Nguyen, T.P., Tran, N.T.M.T., Nguyen, T.N., & Oostendorp, R. (2008). Determinants and impacts of migration in Viet Nam. *Depocen Working Paper Series,* 2008/01.

Nhan, P.T.T., & Tuong, P.D. (2006). Household dietary changes after 6 years (1999-2005) in 6 communes of Dong Anh district - Hanoi. *Journal of food and Nutrition Sciences, 2*(3+4).

Nichols, A. (2007). Causal inference with observational data. *The Stata Journal, 7*(4): 507-541.

Organisation for Economic Cooperation and Development (n.d., a). OECD Data [Database]. Retrieved from https://data.oecd.org

Organisation for Economic Cooperation and Development (n.d., b). What are equivalence scales? Retrieved from http://www.oecd.org

Oster, E. (2016). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics.*

Oxford Dictionaries (n.d.). Diet[1]. Retrieved from https://en.oxforddictionaries.com

Phan, D., & Coxhead, I. (2010). Inter-provincial migration and inequality during Viet Nam's transition. *Journal of Development Economics, 91*: 100-112.

Phung, D.T., & Nguyen, P. (n.d.). Viet Nam Household Living Standards Survey (VHLSS), 2002 and 2004: Basic Information. Retrieved from http://www.worldbank.org/

Popkin, B.M. (1997). The nutrition transition and its health implications in lower-income countries. *Public Health Nutrition, 1*(1): 5-21.

Popkin, B.M., Adair, L.S., & Ng, S.W. (2010). Global nutrition transition and the pandemic of obesity in developing countries. *Nutrition review, 70*(1): 3-21.

Popkin, B.M., Horton, S., & Kim, S. (2001). The nutritional transition and diet-related chronic diseases in Asia: Implications for prevention. *FCND Discussion Paper*, 105. Washington, D.C.: International Food Policy Research Institute.

PopulationPyramid.net (2016). Viet Nam. Retrieved from https://www.populationpyramid.net/viet-nam/2016/

Sparrow, R. (2017). Regressions and Matching Designs [Powerpoint Slides]. *DEC32806: Impact Assessment of Policies and Programmes.* Wageningen, The Netherlands: Wageningen University & Research.

Swindale, A., & Bilinski, P. (2006). Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide [Version 2]. Washington, D.C.: Food and Nutrition Technical Assistance III Project.

Quang, A.T.P. (2015). Multidimensional Poverty and Migrant Households: A Case Study of Viet Nam. *Unpublished Master Thesis*. Sydney, Australia: Macquarie University.

Renzaho, A.M.N., & Burnbs, C. (2006). Post-migration food habits of sub-Saharan African migrants in Victoria: A cross-sectional study. *Nutrition & Dietetics, 63*: 91-102.

Rosenbaum, P.R. (2010). *Design of Observational Studies.* New York, USA: Springer New York.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1): 41-55.

Rosenbaum, P.R., & Rubin, D.B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician, 39*(1): 33-38.

Satia-Abouta, J. (2003). Dietary Acculturation: Definition, Process, Assessment, and Implications. *International Journal of Human Ecology, 4*(1).

Siegel, K.R., Ali, M.K., Srinivasiah, A., Nugent, R.A., & Narayan, K.M.V. (2014). Do We Produce Enough Fruits and Vegetables to Meet Global Health Need? *PLoS ONE, 9*(8): e109059.

Stolzenberg, R.M., & Relles, D.A. (1990). Theory Testing in a World of Constrained Research Design: The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental Research. *Sociological Methods and Research, 18*: 394-415.

Tacoli, C., McGranahan, G., Satterthwaite, D. (2014). Urbanization, Rural-urban Migration and Urban Poverty. *World Migration Report 2015, Background paper.* London, England: International Organization for Migration.

Talsma, E., Brouwer, I., Nijhuis, E., Mekonnen, D., & Achterbosch, T. (2017). Dietary Gap Analysis [Powerpoint Slides]. Wageningen, The Netherlands: Human Nutrition Department and Wageningen Economic Research.

Thang, N.M., & Popkin, B.M. (2004). Patterns of food consumption in Viet Nam: effects on socioeconomic groups during an era of economic growth. *European Journal of Clinical Nutrition, 58*: 145-153.

Thanh, T.D. (2003). Researches in Estuarine Environment and Ecosystem of the Red River: An Overview on Activities and Results. *Marine Environment and Resources*: 34-53.

Thi, H.L., Hong, P.L., Hoang, T. N., Thu, G. N., Thanh, X. L. T., Kim, C. N. T., & Vu, N. T. T. (2015). Changes of Food Expenditure and Food Consumption of People Living in Ba Vi District, Hanoi, Viet Nam from 1999 to 2013. *Health*, *7*: 1696-1702.

Thu, V.T., & Dinh, T.T. (2011). The interactions between agricultural production, food technology, nutrition in production and human life. *Journal of Food and Nutrition Sciences, 7*(2).

Torun, B., Stein, A.D., Schroeder, D., Grajeda, R., Conlisk, A., Rodriguez, M., Mendez, H., & Martorell, R. (2002). Rural-to-urban migration and cardiovascular disease risk factors in young Guatemalan adults. *International Journal of Epidemiology, 31*: 218-226.

Trinh, H.T. (2017). Price of calorie, Food expenditure by food categories: to find the reason why PCCI in 2008 have decreased. *Unpublished Working Paper*.

Trinh, H.T., & Morais, J. (2017). Impact of socioeconomic factors on nutritional diet in Viet Nam from 2004 to 2014: new insights using compositional data analysis. *Toulouse School of Economics Working Papers,* 17-825: 1-26.

Trinh, H.T., Simioni, M., & Thomas-Agnan, C. (2017). A Fresh Look at the Nutritional Transition in Viet Nam using Semiparametric Modeling. *Toulouse School of Economics Working Papers*, 17-842: 1-21.

Tu, T.A., Nguyen, T.A., & Huong, X.T. (2008). Migration to Competing Destinations and Off-Farm Employment in Rural Viet Nam: A Conditional Logit Analysis. *Depocen Working Paper Series,* 2008/22.

United Nations Population Fund (2007). Internal Migration in Viet Nam: The Current Situation. Hanoi, Viet Nam: United Nations Population Fund Viet Nam.

Van der Weele, T.J., & Arah, O.A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology, 22*: 42-52.

Varian, H.R. (2010). *Intermediate Microeconomics: A Modern Approach (8th edition).* New York, USA: W.W. Norten & Company.

Verbeek, M. (2012). *A guide to modern econometrics* (4th edition). Chichester, England: John Wiley & Sons.

Vietnam Trade Promotion Agency (2008). Report on Vietnamese Vegetable and Fruit Sector. Hanoi, Viet Nam: Export Promotion Center.

Vogt, P.W. (2011). *SAGE Quantitative Research Methods* (1st edition). Sage Publications Ltd.

Vu, N.A. (2014). Sugar Industry in Viet Nam. *VietinBankSc Industry Report.* VietinBankSc.

Warne, J.P., & Dallman, M.F. (2007). Stress, diet and abdominal obesity: Y? *Nature Medicine, 13*: 781-783.

White, H., & Masset, E. (2002). Child poverty in Viet Nam: using adult equivalence scales to estimate income-poverty for different age groups. *Munich Personal RePEc Archive Paper,* 777(6).

Wikipedia (n.d.). List of regions of Viet Nam [image]. Retrieved from https://en.wikipedia.org/wiki/List_of_regions_of_Viet Nam

World Bank (2008). Reshaping Economic Geography. *World Development Report,* 43738. Washington, D.C.: The World Bank.

World Bank (2016a). Transforming Vietnamese Agriculture: Gaining More from Less. *Viet Nam Development Report 2016*. Hanoi, Viet Nam: Hong Duc Publishing House.

World Bank (2016b). Viet Nam's Household Registration System. Hanoi, Viet Nam: Hong Duc Publishing House.

World Bank (n.d., a). The World Bank in Viet Nam: Overview. Retrieved from http://www.worldbank.org

World Bank (n.d., b). World Development Indicators [Database]. Retrieved from http://www.worldbank.org

World Food Program, & Vulnerability Analysis and Mapping Branch (2008). Food Consumption Analysis - Calculation and use of the food consumption score in food security analysis. *Technical Guidance Sheet*.

World Health Organization (2003). Diet, nutrition and the prevention of chronic diseases. *WHO Technical Report Series,* 916*.*

World Health Organization (2015). Healthy diet. Retrieved from http://www.who.int/en/

World Health Organization (n.d., a). Diet. Retrieved from http://www.who.int/en/

World Health Organization (n.d., b). Increasing fruit and vegetable consumption to reduce the risk of noncommunicable diseases. Retrieved from http://www.who.int/en/

Zezza, A., Carletto, C., Fiedler, J.L., Gennari, P., & Jolliffe, D. (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.

Appendix A

Table A.1: List of food items included in each food group

Food Group	Food Code	Food item
Rice	101	Plain rice, including fragrant and speciality rice
	102	Sticky rice
Other staples	103	Maize
	104	Cassava
	105	Potato of various kinds
	106	Wheat grains, bread, wheat powder
	107	Flour noodle, instant rice noodle/porridge
	108	Fresh rice noodle, dried rice noodle
	109	Vermicelli
Meat	110	Pork
	111	Beef
	112	Buffalo meat
	113	Chicken meat
	114	Duck and other poultry meat
	115	Other types of meat (goats, dogs, sheep, wild animals, birds, …)
	116	Processed meat (boiled pork pies, fried pork pies, roasted pork, sausages, …)
Fish and seafood	118	Fresh shrimps, fish
	119	Dried and processed shrimps, fish
	120	Other aquatic products and seafood (crabs, snails, …)
Eggs	121	Eggs of chickens, ducks, muscovy ducks, geese
Tofu	122	Tofu
Vegetables	124	Beans of various kinds
	125	Fresh peas of various kinds
	126	Morning glory vegetables
	127	Kohlrabi
	128	Cabbage
	129	Tomato
	130	Other vegetables (gourd, winter melon, cucumber, cabbage, squash, …)
Fruits	131	Orange
	132	Banana
	133	Mango
	134	Other fruits (rambutan melon, papaya, guava, litchi, grapes, …)
Dairy	141	Condensed milk, milk powder

	142	Ice cream, yoghurt
	143	Fresh milk
Food away from home	153	Outdoor meals and drinks (breakfast, lunch, dinner)
Other	117	Lard cooking oil
	123	Peanuts, sesame
	135	Fish sauce
	136	Salt
	137	MSG
	138	Glutamate
	139	Sugar, molasses
	140	Confectionery
	144	Alcohol of various kinds
	145	Beer of various kinds
	147	Instant coffee
	148	Coffee powder
	149	Instant tea powder
	150	Other dried tea
	152	Betel leaves, areca nuts, lime, betel pieces
	154	Other food and drinks (other processed food and foodstuff, additives, seasonings, …)

Notes: Food items in the VHLSS of 2014 are grouped into food groups based on food groups by other papers in the field of food consumption patterns in Viet Nam (i.e. GSO (2012), Hoang (2009b), Mishra & Ray (2009), Nguyen & Winters (2011), and Thang & Popkin (2004)), trends seen in food consumption patterns in Viet Nam over the last years (see next chapter), and the trade-off between the significance of each food group and limiting the total number of food groups.

