Three Essays in Applied Microeconomics

Ling Yu

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Klaus Moeltner, Co-chair
Wen You, Co-chair
Valerie A. Thomas
Leanna L. House

August 14, 2017
Blacksburg, Virginia

Keywords: Applied Microeconomics, Applied Econometrics, Human Behavior

Three Essays in Applied Microeconomics

Ling Yu

(ABSTRACT)

This dissertation consists of three research papers in Applied Microeconomics. Each paper uses an econometric technique to analyze a problem related to human behavior. The first paper examines the separate effects of time and location of the School Breakfast Program on participation and consumption of breakfast by elementary school children in northern Nevada. Controlling for potential selection bias and unobserved individual fixed effects with a panel version of the Heckman sample selection model, it is shown that extra time allowed for breakfast leads to an approximately 20% increase in average participation, and the transition from cafeteria to classroom adds another 40% for the typical student. The second paper uses the Hedonic Property Valuation Method to quantify the willingness-to-pay of residents in the Dan River region for three dimensions of an improved food environment—availability, accessibility, and acceptability of food. This paper accounts for potential omitted variables issue in the hedonic analysis by applying a spatial-lag model, and finds an overall negative or null preference of residents in this region for an improved food environment. The third paper investigates the effects of characteristics of human interpreters and images on the accuracy of cloud interpretation for satellite images in an online experiment, using a fractional logit model. The results indicate that an image with higher cloud coverage and/or larger brightness is more likely to receive higher accuracy, and the more time spent on the image and more image completed are also beneficial for improving the accuracy. This paper also uses a logistic regression model to compare the performance of human interpreters to that of an automated algorithm, and finds that human interpreters outperform the automated algorithm for an average satellite image out of our twelve selected images.

Three Essays in Applied Microeconomics

Ling Yu

(GENERAL AUDIENCE ABSTRACT)

This dissertation consists of three research papers with each focusing on a problem related to human economic behavior. The first paper examines the separate effects of time and location of the School Breakfast Program on participation and consumption of breakfast by elementary school children in northern Nevada. The School Breakfast Program is a national food assistance program that provides subsidized breakfast to children from low-income families at participating institutions. It is shown that extra time allowed for breakfast leads to an approximately 20% increase in average participation, and the transition from cafeteria to classroom adds another 40% for the typical student. The second paper quantifies the preferences of residents in the Dan River region for three dimensions of an improved food environment—availability, accessibility, and acceptability of food. The results suggest an overall negative or null preference of residents in this region for an improved food environment. This is consistent with the focus group findings with local residents showing a desire for high fat and high energy-dense "comfort foods" and little social/cultural norms around healthful foods. The third paper investigates the effects of characteristics of human interpreters and images on the accuracy of cloud interpretation for satellite images in an online experiment. The results indicate that an image with higher cloud coverage and/or larger brightness is more likely to receive higher accuracy, and the more time spent on the image and more image completed are also beneficial for improving the accuracy. This paper also compares the performance of human interpreters to that of an automated algorithm, and finds that human interpreters outperform the automated algorithm for an average satellite image out of our twelve selected images.

Acknowledgments

Graduate life in Blacksburg in the past five years has been a wonderful journey and will prove to be an invaluable asset in my life. At the end of this journey, I am grateful to all people who have helped and supported me during this time. I especially want to thank my main advisor, Dr. Klaus Moeltner, who has always given me guidance and suggestion without reserve. I have benefited a lot from collaborating with him. I learned to be a rigorous independent researcher and a better scientific article writer. Furthermore, I want to thank Dr. Wen You, who has inspired me to start the research on food environments. During the time collaborating with her, I learned to be a better problem solver and a focused researcher. I also want to thank my other two committee members, Dr. Valerie A. Thomas and Dr. Leanna L. House, for their valuable comments on this dissertation. In addition, I would like to thank Dr. Karen Spears, Dr. Jennie L. Hill, Dr. Randolph H. Wynne, Dr. Christine E. Blinn, Seth Peery, and Jill Derwin. This dissertation would not have been possible without the invaluable input and effort from them. I also would like to extend my thanks to my friends and fellow graduate students, Jackie Yenerall, Yu Sun, Chaoping Xie, Yanliang Yang, Vartan Kesiz Abnousi, Xin Ning, Weizhe Weng, and Ruoding Shi, for their support in my difficult times. Finally, I would like to thank my husband, Xiangwei Zheng, my parents, and my family for their continued support and love. I would like to dedicate this work to Xiangwei, and my little girl, Stella.

Contents

List of Figures

List of Tables

Introduction

This dissertation presents three research projects in the field of Applied Microeconomics. Despite different topics and methods, the three projects aim to understand and explain the behavior of humans in some specific circumstances. Specifically, Chapter 1 focuses on participation decision and breakfast consumption of elementary school students in the School Breakfast Program in northern Nevada; Chapter 2 examines the preferences of residents living in the Dan River region for multiple dimensions of an improved food environment based on their home purchase decisions; Chapter 3 compares the performance of human interpreters in a cloud interpretation task for satellite images to that of an automated algorithm. The findings provide policy implications and recommendations for the School Breakfast Program, food environment improving projects, and a hybrid human-algorithm technique for cloud interpretation in the future.

