NUTRITION-RELATED OUTCOMES AND FOOD ENVIRONMENTS
IN AN INCREASINGLY PROCESSED GLOBAL FOOD LANDSCAPE

By

Rahul Dhar

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Agricultural, Food and Resource Economics – Doctor of Philosophy
Economics – Dual Major

2023

ABSTRACT

Globalization and modern food production, processing, and marketing techniques have increased access to food in many regions in the world, reducing food insecurity. However, there are some serious consequences that have, until recently, been largely ignored in many developing countries. Some methods of food processing have produced foods with high caloric density, but minimal nutrient content. This has manifested as decreasing rates of various forms of undernutrition yet increasing overweight and obesity. Policy and program efforts by governments and NGOs try to find simple solutions, but are often created with only partial understanding of these complex issues.

It may be too difficult due to resource constraints to acquire full information about various nutrition-related problems. One must possess knowledge of consumer behavior, the impacts various choices might have on health, and how external drivers influence the consumer and their food environment. Obtaining this information is further complicated by the rapid transitions in developing countries. Diets are becoming increasingly processed, and daily habits are trending towards more sedentary activities. At the same time, incomes are rising, food insecurity is falling, and poverty is declining in general. There is much need to understand these complex issues to better assist designers of policy and programs to provide the greatest benefit while mitigating negative unintended consequences.

This dissertation examines three nutrition- and nutrition-related issues. A variety of econometric techniques are applied to secondary data to explore processed food consumption, consumer shopping behavior, and the relationship between nutrition outcomes and processed food imports. This work contributes to the literature by providing a better understanding of the impact of processed food consumption on overweight or obesity, mechanisms that might drive rapid

obesity increases, and how the food purchasing behavior of low-income households impacts the cost of their diet.

In the first paper, we examine the complex relationship between activity, diet, and the likelihood of adults being overweight or obese using a three-round nationally representative panel data set from Tanzania and a correlated random effects probit model to account for unobserved heterogeneity. Our results suggest large differences in the effects of diet and activity choice between rural and urban areas.

In the second paper, we explore the existence of a two-way linkage between trade and health that has long been discussed, but never investigated empirically. Processed food imports are usually accompanied by robust marketing campaigns, which suggests there might be induced socio-cultural changes that may drive increased processed food importation. The socio-cultural changes may not be directly observable but should manifest in increased rates of overweight or obesity. Our evidence suggests a two-way relationship might exist, which is a serious concern for developing countries where exposure to ultra-processed food to date is below that of developed nations but is rising rapidly.

In the third paper, we examine whether shopping behavior impacts food prices households pay. We develop a model of prices consumers face that accounts for two dimensions of food procurement: spatiality of shopping behavior and frequency of shopping trips. We explore this issue using a recent survey conducted by MSU in low-income areas of Nairobi, Kenya. Our results suggest two prominent points. First, households seem to perform very well controlling the cost of their food basket given their likely time, budget, and information constraints. Second, while we find substantial spatial price differences, we do not find any significant price benefit to those that shop outside of their local food environment.

with zero economic experience and no research to my name, which led to me starting an amazing family and [hopefully] earning a PhD. Dr. Swinton, I thank you for your support when the flow of life started inhibiting my ability to complete the PhD within reasonable time.

I would be remiss if I did not mention the numerous graduate students that not only helped me along the way, but also made studying for prelims less than terrible. These include Danielle Ufer, Salim Nuhu, Abubakr Ayesh, Charuta Parkhi, Hyungjung Kim, Ramyani Mukherjee, Christine Sauer, and Aakanksha Melkani. Thank you for listening to my rants, helping me understand new concepts, studying for the prelims, collaborating to finish homework sets, and all the great talks we had over the years. Thanks to my Macro study group, Josh Browstein, and Mehmet Karaca. I do not think I would have passed the Macro prelim without your help.

My final thank you is to my family. Thanks to my loving wife that endured the stress and anxiety that came with exams and presentations. Thanks to my kids for always being my guiding light, without whom I would not have the drive nor the purpose to persevere.

TABLE OF CONTENTS

CHAPTER 1: Diet or Activity? An Examination of Adult Overweight and Obesity in Tanzania using Panel Data

1.1   Introduction

Africa's food system transformation is driving a rapid change in the kinds of nutritional challenges it faces. A key feature of food system transformation across the world is a rise in processed food production and consumption (Pingali, 2007; Reardon et al., 2021; Tschirley et al., 2015). The modernization of the agrifood systems may reduce labor burdens, increase nutritional access, and increase food safety[1] (Qaim, 2017; Tschirley, Snyder, and Kondo, 2017, Reardon et al., 2010), however, it is also related to the rapid rise of ultra-processed foods and beverages (UPFs)[2] in the global food economy (Popkin, Adair, and Ng, 2012). A plethora of studies have established a strong association between consumption of UPFs and levels of overweight and obesity and related non-communicable diseases (NCDs; Pagliai et al., 2021; Elizabeth et al., 2020; Askari et al., 2020).

It is widely presumed that UPF consumption is a major driver of rising overweight and obesity, yet the empirical literature examining differential impacts between both UPFs and activity on body weight is extremely thin[3]. Poti et. al. (2017) note there is "fairly consistent support" for the link between UPF and weight outcomes, however, it is not definitive and there is a need to

---

[1] Mycotoxin contamination is prevalent in many of the usual staple grains consumed in developing countries (Bullerman, 1979; Marrez and Ayesh, 2021) and the food processing associated with the modernization of the agrifood systems can mitigate or eliminate mycotoxins (Adebo et al., 2021; Ademola et al., 2021).

[2] UPFs are defined as multi-ingredient mixtures formulated by manufacturers (Monteiro et al., 2018)

[3] Even in the United States, there are still significant limitations to our understanding of obesity and how to reduce it at a population level. Brown et al. (2019) highlight major gaps in the current pool of knowledge and call for research to be done in different contexts with varying methods to better understand drivers of food insecurity and obesity.

"confirm these findings in different population locations." In addition, studies that examine the impact on weight outcomes may include detailed measures of either diet or activity while including the other in less detail (see Hall et al., 2019; Kolodinsky et al., 2017). Using less detailed measures may exclude variation needed to adequately estimate the relationships between diet, activity, and body mass.

Moreover, there is less literature that compares the effects diet and activity have on weight outcomes in Africa, even though UPFs are now prevalent in African cities and towns and even rural areas (Reardon et al., 2021; Tschirley et al., 2015). We know of only one article that directly evaluates the influence of processed food consumption (through its association with dietary patterns) with overweight and obesity in Africa (Sarfo et al., 2021), and it does so only for women in rural areas, does not control for physical activity, and does not employ panel estimation methods in its analysis. Others separately examine the impact of shopping in supermarkets on weight outcomes and on "highly processed" food consumption (Debela et al., 2020; Demmler et al., 2018, Khonje and Qaim, 2019; Khonje et al., 2020) but do not directly show the association between weight outcomes and consumption of these foods.

This paper seeks to fill this gap in the literature by assessing the differential impacts of diet and activity on overweight and obesity, using a household level panel data set from Tanzania, the Tanzania National Panel Survey. By applying panel estimation methods and addressing common econometric problems such as unobserved heterogeneity ("missing variables") and endogeneity, we generate robust estimates of the impact of level of processing of home-consumed foods (low and high[4]), of meals away from home (MAFH), and of physical activity on the likelihood of being

---

[4] UPFs are included in the high-processed category.

overweight or obese. We do this separately for rural and urban areas in addition to the full sample. Importantly, we control for macronutrient intake (carbohydrates, protein, and fat), which strengthens our interpretation of results for processed foods.

The paper proceeds as follows. In the next section we describe our methods, which includes a description of our model and the econometric approaches used. Next, we detail and explore the data, followed by a discussion of our results. We conclude the paper with a brief summary of our findings.

1.2 Methods

1.2.1 Diet and activity choices and body weight outcomes

Our analytical approach is informed by the literature around energy balance and body weight outcomes, and food and activity environments. Energy imbalance, where calories consumed exceed calories expended, can lead to undesirable weight gain (Swinburn et al., 2011; Kolodinsky et al., 2017). This balance is the result of dietary and activity choices, which are driven by the interaction of personal characteristics and the food and activity environments that consumers are exposed to.

The composition of chosen foods may matter in addition to the amount consumed in determining energy balance (Lustig 2006; Wells and Siervo, 2011, Ludwig et al., 2021). Per Ludwig et al. (2021), "rapidly digestible carbohydrates … cause increased fat deposition, and thereby drive a positive energy balance." Many processed foods, and in particular UPFs such as sugar-sweetened beverages, snack foods, and fast-food meals, are high in these rapidly digestible

carbohydrates, suggesting a potential metabolic role of UPFs in promoting weight gain. Hall et al. (2019) show increased energy intake as a result of a diet high in UPFs in a controlled, in-lab dietary experiment. UPFs may thus doubly affect weight gain by promoting over-consumption and through metabolic mechanisms.

Food environments are characterized by the availability and prices of food on offer, by the characteristics of the products being sold (e.g., convenience, appeal, quality, safety) and the vendors selling them (e.g., small or large, formal or informal, cleanliness, range of offerings), and by the food and lifestyle messaging prevalent in the areas where consumers buy their food (Fanzo et al., 2020). Because many of these factors vary over space (e.g., richer and poor neighborhoods; dense central cities or outlying neighborhoods or rural areas), individuals operating in the same food system can be exposed to very different food environments.

Individuals' response to their food environment depends on their own characteristics (Fanzo et al., 2020). These include their income (which together with prices drives purchasing power), the food-related information and knowledge they bring to their shopping, values and lifestyle aspirations linked (and promoted by advertising) to food, and the circumstances of their life such as the kind of work they do (more or less sedentary), how they commute to work or school (walking; public transport; private transport), the kind of food environments they come into contact with during that commute or during their working or school hours, the time they have to make food decisions, and their health status as it affects needed diet and mobility and thus ability to choose their own food. Because these factors vary over individuals, different individuals shopping in the same food environment can make very different dietary choices.

The activity environment relates to the factors that drive how much physical activity a person can or must engage in (Casey et al., 2008): the availability and cost of motorized transport,

including public transport; the residential settlement pattern and its proximity to places of work and school; the labor intensity of jobs on offer; and the availability and cost of labor-saving machines for domestic work. Personal factors also mediate how this environment shapes a person's activity choices: income and wealth, knowledge about exercise and health, concepts of healthfulness and desirable body types, and the circumstances of one's life drive differing activity choices for people exposed to similar activity environments.

1.2.2 Binary Choice Model

We wish to understand the factors that drive the likelihood of an individual being overweight or obese. In particular, we are interested in the role that consumption of different types of processed foods may play in this outcome. We employ a variation of the probit model that allows use of an unbalanced panel and accounts for unobserved heterogeneity. First, consider a pooled probit with an index function for the latent variable given by,

$$y_{it}^* = \Theta' O_{it} + \Delta' D_{it} + \Gamma' A_{it} + \Lambda' H_{it} + \Upsilon' W_{it} + \varepsilon_{it} \qquad 1$$

$$y_{it} = \begin{cases} 1, & if\ y_{it}^* > 0 \\ 0, & if\ y_{it}^* \leq 0 \end{cases}$$

Where y_{it} equals 1 if individual i is overweight or obese (BMI>=25) in period t. O_{it} is a (1 x 3) vector of obesogenic environment variables that includes the share of the household's sample cluster that is overweight or obese, an indicator if the household is in a rural locality, and the distance to nearest population center. D_{it} is a (1 x 5) vector of all the diet variables that include

shares of high and ultra-processed foods[5] (HPF) and meals purchased away from home (MAFH) in total food consumption, in addition to the average daily grams of protein, carbohydrates, and fat per adult equivalents consumed. The (1 x 4) vector A_{it} contains activity proxies, which includes hours of high energy activity together with indicators for whether a person owns a motor vehicle, engaged in some sort of agricultural activity in the past 12 months, or works in a sector associated with office work. H_{it} is a (1 x 15) vector of household and individual level socioeconomic and welfare variables that, in addition to the variables outlined in the data section below, includes squared terms for age and log expenditure, and a constant. W_{it} contains survey period dummy variables and region fixed effects. ε_{it} is a normally distributed random error.

1.2.3   Unobserved Heterogeneity

In any analysis based on survey data, many variables are unobserved, such as innate ability. We may use observable proxies for these variables, but the true effect may not be fully captured. This implies there may be unobserved factors that vary across our sample driving consumption and labor decisions. So, we account for unobserved heterogeneity with the correlated random effects probit model using Mundlak's adaptation of Chamberlain's (1980) random effects specification (Mundlak, 1978). The unobserved heterogenous effects, c_i, are modelled as depending on the time averages of the covariates for each individual in the CRE. We estimate a pooled probit using the modified index function,

---

[5] While this category contains UPF, we will only refer to this category as high-processed foods or HPF for the remainder of the paper. Examples of food items in this category include breads, buns, cakes, and biscuits, sugar, sweets, soda, sausage, and dried/canned/salted fish and seafood.

$$y_{it}^* = \Theta'O_{it} + \Delta'D_{it} + \Gamma'A_{it} + \Lambda'H_{it} + \Upsilon'W_{it} + c_i + \varepsilon_{it} \qquad 2$$

where,

$$c_i = \psi + \bar{O}_i + \bar{D}_i + \bar{A}_i + \bar{H}_i + v_i \qquad 3$$

and,

$$\bar{O}_i = \frac{1}{T}\sum_{t=1}^{T} O_{it} \ , \bar{D}_i = \frac{1}{T}\sum_{t=1}^{T} D_{it} \ , \bar{A}_i = \frac{1}{T}\sum_{t=1}^{T} A_{it} \ , \bar{H}_i = \frac{1}{T}\sum_{t=1}^{T} H_{it} \qquad 4$$

Any variables that do not vary over time such as gender are excluded from the averages. Since we have an unbalanced panel, we adopt the method developed by Wooldridge (2019). We assume the data to be missing at random and condition on selection in addition to our covariates. To allow observations that are only observed in 1 or 2 periods, we modify our unobserved heterogeneity in the following way:

$$E(c_i|S_i, S_i'X_i) = \sum_{r=1}^{T} \psi_r \, 1[T_i = r] + \sum_{r=1}^{T} (\bar{O}_i + \bar{D}_i + \bar{A}_i + \bar{X}_i) \, 1[T_i = r] \qquad 5$$

Where X_i is the full matrix of covariates for individual i, T_i is the number of periods an individual has a full set of data, and S_i are the associated selection variables for whether there is a full set of data in period r. Additionally, we assume conditional normality of the heterogeneity and we allow the variance to change in a similar manner:

$$Var(c_i|S_i, S_i'X_i) = \exp\left(\sum_{r=2}^{T} \psi_r \, 1[T_i = r] + \sum_{r=2}^{T} (\bar{O}_i + \bar{D}_i + \bar{A}_i + \bar{X}_i) \, 1[T_i = r]\right) \qquad 6$$

This formulation allows us to estimate our model as a heteroskedastic probit. We focus on the average marginal effects (AME) of the covariates on the probability of being overweight or obese, so the parameters that appear in the modelled heterogeneity are averaged out, allowing us to identify the AMEs of the covariates of interest. (Wooldridge, 2019).

We estimate two probit models. The first is a naïve probit with only the regressors and fixed effects. We incorporate correlated random effects and allow for multiplicative heteroskedasticity in the form of equation 6 in the second. Each model is applied to the full adult sample, and to rural and urban sub-samples. We estimate the probit with the STATA commands *glm* and *hetprobit*, and cluster standard errors at the household level. After each estimation the average marginal effects are calculated using the *margins* command, with standard errors calculated using the delta method.

To compare the differences between the marginal effects in different ranges of BMI, we use ordered probit[6] versions of both previously discussed probit models. The probability we estimate is given by:

$$Pr(y_j = i) = Pr\big(\kappa_{i-1} < B'X_{jt} + c_j + \varepsilon_{jt} \leq \kappa_i\big) \qquad 7$$

$$= \Phi\big(\kappa_i - B'X_{jt}\big) - \Phi\big(\kappa_{i-1} - B'X_{jt}\big) \qquad 8$$

---

[6] Wooldridge (2019) notes that using correlated random effects in an ordered probit model (with or without heteroskedasticity) can be implemented in the same manner as we might with the traditional probit.

Where $\kappa_1, \kappa_2, \ldots, \kappa_i$ are the cut points between underweight, normal, overweight, and obese weight categories and X_{it} contains all regressors included in the probit index equation previously mentioned. To account for unobserved heterogeneity, we model the logarithm of the variance as depending on a linear equation of the form $\ln(\sigma_j) = z_j\gamma$, which implies our probability for the heteroskedastic ordered probit becomes:

$$Pr(y_j = i) = \Phi\left(\frac{\kappa_i - B'X_{jt}}{e^{z_j\gamma}}\right) - \Phi\left(\frac{\kappa_{i-1} - B'X_{jt}}{e^{z_j\gamma}}\right) \qquad 9$$

1.2.4   Endogeneity

1.2.4.1   Instrument Choice

Our second method to explore the robustness of our results aims to show a more causal impact. We do not claim to fully identify the effect from processed food or activity level, but we use these results to show that controlling for some endogeneity does not change the pattern seen in the results. We focus on the potential endogeneity of our main diet and activity choice variables. Specifically, we only consider consumption shares and active hours endogenous.

To account for the endogeneity, we need instruments related to diet and activity choices, but not related to obesity other than through those choices. The instruments must also make sense in the context of our model, meaning that we need instruments that are associated with the food and activity environments. Since the NPS does not contain any variable closely suited to our needs, we construct instruments as follows. First, we sum the total value of food consumption for each processing category in an enumeration area and subtract out the individual's consumption in that

category. We do the same with active hours. For our consumption variables, we divide this number by the total value of all food consumption in the enumeration area. Keeping in mind that sample households shop extremely locally for their food[7], these estimates of the shares of all food consumption in different processing categories among nearby households give us a measure of the food environment for each individual outside of their own consumption. For active hours, we divide total active hours in an enumeration area by the population of that enumeration area, excluding the individual. This proxies for the activity environment as a measure of the average behavior of the individual's neighbors.

Seasonality is an important point not accounted for in our estimation. Households were surveyed during different months of the year. This is particularly important for agricultural and lower income households. Households are likely to have more disposable cash for purchasing food after harvesting their crop, than they might immediately following planting season. Similarly, someone may rely on agricultural labor outside of the household farm and may do this work only during certain seasons. Food available for purchase and off-farm labor opportunities may also vary by month in both urban and rural areas. So, depending on when the household was surveyed, we expect food and activity choices available to them to be different. We use indicators for quarter of the year the household was surveyed as an instrument for food and activity choices to account for this seasonality.

Food and activity environments are also shaped by the physical environment. For instance, geography determines climate, which will shape the type of crop grown in an area, which will

---

[7] For discussions of food shopping in Africa, see Neven et al. (2006), Mensah and Oyebode (2022), Holdsworth and Landais (2022), and Bannor et al. (2022).

impact food and type of work available. Location drives transaction costs which may impact the distribution of imports, potentially limiting the type of convenience foods available for purchase. To account for this variation in the food and activity environments, we include the percent of agricultural cover within one kilometer of the household and annual precipitation.

We argue that these variables should not impact overweight or obesity directly, which is generally not seasonal, and likely persistent. Even across survey periods there is not excessive variation in weight status. So, seasonality should only impact obesity through food and activity choices. It is also unlikely that our enumeration level variables impact obesity through channels other than food or activity choices. If there were an influence from factors outside of choices, those effects are likely absorbed by our enumeration area prevalence variable. And while it may be true that rainfall and land cover around a household may be related to income of the household, this works through consumption and labor decisions, and so should not impact obesity other than through our endogenous variables.

1.2.4.2 Joint Maximum Likelihood

We follow our conceptual model and allow food and activity to be chosen in the first step. Since the data exhibit some censoring at zero from only observing choices made in the 7 days prior to the survey for each respondent, we estimate the first stage with a Tobit for each of our endogenous variables. We use the same CRE methods as previously described. This gives us an estimating equation of the form:

$$y_{it}^* = B'X_{it} + c_i + z_{it} + \varepsilon_{it} \qquad\qquad 10$$

$$y_{it} = \max(0, y_{it}^*)$$

Where y_{it} is the observed value the endogenous variables (HPF and MAFH consumption shares or active hours). X_i contains all our exogenous variables, c_i is the same as previously defined, ε_{it} is a normally distributed random error, and z_i includes our instruments (J enumeration level variables, quarter of the year indicators, % agricultural cover in 1km radius, and annual precipitation):

$$z_{it} = \sum_{j=1}^{J} Share_{ijt} + \sum_{k=2}^{4} Quarter_{ikt} + AgCover_{it} + Precip_{it} \qquad\qquad 11$$

The second step is a probit (or ordered probit), as we have discussed previously. We estimate the system jointly in one step. This is done using the STATA command *cmp* (Roodman, 2007), which fits recursive mixed process models via full information maximum likelihood and allows the error terms between equations to have non-zero covariances.

To test the relevance of our instruments we employ some approximations based on different structural models, since our estimation method does not allow a convenient way to adequately detect weak instruments. However, we are still able to use the F-test from our first step Tobits, as well as an overall F-test for the excluded instruments in the joint maximum likelihood estimation. While this is not a traditional weak instrument test, we get comparable results if we were to estimate the model with different structural equations. The costs of losing our ability to use a traditional test is offset by the gains in accounting for issues such as censoring.

In Panel a of Table 1, we calculate F test statistics from Wald tests of the excluded instruments in each first stage equation. We do this for both linear and Tobit equations. Stock and Yogo (2005) argue for a general rule of thumb of an F statistic above 10, which differs when using large numbers of instruments and multiple endogenous regressors. Using table 1 of their paper, for 10 instruments and 1 endogenous variable, the critical value for 2SLS bias under 5% is an F statistic of 20.74. In the linear case the F statistics are all above 21.81.

The F statistics for each first stage may not adequately detect weak instruments in a model with multiple endogenous explanatory variables, so we employ three tests for this purpose. In the first, we generate the Cragg-Donald minimum eigen value statistic as if we were estimating our model as 2SLS (Cragg and Donald, 1993)[8]. Second, we use the Anderson-Rubin structural test as if we were estimating a probit with linear first step (Anderson and Rubin, 1949)[9]. And lastly, we use a Wald test of the excluded instruments in our Joint MLE estimation. Stock and Yogo (2005) generate critical values for models with up to three endogenous variables. We see as the number of endogenous regressors increase, the critical value also decreases. For three endogenous variables and ten instruments, we would need a critical value of at least 16.8. In all cases, we have sufficiently large F statistics, which implies we likely do not have weak instruments.

---

[8] The Cragg-Donald minimum eigen value statistic is generated by using the same specification as our model in a 2SLS estimation using STATA command *ivregress*, followed by STATA command *estat firststage*.
[9] The Anderson-Rubin test statistic is generated by the user-written STATA command *weakiv* (Finlay et al., 2013)

Table 1: Test Statistics for excluded instruments

Panel a: Tests of excluded instruments for single equations		
Endogenous Variable	Linear	Tobit
High-processed share	37.63	37.86
MAFH share	23.45	24.09
Hours spent in vigorous activity (100 hrs.)	21.81	19.44
Panel b: Tests of excluded instruments for multiple equations		
Test Statistic		Value
Cragg-Donald Minimum Eigen Value Statistic		44.78
Anderson-Rubin Test Statistic		17.83
Joint MLE Wald Statistic		26.58

Sources: TZA NPS (2008/9, 2010/11, 2012/13)

1.3 Data

We use individual level data from the first three waves of the Tanzania National Panel Survey (NPS), conducted in 2008/09, 2010/11 and 2012/13. The most recent survey, conducted in 2014/15, used a new sample that did not overlap with previous panels, so we exclude it to take advantage of panel estimation methods[10].

We followed several steps to clean the data. First, because the definition of overweight and obesity in younger individuals is based on reference means that vary by population and organization, we limited our analysis to adults, defined per WHO as individuals 19 years of age or older. Doing this reduced the total number of available observations to 23,895 from 45,863. Second, we removed any observations with either missing data, or unrealistic measurements likely due to errors during data collection or processing. This left us with 17,498 observations, distributed

---

[10] In our analysis we estimated numerous models which included pooled OLS, fixed effects, dynamic variations of the probit and other variations of the correlated random effects probit. We chose the unbalanced method because it allowed us to use more data for our analysis. The results were generally consistent across all estimations.

across rounds as shown in Table 2. The slight variation in number of individuals per round is due to attrition, missing data, and adolescents aging to 19 years between survey periods and thus becoming classified as adults. Attrition was very low, at 3% for round 2 and 4% for round 3 (National Bureau of Statistics, 2014) and we therefore ignore it in this analysis. Table 3 describes each variable in our main regression and Table 4 presents the sample-weighted means of our dependent variables and covariates. We grouped our variables into four categories: the individual's obesogenic environment, socioeconomic environment and welfare, food choice, and activity choice.

Table 2: Panel Statistics

	2008-09	2010-2011	2012-2013	Total
Households*	2,904	2,953	2,864	3,237
Number of Individuals with complete data				
1 period	1,044	470	639	2,153
2 periods	1,220	1,824	1,504	4,548
3 periods	3,599	3,599	3,599	10,797
Total	5,863	5,893	5,742	17,498

*Note: Households are defined by the identification number of the first period. The total number of households is not the sum of all columns, but rather the total number of unique households identified in the survey.

NBS collected anthropometric data for every household member present during the interview, which we use to calculate BMI. We estimate our model using WHO guidelines for underweight (BMI<18.5), normal (18.5<=BMI<25), overweight or obese (BMI>=25), and for only obese (BMI>=30). We estimate both because evidence of the relationship between body weight and health outcomes is somewhat mixed for overweight but much clearer for obesity (Flegal et al., 2013; Afzal et al., 2016; Tobias and Hu, 2018; Bhaskaran et al., 2018). Average BMI and the proportion of overweight or obese individuals all increased slightly in each wave and were always higher in urban areas than in rural areas. Following USAID guidelines for analyzing

anthropometric data from DHS surveys, we removed 51 observations that had a BMI above 60 or below 12 to reduce bias from measurement error. We also excluded 4,484 individuals who had no anthropometric data.

Table 3: Variable Description

Dependent variables	
Overweight or Obese	1 if individual has a BMI of 25 or greater
Obese	1 if individual has a BMI of 30 or greater
Obesogenic Environment	
Enumeration area prevalence	Average of positive outcomes of dependent variables (either Overweight/Obese or Obese) out of total number of adults within enumeration area
Distance to population center	The km distance to nearest population center of 20,000 or more
Rural	Indicator for rural locality
Socioeconomic Environment and Welfare	
Per Capita Total expenditure (ln)*	Natural Log of total daily per capita expenditure in 2012 dollars (includes food and nonfood).
Asset score*	Asset score calculated using principal component analysis using all available assets in the durables section, excluding motor vehicles
Education years	Years of education
Age in years	Age in Years
Female (0/1)	Female=1
Married or cohabit (0/1)	Married or Cohabit=1
HH size	Total size of household in absolute terms
Child 0-5 (0/1)	Indicator whether household has a child between 0 and 5 years old
Child 6-14 (0/1)	Indicator whether household has a child between 6 and 14
Self Employed (0/1)	1 if individual earned money from self-employment in the past 12 months
Wage Labor (0/1)	1 if individual earned money from wage labor in the past 12 months
Gave Birth in past 24 mo. (0/1)	Indicator whether a woman gave birth in the past 24 months
Overnight Stay in Hospital (0/1)	Indicator whether individual spent a night in a hospital or medical facility for an illness or injury in the past week
Overnight Stay with Healer (0/1)	Indicator whether individual spent a night with a traditional healer for an illness or injury in the past week
Unsatisfactory View of Life (0/1)	Indicator whether individual is dissatisfied or very dissatisfied with their life in general
Food Choice*	
Un-processed share	Value of un-processed food expenditure as share of total food expenditure
Low-processed share	Value of low-processed food expenditure as share of total food expenditure
High-processed share	Value of high-processed food expenditure as share of the total value of expenditure
MAFH share	Value of meals consumed away from home as share of the total value of expenditure
Carbs/AE	Per adult equivalent daily amount of grams of carbohydrates consumed

Table 3 (cont'd)

Fat/AE	Per adult equivalent daily amount of grams of fat consumed
Protein/AE	Per adult equivalent daily amount of grams of protein consumed
Activity Choice	
Hours Spent in Vigorous Activity	Hours spent working in mining, agriculture, or construction (wage labor, self-employed, unpaid, and family farm), colleting firewood, and collecting water over the past week
Owns a Motor Vehicle (0/1)	1 if household has a motorcycle or car
Agricultural work (0/1)	1 if individual worked in agriculture in the past 12 months
Office Work (0/1)	1 if individual worked in a field with ISIC code associated with office work

* Food, asset, and expenditure variables were collected at the household level, and averages are used to proxy for consumption, wealth, and expenditure of the individual within a household.

We control for the local environment with obesogenic environment variables, which include the share of overweight/obese individuals in the smallest unit above the household recorded in the survey, whether the individual lives in a rural or urban area, and their distance to the nearest population center[11]. We interpret the proportion of overweight/obese within an enumeration area as a measure of the social acceptability of obesity in one's immediate area combined with an enumeration area fixed effect.

Individual and household socioeconomic variables show a trend we should expect from a sample surveyed over several years, with mean ages rising for rural households but remaining relatively stable for urban households: rural-to-urban migrants are likely to be younger, counteracting the growing average age from the population already living in urban areas. Urban households are younger than rural households on average.

---

[11] Urban households had a positive value for distance to population center since the measure was to the center of the population center. We did not believe this was a relevant measure for urban households since they all had similar access but were just located different distances from the center of the city. We replaced this value with zero for urban households in the full sample and excluded the variable in the urban estimation.

Table 4: Descriptive Statistics (Means using sample weights)

	All Periods		2008-2009		2010-2011		2012-2013	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Obesogenic Environment								
Body Mass Index (kg/m^2)	24.21	21.88	24.10	21.78	24.03	21.93	24.50	21.94
Overweight or Obese (BMI>=25)	0.35	0.15	0.35	0.14	0.33	0.15	0.36	0.15
Obese (BMI>=30)	0.13	0.03	0.13	0.03	0.12	0.03	0.14	0.04
Distance to population center	0.00	53.55	0.00	52.98	0.00	53.73	0.00	53.93
Socioeconomic Environment and Welfare								
PC Food Expenditure (TZS)	1,382	508	1,608	513	1,253	523	1,282	489
PC Non-Food Expenditure (TZS)	1,371	369	1,537	358	1,263	380	1,311	367
PC Total expenditure (ln TZS)	7.62	6.38	7.74	6.37	7.54	6.41	7.57	6.37
Asset score (0-1)	0.57	0.26	0.55	0.22	0.56	0.27	0.60	0.29
Education years	7.26	4.71	7.37	4.56	7.03	4.69	7.39	4.89
Age in years	37.77	41.24	36.68	40.48	38.06	41.25	38.59	41.98
Female (0/1)	0.56	0.54	0.56	0.54	0.55	0.55	0.57	0.53
Married or cohabit (0/1)	0.60	0.72	0.59	0.71	0.61	0.73	0.60	0.72
HH size	5.21	6.10	4.97	5.91	5.46	6.30	5.21	6.08
Child 0-5 (0/1)	0.52	0.66	0.52	0.66	0.53	0.67	0.50	0.64
Child 6-14 (0/1)	0.59	0.72	0.57	0.72	0.62	0.73	0.59	0.72
Self Employed (0/1)	0.35	0.18	0.36	0.18	0.39	0.22	0.31	0.13
Wage Labor (0/1)	0.27	0.11	0.25	0.07	0.28	0.10	0.28	0.16
Gave Birth in past 24 mo. (0/1)	0.12	0.16	0.12	0.17	0.11	0.16	0.12	0.15
Overnight Stay in Hospital (0/1)	0.08	0.06	0.07	0.05	0.09	0.08	0.08	0.06
Overnight Stay with Healer (0/1)	0.01	0.03	0.01	0.03	0.02	0.03	0.01	0.02
Unsatisfactory View of Life (0/1)	0.54	0.45	0.54	0.45	0.54	0.44	0.52	0.47
Food Choice								
Un-processed expenditure share	0.04	0.03	0.04	0.03	0.04	0.04	0.04	0.03
Low-processed expenditure share	0.56	0.75	0.58	0.77	0.55	0.73	0.56	0.75
High-processed expenditure share	0.25	0.16	0.25	0.16	0.26	0.17	0.23	0.14
MAFH expenditure share	0.15	0.06	0.13	0.04	0.16	0.06	0.17	0.08
Carbs/AE (100g)	4.54	5.03	5.13	5.82	4.20	4.65	4.29	4.63
Fat/AE (100g)	0.70	0.55	0.79	0.61	0.64	0.52	0.66	0.53
Protein/AE (100g)	0.70	0.71	0.78	0.81	0.64	0.66	0.67	0.66
Activity Choice								
Hours Spent in Vigorous Activity	7.9	23.9	6.4	23.7	8.3	23.6	9.0	24.4
Owns a Motor Vehicle (0/1)	0.10	0.04	0.08	0.02	0.09	0.05	0.12	0.06
Agricultural work (0/1)	0.25	0.75	0.19	0.74	0.26	0.74	0.30	0.79
Office Work (0/1)	0.14	0.04	0.14	0.03	0.16	0.06	0.11	0.03
Observations	5,682	11,816	2,056	3,807	1,761	4,132	1,865	3,877

We proxy cognitive and aspirational factors with years of education, age, and subjective welfare. Education remained stable at just under 5 years on average for rural households, and

slightly over 7 years for individuals in urban areas. The rural population tends to be older and happier on average than urban in these data, and average age trends upwards only for rural populations.