Appendix B

Table B.1: List of food consumption indicators explored but not included in the analysis

Indicator	Reason not included in the analysis
Food Consumption Score	The Food Consumption Score (FCS), proposed by the World Food Programme (2008), is a composite score based on dietary diversity, food frequency, and relative nutritional importance of different food groups (World Food Programme [WFP] & Vulnerability Analysis and Mapping Branch [VAM], 2008). One of the advantages of the FCS over the Household Dietary Diversity Score (HDDS) and Food Variety Score (FVS) is that it captures both dietary diversity and food frequency. Its calculation follows five steps: (1) using 7-day food frequency data, all food items are grouped into nine specific food groups, (2) all consumption frequencies of food items of the same group are summed and the value of each group above 7 is regrouped as 7, (3) the value obtained for each food group is multiplied by its weight and new weighted food group scores are created, (4) the weighted food group scores are summed, and thus the FCS is created, and (5) using appropriate thresholds the FCS is recoded from a continuous variable into a categorical variable (WFP & VAM, 2008). However, food consumption data in the Vietnam Household Living Standards Survey (VHLSS) is reported in consumed quantity and value over the last 30 days. The data is reported as such that we do not know the frequency of each item consumed. This is why unfortunately the Food Consumption Score could not be calculated, and thus is excluded in the analysis.
Dietary Gap on: **a) food group level** **b) nutrient level**	Within the flagship Food Systems for Healthier Diets (FSHD) the Human Nutrition Department of Wageningen University & Research (WUR) and Wageningen Economic Research (WECR) are working on a dietary gap analysis with the help of Living Standard Measurement Survey (LSMS) data. In this analysis they calculate the percentage of households meeting their (a) food group recommendations per food group, and (b) Estimated Average Requirements (EAR) on nutrient level. The calculation of the Dietary Gap on food group level is as follows: (1) total household consumption in grams per food group is calculated using Dietary Approaches to Stop Hypertension (DASH) Diet food groups, (2) DASH food group recommendations in grams for each households are calculated using a consumer unit, (3) household consumption of food groups in grams is compared to food group recommendations in grams, (4) the percentage of households meeting their food group recommendations is calculated (Talsma et al., 2017). The calculation of the Dietary Gap on nutrient level is as follows: (1) total household consumption in grams per food item is calculated, (2) food items from the survey are matched with food items in a food composition table (e.g. the SMILING D3.5-a food composition table for Viet Nam), (3) nutrient intake on household level is calculated, (4) using a consumer unit nutrient intake on individual level is calculated, (5) the EAR for each individual in the household is calculated, (6) the nutrient adequacies for each individual are calculated (Talsma et al., 2017). However, it is beyond the scope of this thesis to convert the DASH Diet, which is an US Diet Eating Plan, to the Vietnamese context - especially given the fact that I am not a nutritionist. I argue that using the DASH Diet in its current format does not fit with the local context in Viet Nam, e.g. considering its recommendations of 2 to 4 servings of low-fat or non-fat dairy food a week (Heller, 2017). Unfortunately there currently is no specified Vietnamese diet plan available which could do the trick as well. Furthermore, there are quite a few food items in the VHLSS which are aggregates for which no quantity information is available (e.g. 'food away from home', 'other fruits', and 'other vegetables'). As these aggregates impose difficulties with both the calculation of consumed grams and the calculation of nutrient intake, I choose to not identify the Dietary Gap indicators.
Micronutrient indicators: **a) Vitamin A** **b) Iron**	Each food item has different levels of micronutrients. Some foods are good sources of e.g. vitamin A or iron, which are important nutrients for a healthy diet. The Food and Agricultural Organisation of the United Nations (FAO) (2011b) provides a list of food items rich in vitamin A and food items rich in iron. By using this list the total consumed grams of vitamin A and iron could be calculated. This could be done as follows: (1) total household consumption in grams per food item is calculated, (2) food items from the survey are matched with food items in the SMILING D3.5-a food composition table for Viet Nam, (3) nutrient intake on vitamin A and iron on household level is calculated, (4) vitamin A and iron

	intake on individual level is calculated using adult equivalence scales. Furthermore, as an indicator at the aggregated level, the percentage of individuals/households consuming food items rich in vitamin A or iron could be calculated as well (FAO, 2011b). However, both vitamin A and iron indicators could not be calculated as first of all the conversion to consumed grams of micronutrients requires data about dietary intake at a detailed level and the VHLSS has quite a few aggregated food items. For example, in the VHLSS questions are asked about the consumed grams of pork meat, but no specification is made between organ (rich in vitamin A) and flesh meat. Therefore I could not calculate the micronutrient intake in grams, as well as the percentage of households consuming foods rich in these micronutrients. This also means that I could not include both indicators into the analysis.
Soft drinks indicator	Sugar-sweetened beverages (SSBs), and in particular soft drinks, are associated with high added sugar content, low satiety, and incomplete compensation for total energy (Malik et al., 2006). Therefore it is often discouraged to consume sugary drinks in favour of a healthy lifestyle. It would be interesting to analyse the SSB consumption of Vietnamese households. However, in the VHLSS only an aggregated food item is included containing all bottled, canned, and boxed beverages – from pure water to juice fruit smoothies to soft drinks. Because of this it was not possible to analyse SSB consumption as an indicator on its own, which is why I could not include a soft drinks indicator into the analysis.

Appendix C

Chapter 3 described the methodological approach of Propensity Score Matching (PSM). In this Appendix the propensity score estimation is shown for all matching exercises of this thesis. Furthermore, balance tests are performed on each executed matching algorithm and for each identified subsample.

C.1 Matching for Estimating the Effect of Rural-Urban Migration on Diet

C.1.1 Estimation of Propensity Score

When estimating the propensity score, first of all a choice has to be made for which functional form to use for the estimation. The literature proposes the use of logit or probit models (Caliendo & Kopeinig, 2005; Ham et al., 2005; Sparrow, 2017), due to shortcomings of the linear probability model for this particular case. Caliendo & Kopeinig (2005) argue that both logit and probit models usually yield similar results, which induces that the choice is not too critical. In this thesis the propensity score $P(Xi)$ is estimated by means of logit estimation as this provides favorable characteristics for implementing the Rosenbaum bounds as a sensitivity analysis of the estimated treatment effects (more on this in subsection 3.2.5).

As described before, all covariates Xi should be included in the matching exercise which affect both the propensity to migrate and outcome variables of households. In this way the covariates control for the observable differences between the treatment (migrant) and control (rural non-migrant) group. However, the chosen variables should either be fixed over time or measured before migration to ensure that they are unaffected by (anticipation of) rural-urban migration (Caliendo & Kopeinig, 2005). Furthermore, the choice for each variable should be based on economic theory and previous empirical findings. The number of variables included into the model should be based on a trade-off between bias and efficiency; the inclusion of more (less) variables decreases (increases) bias, and increases (decreases) variance. Because of this trade-off, I tried to limit the inclusion of covariates which are correlated with each other. In order to make sure that rural-urban migrant households are matched to rural households coming from the same region, I selected a subsample consisting of only those households living in rural areas in the Red River Delta (RRD) (N= 1.345) and those migrant households from which the head of the household was born in the RRD (N= 194). This was needed in order to control for regional differences, e.g. climate, geography, food environment and availability of public goods, which might both affect the propensity to migrate and dietary intake. The RRD, which includes Viet Nam's capital Hanoi, is one the regions in Viet Nam experiencing increases in industrial capital after implementation of the *Doi Moi* economic and political reforms - more than other regions within the country (Anh et al., 2012; General Statistics Office of Viet Nam [GSO] & United Nations Population Fund [UNFPA], 2016b). The region even became one of Viet Nam's main locations of industrial zones. It is a highly populated area, with active economic development (Thanh, 2003). Furthermore, the implementation of the Agriculture for Nutrition and Health (A4NH) Food Systems for Healthier Diets (FSHD) project by the International Center for Tropical Agriculture (CIAT) has three areas of focus that are all located in the Red River Delta. These are (1) Cầu Giấy district, Hanoi (2) Đông Anh district, Hanoi, and (3) Mộc Châu, Sơn La province. By focusing on the Red River Delta in some parts of this thesis, this increases the ways in which this thesis can contribute to the A4NH FSHD project. Within this subsample the propensity of being a rural-urban migrant household compared to having stayed in the rural area was estimated.

Several estimations have been executed. Covariates that were included are:

- *Gender of household head*: I expect the head of the household to be heavily involved in making migration decisions of the household. Men and women are often said to have different occupations (Agergaard, 2011), and as rural-urban migration is often the consequence of an economic decision (GSO & UNFPA, 2016b), I expect the gender of the household head to influence the propensity to migrate. Anh et al. (2012) argue that 53.4 percent of all interprovincial migrants which migrate from rural to urban areas are female, which adds support to the argument that the gender of the household head influences migration. Furthermore, in traditional households, women are often accountable for domestic duties such as taking care of the children and preparing and cooking meals. Women are often seen as having more nurturing and caring characteristics than men as well. Therefore I expect the gender of the household head, who I expect to be involved in making decisions about household consumption expenditures, to be of influence for the household's dietary intake.
- *Age of household head and spouse:* According to the GSO & UNFPA (2016b) migrants are relatively young, which is related to work or education opportunities. Furthermore they argued that economic reasons are the main reasons for migration. This shows that age is negatively correlated to migration, and that, thus, migration is influenced by age. However, although migration has been an event in the past and we do not know when households have migrated, I argue that age is still an important variable which we should control for, especially as it influences dietary intake as well. The model of dietary determinants in chapter 2 shows that age is part of the biological determinant of diets (e.g. babies eat differently than adults). Furthermore, I expect that older people might put higher importance on more traditional diets than younger people.
- *Age of household head and spouse, squared:* The squared term of age of the household head and age of the spouse is also included as a covariate. This as I expect age to be non-linearly related to the propensity to migrate and decision-making about food consumption expenditures. I expect that when the head of the household and the spouse get older, the effect of age on migration and on decision-making about food consumption expenditures is lessened.
- *Age of household head and spouse, cubed:* The cubed term of age of the household head and age of the spouse is included as a covariate as well, in order to achieve balance across covariates (more on this below).
- *Education of household head and spouse:* Education of the household head and spouse might influence job opportunities, as generally speaking a higher level of education is associated with 'white collar' occupations (i.e. professional jobs), whereas a lower level of education is associated with 'blue collar' labour (i.e. manual labour) I expect people with higher levels of education to have a higher propensity to migrate, as often urban areas are expected to provide better employment opportunities (Duc et al., 2012). Moreover, education might influence someone's social network, which is also associated with migration (Acosta, 2006; Tu et al., 2008). Furthermore, education is argued to influence diet via the diet determinants of 'physical' and 'attitudes, beliefs and knowledge' (see model of dietary determinants in chapter 2). For example, knowledge about food nutrients might induce a decreased consumption of fast food with high proportions of unhealthy calories. It is assumed that education is positively correlated with knowledge about diet, and that the head and spouse of the household are both highly involved in decision making about food consumption expenditures. Furthermore, education of the household head and spouse can be seen as a proxy for the earnings of the household via earning aspirations, which in turn influences a household's expenditures and thus dietary intake.