Chapter 1 studies separate effects of time and location of school breakfast on participation and consumption of elementary school students. Participation in the subsidized School Breakfast Program has traditionally been unsatisfactory. Universally free breakfast service in the classroom has boosted participation, but is not financially feasible for many schools. Furthermore, it is unclear to what extent participation under the standard cafeteria setting is hampered due to insufficient time to eat. This study separately identifies time and location effects using a unique, individual-level panel data set of elementary school students under three experimental treatments: original setup in the cafeteria, original setup plus ten minutes of mandatory presence in the cafeteria, and classroom service. We control for potential selection bias problem and unobserved individual fixed effects with a panel version of Heckman sample selection model. We find that the extra time plus the "quarantine effect" in the cafeteria increases average daily participation by approximately 20%, while the transition to classroom implementation adds another 40% for the typical student. We also collect detailed data on nutritional intake, and find only minor changes in breakfast consumption for the average participant compared to the baseline. Our results are encouraging especially for the schools that are not able to switch to free breakfast because of financial issue. By simply adding ten minutes, these schools can increase average participation by approximately 20% with almost no additional cost.

Chapter 2 quantifies the preferences of residents in the Dan River region for an improved food environment based on their home purchase decisions. As of today, many people in the U.S.

live in areas with limited access to nutritious and affordable food. Some policy initiatives aim to bring more grocery stores and healthy food retailers to these areas to solve this problem. However, some quasi-experimental studies find that the introduction of a new supermarket in these deprived areas has no impact on the fruits and vegetables consumption of residents. This study uses the hedonic property valuation method to quantify the willingness-to-pay of residents in the Dan River region near the border of Virginia and North Carolina, for multiple dimensions of an improved food environment—availability, accessibility, and acceptability of food. We control for potential omitted variables issue in the hedonic analysis with a spatial-lag model. Our findings indicate that, overall, the willingness-to-pay for a small improvement in food environment, for example, introducing a new fast casual restaurant or a new grocery store in a given neighborhood, or reducing the average traveling distance to grocery stores, or improving healthy food offerings in local food outlets, is either null or negative for residents in this region. It provides one explanation to the findings of ineffectiveness of the food environment improving projects in the literature, and suggests that more monetary and non-monetary incentives along with environment changes will be needed to achieve program goals.

Chapter 3 investigates the effects of characteristics of human interpreters and images on cloud interpretation accuracy for satellite images, and at the same time, compares the performance of human interpreters to that of an automated algorithm. Satellite images are a valuable resource for research and applications in many fields, such as, forestry, agriculture, and regional planning. However, many of the satellite images are covered by clouds and/or cloud shadows. The existence of clouds and cloud shadows could lead to misleading results in remote sensing analysis. Some automated algorithm are developed for this purpose to accommodate the large volume of satellite images. However, our earlier study finds that the human interpreters recruited through an online labor market with limited training and relatively low pay perform better than a popular automated algorithm. This study examines how the characteristics of human interpreters and images affect the accuracy of cloud interpretation, and whether human interpreters outperform the automated algorithm for a set of selected twelve satellite images. The results suggest that an image with higher degree of cloud coverage and brightness receives higher accuracy, and more time spent on the image and more images completed are beneficial for improving the accuracy. For the twelve images, human interpreters are proven to be more accurate than the automated algorithm for an average satellite image in the cloud interpretation task.

Chapter 1

Breakfast at School: The Role of Time and Location for Participation and Nutritional Intake

1.1 Introduction

The School Breakfast Program (SBP) was established under the Child Nutrition Act of 1966 as a two-year pilot program to "safeguard the health and well-being of the Nation's children" (42 U.S.C. 1771). The program became permanent in 1975, providing subsidized breakfast to children from low-income families at participating institutions. Today, the SBP is available in all states. In the 2014-2015 school year an average of close to 12 million low-income children ate daily breakfast at school, which constitutes an increase of 4.2% over the previous year [54].

Breakfast has found to be beneficial to children's nutrient intake and health outcomes, such as better supply of vitamins and minerals, lower consumption of cholesterol and added sugars, and lower body mass index (BMI) [3, 10, 13, 33, 51, 90]. School breakfast programs have also been shown to increase cognitive and academic achievement for participating students [58, 69]. In a recent article, Frisvold [45] uses mandated institutional participation in the SPB for certain states as identification tool to examine the effect of the SBP on scholastic achievements. He finds that the availability of the SBP increases math, reading, and science scores.

However, despite these benefits the uptake rate of the SBP has been substantially lower than that for the National School Lunch Program (NSLP). In the school year 2007-2008, for example, 45.9 children ate breakfast at school for every 100 children that participated in school lunch, and in the school year 2014-2015 the ratio was 54.3/100 [26, 54]. To address potential barriers to breakfast participation, such as the stigma associated with a subsidized meal [65, 75], tight bus schedules, or capacity constraints in the cafeteria the Breakfast in the Classroom (BIC) program started in September 2006 as a pilot program in 20 schools in Houston, TX [44]. Since then, the BIC program has been adopted by selected public school districts in 18 States [74].

Under the traditional SBP, breakfast is served in a cafeteria setting, and household income determines the federal rate of subsidization. Specifically, breakfast is available at a heavily reduced rate (approximately 15-20% of the full price) for families with household income between 130% and 185% of the federal poverty level, and available for free for households below 130% of that level. For example, for the 2014-2015 school year schools received \$1.62 per fully subsidized breakfast, \$1.32 per reduced-price breakfast, and \$0.28 per "paid" breakfast. *Severe need* schools, that is schools with at least 40% of students qualifying for reduced-price of free meals received an additional \$0.31 for each free or reduced-price breakfast served [54].