The family environment is determined by the stability of the household and composition of its inhabitants. To account for this, we use an indicator for whether the individual is married or cohabiting (stability) and indicators for whether the household has children between 0 and 5, and 6 and 14 in addition to household size (household composition). The proportion of married and household composition remained relatively stable across the survey periods, with fewer households having children in successive periods, which is expected.

We proxy economic factors by per capita expenditure and an asset score. Per capita expenditure was calculated using 7-day recall for food, and 12-month and 30-day recalls for nonfood expenditures. The asset score is calculated using principal component analysis then normalized to lie between 0 and 1. Both show a weakly increasing trend for urban households, but more stable for rural households. Urban households are wealthier and spend more on average.

Table 5 displays the means for rural and urban dwellers by expenditure quartiles. We used the entire data set to generate the expenditure quartiles which is why the lowest quartile has more rural residents while the highest quartile contains more urban dwellers. Even controlling for quartile, rural rates of overweight/obesity are lower in every case than urban rates, especially in the 2nd and 3rd quartiles. We see a similar pattern with diet, with high processed and MAFH consistently higher in urban areas. Activity falls with quartile in both rural and urban. In the first (poorest) quartile it is only slightly higher in rural than urban areas but by the top quartile rural residents are nearly four times as active as urban residents,

Table 5: Means by Expenditure Quartiles (using sample weights) 2012-2013

	1st Quartile		2nd Quartile		3rd Quartile		4th Quartile	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Overweight or Obese (BMI>=25)	0.11	0.09	0.27	0.14	0.30	0.19	0.48	0.42
Un-processed expenditure share	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.04
Low-processed expenditure share	0.67	0.80	0.63	0.74	0.57	0.69	0.51	0.58
High-processed expenditure share	0.15	0.10	0.20	0.15	0.23	0.18	0.25	0.23
MAFH expenditure share	0.16	0.07	0.14	0.07	0.16	0.09	0.19	0.15
Hours Spent in Vigorous Activity	21.42	26.14	14.93	24.22	9.74	22.93	4.87	18.13
Observations	95	1,482	238	1,203	606	897	926	295

Figure 1 shows the prevalence of overweight or obese individuals by survey period plotted against the log of total per capita daily expenditure (in 2012 TZS) by survey period. Beyond the expected strong positive relationship, the most noteworthy pattern is that the rising trend seen globally emerges only around the middle of the income distribution, then becomes more accentuated as incomes rise: middle- and higher income Tanzanians clearly became more overweight and obese over the panel period, but those lower in the income distribution did not.

Figure 1: Prevalence of overweight or obese individuals vs. log of per capita total daily expenditure (2012 TZS) by survey period
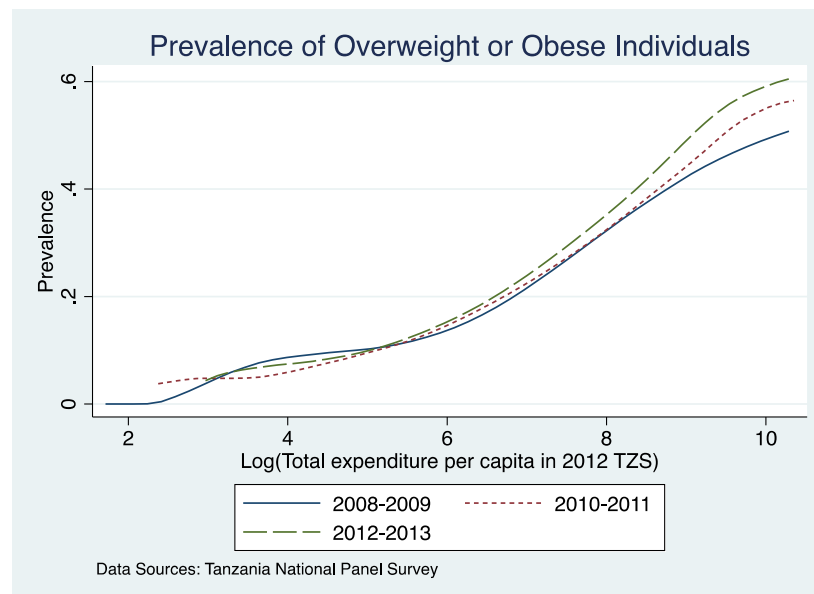
Figure 2 displays the same relationship comparing urban and rural areas. The lowest income rural residents are more likely to be overweight or obese than the lowest income urban residents, however, our results have limited validity because there are so few urban households at this income level[12]. From near the middle of the expenditure distribution, urban rates exceed rural rates, and this difference remains consistent with increasing income.

Figure 2: Prevalence of overweight or obese individuals vs. log of per capita total daily expenditure (2012 TZS) comparing individuals living in rural and urban areas
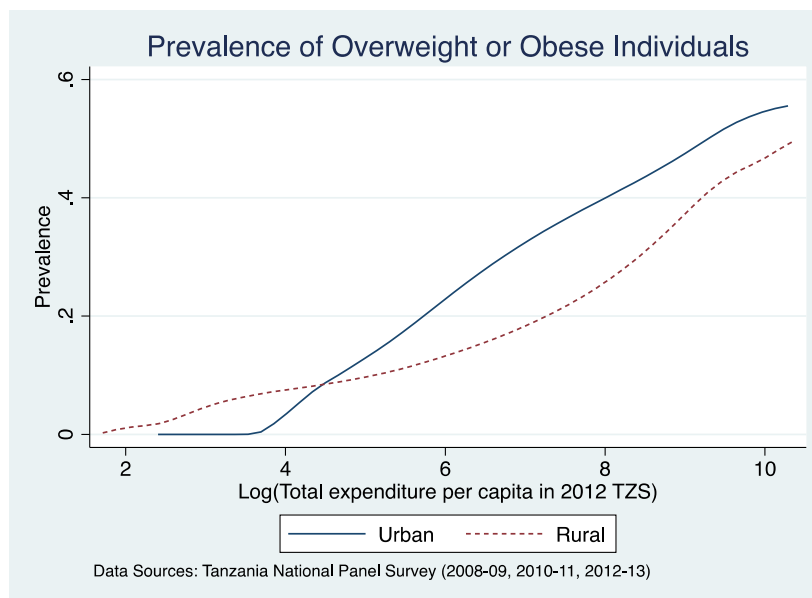
One might worry that we don't have sufficient variation to detect the impact of various drivers. A comparable number of individuals became overweight or obese as moved into the

---

[12]There are 693 rural inhabitants and only 26 urban residents below a log per capita household expenditure of 5 (~148 tzs).

normal/underweight weight range[13]. 1,003 out of 8,026, or 12% of the sample transitioned between weight categories, which we believe is large enough to detect drivers that increase or reduce body mass outcomes as categorized by BMI.

Income-expenditure surveys such as the NPS have well-known limitations as sources of consumption data, especially those that depend on brief periods of recall. Yet Fiedler and colleagues (Fiedler et al., 2012a; Fiedler et al., 2012b; Fiedler, 2013) make the case and provide guidelines for their use in such work. Diet variables in our regression are all at the household level, since individual consumption was not collected in the NPS. Our consumption share variables thus represent "average" consumption of an individual within the household by survey period. Consumption includes food expenditures, gifts, and own-production[14]. Food expenditure was recalled by the household head over the past 7 days. To make the values comparable we generated daily values per adult equivalent measured in real 2012 Tanzanian shillings (TZS)[15]. We drop transactions where the total consumption does not equal the sum of the purchased, gift, and own-production sources, which suggest some sort of clerical error in taking or inputting the survey data, or a misunderstanding on the part of the subject.

Our diet variables include shares of total food consumption of different processed- and prepared foods in the household's diet (proxy for diet composition), and the macronutrients (per AE) in the food consumed (proxy for amount of food). The processed food categories are based

---

[13] It is important to note that this only includes individuals with full data in either 2 or 3 periods. And while there are 994 total transitions, the number of unique individuals that transitioned is 867. This implies some individuals observed in 3 periods transitioned in both waves, going the opposite direction they transitioned in the previous period.

[14] We chose to use the value of the food consumed rather than the calories to calculate the shares because doing so would prohibit the inclusion of food away from home as a share of total consumption in our analysis. Food away from home is not measured in quantities, just dollar amounts, so there is no way to adequately determine total calories consumed as food away from home.

[15] We used the average inflation rate from the World Bank to adjust the TZS to 2012 TZS.

on Sarfo et al. (2021), where processed food categorization is based on the state of the food at the point of consumption. This means that any food cooked at home is at least considered low-processed, and the only unprocessed foods are fruits. While this categorization is more defensible from a nutritional perspective it does come with drawbacks. Due to the high number of zeros in fruit consumption, we group unprocessed and low processed together, similar to the NOVA categorization of unprocessed and minimally processed. Another drawback is that we do not distinguish between high and ultra-processed foods. Meals away from home (MAFH) is any meal prepared and consumed outside the home[16]. Urban households consume a larger share in the form of high-processed food and meals away from home. Unprocessed shares are nearly identical in rural and urban areas, while low-processed shares are higher in rural areas, and high processed shares higher in urban areas.

Macronutrient content of the diet comes from matching food items in the NPS to the Tanzania Food Composition Tables and using averages for NPS items with multiple matching items in the food tables (Lukmanji et al., 2008). We follow (Larsen et al., 2018), and remove households with total caloric consumption far above or below the total household energy requirement. To do so, we first used the method outlined in the joint FAO/WHO/UNU report (2001) to calculate the individual energy requirement, then calculated the energy requirement of the household. We used a lower limit of p25-1.5*IQR and an upper limit of p75+1.5*IQR and removed 393 observations outside of this range[17]. Urban households, which are richer on average

---

[16] The NPS contains expenditures on 7 categories of food away from home. These include meals, barbecued snacks, locally brewed alcohol, commercially produced alcohol, non-alcoholic sugary beverages, sweets, and snacks. We grouped barbecue and meals together for the meals away from home category, and add the remaining 5 categories to the high-processed food group.
[17] p25= 25th percentile, p75=75th percentile, IQR= inner quartile range.

than rural households, tend to eat more fats and protein, while rural households eat a larger amount of carbohydrates, which is consistent with Bennet's Law.

We attempt to capture general health using 4 indicators. The first equals one if the individual was pregnant within the past 24 months. The effect may be ambiguous as those closer to their previous birth might have a higher BMI, but those closer to the 2-year mark may no longer be physically impacted by the pregnancy. The next are two indicators for whether an individual spent the night in a hospital or with a traditional healer for an injury or illness in the past week. We can think of these indicators as measures of general health, which will influence both diet and activity choice. Someone in poor health is more likely to suffer an illness that requires hospitalization, which also would limit their ability to engage in activities. And lastly, we include an indicator if the person answered they were unsatisfied, or very unsatisfied with their life.

To proxy for activity choice, we include hours of vigorous activity as well as indicators for assets or work that is associated with higher levels of energy requirements. We define hours spent in a vigorous activity as hours spent working over the past seven days in a labor-intensive job (agriculture, forestry/wildlife, mining work, or construction) or in household activities such as firewood and water collection, as well as working on the household farm. We remove any quantities totaling greater than 18-hour days (assuming a minimum of 6 hours of sleep per day). Rural households spend a significantly larger portion of their time engaged in vigorous activities.

1.4   Results and Discussion

Table 6 displays the average marginal effects of four variations of the probit using the full sample, while tables 7 and 8 are from the urban and rural subsamples, respectively. Column (1) shows the

24

results from a pooled probit; in column (2) we add correlated random effects (CRE) and allow for heteroskedasticity as outlined in the methods section; and in column (3) we apply an instrumental variable estimation using joint MLE with CRE (this is the structure for every table that reports regression results with numerical column headings). Results are generally in line with our expectations. High-processed foods and meals eaten away from home show a strong positive relationship with the likelihood of being overweight or obese in the full sample, while only HPF is marginally relevant in urban areas, and MAFH in rural areas. Hours spent in a vigorous activity reduce the likelihood of being overweight, and the effect is significant mainly in rural areas. Overweight and obesity seem to be spatially clustered in rural areas, as shown by the significant positive coefficient on overweight and obesity prevalence in a household's sample cluster. Expenditure and asset score show a strong positive impact, and have a much larger effect in urban areas compared to rural. The effects from age and education are positive across most estimations with age showing a larger effect in urban areas, and education in rural areas. Lastly, we estimate both gender and marital status have a positive impact, with females being nearly twice as likely to be overweight or obese in urban areas.

1.4.1   Full Sample

Our estimates suggest dietary consumption patterns, when categorized by level of processing, have a large and significant effect on body mass outcomes. A BMI greater than 25 is strongly related to the proportion of both high-processed food and meals eaten away from home, after controlling for macronutrient content and income. We can interpret the results in column (1) from table 6 as follows: a 10 percentage point increase in the share of HPF or MAFH is associated,

respectively, with a 0.9 or 0.5 percentage point increase in the likelihood of being overweight or obese. These effects appear smaller when we control for unobserved heterogeneity, but still significant. When we instrument for processed food expenditures and activity level, we find that HPF consumption is the only diet variable that has a substantial impact on the likelihood of being overweight or obese.

Lifestyle choices that impact level of daily activity are also an important factor in determining risk of overweight/obesity. For every 20 hours per week spent in a vigorous activity we estimate a between a 0.5 and 1.7 percentage point decrease in the likelihood of increasing body mass over a BMI of 25. This may seem like a small number. However, if an individual is active for half of their available time—assuming six hours of sleep per night, there are 126 hours available for work or leisure— then the likelihood is reduced by between 1.5 and 5.4 percentage points, a practically large result.

We find that our economic variables have a consistent and important effect even after controlling for type and amount of food consumed as well as activity. In all three models, the effects are positive and significant suggesting both wealth and income are relevant in determining body mass outcomes. The impacts from both expenditure and asset score diminish when we control for unobserved heterogeneity and endogeneity, dropping roughly be one-half. A 10 percentage point increase in asset score is associated with a 0.6 to 1.3 percentage point increase in overweight/obesity risk. Increasing log expenditure by one around the mean (6.88) is equivalent to roughly a 100 TZS increase in expenditure, which increases overweight/obesity likelihood by 2.7 to 5.6 percentage points, a larger effect than a comparable change in asset score.

Table 6: Average Marginal Effects on the probability of being overweight or obese (BMI>=25) from variations of the probit using the full sample

VARIABLES	(1)	(2)	(3)
Obesogenic Environment			
Enumeration area prevalence	0.0743***	0.0353	0.0566**
Distance to population center	0.0258	-0.0505	-0.0487
Rural	-0.0275*	-0.0121	-0.0132
Socioeconomic Environment and Welfare			
Per Capita Total expenditure (ln)	0.0580***	0.0376***	0.0352***
Household Asset score (0-1)	0.127***	0.0568***	0.0571***
Education years	0.00396***	0.00363**	0.00427**
Age in years	0.00515***	0.00380***	0.00495***
Female (0/1)	0.163***	0.154***	0.163***
Married or cohabit (0/1)	0.0636***	0.0431***	0.0470***
HH size	0.00247	0.00240	0.00257
Child 0-5 (0/1)	0.0124	0.00200	-0.00184
Child 6-14 (0/1)	0.0111	-1.76e-05	0.00759
Self-employed (0/1)	0.0448***	0.0150**	0.0123
Wage labor (0/1)	-0.00372	0.00186	-0.00277
Gave birth in past 24 mo. (0/1)	-0.0346***	-0.0389***	-0.0367***
Overnight stay in hospital (0/1)	-0.000862	0.0164	0.0145
Overnight stay with healer (0/1)	-0.00129	-0.0147	-0.00345
Unsatisfactory view of life (0/1)	-0.00490	-0.00834	-0.0124**
Food Choice			
High-processed share	0.0922***	0.0247	0.282**
MAFH share	0.0542**	0.0454***	0.0667
Carbs/AE	0.00208	0.00102	0.00261
Fat/AE	0.0339***	0.00964	-0.00772
Protein/AE	-0.0283*	-0.0168	-0.00350
Activity Choice			
Hours spent in vigorous activity (100 hrs.)	-0.0866***	-0.0313*	-0.0256
Owns a motor vehicle (0/1)	0.0437***	0.0189	0.0239*
Agricultural work (0/1)	-0.0190**	-0.00735	-0.0103
Office work (0/1)	-0.00803	0.00794	0.00730
Model Characteristics			
Correlated Random Effects	No	Yes	Yes
Instrumental Variables	No	No	Yes
Heteroskedasticity	No	Yes	No
Region Indicators	Yes	Yes	Yes
Survey Period Indicators	Yes	Yes	Yes
Number of overweight/obese	4,107	4,107	4,093
Observations	17,498	17,498	17,453
Prevalence	23.47%	23.47%	23.45%

*** p<0.01, ** p<0.05, * p<0.1

Our proxies for cognitive ability and experience, using years of education and age, are positively related to the risk of excessive body mass. The effects are similar from both variables and do not change much between models. At the margins, increasing by 1 year in age results in an increase in the likelihood of being overweight or obese by 0.4 to 0.5 percentage points, while there

is a comparable change from an increase in education years. This is consistent with our expectations as we expect the risk of overweight/obesity to increase greatly in the latter half of one's life.

The most important aspect of the family environment in our estimation is marital status. Those that are married or cohabiting are more likely to be overweight or obese by roughly 4.3-6.3 percentage points, suggesting a stable home environment is positively related to body mass increases. Household composition, however, seems to have no impact on overweight or obesity. This is interesting, considering we might expect household composition to impact health through similar channels as marriage.

We find evidence of spatial clustering given by the highly significant effect from enumeration area prevalence. A 10 percentage point increase in the proportion of individuals overweight or obese within the enumeration area, is associated with a 0.4 to 0.7 percentage point increase in the probability of an individual in that geographical area being overweight or obese. It is interesting that both the rural indicator and distance to population center do not show a significant relationship, which suggests the other variables in our model may adequately account for rural-urban differences relevant to body mass outcomes.

The remaining estimates included both expected and unexpected outcomes. Females have an increased likelihood of being overweight or obese by 15.4 to 16.3 percentage points – the largest effect from any of our covariates – which is consistent with the literature. The signs on the health and welfare variables, however, are surprising. One might expect giving birth recently to have a positive impact on body mass due to the physiological characteristics of pregnancy, but our estimates are negative. Subjective welfare is only marginally significant in one of the three models. We also find overnight medical visits do not have an effect statistically different from zero with

the exception of a marginally significant effect in one model and only one of the variables, which we might expect otherwise.

1.4.2   Rural-Urban Differences

The most intriguing result from the Rural and Urban estimations are the differential effects of activity and diet choices. The impact of HPF and MAFH consumption is much larger in the urban subsample compared to rural (where the effect of HPF is not statistically significant at all), which is likely driving the significance in the full sample. These effects may be impacted by availability, as we might expect the variety and abundance of processed foods to be greater in urban areas. This stark difference may be explained by the difference in average consumption patterns between the two locality types. Both rural and urban households consume similar proportions of un-processed foods, however urban households consume twice as large a share of HPF and nearly three times as large a share of MAFH, while consuming only two-thirds as large a share of low-processed food when compared with rural households.

We see similar differences in activity patterns, with rural households spending three times as many hours per week engaged in a physically demanding activity than urban households. Some rural agricultural households, that collect their own firewood and water, likely spend a majority of their available time physically engaged, while the opposite may be true for urban households with running water and electricity. This is perhaps why the coefficient in the rural sample is much more significant than in the urban sample. It is interesting to note that the coefficient is much larger in the urban sample (and even positive in the JMLE), but the lower significance suggests the relationship may not be accurately estimated.

Table 7: Average Marginal Effects on the probability of being overweight or obese (BMI>=25) from variations of the probit using the urban sub-sample

VARIABLES	(1)	(2)	(3)
Obesogenic Environment			
Enumeration area prevalence	0.0365	0.00294	-0.00163
Socioeconomic Environment and Welfare			
Per Capita Total expenditure (ln)	0.0934***	0.0755***	0.0649***
Household Asset score (0-1)	0.0813**	0.0550	0.0690**
Education years	0.00767***	0.00331	0.00450
Age in years	0.00998***	0.00782***	0.00915***
Female (0/1)	0.227***	0.217***	0.227***
Married or cohabit (0/1)	0.0801***	0.0443**	0.0498**
HH size	0.00335	0.00456	0.00231
Child 0-5 (0/1)	0.0350**	0.0179	0.0199
Child 6-14 (0/1)	0.00307	-0.00275	-0.00510
Self-employed (0/1)	0.0565***	0.0215	0.0276*
Wage labor (0/1)	-0.00606	0.00391	-0.00730
Gave birth in past 24 mo. (0/1)	-0.0581***	-0.0296	-0.0326*
Overnight stay in hospital (0/1)	-0.0130	-0.000255	-0.0134
Overnight stay with healer (0/1)	0.0341	-0.000404	0.0498
Unsatisfactory view of life (0/1)	-0.000453	-0.0163	-0.0238**
Food Choice			
High-processed share	0.174***	0.0241	0.561*
MAFH share	0.104***	0.0453	-0.0159
Carbs/AE	-0.00462	-0.00196	-0.00225
Fat/AE	0.0830***	0.0184	-0.000172
Protein/AE	-0.0516	-0.0225	-0.00207
Activity Choice			
Hours spent in vigorous activity (100 hrs.)	-0.0984*	-0.0169	0.113
Owns a motor vehicle (0/1)	0.0608**	0.0397	0.0516**
Agricultural work (0/1)	-0.0437**	-0.0379**	-0.0544**
Office work (0/1)	-0.0224	-0.00130	-0.000163
Model Characteristics			
Correlated Random Effects	No	Yes	Yes
Instrumental Variables	No	No	Yes
Heteroskedasticity	No	Yes	No
Region Indicators	Yes	Yes	Yes
Survey Period Indicators	Yes	Yes	Yes
Number of overweight/obese	2,120	2,120	2,113
Observations	5,682	5,682	5,663
Prevalence	37.31%	37.31%	37.31%

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

While the relationship between our economic proxies (total expenditure and asset score) and overweight/obesity is positive in all estimations, the effect from expenditure is larger in the urban subsample, which is much wealthier on average. It is possible there are some differences in diet and activity that are related to income and wealth that our variables do not capture. One such explanation might be that the types and amounts of foods available, and in particular their energy

density, differ between rural and urban localities. This would make sense if a large portion of high-processed foods are imported through, or processed near, the relevant urban center. It is also likely that there are larger and more diverse markets for MAFH in urban areas, which could potentially be more obesogenic.

Table 8: Average Marginal Effects on the probability of being overweight or obese (BMI>=25) from variations of the probit using the rural sub-sample

VARIABLES	(1)	(2)	(3)
Obesogenic Environment			
Enumeration area prevalence	0.0765***	0.0701**	0.0575*
Distance to population center	0.0211	-0.0127	-0.00820
Socioeconomic Environment and Welfare			
Per Capita Total expenditure (ln)	0.0436***	0.0278***	0.0201**
Household Asset score (0-1)	0.155***	0.0338	0.0376
Education years	0.00237*	0.00597***	0.00628***
Age in years	0.00312***	0.00308***	0.00353***
Female (0/1)	0.133***	0.131***	0.131***
Married or cohabit (0/1)	0.0494***	0.0507***	0.0534***
HH size	0.00162	0.00194	0.00227
Child 0-5 (0/1)	0.00124	-0.0165*	-0.0182*
Child 6-14 (0/1)	0.0170*	0.0132	0.0199**
Self-employed (0/1)	0.0295***	0.00322	0.00262
Wage labor (0/1)	-0.00441	0.00544	0.00526
Gave birth in past 24 mo. (0/1)	-0.0277***	-0.0345***	-0.0345***
Overnight stay in hospital (0/1)	0.00443	0.0350**	0.0284**
Overnight stay with healer (0/1)	-0.0130	-0.0213	-0.0171
Unsatisfactory view of life (0/1)	-0.00867	-0.000817	-0.00489
Food Choice			
High-processed share	0.0539	0.0364	0.250
MAFH share	0.0281	0.0260	0.0907*
Carbs/AE	0.00336	0.00163	0.00422*
Fat/AE	0.00919	-0.000922	-0.0145
Protein/AE	-0.0120	-0.00897	0.00174
Activity Choice			
Hours spent in vigorous activity	-0.0839***	-0.0439**	-0.0978
Owns a motor vehicle (0/1)	0.0432**	0.0133	0.0206
Agricultural work (0/1)	-0.00703	0.00386	0.0149
Office work (0/1)	-0.00414	0.00827	0.000205
Model Characteristics			
Correlated Random Effects	No	Yes	Yes
Instrumental Variables	No	No	Yes
Heteroskedasticity	No	Yes	No
Region Indicators	Yes	Yes	Yes
Survey Period Indicators	Yes	Yes	Yes
Number of overweight/obese	1,987	1,987	1,980
Observations	11,816	11,816	11,790
Prevalence	16.82%	16.82%	16.79%

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

We find age and education are both positively related to overweight/obesity, while education is only significant for the base probit in the urban subsample. This effect is larger than any of the estimated effects in the rural population, however, the effect from education is more consistent in the rural estimations. The relation of age and overweight/obesity is twice as strong in urban areas. Similarly, the average age is lower in urban areas, so the marginal increases in body mass from aging are larger than in rural areas.

In rural localities we find evidence of spatial clustering given by the significant effect from enumeration area prevalence. A 10 percentage point increase in the proportion of individuals overweight or obese within the enumeration area, is associated with a 0.6 to 0.8 percentage point increase in the probability of an individual in that geographical area being overweight or obese. An enumeration area in a rural setting is more likely closer to a village, while there are numerous enumeration areas within an urban locality. For the rural person, this means the totality of their social experience likely occurs within an enumeration area, while those in a city will have contact with many outside of their designated enumeration area. This is perhaps why we only see a significant relationship from enumeration area prevalence in the rural context. This is the only local environmental factor we find relevant.

These differences are further showcased when we estimate the same models with interactions between the rural/urban indicator and our main diet and activity choice variables (HPF, MAFH, Active Hours). The results are displayed in Table 9, which exhibit the same patter we see when we estimate the subsamples separately. This is interesting given the probit model is nonlinear, and adding nonlinearities into the specification can sometimes produce different results. However, we find that MAFH and HPF only have an estimated effect in urban areas, whereas there is no evidence of an impact from activity hours outside of the naive probit.

Table 9: Average Marginal Effects for Urban and Rural Households on the probability of being overweight or obese (BMI>=25) from variations of the probit using the full sample with interactions between rural/urban indicators and HPF, MAFH, and Activity hours

VARIABLES	(1)	(2)	(3)
High-processed share			
Rural	0.0522	-0.00104	0.149
Urban	0.145***	0.0560	0.253*
MAFH share			
Rural	0.0414	0.0375	0.0472
Urban	0.0698**	0.0498**	0.0794*
Hours spent in vigorous activity			
Rural	-0.0848***	-0.0325	-0.0108
Urban	-0.0883**	-0.0193	0.0472
Number of overweight or obese	1,325	1,325	1,318
Observations	17,498	17,498	17,453
Prevalence	7.57%	7.57%	7.55%

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

There is one last point worth noting. The lack of an effect of diet choice in rural areas might seem odd, but it may have to do with (1) the available HPF and MAFH is less obesogenic in rural areas, which is not measured in the NPS, (2) the much higher activity levels observed in rural areas may be enough to prevent energy imbalance from tilting toward excess calorie intake, and (3) consumption shares of high-processed and MAFH need to pass a certain threshold before they exhibit obesogenic effects. These issues cannot be adequately answered in this paper and warrant further research.

1.4.3 Using Continuous BMI as the Dependent Variable

We opted to use a binary indicator as our main variable since we believe it is more important to understand what drives transitions at the margins between weight categories. However, we include results using a fixed effects estimation in table 10 with additional interactions between rural and urban households.

Table 10: Effects of Diet and Activity on Log(BMI), using the Full sample, with interactions between diet/activity and the rural/urban indicator

VARIABLES	(1)	(2)	(3)
Food Choice			
High-processed share	0.00496		0.0115
High-processed share x Urban		-0.00905	-0.0188
MAFH share	0.00890		0.00723
MAFH share x Urban		0.00888	0.00264
Carbs/AE	-0.000276	-0.000354	-0.000244
Fat/AE	0.000127	0.000413	-1.74e-05
Protein/AE	0.00717	0.00658	0.00712
Activity Choice			
Hours spent in vigorous activity	0.000200		-0.00234
Hours spent in vigorous activity x Urban		0.0128	0.0143
Owns a motor vehicle (0/1)	0.00401	0.00393	0.00397
Agricultural work (0/1)	-0.00480**	-0.00550***	-0.00501**
Office work (0/1)	-0.00135	-0.00106	-0.00113
Number of obese	4,107	4,107	4,107
Observations	17,498	17,498	17,453
Prevalence	23.47%	23.47%	23.53%

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

We do not find any impacts from our main diet and activity variables; however, we do find agricultural labor shows a very strong negative relationship with BMI. In part, the reason for the lack of results may be explained if we segment our analysis by weight category, which is shown in Table 11.

Interestingly, the only weight category where we estimate any impacts from diet or activity choices are for the normal weight category. We find there is a general positive association with HPF and BMI, however, the interaction with the urban indicator is slightly larger in magnitude and negative, which tells a different story than our previous results. Urban households consuming HPF have lower BMIs than rural households that do so. We do find a negative association with hours in a vigorous activity, but the interaction with the urban indicator is not significant.

One might infer these results conflict with our conclusion, however, I do not think it invalidates our previous results. While Tables 10 and 11 suggest there are limited impacts from diet and activity on BMI, we should be more cautious interpreting these results. Using a continuous

BMI dependent variable exhibits much greater amount of variation, and likely is heterogeneously impacted by many factors we cannot control for, such as genetics or basal metabolic rate. The benefit to focusing on binary indicators is we can look at probabilities of either being overweight or not, which is less sensitive to the varied impacts from diet and activity choices on weight outcomes. This allows us to focus on the marginal impacts of diet and activity choices for individuals near weight category cut offs, which are likely the most susceptible to becoming overweight or obese. This idea is supported by the fact we only estimate significant impacts from diet and activity on BMI in the Normal weight range.

Table 11: Fixed Effects estimation of Diet and Activity on Log(BMI), using the Full sample, with interactions between diet/activity and the rural/urban indicator, by weight category

VARIABLES	Underweight	Normal	Overweight	Obese
Food Choice				
High-processed share	-0.0183	0.0142*	0.0224	0.0274
High-processed share x Urban	0.0401	-0.0438***	-0.0386	-0.0373
MAFH share	-0.0132	-0.000280	0.0180	-0.0172
MAFH share x Urban	0.0382	-0.00413	-0.0232	0.0158
Carbs/AE	-0.000285	3.22e-05	-0.000118	4.43e-05
Fat/AE	-0.0108	0.00412	0.00385	0.000128
Protein/AE	0.0127*	0.00125	0.000118	0.00478
Activity Choice				
Hours spent in vigorous activity	-0.00622	-0.00749**	0.00762	-0.0275
Hours spent in vigorous activity x Urban	0.0153	0.0105	-0.00127	0.0591***
Owns a motor vehicle (0/1)	-0.0103	0.00515	0.00512	0.0120
Agricultural work (0/1)	-0.00528	-0.00228	-0.00755*	-0.00579
Office work (0/1)	0.0108	-0.00262	-0.00900*	0.0105
Number of Individuals in Weight Category	1,864	11,496	2,775	1,318
Observations	17,453	17,453	17,453	17,453
Prevalence	10.68%	65.87%	15.90%	7.55%

*** p<0.01, ** p<0.05, * p<0.1

1.4.4 Diet and Activity Choice Simulation

To better understand the relative contribution of dietary and activity factors we estimate the impact of changes in dietary and activity choices on the likelihoods of being overweight or

35

obese (table 12). To do this we estimate the separate impact of a person moving from the 50th percentile to the 75th percentile in the combined consumption of HPF and MAFH, and also in the number of hours spent in a vigorous activity. In rural areas, the change in physical activity reduces the likelihood of being overweight or obese by roughly 0.8 to 1.9 percentage points (a 4%-11% decrease), while increasing consumption of HPF and MAFH increases the likelihood by 0.3 to 1.8 percentage points (an increase between 1.8% and 11%), only showing a practically large effect when diet is instrumented. We should consider the lack of significance from the marginal effect of HPF and MAFH in the rural estimation as a reason to question the magnitude of this result. The effect of diet is much larger in urban areas, with an increase in HPF and MAFH consumption from the 50th to 75th percentile increasing the likelihood of being overweight or obese by 0.4 to 2.9 percentage points, or between a 1.1% and 7.3% increase. These results echo what we discussed above, namely that processed food consumption seems to pose more of a risk for individuals living in urban localities.