Initially *ethnicity* was included in the propensity estimation as well, as Coxhead et al. (2015) found that members of ethnic minority groups, which are groups other than 'Kinh' (82.58 percent of the Vietnamese population belongs to the 'Kinh' ethnic group), are much less likely to migrate. Furthermore I expected ethnicity to be of large influence for dietary intake, as traditions and culture influence diet (see model of dietary determinants in chapter 2), and most ethnic minority groups live in mountainous areas in the north - an area with a different geography and climate than the rest of Viet Nam. Furthermore, the National Institute of Nutrition (2011) argued that especially mothers and family members living in rural, remote/isolated areas and ethinc groups, have improper nutrition knowledge and practices, providing additional support for the argument that ethnicity likely influences dietary intake. However, in the mentioned subsample all migrant household belong to the 'Kinh' ethnic group, and therefore ethnicity was omitted from the estimation model as it predicted failure perfectly. A list of the included covariates and their properties is included in table C.1 below.

Table C.1: Covariates and their description

Covariate	Description
Gender of household head	Dummy variable for the gender of the household head: =0 if female =1 if male.
Age of household head and spouse	Continuous variable for age of the household head and spouse at time of the survey.
Age of household head and spouse, squared	Continuous variable for the squared term of the household head and spouse at time of the survey.
Age of household head and spouse, cubed	Continuous variable for the cubed term of the household head and spouse at time of the survey.
Education of household head and spouse	Categorical variable for highest obtained qualification: =0 if no qualification =1 if primary qualification =2 if lower secondary qualification =3 if higher secondary qualification =4 if college, university, MA/MSc, or PhD. Each category is included in the propensity score estimation as dummy variable.
Ethnicity	Dummy variable for the ethnicity of the household: =0 if belonging to the 'Kinh' ethnic group =1 if belonging to an ethnic minority group.

The estimation results of the probability of having migrated to an urban area compared to having stayed behind in a rural area are shown in table C.2 below.

Table C.2: Logit estimation for the propensity of rural-urban migration compared to having stayed behind in a rural area

Covariate		Coefficient
Gender of head		-1.29*** (0.34)
Age of head		0.31 (0.41)
Age of head, squared		-0.00 (0.01)
Age of head, cubed		0.00 (0.00)
Education of head	**=1**	-0.73 (0.53)
	=2	-0.65 (0.50)
	=3	-0.02 (0.52)
	=4	1.24** (0.57)

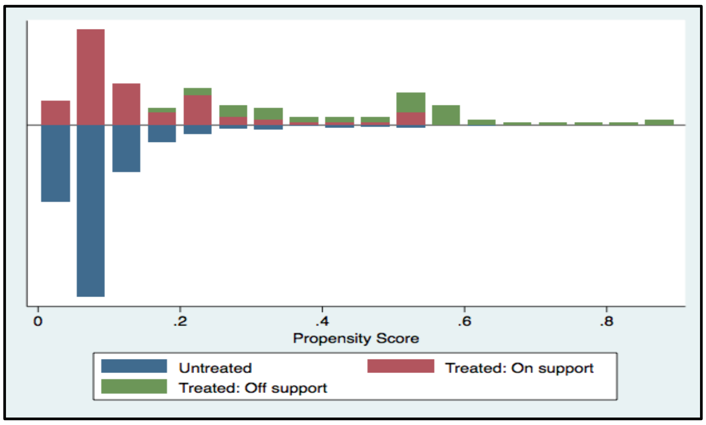
Age of spouse	-0.46 (0.33)
Age of spouse, squared	0.01 (0.01)
Age of spouse, cubed	-0.00 (0.00)
Education of spouse =1	-0.06 (0.48)
=2	-0.41 (0.48)
=3	0.48 (0.52)
=4	0.77 (0.57)
Constant	-0.07 (4.78)
N= 1.182, Ps R^2= 0.1550, LR test (prob)= 135.95 (0.000)*** Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1	

The estimation results in table C.2 above show first of all that the selection model is significant (p= 0.000). Furthermore, the propensity to be a rural-urban migrant household is positively and significantly influenced at the 5 percent level by the household head having obtained educational qualification at either college, university, MA/MSc or PhD level. The propensity to be a rural-urban migrant household is negatively and significantly influenced at the 1 percent level if the household head is male. The average propensity score is estimated to equal 0.1193, with a standard deviation of 0.1280. However, Caliendo & Kopeinig (2005) and Rosenbaum (2010) argue that it should be kept in mind that the main purpose of propensity score estimation is to balance all covariates across treatment and control group, instead of predicting selection into treatment as best as possible. We turn to matching and balance checks in the next subsection.

C.1.2 Matching Methods and Balance Tests

Following from the common support assumption, matching is only possible over a similar range of distribution of $P(Xi)$ for migrant and rural non-migrant households. Therefore the matching exercise is restricted to the range of common support (i.e. the range where the distributions of the propensity score of both groups overlap), which induces households which fall outside this range to be dropped from the sample. In this case the common support in the propensity score ranges from 0.024811 to 0.740591, inducing that 45 treated and 0 control households are dropped from the sample, leaving us with 96 treated households and 1.041 control households. The other 53 treated households that are dropped from the sample are households which have no spouse, as these households have missing values for the matching variables associated with the spouse. Figure 3.2 below shows a graphical representation of the common support. As can be seen from the figure, most of the treated households that are dropped from the sample have a relatively high propensity score compared to all other households. The common support assumption add to the generalizability of the estimated treatment effects (Nichols, 2007).

Figure C.1: Common support



Several matching methods exist for propensity score matching: (1) nearest neighbour matching, (2) caliper and radius matching, (3) stratification and interval matching, (4) kernel and local linear matching, and (5) weighting by the propensity score (Caliendo & Kopeinig, 2005). Each method has its own way of defining the neighbourhood for each treated individual, handling the common support problem, and assigning weights to neighbours. Basically all these methods aim to recreate an experimental design ex-post, by re-weighting the data based on $P(Xi)$ (Sparrow, 2017). However, the performance of each matching method depends largely on the situation at hand (Caliendo & Kopeinig, 2005), inducing that there is no general guideline on which matching method to choose. Again trade-offs exist in terms of bias and efficiency (Caliendo & Kopeinig, 2005), which is why I have chosen to try and compare several matching approaches.

In the case of this subsample we have 96 treated and 1.041 control households, so it made sense to look into cases of oversampling or kernel matching in order to gain more precision in the estimates (Caliendo & Kopeinig, 2005). Matching has been executed by (1) nearest neighbour matching with five neighbours, (2) nearest neighbour matching with five neighbours within a specified caliper (caliper= 0.001), (3) kernel matching, and (4) caliper and radius matching (caliper= 0.001), where the tolerance level of 0.001 is based on a rule of thumb widely used in the PSM literature (Anders & Schroeter, 2010; Best & Wolf, 2014; Groth et al., 2017; Vogt, 2011). When analysing each of these methods, all approaches gave more or less similar results. However, the matching method of caliper and radius matching produced the best balance across covariates (more on this below), which is why this method has been chosen for the main analysis in chapter 5. Moreover, Austin (2013) has compared twelve different algorithms for propensity score matching, and argues in favour of some desirable characteristics of caliper and radius matching.

The method of caliper matching imposes a tolerance level, also known as the 'caliper', on the maximum propensity score distance between treated and control households (c= 0.001). Caliper matching ensures that those households from the comparison group are chosen as a matching partner for a treated household that lie within the caliper ('propensity range') and are closest in terms of propensity score (Caliendo & Kopeinig, 2005). By using a caliper we avoid the making of bad matches, which increases the quality of the matching. In this thesis we combine caliper with radius matching, an approach suggested by Dehejia & Wahba (2002), which imposes that all of the comparison households within the caliper are used as a matching partner for the treated household. In our case this is a favourable approach in terms of variance, due to the high number of comparison observations compared to treatment observations. The approach allows for usage of extra (fewer) units when good matches are (not) available, which also reduces the bias of the estimates. In this thesis PSM is implemented in STATA by the command **psmatch2**, developed by Leuven and Sianesi (2003).

Matching has been executed by nearest neighbour matching with only one neighbour as well, as this is the only matching algorithm which allows for the implementation of Rosenbaum bounds using the STATA command **rbounds** (see also subsection 3.2.5). In this matching exercise one household from the rural-urban migrant household group is matched to only one household in the rural non-migrant group (Caliendo & Kopeinig, 2005). Again, the procedure is implemented in STATA by the command **psmatch2**.

As stressed before, an important condition for having a correctly specified model is balanced covariates across treatment (migrant) and control (rural non-migrant) group. This means that the mean of each element of Xi should be identically distributed across both groups (Caliendo & Kopeinig, 2005; Ham et al., 2005). There are several ways to assess the balance, and we perform multiple tests of them. The output of the balancing test for PSM using radius caliper matching is shown in table C.3 below. In this case 96 rural-urban migrant households are matched to 1041 rural non-migrant households. These rural non-migrant households are households that live in the Red River Delta (RRD).

Table C.3: Output of balancing test for PSM, radius caliper matching

	Sample	Mean Treated	Mean Control	Difference	%Bias	V(T)/V(C)
Gender of head	Unmatched	0.79	0.95	-0.16***	-49.4	
	Matched	0.85	0.89	-0.04	-10.3	
Age of head	Unmatched	51.65	50.47	1.18	9.6	0.90
	Matched	51.83	52.44	-0.61	-4.9	1.23
Age of head, squared	Unmatched	2808.10	2704.80	103.30	7.9	0.87
	Matched	2832.40	2867.90	-35.50	-2.7	1.22
Age of head, cubed	Unmatched	1.6e+05	1.5e+05	1.0e+04	5.8	0.83
	Matched	1.6e+05	1.6e+05	0.00	-0.6	1.21
Education of head =1	Unmatched	0.10	0.18	-0.08**	-22.6	
	Matched	0.14	0.15	-0.01	-4.8	
=2	Unmatched	0.32	0.54	-0.22***	-46.5	
	Matched	0.44	0.43	0.01	1.5	
=3	Unmatched	0.28	0.20	0.08**	19.3	
	Matched	0.28	0.30	-0.02	-5.2	
=4	Unmatched	0.26	0.04	0.22***	64.6	
	Matched	0.07	0.06	0.01	4.1	
Age of spouse	Unmatched	49.70	47.85	1.85	14.6	0.99
	Matched	49.65	50.05	-0.40	-3.2	1.18
Age of spouse, squared	Unmatched	2627.40	2450.30	177.10	13.8	0.98
	Matched	2623.00	2639.10	-16.10	-1.2	1.18
Age of spouse, cubed	Unmatched	1.5e+05	1.3e+05	2.0e+04	12.3	0.95
	Matched	1.5e+05	1.5e+05	0.00	0.5	1.17
Education of spouse =1	Unmatched	0.15	0.20	-0.05	-14.4	
	Matched	0.20	0.21	-0.01	-1.9	
=2	Unmatched	0.31	0.55	-0.24***	-50.3	
	Matched	0.44	0.42	0.02	3.4	