In contrast, in the BIC program breakfast is free of charge to all students and consumed in the classroom during the first 10 minutes of the official school day. The BIC implementation falls under the category of "universally free" school breakfast programs, in that breakfast is served at no cost to any students. However, the federal reimbursement policy remains unchanged. This implies that excess proceeds from federally subsidized meals must be used by school administrators to cover meals for students that otherwise would have to pay partly or in full.

The BIC program has indeed been found effective in increasing SBP participation, to the order of 30-60% over schools with traditional implementation [27, 83]. Furthermore, as is becoming evident from individual school district reports, BIC implementation also appears to be more cost-effective per meal served, despite increased costs for waste management [34]. Some schools also recruit students and volunteer parents to help with meal set-up and delivery [68, 87]. Ultimately, though, the financial feasibility of running a universally free breakfast program such as BIC at the school level hinges crucially on the proportion of subsidized to unsubsidized meals at any level of participation. For example, using actual budget figures given in Dickl [34], if the average cost per meal under BIC amounts to \$1.48 and the average federal subsidy is \$1.62, the proportion of subsidized meals needs to be (1.48/1.62), that is over 91% for the school to break even.

Thus, for institutions with a lower percentage of students that qualify for meal subsidies, BIC implementation is not feasible from a purely financial perspective. It is therefore meaningful to ask how breakfast participation can be increased under the conventional SBP in a cafeteria setting. This study examines the separate effects of additional time to eat and change-of-location to the classroom on breakfast participation and nutritional intake of third to fifth grade elementary students in the Reno/Sparks metropolitan area of northern Nevada. We benefit from a natural experiment as all of our three targeted schools were slated to move from conventional SBP to BIC for the second year of our study. Therefore, in the first year we collect baseline data under conventional implementation, followed by two weeks of allowing for an additional ten minutes beyond usual cafeteria hours for students to have breakfast ("C+10"). This mimics the "breakfast after the bell" implementation of BIC, but with an unchanged cafeteria location.

Each experimental treatment can be broken into four operative components, as captured in Table 1.1. The first component is a pure "time to eat" effect - presumably participation and possibly nutritional intake will increase with available time. As is evident from the first row of the table, time to eat is variable under the baseline period (henceforth referred to as T0), depending on students' arrival and cafeteria hours. The C+10 treatment (henceforth also referred to as T1) simply adds ten minutes to baseline time, thus keeping total available time student-specific. In contrast, the BIC treatment (henceforth T2) restricts time to eat to exactly ten minutes for all students.

Table 1.1: Treatment details

treatment component	baseline (T0)	C+10 (T1)	BIC (T2)
time	variable	variable + 10 min	10 min for all
quarantine	none	yes, 10 min	yes, 10 min
location	cafeteria	cafeteria	classroom
price	full / reduced / free	full / reduced / free	universally free

min = minutes
C+10 = cafeteria, 10 extra minutes
BIC = breakfast in classroom

The second treatment component is a "quarantine" effect, resulting from forcing all students to remain in the same location where breakfast is served for a specific period of time, after the school bell. As captured in the table, such a quarantine period is absent under T0, but built into both T1 and T2, for a length of 10 minutes. We hypothesize that this quarantine effect will limit distractions, such as access to the playground, and increase participation and, possibly, intake.

The third treatment component is the pure location effect, that is a cafeteria environment under T0 and T1, and the classroom under T2. The fourth and final component is the pricing structure under each treatment, which is the standard full / reduced / free pricing based on eligibility under T0 and T1, and universally free pricing under T2.

A crucial feature of our experiment is that we capture each student's *daily arrival time* at school, which allows us to compute the exact time available for breakfast under T0, and thus, trivially, under T1. This allows for the identification of the pure time-to-eat component of breakfast implementation. In turn, the incremental impact of T1, after controlling for time-to-eat, can then be interpreted as a pure quarantine effect, since both pricing and location remain unchanged compared to T0. In consequence, the incremental effect of T2 over T1 can be interpreted as a combination of a location and price effect. As we argue below, price effects will likely be small for our sample, so the location effect should dominate in this comparative analysis.

Our results show that the C+10 treatment increases average daily participation for the typical student by approximately 20%, of which 5-6% can be attributed to a pure quarantine effect. It virtually eliminates the probability that a child "goes hungry" all morning, as experienced by 2-4% of students in our sample. Consistent with the bulk of the literature, moving breakfast to the classroom adds an additional 40% increase in average daily participation. Overall, we obtain an upper bound of close to 60% for the pure classroom location effect, after controlling for time to eat and quarantine effects. Treatment one produces a slight decrease in caloric intake for the typical participant, while the intake effect under T2 remains insignificant.

To our knowledge, this is the first study that examines the effect of different modes of SBP implementation on participation and nutritional intake using a panel of daily, individual-level data on arrival time at school, preferences for the daily menu, and exact nutritional intake, in addition to numerous demographic and breakfast-related background information.

The remainder of this paper is organized as follows: The next section describes our econometric approach. This is followed by an empirical section that introduces the data, provides descriptive statistics, and presents estimation results. The final section concludes.