We explore the same simulation using obesity as the dependent variable in the binary choice model (Table A1). We find similar effects from HPF and MAFH in the urban subsample, while the size of the effects from changes in diet and activity choices elsewhere were very small, which suggest these changes have little impact on the probability of being obese. This stands contrary to much of the literature. We believe the small number of positive outcomes likely contributes to the lack of a statistically measurable impact from diet and activity choices in the rural sample, which likely influence results from the full sample. Our simulation reflects a relevant impact from changing consumption patterns in urban localities on the likelihood of being obese. Increasing consumption of high-processed food and meals eaten outside of the home increases obesity likelihood 0.1 to 3.1 percentage points for urban households (a 0.9% to 19.4% increase), while a

similar increase in activity hours in rural areas decreases the likelihood by roughly 0.2 to 0.6 percentage points (a 4.2% to 13.4% decrease). These results suggest processed food consumed in urban environments may have a more obesogenic impact. The reasons for this are outside of the scope of this study and should be further investigated.

Table 12: Simulation on the impact of diet and activity on the probability of being overweight or obese (BMI>=25)

		Variable: 50th to 75th Pct.	at 50th Pct. (%)	at 75th Pct. (%)	Change in Probability
Full	Probit	Processing & MAFH	23.07	24.08	1.01
		Active Hours	24.12	22.03	-2.09
	Het Probit CRE	Processing & MAFH	23.04	23.77	0.73
		Active Hours	23.67	22.92	-0.75
	Probit JMLE	Processing & MAFH	22.70	25.58	2.89
		Active Hours	23.81	23.06	-0.75
Urban	Probit	Processing & MAFH	37.13	39.35	2.22
		Active Hours	37.87	37.74	-0.12
	Het Probit CRE	Processing & MAFH	37.39	37.81	0.41
		Active Hours	37.43	37.43	0.00
	Probit JMLE	Processing & MAFH	39.29	42.14	2.86
		Active Hours	36.93	37.08	0.15
Rural	Probit	Processing & MAFH	16.69	17.01	0.32
		Active Hours	16.90	15.42	-1.48
	Het Probit CRE	Processing & MAFH	16.61	16.91	0.31
		Active Hours	16.81	16.06	-0.75
	Probit JMLE	Processing & MAFH	16.29	18.10	1.81
		Active Hours	17.12	15.27	-1.85

Note: Processing changes is in high processed share (Full sample: 0.21 to 0.28; urban subsample: 0.27 to 0.30; rural subsample: 0.16 to 0.24) and MAFH share (Full sample: 0.02 to 0.10; urban subsample: 0.10 to 0.25; rural subsample: 0.02 to 0.05). Activity change is of active hours ((Full sample: 7 to 30 hours per week; urban subsample: 0 to 1 hours per week; rural subsample: 18 to 36 hours per week)

1.4.5   Robustness Checks

1.4.5.1   Obesity (BMI>=30)

Table 13 shows our estimation using an indicator of obesity (BMI>=30) as the dependent variable, which serves as a robustness check as to whether the relationships hold at higher body

mass measures. The most prominent result is that we see a similar effect from processed food, albeit smaller in magnitude and not all statistically significant. We do see effects from HPF and MAFH for urban households. Hours spent in a vigorous activity is only marginally significant in the urban subsample, with a similar magnitude to the overweight/obese estimation[18]. We also see the same patterns from asset score and total expenditure, as well as education, age, and gender[19]. Interestingly, we find that protein consumption is negatively related to obesity, which echoes a portion of the results from Larsen et al. (2018). This suggests the relationships we found previously hold at higher levels of body mass.

Table 13: Average Marginal Effects on the probability of being obese (BMI>=30) from variations of the probit using the full sample

VARIABLES	(1)	(2)	(3)
Food Choice			
High-processed share	0.0682***	-0.00567	0.182**
MAFH share	0.0138	0.00846	0.0589**
Carbs/AE	8.95e-05	0.00134	0.00197
Fat/AE	0.0163*	0.00765	0.00472
Protein/AE	-0.0127	-0.0211**	-0.00842
Activity Choice			
Hours spent in vigorous activity	-0.0413**	-0.0134	-0.0310
Owns a motor vehicle (0/1)	0.0232**	0.0115	0.0163*
Agricultural work (0/1)	-0.0112*	-0.000598	0.00105
Office work (0/1)	-0.000563	0.00561	0.00276
Number of obese	1,325	1,325	1,318
Observations	17,498	17,498	17,453
Prevalence	7.57%	7.57%	7.55%

*** p<0.01, ** p<0.05, * p<0.1

---

[18] The lower level of statistical significance in some cases is likely the result of the small number of obese individuals limiting the variation necessary to detect an impact.

[19] Results for categories other than food and activity choice, as well as the urban and rural sub-samples, are available upon request from the authors.

1.4.5.2   Comparison with Normal BMI

Next, we explore excluding observations below the threshold BMI of 18.5 (Table 14), which is the cutoff for a normal weight status[20]. This method enables us to compare the "good" or normal range (BMI 18.5-25) to the "bad" or overweight/obese range (BMI 25+). For brevity, we only include the effects from the diet and activity choices, since these variables are the focus of this paper[21].

Table 14: Average Marginal Effects on the probability of being overweight or obese (BMI>=25) from variations of the probit using the full sample (with BMI<=18.5 excluded)

VARIABLES	(1)	(2)	(3)
Food Choice			
High-processed share	0.102***	0.0252	0.368**
MAFH share	0.0470**	0.0485***	0.0596
Carbs/AE	0.00277	0.000794	0.00287
Fat/AE	0.0315**	0.00769	-0.0147
Protein/AE	-0.0305	-0.0126	0.00293
Activity Choice			
Hours spent in vigorous activity	-0.0964***	-0.0416**	-0.0480
Owns a motor vehicle (0/1)	0.0409**	0.0153	0.0230
Agricultural work (0/1)	-0.0214**	-0.0116	-0.0107
Office work (0/1)	-0.00891	0.00587	0.00482
Number of overweight/obese	4,107	4,107	4,093
Observations	15,632	15,632	15,589
Prevalence	26.27%	26.27%	26.26%

*** p<0.01, ** p<0.05, * p<0.1

We can see that the effect from processed food consumption is still significant. The magnitude of the effect from high-processed foods is larger while the effect from MAFH is smaller than when we include the full BMI range, which suggests including the underweight population

---

[20] There are 434 (8%) observations excluded from the urban subsample and 1,432 (12%) observations excluded from the rural subsample.

[21] Results for categories other than food and activity choice, as well as the urban and rural sub-samples, are available upon request from the authors.

reduces the measurable impact from processed food consumption. The estimate of the effect from activity level is comparable to both our previous results from the urban and rural sub-samples. This is likely due to the higher proportion of agricultural households from rural areas (higher activity levels on average) that are excluded when restricting the sample to a BMI greater than or equal to 18.5.

1.4.5.3   Balanced Panel[22]

Previously, we argued that the data are likely not missing in a systematic manner, but we cannot know this with complete certainty. So, we explore how our results may change if we only include those individuals with complete data in all three survey periods in table 15.

Table 15: Average Marginal Effects on the probability of being overweight or obese (BMI>=25) from variations of the probit using the full sample (with a balanced panel)

VARIABLES	(1)	(2)	(3)
Food Choice			
High-processed share	0.113**	0.0355	0.306
MAFH share	0.0530*	0.0206	0.0579
Carbs/AE	0.000971	-0.00107	0.00182
Fat/AE	0.0176	-0.00816	-0.0311*
Protein/AE	-0.0145	0.0186	0.0314
Activity Choice			
Hours spent in vigorous activity	-0.0819***	0.00190	0.0367
Owns a motor vehicle (0/1)	0.0348	0.00220	-0.00208
Agricultural work (0/1)	-0.0301**	-0.0175**	-0.0277
Office work (0/1)	-0.000837	0.0179*	0.0179
Number of overweight/obese	2,669	2,669	2,660
Observations	10,797	10,797	10,770
Prevalence	24.72%	24.72%	24.70%

*** p<0.01, ** p<0.05, * p<0.1

---

[22] Results for categories other than food and activity choice, as well as the urban and rural sub-samples, are available upon request from the authors.

For us to use a balanced panel, we reduce our total sample size from 17,498 to 10,797, which is slightly more than half. We also adjust our CRE framework slightly and revert to the traditional Mundlak device. With so many fewer observations it is not surprising that we see fewer significant effects. We do, however, see a positive and at least marginally significant impact from HPF and MAFH in the base probit. We see a similar lack of significance for activity choice, with agricultural work the only variable with a significant effect in models other than the base probit.

1.4.5.4 Alternative BMI Cutoff[23]

We previously noted the health consequences of BMIs exceeding 25 are not clear, although it is used as a starting point in much of the literature that examines nutritional outcomes. Recently, Teufel et al. (2021) examined BMI and the risk of diabetes. They estimated optimal BMI cutoffs that were associated with the greatest marginal increases in the risk of diabetes by World Bank Regions. For Sub-Saharan Africa, the optimal cutoffs were found to be 27.3 for women and 25.4 for men. This implies that using a BMI cutoff of 25 may not be directly related to an increased risk of diabetes. However, it may still be true that the risk of other nutrition related health outcomes increases above a BMI of 25. We explore how our results change using these optimal cutoffs in Table 16, which are directly associated with an increased risk of diabetes. HPF and activity level is only significant in the base probit. This suggests the effect from either diet or activity on the likelihood of being in a BMI range associated with negative health risks is not clear.

---

[23] Results for categories other than food and activity choice, as well as the urban and rural sub-samples, are available upon request from the authors.

Table 16: Average Marginal Effects on the probability of being above optimal BMI cutoff (BMI>=27.3 for females and BMI>=25.4 for males) from variations of the probit using the full sample

VARIABLES	(1)	(2)	(3)
Food Choice			
High-processed share	0.0886***	0.0123	0.150
MAFH share	0.0260	0.0230	0.0319
Carbs/AE	0.000814	0.00171	0.00137
Fat/AE	0.0281**	0.00565	0.00200
Protein/AE	-0.0297*	-0.0209*	-0.0130
Activity Choice			
Hours spent in vigorous activity	-0.0529**	-0.00740	-0.0163
Owns a motor vehicle (0/1)	0.0459***	0.00895	0.0137
Agricultural work (0/1)	-0.0231***	-0.00862	-0.0117
Office work (0/1)	-0.00894	0.0138	0.0118
Number above BMI cutoff	2,897	2,897	2,883
Observations	17,453	17,453	17,453
Prevalence	16.60%	16.60%	16.52%

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.4.5.5 Controlling for endogeneity and heteroskedasticity[24]

We have three models, one "naïve" probit that does not account for any unobserved factors, and two that attempt to account for unobserved factors. Between the latter two models, one allows for heteroskedasticity and in the other we instrument for endogeneity. This final approach allows us to account for not only unobserved time-invariant heterogeneity, but also endogeneity while allowing the variance to vary. We augment our correlated random effects model using control functions to account for endogeneity as outlined in Lin and Wooldridge (2019). We include residuals from the first step as regressors in the heteroskedastic probit, as well as in the variance function. Just as in our JMLE approach, we use a Tobit for the first step of each endogenous variable. Due to the nature of the Tobit model, we cannot use simple residuals as in the linear case,

---

[24] Results for categories other than food and activity choice are available upon request from the authors.

so we use generalized Tobit residuals given in Gourieroux et al. (1987). Our index function is given in equation 12,

$$y_{it}^* = \Theta'^{O_{it}} + \Delta'^{D_{it}} + \Gamma'^{A_{it}} + \Lambda'^{H_{it}} + \Upsilon'^{W_{it}} + c_i + \delta R + \varepsilon_{it} \qquad 12$$

where all variables are as previously defined, and R is the vector of generalized residuals from the first step Tobit. To generate standard errors, we bootstrap the results using 1000 replications, including the estimation of both steps in each replication.

We would prefer to make this our main model; however, we must simplify our specification in order to achieve a sufficient number of successful bootstrap replications for the urban and rural sub-samples. This means reducing our variance function given in equation 6 in the methods section to:

$$Var(c_i|S_i, S_i'X_i) = \exp\left(\delta R + \sum_{r=2}^{T} \psi_r \, 1[T_i = r] \right) \qquad 13$$

Using this simplified form of the variance function, the significance of the heteroskedasticity is much less than our main heteroskedastic probit model, so we cannot justify using this as one of our main specifications. Nonetheless, the results given in Table 17 are useful as a point of comparison to our main results.

Table 17: Average Marginal Effects on the probability of being overweight or obese (BMI>=25) from heteroskedastic probit with CRE using control functions

VARIABLES	Full	Urban	Rural
Food Choice			
High-processed share	0.211	0.501	0.219
MAFH share	0.0417	-0.0511	0.0987*
Carbs/AE	0.00176	-0.00318	0.00408*
Fat/AE	-0.000829	0.000558	-0.0143
Protein/AE	-0.00909	-0.00313	0.00207
Activity Choice			
Hours spent in vigorous activity	-0.00142	0.197**	-0.100
Owns a motor vehicle (0/1)	0.0199	0.0449*	0.0167
Agricultural work (0/1)	-0.0133	-0.0681**	0.0153
Office work (0/1)	0.0104	0.00709	-0.00312
Number above BMI cutoff	4,093	2,113	1,980
Observations	17,453	5,663	11,790
Prevalence	23.45%	37.31%	16.79%

*** p<0.01, ** p<0.05, * p<0.1

We see that the effect from HPF is no longer significant, however MAFH has a marginally significant impact in rural areas. The most puzzling result is the large and positive effect of activity in urban areas. One potential explanation may stem from the small number of agricultural households in urban areas. Those households may generate more income than poor households that do not engage in agricultural activities, which could result in increased food expenditures, which are more processed in urban areas on average. This might cause stark differences in weight outcomes between households that engage in physically demanding labor versus those that don't. Our instruments are also related to income generated through agricultural activities, which could be the reason we only see this when attempting to control for endogeneity. It is worth noting we get a positive effect in the JMLE in the urban sub-sample, but it is not significant.

1.4.5.6 Ordered Probit

In table 18 we explore the results from variations of the ordered probit using the full sample. There are a couple main points. First, the evidence for processed shares having an impact is very limited,

as the coefficient is only significant in the naïve probit. Second, MAFH is the only diet or activity

variable that is significant in all three regressions. This is evidence in favor of diet playing a large

role in body-mass outcomes.

Table 18: Average Marginal Effects on the probability of being in a specific BMI range from variations of the ordered probit using the full sample

VARIABLES	Underweight	Normal	Overweight	Obese
Ordered Probit				
Food Choice				
High-processed share	-0.0359*	-0.0217*	0.0321*	0.0255*
MAFH share	-0.0353***	-0.0214***	0.0316***	0.0251***
Carbs/AE	-0.000586	-0.000354	0.000524	0.000416
Fat/AE	-0.0225***	-0.0136***	0.0201***	0.0160***
Protein/AE	0.0108	0.00653	-0.00966	-0.00768
Activity Choice				
Hours spent in vigorous activity	0.0254***	0.0154**	-0.0227***	-0.0181***
Owns a motor vehicle (0/1)	-0.0291***	-0.0254***	0.0293***	0.0252***
Agricultural work (0/1)	0.00783	0.00485	-0.00710	-0.00558
Office work (0/1)	0.00127	0.000757	-0.00113	-0.000897
Heteroskedastic Ordered Probit with CRE				
Food Choice				
High-processed share	-0.00828	-0.00370	0.00725	0.00472
MAFH share	-0.0283***	-0.0126***	0.0248***	0.0161***
Carbs/AE	-0.00123	-0.000548	0.00107	0.000700
Fat/AE	-0.0123*	-0.00552*	0.0108*	0.00704*
Protein/AE	0.0154**	0.00690**	-0.0135**	-0.00881**
Activity Choice				
Hours spent in vigorous activity	0.00747	0.00334	-0.00655	-0.00427
Owns a motor vehicle (0/1)	-0.0179**	-0.0107**	0.0171**	0.0115**
Agricultural work (0/1)	0.00333	0.00152	-0.00294	-0.00191
Office work (0/1)	-0.000204	-9.16e-05	0.000974	0.000117
JMLE Ordered Probit with CRE				
Food Choice				
High-processed share	-0.0658	-0.0386	0.0582	0.0462
MAFH share	-0.0551**	-0.0323**	0.0488**	0.0387**
Carbs/AE	-0.00159*	-0.000930*	0.00140*	0.00111*
Fat/AE	-0.00958	-0.00562	0.00848	0.00672
Protein/AE	0.0105	0.00613	-0.00925	-0.00733
Activity Choice				
Hours spent in vigorous activity	-0.0298	-0.0175	0.0263	0.0209
Owns a motor vehicle (0/1)	-0.0178***	-0.0131**	0.0169**	0.0140**
Agricultural work (0/1)	0.0117	0.00713	-0.0106	-0.00826
Office work (0/1)	-0.00435	-0.00269	0.00390	0.00313
Number of overweight or obese	4,093	4,093	4,093	4,093
Observations	17,453	17,453	17,453	17,453
Prevalence	23.45%	23.45%	23.45%	23.45%

*** p<0.01, ** p<0.05, * p<0.1
Note: Underweight: BMI<18.5; Normal: 18.5<=BMI<25; Overweight: 25<=BMI<30; Obese: BMI>=30

We can see in interesting pattern. In all cases, the effects from relevant regressors change in direction going from normal to overweight. Specifically, if we consider the marginal impact from high-processed food shares in the base probit, we can see a negative effect for both underweight and normal, while a positive effect on both overweight and obese. This pattern is consistent, and generally the sizes of the effects are comparable when comparing marginal effects on different weight categories.

1.4.6   Policy Implications

Our results suggest diet has a bigger impact than activity on weight outcomes, which presents an avenue for policy makers to induce healthier diet choices. Such policies may target different aspects of both the food environments and individual or household preferences. Adom et al. (2021) review numerous policies employed in Africa that have targeted different levels of the food environment, which include both the macro and micro level. Policies aimed at the micro level are crafted to induce changes in the family, community, or school, which may include information interventions, financial inducements towards healthy diet choices, or programs that encourage exercise where it may be lacking. Interventions at the macro level target production, availability, and access to certain foods, which might include tariffs or taxes, regulations, or incentive programs to induce producers to make healthier food.

The effectiveness of various policy choices is often dependent on the population to which they will be applied. For instance, let us consider inducement via taxation. Such a policy typically increases the cost of unhealthy foods via an excise tax, much like that used for cigarettes or tobacco. This strategy was attempted in Hungary, where a junk food tax went into effect in 2011,

and was found to reduce the consumption of unhealthy junk foods (Bíró, 2015). The authors found a slight increase in the consumption of unprocessed foods; however, the estimates were not significant. In sub-Saharan Africa, tariffs on highly processed foods had a similar impact, yet there were potential downsides. Boysen et al. (2019) found higher tariffs were associated with reductions in the consumption of highly processed foods, as well as a reduction in obesity rates. On the other hand, the authors found there was a measurable increase in underweight, likely due to the most food insecure being able to afford less calorie dense food like UPF. We found similar results in our ordered probit analysis in table 15, where MAFH was found to reduce the likelihood of being underweight. This poses a serious challenge for policy makers, and a need for numerous studies to understand the complexities of diet, nutrition, health, and food security.

Policy decisions are further complicated by the nature of the human mind. Just and Payne (2009) point out that food decisions are not always rational in terms on long-term health, but more often hedonic, seeking immediate gratification. This means that marketing from UPF manufacturers can be extremely effective at encouraging UPF consumption, and in part can explain why macro level policies tend to have a limited impact when they are successful. Perhaps a more effective strategy might be using various information interventions and micro level inducements to encourage healthier habits rather that punish unhealthy choices. This is indeed suggested in a recent paper discussing interventions for healthy adolescent growth. Hargreaves et al. (2021) discuss various methods targeting individuals when they are adolescents, before they develop deeply rooted habits, which include using various social channels to promote nutrition education as well as increasing access and availability of healthy foods. They suggest it is most effective if the interventions are led by members in community rather than external actors.

1.5   Conclusion

As Tanzania's economy grows, access to processed and ultra-processed food will increase, and the risk of obesity will move in tandem. So, understanding what drives overweight/obesity is imperative for policy makers. In this paper, we estimate the relationship between processed food and overweight and obesity, while also estimating the effect from many other known drivers.

Using panel data from the National Panel Survey, we leveraged various econometric techniques to account for unobserved heterogeneity, heteroscedasticity and endogeneity and found our results were relatively robust. We find processed food consumption and meals eaten away from home are positively related to overweight/obesity in urban areas, while we only found limited evidence of an effect from MAFH in rural areas. High-processed food and meals eaten outside of the home have the largest impact on the likelihood of being overweight or obese. We also find that activity level matters in both rural and urban areas, but the effect is twice as large in urban areas. Proxies for income and wealth have a strong and positive effect on the likelihood of being overweight or obese. And while household composition does not have a statistically significant effect, gender is the most significant individual characteristic in determining risk for obesity, as females are more likely to be overweight or obese compared to male counterparts, all else equal.

This provides key insights for future researchers and policy makers. First, the survey was not designed to study an issue as complex as the focus of this paper, yet we were able to elicit relevant and meaningful information. Second, drivers for overweight/obesity can differ greatly between rural and urban areas, so policies need to be geographically tailored. Third, data collection and generation should be improved to better understand the role diet composition plays in determining health outcomes. Much of the variation of actual consumption is lost in using the

broad groups used by the NPS, limiting the precision of our estimates. Also, collecting consumption data at the household level rather than the individual level erases individual level variation and leaves much room to omit consumption outside of the house if not all household members are present at the time of the survey.

Possible areas of future research include designing a survey to capture more detailed consumption patterns that include more food items to allow for a more precise processing categorization. It is also important these surveys be done at the individual level so one can establish a definitive association between individual habits and nutrition-related health outcomes. The feasibility of such a survey is much greater due to the widespread adoption of smart phone technologies. There is a Household Budget survey that does include a detailed food diary over the course of a month at the household level in Tanzania, but it does not contain anthropometric data, making any weight related inference impossible. Another area for research is better understanding daily energy expenditures through more detailed surveys, as the proxies in our analysis do not capture all physically demanding activities in which an individual might engage. Lastly, processed food is not all equally obesogenic, and may have different health effects depending on the brand or origin. So, capturing a more detailed understanding of the processed food available in a given food environment can complement any analysis searching for relationships between diet and health. This may require field support, however, respondents with smartphones can capture brand names, nutrition labels, and countries of origin. Understanding what processed food is available, and whether or not certain types are more obesogenic will aid interested parties in developing long term solutions.

REFERENCES

Abarca-Gómez, L., Z. A. Abdeen, Z. A. Hamid, N. M. Abu-Rmeileh, B. Acosta-Cazares, C. Acuin, … M. Ezzati. 'Worldwide Trends in Body-Mass Index, Underweight, Overweight, and Obesity from 1975 to 2016: A Pooled Analysis of 2416 Population-Based Measurement Studies in 128·9 Million Children, Adolescents, and Adults'. *The Lancet* 390,10113(2017):2627–2642.

Adebo, O. A., T. Molelekoa, R. Makhuvele, J. A. Adebiyi, A. B. Oyedeji, S. Gbashi, … P. B. Njobeh. 'A Review on Novel Non-Thermal Food Processing Techniques for Mycotoxin Reduction'. *International Journal of Food Science & Technology* 56,1(January 2021):13–27.

Ademola, O., N. Saha Turna, L. S. O. Liverpool-Tasie, A. Obadina, and F. Wu. 'Mycotoxin Reduction through Lactic Acid Fermentation: Evidence from Commercial Ogi Processors in Southwest Nigeria'. *Food Control* 121(March 2021):107620.

Adom, T., A. De Villiers, T. Puoane, and A. P. Kengne. 'A Scoping Review of Policies Related to the Prevention and Control of Overweight and Obesity in Africa'. *Nutrients* 13,11(November 2021):4028.

Afzal, S., A. Tybjærg-Hansen, G. B. Jensen, and B. G. Nordestgaard. 'Change in Body Mass Index Associated With Lowest Mortality in Denmark, 1976-2013'. *JAMA* 315,18(May 2016):1989–1996.

Anderson, T. W., and H. Rubin. 'Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations'. *Ann. Math. Statist.* 20,1(1949):46–63.

Askari, M., J. Heshmati, H. Shahinfar, N. Tripathi, and E. Daneshzad. 'Ultra-processed Food and the Risk of Overweight and Obesity: A Systematic Review and Meta-Analysis of Observational Studies'. *International Journal of Obesity* 44,10(October 2020):2080–2091.

Bannor, R. K., B. Amfo, H. Oppong-Kyeremeh, and S. K. Chaa Kyire. 'Choice of Supermarkets as a Marketing Outlet for Purchasing Fresh Agricultural Products in Urban Ghana'. *Nankai Business Review International* 13,4(October 2022):545–566.

Bhaskaran, K., I. Dos-Santos-Silva, D. A. Leon, I. J. Douglas, and L. Smeeth. 'Association of BMI with Overall and Cause-Specific Mortality: A Population-Based Cohort Study of 3.6 million Adults in the UK'. *The Lancet Diabetes & Endocrinology* 6,12(December 2018):944–953.

Bíró, A. 'Did the Junk Food Tax Make the Hungarians Eat Healthier?' *Food Policy* 54(July 2015):107–115.

Brown, A. G. M., L. E. Esposito, R. A. Fisher, H. L. Nicastro, D. C. Tabor, and J. R. Walker. 'Food Insecurity and Obesity: Research Gaps, Opportunities, and Challenges'. *Translational Behavioral Medicine* 9,5(October 2019):980–987.

Boysen, O., K. Boysen-Urban, H. Bradford, and J. Balié. 'Taxing Highly Processed Foods: What Could be the Impacts on Obesity and Underweight in Sub-Saharan Africa?' *World Development* 119(July 2019):55–67.

Bullerman, L. B. 'Significance of Mycotoxins to Food Safety and Human Health1,2'. *Journal of Food Protection* 42,1(January 1979):65–86.

Casey, A. A., M. Elliott, K. Glanz, D. Haire-Joshu, S. L. Lovegreen, B. E. Saelens, … R. C. Brownson. 'Impact of the Food Environment and Physical Activity Environment on Behaviors and Weight Status in Rural U.S. Communities'. *Preventive Medicine* 47,6(December 2008):600–604.

Chamberlain, G. 'Analysis of Covariance with Qualitative Data'. *The Review of Economic Studies* 47,1(January 1980):225–238.

Cragg, J. G., and S. G. Donald. 'Testing Identifiability and Specification in Instrumental Variable Models'. *Econometric Theory* 9,2(April 1993):222–240.

Debela, B. L., K. M. Demmler, S. Klasen, and M. Qaim. 'Supermarket Food Purchases and Child Nutrition in Kenya'. *Global Food Security* 25(June 2020):100341.

Demmler, K. M., O. Ecker, and M. Qaim. 'Supermarket Shopping and Nutritional Outcomes: A Panel Data Analysis for Urban Kenya'. *World Development* 102(February 2018):292–303.

Elizabeth, L., P. Machado, M. Zinöcker, P. Baker, and M. Lawrence. 'Ultra-Processed Foods and Health Outcomes: A Narrative Review'. *Nutrients* 12,7(June 2020):1955.

Fanzo, J., L. Haddad, R. McLaren, Q. Marshall, C. Davis, A. Herforth, … A. Kapuria. 'The Food Systems Dashboard is a New Tool to Inform Better Food Policy'. *Nature Food* 1,5(May 2020):243–246.

Fiedler, J. L. 'Towards Overcoming the Food Consumption Information Gap: Strengthening Household Consumption and Expenditures Surveys for Food and Nutrition Policymaking'. *Global Food Security* 2,1(March 2013):56–63.

Fiedler, J. L., C. Carletto, and O. Dupriez. 'Still Waiting for Godot? Improving Household Consumption and Expenditures Surveys (HCES) to Enable More Evidence-Based Nutrition Policies'. *Food and Nutrition Bulletin* 33,3_suppl2(September 2012):S242–S251.

Fiedler, J. L., K. Lividini, O. I. Bermudez, and M.-F. Smitz. 'Household Consumption and Expenditures Surveys (HCES): A Primer for Food and Nutrition Analysts in Low- and Middle-Income Countries'. *Food and Nutrition Bulletin* 33,3_suppl2(September 2012):S170–S184.

Finlay, K., L. Magnusson, and M. E. Schaffer. 'WEAKIV: Stata Module to Perform Weak-Instrument-Robust Tests and Confidence Intervals for Instrumental-Variable (IV) Estimation of Linear, Probit and Tobit Models'. Boston College Department of Economics. 2013.

51

Flegal, K. M., B. K. Kit, H. Orpana, and B. I. Graubard. 'Association of All-Cause Mortality with Overweight and Obesity Using Standard Body Mass Index Categories: A Systematic Review and Meta-analysis'. *JAMA* 309,1(January 2013):71–82.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon. 'Generalised Residuals'. *Journal of Econometrics* 34,1–2(January 1987):5–32.

Government of Tanzania. 'National Panel Survey'. National Bureau of Statistics. 2015.

Hall, K. D., A. Ayuketah, R. Brychta, H. Cai, T. Cassimatis, K. Y. Chen, … M. Zhou. 'Ultra-Processed Diets Cause Excess Calorie Intake and Weight Gain: An Inpatient Randomized Controlled Trial of Ad Libitum Food Intake'. *Cell Metabolism* 30,1(July 2019):67-77.e3.

Hargreaves, D., E. Mates, P. Menon, H. Alderman, D. Devakumar, W. Fawzi, … G. C. Patton. 'Strategies and Interventions for Healthy Adolescent Growth, Nutrition, and Development'. *The Lancet* 399,10320(January 2022):198–210.

Holdsworth, M., and E. Landais. 'Urban Food Environments in Africa: Implications for Policy and Research'. *Proceedings of the Nutrition Society* 78,4(November 2019):513–525.

Joint, F. *Human energy requirements. Report of a Joint FAO/WHO/UNU Expert Consultation*. Rome: FAO/WHO/UNU. 2001.

Just, D. R., and C. R. Payne. 'Obesity: Can Behavioral Economics Help?' *Annals of Behavioral Medicine* 38,S1(December 2009):47–55.

Khonje, M. G., O. Ecker, and M. Qaim. 'Effects of Modern Food Retailers on Adult and Child Diets and Nutrition'. *Nutrients* 12,6(June 2020):1714.

Khonje, M. G., and M. Qaim. 'Modernization of African Food Retailing and (Un)healthy Food Consumption'. *Sustainability* 11,16(August 2019):4306.

Kolodinsky, J. M., G. Battista, E. Roche, B. H. Y. Lee, and R. K. Johnson. 'Estimating the Effect of Mobility and Food Choice on Obesity in a Rural, Northern Environment'. *Journal of Transport Geography* 61(May 2017):30–39.

Larsen, A. F., D. Headey, and J. Heckert. *Changes in Women's Body Mass, Overweight and Obesity Status in Tanzania: A Quantitative Analysis from the 2008/09 and 2014/15 National Panel Survey* (pp. 65–82). IFPRI. 2018.

Lin, W., and J. M. Wooldridge. 'Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models'. In *Panel Data Econometrics* (pp. 21–43). Elsevier. 2019.

Ludwig, D. S., L. J. Aronne, A. Astrup, R. de Cabo, L. C. Cantley, M. I. Friedman, … C. B. Ebbeling. 'The Carbohydrate-Insulin Model: A Physiological Perspective on the Obesity Pandemic'. *The American Journal of Clinical Nutrition* 114,6(December 2021):1873–1885.

Ludwig, D. S., S. L. Dickinson, B. Henschel, C. B. Ebbeling, and D. B. Allison. 'Do Lower-Carbohydrate Diets Increase Total Energy Expenditure? An Updated and Reanalyzed Meta-Analysis of 29 Controlled-Feeding Studies'. *The Journal of Nutrition* 151,3(March 2021):482–490.

Lukmanji, Z., E. Hertzmark, N. Mlingi, V. Assey, G. Ndossi, and W. Fawzi. *Tanzania Food Composition Tables*. Muhimbili University of Health and Allied Sciences, Tanzania Food and Nutrition Centre, and Harvard School of Public Health. 2008.

Lustig, R. H. 'Childhood Obesity: Behavioral Aberration or Biochemical Drive? Reinterpreting the First Law of Thermodynamics'. *Nature Clinical Practice Endocrinology & Metabolism* 2,8(August 2006):447–458.

Marrez, D. A., and A. M. Ayesh. 'Mycotoxins: The Threat to Food Safety'. *Egyptian Journal of Chemistry* 0,0(July 2021):0–0.

Mensah, D. O., and O. Oyebode. '"We Think about the Quantity More": Factors Influencing Emerging Adults' Food Outlet Choice In A University Food Environment, A Qualitative Enquiry'. *Nutrition Journal* 21,1(December 2022):49.