=3	Unmatched	0.26	0.12	0.14***	34.5	
	Matched	0.21	0.22	-0.01	-4.3	
=4	Unmatched	0.23	0.05	0.18***	53.3	
	Matched	0.07	0.08	-0.01	-2.9	

*** p<0.01, ** p<0.05, * p<0.1
~ if variance ratio outside [0.72; 1.39] for unmatched and [0.67; 1.50] for matched

Unmatched: Ps R^2= 0.156, LR test (prob)= 134.44 (0.000)***, MeanBias= 27.9, MedBias= 19.3, B= 101.3, R= 2.71
Matched: Ps R^2= 0.010, LR test (prob)= 2.71 (1.000), MeanBias= 3.4, MedBias= 3.2, B= 23.7, R= 1.36

First of all the standardised bias is analysed for each covariate, before and after matching. The standardised bias is the difference of sample means in the treated and matched control subsample as a percentage of the square root of the average of sample variances in both groups (Caliendo & Kopeinig, 2005; Rosenbaum & Rubin, 1985). We follow Caliendo & Kopeinig (2005) who argue that a bias reduction below 3 or 5 percent is usually seen as sufficient for arguing in favour of the success of the matching procedure. As can be seen in table C.3, all covariates have a bias reduction less than 5 percent, except for the covariate of gender of the head of the household. However, as the associated bias reduction is not too far from 5 percent (i.e. 10.3), I argue that according to the standardized difference analysis balance has been achieved across treatment and control group. Furthermore all bias percentages were above 5 percent before matching, which indicates that matching has helped reducing bias associated with observable characteristics. Nevertheless, other balance tests are performed as well in order to further validate the balance assessment.

The next balancing test I turn to is two-sample t-tests. When performing these tests we compare differences in covariate means for both groups. As can be seen from table C.3 for all covariates we fail to reject the null hypothesis that the mean of the covariate is different for treatment and control group. The table also shows that, with the exception of five covariates (i.e. age and age squared of the household head and spouse, and the spouse having obtained primary qualification), the means of the covariates were significantly different from each other before matching. The two-sample t-tests provide additional support that the matching exercise has helped to reduce bias associated with observable characteristics.

Another balancing test that can be performed is the evaluation of the ratio of variances. The ratio of the variance of the covariate in the treated group over the control group should be near one if the covariate is balanced. The ratio of variances in table C.3 is only reported for continuous variables. The variance ratio of these covariates is near to one, and therefore I argue that this again supports the argument that the matching exercise has been successful in reducing bias. However, in this case this test is less strong as the variance ratio was also near to one before matching.

Lastly we turn to some summary statistics, illustrated in the bottom row of table C.3. Before matching the mean and median bias were estimated to be 27.9 and 19.3 respectively, whereas they decreased to 3.4 and 3.2 after matching. This suggest great reduction of bias due to matching. Furthermore after matching the pseudo R^2 should be close to zero to ensure that the covariates have no explanatory power in the matched samples, and given that this statistic equals 0.010 after matching I argue that this is another argument for balance across samples (p= 1.000). As after matching Rubin's B (i.e. the absolute standardized difference of the means of the linear index of the propensity score in both groups) is below 25 percent, and Rubin's R (i.e. the ratio of treated to control variances of the propensity score index) is between 0.5 and 2, this again provides additional support for sufficient balance of the matched sample.

As multiple tests have indicated balance across covariates in both treatment and control group, I argue that balance has been achieved. This indicates that matching has helped creating a balanced sample, which induces that based on observable characteristics the only thing that makes the treatment (migrant) and control (rural non-migrant) group different is participation into treatment. As propensity score matching assumes that unobservables are equal in expectation for treatment and control observations, we conclude by arguing that the matching exercise has produced two comparable groups.

Balancing tests for one to one nearest neighbour matching have been executed as well. The output of the balancing tests is shown in table C.4 below. In this case 96 rural-urban migrant households are matched to 96 rural non-migrant households. These rural non-migrant households are households that live in the Red River Delta (RRD).

Table C.4: Output of balancing test for PSM, nearest neighbour matching

	Sample	Mean Treated	Mean Control	Difference	%Bias	V(T)/V(C)
Gender of head	Unmatched	0.79	0.95	-0.16***	-49.4	
	Matched	0.85	0.90	-0.05	-12.9	
Age of head	Unmatched	51.65	50.47	1.18	9.6	0.90
	Matched	51.83	51.63	0.20	1.7	1.62~
Age of head, squared	Unmatched	2808.10	2704.80	103.30	7.9	0.87
	Matched	2832.40	2755.30	77.10	5.9	1.61~
Age of head, cubed	Unmatched	1.6e+05	1.5e+05	1.0e+04	5.8	0.83
	Matched	1.6e+05	1.5e+05	1.0e+04	9.3	1.63~
Education of head =1	Unmatched	0.10	0.18	-0.08**	-22.6	
	Matched	0.14	0.13	0.01	3.0	
=2	Unmatched	0.32	0.54	-0.22***	-46.5	
	Matched	0.44	0.42	0.02	4.3	
=3	Unmatched	0.28	0.20	0.08**	19.3	
	Matched	0.28	0.35	-0.07	-17.2	
=4	Unmatched	0.26	0.04	0.22***	64.6	
	Matched	0.07	0.05	0.02	6.2	
Age of spouse	Unmatched	49.70	47.85	1.85	14.6	0.99
	Matched	49.65	49.10	0.55	4.3	1.58~
Age of spouse, squared	Unmatched	2627.40	2450.30	177.10	13.8	0.98
	Matched	2623.00	2511.70	111.30	8.7	1.62~
Age of spouse, cubed	Unmatched	1.5e+05	1.3e+05	2.0e+04	12.3	0.95
	Matched	1.5e+05	1.3e+05	2.0e+04	12.1	1.75~
Education of spouse =1	Unmatched	0.15	0.20	-0.05	-14.4	
	Matched	0.20	0.19	0.01	2.7	
=2	Unmatched	0.31	0.55	-0.24***	-50.3	
	Matched	0.44	0.45	-0.01	-2.2	

=3	Unmatched	0.26	0.12	0.14***	34.5	
	Matched	0.21	0.23	-0.02	-5.4	
=4	Unmatched	0.23	0.05	0.18***	53.3	
	Matched	0.07	0.09	-0.02	-6.2	

*** p<0.01, ** p<0.05, * p<0.1
~ if variance ratio outside [0.72; 1.39] for unmatched and [0.67; 1.50] for matched

Unmatched: Ps R^2= 0.156, LR test (prob)= 134.44 (0.000)***, MeanBias= 27.9, MedBias= 19.3, B= 101.3, R= 2.71
Matched: Ps R^2= 0.042, LR test (prob)= 11.26 (1.000), MeanBias= 6.8, MedBias= 5.9, B= 48.5, R= 1.85

First of all, the standardised bias tests indicate that for the household head the covariates of gender, age squared and cubed, and the third and fourth obtained level of education have a bias reduction of more than 5 percent. The associated bias reduction ranges from 5.9 to -17.2. For the spouse of the household head the covariates of age squared and cubed and the third and fourth obtained level of education have a bias reduction which still is too large after matching. However, the associated bias reduction is only 5.4 and 6.2 for the latter two covariates. As quite some covariates have a bias reduction that is too large after matching, the standardized difference analysis does not indicate balance across both groups.

The next balancing test I turn to is two-sample t-tests. Table C.4 shows that for all covariates we fail to reject the null hypothesis that the mean of the covariates is different for treatment and control group. However, for seven of the fifteen covariates this was the case before matching as well. For the other eight covariates, two-sample t-tests provide support for the argument that the matching exercise has helped to reduce bias associated with observable characteristics.

The ratio of variance test shows that for all continuous variables, the ratio is not near to one after matching. On the contrary, this was the case before matching. Therefore the ratio of variance tests argues in favour of imbalance after matching.

Lastly we turn to some summary statistics. Before matching the mean and median bias were estimated to be 27.9 and 19.3 respectively, whereas they decreased to 6.8 and 5.9 after matching. This suggest reduction of bias due to matching. Furthermore after matching the pseudo R^2 equals 0.042, which is quite close to zero. This is an argument in favour of balance across samples (p= 1.000). Rubin's B is not below 25 percent after matching, while Rubin's R is between 0.5 and 2. This latter statistic provides support for balanced groups, whereas the former does not.

Multiple tests are not too convincing in arguing in favour of balanced covariates in both treatment and control group. However, the main analysis of this thesis focuses on the results produced using radius caliper matching. The nearest neighbour matching procedure is used only for the implementation of Rosenbaum bounds.

C.2 Matching for Estimating the Effect of Having Rural Origins on Diet

C.2.1 Estimation of Propensity Score

For estimating the effect of having rural origins compared to being born in urban areas, rural-urban migrant households are matched to (a) urban non-migrant households living in the Red River Delta (RRD) and (b) urban non-migrant households living in the South-Eastern Area (SEA) of Viet Nam. By matching to urban non-migrant households it is ensured that the food environment, i.e. the economic and physical determinants of diet, are similar across households. First of all the subsample of the RRD was selected, which contains 66 rural-urban migrant households and 498 urban non-migrant households. However, as most rural-urban migrant households migrated to the South-Eastern Area (SEA), the analysis has been executed also on this second subsample. This sample contains 203 rural-urban migrant households and 362 urban non-migrant households. The SEA has the highest number of migrants from all regions in the country (GSO, 2016; GSO and UNFPA, 2016a). This as it contains large cities, like Ho Chi Minh City, and has new industrial zones and vibrant economies which make the region economically attractive. The region has even attracted a large number of intra-regional migrants and migrants from other regions (GSO and UNFPA, 2016a). Within both subsamples the propensity of having rural origins compared to being born in the urban area was estimated.

Like when matching rural-urban migrant households to rural non-migrant households, the propensity score $P(Xi)$ is estimated by means of logit estimation. Furthermore, the same covariates are included in the matching exercise in the case for the RRD. In the case of the SEA, however, the cubed terms of the age of the household head and spouse are excluded. These cubed terms are excluded in the matching exercise as they were not needed to create balance across covariates (more on this below). The results of the propensity score estimation in both subsamples is shown in table C.5 below.