1.2 Methods

The point of departure for our estimation approach is the panel version of the Heckman sample selection model as described in Wooldridge [97], with a participation equation and a consumption equation. We ex ante allow the two equations to be correlated via unobservables. Formally, the model can be written as:

$$\begin{aligned}
p_{it}^* &= \gamma_0 + C_{it}\gamma_1 + B_{it}\gamma_2 + \mathbf{Z}_{it}'\boldsymbol{\gamma}_3 + \xi_i + a_{it} \\
y_{it}^* &= \beta_0 + C_{it}\beta_1 + B_{it}\beta_2 + \mathbf{X}_{it}'\boldsymbol{\beta}_3 + \alpha_i + u_{it} \\
p_{it} &= 1 \text{ if } p_{it}^* > 0,\ 0 \text{ otherwise} \\
y_{it} &= y_{it}^* \text{ if } p_{it} = 1,\ \text{unobserved otherwise} \\
i &= 1, \ldots, N;\ t = 1, \ldots, T,
\end{aligned} \tag{1.1}$$

where y_{it} denotes the caloric intake from breakfast at school by student i on day t. This measure is only observed if the student participates in school breakfast that day, that is if the participation indicator p_{it} is equal to one. The binary indicators C_{it} and B_{it} denote, respectively, C+10 and BIC treatment for student i on day t. Vector $\mathbf{Z}_{it}$ in the participation equation contains observed explanatory variables, such as time to eat, mode of transportation to school, food consumption prior to arrival, hunger level, demographic characteristics, menu features, and day-of-week indicators. Vector $\mathbf{X}_{it}$ in the intake equation comprises the same elements, minus one to satisfy the standard exclusion restriction of the selection model.[1] Terms ξ_i and α_i capture individual fixed effects, and a_{it} and u_{it} are idiosyncratic error terms that are jointly normally distributed with zero means.

We first test for selection bias. As a first step, this requires the separate estimation of the participation equation. To circumvent the incidental parameter problem in our fixed effects probit we use the approach proposed in Mundlak [70], Nijman and Verbeek [71], and Zabel [100] by expressing the fixed effect as a function of the time-average of the explanatory variables, that is

$$\xi_i = \eta_0 + \bar{C}_i\eta_1 + \bar{B}_i\eta_2 + \bar{\mathbf{Z}}_i'\boldsymbol{\eta}_3 + c_i, \tag{1.2}$$

where c_i is a standard normal idiosyncratic error term that is independent of included variables. This yields the modified participation equation

$$p_{it}^* = \delta_0 + \bar{C}_i\eta_1 + \bar{B}_i\eta_2 + \bar{\mathbf{Z}}_i'\boldsymbol{\eta}_3 + C_{it}\gamma_1 + B_{it}\gamma_2 + \mathbf{Z}_{it}'\boldsymbol{\gamma}_3 + \nu_{it}, \tag{1.3}$$

where $\delta_0 = \gamma_0 + \eta_0$, and $v_{it} = a_{it} + c_i$. This equation can be consistently estimated via pooled probit and produces an inverse Mills ratio (IMR), that is the conditional expectation of ν_{it}, given participation. In the second step of the test, this IMR is inserted in the consumption equation. The selection test is then based on the significance of its coefficient.

For our application we do not find any indication of selection problems, probably due to the exhaustive list of covariates employed in both the participation and intake equations.[2] We therefore estimate the participation equation via pooled probit as discussed above, and the consumption equation as a basic fixed-effects panel model with a full set of regressors (that is, replacing $\mathbf{X}_{it}$ with $\mathbf{Z}_{it}$ in the second equation of (1.1)).

1.3 Empirical Application

The study was conducted in three elementary schools of the Washoe County school district in the Reno/Sparks metropolitan area of northern Nevada over three two-week periods during the 2010-2011 and 2011-2012 school years.[3] As mentioned above, all three schools operated under the standard SBP in 2010-2011, and switched to universally free breakfast with BIC implementation in the second year. While we do not have information on the exact proportion of fully and partially subsidized students, it can be assumed to be at least 80% or higher, as it would otherwise not be cost-effective to move into BIC implementation under typical cost structures [43].

[1] We experiment with different exclusion restrictions, with no measurable effect on our results.

[2] The t-tests for selection bias produced p-values of 0.612, 0.872, 0.836, and 0.988 for the three schools and the total sample, respectively.

[3] The study was approved by the Institutional Review Board of the University of Nevada, Reno, and the Washoe County school district department of public policy, accountability, and assessment.

1.3.1 Data

In the first year we recruited third and fourth graders based on parental consent and student's assent. In the following year, the same group of students was included, in addition to the new cohort of third graders.[4] In addition to consent and assent, a third requirement for participation was the absence of any other reasons that would preempt breakfast consumption at school, such as health issues or religious beliefs. Both parents and students had the right to withdraw from the study at any point in time.

Data under baseline conditions (T0) were collected at each location during a two-week period in spring 2011. The baseline period was immediately followed by two weeks of C+10 (T1). Data under BIC (T2) were collected in late fall 2011 at schools one and three, and in early 2012 at school two. The detailed sampling plan is given in Table 1.2. Due to computer calibration or weather issues the actual days of successful implementation of each treatment varied between eight and ten across the three schools, as captured in the last column of the table.

Table 1.2: Sampling plan

school	treatment	begin date	end date	valid days
1	baseline	18-Jan-11	31-Jan-11	9
	C+10	1-Feb-11	15-Feb-11	9
	BIC	28-Nov-11	12-Dec-11	10
2	baseline	28-Feb-11	14-Mar-11	10
	C+10	15-Mar-11	28-Mar-11	8
	BIC	17-Jan-12	31-Jan-12	8
3	baseline	16-May-11	31-May-11	10
	C+10	1-Jun-11	14-Jun-11	9
	BIC	19-Sep-11	3-Oct-11	10

basic = cafeteria, regular time
C+10 = cafeteria, 10 extra minutes
BIC = breakfast in classroom

For the remainder of this text, we use the term "enrolled student" to indicate a student who was recruited into our study, and the term "participating student" to indicate an enrolled student who had some food for breakfast at school that day.