Monteiro, C. A., J.-C. Moubarac, R. B. Levy, D. S. Canella, M. L. da C. Louzada, and G. Cannon. 'Household Availability of Ultra-processed Foods and Obesity in Nineteen European Countries'. *Public Health Nutrition* 21,1(January 2018):18–26.

Mundlak, Y. 'On the Pooling of Time Series and Cross Section Data'. *Econométrica* 46, 1 (January 1978):69–85.

Neven, D., T. Reardon, J. Chege, and H. Wang. 'Supermarkets and Consumers in Africa: The Case of Nairobi, Kenya'. *Journal of International Food & Agribusiness Marketing* 18,1–2(July 2006):103–123.

Pagliai, G., M. Dinu, M. P. Madarena, M. Bonaccio, L. Iacoviello, and F. Sofi. 'Consumption of Ultra-processed Foods and Health Status: A Systematic Review and Meta-analysis'. *British Journal of Nutrition* 125,3(February 2021):308–318.

Pingali, P. 'Westernization of Asian Diets and the Transformation of Food Systems: Implications for Research and Policy'. *Food Policy* 32,3(June 2007):281–298.

Popkin, B. M., L. S. Adair, and S. W. Ng. 'Global Nutrition Transition and the Pandemic of Obesity in Developing Countries'. *Nutrition Reviews* 70,1(January 2012):3–21.

Poti, J. M., B. Braga, and B. Qin. 'Ultra-processed Food Intake and Obesity: What Really Matters for Health-Processing or Nutrient Content?' *Current Obesity Reports* 6,4(2017):420–431.

Qaim, M. 'Globalisation of Agrifood Systems and Sustainable Nutrition'. *Proceedings of the Nutrition Society* 76,1(February 2017):12–21.

Reardon, T., D. Tschirley, L. S. O. Liverpool-Tasie, T. Awokuse, J. Fanzo, B. Minten, … B. M. Popkin. 'The Processed Food Revolution in African Food Systems and the Double Burden of Malnutrition'. *Global Food Security* 28(March 2021):100466.

Roodman, D. 'CMP: Stata Module to Implement Conditional (Recursive) Mixed Process Estimator'. Boston College Department of Economics. 2007.

Sarfo, J., E. Pawelzik, and G. B. Keding. 'Dietary Patterns as Characterized by Food Processing Levels and Their Association with the Health Outcomes of Rural Women in East Africa'. *Nutrients* 13,8(August 2021):2866.

Stock, J., and M. Yogo. 'Testing for Weak Instruments in Linear IV Regression'. *Identification and Inference for Econometric Models* (2005):53.

Swinburn, B. A., G. Sacks, K. D. Hall, K. McPherson, D. T. Finegood, M. L. Moodie, and S. L. Gortmaker. 'The Global Obesity Pandemic: Shaped by Global Drivers and Local Environments'. *The Lancet* 378,9793(August 2011):804–814.

Teufel, F., J. A. Seiglie, P. Geldsetzer, M. Theilmann, M. E. Marcus, C. Ebert, … J. Manne-Goehler. 'Body-mass Index and Diabetes Risk in 57 Low-income and Middle-income Countries: A Cross-Sectional Study of Nationally Representative, Individual-Level Data in 685 616 Adults'. *The Lancet* 398,10296(July 2021):238–248.

Tobias, D. K., and F. B. Hu. 'The Association Between BMI and Mortality: Implications for Obesity Prevention'. *The Lancet Diabetes & Endocrinology* 6,12(December 2018):916–917.

Tschirley, D. L., J. Snyder, M. Dolislager, T. Reardon, S. Haggblade, J. Goeb, … F. Meyer. 'Africa's Unfolding Diet Transformation: Implications for Agrifood System Employment'. *Journal of Agribusiness in Developing and Emerging Economies* 5,2(November 2015):102–136.

Tschirley, D., T. Reardon, M. Dolislager, and J. Snyder. 'The Rise of a Middle Class in East and Southern Africa: Implications for Food System Transformation'. *Journal of International Development* 27,5(July 2015):628–646.

Tschirley, D., J. Snyder, C. Ijumba, and M. Kondo. 'Employment Intensity and Scale of Operation in Agro-processing: A Case of Cereal Millers in Tanzania.' (2017).

Wells, J. C. K., and M. Siervo. 'Obesity and Energy Balance: Is the Tail Wagging the Dog?' *European Journal of Clinical Nutrition* 65,11(November 2011):1173–1189.

Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. MIT press. 2010.

Wooldridge, J. M. 'Control Function Methods in Applied Econometrics'. *Journal of Human Resources* 50,2(2015):420–445.

Wooldridge, J. M. 'Correlated Random Effects Models with Unbalanced Panels'. *Journal of Econometrics* 211,1(July 2019):137–150.

‘World Bank Development Indicators’. n.d.

Tables

Table A1: Simulation on the impact of diet and activity on the probability of being obese (BMI>=30)

		Variable: 50th to 75th Pct.	at 50th Pct. (%)	at 75th Pct. (%)	Change in Probability
-	-				
Full	Probit	Processing & MAFH	7.28	7.78	0.51
		Active Hours	7.70	6.72	-0.98
	Het Probit CRE	Processing & MAFH	7.42	7.46	0.04
		Active Hours	7.52	7.22	-0.31
	Probit JMLE	Processing & MAFH	6.79	8.27	1.48
		Active Hours	7.78	7.00	-0.78
Urban	Probit	Processing & MAFH	14.32	15.60	1.28
		Active Hours	14.95	14.83	-0.12
	Het Probit CRE	Processing & MAFH	14.58	14.71	0.13
		Active Hours	14.42	14.44	0.02
	Probit JMLE	Processing & MAFH	16.15	19.29	3.14
		Active Hours	15.81	15.82	0.00
Rural	Probit	Processing & MAFH	4.37	4.44	0.07
		Active Hours	4.27	3.87	-0.39
	Het Probit CRE	Processing & MAFH	4.52	4.39	-0.13
		Active Hours	4.38	4.20	-0.19
	Probit JMLE	Processing & MAFH	4.17	4.65	0.48
		Active Hours	4.32	3.74	-0.58

Note: Processing changes is in high processed share (Full sample: 0.21 to 0.28; urban subsample: 0.27 to 0.30; rural subsample: 0.16 to 0.24) and MAFH share (Full sample: 0.02 to 0.10; urban subsample: 0.10 to 0.25; rural subsample: 0.02 to 0.05). Activity change is of active hours ((Full sample: 7 to 30 hours per week; urban subsample: 0 to 1 hours per week; rural subsample: 18 to 36 hours per week)

CHAPTER 2: <u>**The Relationship between Ultra-Processed Food Imports and Nutrition-related Outcomes**</u>

2.1   Introduction

The rise of obesity across the developing world has driven researchers to explore potential mechanisms responsible for the rapid increase, resulting in the birth of an entire strand of literature examining the dangers of the obesity pandemic and how to stem its rise (Abarca-Gómez et al., 2017; Swinburn et al., 2011; Meldrum et al., 2017). Much of the evidence to date suggests the surging obesity in developing countries is partly driven by rising imports of unhealthy ultra-processed foods (UPF)[25] often related to various factors associated with increasing globalization (Popkin et al., 2012; McNamara, 2017; Cuevas Garcia-Dorado et al., 2019; An et al., 2020). UPF have also been shown to be related to increases in body-mass (Hall et al., 2019; Canella et al., 2014; Giuntella et al., 2020; Lin et al., 2018; Louzada et al., 2015b; Rauber et al., 2020), type-2 diabetes (Srour et al., 2020; Levy et al., 2021), hypertension (De Deus Mendoca et al., 2017), and others health issues including cancer and depression (Lane et al., 2021). It is evident, that diets with higher proportions of UPF are associated with poor health outcomes related to excessive weight gain, however, the mechanism is not clear (Poti et al., 2017).

---

[25]We categorize processed food using the NOVA classification system (Monteiro et al., 2019). This uses four main categories: 1) Unprocessed and minimally processed, 2) Culinary Ingredients, 3) Processed Food, 4) Ultra-processed Food.

Demand for ultra-processed foods (UPF) has risen dramatically in recent years[26]. These increases are evident in both SSA (Reardon et al., 2021: Tschirley et. al., 2015), and Asia (Baker and Friel, 2016; Pingali, 2017; Sievert et al. 2019). The rise in demand for UPF in developing countries typically outpaces the ability of local food systems to respond, leading to higher UPF imports (Ravuvu et al., 2017; Friel et al., 2013; Munthali et al., 2021)[27]. For example, in the Southern African Development Community, imports of sugary beverages and processed snack foods from outside the region increased by 1,200% (from 1,408 to 16,713 metric tons) and 750% (2,467 to 18,522 metric tons), respectively, between 1995 and 2010 (Thow et al., 2015). The increasing demand is expected to continue to outpace domestic production[28], which implies UPF imports in developing countries will continue to rise in the near term (Zhou and Staatz, 2016).

UPF is usually imported from large transnational food companies, who employ "coordinated and comprehensive" marketing campaigns (Hawkes, 2006), which may shift diet preferences (Kremer et al., 2019), which would shift demand through food-related social channels (Oberlander et al., 2017). In developing countries, this necessitates an increase in imports, since the domestic capacity to produce such foods is usually limited. In other words, the process of importing UPF, creates a positive feedback loop, which we do not observe directly. However, we

---

[26] The ultra-processed foods category contains, "…carbonated soft drinks; sweet, fatty or salty packaged snacks; candies (confectionery); mass produced packaged breads and buns, cookies (biscuits), pastries, cakes and cake mixes; margarine and other spreads; sweetened breakfast 'cereals' and fruit yoghurt and 'energy' drinks; pre-prepared meat, cheese, pasta and pizza dishes; poultry and fish 'nuggets' and 'sticks'; sausages, burgers, hot dogs and other reconstituted meat products; powdered and packaged 'instant' soups, noodles and desserts; baby formula; and many other types of products…" (Monteiro et al., 2019).

[27]

[28] In part, the increasing demand for UPF is driven by the modernization of the agrifood systems of these countries, which brings many benefits, such as lower transportation costs or improved food access and safety (Qaim, 2017; Reardon et al., 2021).

do observe changes in body-mass, which serves as a proxy for the impact that exposure to UPF and its marketing has on food preferences and norms.

This paper seeks to understand the relationship between nutrition-related outcomes (NROs) and UPF imports, and whether there is a mechanism through which increases in UPF imports can create a feedback loop that leads to rapid increases in negative nutrition-related outcomes. There is limited literature that discusses such pathways, and no empirical investigations for the existence of such multi-directional impacts whether direct or indirect[29]. Understanding this relationship, and whether it matters, is critical for policy makers and organizations committed to mitigating the effects from the obesity epidemic. If this feedback loop exists and has a measurable impact on trade, then it creates a mechanism through which UPF imports and worsening NROs can feed into one another. Interventions ignorant of this mechanism will have limited success in stemming the rapid rise in obesity.

The rest of the paper is organized as follows. In the next section we discuss our model, and the data used in our analysis. We then discuss the results of our main specification, and highlight their robustness. Next, we discuss policy implications of our results, and end with concluding remarks.

---

[29] Drichoutis et al. (2009) explored a bi-directional relationship between food eaten away from home and BMI in a simultaneous model. However, they dropped BMI as an endogenous regressor in the FAFH equation in favor of 2SLS in their published version (Drichoutis et al., 2012).

2.2   Data and Methods

2.2.1   Conceptual Model

We model this relationship drawing from Woodard et al. (2001) and Hawkes (2006). Our model relates external and internal influences to food consumption decisions, which then change some food norms while reinforcing others, which lead to shifts in food demand and thus imports (Figure 3).

Figure 3: Relationship Between Food Imports and NROs
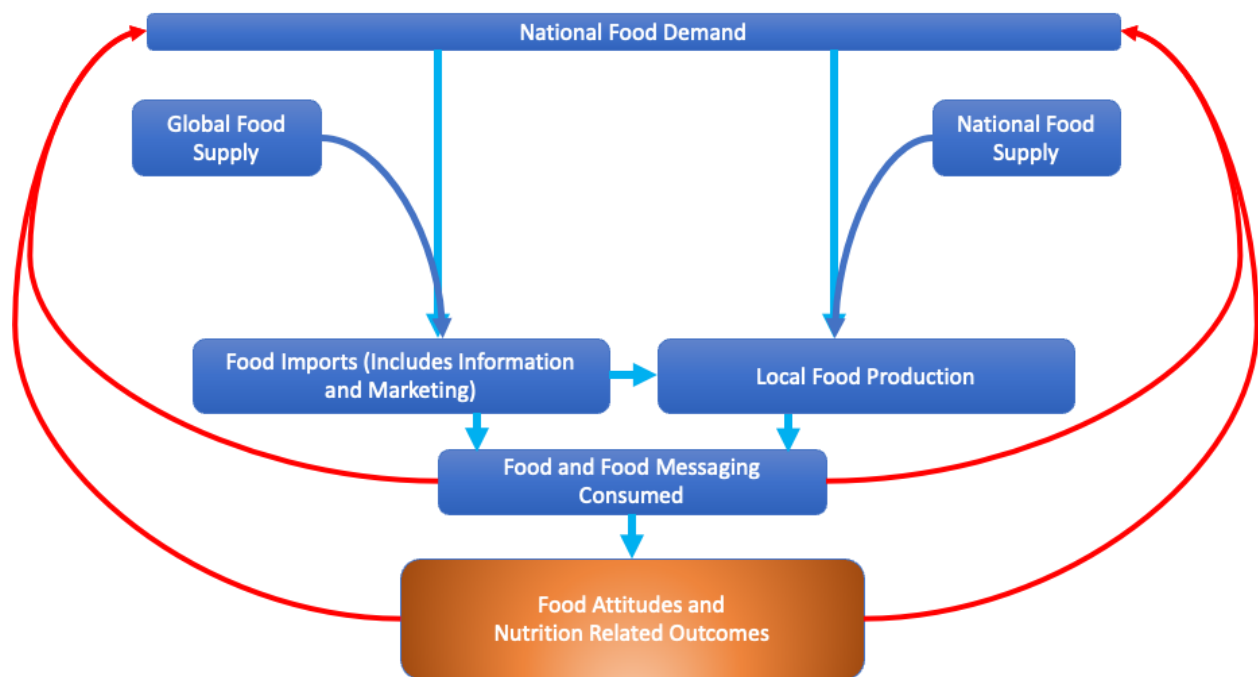
Consider first a developing country at the beginning of the nutrition transition. Food production is likely to be based on small scale farmers with a limited post-farm food processing

sector. Due to low purchasing power and limited infrastructure (both physical and virtual), multinational corporations (MNCs) may not find it profitable to export their goods to such a country, or invest in local productive means to produce their UPF domestically. Consequently, the population of this country may not have encountered sufficient UPF to develop strong preferences for it. National demand thus might be met by the minimally processed food produced through local farmers and processors. We might also expect overweight and obesity to be limited to the wealthiest household consuming enough calories to outpace daily caloric expenditures.

As the country develops, incomes will likely rise and a higher proportion of the population will move to urban centers, which should incentivize MNCs that produce UPF to do business locally. Increased abundance of UPF leads to increased exposure to these unhealthy foods[30] and their associated food messaging, which may target aspects such as palatability, cost, and convenience that are especially important to lower income households (Moran et al., 2019)[31]. In addition, psychological aspects such as addiction to UPF (Ifland et al. 2015), present bias, inability to delay gratification, or inaccurate time preferences may make some populations especially susceptible to UPF promotion (Epstein et al., 2010; Bickel et al., 2016; Stojek and MacKillop, 2017). Marketing campaigns aim to exploit these psychological factors and encourage cultural changes related to leisure time or convenience, shifting diets away from traditional foods to UPF and – in countries where the local food manufacturing sector cannot keep up – increasing the imports of those foods. Eventually MNCs will develop domestic production (Moodie et al., 2021),

---

[30] UPFs are typically characterized by higher amounts of simple carbohydrates such as sugars, saturated fats, and sodium, while containing micronutrient levels strongly associated with poor diet quality (Louzada et al., 2015a; Martinez Steele et al., 2017; Moubarac et al., 2017; Cornwell et al., 2018; Rauber et al., 2018).

[31] There are also food safety benefits to processing, such as reduction in mycotoxins in fruit products (see Pal et al., 2019) or maize (see Ademola et al., 2021).

however, such development takes time, and demand will likely be satisfied through importation when lower income populations are the most vulnerable to changes in nutrition related outcomes.[32]

Body norms and preferences in developing regions may offer limited resistance to the changing food norms, and may in fact be reinforced by the health impacts of increased processed food importation. Larger body sizes are preferred in regions emerging from widespread poverty and communicable diseases just as they were in the west in the late 19th to early 20th century, as larger body sizes were signals of disease absence (Renzaho, 2003). This preference is observed in much of Sub-Saharan Africa where many countries still face high disease prevalence (Okop et al., 2016; Macia et al., 2017; Naigaga et al., 2018; Christian and Frempong, 2020; Pradeilles et al., 2021). This is further documented as a consistent tendency of many across the region to underestimate their weight (Puoane et al., 2002; Muhihi et al., 2012; Tateyama et al., 2018). Education and wealth may lower this tendency (Alwan et al., 2010; Christian and Frempong, 2020) but in some cases it still occurs to a very high degree even among university students (Peltzer and Penpid, 2015).

Linkages in both directions imply a positive feedback loop where the increasing presence of UPF and its marketing drives changes in socio-cultural norms, leading to a shift in demand and potentially further increases in UPF imports. The presence of such a feedback loop matters for two reasons. First, methodologically this implies endogeneity through simultaneity, and ignoring it may bias estimation results, since both outcomes are related to omitted causal factors (Wooldridge, 2010). Second, policy tailored to solve health problems by targeting only one pathway may fall

---

[32] Once countries develop large scale food processing capabilities, they too become exporters of UPF. Mexico is a substantial exporter to less developed countries in Latin America (Popkin and Reardon, 2018).

short of the intended goal or produce unintended consequences that may negate any intended benefits.

Shifts in individual preferences related to food drive changes in national food demand, which feed into increased UPF-related imports. It is not only imports of processed foods that increase, but also foreign direct investment (FDI) from MNCs in the form of manufacturing plants to domestically produce their products. In addition, any ingredients unable to be sourced from local firms must be imported. So, we expect the changes in food norms to impact imports of these ingredients in addition to UPF. We can model this relationship as a dynamic system of equations,

$$N_{it} = f(T_{it-1}, N_{it-1}, X_{it}, Z_{N,it})$$

$$T_{it} = g(N_{it-1}, T_{it-1}, X_{it}, Z_{T,it})$$

Where N is an outcome related to nutrition (e.g., obesity rate), T is a measure of trade (e.g., import value of UPF), X is a vector of characteristics of a country that affect both nutrition-related outcomes (NROs) and trade, Z_N is a vector of factors related to NROs but not necessarily trade, and Z_T contains factors related to trade but not necessarily NROs. We use the first lags of trade and NROs as regressors for two reasons. First, both trade and NROs are likely persistent to some extent[33], and second, any changes in diet norms or trade are likely take time to show their impact on the other. In our model, NROs function as measures of the impact that exposure to UPF has had on a population, which can function as a proxy for the shift in societal and cultural norms

---

[33] Wilkins (2018) finds excluding lagged dependent variables in a dynamic model can lead to severe bias.

about food. For instance, two countries may be exposed to processed foods but have drastically different levels of obesity. The differences in processed food consumption and importation might be driven by the different norms regarding food or body type, which leads to very different diets. One can imagine two such countries might have measurably different UPF imports.

2.2.2 Data

We use three main sources of data for 151 countries: World Bank (WB), Centre for Prospective Studies and International Information (CEPII), and the Non-Communicable Diseases Risk Factor Collaboration (NCD-RiSC). In addition to variables that are related to NROs and trade, we must account for national characteristics and the local food environment. So, in our analysis we include variables that control for level of income, domestic food production, and various other controls that account for the general level of development and agricultural productivity. Due to data limitations, we chose select variables to include the maximum number of countries while still adequately controlling for different characteristics. This meant dropping variables we would have like to use in favor of increasing our sample size. The variables we chose and their definitions are displayed in table 19, with discussions below.

NRO Variables: We use various national measures from NCD-RiSC for our main dependent NRO variables. We focus on BMI, overweight/obesity (BMI>=25), and obesity (BMI>=30), and examine whether our hypothesized relationship can be shown between UPF and other NRO measures. These include underweight, high blood pressure, diabetes, and cholesterol. Measures such as cholesterol or overweight/obesity will likely see the quickest changes in response to UPF consumption, whereas high blood pressure and cholesterol levels can take longer to manifest in

measurable changes. So, we expect any relationships between the latter and UPF-related trade

will be difficult to measure and likely less strong than the former.

Table 19: Variable Definitions

Dependent NRO Variables (N)	
Overweight Obesity (Male)	Proportion of Male Population with BMI>=25
Overweight Obesity (Female)	Proportion of Female Population with BMI>=25
Obesity (Male)	Proportion of Male Population with BMI>=30
Obesity (Female)	Proportion of Female Population with BMI>=30
Underweight (Male)	Proportion of Male Population with BMI<=18.5
Underweight (Female)	Proportion of Female Population with BMI<=18.5
BMI(Male)	Average Male Body Mass Index (kg/m^2)
BMI(Female)	Average Female Body Mass Index (kg/m^2)
Dependent Trade Variables (T)	
Total PC Imports	Total PC imports in 2010 USD
Non-Food PC Imports	Non-Food PC imports in 2010 USD
Un/Min Processed PC Imports	Unprocessed and minimally processed PC imports in thousand 2010 USD
Culinary Ingredients PC Imports	Culinary Ingredients PC imports in thousand 2010 USD
Processed PC Imports	Non-UPF Processed Food PC imports in thousand 2010 USD
UPF PC Imports	Ultra-Processed Food PC imports in thousand 2010 USD
Excluded Trade and NRO Variables (Z)	
NRO Variables (Zn)	
Incidence of Anemia	Incidence of anemia per 1000
Death Rate	Death rate per 1000
Incidence of Tuberculosis	Incidence of tuberculosis per 1000
OOP Health Exp Ratio	Ratio of out-of-pocket health expenditures per capita to total health expenditure per capita
Trade Variables (Zt)	
Net FDI Inflows PC	Net Foreign Direct Investment PC in 2010 USD
Manufacturing Trade Flows PC	Total bi-lateral manufacturing trade flows PC in 2010 USD
Largest City (% of Urban Pop)	Largest City as a percent of total urban Population
Ag and Other Controls (X)	
Agricultural Variables	
Ag Land (% of Total)	Percent of total land dedicated to agricultural uses
Ag Employment (% of Total)	Percent of labor force employed in agriculture-related activities
Food Production Index	Relative level of the aggregate volume of agricultural production for each year in comparison with the base period 2014-2016.
Other Controls	
GDP PC	Gross Domestic Product PC in thousand 2010 USD
KOF Index	Measures the economic, social, and political dimensions of globalization
Electrification	Percent of population with access to electricity
% Rural	Percent of population living in rural areas
Mobile Subscriptions (per 100)	Number of mobile subscriptions per 100
% Female	Percent of population Female

We supplement our NRO equation with four explanatory variables from the World Bank

Development Indicators to approximate the health environments in each country. The first is

incidence of anemia, which is related to the general level of nutrition in a country. The second is the incidence of tuberculosis, which approximates the incidence of communicable diseases. The third is the ratio of out-of-pocket health expenditure to total health expenditure which we can think of as a measure of the quality of the health care system. A robust health care system is likely to feature better insurance coverage, which implies lower out-of-pocket expenses for the average consumer. So, we take a lower ratio to loosely imply a higher quality of health care for the average person. We would expect a lower ratio to be related to lower levels of these negative health outcomes (and potentially lower levels of NROs). The fourth is death rate which can approximate the health environment to some extent. While one might argue this could be used as another control variable for both equations, the correlation between death rate and UPF is very small. We display correlation between variables in table A2 in the appendix.

Trade Variables: For our main dependent trade variable, we focus on imports rather than bilateral trade flows, since our model focuses on NROs and diet norms within a country. We categorize imports by level of processing using the NOVA classification (Monteiro et al., 2019). To do this we use trade data aggregated to the 6-digit level of the harmonized system (HS) to allow for more accurate processing categorization. We use UN COMTRADE bilateral trade flow data maintained by CEPII[34]. The data are in thousand current US dollars per capita, so we deflate the value of imports by the US CPI. We also examine relationships with UPF measured as quantity imported (1000 metric tons per capita), share of total imports, and share of food imports.

---

[34] Gaulier and Zignago (2010)

We also include trade specific variables from the CEPII gravity database[35] as explanatory variables to control for external influences in our model. The first two control for some aspects of MNC investments in countries to which they may export their UPF. We use manufacturing trade flows and net FDI inflows, which usually originate from MNCs[36]. As we discussed in our conceptual model, UPF usually includes foreign investment from MNCs in both marketing their product and developing local production capabilities to reduce costs. We should capture some of this effect with these variables. In addition, we include population of the largest city as a percent of total urban population from world bank development indicators in the trade equation but not the NRO equation[37]. If a large portion of the urban population resides in the largest city, we might expect an increase UPF consumption by reducing costs of distribution and by demonstration effects due the high population density of such an urban center.

Agricultural Variables: Our analysis focuses on food imports and NROs, so we need to control for characteristics of the domestic food system. We do this by including three regressors that account for different aspects of agricultural development: percent of land area dedicated to agricultural production, share of labor force working in agriculture, and a food production index calculated by FAO. The percent of total land area in a country designated as agricultural land will depend on the type of agriculture and the level of sophistication of agricultural technology. Countries that rely on smaller subsistence farmers, may require more land to feed their populations.

---

[35] Conte et al., (2021)

[36] It is possible that FDI inflows may also come from governments or NGOs, however, we do not believe that matters for the sake of our analysis. Net FDI inflows are merely used a control used to better isolate the trade and NRO relationships.

[37] We originally included this variable in both equations, but it was not significant in the health equation. This is likely due in part to the low correlation with UPF as seen in table A2.

We might expect those same countries to haver lower incidence of overweight obesity and to import less food. In table A2 of the appendix, we present a correlation matrix of variables in log form. We can see clearly that the percent of agricultural land is negatively correlated with UPF imports per capita, GDP, and female overweight/obesity.

Agricultural systems in developed countries are typically capital intensive and have lower labor requirements, so we include the percent of the labor force employed in agriculture to approximate this aspect of domestic food production. Countries that produce high amounts of UPF require highly industrial manufacturing equipment, and so may be likely also to employ similarly advanced machinery in their agricultural production. We expect the percent employed in agriculture to be strongly related to level of development, which is clearly shown in table A2, where we can see a strong inverse relationship with female overweight, GDP, and UPF imports.

Lastly, we include the food production index calculated by FAO (WB Indicator) to control for growth of agricultural production. We expect this to generally increase, with higher growth for countries increasing rapidly in population and wealth, while remaining somewhat flat for the most developed countries. Food production index is positively correlated with NROs, GDP, and trade.

Other controls: Many of the variables we have already described can be argued to have a relationship with level of development. However, since we aim to identify a relationship between trade and NROs, we include more explanatory variables in both the NRO and trade equations that may adequately control for the various characteristics of a country's state of development. Just as above, we limit the number of variables to allow for a greater number of countries in the analysis. The variables we use include the KOF globalization index, GDP per capita (taken from penn world tables), percent of the population with access to electricity (electrification), rural population as a

percent of total population, mobile cellular subscriptions per 100, and percent of the population female.

Table 20: Variable Means

Dependent Variables			Explanatory Variables		
	Mean	SD		Mean	SD
Dependent NRO Variables (N)			**Trade and NRO Variables (Z)**		
Overweight Obesity (Male)	0.47	0.14	**NRO Variables (Zn)**		
Overweight Obesity (Female)	0.45	0.20	Incidence of Anemia	31.63	13.49
Obesity (Male)	0.19	0.10	Death Rate	8.28	3.22
Obesity (Female)	0.13	0.08	Incidence of Tuberculosis	117.22	144.34
Underweight (Male)	0.06	0.05	OOP Health Exp Ratio	0.37	0.19
Underweight (Female)	0.05	0.06	**Trade Variables (Zt)**		
BMI(Male)	25.44	2.11	Net FDI Inflows PC	6.24	24.17
BMI(Female)	24.99	2.18	Manufacturing Trade Flows PC	33.58	48.60
Dependent Trade Variables (T)			Largest City (% of Urban Pop)	31.27	15.10
Total PC Imports	3,944	5,616	**Ag and Other Controls (X)**		
Non-Food PC Imports	3,611	5,224	**Agricultural Variables**		
Un/Min Processed PC Imports	154.21	201.89	Ag Land (% of Total)	12.36	17.76
Culinary Ingredients PC Imports	47.00	62.33	Ag Employment (% of Total)	41.97	20.47
Processed PC Imports	27.83	43.02	Food Production Index	27.29	22.39
UPF PC Imports	103.89	147.81	**Other Controls**		
			GDP PC	16,337	18,271
			KOF Index	62.56	14.32
			Electrification	73.99	35.70
			% Rural	41.22	21.46
			Mobile Subscriptions (per 100)	78.50	45.52
			% Female	50.00	3.58

GDP per capita is a rough measure of wealth in a given country. Countries with greater capital and other resources will typically produce more, and thus have a higher GDP. Electrification should improve health (and potentially NROs) through avenues as varied as reduced use of firewood and charcoal in cooking to improved performance of hospitals and health clinics. It may increase trade through access to information and to the advertising of MNC food companies. Mobile subscriptions will likely increase with level of development, but they also offer a crude measure of the potential level of information sharing within a country. Mobile phone usage has been shown to increase information sharing, and we see it has a positive correlation with our main trade and NRO variables.

The KOF index was first created by Dreher, A. (2006) to measure a countries level of globalization by economic, social, and political dimensions. It has since been updated and uses 43 variables to index level of globalization (Gygli et. al., 2019), which is expected to have positive relationship with weight-related measures and level of trade. This is evident in our data, as the KOF index has a very high correlation with our main dependent variables.

Rural population as a percent of total population is an inverse measure of the absolute level of urbanization, and has been shown to be negatively related to level of imports as well as NROs, which is echoed in table A2. Lastly, we include the percent of the population that is female since male and females have both physical and mental differences, as well as different norms across countries, which might influence preferences regarding UPF.

2.2.3   Trade and NRO patterns

In this section we limit our discussion to overweight/obesity for females as rates for males exhibit similar patterns. Figure 4 shows trends of the population weighted means of overweight and obesity rates by income group and region. We see three prominent patterns.   First, overweight/obesity is increasing across all regions and income groups. Second, the increases are relatively smooth, which we might expect since NROs can be persistent when measured at national level. Third, overweight/obesity is higher in higher income regions, with the exception that rates are higher for the low-income group than the lower middle-income group. This is echoed in the regional graph if we compare Sub-Saharan Africa (SSA) to South Asia. These results are largely driven by the higher prevalence of underweight in South Asia (mostly driven by lower-middle income India), which contains roughly one-seventh of the world's population. Since the means are

population weighted, India drives the regional and lower-middle income averages lower. We expect higher income countries with higher obesity rates will consume more processed food.

Figure 4: Population-weighted Means of Overweight/obesity rates by Income (left) Group and Region (right)
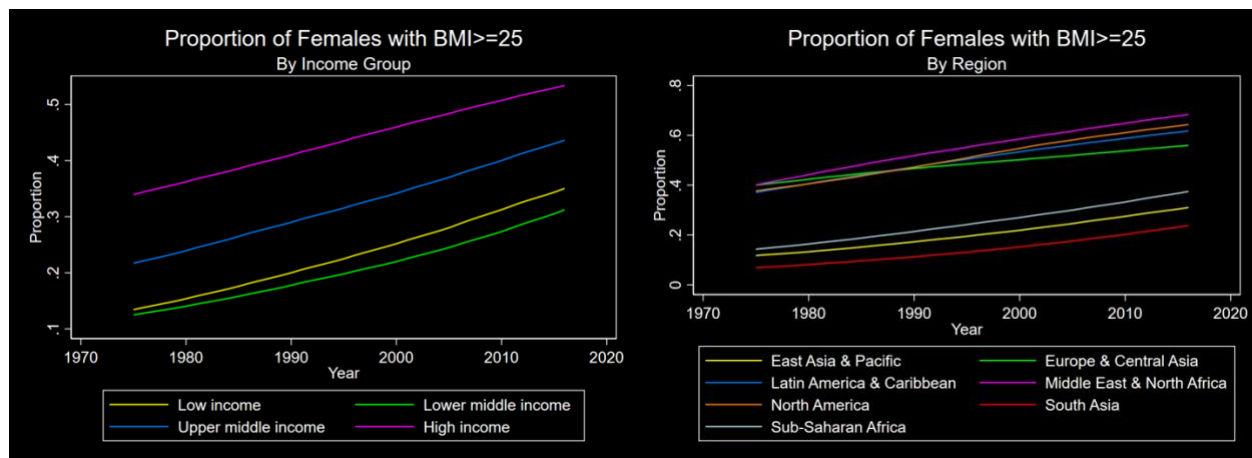
UPF imports are also increasing though not as smoothly. Figure 5 displays population weighted means of per capita UPF imports by income group (left) and region (right). Two prominent points echo what we see in figure 4. First, there is a general increase in UPF imports over time across income groups and regions. The complexities of trade lead to less consistent year-to-year changes in total UPF imports, which is evident in the large variability we see over time. So, we see far more variation than in our NRO measures, but in general we still see an upward trend in UPF imports. Second, UPF imports per capita generally follow income levels, on average. However, SSA and South Asia have very similar levels of UPF imports. This runs contrary to our expectations, but is consistent with the lower rates of overweight/obesity observed in South Asia. We cannot know however if it reflects lower UPF consumption in these countries or higher local production of UPF that displaces imports.