Table C.5: Logit estimation for the propensity of having rural origins compared to being born in an urban area

		RRD	*SEA*
Covariate		**Coefficient**	**Coefficient**
Gender of head		0.33 (0.56)	0.17 (0.31)
Age of head		-1.20* (0.70)	-0.08 (0.10)
Age of head, squared		0.02* (0.01)	0.00 (0.00)
Age of head, cubed		-0.00* (0.00)	
Education of head	**=1**	-0.59 (1.58)	-0.27 (0.46)
	=2	0.11 (1.29)	-0.12 (0.49)
	=3	0.56 (1.32)	-0.18 (0.52)
	=4	1.67 (1.35)	0.52 (0.59)
Age of spouse		-0.47 (0.62)	-0.13 (0.10)
Age of spouse, squared		0.01 (0.01)	0.00 (0.00)
Age of spouse, cubed		-0.00 (0.00)	
Education of spouse	**=1**	0.13 (1.37)	0.33 (0.44)
	=2	-0.63 (1.35)	1.10** (0.48)
	=3	-1.03 (1.38)	0.56 (0.51)

=4	-0.12 (1.38)	-0.25 (0.62)
Constant	22.34*** (7.67)	4.33*** (1.51)
RRD: N= 436, Ps R^2= 0.1470, LR test (prob)= 43.20 (0.0001)*** SEA: N= 385, Ps R^2= 0.0706, LR test (prob)= 35.13 (0.008)*** Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

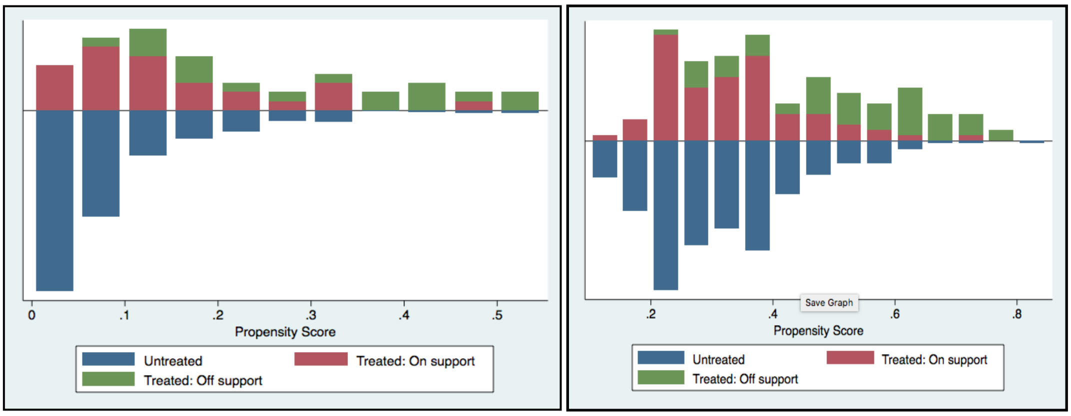
The estimation results in table C.5 above show first of all that both selection models are significant (p= 0.0001 and p= 0.008 for the RRD and SEA respectively). Furthermore, for the RRD the coefficients of no covariates are significant at at least the 5 percent level. For the SEA, however, the propensity to be a rural-urban migrant household is positively and significantly influenced at the 5 percent level by the spouse of the household head having obtained lower secondary qualification.

The average propensity score is estimated to equal 0.1055 for the RRD and 0.3481 for the SEA, with a standard deviation of 0.1066 and 0.1435 respectively. However, Caliendo & Kopeinig (2005) and Rosenbaum (2010) argue that it should be kept in mind that the main purpose of propensity score estimation is to balance all covariates across treatment and control group, instead of predicting selection into treatment as best as possible. We turn to matching and balance checks in the next subsection.

C.2.2 Matching Methods and Balance Tests

Similar to the matching exercise of matching to rural non-migrant households, the matching in this case is restricted to the range of common support (i.e. the range where the distributions of the propensity score of both groups overlap), which induces households which fall outside this range to be dropped from the sample. In the case of the RRD the common support in the propensity score ranges from 0.0297714 to 0.5187103, inducing that 18 treated and 0 control households are dropped from the sample, leaving us with 28 treated households and 390 control households. The other 20 treated and 108 control households that are dropped from the sample are households which have no spouse, as these households have missing values for the matching variables associated with the spouse. In the case of the SEA the common support of the propensity score ranges from 0.111502 to 0.781999, inducing that 54 treated and 0 control households are dropped from the sample, leaving us with 80 treated households and 251 control households. The other 69 treated households and 111 control households have missing values for the matching variables (e.g. matching variables associated with the spouse when households have no spouse). Figure C.2 below shows a graphical representation of the common support of both subsamples (RRD on the left; SEA on the right). As can be seen from the figure, most of the treated households that are dropped from the sample have a relatively high propensity score compared to all other households. The common support assumption add to the generalizability of the estimated treatment effects (Nichols, 2007).

Figure C.2: Common support



RRD *SEA*

As described in subsection C.1.2, several matching methods exist for propensity score matching. For the matching of rural-urban migrant households to urban non-migrant households, the same rationale is followed for choosing the matching algorithm as for the matching of rural-urban migrant households to rural non-migrant households (e.g. oversampling and comparing the balance achievement of different matching methods). The matching method of caliper (c= 0.001) and radius matching produced the best balance across covariates (more on this below), which is why this method has been chosen for the analysis. Furthermore matching has again been exectued by nearest neighbour matching with only one neighbour as well, as this is the only matching algorithm which allows for the implementation of Rosenbaum bounds using the STATA command **rbounds**. PSM has been implemented in Stata by the command **psmatch2**, developed by Leuven and Sianesi (2003).

As stressed in chapter 3, an important condition for having a correctly specified model is balanced covariates across treatment (migrant) and control (urban non-migrant) group. This means that the mean of each element of Xi should be identically distributed across both groups (Caliendo & Kopeinig, 2005; Ham et al., 2005). There are several ways to assess the balance, and we perform multiple tests of them. The output of the balancing tests for radius caliper and nearest neighbour matching on both subsamples RRD and SEA is shown in tale C.6 to C.9 below. In the case of radius caliper matching within the subsample of the RRD, 28 rural-urban migrant households are matched to 390 urban non-migrant households.

Table C.6: Output of balancing test for PSM, radius caliper matching (RRD)

	Sample	Mean Treated	Mean Control	Difference	%Bias	V(T)/V(C)
Gender of head	Unmatched	0.80	0.82	-0.02	-4.8	
	Matched	0.75	0.83	-0.08	-19.5	
Age of head	Unmatched	53.00	50.41	2.59	21.3	1.25
	Matched	52.46	51.33	1.13	9.4	1.12
Age of head, squared	Unmatched	2969.80	2672.30	297.50	23.9	1.11
	Matched	2893.50	2760.10	113.40	10.7	1.25
Age of head, cubed	Unmatched	1.7e+05	1.5e+05	2.0e+04	24.2	0.99
	Matched	1.7e+05	1.5e+05	2.0e+04	12.1	1.44
Education of head =1	Unmatched	0.02	0.09	-0.07	-29.0	
	Matched	0.04	0.03	0.01	1.4	
=2	Unmatched	0.20	0.35	-0.15**	-34.2	

		Matched	0.29	0.27	0.02	4.5	
	=3	Unmatched	0.26	0.37	-0.11	-22.8	
		Matched	0.25	0.33	-0.08	-17.0	
	=4	Unmatched	0.50	0.17	0.33***	72.8	
		Matched	0.43	0.35	0.08	17.3	
Age of spouse		Unmatched	50.78	47.82	2.96	23.1	1.21
		Matched	50.32	48.99	1.33	10.4	0.89
Age of spouse, squared		Unmatched	2754.90	2434.60	322.30*	25.4	1.123
		Matched	2684.00	2570.70	113.30	9.0	0.99
Age of spouse, cubed		Unmatched	1.6e+05	1.3e+05	3.0e+04	25.2	1.04
		Matched	1.5e+05	1.4e+05	1.0e+04	8.0	1.08
Education of spouse	=1	Unmatched	0.07	0.08	-0.01	-7.3	
		Matched	0.07	0.10	-0.03	-12.3	
	=2	Unmatched	0.24	0.37	-0.13*	-28.9	
		Matched	0.29	0.30	-0.01	-2.9	
	=3	Unmatched	0.20	0.34	-0.14	-32.5	
		Matched	0.25	0.28	-0.03	-7.7	
	=4	Unmatched	0.48	0.18	0.30***	65.2	
		Matched	0.39	0.30	0.09	19.7	

*** p<0.01, ** p<0.05, * p<0.1
~ if variance ratio outside [0.55; 1.81] for unmatched and [0.46; 2.16] for matched

Unmatched: Ps R^2= 0.150, LR test (prob)= 43.94 (0.000)***, MeanBias= 29.4, MedBias= 25.2, B= 108.8, R= 0.90
Matched: Ps R^2= 0.073, LR test (prob)= 5.60 (0.960), MeanBias= 10.8, MedBias= 10.4, B= 64.3, R= 0.90

First of all, the standardised bias tests indicate that only three of the fifteen covariates (i.e. first and second obtained level of education of the household head and second obtained level of education of the spouse) have a bias reduction less than 5 percent after matching. All other covariates have a bias reduction which still is too large after matching for arguing in favour of balance across treatment and control group.

The next balancing test I turn to is two-sample t-tests. As can be seen from table C.8 for all covariates we fail to reject the null hypothesis that the mean of the covariate is different for treatment and control group. However, for twelve of the fifteen covariates this was the case before matching as well (at the 5 percent level). Therefore two-samplte t-tests are not too convincing in arguing in favour of balance after matching.

Another balancing test that can be performed is the evaluation of the ratio of variances. The variance ratio of the (continuous) covariates is near to one after matching, and therefore I argue that this supports the argument of balance across groups. However, in this case this test is less strong as the variance ratio was also near to one before matching.

Lastly we turn to some summary statistics, illustrated in the bottom row of table C.8. Before matching the mean and median bias were estimated to be 19.4 and 25.2 respectively, whereas they decreased to 10.8 and 10.4 after matching. This suggest small reduction of bias due to matching. Furthermore after matching the pseudo R^2 should be close to zero to ensure that the covariates have no explanatory power in the matched samples, and given that this statistic equals 0.073 after matching I argue that this is another argument for balance across samples (p= 0.960). However, although Rubin's R is between 0.5 and 2 after matching, Rubin's B is not below 25 percent.

Multiple tests are not too convincing in arguing in favour of balanced covariates in both treatment and control group. In fact, I would argue that balance has not been achieved and that, thus, treatment and control group are not comparable enough to produce unbiased results.

Balancing tests for one to one nearest neighbour matching within the RRD have been executed as well. The output of the balancing tests is shown in table C.7 below. In this case 28 rural-urban migrant households are matched to 28 urban non-migrant households.