[4] Parents received a recruitment form in both English and Spanish. They were asked to return the form if they did not want their child to participate in the study. In addition to obtaining parents' consent, an assent script was read to students in English or Spanish, depending on their primary language. The assent script was designed according to the Flesch readability scale for a fourth grade child [42].

Demographic data

Information on gender, race/ethnicity and age of enrolled students was obtained from school records. Students' height and weight were measured annually as part of our study. Each student's BMI was then computed as weight divided by the square of height (kg/m^2).

Arrival time data

Research staff were positioned at every school entry point to record the arrival time of enrolled students, regardless of transportation mode. As discussed above, this information was used to compute the exact available time to eat, in minutes, under T0 and T1. For T2, the time available to eat breakfast was ten minutes for all students and schools.

Daily food questionnaire

A daily food questionnaire was administered to inquire about students' hunger level, how they traveled to school that day, what, if anything, they had eaten before school, and preferences for food items being served at breakfast that day. Hunger level was measured on a Likert scale ranging from zero ("not hungry") to 17 ("very hungry"). The transportation mode choices for getting to school were "bus," "dropped off by car," "walk," "bike," and "other." The students were also asked to describe any food or drinks they had before coming to school based on a check list with popular items (e.g."cereal," "toast," "burrito," "eggs," "milk," "juice," and "soda"). Preferences for breakfast items served at school that day were captured on a five-point Likert scale ("do not like it," "do not like it a little," "like it a little," and "like it a lot"). All of this information was collected from all enrolled students, regardless of their participation in school breakfast that day. The daily food questionnaires were distributed to all enrolled students in the classroom under T0 and T2, and in the cafeteria under T1. Research staff reviewed the responses and inquired about missing items upon collecting the questionnaires. The questionnaire is available in a supplementary online appendix.

Dietary intake data

The Spears Dietary Assessment Tool (Spears-DAT) was used to measure dietary intake of participating students [86]. This method is based on a computerized setup with a weighing scale that calculates and records calorie and nutrient intake based on the type and weight of food consumed. Under T0 and T1 the students placed selected food and beverage items on their barcoded trays and paid corresponding prices (free, reduced, or full) at checkout. They then immediately proceeded to the weigh station, where the research team scanned their ID number, tray bar code, and food item bar codes, and obtained the weight of all food items. During breakfast research staff documented any food trading and spillage. After breakfast students left all remaining food, drinks, and packages on their tray. Trays were then collected by staff, and all left items were weighed back to obtain the net weight of food consumed.

For the BIC implementation under T2 food items with variable weight, such as fruit, were weighed before breakfast and placed in a zip-lock bag with weights recorded on the bag. All food items were transported to the classroom with appropriate packaging. During the first 10 minutes of a given school day, participating students picked up bar-coded empty containers, placed their pre-printed name labels on the container, chose their breakfast food items, and placed them in the container. After breakfast was completed, containers were collected by staff and returned to the cafeteria for weigh-back. For those food items with varied weights, weights indicated on zip-lock bags were entered into the computer manually before weighing back. As before, the amount of food consumed was obtained by subtracting after-breakfast weight from before-breakfast weight for each item.

The nutrient intake of a given student on a given breakfast occasion was obtained by multiplying the consumed amount by the food item's nutritional composition, obtained either from the USDA National Nutrient Database for Standard Reference [94], the manufacturers' label, or the school nutrition service. For this study, we focus exclusively on total calories consumed, leaving a detailed examination of intake of different nutritional components to separate work.[5]

Menu data

In order to get reimbursed schools must provide a breakfast that meets the recommendations of the Dietary Guidelines for Americans [95] on the percent of calories from fat (less than 30 %) and saturated fat (less than 10%), and provides one-fourth of the Recommended Dietary Allowance (RDA) suggested in the Dietary Guidelines for Americans [95] for certain minerals and vitamins. Within these parameters, the selection of specific foods to be served lies with each individual school's food service.

A typical breakfast menu is usually comprised of an entree, fruit, and a drink. For all three schools in our study, more food choices were offered in the cafeteria than in the classroom. For example, cereal was only served in the cafeteria, and there were at least two different drinks offered in the cafeteria, but often only one in the classroom. Typical breakfast menus in the cafeteria and in the classroom are shown in Table 1.3. In total, 13 different types of entrees, 10 different types of fruit, and 10 different types of drinks were served across schools and treatments.

We capture basic menu items with binary indicators in our econometric specifications. For example, for the classroom menu shown in Table 1.3, we would set variables "yogurt"=1,

[5] The nutrients recorded in the experiment are calories, fat (g), saturated fat (g), carbohydrate (g), protein (g), vitamin C (mg), calcium (mg), iron (mg), fiber (g), cholesterol (mg) and sodium (mg).

"apple"=1, "1% milk"=1 and all other food item indicators to zeros. While menu service did not follow a systematic rotation at any of the schools, we include basic day-of-week indicators in both the participation and consumption models to control for unobserved day-of-week-specific menu (and other) effects.[6]

Table 1.3: Typical breakfast menus

	cafeteria	classroom
main entrée	burrito	yogurt with granola
	cereal [a]	
fruit	banana	apple
	tangerine	
drink	orange juice	plain 1% milk
	plain skim milk	

[a] Cereal was offered everyday, but only in the cafeteria.