Figure 5: Means of Per Capita UPF Imports by Income Group Excluding High Income (left) and Region Excluding North America and Europe/Central Asia (right)
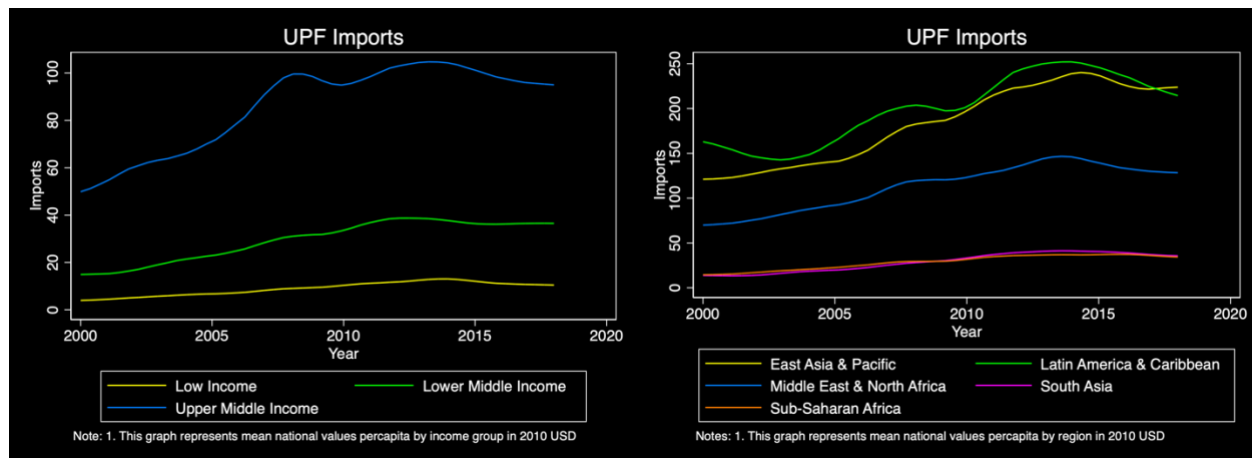
In figure 6, we examine the trends of UPF imports as a share of total imports by both income group and region. There are two distinctions we can see from our previous figures. Lower income countries tend to import higher shares of UPF imports, and the similarities between SSA and South Asia disappear when we examine imports as shares rather than absolute value. SSA imports a much larger share of UPF than South Asia, while they both import comparable per capita amounts of UPF, on average.

Figure 6: Means of UPF Imports as a share of total imports by Income Group (left) and Region (right)
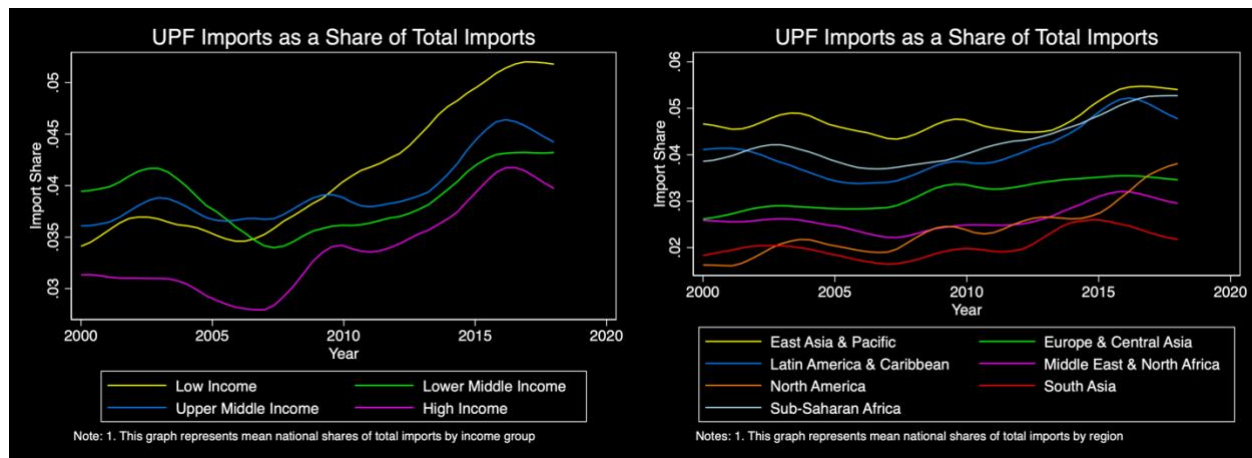
2.2.4   Methods

Our model requires a simultaneous estimation approach since NROs, as a proxy for changes in food norms due to exposure to UPF, and trade flows of processed food both exert a continuous effect on one another. Examples of recent papers that have used simultaneous equations with trade or health models include linking FDI and trade (Kahouli and Omri, 2017); the water-energy-food nexus (Fan et al., 2019; Huang et al., 2020); conflict and food prices (Raleigh et al., 2015); trade and energy (Omri et al., 2015; Tiba and Frikha, 2018); CO2 emissions, FDI, and growth (Omri et al., 2014); and CO2 emissions and health spending (Chaabouni and Saidi 2017; Ullah et al., 2019). To measure the NROs of a population, we first focus on overweight/obesity as our main dependent variable, which has been shown to have a positive relationship with UPF (Hall et al., 2019). We examine the effects using other measures of nutrition-related outcomes as dependent variables to ensure the robustness of our results.

We expect the effect of changes in NROs or trade on the other to take time, and therefore use a dynamic model, as opposed to a truly simultaneous model, that includes the lags of each as a determinant of the other[38]. We also assume there may be some persistence in both trade and NROs, so we include lags of the dependent variable in each equation as well. We use a log-log form for both equations, so estimates are elasticities:

$$\ln(Trade_{it}^n) = \alpha_0 + \alpha_1 \ln(NRO_{i\,t-1}) + \alpha_2 \ln(Trade_{i\,t-1}^n) + \delta_1' \ln(X_{it}) \qquad 5$$
$$+ \delta_2' \ln(Z_{it}^T) + t + \varepsilon_{ijt}$$

$$\ln(NRO_{it}) = \beta_0 + \beta_1 \ln(Trade_{i\,t-1}^n) + \beta_2 \ln(NRO_{i\,t-1}) + \gamma_1' \ln(X_{it}) \qquad 2$$
$$+ \gamma_2' \ln(Z_{it}^N) + t + u_{it}$$

Where α_1, β_1 are our parameters of interest that represent estimates of impact of changes in diet and body norms (proxied by NROs) on trade and the impact of changes in trade on NROs, respectively. While it may be imprecise to do so, we will call these the NRO-to-trade effect (α_1) and the trade-to-NRO effect (β_1) for convenience. $Trade_{it}^n$ is a measure of imports (value, share, or quantity) into country i for processing category n (grouped by processing level), NRO_{it} is a measure of the level of NROs (proportion of overweight/obese, or other nutrition related measure) of country i, X_{it} include demographic and other national characteristics, such as GDP. Z_{it}^N and Z_{it}^T are included only in the NRO or trade equations, respectively. ε_{ijt} and u_{it} are random errors.

---

[38] We estimated our model with the 2nd and 3rd lags and found them to not be significant in most estimations. However, we did find other lags of dependent variables were significant in some cases, but the estimates did not change in any meaningful ways. Additional lags also require additional instruments, which also reduces sample size, so we chose to use only the first lagged difference of each, since our data did not allow for using any further lags beyond the 3rd.

We deal with endogeneity by using the Generalized Method of Moments (GMM)[39] estimator. This allows us to account for endogeneity and generate consistent estimates, while also being robust to heteroskedasticity. Following an approach outlined in Hsiao and Zhou (2015), we use first differences of both the regressors and dependent variables in our main specification to eliminate unobserved heterogeneity. Finding exogenous instruments in a model such as ours is a challenge, so we use the second and third lags of trade and NRO measures in levels as instruments for lagged differences of the respective variables when they appear as explanatory variables, and only the 2nd lags to instrument the respective lagged dependent variables[40]. We use the minimum number of instruments that allow for overidentification, while limiting bias from adding many potentially weak instruments (Roodman, 2009). In all estimations, standard errors are clustered at the country level, and are robust to heteroskedasticity and arbitrary correlation within panels. We estimate GMM in two steps allowing clustering in the weight matrix in the second step.

2.3   Results

Two main patterns can be seen in the results. First, UPF imports are estimated to have a strong relationship with nutrition related outcomes that are mostly impacted in the short term (such as overweight and obesity rates). Second, the NRO-to-trade effect is large and significant across the

---

[39] In addition to simultaneous GMM, we estimated single equations via GMM for both the NRO and trade equations to serve as a point of comparison. We also estimate each equation using the fixed effects estimator, as well as the two-stage least squares variant. These results can be provided by the authors upon request.

[40] We attempted to include further lags of the dependent variable as instruments, however none of our estimations had sufficiently low Hansen J test statistics to confirm instrument validity. As Roodman (2009) points out, there is a tradeoff between including further lags and sample size. Models estimated with the 2nd and 3rd lagged regressors also performed better than only including the 2nd lag, but there were no gains from increasing the instruments beyond the 3rd lag.

majority of estimations. This shows strong evidence of the existence of both direct and indirect pathways between trade and NROs, with NROs serving as a proxy for changing diet and body norms.

Table 21 shows the results between UPF imports and overweight, obesity, and BMI for males and females[41]. The first important point is that we see significant and positive relationships in most estimations in both NRO and trade equations, apart from BMI for females and quantity, for both males and females. The elasticities for the effect from NROs to trade, however, are much larger than the effect from trade to NROs. This is seen if we look at the "Overweight + Obesity" column in Table 19, which represents our main specification. The three rows labeled value, quantity, and share represent UPF imports per capita in 2010 USD, in 1000 metric tons, and UPF imports as a share of total imports, respectively. The NRO-to-trade effect (α_1 in equation 1) is displayed in first three rows, and the trade-to-NRO effect (β_1 in equation 2) is displayed in the three rows below. When we examine the relationship between overweight/obesity and the value of UPF imports, our estimated NRO-to-trade effect is 0.498% while the trade-to-NRO effect is estimated to be 0.00036%. These two effects may differ by orders of magnitude. Changes in NROs take time to manifest, however, preferences and norms can shift rapidly in comparison. When we observe changes in nutrition related outcomes, it is usually the result of diets that have already shifted. So, we would see a much more pronounced increase in UPF imports than we would changes in the NROs of a country, and thus a larger elasticity.

---

[41] Table A3 contains the full set of results for the relationship between UPF imports in per capita values, and both male and female overweight/obesity.

We estimate at least marginally significant NRO-to-trade effects for every single estimation using measures of weight-related NROs of the male population. However, we do not estimate any trade-to-NRO effect when considering male overweight or obesity using quantity as the measure of UPF imports. This is perhaps slightly stronger evidence for the NRO-to-trade effect than the trade-to-NRO effect, which has already been well established in the literature.

Table 21: UPF trade and NROs: NRO-to-trade and trade-to-NRO effects with NRO as body weight measures and UPF trade in value, quantity, and shares

	Male			Female		
	Overweight + Obesity	Obesity	BMI	Overweight + Obesity	Obesity	BMI
NRO-to-trade, UPF trade as:						
Value	0.498**	0.256**	1.608*	0.435**	0.230**	1.848
Quantity	1.236***	0.604***	2.066*	0.512*	0.375**	0.308
Share	0.640**	0.342***	1.814*	0.581***	0.313***	1.644
Trade-to-NRO, UPF trade as:						
Value	0.000360**	0.000560*	0.000253***	0.000494**	0.000882*	0.000340***
Quantity	8.54e-05	8.99e-05	4.95e-05**	0.000100	0.000227	6.91e-05**
Share	0.000296***	0.000543***	0.000226***	0.000496***	0.000880***	0.000276***

*** p<0.01, ** p<0.05, * p<0.1

Lagged instruments pose some risks for identifying effects, and have been known to produce bias if the instruments are not valid. To test the validity of our over-identifying restrictions we employ Hansen J test (Hansen, 1982). The p-values range from 0.157 to 0.813, which suggest our instruments are valid in all cases.

2.4 Other Trade and NRO Pathways

To ensure we are truly measuring the effect we claim, we need to explore whether this is only detectable for weight related categorizations or other nutrition-related measures related to

nutrition. We also need to examine whether we see this effect in all trade, and if so, how that compares with UPF. So, we examine the same model with different sets of NROs and trade dependent variables.

2.4.1 Other Trade Variables

We first examine imports of each of the other processing categories, as well as non-food imports. We estimate the same model as described in the methods section using each of these in place of UPF imports.

Table 22 shows our results for the other processing categories in the NOVA classification. We do not estimate any significant trade-to-NRO effects for any other category of imports on male or female overweight/obesity using per capita values, which is evidence in favor of UPF being the main driver of changes in diet norms. We also estimate significant and large effects on the NRO-to-trade pathway. Interestingly, the effect on UPF (0.498) from changes in male overweight/obesity is not as large as the estimated effect on culinary ingredients (0.936) or minimally/un-processed foods (1.313). Countries also process their own food, which requires inputs such as raw grain or butter. These changes in diet norms not only bring in UPF, but also the ingredients necessary to domestically produce UPF.

We estimate no effect in either direction from Non-food, which suggests this relationship only exists between food imports and NROs. In addition, no trade-to-NRO effects are estimated for anything other than UPF, which also means our hypothesized relationship is only estimated with UPF.

Table 22: Comparing UPF trade to other trade: NRO-to-trade and trade-to-NRO effects with NRO as overweight + obesity and alternative measures of trade (all trade in value terms)

	Male	Female
NRO-to-trade, trade value of:		
Non-food	0.0107	-0.120
Un/minimally processed foods	1.313***	1.242***
Culinary ingredients	0.936***	0.919***
Non-UPF processed foods	0.122	-0.0480
UPF	0.498**	0.435**
Trade-to-NRO, trade value of:		
Non-food	-4.43e-06	-3.50e-05
Un/minimally processed foods	1.36e-05	6.74e-05
Culinary ingredients	-3.57e-05	-7.75e-05
Non-UPF processed foods	-3.69e-05	-0.000122
UPF	0.000360**	0.000494**

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

2.4.2   Other NRO Variables

We repeat the same exercise as above with other NROs. This entails replacing overweight/obesity with BMI, cholesterol levels, prevalence of high blood pressure, diabetes, and underweight. The results for male and female nutrition-related outcomes are displayed in Table 23.

Table 23: UPF trade and NROs: NRO-to-trade and trade-to-NRO effects with alternative measures of NROs (all trade in value terms)

	Male	Female
NRO-to-trade, NRO as:		
Cholesterol	0.472	0.787
Diabetes	0.558***	0.545***
High blood pressure	0.291	0.136
Underweight	-0.142	0.104
Trade-to-NRO, NRO as:		
Cholesterol	-0.000103	2.66e-05
Diabetes	0.00214***	0.00294***
High blood pressure	7.03e-05	-9.45e-05
Underweight	-0.00101***	0.000502

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

UPF has positive effect on male and female diabetes prevalence (0.00214, 0.00294), while negatively related to male underweight (-0.00101). We only see the two-way linkage with diabetes

since this is the only estimation where the lag of NRO is significant in the trade equation (0.558 for males, and 0.545 for females), which is comparable to the estimated effect from overweight/obesity we see in table 21 (0.498 for males, and 0.435 for females). The only other significant estimate we see in table 23 is the trade-to-NRO effect from increases in UPF imports on male underweight. We find the elasticity is negative and larger in absolute value than the estimated effect on male overweight/obesity. This highlights the double edged sword of importing calorie dense food for developing countries with high rates of undernutrition.

We do not estimate NRO-to-trade or trade-to-NRO effects from any other variables, High-blood pressure may have other causal factors beyond nutrition and cholesterol levels may take time to develop. These results all suggest we find the much more evidence of a relationship with weight related measures and UPF than any other combination of NRO and trade measures.

2.4.3   Income Group Heterogeneity

To examine whether we find heterogenous impacts between World Bank income groups (low, lower middle, upper middle, and high), we estimate our model using subsamples of combinations of World Bank income groups. We are unable to estimate individual income groups due to the sample size being too small given the number of parameters in our model. So, we use a minimum of two income groups together in an estimation. Table 24 displays our results.

Two prominent points emerge. First, we find that when we exclude high income countries the magnitude of the effects increase. For instance, we estimate the male NRO to trade effect for Overweight/obesity on UPF imports as 0.498 using all countries, and 0.683 when we exclude high-income countries. However, the 95% confidence intervals overlap, so we cannot say these values

80

are statistically different. Second, we find significant 2-way impacts only when our subsample

consists of low, lower middle, and upper middle income countries. Perhaps, we can only estimate

impacts when countries that are currently in a transitionary state, likely still reliant on imported

UPFs to a greater extent than those produced domestically.

Table 24: Comparing UPF trade and Overweight/Obesity by income groups: NRO-to-trade and trade-to-NRO effects with NRO as overweight + obesity and trade as UPF imports using various combinations of World Bank income groups

	Male	Female
NRO-to-trade, subsamples by income groups:		
Low and Lower-Middle Income	0.260	0.212
Lower and Upper Middle Income	0.591*	0.511**
Upper Middle and High Income	0.331	0.326
Low, Lower Middle, and Upper Middle Income	0.683**	0.656***
Lower Middle, Upper Middle, and High Income	0.473*	0.413**
All Countries	0.498**	0.435**
Trade-to-NRO, subsamples by income groups:		
Low and Lower-Middle Income	-0.000301	8.54e-05
Lower and Upper Middle Income	0.000546	0.000499
Upper Middle and High Income	7.55e-05	-5.89e-05
Low, Lower Middle, and Upper Middle Income	0.000432**	0.000509**
Lower Middle, Upper Middle, and High Income	0.000214	4.68e-05
All Countries	0.000360**	0.000494**

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

2.5   Discussion and Policy Implications

We choose to focus on overweight or obesity since these measures are often used to predict

negative health outcomes related to high calorie diets with poor nutrition that are typically

associated with over consumption of UPF[42]. Body mass measures are likely the first nutrition-related metrics to exhibit noticeable changes from shifts in diet norms. We can think of using these variables as measuring the impact at the margins where the effect from diet shifts BMI into ranges often associated with negative nutrition-related health outcomes.

Our results suggest there is a very significant (although arguably not always practically large) interaction between changes in NROs and trade. While the elasticities for the trade-to-NRO effect may seem small, they are comparable to findings from recent literature. Lin et al (2018) examine a single equation model with average BMI as the dependent variable and find a 10% increase in sugar and UPF imports is associated with a 0.0002 increase in average BMI. When examining a linear model in levels Goryakin et al. (2017) find a one-liter increase in annual per capita sales of SSBs are associated with 0.019, 0.116, and 0.098 increases in BMI, overweight (%), and obese (%), respectively. These are larger than our analysis, but when they control for country fixed effects, the point estimates are smaller and no longer statistically significant (0.001, -0.002, and 0.006, respectively). It is possible that the discrepancies in magnitude are due to the smaller sample size and differences in the model and specification. Vandevijvere et al. (2019) examine the effect of total sales of UPF (food and drink) on mean BMI in a mixed model, and find every one kilogram per capita increase annual of sales of UPF is associated with an increase of 0.0003 for male BMI and a 0.0004 increase in female BMI. While these studies apply different methods and use different regressors, we see results of comparable magnitude and direction.

---

[42] While it may be true that a BMI of over 25 (but below 30) has an ambiguous relationship with health, a BMI greater than 30 (obesity) is generally shown to have a negative relationship with health. Teufel et al. (2021) show that optimal BMI cutoffs used to predict negative health outcomes may vary by region, but do not exceed 28.3. This implies using obesity would adequately serve to associate nutrition-related weight outcomes with negative health.

There are other reasons why the effect from UPF may appear small. For example, UPF imports are usually not distributed evenly across a country, but enter the limited number of ports of entry. It is likely that the urban areas are the first to gain access, and reach a level of abundance much quicker than rural areas in countries that primarily import UPF. So, the impact of increases in UPF imports is likely heterogenous within a country, which implies we should expect overweight/obesity rates to rise quicker in heavily populated urban areas, meaning the effect would appear diminished when averaged at the national level. It is also important to consider that not everyone that consumes UPF will become obese, and so we might expect this effect to be small. It is also true that the size of this effect is dampened by developed countries. Countries in the North America and the European Union are some of the largest UPF importers. Many of these countries already have very high overweight/obesity rates, and might require much higher amounts of UPF imports to see any drastic changes in average weight outcomes. So, it might be the case that the effect is so small for those countries it pulls our estimates downward.

On the other hand, the NRO-to-trade effect is much larger. Consider the results from Table 21. The NRO elasticity estimate shows 1% increase in overweight/obesity rates is associated with a 0.498% increase in next period UPF imports, which is several orders of magnitude larger than the trade-to-NRO effect we estimated in the same model (0.0004%). We do not have literature to which we can compare these values, so we must rely on our intuition to explain why this might occur. Observed changes in weight result from choices that have already been made, or UPF already consumed. We can think of that as resulting from demand previously satisfied. This implies our measured increases in body-mass outcomes are the result of diets that have already begun shifting towards calorie dense foods such as UPF. In other words, demand has increased, and if domestic production capacity is not sufficient, a country must import. If we consider a

developed country with already high rates of overweight/obesity, their rates change at much slower rates than developing countries due to the diminishing returns on body mass gains. We might still see increases in UPF, however, which would be associated with very small changes in BMI or other related weight measures. On the other hand, for the average country in SSA, we would expect much larger changes in weight related indicators, and comparably large increases in UPF imports. As a result, the changes in NROs due to shifts in diet preferences would appear much smaller than the changes in imports following those same changes in levels of NROs.

There are three potential issues with our analysis. First[43], there is a known underdiagnosis of diabetes in developing countries (Misra et al., 2019), so one might conclude these results are merely the result of better diagnostic tools and procedures for diabetes detection that typically follow increases in wealth and income. I do not believe this is an issue in our analysis. We control for various factors related wealth, income, and development, which should capture at least some of the effects from increased diabetes detection related to growth. If the estimates are driven primarily by increases in the ability to detect diabetes, then we might also pick up similar relationships with other NRO variables since there are also issues detecting and treating high cholesterol (Murphy et al., 2017) and hypertension (Mohsen Ibrahim, 2018) in low- and middle-income countries.

Second[44], we do not have data on the proportion of UPF consumed that are imported versus domestically produced. This poses an issue for our analysis, since we might not be able to truly link the increase in UPF imports as the main driver of obesity increases. For example, increases in

---

[43] This issue was brought to my attention by one of my dissertation committee members, Dr. Felicia Wu.
[44] The second and third issues were noted by Dr. Thomas Reardon, another dissertation committee member.

overweight and obesity in Mexico were linked to the unhealthy exports from the US, however Mexico is a significant UPF producer and exporter (Popkin and Reardon, 2018). So, one might conclude imports into Mexico have a less substantial role in the obesity increases. That may be true currently, however the increases in obesity likely occurred long before Mexico developed such a robust food processing industry, which means the initial UPF imports were the main causes of such increases. In addition, Mexico is largely exporting their UPF to other less developed and less obese countries, which are all becoming more obese. So, I do not believe this is in contradiction to our model, and is in fact contained within it.

Third, we do not have data on the distribution of UPF imports within each country or who is actually consuming it. This may potentially limit our ability to claim any relationship we find is consistent across the entire population. For instance, if wealthy households consume the imported UPF and poor households consume the domestically produced UPF, and obesity rates are increasing for both, we cannot say it is imports that are responsible for anything other than increasing obesity rates amongst the wealthy. While it is true that gaps in obesity rates amongst rural and urban dwellers are falling (Jiwani et al., 2019), the distribution of UPF consumption is largely dependent on context. In Canada, income groups were found to consume an equally diverse set of UPFs (Seale et al., 2019), while in Portugal, the wealthiest households consumed the highest proportion of their diet as UPF (Costa de Miranda et al., 2021). Let us assume that the wealthiest households are the only ones to consume UPF imports. Then we can use the same logic as above, to describe the initial gains in obesity are largely due to UPF imports, and only prevalent amongst wealthy household. In developing countries, it is well known that larger body sizes signal wealth as much as they do disease absence (Renzaho, 2004). Just as with other new technologies, the wealthy adopt imported UPF early, which will become cheaper and more accessible to poor

85

households as adoption increases, and eventually domestic production will follow. So, I do not believe this diminishes our results in any way.

Nonetheless, the NRO-to-trade effect has not been explored prior to our research, and we see very strong indication that it does exist. This is important especially for developing countries where exposure to UPF is not as prevalent as it might be compared with developed countries. As more of the population is exposed, we are likely to see increased uptake of UPF consumption, leading to changes in diet norms satisfied through increased UPF imports, which will manifest in increased rates of nutrition-related issues that accompany UPF consumption.

2.5.1   Policy Implications

UPF has some potential benefit for developing countries, so banning UPF would not necessarily result in optimal outcomes. This presents a double edge sword for policy makers, as there are both negative and positive consequences of importing and producing UPF. Our results showed a significant reduction in underweight rates with increased UPF imports. While trade costs may imply UPF imports may not have the same low-cost benefits they might in their country of origin, domestic production is likely to increase over time. The FDI associated with increased UPF imports, through channels noted by Hawkes et al. (2006), can also provide numerous economic benefits which might include infrastructure development, new labor opportunities, and increases in GDP. UPF typically has a longer shelf life, and can reduce a household's food waste. All these ancillary effects can be beneficial to a developing country. The risk lies with an uninformed population that may think a bag of chips has comparable nutrition to the product from which they

originate. Lack of knowledge can allow for overconsumption of UPF, leading to many of the negative nutrition-related outcomes discussed in this paper.

Some countries have tried to limit sugar consumption by introducing excise taxes with some degree of success (Backholer et al., 2016). However, the immediate gratification from consuming temptation goods might outweigh the increase in price, thus leading to little effect other than reducing disposable income for the poor or shifting sales to other products or regions (Cawley et al., 2019). This type of policy ignores the socio-cultural aspects that led to that preference in the first place. In Oakland, California, a tax on SSBs was found to reduce the volume of SSBs purchased by 14%, however this was offset by an 8% increase in sales at the border among other tax avoidance strategies (Léger and Powell, 2021).

Further complication arises in developing countries where processing is not dominated by few large processors, but rather provided by many small and medium enterprises primarily in the informal sector (Reardon et al., 2021). Taxation in this these countries may be limited, but there has been some success in places such as Mexico, one of the first countries to enact a SSB tax (Colchero et al., 2017). In addition, other policies such as labeling may help induce shifts towards using healthy ingredients in processed food production. Chile is among the first countries to adopt strict labeling and advertising policies around food. The restrictions on advertising were found to have driven large reductions in child-directed marketing for unhealthy cereals (Stoltze et al., 2019). However, firms were found to make little changes to the composition of their products in anticipation of these laws going into effect (Kanter et al., 2019). So, while there were positive benefits in reducing messaging aimed at vulnerable populations, such as children, the food environment did not change in a comparable manner. The limited success of the Chilean labeling

policies may be a prime example of the need for policy makers to use multi-faceted approaches to solve nutrition-related issues.

A combined approach is necessary to induce lasting changes. Our model implies the norms around food are at least equally important to understanding and curbing the rapid rise in obesity as UPF itself. Just as in the Chilean example above, it is not enough to target one aspect of unhealthy food production (e.g., labeling), but it is also necessary to encourage producers to use healthier ingredients, and also nudge food norms of consumers toward healthier diets. In a recent review by Laiou et al. (2021), there is evidence that some "nudge" interventions were successful at encouraging shifts toward healthier lifestyle choices such as changing the presentation and proximity of healthy food items, while others were not (labelling, availability, prompting, functional design and sizing nudge- related interventions). One might also be able to convince consumers to alter their diet habits through science-based recommendations. Zeraatkar et al. (2019) discuss some of the issues and consequences of government dietary recommendations, with a set of solutions to develop dietary guidelines. As it stands, there is significant variability in government recommendations, so individuals may have a difficult time determining whether to trust international organizations (e.g., the WHO) or their own governments (Herforth et al., 2019). Often nutrition policy and recommendations are subject to the same rent-seeking behaviors prevalent in many other aspects of the economy where governments are involved. To alleviate the dangers of establishing arbitrary guidelines or ignoring conflicts of interest, rigorous methods are necessary to provide the public with reliable nutrition knowledge to encourage healthy behaviors. A track record of well-researched guidelines may lead to consumer confidence in those recommendations, and potentially even lead to healthier food norms.

2.6 Conclusion

In this paper, we presented the case for a two-way (direct and indirect) relationship between UPF and nutrition related outcomes, which serve as a proxy for the impact from UPF exposure on diet and body norms, and later tested for its existence. We summarize our conceptual model as follows. When developing or low-income countries import UPF, portions of the population may become exposed that previously had never consumed UPF. Exposure to greater abundance and new types of UPF in low- and middle-income countries, typically leads to a shift in demand towards higher consumption of UPF. This demand shift leads to higher imports of UPF, which exposes a higher degree of the population to UPF, resulting in a feedback loop where increases in UPF demand and imports feed into one another. We proxied for this change in preferences using national measures of nutrition related outcomes.

To test for the existence of the relationship between UPF and NROs, we used dynamic trade and NRO equations in log form with lagged trade and NRO effects. We estimated a dynamic model using first differenced equations and the generalized method of moments estimator. We controlled for various socio-economic characteristics, agricultural development, and factors relating to trade and NROs. In addition to measuring the relationship between UPF and weight-related measures, we also tested the relationship between UPF and other nutrition related measures such as diabetes. We examined the robustness of our results by also testing for the existence of a two-way relationship between nutrition-related outcomes and imports of other NOVA categories as well as non-food imports.

We found the relationship was positive and significant between UPF and weight-related measures such as overweight, obesity, and BMI. Interestingly, this relationship also existed

between UPF and diabetes but not for any other nutrition-related measures. We also found that overweight/obesity did not have a significant two-way relationship with any other categories of imports, which makes the results on UPF more compelling.

Our evidence suggests that a feedback loop may exist between UPF imports and shifts in diet preferences. This manifests in declining nutrition-related health and may pose a serious challenge for both policy makers and public health officials as externalities may emerge as both UPF consumption and the unhealthy proportion of the population within a country both grow.

There are some limitations of our analysis. First, we do not have data on UPF distribution within a country, nor do we have similarly disaggregated data on nutrition-related outcomes within a country, so we could be underestimating the effects. One potential explanation for our underestimation could be related to how UPF enters and is transmitted throughout a country. For example, it could be the case that the populations of lower- and middle-income countries that reside in main trading ports of entry, will be the first to see rapid increases in UPF consumption alongside increases in overweight and obesity. So, UPF and obesity increases would only rise rapidly in a small portion of the country, which when examined in the aggregate, appears smaller due to the slow increases seen in the larger rural population. Second, other cultural phenomena, such as the advent of social media, may also play a role in diminishing activity or shifting diets of a population, but we could not analyze that in this analysis.

REFERENCES

Abarca-Gómez, L., Z. A. Abdeen, Z. A. Hamid, N. M. Abu-Rmeileh, B. Acosta-Cazares, C. Acuin, … M. Ezzati. 'Worldwide Trends in Body-Mass Index, Underweight, Overweight, and Obesity from 1975 to 2016: A Pooled Analysis of 2416 Population-Based Measurement Studies in 128·9 Million Children, Adolescents, and Adults'. *The Lancet* 390,10113(2017):2627–2642.

Ademola, O., N. Saha Turna, L. S. O. Liverpool-Tasie, A. Obadina, and F. Wu. 'Mycotoxin Reduction through Lactic Acid Fermentation: Evidence from Commercial Ogi Processors in Southwest Nigeria'. *Food Control* 121(March 2021):107620.

Alwan, H., B. Viswanathan, J. Williams, F. Paccaud, and P. Bovet. 'Association Between Weight Perception and Socioeconomic Status Among Adults in the Seychelles'. *BMC Public Health* 10,1(2010):467.

An, R., J. Shen, T. Bullard, Y. Han, D. Qiu, and S. Wang. 'A Scoping Review on Economic Globalization in Relation to the Obesity Epidemic'. *Obesity Reviews* 21,3(March 2020):e12969.

Backholer, K., D. Sarink, A. Beauchamp, C. Keating, V. Loh, K. Ball, … A. Peeters. 'The Impact of a Tax on Sugar-Sweetened Beverages According to Socio-Economic Position: A Systematic Review of the Evidence'. *Public Health Nutrition* 19,17(December 2016):3070–3084.