Table C.7: Output of balancing test for PSM, nearest neighbour matching (RRD)

	Sample	Mean Treated	Mean Control	Difference	%Bias	V(T)/V(C)
Gender of head	Unmatched	0.80	0.82	-0.02	-4.8	
	Matched	0.75	0.79	-0.04	-9.1	
Age of head	Unmatched	53.00	50.41	2.59	21.3	1.25
	Matched	52.46	49.57	2.89	23.8	1.31
Age of head, squared	Unmatched	2969.80	2672.30	297.50	23.9	1.11
	Matched	2893.50	2564.80	328.70	26.4	1.60
Age of head, cubed	Unmatched	1.7e+05	1.5e+05	2.0e+04	24.2	0.99
	Matched	1.7e+05	1.4e+05	3.0e+04	28.4	2.02
Education of head =1	Unmatched	0.02	0.09	-0.07	-29.0	
	Matched	0.04	0.00	0.04	15.9	
=2	Unmatched	0.20	0.35	-0.15*	-34.2	
	Matched	0.29	0.25	0.04	8.1	
=3	Unmatched	0.26	0.37	-0.11	-22.8	
	Matched	0.25	0.36	-0.11	-23.1	
=4	Unmatched	0.50	0.17	0.33***	72.8	
	Matched	0.43	0.36	0.07	16.0	
Age of spouse	Unmatched	50.78	47.82	2.96	23.1	1.21
	Matched	50.32	47.86	2.46	19.2	1.03
Age of spouse, squared	Unmatched	2754.90	2434.60	322.30*	25.4	1.12
	Matched	2684.00	2438.00	246.00	19.5	1.32
Age of spouse, cubed	Unmatched	1.6e+05	1.3e+05	3.0e+04	25.2	1.04
	Matched	1.5e+05	1.3e+05	2.0e+04	20.1	1.79
Education of spouse =1	Unmatched	0.07	0.08	-0.01	-7.3	
	Matched	0.07	0.07	0.00	0.0	
=2	Unmatched	0.24	0.37	-0.13*	-28.9	
	Matched	0.29	0.25	0.04	7.8	

	Sample	Mean Treated	Mean Control	Difference	%Bias	V(T)/V(C)
=3	Unmatched	0.20	0.34	-0.14*	-32.5	
	Matched	0.25	0.32	-0.07	-16.3	
=4	Unmatched	0.48	0.18	0.30***	65.2	
	Matched	0.39	0.36	0.03	7.9	

*** p<0.01, ** p<0.05, * p<0.1
~ if variance ratio outside [0.55; 1.81] for unmatched and [0.46; 2.16] for matched

Unmatched: Ps R^2= 0.150, LR test (prob)= 43.94 (0.000)***, MeanBias= 29.4, MedBias= 25.2, B= 108.8, R= 0.90
Matched: Ps R^2= 0.149, LR test (prob)= 11.12 (1.000), MeanBias= 16.1, MedBias= 16.3, B= 90.4, R= 0.77

First of all, the standardised bias tests indicate that only one of the fifteen covariates (i.e. first obtained level of education of the spouse) has a bias reduction less than 5 percent after matching. All other covariates have a bias reduction which still is too large after matching for arguing in favour of balance across treatment and control group.

The next balancing test I turn to is two-sample t-tests. Table C.7 shows that for all covariates we fail to reject the null hypothesis that the mean of the covariates is different for treatment and control group. However, for thirteen of the fifteen covariates this was the case before matching as well (at the 5 percent level). Therefore two-samplte t-tests are not too convincing in arguing in favour of balance after matching.

Furthermore, the variance ratio of these covariates is near to one. However, in this case this test is less strong as the variance ratio was also near to one before matching.

Lastly we turn to some summary statistics. Before matching the mean and median bias were estimated to be 29.4 and 25.2 respectively, whereas they decreased to 16.1 and 16.3 after matching. This suggest small reduction of bias due to matching. Furthermore after matching the pseudo R^2 equals 0.149, which is not that close to zero. Rubin's B is not below 25 percent after matching, whereas Rubin's R is between 0.5 and 2.

Multiple tests are not convincing in arguing in favour of balanced covariates in both treatment and control group. In fact, I would argue that balance has not been achieved. However, the main analysis of this thesis focuses on the results produced using radius caliper matching. The nearest neighbour matching procedure is used only in order to be able to implement Rosenbaum bounds.

Turning to the subsample of the SEA, we first present the balancing tests results for using radius caliper matching (see table C.8 below). In this case 80 rural-urban migrant households are matched to 251 urban non-migrant households.

Table C.8: Output of balancing test for PSM, radius caliper matching (SEA)

	Sample	Mean Treated	Mean Control	Difference	%Bias	V(T)/V(C)
Gender of head	Unmatched	0.76	0.73	0.03	7.4	
	Matched	0.80	0.69	0.11	25.0	
Age of head	Unmatched	44.46	48.85	-4.39***	-35.2	1.16
	Matched	48.03	47.62	0.41	3.3	1.01
Age of head, squared	Unmatched	2142.70	2529.10	-386.40***	-30.8	0.99
	Matched	2446.60	2406.50	40.1	3.2	0.95
Education of head =1	Unmatched	0.21	0.29	-0.08*	-19.8	
	Matched	0.29	0.23	0.06	14.4	

	=2	Unmatched	0.24	0.21	0.03	7.6	
		Matched	0.23	0.30	-0.07	-17.9	
	=3	Unmatched	0.26	0.26	0.00	0.5	
		Matched	0.28	0.20	0.08	16.5	
	=4	Unmatched	0.20	0.14	0.06	16.5	
		Matched	0.11	0.12	-0.01	-2.2	
Age of spouse		Unmatched	42.65	47.14	-4.49***	-36.6	1.17
		Matched	45.78	46.27	-0.49	-4.0	0.88
Age of spouse, squared		Unmatched	1981.50	2363.70	-379.2***	-31.8	1.06
		Matched	2230.50	2293.70	-63.2	-5.3	0.78
Education of spouse	=1	Unmatched	0.21	0.29	-0.08*	-19.8	
		Matched	0.29	0.33	-0.04	-10.4	
	=2	Unmatched	0.33	0.20	0.13***	29.5	
		Matched	0.24	0.22	0.02	3.4	
	=3	Unmatched	0.25	0.22	0.03	8.1	
		Matched	0.24	0.20	0.04	8.6	
	=4	Unmatched	0.13	0.15	-0.02	-7.1	
		Matched	0.11	0.11	0.00	-0.6	

*** p<0.01, ** p<0.05, * p<0.1
~ if variance ratio outside [0.71; 1.41] for unmatched and [0.64; 1.56] for matched

Unmatched: Ps R^2= 0.071, LR test (prob)= 35.09 (0.001)***, MeanBias= 19.3, MedBias= 19.8, B= 64.0, R= 1.42
Matched: Ps R^2= 0.038, LR test (prob)= 8.41 (0.816), MeanBias= 8.8, MedBias= 5.3, B= 46.1, R= 0.53

First of all the standardized differences in the matched sample are analysed. As can be seen in table C.8, the covariates associated with age of the household head and spouse have a bias reduction (of more or less) less than 5 percent, as well as the covariates associated with the fourth obtained level of education of the head of the household, and the second and fourth obtained level of education of the spouse. For these covariates matching has helped reducing bias associated with observable characteristics. All other covariates have a bias reduction ranging between 8.6 and 25.0, indicating that just relying on the standardized difference analysis is not too convincing for arguing that balance has been achieved.

The next balancing test I turn to is two-sample t-tests. As can be seen from table C.8 for all covariates we fail to reject the null hypothesis that the mean of the covariate is different for treatment and control group. However the table also shows that with the exception of seven covariates, the means of the covariates were significantly different from each other before matching. The two-sample t-tests provide support for the argument that the matching exercise has helped to reduce bias associated with observable characteristics.

Another balancing test that can be performed is the evaluation of the ratio of variances. The variance ratio of the (continuous) covariates is near to one, and therefore I argue that this again supports the argument that the matching exercise has helped reducing bias. However, in this case this test is less strong as the variance ratio was also near to one before matching.

Lastly we turn to some summary statistics, illustrated in the bottom row of table C.8. Before matching the mean and median bias were estimated to be 19.3 and 19.8 respectively, whereas they decreased to 8.8 and 5.3 after matching. This suggest reduction of bias due to matching. Furthermore after matching the pseudo R^2 should be close to zero to ensure that the covariates have no explanatory power in the matched samples, and given that this statistic equals 0.038 after matching I argue that this is another argument for balance across samples (p= 0.816). However, although Rubin's R is between 0.5 and 2 after matching, Rubin's B is not below 25 percent.

Multiple tests have indicated balance across covariates in both treatment and control group, although some tests were stronger than others. However, as multiple tests point in more or less the same direction, I argue that balance has been achieved. This indicates that matching has helped creating a balanced sample, which induces that based on observable characteristics the matching exercise has produced two comparable groups.

Balancing tests for one to one nearest neighbour matching within the SEA have been executed as well. The output of the balancing tests is shown in table C.9 below. In this case 80 rural-urban migrant households are matched to 80 urban non-migrant households.

Table C.9: Output of balancing test for PSM, nearest neighbour matching (SEA)

	Sample	Mean Treated	Mean Control	Difference	%Bias	V(T)/V(C)
Gender of head	Unmatched	0.76	0.73	0.03	7.4	
	Matched	0.80	0.70	0.10	22.9	
Age of head	Unmatched	44.46	48.85	-4.39***	-35.2	1.16
	Matched	48.03	47.09	0.94	7.5	1.15
Age of head, squared	Unmatched	2142.70	2529.10	-386.40***	-30.8	0.99
	Matched	2446.60	2338.60	108.00	8.6	1.17
Education of head =1	Unmatched	0.21	0.29	-0.08*	-19.8	
	Matched	0.29	0.23	0.06	14.4	
=2	Unmatched	0.24	0.21	0.03	7.6	
	Matched	0.23	0.30	-0.07	-18.0	
=3	Unmatched	0.26	0.26	0.00	0.5	
	Matched	0.28	0.23	0.05	11.4	
=4	Unmatched	0.20	0.14	0.06	16.5	
	Matched	0.11	0.13	-0.02	-3.3	
Age of spouse	Unmatched	42.65	47.16	-4.51***	-36.6	1.17
	Matched	45.78	45.46	0.32	2.5	1.04
Age of spouse, squared	Unmatched	1981.50	2363.70	-342.20***	-31.8	1.06
	Matched	2230.50	2197.30	33.20	2.8	0.96
Education of spouse =1	Unmatched	0.21	0.29	-0.08*	-19.8	
	Matched	0.29	0.35	-0.06	-14.4	
=2	Unmatched	0.33	0.20	0.13***	29.5	
	Matched	0.24	0.23	0.01	2.9	
=3	Unmatched	0.25	0.22	0.03	8.1	

	Matched	0.24	0.21	0.03	5.9	
=4	Unmatched	0.13	0.15	-0.02	-7.1	
	Matched	0.11	0.11	0.00	0.0	

*** p<0.01, ** p<0.05, * p<0.1
~ if variance ratio outside [0.71; 1.41] for unmatched and [0.64; 1.56] for matched

Unmatched: Ps R^2= 0.071, LR test (prob)= 35.09 (0.001)***, MeanBias= 19.3, MedBias= 19.8, B= 64.0, R= 1.42
Matched: Ps R^2= 0.032, LR test (prob)= 7.16 (0.894), MeanBias= 8.8, MedBias= 7.5, B= 42.6, R= 0.73

First of all, the standardised bias tests indicate that for the household head the covariates of gender, age and age squared, and the first, second and third obtained level of education have a bias reduction of more than 5 percent. The associated bias reduction ranges from 7.5 to 22.9. For the spouse of the household head the covariates of the first obtained level of education has a bias reduction which still is too large after matching; all other reductions are below 5 percent. More balance tests need to be done in order to say something about achieved balance or not.

The next balancing test I turn to is two-sample t-tests. Table C.9 shows that for all covariates we fail to reject the null hypothesis that the mean of the covariates is different for treatment and control group. However, for ten of the fifteen covariates this was the case before matching as well (at the 5 percent level). For the other five covariates, two-sample t-tests provide support for the argument that the matching exercise has helped to reduce bias associated with observable characteristics.