1.3.2 Descriptive Statistics

Table 1.4 shows enrollment counts and demographic statistics for each school and the sample at large. In total there were 236 students enrolled in the first year of the study out of 355 eligible students, for a participation rate of 66%. This number increased to 371 in year two out of 575 eligible (65% participation rate), primarily due to the expansion to three scholastic grades (new cohort of third graders, plus continuing cohort of - by that time - fourth and fifth-graders).[7] Overall, 166 students were enrolled in both research years, and thus under all three treatments. We use the total sample of enrolled students in our analysis.[8] Enrollment counts are highest for school three and lowest at school one, which reflects the relative size of the student body at these institutions. Participation rates (enrolled/eligible) are in the 60-70% range for all schools and time periods.

By far the dominant race/ethnic group is Hispanic for all three schools, with White students comprising the second largest contingent. Age and BMI statistics are also comparable at all three locations.

Table 1.5 presents sample statistics and results of basic t-tests for mean-equality for responses to the daily food questionnaire as well as available time to eat school breakfast based on arrival time. Perhaps the most important insight gained form the table is that the percentage of students who had food or drinks before arriving at school declines significantly going from

[6] Detailed daily menus for each school are available from the authors upon request.

[7] Nine students had to be dropped due to missing age and/or weight information in the first year, and 12 in the second year.

[8] An analysis using only the overlapping students produces very similar results.

Table 1.4: Sample statistics, part I

	school 1				school 2			
	year 1		year 2		year 1		year 2	
eligible students	80		128		109		184	
enrolled students*	48		78		73		129	
overlap students			33				53	
male	50.00%		42.31%		60.27%		53.49%	
white	8.33%		15.38%		13.70%		12.40%	
black	2.08%		6.41%		6.85%		6.20%	
native	0.00%		0.00%		1.37%		0.78%	
asian	2.08%		2.56%		4.11%		2.33%	
hispanic	81.25%		61.54%		49.32%		62.02%	
other	6.25%		14.10%		24.66%		16.28%	
	mean	std.	mean	std.	mean	std.	mean	std.
age	9.38	0.56	9.94	0.88	9.38	0.60	9.73	0.85
BMI	19.81	5.03	20.37	5.03	19.90	4.44	20.77	4.84
	school 3				all			
	year 1		year 2		year 1		year 2	
eligible students	166		263		355		575	
enrolled students	115		164		236		371	
overlap students			80				166	
male	44.35%		46.95%		50.42%		48.25%	
white	23.48%		24.39%		17.37%		18.33%	
black	0.87%		0.61%		2.97%		3.77%	
native	0.87%		1.22%		0.85%		0.81%	
asian	0.00%		0.00%		1.69%		1.35%	
hispanic	61.74%		64.63%		61.86%		63.07%	
other	13.04%		8.54%		15.25%		12.40%	
	mean	std.	mean	std.	mean	std.	mean	std.
age	9.68	0.63	9.59	0.90	9.53	0.62	9.72	0.89
BMI	19.04	3.99	19.36	4.04	19.46	4.36	20.06	4.58

*enrolled and completed study (did not drop out)
std. = standard deviation

T0 to T1, and from T1 to T2. This points at a substitution effect from breakfast at home to breakfast at school under C+10 and BIC compared to the baseline. Results based on the overlap sample only (available upon request) show the same substitution effect, with very similar percentages.

Except for school one, which is a small neighborhood school that is not serviced by school buses, the first block of rows in the table point at a shift from “car” to “bus” and “walk / bike” under T2 in terms of transportation mode. This result, which holds for the overlap sample as well, suggests that using the school bus or walking may have made it difficult to arrive in time for breakfast under T0 for some students. Interestingly, this mode-shifting effect is not present under T1, possibly because it takes some time for households to adjust their daily commuting schedule, and T1 immediately followed T0. In contrast, parents had at the very least the summer break between school years to prepare for a changed morning routine under T2.

Table 1.5 also shows that there is substantial variability in the available time to eat under baseline implementation across the school, from just over 11 minutes at school one to over 20 minutes at school three. This is primarily related to differences in cafeteria hours, with school three offering the longest breakfast service (8:15 - 9:20 am), followed by school two (8:15 - 8:50 am) and school one (8:30 - 8:50 am). These gaps narrow somewhat under C+10, with the typical student having between 23 minutes (school 1) and 29 minutes (school 3) to eat breakfast. As mentioned before, breakfast time is universally set to exactly ten minutes under the classroom implementation.

Table 1.6 presents similar descriptive statistics for the sub-sample of student-days with breakfast participation. The general patterns of substitution in transportation mode and from breakfast at home to breakfast at school observed for the full sample also hold for this group. Looking at the “time to eat” rows, it becomes evident that participating students had more time for school breakfast than the sample at large at all three schools. Specifically, across all three locations the typical participating student had 28 minutes to consume school breakfast under T0, and over 31 minutes under T1. This compares to 18 minutes under T0 and 28 minutes under T1 for the full sample. This gives a first indication of a positive linkage between available time and school breakfast participation.

A general overview of participation and related outcomes under the three treatments is given in Table 1.7. As is evident from the first row for each school the C+10 implementation (T1) had a significant positive effect on average daily participation, ranging form approximately 13% for school three to close to 29% for school one. For the sample at large close to 18% more students ate breakfast at school on a typical day than under baseline conditions. Consistent with existing empirical evidence from the literature, participation approaches 100% under BIC (T2), implying a 60% increase over baseline and a 44% increase over T1 for the sample at large.