Baker, P., and S. Friel. 'Food Systems Transformations, Ultra-Processed Food Markets and the Nutrition Transition in Asia'. *Globalization and Health* 12,1(2016):80.

Bickel, W. K., L. Moody, and S. T. Higgins. 'Some Current Dimensions of the Behavioral Economics of Health-Related Behavior Change'. *Preventive Medicine* 92(2016):16–23.

Canella, D. S., R. B. Levy, A. P. B. Martins, R. M. Claro, J. C. Moubarac, L. G. Baraldi, … C. A. Monteiro. 'Ultra-processed Food Products and Obesity in Brazilian Households (2008-2009)'. *PLoS ONE* 9,3(March 2014):e92752.

Cawley, J., A. M. Thow, K. Wen, and D. Frisvold. 'The Economics of Taxes on Sugar-Sweetened Beverages: A Review of the Effects on Prices, Sales, Cross-Border Shopping, and Consumption'. *Annual Review of Nutrition* 39,1(August 2019):317–338.

Chaabouni, S., and K. Saidi. 'The Dynamic Links Between Carbon Dioxide (CO2) Emissions, Health Spending and GDP Growth: A Case Study for 51 Countries'. *Environmental Research* 158(October 2017):137–144.

Christian, A. K., and G. A. Frempong. 'Correlates of Over- or Under-Estimation of Body Size Among Resource-Poor Urban Dwellers in a Sub-Saharan African City'. *Annals of Human Biology* 47,7–8(November 2020):602–609.

Colchero, M. A., J. Rivera-Dommarco, B. M. Popkin, and S. W. Ng. 'In Mexico, Evidence of Sustained Consumer Response Two Years After Implementing a Sugar-Sweetened Beverage Tax'. *Health Affairs* 36,3(March 2017):564–571.

Conte, M., P. Cotterlaz, and T. Mayer. 'The CEPII gravity database'. *CEPII: Paris, France* (2021).

Cornwell, B., E. Villamor, M. Mora-Plazas, C. Marin, C. A. Monteiro, and A. Baylin. 'Processed and Ultra-Processed Foods are Associated with Lower-Quality Nutrient Profiles in Children from Colombia'. *Public Health Nutrition* 21,1(2018):142–147.

Costa de Miranda, R., F. Rauber, M. M. de Moraes, C. Afonso, C. Santos, S. Rodrigues, and R. B. Levy. 'Consumption of Ultra-Processed Foods and Non-Communicable Disease-Related Nutrient Profile in Portuguese Adults and Elderly (2015–2016): The UPPER Project'. *British Journal of Nutrition* 125,10(May 2021):1177–1187.

Cuevas García-Dorado, S., L. Cornselsen, R. Smith, and H. Walls. 'Economic Globalization, Nutrition and Health: a Review of Quantitative Evidence'. *Globalization and Health* 15,1(December 2019):15.

De Deus Mendonca, R., A. C. Souza Lopes, A. M. Pimenta, A. Gea, M. A. Martinez-Gonzalez, and M. Bes-Rastrollo. 'Ultra-Processed Food Consumption and the Incidence of Hypertension in a Mediterranean Cohort: The Seguimiento Universidad de Navarra Project'. *American Journal of Hypertension* 30,4(April 2017):358–366.

Dreher, A. 'Does Globalization Affect Growth? Evidence from a new Index of Globalization'. *Applied Economics* 38,10(June 2006):1091–1110.

Drichoutis, A. C., P. Lazaridis, and R. M. Nayga. *Body Weight Outcomes and Food Expenditures Among Older Europeans : A Simultaneous Equation Approach* (pp. 1–17). 2009.

Drichoutis, A. C., R. M. Nayga, and P. Lazaridis. 'Food Away From Home Expenditures and Obesity Among Older Europeans: Are there Gender Differences?' *Empirical Economics* 42,3(June 2012):1051–1078.

Epstein, L. H., S. J. Salvy, K. A. Carr, K. K. Dearing, and W. K. Bickel. 'Food Reinforcement, Delay Discounting and Obesity'. *Physiology and Behavior* 100,5(July 2010):438–445.

Fan, C., C. Y. Lin, and M. C. Hu. 'Empirical Framework for a Relative Sustainability Evaluation of Urbanization on the Water–Energy–Food Nexus Using Simultaneous Equation Analysis'. *International Journal of Environmental Research and Public Health* 16,6(2019).

Friel, S., L. Hattersley, W. Snowdon, A. M. Thow, T. Lobstein, D. Sanders, … C. Walker. 'Monitoring the Impacts of Trade Agreements on Food Environments'. *Obesity Reviews* 14,S1(October 2013):120–134.

Gaulier, G., and S. Zignago. 'Baci: international trade database at the product-level (the 1994-2007 version)' (2010).

Giuntella, O., M. Rieger, and L. Rotunno. 'Weight Gains from Trade in Foods: Evidence from Mexico'. *Journal of International Economics* 122(2020):103277.

Goryakin, Y., P. Monsivais, and M. Suhrcke. 'Soft Drink Prices, Sales, Body Mass Index and Diabetes: Evidence from a Panel of Low-, Middle- and High-income Countries'. *Food Policy* 73(December 2017):88–94.

Gygli, S., F. Haelg, N. Potrafke, and J.-E. Sturm. 'The KOF Globalisation Index – Revisited'. *The Review of International Organizations* 14,3(September 2019):543–574.

Hall, K. D., A. Ayuketah, R. Brychta, H. Cai, T. Cassimatis, K. Y. Chen, … M. Zhou. 'Ultra-Processed Diets Cause Excess Calorie Intake and Weight Gain: An Inpatient Randomized Controlled Trial of Ad Libitum Food Intake'. *Cell Metabolism* 30,1(July 2019):67-77.e3.

Hansen, L. P. 'Large Sample Properties of Generalized Method of Moments Estimators'. *Econometrica* 50,4(July 1982):1029–1054.

Hawkes, C. 'Uneven Dietary Development: Linking the Policies and Processes of Globalization with the Nutrition Transition, Obesity and Diet-Related Chronic Diseases'. *Globalization and Health* 2,1(2006):4.

Hawkes, C., and M. Ruel. 'The Links Between Agriculture and Health: An Intersectoral Opportunity to Improve the Health and Livelihoods of the Poor'. *Bulletin of the World Health Organization* 84,12(2006):985–991.

Hawkes, C., M. T. Ruel, L. Salm, B. Sinclair, and F. Branca. 'Double-duty Actions: Seizing Programme and Policy Opportunities to Address Malnutrition in all its Forms'. *The Lancet* 395,10218(2020):142–155.

Herforth, A., M. Arimond, C. Álvarez-Sánchez, J. Coates, K. Christianson, and E. Muehlhoff. 'A Global Review of Food-Based Dietary Guidelines'. *Advances in Nutrition* 10,4(July 2019):590–605.

Hsiao, C., and Q. Zhou. 'Statistical Inference for Panel Dynamic Simultaneous Equations Models'. *Journal of Econometrics* 189,2(2015):383–396.

Huang, D., G. Li, C. Sun, and Q. Liu. 'Exploring Interactions in the Local Water-Energy-Food Nexus (WEF-Nexus) using a Simultaneous Equations Model'. *Science of the Total Environment* 703(February 2020):135034.

Ifland, J., H. G. Preuss, M. T. Marcus, K. M. Rourke, W. Taylor, and H. Theresa Wright. 'Clearing the Confusion around Processed Food Addiction'. *Journal of the American College of Nutrition* 34,3(May 2015):240–243.

Jiwani, S. S., R. M. Carrillo-Larco, A. Hernández-Vásquez, T. Barrientos-Gutiérrez, A. Basto-Abreu, L. Gutierrez, … J. J. Miranda. 'The Shift of Obesity Burden by Socioeconomic Status Between 1998 and 2017 in Latin America and the Caribbean: A Cross-Sectional Series Study'. *The Lancet Global Health* 7,12(December 2019):e1644–e1654.

Kahouli, B., and A. Omri. 'Foreign Direct Investment, Foreign Trade and Environment: New Evidence from Simultaneous-Equation System of Gravity Models'. *Research in International Business and Finance* 42(December 2017):353–364.

Kanter, R., M. Reyes, S. Vandevijvere, B. Swinburn, and C. Corvalán. 'Anticipatory Effects of the Implementation of the Chilean Law of Food Labeling and Advertising on Food and Beverage Product Reformulation'. *Obesity Reviews* 20,S2(November 2019):129–140.

Kremer, M., G. Rao, and F. Schilbach. 'Behavioral development economics'. In B. D. Bernheim, S. DellaVigna, & D. B. T.-H. of B. E. A. and F. 1 Laibson (Eds.), *Handbook of Behavioral Economics - Foundations and Applications 2* (Vol. 2, pp. 345–458). Elsevier. 2019.

Laiou, E., I. Rapti, R. Schwarzer, L. Fleig, L. Cianferotti, J. Ngo, … E. E. Ntzani. 'Review: Nudge Interventions to Promote Healthy Diets and Physical Activity'. *Food Policy* 102(July 2021):102103.

Lane, M. M., J. A. Davis, S. Beattie, C. Gómez-Donoso, A. Loughman, A. O'Neil, … T. Rocks. 'Ultraprocessed Food and Chronic Noncommunicable Diseases: A Systematic Review and Meta-Analysis of 43 Observational Studies'. *Obesity Reviews* 22,3(March 2020):e13146.

Léger, P. T., and L. M. Powell. 'The Impact of the Oakland SSB tax on Prices and Volume sold: A Study of Intended and Unintended Consequences'. *Health Economics* 30,8(August 2021):1745–1771.

Levy, R. B., F. Rauber, K. Chang, M. L. da C. Louzada, C. A. Monteiro, C. Millett, and E. P. Vamos. 'Ultra-processed Food Consumption and Type 2 Diabetes Incidence: A Prospective Cohort Study'. *Clinical Nutrition* (2021).

Lin, T. K., Y. Teymourian, and M. S. Tursini. 'The Effect of Sugar and Processed Food Imports on the Prevalence of Overweight and Obesity in 172 Countries'. *Globalization and Health* 14,1(2018):1–14.

Louzada, M. L. da C., L. G. Baraldi, E. M. Steele, A. P. B. Martins, D. S. Canella, J. C. Moubarac, … C. A. Monteiro. 'Consumption of Ultra-Processed Foods and Obesity in Brazilian Adolescents and Adults'. *Preventive Medicine* 81(December 2015):9–15.

Louzada, M. L. da C., A. P. B. Martins, D. S. Canella, L. G. Baraldi, R. B. Levy, R. M. Claro, … C. A. Monteiro. 'Impact of Ultra-Processed Foods on Micronutrient Content in the Brazilian Diet'. *Revista de Saude Publica* 49(2015).

Louzada, M. L. da C., A. P. B. Martins, D. S. Canella, L. G. Baraldi, R. B. Levy, R. M. Claro, … C. A. Monteiro. 'Ultra-processed Foods and the Nutritional Dietary Profile in Brazil'. *Revista de Saude Publica* 49(2015).

Macia, E., E. Cohen, L. Gueye, G. Boetsch, and P. Duboz. 'Prevalence of Obesity and Body Size Perceptions in Urban and Rural Senegal: New Insight on the Epidemiological Transition in West Africa'. *Cardiovascular Journal of Africa* 28,5(October 2017):324–330.

Martínez Steele, E., B. M. Popkin, B. Swinburn, and C. A. Monteiro. 'The Share of Ultra-Processed Foods and the Overall Nutritional Quality of Diets in the US: Evidence from a Nationally Representative Cross-Sectional Study'. *Population Health Metrics* 15,1(2017):6.

McNamara, C. 'Trade Liberalization and Social Determinants of Health: A State of the Literature Review'. *Social Science and Medicine* 176(March 2017):1–13.

Meldrum, D. R., M. A. Morris, and J. C. Gambone. 'Obesity Pandemic: Causes, Consequences, and Solutions—but Do We Have the Will?' *Fertility and Sterility* 107,4(2017):833–839.

Misra, A., H. Gopalan, R. Jayawardena, A. P. Hills, M. Soares, A. A. Reza-Albarrán, and K. L. Ramaiya. 'Diabetes in Developing Countries'. *Journal of Diabetes* 11,7(July 2019):522–539.

Mohsen Ibrahim, M. 'Hypertension in Developing Countries: A Major Challenge for the Future'. *Current Hypertension Reports* 20,5(May 2018):38.

Monteiro, C. A., G. Cannon, M. Lawrence, M. L. da C. Louzada, and P. P. Machado. 'Ultra-Processed Foods, Diet Quality, and Health using the NOVA Classification System'. *Rome: FAO* 48,August(2019):48.

Moodie, R., E. Bennett, E. J. L. Kwong, T. M. Santos, L. Pratiwi, J. Williams, and P. Baker. 'Ultra-Processed Profits: The Political Economy of Countering the Global Spread of Ultra-Processed Foods – A Synthesis Review on the Market and Political Practices of Transnational Food Corporations and Strategic Public Health Responses'. *International Journal of Health Policy and Management* (May 2021):1.

Moran, A. J., N. Khandpur, M. Polacsek, and E. B. Rimm. 'What Factors Influence Ultra-Processed Food Purchases and Consumption in Households with Children? A Comparison Between Participants and Non-Participants in the Supplemental Nutrition Assistance Program (SNAP)'. *Appetite* 134(March 2019):1–8.

Moubarac, J. C., M. Batal, M. L. Louzada, E. Martinez Steele, and C. A. Monteiro. 'Consumption of Ultra-Processed Foods Predicts Diet Quality in Canada'. *Appetite* 108(January 2017):512–520.

Muhihi, A. J., M. A. Njelekela, R. Mpembeni, R. S. Mwiru, N. Mligiliche, and J. Mtabaji. 'Obesity, Overweight, and Perceptions about Body Weight among Middle-Aged Adults in Dar es Salaam, Tanzania'. *International Scholarly Research Notices* 2012(August 2012):1–6.

Munthali, M., C. Nyondo, M. Muyanga, S. Chimatiro, R. Chaweza, L. Chiwaula, … F. Zhuwao. 'Food Imports in Malawi: Trends, Drivers, and Policy Implications' (2021).

Murphy, A., J. R. Faria-Neto, K. Al-Rasadi, D. Blom, A. Catapano, A. Cuevas, … D. Wood. 'World Heart Federation Cholesterol Roadmap'. *Global Heart* 12,3(September 2017):179.

Naigaga, D. A., D. Jahanlu, H. M. Claudius, A. K. Gjerlaug, I. Barikmo, and S. Henjum. 'Body Size Perceptions and Preferences Favor Overweight in Adult Saharawi Refugees'. *Nutrition Journal* 17,1(2018):17.

Oberlander, L., A. C. Disdier, and F. Etilé. 'Globalisation and National Trends in Nutrition and Health: A Grouped Fixed-Effects Approach to Intercountry Heterogeneity'. *Health Economics (United Kingdom)* 26,9(2017):1146–1161.

Okop, K. J., F. C. Mukumbang, T. Mathole, N. Levitt, and T. Puoane. 'Perceptions of Body Size, Obesity Threat and the Willingness to Lose Weight among Black South African Adults: A Qualitative Study'. *BMC Public Health* 16,1(2016):365.

Omri, A., N. Ben Mabrouk, and A. Sassi-Tmar. 'Modeling the Causal Linkages Between Nuclear Energy, Renewable Energy and Economic Growth in Developed and Developing Countries'. *Renewable and Sustainable Energy Reviews* 42(2015):1012–1022.

Omri, A., D. K. Nguyen, and C. Rault. 'Causal Interactions Between CO2 Emissions, FDI, and Economic Growth: Evidence from Dynamic Simultaneous-Equation Models'. *Economic Modelling* 42(October 2014):382–389.

Pal, P., J. P. Pandey, and G. Sen. 'Processing Techniques or Mycotoxins—A Balancing Act of Food Safety and Preservation'. In *Preservatives and Preservation Approaches in Beverages* (pp. 375–425). Elsevier. 2019.

Peltzer, K., and S. Pengpid. 'Underestimation of Weight and its Associated Factors in Overweight and Obese University Students from 21 Low, Middle and Emerging Economy Countries'. *Obesity Research and Clinical Practice* 9,3(May 2015):234–242.

Pingali, P., B. Mittra, and A. Rahman. 'The Bumpy Road from food to Nutrition Security – Slow Evolution of India's Food Policy'. *Global Food Security* 15(December 2017):77–84.

Popkin, B. M., and T. Reardon. 'Obesity and the Food System Transformation in Latin America: Obesity and Food System Transformation'. *Obesity Reviews* 19,8(August 2018):1028–1064.

Popkin, Barry M., L. S. Adair, and S. W. Ng. 'Global Nutrition Transition and the Pandemic of Obesity in Developing Countries'. *Nutrition Reviews* 70,1(January 2012):3–21.

Poti, J. M., B. Braga, and B. Qin. 'Ultra-processed Food Intake and Obesity: What Really Matters for Health-Processing or Nutrient Content?' *Current Obesity Reports* 6,4(2017):420–431.

Pradeilles, R., M. Holdsworth, O. Olaitan, A. Irache, H. A. Osei-Kwasi, C. B. Ngandu, and E. Cohen. 'Body Size Preferences for women and Adolescent Girls Living in Africa: A Mixed-Methods Systematic Review'. *Public Health Nutrition* (2021):1–22.

Puoane, T., K. Steyn, D. Bradshaw, R. Laubscher, J. Fourie, V. Lambert, and N. Mbananga. 'Obesity in South Africa: The South African Demographic and Health Survey'. *Obesity Research* 10,10(October 2002):1038–1048.

Qaim, M. 'Globalisation of Agrifood Systems and Sustainable Nutrition'. *Proceedings of the Nutrition Society* 76,1(February 2017):12–21.

Raleigh, C., H. J. Choi, and D. Kniveton. 'The Devil is in the Details: An Investigation of the Relationships Between Conflict, Food Price and Climate across Africa'. *Global Environmental Change* 32(2015):187–199.

Rauber, F., M. L. da C. Louzada, E. M. Steele, C. Millett, C. A. Monteiro, and R. B. Levy. 'Ultra-processed Food Consumption and Chronic Non-Communicable Diseases-Related Dietary Nutrient Profile in the UK (2008–2014)'. *Nutrients* 10,5(2018).

Rauber, F., E. M. Steele, M. L. da Costa Louzada, C. Millett, C. A. Monteiro, and R. B. Levy. 'Ultra-processed Food Consumption and Indicators of Obesity in the United Kingdom Population (2008-2016)'. *PLoS ONE* 15,5(May 2020):e0232676.

Ravuvu, A., S. Friel, A. M. Thow, W. Snowdon, and J. Wate. 'Monitoring the Impact of trade Agreements on National Food Environments: Trade Imports and Population Nutrition Risks in Fiji'. *Globalization and Health* 13,1(2017):33.

Reardon, T., S. Henson, and A. Gulati. 'Links Between Supermarkets and Food Prices, Diet Diversity and Food Safety in Developing Countries'. *Trade, Food, Diet and Health: Perspectives and Policy Options* (2010):111–130.

Reardon, T., D. Tschirley, L. S. O. Liverpool-Tasie, T. Awokuse, J. Fanzo, B. Minten, … B. M. Popkin. 'The Processed Food Revolution in African Food Systems and the Double Burden of Malnutrition'. *Global Food Security* 28(March 2021):100466.

Renzaho, A. M. N. 'Fat, Rich and Beautiful: Changing Socio-Cultural Paradigms Associated with Obesity Risk, Nutritional Status and Refugee Children from Sub-Saharan Africa'. *Health and Place* 10,1(March 2004):105–113.

Roodman, D. 'A Note on the Theme of Too Many Instruments'. *Oxford Bulletin of Economics and Statistics* 71,1(February 2009):135–158.

Seale, E., L. S. Greene-Finestone, and M. de Groh. 'Examining the Diversity of Ultra-processed Food Consumption and Associated Factors in Canadian Adults'. *Applied Physiology, Nutrition, and Metabolism* 45,8(August 2020):857–864.

Sievert, K., M. Lawrence, A. Naika, and P. Baker. 'Processed Foods and Nutrition Transition in the Pacific: Regional Trends, Patterns and Food System Drivers'. *Nutrients* 11,6(2019):1328.

Srour, B., L. K. Fezeu, E. Kesse-Guyot, B. Allès, C. Debras, N. Druesne-Pecollo, … M. Touvier. 'Ultraprocessed Food Consumption and Risk of Type 2 Diabetes among Participants of the NutriNet-Santé Prospective Cohort'. *JAMA Internal Medicine* 180,2(February 2020):283–291.

Stojek, M. M. K., and J. MacKillop. 'Relative Reinforcing Value of Food and Delayed Reward Discounting in Obesity and Disordered Eating: A Systematic Review'. *Clinical Psychology Review* 55(July 2017):1–11.

Stoltze, F. M., M. Reyes, T. L. Smith, T. Correa, C. Corvalán, and F. R. D. Carpentier. 'Prevalence of Child-Directed Marketing on Breakfast Cereal Packages before and after Chile's Food Marketing Law: A Pre- and Post-Quantitative Content Analysis'. *International Journal of Environmental Research and Public Health* 16,22(November 2019):4501.

Swinburn, B. A., G. Sacks, K. D. Hall, K. McPherson, D. T. Finegood, M. L. Moodie, and S. L. Gortmaker. 'The Global Obesity Pandemic: Shaped by Global Drivers and Local Environments'. *The Lancet* 378,9793(August 2011):804–814.

Tateyama, Y., T. Techasrivichien, P. M. Musumari, S. P. Suguimoto, R. Zulu, M. Macwan'gi, … M. Kihara. 'Obesity Matters but is not Perceived: A Cross-Sectional Study on Cardiovascular Disease Risk Factors among a Population-Based Probability Sample in Rural Zambia'. *PLoS ONE* 13,11(November 2018):e0208176.

Teufel, F., J. A. Seiglie, P. Geldsetzer, M. Theilmann, M. E. Marcus, C. Ebert, … J. Manne-Goehler. 'Body-mass Index and Diabetes Risk in 57 Low-income and Middle-income Countries: A Cross-Sectional Study of Nationally Representative, Individual-Level Data in 685 616 Adults'. *The Lancet* 398,10296(July 2021):238–248.

Thow, A. M., D. Sanders, E. Drury, T. Puoane, S. N. Chowdhury, L. Tsolekile, and J. Negin. 'Regional Trade and the Nutrition Transition: Opportunities to Strengthen NCD Prevention Policy in the Southern African Development Community'. *Global Health Action* 8,1(December 2015):28338.

Tiba, S., and M. Frikha. 'Income, Trade Openness and Energy Interactions: Evidence from Simultaneous Equation Modeling'. *Energy* 147(March 2018):799–811.

Tschirley, D. L., J. Snyder, M. Dolislager, T. Reardon, S. Haggblade, J. Goeb, … F. Meyer. 'Africa's Unfolding Diet Transformation: Implications for Agrifood System Employment'. *Journal of Agribusiness in Developing and Emerging Economies* 5,2(November 2015):102–136.

Ullah, I., S. Ali, M. H. Shah, F. Yasim, A. Rehman, and B. M. Al-Ghazali. 'Linkages Between Trade, CO2 Emissions and Healthcare Spending in China'. *International Journal of Environmental Research and Public Health* 16,21(2019).

Vandevijvere, S., L. M. Jaacks, C. A. Monteiro, J.-C. Moubarac, M. Girling-Butcher, A. C. Lee, … B. Swinburn. 'Global Trends in Ultraprocessed Food and Drink Product Sales and their Association with Adult Body Mass Index Trajectories'. *Obesity Reviews* 20,S2(November 2019):10–19.

Wilkins, A. S. 'To Lag or Not to Lag?: Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis'. *Political Science Research and Methods* 6,2(April 2018):393–411.

Woodward, D., N. Drager, R. Beaglehole, and D. Lipson. 'Globalization and Health: A Framework for Analysis and Action'. *Bulletin of the World Health Organization* 79,9(2001):875–881.

Wooldridge, J. M. *Econometric analysis of cross section and panel data*. MIT press. 2010.

'World Bank Development Indicators'. n.d.

Zeraatkar, D., B. C. Johnston, and G. Guyatt. 'Evidence Collection and Evaluation for the Development of Dietary Guidelines and Public Policy on Nutrition'. *Annual Review of Nutrition* 39,1(August 2019):227–247.

Zhou, Y., and J. Staatz. 'Projected Demand and Supply for Various Foods in West Africa: Implications for Investments and Food Policy'. *Food Policy* 61(May 2016):198–212.

Tables

Table A2: Correlation Matrix of variables in log form

	UPF PC Imports	Overweight (Female)	GDP PC	Net FDI Inflows PC	Manufacturing Trade Flows PC	Largest City (% of Urban Pop)	Incidence of Anemia	Death Rate	Incidence of Tuberculosis
UPF PC Imports	1.00								
Overweight (Female)	0.54	1.00							
GDP PC	0.77	0.78	1.00						
Net FDI Inflows PC	0.27	0.20	0.30	1.00					
Manufacturing Trade Flows PC	0.72	0.76	0.91	0.34	1.00				
Largest City (% of Urban Pop)	-0.50	-0.10	-0.23	-0.10	-0.08	1.00			
Incidence of Anemia	-0.70	-0.74	-0.77	-0.26	-0.74	0.23	1.00		
Death Rate	-0.06	-0.23	-0.27	-0.03	-0.20	-0.04	0.12	1.00	
Incidence of Tuberculosis	-0.62	-0.74	-0.76	-0.26	-0.75	0.17	0.76	0.25	1.00
OOP Health Exp Ratio	-0.48	-0.36	-0.52	-0.28	-0.57	0.17	0.51	0.06	0.50
Ag Land (% of Total)	-0.10	-0.11	-0.30	0.00	-0.25	-0.07	0.08	0.28	0.21
Ag Employment (% of Total)	-0.71	-0.72	-0.86	-0.27	-0.84	0.22	0.72	0.21	0.80
Food Production Index	0.38	0.26	0.32	0.06	0.34	-0.05	-0.30	-0.06	-0.29
KOF Index	0.78	0.71	0.82	0.29	0.86	-0.23	-0.79	-0.06	-0.72
Electrification	0.56	0.73	0.71	0.13	0.62	-0.24	-0.63	-0.33	-0.55
% Rural	-0.48	-0.66	-0.74	-0.19	-0.66	0.08	0.54	0.36	0.57
Mobile Subs (per 100)	0.57	0.54	0.61	0.14	0.64	-0.09	-0.46	-0.19	-0.45
% Female	0.00	-0.09	-0.26	-0.04	-0.19	-0.01	-0.06	0.68	0.17

	OOP Health Exp Ratio	Ag Land (% of Total)	Ag Employment (% of Total)	Food Production Index	KOF Index	Electrification	% Rural	Mobile Subs (per 100)	% Female
UPF PC Imports									
Overweight (Female)									
GDP PC									
Net FDI Inflows PC									
Manufacturing Trade Flows PC									
Largest City (% of Urban Pop)									
Incidence of Anemia									
Death Rate									
Incidence of Tuberculosis									
OOP Health Exp Ratio	1.00								
Ag Land (% of Total)	0.13	1.00							
Ag Employment (% of Total)	0.60	0.27	1.00						
Food Production Index	-0.21	0.02	-0.24	1.00					
KOF Index	-0.54	-0.10	-0.77	0.35	1.00				
Electrification	-0.13	-0.10	-0.56	0.22	0.64	1.00			
% Rural	0.48	0.30	0.77	-0.17	-0.58	-0.46	1.00		
Mobile Subs (per 100)	-0.32	-0.10	-0.49	0.57	0.67	0.46	-0.41	1.00	
% Female	0.15	0.39	0.22	0.02	-0.02	-0.07	0.39	-0.07	1.00

Table A3: Main Results

	Male		Female	
	Trade Eq	NRO Eq	Trade Eq	NRO Eq
Lagged NRO Variables (H)				
Overweight Obesity	0.498**	0.995***	0.435**	1.007***
	(0.236)	(0.00171)	(0.203)	(0.00226)
Lagged Trade Variables (T)				
UPF PC Imports	0.359***	0.000360**	0.368***	0.000494**
	(0.0508)	(0.000179)	(0.0520)	(0.000238)
NRO Variables (Zn)				
Incidence of Anemia		0.000771		0.00319**
		(0.00112)		(0.00142)
Death Rate		0.00103**		-0.000505
		(0.000439)		(0.000557)
Incidence of Tuberculosis		-0.000113*		-0.000156*
		(6.78e-05)		(8.14e-05)
OOP Health Exp Ratio		7.77e-05*		9.29e-05*
		(4.25e-05)		(5.27e-05)
Trade Variables (Zt)				
Net FDI Inflows PC	-0.354*		-0.244	
	(0.208)		(0.202)	
Manufacturing Trade Flows PC	0.552***		0.558***	
	(0.0473)		(0.0469)	
Largest City (% of Urban Pop)	0.102		0.0603	
	(0.105)		(0.107)	
Agricultural Variables				
Ag Land (% of Total)	-0.208	-0.000360	-0.304**	-0.000513
	(0.135)	(0.000255)	(0.132)	(0.000334)
Ag Employment (% of Total)	-0.105***	-6.71e-06	-0.0971**	6.44e-05
	(0.0392)	(8.63e-05)	(0.0383)	(0.000110)
Food Production Index	-0.0116	-6.13e-06	-0.0201	-5.18e-05
	(0.0498)	(5.68e-05)	(0.0497)	(7.75e-05)
Other Controls				
GDP PC	0.359***	0.000360**	0.0723	0.000272*
	(0.0508)	(0.000179)	(0.0578)	(0.000144)
KOF Index	-0.0642	0.000844***	-0.133	0.000596
	(0.241)	(0.000324)	(0.240)	(0.000441)
Electrification	-0.00514	-6.13e-06	-0.00247	7.65e-06
	(0.0157)	(1.17e-05)	(0.0156)	(1.58e-05)
% Rural	0.0647***	7.87e-05	0.0709***	4.90e-05
	(0.0145)	(0.000102)	(0.0129)	(0.000111)
Mobile Subscriptions (per 100)	-0.0230	0.000105**	-0.0203	0.000116**
	(0.0142)	(4.45e-05)	(0.0153)	(5.75e-05)
% Female	0.541**	-0.00336***	0.650***	-0.00425***
	(0.222)	(0.000942)	(0.244)	(0.00120)
Year Dummies	X	X	X	X
Observations	1,901	1,901	1,901	1,901

Standard errors in parentheses are robust to heteroskedasticity and arbitrary within panel correlation, *** p<0.01, ** p<0.05, * p<0.1

CHAPTER 3:   How Do Low-income Urban Consumers Obtain Their Food and Does This Impact the Prices They Pay?

3.1   Introduction

In this paper, we seek to understand the food prices poor households face, as they spend a large portion of their budget on food. There are two strands of literature that attempt to explain how the poor interact with their food environments (FE)[45]. The first is poverty penalty literature, and in the context of food prices, researchers seek to answer the question, "do the poor pay more?"[46]. The second is store choice literature, which often looks to investigate why shoppers choose certain retail food outlets. While many of these poverty-focused papers attempt to deal with consumer behavior, they tend to focus on characteristics of stores, households, and individuals with little attention paid to the way in which consumers do their shopping and whether that behavior impacts the price they pay for food.

The evidence on whether low-income households pay more is inconsistent. Typically, poor households are believed to face liquidity constraints, and so are unable to take advantage of scale economies by buying in bulk, resulting in higher unit prices paid than wealthier households (Attanasio and Frayne, 2006; Mussa, 2015: Gibson and Kim, 2018; Gibson and Kim, 2013). Some papers find evidence the poor pay more (Goodman, 1968; Alcaly and Klevoric, 1971; Kunreuther,

---

[45] For a discussion on the FE see Fanzo et al. (2020)

[46] The 'poor pay more' literature is a sub-strand of the poverty penalty literature. For theoretical explanations that detail types of poverty penalties and reasons they might exist, see Mendoza (2011)

1972; Chung and Myers, 1999;), while others find the opposite (Kaufman et al., 1997; Beatty, 2010) . Dillon et al. (2021) examine this question in Tanzania. They find that, even though edible oils are typically cheap enough to not be limited by liquidity constraints, poor households still purchase small amounts, failing to take advantage of bulk discounting, yet could save if they were reduce purchase frequency. Using a different data set than Dillon et a. (2021) and addressing the problem using various goods in hedonic price regressions, Sauer et al. (2021) find the poor pay less in Tanzania, which stands in contrast to Dillon et al.'s (2021) findings.

Alternatively, store choice literature often focuses on the source from which shoppers obtain their food but pays less attention to what drives the prices they face or whether shopper behavior impacts those prices. Previous work has categorized shopping behavior by motivation for shopping (Jayasankaraprasada and Kathyayani, 2014), their preference for traditional or modern retailers (Hai Tran and Sirieix, 2020), and economic characteristics and eating patterns (Hino, 2014). While price is considered in various parts of these papers, it is typically only explored as a determinant of store choice and not as an outcome of shopping strategies. In fact, poor households may choose specific outlets due to prices available, engaging in local arbitrage behavior (MacNeil, 2018; Kaiser et al., 2019; Darko et al., 2013).