Furthermore, the variance ratio of these covariates is near to one. However, in this case this test is less strong as the variance ratio was also near to one before matching.

Lastly we turn to some summary statistics. Before matching the mean and median bias were estimated to be 19.3 and 19.8 respectively, whereas they decreased to 8.8 and 7.5 after matching. This suggest reduction of bias due to matching. Furthermore after matching the pseudo R^2 equals 0.032, which is quite close to zero. This is an argument in favour of balance across samples (p= 0.894). Rubin's B is not below 25 percent after matching, while Rubin's R is between 0.5 and 2. This latter statistic provides support for balanced groups, whereas the former does not.

Multiple tests are not too convincing in arguing in favour of balanced covariates in both treatment and control group. However, the main analysis of this thesis focuses on the results produced using radius caliper matching. The nearest neighbour matching procedure is used only in order to be able to implement Rosenbaum bounds.

Appendix D

Table D.1: Tukey test results for pairwise comparisons of means (characteristics of different household types)

	Urban vs Rural			RU Migrants vs Rural			RU Migrants vs Urban		
Characteristics	**Difference**	**t**	**P>|t|**	**Difference**	**t**	**P>|t|**	**Difference**	**t**	**P>|t|**
Ethnic minority (%)	-13.17 (0.93)	-14.16	0.000	-20.64 (1.70)	-12.13	0.000	-7.47 (1.82)	-4.10	0.000
Age of household head	1.22 (0.35)	3.51	0.001	0.89 (0.64)	1.40	0.340	-0.33 (0.68)	-0.48	0.881
Age of spouse	1.53 (0.35)	4.38	0.000	2.02 (0.67)	3.01	0.007	0.48 (0.71)	0.68	0.778
Male household head (%)	-13.92 (1.06)	-13.10	0.000	-17.86 (1.94)	-9.18	0.000	-3.94 (2.08)	-1.89	0.141
Married household head (%)	-3.08 (0.97)	-3.17	0.004	-7.98 (1.78)	-4.49	0.000	-4.90 (1.90)	-2.57	0.027
Education of household head (highest obtained level)	1.16 (0.05)	22.42	0.000	1.89 (0.09)	20.07	0.000	0.73 (0.10)	7.23	0.000
Education of spouse (highest obtained level)	1.31 (0.06)	22.03	0.000	1.70 (0.11)	15.19	0.000	3.89 (0.12)	3.26	0.003
Registration status in resident area* (%)	-2.71 (0.41)	-6.53	0.000	-19.21 (0.76)	-25.31	0.000	-16.50 (0.81)	-20.29	0.000
Household size	-0.06 (0.04)	-1.64	0.229	-0.29 (0.07)	-4.09	0.000	-0.22 (0.07)	-2.98	0.008
Number of children	-0.13 (0.03)	-5.22	0.000	-0.26 (0.05)	-5.68	0.000	-0.13 (0.05)	-2.63	0.023
Classified as 'poor' in 2014 (%)	-9.66 (0.75)	-12.96	0.000	-11.65 (1.36)	-8.54	0.000	-1.99 (1.46)	-1.36	0.362
Insufficient food* (%)	-1.19 (0.34)	-3.51	0.001	-1.92 (0.62)	-3.08	0.006	-0.73 (0.67)	-1.09	0.520
Insufficient foodstuff* (%)	-4.23 (0.55)	-7.79	0.000	-4.98 (1.01)	-4.94	0.000	-0.67 (1.08)	-0.64	0.799
Savings at bank (%)	23.11 (0.95)	24.44	0.000	39.99 (1.73)	23.11	0.000	16.88 (1.85)	9.11	0.000
Residential area (m2)	15.44 (1.24)	12.44	0.000	12.50 (2.27)	5.50	0.000	-2.94 (2.43)	-1.21	0.449
Value of accommodation (1000 VND)	792488.10 (26667.00)	29.72	0.000	105143.00 (48803.61)	21.54	0.000	258943.10 (52294.76)	4.95	0.000
Total food consumption expenditure, regular (1000 VND)**	2044.65 (64.17)	31.86	0.000	2566.41 (117.44)	21.85	0.000	521.76 (125.85)	4.15	0.027
Total food consumption expenditure, festive (1000 VND)**	842.75 (52.14)	16.14	0.000	1104.76 (95.48)	11.57	0.000	6305.59 (112.95)	2.57	0.000
Total calorie intake, regular (kcal)**	210.09 (2870.22)	0.07	0.997	-3289.97 (5253.34)	-0.63	0.806	-3500.07 (5629.37)	-0.62	0.808
Total revenue (1000 VND)**	58097.24	5.33	0.000	67005.22	3.36	0.002	8907.98	0.42	0.909

	(10891.79)			(19935.12)			(21362.05)		
Total income (1000 VND)**	52792.01 (2554.91)	20.66	0.000	77557.88 (4647.23)	16.59	0.000	24765.87 (5010.94)	4.94	0.000

*Notes: Standard errors in parentheses. The sample size for 'Rural' equals 6.415, 2.157 for 'Urban' (excl. RU Migrants), and 521 for 'RU Migrants'. *The percentage of households having registration status in the resident area (in commune or ward) is based on the registration status of the head of the household. **Estimated over the past 30 days.*
Source: Authors calculations based on the VHLSS of 2014.

Table D.2: Tukey test results for pairwise comparisons of means (food expenditure and calorie intake shares of different migrant household types)

		H&S born in same province vs H&S born in difference province			No spouse vs H&S born in same province			No spouse vs H&S born in different province		
		Difference	t	P>|t|	Difference	t	P>|t|	Difference	t	P>|t|
Rice	**Food expenditure share**	-1.26 (1.48)	-0.85	-0.830	-1.29 (1.55)	-0.83	0.838	-0.03 (1.58)	-0.02	1.000
	Calorie intake share	-1.35 (1.84)	-0.73	0.883	-6.55 (1.92)	-3.41	0.004	-5.20 (1.97)	-2.65	0.041
Other staples	**Food expenditure share**	-0.15 (0.26)	-0.58	0.939	0.03 (0.27)	0.10	1.000	0.18 (0.28)	0.64	0.919
	Calorie intake share	-0.14 (0.57)	-0.25	0.995	-0.04 (0.59)	-0.06	1.000	0.11 (0.60	0.17	0.998
Meat	**Food expenditure share**	-0.30 (0.92)	-0.33	0.988	-0.81 (0.96)	-0.85	0.830	-0.51 (0.98)	-0.52	0.953
	Calorie intake share	0.21 (0.46)	0.45	0.969	-0.53 (0.48)	-1.10	0.690	-0.74 (0.49)	-1.50	0.438
Fish and seafood	**Food expenditure share**	-0.62 (0.99)	-0.62	0.924	-0.58 (1.03)	-0.56	0.944	0.04 (1.05)	0.04	1.000
	Calorie intake share	-0.09 (0.20)	-0.48	0.963	-0.13 (0.20)	-0.63	0.922	-0.03 (0.21)	-0.16	0.998
Eggs	**Food expenditure share**	-0.07 (0.11)	-0.63	0.921	0.14 (0.12)	1.16	0.652	0.21 (0.12)	1.73	0.307
	Calorie intake share	-0.00 (0.04)	-0.04)	1.000	0.08 (0.04)	1.93	0.214	0.08 (0.04)	1.93	0.215
Tofu	**Food expenditure share**	-0.03 (0.11)	-0.25	0.994	0.12 (0.11)	1.09	0.697	0.15 (0.11)	1.30	0.562
	Calorie intake share	-0.04 (0.09)	-0.41	0.977	0.03 (0.09)	0.29	0.991	0.07 (0.09)	0.67	0.908
Vegetables	**Food expenditure share**	0.26 (0.29)	0.92	0.795	0.24 (0.30)	0.80	0.854	-0.02 (0.31)	-0.08	1.000
	Calorie intake share	0.00 (0.00)	0.00	1.000	0.52 (0.20)	2.57	0.050	0.52 (0.21)	2.51	0.058
Fruits	**Food expenditure share**	0.22 (0.24)	0.93	0.787	0.25 (0.25)	0.99	0.754	0.02 (0.26)	0.10	1.000
	Calorie intake share	0.23 (0.33)	0.70	0.897	0.16 (0.35)	0.47	0.967	-0.07 (0.35)	-0.20	0.997
Dairy	**Food expenditure share**	2.05 (0.83)	2.46	0.066	-1.26 (0.87)	-1.45	0.470	-3.31 (0.89)	-3.73	0.001

	Calorie intake share	0.41 (0.20)	2.09	0.156	-0.24 (0.21)	-1.17	0.649	-0.66 (0.21)	-3.10	0.010
Food away from home	**Food expenditure share**	0.87 2.17)	0.40	0.978	3.24 (2.26)	1.44	0.476	2.37 (2.31)	1.03	0.732
	Calorie intake share	1.74 (2.00)	0.87	0.820	6.37 (2.08)	3.06	0.012	4.64 (2.13)	2.18	0.129
Other	**Food expenditure share**	-0.99 (0.57)	-1.72	0.313	-0.08 (0.60)	-0.14	0.999	0.90 (0.61)	1.48	0.449
	Calorie intake share	-0.96 (0.64)	-1.50	0.440	0.33 (0.67)	0.49	0.962	1.29 (0.59)	1.88	0.237

Notes: Standard errors in parentheses. The sample size for households for which the head and spouse are born in the same province equals 194, 176 for households for which the head and spouse are born in a different province, and 151 for households which have no spouse.
Source: Author's calculations based on the VHLSS of 2014.

Appendix E

Table E.1: Price of 1000 kcal in 2014

Food Code	Food item	Price of 1000 kcal (1000 VND)
101	Plain rice, including fragrant and speciality rice	1.10
102	Sticky rice	1.88
103	Maize	1.33
104	Cassava	1.98
105	Potato of various kinds	3.73
106	Wheat grains, bread, wheat powder	3.45
107	Flour noodle, instant rice noodle/porridge	3.69
108	Fresh rice noodle, dried rice noodle	3.02
109	Vermicelli	10.03
110	Pork	9.79
111	Beef	47.69
112	Buffalo meat	52.71
113	Chicken meat	15.19
114	Duck and other poultry meat	7.10
115	Other types of meat (goats, dogs, sheep, wild animals, birds, …)	12.50
116	Processed meat (boiled pork pies, fried pork pies, roasted pork, sausages, …)	12.36
117	Lard cooking oil	1.51
118	Fresh shrimps, fish	21.80
119	Dried and processed shrimps, fish	9.67
120	Other aquatic products and seafood (crabs, snails, …)	22.35
121	Eggs of chickens, ducks, muscovy ducks, geese	12.16
122	Tofu	6.01
123	Peanuts, sesame	2.41
124	Beans of various kinds	15.00
125	Fresh peas of various kinds	7.04
126	Morning glory vegetables	9.18
127	Kohlrabi	8.41
128	Cabbage	10.11
129	Tomato	18.25
130	Other vegetables (gourd, winter melon, cucumber, cabbage, squash, …)	9.66
131	Orange	23.61
132	Banana	4.01

133	Mango	9.99
134	Other fruits (rambutan melon, papaya, guava, litchi, grapes, …)	8.08
135	Fish sauce	14.37
139	Sugar, molasses	1.45
140	Confectionery	5.19
141	Condensed milk, milk powder	19.50
142	Ice cream, yoghurt	18.64
143	Fresh milk	18.55
144	Alcohol of various kinds	13.98
145	Beer of various kinds	69.75
147	Instant coffee	10.64
148	Coffee powder	9.42
153	Outdoor meals and drinks (breakfast, lunch, dinner)	3.56
154	Other food and drinks (other processed food and foodstuff, additives, seasonings, …)	3.39

Notes: Mentioned prices of 1000 kcal are the averages for the whole population. In practice each household faces its own household price, based on e.g. region, taste, quality of the product, and availability of products. However, these average prices can function as a reference price.
Source: Table is based on data from Huong Thi Trinh, PhD candidate at the Toulouse School of Economics (TSE) (Trinh, 2017).