As is clear from the remaining rows in the table, T2 virtually eliminates the share of students that skipped breakfast, i.e. decided not to participate even though they had sufficient time to eat and no food before arriving at school. Treatment one also reduces this contingent, but only by a few percentage points compared to baseline. Importantly, both treatments reduce the percentage of students that had to “go hungry” on a typical day - i.e. that did not have any food before arrival at school and did not participate in school breakfast *due*

Table 1.5: Sample statistics, part II

	school 1					school 2				
	treatment			t-tests[a]		treatment			t-tests	
	T0	T1	T2	T1	T2	T0	T1	T2	T1	T2
bus	0.00%	0.00%	0.43%		*	3.25%	3.53%	6.80%		***
car	41.46%	43.49%	48.34%		**	74.14%	75.68%	63.27%		***
walk / bike	57.42%	55.68%	51.23%		*	20.98%	19.54%	27.08%		***
other	1.12%	0.83%	0.00%		**	1.63%	1.25%	2.85%		
food before	73.67%	59.56%	42.14%	***	***	64.92%	55.72%	31.80%	***	***
hunger (0-17)	5.56	5.42	8.37		***	6.36	6.28	10.27		***
	(5.42)	(5.80)	(4.43)			(5.46)	(5.49)	(4.89)		
time to eat (min.)	11.01	23.38	10.00	***	**	16.68	28.33	10.00	***	***
	(9.70)	(10.00)	(0.00)			(11.85)	(12.54)	(0.00)		
observations	357	361	693			553	481	912		
total		1411					1946			

	school 3					all				
	treatment			t-tests		treatment			t-tests	
	T0	T1	T2	T1	T2	T0	T1	T2	T1	T2
bus	8.91%	9.36%	14.73%		***	5.46%	5.69%	9.16%		***
car	57.17%	57.11%	49.02%		***	59.23%	59.49%	53.09%		***
walk / bike	30.11%	30.45%	34.35%		**	32.68%	32.74%	36.00%		**
other	3.80%	3.08%	1.90%		***	2.62%	2.08%	1.75%		**
food before	71.85%	61.97%	45.55%	***	***	70.11%	59.67%	40.71%	***	***
hunger (0-17)	6.26	5.86	9.91		***	6.16	5.88	9.67		***
	(5.48)	(5.56)	(5.01)			(5.47)	(5.60)	(4.90)		
time to eat (min.)	20.77	29.10	10.00	***	***	17.63	27.65	10.00	***	***
	(19.47)	(14.34)	(0.00)			(16.27)	(13.20)	(0.00)		
observations	920	844	1473			1830	1686	3078		
total		3237					6594			

[a] Two-sample t-tests with unequal variances, compared to T0
* (**) [***] significant at 10 (5) [1] percent
(standard deviation)

to time constraints - to close to zero, from 2-4% under T0, depending on location. Thus, the C+10 mode appears to be effective in recruiting the most vulnerable children - those that would otherwise start the day without any food for lack of time to eat - back into the breakfast pool.

The last row for each school, labeled "double breakfast," captures the average daily percentage of students that participated in school breakfast, but also had food before arrival. Under T0, this share amounts to 16-23% across schools. It increases by a few percentage points

Table 1.6: Sample statistics, participating students only

	school 1					school 2				
	treatment			t-tests[a]		treatment			t-tests	
	T0	T1	T2	T1	T2	T0	T1	T2	T1	T2
bus	0.00%	0.00%	0.44%		*	0.42%	3.29%	6.76%	***	***
car	45.16%	35.00%	48.48%			75.42%	76.64%	63.41%		***
walk / bike	54.84%	63.50%	51.09%			22.92%	19.41%	26.94%		
other	0.00%	1.50%	0.00%	*		1.25%	0.66%	2.88%		*
food before	59.14%	46.50%	42.09%	**	***	52.92%	47.37%	31.37%		***
hunger (0-17)	7.77	6.74	8.38			8.57	7.89	10.36		***
	(5.80)	(5.96)	(4.43)			(5.87)	(5.61)	(4.82)		
time to eat (min.)	16.56	26.80	10.00	***	***	21.61	30.64	10.00	***	***
	(5.76)	(8.98)	(0.00)			(9.59)	(12.14)	(0.00)		
observations	93	200	689			240	304	902		
total		982					1446			
	school 3					all				
	treatment			t-tests		treatment			t-tests	
	T0	T1	T2	T1	T2	T0	T1	T2	T1	T2
bus	5.20%	7.31%	14.82%		***	2.80%	4.42%	9.16%	*	***
car	56.94%	50.47%	49.03%	*	***	61.86%	55.71%	53.18%	**	***
walk / bike	33.53%	38.44%	34.21%			32.70%	37.61%	35.88%	**	
other	4.34%	3.77%	1.94%		**	2.65%	2.26%	1.78%		
food before	53.18%	43.87%	45.22%	**	***	53.90%	45.58%	40.40%	***	***
hunger (0-17)	8.06	7.54	10.04		***	8.20	7.48	9.76	**	***
	(5.61)	(5.79)	(4.97)			(5.73)	(5.78)	(4.86)		
time to eat (min.)	35.62	34.21	10.00		***	28.05	31.44	10.00	***	***
	(20.65)	(15.37)	(0.00)			(17.78)	(13.48)	(0.00)		
observations	346	424	1444			679	928	3035		
total		2214					4642			

[a] Two-sample t-tests with unequal variances, compared to T0
* (**) [***] significant at 10 (5) [1] percent

under T1 (5% for the sample at large), and almost doubles under T2 (from 20% to almost 40% for the sample at large). This "double-dipping" is consistent with findings reported in Corcoran, Elbel, and Schwartz [27], though the authors of that study diffuse any concerns related to BIC and obesity in their New York City-based study.