We attempt to bridge the gap between low-income food price literature and consumer shopping behavior, by including a more detailed assessment of household shopping behavior as a driver of prices paid in our analysis. To do this, we first characterize household shopping behavior, or food procurement styles, into 4 main types using the spatial and temporal dimensions: 1) High-frequency hyper-local shoppers, 2) Low-frequency hyper-local shoppers, 3) Low-frequency spatially extensive shoppers, and 4) High-frequency spatially extensive shoppers. This allows us to not only confirm the existence of temporal savings described by Dillon et al. (2021) in a different

context, but also examine whether there is a spatial component to savings available to poor households.

To our knowledge this is the first paper to examine shopping behavior along these dimensions to describe determinants of food costs. We answer the following three questions: 1) What impact does the food procurement style have on the prices paid by consumers? 2) Does our answer to the previous question change with the method we use? 3) Is there a spatial component to savings available for food prices?

The rest of the paper proceeds as follows. In the next section we discuss the data and methods used in our analysis, followed by a comparison of food procurement styles and the average prices they pay for various categories of food. We then present our results, followed by a brief discussion, and conclude our findings.

3.2 Data and Methods

3.2.1 Data

Our main data come from a recent survey conducted by Michigan State University (MSU) in low-income areas of Nairobi, Kenya, as part of an assessment of the Marketplace for Nutritious Foods program run by the Global Alliance for Improved Nutrition (GAIN) to examine food environments and consumer diets in five low-income neighborhoods of Nairobi, Kenya (Gatina, Kabria, Kangemi, Kawangware, and Kibera). The GAIN data allow us to investigate our research questions in detail, as we have GPS coordinates for the household, the area where they do most of their shopping (including coordinates specific to every supermarket), as well as detailed food

purchases, which include type of outlet and distance of certain outlet types from the respective household.

The area of the surveyed neighborhoods was divided into 129 segments of roads containing food outlets, of which 80 were selected for the survey. Within each segment a census of every outlet was conducted on the presence of food on offer across 2,388 stores and 12 store types. To link consumer data to food environments, consumers were randomly recruited among those shopping in randomly selected *dukas* in each segment. This approach was based on information from key informant interviews (Tschirley et al., 2022) that nearly all consumers use dukas on a nearly daily basis in this region of Nairobi. Thus, the consumer segment was designed around a random sample of 321 dukas, from which 1286 total consumers were chosen at random (4 selected with two replacements per duka). We focus on the expenditure and shopping modules of the consumer survey.

The expenditure module captures household food consumption, which includes purchases, own production, and gifts. These data include reported 7-day consumption of 91 food items (including a catchall "other" category). Nearly all consumption came from purchases, with gifts and own production accounting for less than 1%.

The shopping module asked whether households shop at each of a list of nine types of outlets during a typical month. Then for each type that the household uses (and others not listed that they could specify) the main shopper was asked (a) the distance from their home to the shop, (b) how often they shop there in a month, (c) whether they typically buy food from each of 19 food categories and (d) the total amount typically spent per trip. From this module we generate variables that represent our spatial and temporal dimensions of food procurement to categorize shoppers.

For our temporal dimension we use the number of shopping trips to any outlet over a typical month. For our spatial variable, we use the average distance traveled per trip to any outlet.

Most of our analysis is based on the 1274 households that have complete data; price analysis uses only households that consumed an item, which numbered at least 198 in each case,

We use the expenditure module to construct unit prices. Of the 28,655 total purchase observations in the data, 10,992 were in standard units (kg/gram or liter/ml), 3,621 were in semi-standard units (cups, teaspoons, tablespoons, etc.), and 12,491 were in fully non-standard units (bunches, handfuls, piles, etc.). See table A4 for the share of each unit type for the 20 most consumed items.

We convert semi-standard and non-standard units to kg or liters using conversion information from various internet sources; see table A5 for a listing of these factors and their sources. Where we could not find internet sources listing conversion rates, we did the following. First, we used price data collected from shops in the survey area but outside the consumer survey, to compare food item sizes sold in standard units (kg or liters). Second, we use various percentiles of those food item size to estimate the usual purchase quantity of the given non-standard unit, using our best intuition for the choice of percentile. For items sold as packets or pieces (38.7% of all observations) we used the median size (in g or ml) of the item sold within this supplemental price data, and used that size to compute a standard unit price. We chose the median value since it is likely that the commonly purchased form of an item in lower income areas is likely to be a smaller size such as packets. This is evident in studies like Dillon et al. (2020) where low-income households did not exploit bulk discounting. For tea bags and sachet (0.03% of all observations) we used the 25th percentile size (in g or ml) of the given item sold. It is more likely that tea bags or sachets are commonly sold as a package containing multiple items, so the 25th percentile

quantity may be closer to the true size. For bunch and bundle (11.9% of all observations), we used the 75th percentile size (in g or ml) of the given item sold, since bunches and bundles may be larger than the most commonly purchased quantities, in the survey region. For units listed as whole (0.04% of all observations) we used the maximum quantity (in g or ml) of the given item sold, since these items are very rare in these data, and are likely to be toward the higher end of purchase sizes.

3.2.2   Methods

3.2.2.1   Food Procurement Styles

There is much evidence that source of food procurement is related to various factors that influence shopping behavior such as diet and food quality available (Rahkovsky and Snyder, 2015; Krukowski et al., 2013), convenience (He et al., 2012; Ambikapathi et al., 2021), marketing (Chandon and Wansink, 2012; Cairns, 2019), store quality and loyalty (Das, 2014), and cost of food items (Iton, 2015; Balaji, 2017). However, using outlet type to segment shopping strategies created unnecessary complexities in our model, so we choose to focus on the two dimensions (spatial and temporal) directly related to our research questions.

The spatial dimension is the spatial extent a consumer travels to purchase food, which is largely dependent on mobility[47]. The temporal dimension is the frequency of shopping trips or purchases of a given item, which is related to both store choice and nutrition related health outcomes (Minaker et al., 2016), and inversely related to purchase size (Tripathi and Sinha, 2006). We use our spatial and temporal shopping dimensions to generate four shopping styles:

1) High-frequency-local
2) Low-frequency-local
3) Low-frequency spatially extensive
4) High-frequency spatially extensive

This allows us to examine whether increased shopping frequency (or smaller purchase sizes) is related to lower prices and whether shoppers that travel outside of their local FE face different prices. For temporal dimension we use sum of the frequencies households state they go to each type of outlet. Survey answers and our approximate values are given below:

- Daily= 30
- Few times per week=15
- Once a week=4
- 2-3 times a month=2.5
- Once a month=1
- Less than once a month=0.5

---

[47] Poor households generally have few modes of travel available to them, and typically live in areas with underdeveloped infrastructure, restricting their ability expand their spatial shopping extent. This limited mobility is often considered a poverty penalty (Andaleeb, 1995).

For the spatial dimensions we use the average distance per trip. This is calculated by using the usual distance households travel to a given outlet. We weight this distance by the frequency households travel to a given outlet in a month divided by the total shopping trips per month.

Next, we use the median values of each variable to separate local/extensive, infrequent/frequent, which gives us 4 nearly equal groups (See table 25). While it would be ideal to separate each of these categories by the source of procurement as well, we would have too many groups for us to make meaningful comparisons. So, to incorporate the source of procurement, we will examine store choice differences within these categories.

3.2.2.2 Regression Analysis

The opposing conclusions regarding the existence of a food-price poverty penalty suggest that prices paid by poor households depend not only on the characteristics of local food environments but also on the methods used to examine the question. For instance, Hansen et al. (2004) show that using different metrics to measure store choice can affect the significance of the impact of distance in their analysis. Thus, we employ three different sets of regressions using unit values as our dependent variable to proxy for prices, and various household characteristics to control for quality differences.

- First, we estimate unit price OLS regressions and examine the existence of bulk discounts for the entire data set, food categories, and specific food items.

- Second, we estimate the relationship between price differences and distance between households by constructing a dyadic data set from our household level data. This is done to examine whether there are spatial relationships in price differences.

- Third, we look at determinants of various aspects of price search performance to compare the prices paid by households to the best they could do given the prices in these data, and we also compare how they perform relative to neighborhood and survey segment average prices[48].

A challenge in using unit values is the potential for quality differences to drive a portion of the price differences. So, we follow Sauer et al. (2021) and include various household socio-economic and demographic controls. The first set of variables includes measures to account for the household head and main shopper characteristics. The sex, age, and education of either the household head or main shopper could influence the knowledge of prices within the food environment, as well as possession and use of devices such as smart phones to find such information.

Purchase behavior is also likely dependent on household composition. So, we control for the size of the household, the average age of the household, and per capita education, all of which are likely to be related to the amount and types of food purchased. We also account for heterogeneity in food price knowledge by including the total number of shoppers as well as the number of household members employed in the food sector.

---

[48] Results were also estimated using the continuous spatial and temporal variables as regressors in place of the group indicators. These results were not included, but are available upon request.

Lastly, we control for quality differences that might be due to differences in income or wealth. We use the value of total food consumption to approximate income, and a wealth index calculated using principal component analysis on the asset ownership indicated in the poverty module of the survey. We also control for ownership of any type of motorized vehicle.[49]

To account for differences in food environments between households we include controls based on the food environment census conducted within the survey area. The census contains GPS coordinates for every food outlet as well as the type of outlet. We construct two variables to describe the food environment around each household. First, we use the average distance households traveled to the duka where they were surveyed (0.27km) and calculate the number of shops within that radius around the household. This approximates shop density for each household by each outlet type. The second set is a measure of distance to each outlet type. We use the average distance, by outlet type, of the 5 outlets nearest to the household. This accounts for travel distance required for households to visit various shop types to search for the affordable and desirable food items.

To capture seasonal variation in prices we include indicators for the month the respondent completed the survey (which occurred over the course of six months). We might expect prices to vary in relation to the production cycle of a given food item. In addition, we include indicators for each neighborhood to capture any overall regional differences in prices.

---

[49] While we would have like to use these household variables as instruments to predict the shopper type, few showed high enough F-statistics in single endogenous regressor and single instrument equations. Furthermore, when we attempted to include those instruments that did have high enough F stats in an instrumental variable regression with multiple endogenous regressors, we could not find a combination to produce sufficiently high F-statistics that would indicate instrument relevance.

Our first set of regressions are OLS unit price regressions to show there are bulk discounts and differences in prices paid for various items by shopping style. We use the log of unit prices as the dependent variable, where prices are constructed as unit values: cost/quantity purchased. In addition to unit prices, we construct expenditure-weighted average prices for baskets of food items at the household level. This allows us to compare prices paid by households that may have purchased different items. Our chosen specification includes independent variables both in logs and levels[50]:

$$\ln(price_i) = \alpha + \beta' S_i + \gamma' V_i + \delta' X_i + \tau_i + \mu_i + \varepsilon_i \qquad\qquad 6$$

where, $price_i$ is the unit price paid by household i, S_i is a vector representing indicators for shopper categories, V_i is a vector of control variables in log form, X_i is a vector of control variables in linear form, τ_i represents neighborhood fixed effects, μ_i represents survey month fixed effects, and ε_i is the idiosyncratic error. We estimate this model using the full set of consumption data as well as subsets by food category. This allows us to determine if bulk discounts exist generally or for specific types of food, and whether different shopper types pay different prices.

Our aim is to quantify spatial price savings, but previously described methods only tested for differences in prices paid by shopping typologies. Therefore, we first must show the existence of spatial savings before we can quantify them. To do this, we construct a dyadic dataset using

---

[50] IV was not feasible due to the lack of sufficient instruments. We tested many potential candidates and when relevant (high enough F stat) and more than one instrument used, the Hansen J test of overidentifying restrictions was rejected. When we failed to reject, the F stat was too low. Instead, we incorporated potential instruments as controls (this helps us control for quality differences) since they were household level variables.

unique household pairs. We create the dyadic variables for every possible unique combination of households (1274 households makes roughly 810,901 pairs). We use the geographic straight-line distance between households to explore whether there are spatial relationships for price differences. We then take the absolute difference in prices paid by household pairs for the 20 most consumed food items. We construct the absolute difference of prices between households to explore the existence of spatial price savings in our second set of regressions,

$$lnpricediff_{ij} = \ln\left(abs\left(price_i^n - price_j^n\right)\right) \qquad 4$$

This allows us to estimate a gravity-style model, using each household as a representative point for the local food environment. In place of the mass variable used in traditional gravity models (typically GDP) we use count of shops within a radius around each household. To best represent the local food environment of these household we use the average distance traveled to the survey shop where households encountered the enumerator – 0.27 km. This approximates the average willingness to travel for daily shopping needs for households in these data. Since most shopping occurs within this very small radius around one's home, we use the price differences for food items purchased by households to approximate differences in their respective food environments. We also include the average distance of the 5 nearest outlets of each outlet type to further control for variation in local food environments. Our model for our third set of regressions is thus,

$$\begin{aligned}
\ln\left(price_{ij}\right) = {} & \alpha_0 + \alpha_1 \ln\left(DIST_{ij}\right) + \alpha_2 \ln\left(shopcount_{ij}\right) + \beta_1' S_i + \beta_2' S_j + \gamma_1' V_i \qquad 5 \\
& + \gamma_2' V_j + \delta_1' X_i + \delta_2' X_j + \tau_i + \tau_j + \mu_i + \mu_j + \varepsilon_{ij}
\end{aligned}$$

Where $price_{ij}$ and $shopcount_{ij}$ are absolute value of the differences of price and total retail outlet count between households, respectively. For the total retail count difference, we only include outlets that are likely to sell the specific food item for which we are taking the price difference. We include the food environment controls in log form.

In our third and final set of regressions, we explore the effectiveness of various shopping strategies at achieving savings by using two different measures. The first measure is the deviation from the average price of the survey segment or neighborhood that the household is within:

$$pricedev_i^n = avgprice_C^n - price_i^n \qquad\qquad 7$$

Which is a measure of the prices paid by households relative to the average in their location (survey segment or neighborhood). Negative means the household paid less than the average[51]. This gives us a measure of how households compare to their neighbors in the prices they pay. The second measure we use is called the price search effectiveness (PSE), which is the ratio of (1) the difference between the maximum the house could have paid and how much they actually paid and (2) the difference between the max the household could have paid and the minimum they could have paid[52]. This measure thus shows how much of the maximum savings available to the household that the household captured. For all N food items our PSE is,

---

[51] We use a similar idea to Binkley and Chen (2016), however instead of focusing on chains or types of stores, we use the survey segment and neighborhood averages.
[52] We develop this measure from Gauri et al. (2008). In that paper, the authors had data on individual shopping trips, so we modified our measure accommodate our data limitations.

$$PSE = \frac{\sum_{i=1}^{N} P_{max}Q_i - \sum_{i=1}^{N} P_{actual}Q_i}{\sum_{i=1}^{N} P_{max}Q_i - \sum_{i=1}^{N} P_{min}Q_i} \hspace{2cm} 8$$

and for a specific food item, n, we use,

$$PSE_i^n = \frac{P_{max}^n Q_n - P_{actual}^n Q_n}{P_{max}^n Q_n - P_{min}^n Q_n} \hspace{2cm} 9$$

This is a measure of the prices the household paid relative to the best- and worst-case prices within the survey area. We focus on the top 20 consumed items for item specific variables and include all food items for general price measures.

We regress the household's price deviation measure and its measure of price search effectiveness on the following specification at the household level:

$$P = \alpha + \beta' S_i + \delta' X_i + \tau_i + \mu_i + \varepsilon_i \hspace{2cm} 10$$

where P is a measure of price search effectiveness, and all other variables are as previously defined. This gives us a measure of how well households were able to find lower prices as determined by these data. A higher value implies a greater ratio between actual savings and maximum savings, and thus more effective price search strategies.

3.3   Food Procurement Styles

Table 25 displays group means of selected household characteristics by food procurement style. The dominant finding from the table is the great similarity across households in each shopping

category, seen in three ways. First, despite categorizing by frequency and spatial extent of shopping, nearly all shoppers are extremely local and frequent: even the most extensive shoppers travel on average only about half a kilometer (0.44 – 0.61) for their purchases while the most local travel less than one-tenth of a kilometer, and the least frequent shoppers still make about 60 separate purchases per month (meaning 60 visits to different shops) while the most frequent make about 100.

Second, the groups are nearly demographically identical[53], with little if any meaningful differences in age of the household head, age or gender of the main shopper, household size, literacy, and education. Even modes of transport for shopping differ very little: 94% - 96% of extensive shoppers walk for their shopping compared to 98% for local shoppers.

Three main differences emerge. First, the most local and infrequent shoppers have the lowest wealth and spend the least on food (an indicator of total income) while those that shop most extensively and frequently have the highest wealth and spend the most on food. Second, extensive shoppers spend the most on food away from home (around Ksh 1,200 per month compared to around Ksh 800 for local shoppers). Finally, extensive shoppers rely the most on large-format supermarkets, spending 13%-14% of their total food expenditure in these outlets compared to 5%-6% for local shoppers. Note, however, that even for these more extensive shoppers, dukas, small-format supermarkets (which are more densely and locally distributed) and other local outlets such as *mama mbogas* are far more important than large-format supermarkets.

---

[53] This mitigates any price disparities that might arise due to socio-economic differences, such as those found by Graddy (1997).

Table 25: Mean Characteristics for Shopper Categorization

Variables	Local - Infrequent	Local - Frequent	Extensive - Infrequent	Extensive - Frequent
Number of households in group	361	273	267	373
Wealth and Food Expenditure				
Wealth index (0-1)	0.29	0.33	0.32	0.35
Total HH pc monthly FAFH expenditure (ksh)	761	821	1,294	1,166
Total HH pc monthly food shopping expenditures (ksh)	4,221	6,247	5,843	9,838
Average total purchase per shopping trip (ksh)	1,200	1,436	1,570	2,000
HH Demographics				
Age of HH Head	33.3	34.6	35.0	34.9
Female HH Head (proportion)	0.60	0.51	0.58	0.49
If HH has a male head, he …				
is self-employed (proportion)	0.09	0.09	0.11	0.13
has salaried employment (proportion)	0.20	0.24	0.20	0.26
Age of Main Food Shopper	30.8	31.2	32.9	31.7
Female Main Food Shopper (proportion)	0.98	0.99	0.93	0.95
Main shopper …				
is self-employed (proportion)	0.34	0.33	0.34	0.40
has salaried employment (proportion)	0.24	0.23	0.29	0.23
Household Size	3.3	3.6	3.3	3.7
Household Literacy and Education				
HH Literacy Rate (proportion)	0.87	0.84	0.88	0.86
Per capita Education of HH	9.7	9.4	10.5	9.7
Owns a Car or Motorcycle	0.10	0.13	0.15	0.14
Frequency-weighted Mode of Travel				
Walking (weighted proportion)	0.981	0.982	0.959	0.938
Bus or Other Public Transportation (weighted proportion)	0.008	0.009	0.028	0.036
Other (weighted proportion)	0.011	0.009	0.013	0.025
Food Shopping Behavior				
Frequency-weighted distance traveled for food shopping (km, weighted)	0.06	0.07	0.44	0.61
Shopping frequency across all outlets	62	99	61	106
FAFH consumption frequency	13	18	15	21
Number of unique outlet types used per month	4.1	5.6	4.6	6.1
Market Shares for Outlet Types				
Large Supermarket chain	0.05	0.06	0.13	0.14
Small-format supermarket	0.14	0.22	0.12	0.27
Duka or kiosk	0.46	0.32	0.38	0.18
Mama Mboga	0.11	0.10	0.08	0.07
Market Place	0.10	0.13	0.16	0.15
Other	0.14	0.18	0.14	0.19

Notes: Local<0.13km and Extensive>=0.13km, using frequency-weighted distance. Infrequent<82.43, Frequent>=82.43, estimated shopping trips per month

Source: GAIN demographic survey from the consumer segment

The great similarity across these groups suggests two things. First, large and systematic price differences may be hard to find. Second, those differences that we do find are more likely to

be related to (admittedly small) differences in the extent and frequency of shopping, and resulting differences in the food environments that they access, than to differences in demographics.

3.4   Price Indices

In Table 26 we display indices of prices paid by each shopper type. These are expenditure weighted averages of prices (ksh/g). There are two prominent points worth discussing. The first is that there are many categories with little price variation, which may indicate similar or efficient markets for those goods. Second, differences are generally substantial when present. It is difficult to discern whether these differences are the result of seeking specific goods, due to differences in shopping strategies, or due to different food environments.

Table 26: Indices of Prices Paid by Households by Food Category (ksh/g)

Variables	Local - Infrequent	Local - Frequent	Extensive - Infrequent	Extensive - Frequent
Price Index of Food Category				
Cereals	0.12	0.10	0.12	0.10
Roots, Tubers, Plantains	0.05	0.05	0.05	0.05
Legumes and Nuts	0.34	0.19	0.22	0.19
Fruits	0.06	0.06	0.06	0.06
Vegetables	0.02	0.03	0.03	0.02
Oils and Fats	0.15	0.16	0.16	0.16
Dairy	0.18	0.16	0.16	0.16
Meat, Fish, and Eggs	0.35	0.30	0.58	0.41
Snacks, Drinks, and condiments	0.47	0.52	0.60	0.47

Notes: Local<0.13km and Extensive>=0.13km, using frequency-weighted distance. Infrequent<82.43, Frequent>=82.43, estimated shopping trips per month

Many staples show little price variation at the average of our 4 groups. For example, roots, tubers, and plantains show indistinguishable differences at 2 decimal points (0.05 ksh/g), which implies near identical average prices between shopping strategies. We see large price differences for items such as meat and nuts. Local/infrequent shoppers pay almost double per gram of legumes

119

and nuts compared to other groups (0.34 vs. 19-22 ksh/g). Spatially extensive shoppers pay more

for meat (0.41-0.58 vs. 0.3-0.35 ksh/g). Extensive/infrequent shoppers pay the most for snacks

(0.6 ksh/g), followed by local frequent shoppers (0.52 ksh/g).

3.5   Results

We find some evidence that potential savings exist over space, but also that households may

already do rather well at finding suitable prices given their preferences and constraints. To

understand this point, we start by looking at our first set of regressions in Table 27, where the

dependent variable is unit prices by food category.

We include two subsets of each set of regressions. In one subset, we include dummy variables

for shopper types or food procurement styles. The other set replaces category dummies with the

spatial and temporal variables we use to define those groups. The first column presents the

regression using all available consumption data, while the remaining columns show results by food

category. Our results highlight two important facts. First, bulk discounting exists across all items,

and the estimates are highly significant in every regression. Second, we find food procurement

style impacts the prices households pay depending on the food categories being purchased.

In the first column, using all available data on food purchases, we find that bulk discounting

is present (-0.71), and our estimates are highly significant. We also see that both types of frequent

shoppers pay less than infrequent shoppers, which is implied by the negative and significant

coefficients on the respective group indicator variables (-0.033 and -0.034). This suggests that

overall, frequent shoppers tend to pay less for their food but that moving outside your local area

for shopping generates no reliable savings. This is echoed in the bottom portion of the table, where frequency has a significant negative coefficient, but distance is insignificant.

When we restrict the sample by food category, we find our results depend on the food category. Local/frequent shoppers pay less for legumes/nuts (-0.10) and oils/fats (-0.11), while extensive/infrequent shoppers pay more for dairy (0.09), and extensive/frequent shoppers pay less for snacks/drinks (-0.14). Our continuous variables tell a similar story: shopping frequency is always negative and is significant for cereals (-0.04), legumes/nuts (-0.11), dairy (-0.17), and snacks/drinks (-0.15). Distance traveled is significant for three categories but is positive for two of those: fruit and dairy prices (0.02 and 0.04, respectively).

Table 27: Household Level Unit Price Regressions by Food Category

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All data - Unit Price	Cereals	Roots, Tubers, Plantains	Legumes and Nuts	Fruits	Vegetables	Oils and Fats	Dairy	Meat, Fish, and Eggs	Snacks, Drinks, and condiments
Shopping Category Dummies: Dep Var - ln(price)										
Local - Frequent	-0.0327*	-0.0144	-0.0603	-0.103*	-0.0415	-0.0230	-0.107***	-0.0777	-0.0274	-0.0228
Extensive - Infrequent	0.0122	0.00231	0.0303	-0.0210	0.0383	-0.00855	0.00890	0.0863*	-0.0109	0.00219
Extensive - Frequent	-0.0344**	-0.0233	-0.00233	-0.0765	0.0164	-0.0348	-0.0175	0.0517	-0.0412	-0.140***
ln(qty)	-0.713***	-0.502***	-0.699***	-0.811***	-0.409***	-0.631***	-0.889***	-0.679***	-0.839***	-0.668***
ln(distance to survey shop)	0.00705	-0.00216	-0.0147	0.00733	-0.00667	0.0180**	0.0119	-0.0235	0.00633	0.00240
Observations	26,811	6,191	1,201	1,375	2,902	5,590	1,575	1,351	2,455	4,171
Frequency and Distance Variables: Dep Var - ln(price)										
ln(Monthly Shopping Frequency)	-0.0660***	-0.0399*	-0.0254	-0.112*	-0.0142	-0.0459	-0.0634	-0.168***	-0.0317	-0.150***
ln(Frequency weight distance traveled per trip)	0.00176	0.00368	0.0101	0.000377	0.0212**	-0.000160	0.0115	0.0447***	0.000208	-0.0288**
ln(qty)	-0.713***	-0.502***	-0.698***	-0.811***	-0.409***	-0.631***	-0.888***	-0.676***	-0.839***	-0.668***
ln(distance to survey shop)	0.00718	-0.00208	-0.0143	0.00736	-0.00620	0.0179**	0.0118	-0.0243*	0.00657	0.00289
Observations	26,811	6,191	1,201	1,375	2,902	5,590	1,575	1,351	2,455	4,171

Notes: Dependent variable is price household paid for food item, and we include only observations in their repsective food category for each regression beyond the first column

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Next, we examine determinants of prices households paid for specific food items. We look at the top 20 most consumed items by share of all consumption, and only include households that recorded a purchase of the respective item. These results are split between tables 28 and 29. There are two important points worth noting. First, bulk discounting is again present across the entire set of regressions to varying degrees (-0.2 to -0.94). Second, frequent shopping is generally related to lower prices, while our results suggest that spatially extensive shopping may be related to higher prices where we estimate a significant relationship.

Our second result bears a closer look. First, we look at the estimates from the first subset of these regressions using indicators for shopper types. Local/frequent shoppers pay more for white bread (0.04) and Sukuma wiki (0.06), and less for rice (-0.11); extensive/infrequent shoppers pay more for sugar/honey (0.10) and Sukuma wiki (0.10), and less for wheat chapati (-0.09) and whole maize meal (-0.19); and extensive/frequent shoppers pay more for Sukuma wiki (0.13), and less for refined maize (-0.06). In the second subset, using the continuous variables that defined shopper type, we find that frequency is negatively related to prices of unflavored pasteurized milk and rice, while positively related to whole maize meal. However, we only find distance to be positively related to prices where we find significance. This again suggests that frequent shopping is related to lower prices, while spatially extensive shopping is related to higher prices.

Table 28: Household Level Unit Price Regressions by Specific Food Item (top 10 items)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Refined Maize	Unflv Pkg Past Milk	Beef	Rice	Tomato	White Bread	Cooking Oil	Sugar, Honey	Mandazi, cakes, biscuits, pastries	Wheat Chapati
Shopping Category Dummies: Dep Var - ln(price) HHs paid										
Local - Frequent	-0.0171	-0.109	0.00146	-0.109**	-0.0539	0.0448**	-0.0792*	-0.0120	0.0107	-0.0439
Extensive - Infrequent	0.0199	-0.0304	0.0623	0.0297	-0.0394	0.0364	0.0128	0.0997**	0.0421	-0.0939*
Extensive - Frequent	-0.0568*	-0.0223	0.0544	-0.0538	-0.0193	0.0102	-0.0364	0.0382	0.0758	-0.0521
ln(qty)	-0.201**	-0.669***	-0.676***	-0.577***	-0.250***	-0.245***	-0.938***	-0.753***	-0.373***	-0.478***
ln(distance to survey shop)	0.0118	-0.0167	0.00112	0.00179	0.000347	-0.00543	0.00269	0.00347	-0.0122	0.0131
Frequency and Distance Variables: Dep Var - ln(price) HHs paid										
ln(Monthly Shopping Frequency)	-0.0582	-0.120*	0.0474	-0.114**	-0.0458	0.0214	-0.0618	-0.0412	-0.0598	-0.0267
ln(Frequency weight distance traveled per trip)	-0.00510	0.00952	0.0100	0.00627	0.00673	0.00726	0.00603	0.0243*	0.0683***	-0.00609
ln(qty)	-0.200**	-0.670***	-0.676***	-0.579***	-0.248***	-0.245***	-0.937***	-0.754***	-0.368***	-0.481***
ln(distance to survey shop)	0.0119	-0.0169	0.00135	0.00215	-0.000405	-0.00541	0.00288	0.00356	-0.0110	0.0132
Observations	1,039	757	710	1,071	1,187	884	1,238	1,225	883	707

Notes: Dependent variable is price household paid for food item

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 29: Household Level Unit Price Regressions by Specific Food Item (top 11-20 items)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
									Whole Maize	
	Eggs	DryFish	Unflv Frsh Past Milk	Beans	Sukuma Wiki	Bananas	Potatoes	Fresh Fish	Meal	Chicken
Shopping Category Dummies: Dep Var - ln(price) HHs paid										
Local - Frequent	-0.0229	0.193	0.0843	-0.0621	0.0607*	-0.00995	-0.0101	0.132	-0.0372	-0.0815
Extensive - Infrequent	-0.0192	0.115	0.200	0.0452	0.0969**	-0.00366	0.0998	-0.0256	-0.187*	0.0248
Extensive - Frequent	-0.0146	-0.113	0.196	-0.0354	0.127***	-0.00860	0.103	0.188	-0.0277	-0.125
ln(qty)	-0.0744***	-0.914***	-0.508***	-0.846***	-0.264***	-0.203***	-0.736***	-0.728***	-0.237***	-0.690***
ln(distance to survey shop)	0.00807	-0.0101	-0.0233	0.0164	-0.0113	0.00767	-0.0138	0.0336	0.0209	-0.0215
Frequency and Distance Variables: Dep Var - ln(price) HHs paid										
ln(Monthly Shopping Frequency)	0.0199	-0.0497	-0.0314	-0.0318	0.0314	-0.0416	0.0502	0.153	0.141*	-0.126
ln(Frequency weight distance traveled per trip)	-0.00109	-0.0459	0.0457	-0.00348	0.0296**	0.0197	0.0447**	0.0355	-0.0333	0.0329
ln(qty)	-0.0732***	-0.908***	-0.513***	-0.845***	-0.267***	-0.206***	-0.736***	-0.719***	-0.229***	-0.691***
ln(distance to survey shop)	0.00831	-0.00117	-0.0236	0.0165	-0.0106	0.00792	-0.0123	0.0399	0.0218	-0.0198
Observations	819	313	239	781	1,024	803	583	191	255	176

Notes: Dependent variable is price household paid for food item

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Next, we examine the second set of regressions focusing on price differences in a dyadic framework. We use the log of the absolute value of the price difference between two households for each of the top 20 consumed food items as the dependent variable. This gives us a non-directional price difference, giving us a total of 809,627 unique pairs[54]. Just as before we only focus on households that purchased the given food item.

We estimate two models. Independent variables in both include geographic distance between two households, controls for each household just as they might appear in a typical gravity equation, fixed effects for survey month and neighborhood, and a relative retail count variable, which is the absolute value of the difference of the total number of retail outlets that are likely to sell the respective food item within a 0.27 km radius around each household. The second model adds an interaction between our relative shop count variable and the distance between households to explore whether market size and distance have any cross relationships with prices. The distance, quantity, and food environment variables are in logs, while the demographic controls and fixed effects are in levels. Results are found in Table 30, focusing on the distance and outlet count variables.

We find four key results. First, we find a positive and significant coefficient on the distance between households for 7 out of 20 food items. This suggest there is money left on the table spatially for certain types of food, which likely include some level of travel costs, and costs to obtain information about prices. Perhaps households could save if they travel to other areas, but we cannot know for certain given our limited information about costs of moving between areas.

---

[54] We use a non-directional price difference with unique pairs. Otherwise, we would have a completely symmetric data set. This is akin to using bilateral trade flows in a gravity-based trade model, where the direction of the trade flows is irrelevant and only the total amount of trade between two countries matters for the sake of the analysis.