Appendix F

In a paper on selection on observed and unonbserved variables, Altonji et al. (2005) propose a way to use observables to draw inferences about selection bias. They argue that the degree of selection on the observables is informative about selection on unobserved characteristics. In this way, using Ordinary Least Squares (OLS) estimators, the degree of selection bias can be assessed (Altonji et al., 2005). Bellows and Miguel (2009) follow the intuition proposed by Altonji et al. (2005) when deriving the ratio of the 'influence' of omitted variables relative to the observed control variables that would be needed to fully explain away the treatment effect. The main idea is that the relative importance of omitted variable bias is assessed by analysing how the coeffcients of interest change with the inclusion of additional explanatory variables (Bellows and Miguel, 2009). In the case of this thesis this would imply that if the inclusion of covariates substantially attenuates the coefficient estimates of rural-urban migration, than it is possible that including more control variables (i.e. unobserved characteristics) would reduce the estimated effect even further. If, on the contrary, the inclusion of covariates has no effect on the magnitude of coefficient estimates, then it might be suggested that including even more covariates (i.e. unobserved characteristics) does not alter the magnitude of the coefficient estimates of migration as well. Bellows and Miguel (2009) thus propose an intuitive method of analysing coefficient stability for assessing the importance of omitted variable bias due to selection bias.

In the appendix of their article, Bellows and Miguel (2009) derive a measure of how strong the covariance between the unobserved part of selection and migration must be, relative to the covariance between the observed part of selection and migration, to explain away the entire effect. The measure can easily be calculated using the coefficients from two OLS regressions; one with control variables and one without. I follow the notation of Bellows and Miguel (2009) when defining the ratio as:

$$\hat{\alpha}OLS,C\ /\ (\hat{\alpha}OLS,NC - \ \hat{\alpha}OLS,C) = Cov(c,\tilde{q})\ /\ Cov(c,x'\beta), \tag{F.1}$$

where $\hat{\alpha}OLS$ represent the coefficient estimates of rural-urban migration in the OLS model; C are the estimates of the model with controls, and NC for the estimates of the model with no controls. Furthermore, c denotes rural-urban migration, $\tilde{q}$ the unobserved part of selection into migration, and $x'\beta$ the observed part of selection into migration. For more detailed information on the derivation of formula F.1 I refer the reader to Bellows and Miguel (2009). It is argued that if the set of observed controls is representative for all possible control variables, then a large ratio suggests that it is implausible that the unobservables could explain away the entire treatment effect (Bellows and Miguel, 2009).

The ratio defined in equation F.1 is calculated for all estimated OLS models, using the coefficients of the OLS models with and without covariates (see table 5.8 – 5.11). It should be stressed that for all OLS models the inclusion of covariates raises R^2, although the variance in dependent variable explained by the variance in independent variables remains rather small even after this inclusion of covariates. The ratios for the estimates of rural-urban migration on food expenditure and calorie intake shares of migrants are presented in table F.1 below. Recall that the OLS model produced significant effects at at least the 5 percent level for both the food expenditure and calorie intake shares of rice (-), meat (-), eggs (-), tofu (-), dairy (+) and FAFH (+). For these significant food expenditure shares the coefficient stability ratio in table F.1 ranges from 2.79 for rice to -52.50 for other. This implies that for rice the unobservables would have to have a 2.79 times higher effect on migration than the observables in order to invalidate the significant negative effect that was found. For the other significant food expenditure shares this ratio is even higher, indicating that more even selection bias is 'needed' in order to entirely explain away the significant effects on these expenditure shares. For the significant calorie intake shares the coefficient stability ratio ranges from -1.32 for rice to -5.83 for tofu. The ratios are considerably smaller for calorie intake shares than for food expenditure shares, although for eggs, tofu, dairy and food away from home I argue, following the rationale behind the coefficient stability ratios, that still a large influence of unobservables is needed in order to explain away the effect on the corresponding calorie intake shares.

Table F.1: Coefficient stability ratio of rural-urban migration on food expenditure and calorie intake shares

Share	Food expenditure shares	Calorie intake shares
Rice	2.79	-1.32
Other staples	1.96	-0.38
Meat	-19.67	-1.90
Fish and seafood	1.24	4.50
Eggs	12.25	-2.67
Tofu	35.50	-5.83
Vegetables	-1.65	-2.00
Fruits	0.11	2.92
Dairy	3.76	2.79
Food away from home	7.60	4.21
Other	-52.50	2.29

Notes: Ratios are based on the coefficient estimates for rural-urban migration, which can be found in table 5.8. N= 1.539 in the case without covariates (including 194 rural-urban migrant households); N= 1.182 in the case with covariates (including 141 rural-urban migrant households).

The OLS model produced positive and significant estimates for the effect of rural-urban migration on HDDS and FVS. A negative significant effect was found for the indicator of whether or not a household meets the fruit and vegetable recommendations of the World Health Organization (WHO) (see table 5.9). The corresponding coefficient stability ratios can be found in table F.2 below. The ratios range from 2.05 to 5.30, indicating that for FV the unobservables should at least have a 2.05 times higher effect on migration than the observables in order to invalidate the significant negative effect that was found. For the other two indicators the corresponding coefficient stability ratio is higher, indicating that even more omitted variable bias is 'needed' in order to explain away the effect of migration on HDDS and FVS.

F.2: Coefficient stability ratio of rural-urban migration on HDDS, FVS, and FV

Indicator	
HDDS	5.30
FVS	4.29
FV	2.05

Notes: Ratios are based on the coefficient estimates for rural-urban migration, which can be found in table 5.9. N= 1.539 in the case without covariates (including 194 rural-urban migrant households); N= 1.182 in the case with covariates (including 141 rural-urban migrant households).

For the effect of having rural-origins on food expenditure and calorie intake shares the OLS models found no significant effects for migrant households living in the South Eastern Area (SEA) of Viet Nam. Significant effects were found for migrant households living in the Red River Delta (RRD). The OLS model predicted significant effects at at least the 5 percent level for the expenditure share of rice (-), other staples (-), meat (-), eggs (-), tofu (-), dairy (+), FAFH (+), and other (-), and for the calorie intake share of of rice (-), dairy (+), and FAFH (+). De coefficient stability ratios of food expenditure shares range from 2.78 for tofu to 182.34 for other staples, indicating that the unobservables should at least have a 2.78 times higher effect on migration than the observables in order to explain away the entire effect of having rural-origins on the expenditure share of tofu. For the other categories the ratio is even higher. For calorie intake shares the ratio ranges from 1.70 for dairy to 8.70 for food away from home. Again, the ratios for calorie intake shares are much lower than the ratios for food expenditure shares.

Table F.3: Coefficient stability ratio of having rural origins on food expenditure and calorie intake shares

	RRD		SEA	
Share	**Food expenditure shares**	**Calorie intake shares**	**Food expenditure shares**	**Calorie intake shares**
Rice	4.01	2.51	0.51	-0.24
Other staples	182.34	-1.30	-2.50	-0.24
Meat	-13.55	-1.86	-0.08	-0.47
Fish and seafood	-3.68	5.80	-0.77	-1.33
Eggs	-6.75	-0.71	1.41	-5.00
Tofu	2.78	-42.15	1.33	-8.00
Vegetables	1.09	-2.08	-1.22	2.00
Fruits	-0.95	-1.59	1.19	1.80
Dairy	5.24	1.70	7.86	-1.12
Food away from home	-29.38	8.70	0.29	-.26
Other	33.25	-2.31	-2.00	-1.05

Notes: Ratios are based on the coefficient estimates for rural-urban migration, which can be found in table 5.10. For the subsample of the RRD: N= 564 in the case without covariates (including 66 rural-urban migrant households); N= 436 in the case with covariates (including 46 rural-urban migrant households). For the subsample of the SEA: N= 565 in the case without covariates (including 203 rural-urban migrant households); N= 385 in the case with covariates (including 134 rural-urban migrant households).

The OLS model in table 5.11 shows significant positive effects of having rural origins on HDDS and FVS for migrant households living in the RRD. For the SEA no significant effects were found. The corresponding coefficient stability ratios, which are presented in table F.4 below, are 5.84 for HDDS and 4.83 for FVS. These ratios show that the unobservables should have at least a 4.82 times higher effect on migration than the observables in order to explain away the entire effect of having rural origins on FVS. For explaining away the effect on HDDS the ratio is even higher.

Table F.4: Coefficient stability ratio of having rural origins on HDDS, FVS, and FV

Indicator	RRD	SEA
HDDS	5.84	0.38
FVS	4.82	-0.52
FV	-1.12	2.08

Notes: Ratios are based on the coefficient estimates for rural-urban migration, which can be found in table 5.11. For the subsample of the RRD: N= 564 in the case without covariates (including 66 rural-urban migrant households); N= 436 in the case with covariates (including 46 rural-urban migrant households). For the subsample of the SEA: N= 565 in the case without covariates (including 203 rural-urban migrant households); N= 385 in the case with covariates (including 134 rural-urban migrant households).

The coefficient stability ratios should guide as an intuitive assessment for assessing the importance of selection bias. However, it should be kept in mind that one can never, by definition, know how much the unobservables are influencing the coefficient estimates of interest. I have reported the coefficient stability ratios in order to give some intuitive insights about the possible degree of selection bias, but one should be cautious in making exact statements.

Furthermore, Altonji et al. (2005) argue that it might be dangerous to infer too much about selection on the unobservables from selection on the observables if only a limited number of observables are included in the model, if the observables are only small in explanatory power, or if the observables are unlikely to be representative of the full range of factors that determine food consumption patterns. I argue that all these possible dangers apply to the case of this thesis, given that only a small number of covariates could be included as they should be unaffected by migration, and given that even with inclusion of covariates R-squared does not exceed 0.1655 at maximum. Therefore I argue that the coefficient stability ratios should be interpreted with caution.

Concluding, Oster (2016) builds on the idea of coefficient stability ratios, but argues that R-squared needs to be included into the calculation and comparison as well. She argues that coefficient stability on its own is at best uninformative and at worst very misleading, and that it must be combined with movements in R-squared in order to develop an argument. She developed the command **psacalc** to implement the upgraded coefficient stability calculations in STATA. Future research could explore the use of this command.