Table 30: Household Level Unit Price Regressions (top 20 items)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Refined Maize	Unflv Pkg Past Milk	Beef	Rice	Tomato	White Bread	Cooking Oil	Sugar, Honey	Mandazi, cakes, biscuits, pastries	Wheat Chapati
Dyadic Model										
ln(dist HH1 - HH2)	0.0847	0.0767	-0.0941	0.191	0.0662	0.179	0.0379	0.192**	0.109	0.0452
ln(diffoutlet HH1 - HH2)	-0.162*	-0.0542	0.371	0.230	-0.0543	-0.0954	-0.0667	-0.0491	-0.428	-0.0471
Dyadic Model with interaction										
ln(dist HH1 - HH2)	0.196*	-0.0327	-0.118	0.137	0.139**	0.126	0.0839	0.248*	0.329	0.0592
ln(diffoutlet HH1 - HH2)	-0.123	-0.101	0.344	0.211	-0.0212	-0.118	-0.0504	-0.0294	-0.335	-0.0412
ln(dist HH1 - HH2) x ln(diffoutlet	-0.0491	0.0572	0.0301	0.0239	-0.0399	0.0269	-0.0206	-0.0249	-0.112	-0.00710
Observations	491,658	156,648	178,411	516,872	523,238	131,031	744,238	699,226	285,011	214,852

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Unflv Frsh Past Milk		Sukuma Wiki				Whole Maize Meal	
	Eggs	DryFish		Beans		Bananas	Potatoes	Fresh Fish		Chicken
Dyadic Model										
ln(dist HH1 - HH2)	0.138	0.0383	-0.0736	0.394****	0.157	1.013****	0.215****	0.00737	0.0216	0.0536
ln(diffoutlet HH1 - HH2)	-0.123	-0.0171	-0.358***	0.0804	-0.271	0.165	-0.0538	-0.127***	0.00399	0.113
Dyadic Model with interaction										
ln(dist HH1 - HH2)	0.157	0.0650	-0.176	0.354****	0.830***	0.765***	0.181***	0.0382	-0.0730	0.294
ln(diffoutlet HH1 - HH2)	-0.115	0.00254	-0.400***	0.0502	0.0279	0.0441	-0.0680	-0.103**	-0.0662	0.342
ln(dist HH1 - HH2) x ln(diffoutlet	-0.00920	-0.0233	0.0560	0.0365	-0.369***	0.142	0.0183	-0.0285	0.0826*	-0.266
Observations	268,539	42,598	27,055	270,774	350,341	284,442	160,880	15,485	29,259	12,134

Notes: Dependent variable is the absolute value of the difference of natual log of price index of food vategories between HH1 and HH2

Robust standard errors in parentheses

**** p<0.001, *** p<0.01, ** p<0.05, * p<0.1

Second, we find an odd relationship with our relative count variable and prices. All the elasticities are estimated to be negative where we find significant relationships. This implies that larger differences in market size among a pair of households are associated with smaller differences in price between those households' FE. Intuitively, one might think that larger markets (i.e., more local shops) would have lower prices than smaller markets from increased local competition, suggesting that bigger differences in market size should drive bigger price differences. However, our results indicate that as market sizes diverge, prices converge.

Third, the interaction term, which estimates the influence that distance and relative market size have together on price differences, is only significant for 2 of the 20 food items, which suggests there is likely no general relationship. We find one estimate is positive and significant (Sukuma wiki), while the other is negative and only marginally significant and practically small. We expect to find a negative relationship here, as the increase in market size might reduce the extent to which distance between markets relates to higher price differences.

Fourth, only considering our significant estimates, the elasticities are highest for items such as bananas, which are likely purchased locally so we might not expect households to travel for greater savings. Beans and potatoes, however, are less perishable, and so may be included in spatial price searches by households but missed due to lack of information or restricted by travel costs.

The third and last set of regressions explores what is typically referred to as price search performance or effectiveness. We attempt to quantify how well shoppers find the lowest prices by using the two measures indicated in the data section. For the first, we use the deviation from the mean price at the neighborhood and survey segment level of aggregation, which gives us a comparison of the price paid by households in comparison to peers in the area where they likely

do most of their food shopping. These results (tables 31 and 32) are similar to some of what we have seen, with 2 notable points.

First, we see the coefficient on the continuous spatial variable is generally positive, which might lead to similar conclusions as in in previous results. However, there are some cases, such as unflavored packaged pasteurized milk, where the spatially extensive food procurement styles are estimated to have negative relationship with the deviation from average prices. We can take this to imply that spatially extensive shoppers pay lower prices for some items, and higher for others, than the average household in their neighborhood or within their survey segment.

Second, frequency is negatively related to household deviations from regional averages except for Sukuma wiki. It is difficult to say specifically whether the distance or frequency are more important in the price households pay for Sukuma wiki since the continuous variables are not significant in either subset of regressions. We do find that all groups pay more than the reference group (local/infrequent), which could be the result of a combination of both spatially extensive and frequent shopping strategies.

The second measure we use is the PSE, which ranges from zero to one, with one being the most effective or successful at finding the lowest prices (tables 33 and 34). We also get mixed results on the performance of shopping strategies and find far less significance in our continuous shopping dimensions than we do in our food procurement style indicators. So, it is difficult to say with certainty whether a specific dimension of shopping strategy has a larger impact on the prices those household paid for specific food items.

While spatial price savings exist, we cannot conclude that spatially extensive shoppers pay lower prices. In fact, in cases where we do find a significant difference between groups, we see that spatially extensive shoppers pay more than local shoppers, with some exceptions.

Table 31: Household Level Unit Price Regressions (top 10 items)

	(1) Refined Maize	(2) Unflv Pkg Past Milk	(3) Beef	(4) Rice	(5) Tomato	(6) White Bread	(7) Cooking Oil	(8) Sugar, Honey	(9) Mandazi, cakes, biscuits, pastries	(10) Wheat Chapati
Price minus average survey segment price										
Local - Frequent	-0.000799	-0.417	-0.0821	-0.129	-0.00341*	0.00478	0.00158	0.133	0.00607	-0.00107
Extensive - Infrequent	0.00209	-0.512	1.285	0.124	-0.00269*	0.0180*	-0.0251	0.753	0.0133	-0.00182
Extensive - Frequent	-0.00411*	-0.603*	-0.394	-0.187	-0.00212	0.00603	-0.0340	0.128	0.0314**	-0.0105
distance to survey shop	0.000211	-0.227	0.639	-0.124	-0.00107	-0.00581*	-0.0242	0.229	0.0113	0.00627
Price minus average survey segment price										
Monthly Shopping Frequency	-6.11e-05	-0.00607	-0.000799	0.00209	-5.59e-05***	-6.49e-05	-0.000194	0.00246	9.69e-05	-5.11e-05
Frequency weight distance traveled per trip	0.00124	-0.0602	-1.064	-0.194	0.000846	0.00312	-0.00574	-0.131	0.0198***	-0.00620
distance to survey shop	6.79e-05	-0.184	0.522	-0.124	-0.00113	-0.00603*	-0.0247	0.222	0.0115	0.00574
Price minus average neighborhood price										
Local - Frequent	-6.67e-05	-0.340	0.0508	-0.247	-0.00384*	0.00478*	-0.0138	0.0143	0.00596	-0.00152
Extensive - Infrequent	0.00274	-0.661*	2.534	0.211	-0.00294*	0.0186*	-0.0223	0.843	0.0150	-0.00441
Extensive - Frequent	-0.00492*	-0.759*	0.369	-0.182	-0.00258	0.00767	-0.0340	0.129	0.0358***	-0.00498
distance to survey shop	-0.00171	-0.281	1.769	-0.160	-0.00134	-0.00704*	-0.0240	0.210	0.0106	0.00232
Price minus average neighborhood price										
Monthly Shopping Frequency	-7.60e-05	-0.00653	0.00569	-6.06e-06	-6.05e-05***	-6.50e-05	-0.000350	0.000481	0.000116	-1.66e-05
Frequency weight distance traveled per trip	0.00103	-0.154	-0.0130	-0.0589	0.000879	0.00428	0.00113	-0.0107	0.0243***	-0.00202
distance to survey shop	-0.00194	-0.230	1.764	-0.158	-0.00139	-0.00724*	-0.0243	0.206	0.0110	0.00226
Observations	1,040	757	710	1,072	1,188	884	1,239	1,226	884	708

Notes: 1. Dependent variable is the price hosuehold paid for each item minus either the average price paid within the households survey segment or neighborhood.

2. We estimate the model using two different specifications: 1) we use the shopper categories as our variable of interest, and 2) we use continuous temporal and spatial variables.

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 32: Household Level Unit Price Regressions (top 11-20 items)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Unflv Frsh Past Milk		Sukuma Wiki			Fresh Fish	Whole Maize Meal	
	Eggs	DryFish	Past Milk	Beans	Wiki	Bananas	Potatoes	Fish	Meal	Chicken
Price minus average survey segment price										
Local - Frequent	-0.00223	0.179	-0.00125	-2.705	0.00150**	-0.00433	-0.00410	0.0534	0.00354	6.057
Extensive - Infrequent	-0.00166	0.136	-0.000634	-1.063	0.00237**	0.00391	-0.00966	0.0321	-0.00591	11.64
Extensive - Frequent	-0.00251	-0.0125	-0.00707	-2.314	0.00230***	-0.00281	0.00319	0.0471	-0.00595	8.054
distance to survey shop	0.000814	-0.0921	-0.0151**	1.658	-8.54e-05	0.000514	0.0304	-0.0168	0.00320	-0.847
Price minus average survey segment price										
Monthly Shopping Frequency	-2.25e-05	-0.00172	-0.000274	-0.0309	6.03e-06	-3.56e-05	9.71e-05	0.000776	7.19e-05	0.0129
Frequency weight distance traveled per trip	0.00332	0.0835	-0.00580	-0.677	0.000408	0.00533	0.00399	0.0297	-9.81e-05	0.930
distance to survey shop	0.00122	-0.0737	-0.0177***	1.625	-0.000106	0.000844	0.0312	0.000376	0.00327	-0.792
Price minus average neighborhood price										
Local - Frequent	-0.00234	0.170	0.00680	-3.379	0.00151**	-0.00424	0.000998	0.0156	0.00389	-6.921
Extensive - Infrequent	-0.000665	0.260*	0.0125	-2.343	0.00296**	0.00255	-0.00886	0.0446	-0.0168**	7.913
Extensive - Frequent	-0.00237	0.0584	0.00449	-2.474	0.00303***	-0.00395	0.00783	0.0669	-0.00761	1.153
distance to survey shop	-0.000955	-0.0351	-0.0256*	2.442	-0.000530	0.00145	0.0346	-0.0101	0.000493	-4.215
Price minus average neighborhood price										
Monthly Shopping Frequency	-3.07e-05	-0.00174	-0.000221	-0.0248	6.54e-06	-1.39e-05	0.000193***	0.000124	8.05e-05	-0.0626
Frequency weight distance traveled per trip	0.00365	0.156*	-0.00832	-0.642*	0.000542	0.00506	0.00357	0.0479	-0.00136	1.166
distance to survey shop	-0.000559	-0.0172	-0.0305**	2.420	-0.000544	0.00187	0.0355	0.0114	-0.000297	-3.178
Observations	820	313	240	781	1,025	803	583	191	255	176

Notes: 1. Dependent variable is the price hosuehold paid for each item minus either the average price paid within the households survey segment or neighborhood.

2. We estimate the model using two different specifications: 1) we use the shopper categories as our variable of interest, and 2) we use continuous temporal and spatial variables

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 33: Household Level Unit Price Regressions (top 10 items)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Refined Maize	Unflv Pkg Past Milk	Beef	Rice	Tomato	White Bread	Cooking Oil	Sugar, Honey	Mandazi, cakes, biscuits, pastries	Wheat Chapati
Price Search Effectiveness (PSE) measure										
Local - Frequent	0.000572	0.00605	-9.48e-05	0.00206	0.0108*	-0.00411*	0.00121	-0.000152	-0.00320	0.000774
Extensive - Infrequent	-0.00489	0.0121*	-0.00422	-0.00176	0.00668	-0.0146*	0.00195	-0.00703	-0.00855	0.00214
Extensive - Frequent	0.00833*	0.0137*	-0.000622	0.00152	0.00627	-0.00609	0.00403	-0.00104	-0.0199***	0.00242
distance to survey shop	0.00281	0.00471	-0.00295	0.00134	0.00366	0.00573*	0.00472	-0.00177	0.00126	-0.000622
Price Search Effectiveness (PSE) measure										
Monthly Shopping Frequency	0.000131	0.000119	-9.56e-06	1.62e-08	0.000180***	5.04e-05	4.63e-05	-3.81e-06	-6.17e-05	4.98e-06
Frequency weight distance traveled per trip	-0.00187	0.00280	2.24e-05	0.000503	-0.00331	-0.00342	-0.000570	9.17e-05	-0.0140***	0.000895
distance to survey shop	0.00321	0.00381	-0.00294	0.00132	0.00378	0.00589*	0.00476	-0.00174	0.000419	-0.000605
Observations	1,040	757	710	1,072	1,188	884	1,239	1,226	884	708

Notes: 1. Dependent variable is the price search effectivenes of the household in comparison to the best they could have done given these data

2. We estimate the model using two different specifications: 1) we use the shopper categories as our variable of interest, and 2) we use continuous temporal and spatial variables.

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 34: Household Level Unit Price Regressions (top 11-20 items)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
			Unflv Frsh		Sukuma			Fresh	Whole Maize	
	Eggs	DryFish	Past Milk	Beans	Wiki	Bananas	Potatoes	Fish	Meal	Chicken
Price Search Effectiveness (PSE) measure										
Local - Frequent	0.00550	-0.0239	-0.00928	0.0169	-0.0182**	0.00579	-0.00261	-0.0124	-0.0195	0.0134
Extensive - Infrequent	0.00318	-0.0295	-0.0147	0.0117	-0.0401***	-0.00278	0.0116	-0.0319	0.0488**	-0.0234
Extensive - Frequent	0.00544	-0.0123	-0.00725	0.0122	-0.0428***	0.00933	-0.0112	-0.0343	0.0247	-0.00481
distance to survey shop	0.00247	-0.00474	0.0269**	-0.0125	0.00907	-0.00268	-0.0491	0.000522	-0.00472	0.00851
Price Search Effectiveness (PSE) measure										
Monthly Shopping Frequency	5.12e-05	0.000161	0.000203	0.000121	-0.000115	-4.82e-06	-0.000285***	-3.08e-05	-0.000223	0.000147
Frequency weight distance										
traveled per trip	-0.0101	-0.0473**	0.00823	0.00317*	-0.00843	-0.00945	-0.00474	-0.0287	0.00529	-0.00281
distance to survey shop	0.00133	-0.00781	0.0324**	-0.0124	0.00912	-0.00358	-0.0502	-0.0114	-0.00195	0.00609
Observations	820	313	240	781	1,025	803	583	191	255	176

Notes: 1. Dependent variable is the price search effectivenes of the household in comparison to the best they could have done given these data

2. We estimate the model using two different specifications: 1) we use the shopper categories as our variable of interest, and 2) we use continuous temporal and spatial variables.

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

3.6 Discussion

Our examination of Nairobi shopping behavior aims to identify spatial and temporal relationships with prices using household shopping strategies based on those two dimensions. We focus on prices paid by consumers. Due to factors such as quality or information differences, these prices exhibit a high degree of heterogeneity. While we cannot fully control for quality differences in unit prices, our controls absorb any differences due to household characteristics. We find mixed results on whether different strategies provide savings for consumers.

In the first set of regressions at the level of food categories, we find strong evidence that increased shopping frequency is related to lower prices paid. This is evident by negative and significant coefficients on the shopper types associated with higher shopping frequency and on the frequency variable when it is included in place of the categories. We also find some evidence of higher prices paid by spatially extensive shoppers for locally purchased items such as fruit and dairy.

This pattern is generally consistent in our second subset of the first set of regressions where we focus on individual food item prices at the household level. Results generally hold-up, though with less significance, when we account for spatial autocorrelation in prices: results still show that shopping frequency is negatively related to prices through direct channels, with no influence through indirect channels, while spatially extensive shopping is generally related to higher prices. It may be the case that spatially extensive search strategies are used to find specific items. Quality variation or specific brand type could drive these differences, or even store loyalty, but this is beyond our ability to test with the available data.

Using our dyadic model, in the second set of regressions we find clear evidence of price savings related to distance between households. We are able to use household locations as proxies for food environments due to the highly local shopping behavior of nearly all households. The coefficient on log of distance between households is positive in 7 of 20 food items, which indicates a positive distance elasticity for absolute price differences. More simply, this means that households that are farther apart show larger differences in the prices they paid. This difference is not always large, which could be a factor on why they exist. It may cost more to travel than what they might save. It is larger for many perishable items that are likely to be purchased from local food outlets, which may indicate a preference to purchase daily-consumed or perishable foods locally. Nonetheless, these results suggest a clear spatial difference in prices.

In our third set of regressions, our exploration of price search effectiveness tells a similar story. Using deviations from neighborhood or survey segment mean prices as our dependent variable, we see some evidence that prices paid by households generally have a positive relationship with spatially extensive shopping and a negative relationship with frequent shopping. There are exceptions in both cases that indicate the most effective strategy is largely dependent on the specific food item. We find similar results using PSE. While there are still price differences in these data that may be exploitable, these results suggest some shoppers likely do benefit heterogeneously from both frequent and spatially extensive shopping. This may indicate that barriers such as limited information or cost of travel restrict the ability of some shoppers to find the best food prices.

So, while spatial savings exist and there are systematic spatial channels through which prices can be influenced, the spatially extensive shopping strategy does not perform the best in terms of prices paid by consumers. On the other hand, shoppers seem to benefit from frequent shopping,

though the extent of benefit is highly dependent on the food item. Maybe households generally already do the best given their preferences and constraints. It is possible that spatially extensive households pay more because they are attempting to satisfy a minimum quality for specific food items.

3.7   Conclusion

In this paper, we explored the food purchase behavior of poor households in Nairobi and its relationship to prices paid. We used data collected by MSU in conjunction with GAIN on household shopping and consumption in 5 low-income neighborhoods of Nairobi.

We categorized households by spatial and temporal dimensions of shopping behavior and created 4 groups of shopping strategies based on those two dimensions. We then explored the relationship between prices households paid for food and membership in one of these 4 groups, using local/infrequent shoppers as the base group.

Our analysis found two interesting results. First, households seem to do as well as could be expected given their likely time, budget, and information constraints. While we did find some evidence of exploitable price differences, our comparisons of the performance of our shopping categories yielded conflicting results that were dependent on the method as well as the food item we examined. Second, the local food environments in which households live may provide competitive prices, which makes any financial gains from spatially extensive shopping strategies minimal. We found that there are substantial spatial price differences, however, there does not appear to be a systematic benefit to those that shop beyond their local food environments.

In contrast with Dillon et al. (2019), we found that frequent shopping was beneficial in these data. One of Dillon et al.'s conclusions was that poor households could save if they waited and purchased in bulk. We did find evidence of bulk discounting, however, frequent shoppers tended to outperform infrequent shoppers. Perhaps, households could save if they exploited economies of scale, but we cannot say anything beyond verifying the existence of bulk discounting. The answer requires further research, and a more detailed survey, which records purchases and distances from each food outlet the household visits.

REFERENCES

Alcaly, R. E., and A. K. Klevorick. 'Food Prices in Relation to Income Levels in New York City'. *The Journal of Business* 44,4(July 1971):380–397.

Ambikapathi, R., G. Shively, G. Leyna, D. Mosha, A. Mangara, C. L. Patil, … N. S. Gunaratna. 'Informal Food Environment is Associated with Household Vegetable Purchase Patterns and Dietary Intake in the DECIDE Study: Empirical Evidence from Food Vendor Mapping in Peri-Urban Dar Es Salaam, Tanzania'. *Global Food Security* 28(March 2021):100474.

Andaleeb, S. S. 'Do the Poor Pay More? A Developing Country Perspective'. *Journal of International Consumer Marketing* 7,2(January 1995):59–72.

Attanasio, O., and C. Frayne. 'Do the poor pay more?' (2006).

Balaji, P. 'Retail Store Choice Behaviour : An Empirical Study on Fruits and Vegetables (F&V) Consumers' (2017):9.

Beatty, T. K. M. 'Do the Poor Pay More for Food? Evidence from the United Kingdom'. *American Journal of Agricultural Economics* 92,3(April 2010):608–621.

Binkley, J. K., and S. E. Chen. 'Consumer Shopping Strategies and Prices Paid in Retail Food Markets'. *Journal of Consumer Affairs* 50,3(November 2016):557–584.

Chandon, P., and B. Wansink. 'Does Food Marketing Need to Make Us Fat? A Review and Solutions'. *Nutrition Reviews* 70,10(October 2012):571–593.

Chung, C., and S. L. Myers. 'Do the Poor Pay More for Food? An Analysis of Grocery Store Availability and Food Price Disparities'. *Journal of Consumer Affairs* 33,2(December 1999):276–296.

Darko, J., D. L. Eggett, and R. Richards. 'Shopping Behaviors of Low-income Families During a 1-Month Period of Time'. *Journal of Nutrition Education and Behavior* 45,1(January 2013):20–29.

Das, G. 'Impacts of Retail Brand Personality and Self-Congruity on Store Loyalty: The Moderating Role of Gender'. *Journal of Retailing and Consumer Services* 21,2(March 2014):130–138.

Dillon, B., J. De Weerdt, and T. O'Donoghue. 'Paying More for Less: Why Don't Households in Tanzania Take Advantage of Bulk Discounts?' *The World Bank Economic Review* 35,1(February 2021):148–179.

Gauri, D. K., K. Sudhir, and D. Talukdar. 'The Temporal and Spatial Dimensions of Price Search: Insights from Matching Household Survey and Purchase Data'. *Journal of Marketing Research* 45,2(April 2008):226–240.
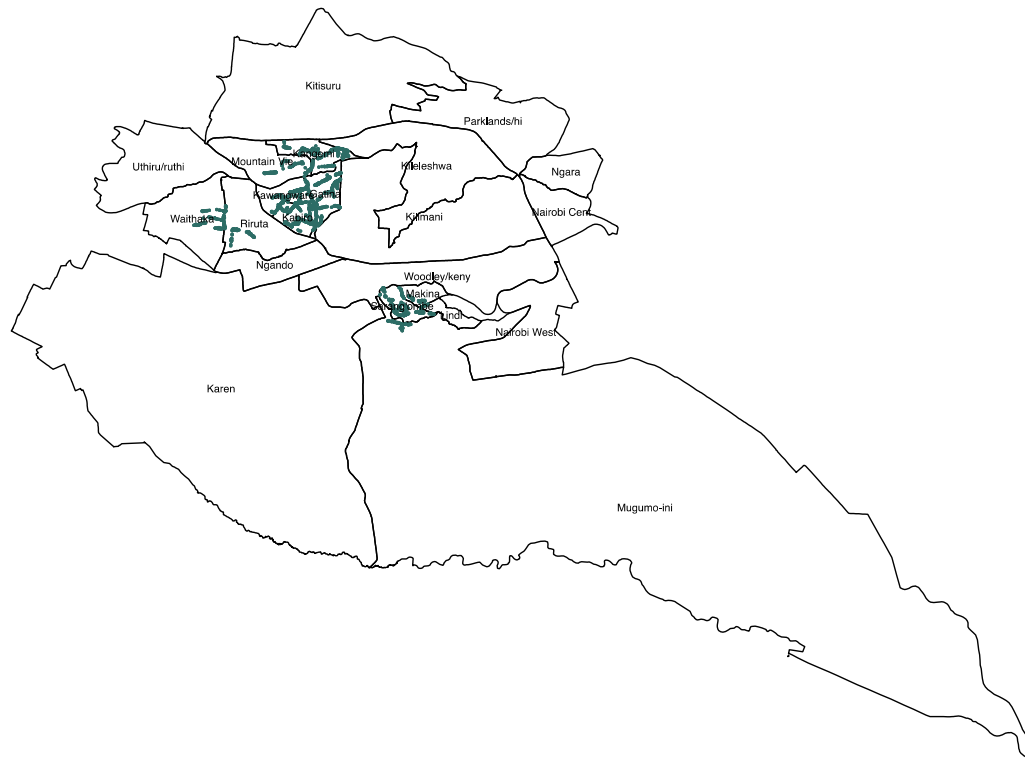
Gibson, J., and B. Kim. 'Do the Urban Poor Face Higher Food Prices? Evidence from Vietnam'. *Food Policy* 41(August 2013):193–203.

Gibson, J., and B. Kim. 'Economies of Scale, Bulk Discounts, and Liquidity Constraints: Comparing Unit Value and Transaction Level Evidence in a Poor Country'. *Review of Economics of the Household* 16,1(March 2018):21–39.

Goodman, C. S. 'Do the Poor Pay More?' *Journal of Marketing* 32,1(January 1968):18–24.

Graddy, K. 'Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area?' *Journal of Business & Economic Statistics* 15,4(October 1997):391–401.

Hai Tran, V., and L. Sirieix. 'Shopping and Cross-Shopping Practices in Hanoi Vietnam: An Emerging Urban Market Context'. *Journal of Retailing and Consumer Services* 56(September 2020):102178.

Hansen, T., F. Cumberland, and H. S. Solgaard. 'How the Measurement of Store Choice Behaviour Moderates the Relationship between Distance and Store Choice Behaviour'. *The Marketing Review* 4.1(2004):99–111.

He, M., P. Tucker, J. Gilliland, J. D. Irwin, K. Larsen, and P. Hess. 'The Influence of Local Food Environments on Adolescents' Food Purchasing Behaviors'. *International Journal of Environmental Research and Public Health* 9,4(April 2012):1458–1471.

Hino, H. 'Shopping at Different Food Retail Formats: Understanding Cross-Shopping Behavior through Retail Format Selective use Patterns'. *European Journal of Marketing* 48,3/4(April 2014):674–698.

Iton, C. W. A. 'Retail Outlet Attributes Influencing Store Choice for Roots and Tubers in Trinidad and Tobago'. *European Journal of Business and Management* 7,15(2015):54–62.

Jayasankaraprasad, C., and G. Kathyayani. 'Cross-format Shopping Motives and Shopper Typologies for Grocery Shopping: A Multivariate Approach'. *The International Review of Retail, Distribution and Consumer Research* 24,1(January 2014):79–115.

Kaiser, M. L., J. K. Carr, and S. Fontanella. 'A Tale of two Food Environments: Differences in Food Availability and Food Shopping Behaviors Between Food Insecure and Food Secure Households'. *Journal of Hunger & Environmental Nutrition* 14,3(May 2019):297–317.

Kaufman, P. R., J. M. MacDonald, S. M. Lutz, and D. M. Smallwood. *Do the Poor Pay More for Food? Item Selection and Price Differences Affect Low-Income Household Food costs* (Economic Report No. AER-759) (p. 27). Economic Research Service, USDA. 1997.

Krukowski, R. A., C. Sparks, M. Dicarlo, J. McSweeney, and D. S. West. 'There's More to Food Store Choice Than Proximity: A Questionnaire Development Study'. *BMC Public Health* 13,1(December 2013):586.

Kunreuther, H. 'Why the Poor May Pay More for Food: Theoretical and Empirical Evidence'. *The Journal of Business* 46,3(July 1973):368–383.

MacNell, L. 'A Geo-Ethnographic Analysis Of Low-Income Rural and Urban Women's Food Shopping Behaviors'. *Appetite* 128(September 2018):311–320.

Minaker, L. M., D. L. Olstad, M. E. Thompson, K. D. Raine, P. Fisher, and L. D. Frank. 'Associations Between Frequency of Food Shopping at Different Store Types and Diet and Weight Outcomes: Findings from the NEWPATH Study'. *Public Health Nutrition* 19,12(August 2016):2268–2277.

Moran, P. A. P. 'Notes on Continuous Stochastic Phenomena'. *Biometrika* 37,1/2(June 1950):17–23.

Mussa, R. 'Do the Poor Pay More for Maize in Malawi?' *Journal of International Development* 27,4(May 2015):546–563.

Rahkovsky, I., and S. Snyder. 'Food Choices and Store Proximity' (2015):36.

Sauer, C. M., T. Reardon, D. Tschirley, B. Waized, D. Ndyetabula, and R. Alphonce. 'The Poor Do Not Pay More: New Evidence from Tanzania' (2022):2.

Seya, H. 'Global and Local Indicators of Spatial Associations'. In *Spatial Analysis Using Big Data* (pp. 33–56). Elsevier. 2020.

Tripathi, S., and P. K. Sinha. 'Family and Store Choice - A Conceptual Framework'. *Indian Institute of Management* ,2006(2006):1–21.

Tschirley, D., Andrew D. Jones, Mywish K. Maredia, John Mungai, Stella Nordhagen, Ahmed Salim Nuhu, … Djeinam Toure. 'Shelf Space Explains Consumer Food Choice Better than Product or Outlet Counts, but not much Overall: Evidence from Low-income Neighborhoods in Nairobi'. *Food Policy*.

Figures

Figure A1: Shop Distribution Within Wards

Table A4: Units for Top 20 Items Consumed (% of total frequency)

	Standard			Semi-standard				Non-standard												
																				Unit of Item Consumed
Food Item	grams	ml	Std %	Tea spoon	Table spoon	Cup	Semi-Std %	Packets	Bundle	Pieces	Bunch	Pinch	Loaf	Plate	Tea bag	Sachet	Whole	Ox Cart	Non-std. %	Total %
Refined (sifted) maize meal	96.3	0.1	96.4	0.0	0.0	0.2	0.2	2.8	0.1	0.5	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	3.5	100.0
Unflavored packaged pasteurized milk	0.6	23.5	24.1	0.0	0.0	0.3	0.3	70.9	0.0	4.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	75.6	100.0
Beef (unprocessed)	99.6	0.1	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.3	100.0
Rice	98.6	0.2	98.8	0.0	0.0	0.7	0.7	0.0	0.1	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.5	100.0
Tomato	1.1	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	98.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.9	100.0
White bread	55.8	0.0	55.8	0.0	0.0	0.0	0.0	10.0	0.0	15.9	0.0	0.0	18.2	0.0	0.0	0.0	0.0	0.0	44.2	100.0
Cooking oil	6.1	92.8	98.8	0.8	0.1	0.2	1.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	100.0
Sugar, honey	97.8	0.1	97.8	2.2	0.0	0.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Mandazi, cakes, biscuits, pastries	2.1	0.2	2.4	0.0	0.0	0.0	0.0	5.0	0.0	92.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	97.6	100.0
Wheat chapati	56.7	0.1	56.8	0.0	0.0	0.0	0.0	1.2	0.1	41.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	43.2	100.0
Eggs	0.1	0.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	99.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.6	100.0
Dry fish	19.9	1.9	21.8	0.0	0.0	4.4	4.4	0.0	0.3	70.1	2.8	0.0	0.0	0.3	0.0	0.0	0.3	0.0	73.8	100.0
Unflavored fresh pasteurized milk	0.4	82.3	82.7	0.0	0.0	6.2	6.2	10.3	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.1	100.0
Beans	57.1	5.3	62.4	0.5	0.0	36.6	37.1	0.0	0.0	0.1	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.5	100.0
Sukuma wiki	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	7.6	2.7	89.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0
Bananas (fruit)	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	95.9	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	100.0
Potatoes	65.7	1.3	67.1	0.0	0.0	0.3	0.3	0.5	0.3	17.7	13.7	0.0	0.0	0.3	0.0	0.0	0.0	0.0	32.6	100.0
Fresh fish	2.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	98.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.0	100.0
Whole maize meal	86.0	0.3	86.3	0.0	0.0	11.2	11.2	0.0	0.0	2.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	2.5	100.0
Chicken meat (unprocessed)	87.4	0.0	87.4	0.0	0.0	0.0	0.0	0.0	0.0	9.6	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	12.6	100.0

Table A5: Unit Conversion

Standard Unit Conversions	

Cup Conversion to grams

Food item	grams/cup
maize meal/grain	150
rice	190
flour	120
beans	180
ground nuts	125
yogurt	245
ice cream	150
sesame other seeds	140
potato	140
mango	265
amaranth	193

Tea/tablespoon Conversion

Food item	grams/teaspoon	grams/tablespoon
Tea	2.4	7.2
coffee	2.7	8.1
beans	3.8	
millet flour	3	9
salt	6	18
sugar/honey	6	18
peanut butter	5.2	15.6
margarine	5.3	15.9
groundnuts	2.6	7.8
chocolate	4.5	
cabbage	4.2	12.6
condiments/ spices	5	

*Sources used:
https://www.aqua-calc.com
https://www.howmany.wiki/
https://coolconversion.com/

Non-Standard Conversion

Food Item	grams/unit	Note/source
Papaya	1000	http://oxfarm.co.ke/tree-fruits/paw-paw/pawpaw-fruit-farming-guide-made-easy-in-kenya/
Wheat Chapati	80	https://www.mdpi.com/2072-6643/13/12/4470
Passion Fruit	45	https://www.agrifarming.in/passion-fruit-farming
Cabbage	3000	https://harvestseason.co.ke/cabbage-farming-kenya/
Fish (whole/plate)	400/100	https://edepot.wur.nl/332041
ox cart	900000	https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.1250&rep=rep1&type=pdf
coffee	350	https://magutaestatecoffee.com/shop/
salt	0.75	https://healthyeating.sfgate.com/much-salt-salt-packet-9210.html