Table 1.7: Participation statistics

	school 1					school 2				
	treatment			t-tests[a]		treatment			t-tests	
	T0	T1	T2	T1	T2	T0	T1	T2	T1	T2
participation	26.05%	55.40%	99.42%	***	***	43.40%	63.20%	98.90%	***	***
skip breakfast	15.69%	10.80%	0.29%	*	***	14.65%	11.02%	0.33%	*	***
go hungry	1.96%	0.28%	0.00%	**	***	3.25%	0.42%	0.00%	***	***
double breakfast	15.41%	25.76%	41.85%	***	***	22.97%	29.94%	31.03%	**	***
	school 3					all				
	treatment			t-tests		treatment			t-tests	
	T0	T1	T2	T1	T2	T0	T1	T2	T1	T2
participation	37.61%	50.24%	98.03%	***	***	37.10%	55.04%	98.60%	***	***
skip breakfast	10.54%	9.83%	0.75%		***	12.79%	10.38%	0.52%	**	***
go hungry	3.80%	0.00%	0.00%	***	***	3.28%	0.18%	0.00%	***	***
double breakfast	20.00%	22.04%	44.33%		***	20.00%	25.09%	39.83%	***	***

All values are averaged over days
[a] Two-sample t-tests with unequal variances, compared to T0
* (**) [***] significant at 10 (5) [1] percent

1.3.3 Estimation Results

We first discuss results for the participation model, followed by the caloric intake model.

Participation equation

We estimate our feasible fixed-effects probit model for the participation equation with treatment indicators for T1 and T2, and a separate continuous variable that measures available time to eat, in minutes. As discussed previously, this allows us to dis-entangle the pure added time effect from quarantine effects, as well as location effects associated with the two experimental treatments. We also include the following variables from the daily food questionnaire: an indicator for walking or biking to school (hypothesizing that this might increase appetite), and indicator if they had food before heading to school, and the measured index on the hunger scale. This set is augmented with the two demographic variables age and BMI. The model also includes day-of-the week indicators (with Monday as omitted baseline) and up to 27 menu indicators for food items served on a given day.

We use conventional methods to translate the raw coefficients from this model into marginal effects on participation probabilities [e.g. 16, section 14]. Specifically, we use continuous derivatives for marginal effects associated with (pseudo-) continuous variables (time, hunger, age, BMI), and differences in cumulative distribution functions to obtain marginal effects for binary regressors (treatments, walk/bike, food before). Each marginal effect is computed for each observation in the sample, then averaged over all observations. Standard errors for each effect are obtained using the Delta method.

Table 1.8 shows marginal effects for the food questionnaire and demographic variables, holding treatment, menu, and day-of-week effects at observed levels. Clearly, available time to eat breakfast matters, with an additional minute boosting participation probabilities by 0.7-0.8%. A marginal increase in hunger level, as captured by our 17-point scale, has a similar positive effect of 0.7-1.2%, depending on location. In contrast, having had food or drinks before arriving at school decreases participation by 4-5% for the typical student. None of the other covariates produce significant effects at the 5% level or higher.

Table 1.8: Marginal effects for moderator variables, participation equation

	school 1		school 2		school 3		all	
T1			set to observed values					
T2			set to observed values					
time to eat	0.007	***	0.007	***	0.008	***	0.007	***
walk / bike	-0.032		0.001		0.008		0.003	
food before	-0.016		-0.048	**	-0.049	***	-0.044	***
hunger	0.007	***	0.010	***	0.012	***	0.010	***
age	-0.031		-0.008		0.006		-0.010	
BMI	0.003		-0.005		0.003*		0.0002	
menu			set to observed values					
day of week			set to observed values					
students	93		149		199		441	
observations	1411		1946		3237		6594	

All values are averaged over observations
* (**) [***] significant at 10 (5) [1] percent

Marginal effects for the two treatments are computed for three different counterfactual scenarios, as shown in Table 1.9. Scenario one sets a baseline of zero time to eat, and zero treatment effect, that is no mandatory gathering in a specific location (i.e. "quarantine"). The marginal effect is then computed for a change from this baseline to ten minutes of eating time, plus quarantine in the cafeteria (T1) and classroom (T2), respectively. Therefore, this scenario can be interpreted as the *full treatment effect* for a student who up to that point would have not been able to participate in school breakfast.

Scenario two repeats this setup, with the single difference of setting available time to eat to the sample mean under T0. Thus, the resulting marginal effect can be interpreted as the *full treatment effect* for a student with typical available eating time.

Scenario three, in contrast, starts form a baseline of zero treatment indicators, but uses actually observed eating time for each student-day. This is then changed to observed eating time, plus a treatment indicator of one. Therefore, the marginal effect obtained from scenario three can be interpreted as the *pure quarantine effect* for T1, and as the combination of quarantine, location, and price effect for T2. As discussed previously, the difference in marginal effects for this last scenario between T1 and T2 can be interpreted as the location-plus-price effect, with the location effect likely dominating the price effect.

To avoid confounding effects from the opposing treatment, we derive marginal effects under T1 using only data from the T0 and T1 periods (that is, the first year of implementation), and marginal effects under T2 using only baseline data (first two weeks of year one) plus T2 data (year two).

Table 1.9: Treatment scenarios for marginal effects

scenario	before	after
	treatment 1 (C+10)	
1	T1 = 0, time = 0	T1 = 1, time = 10
2	T1 = 0, time = avg. under T0	T1 = 0, time = avg. under T0 + 10
3	T1 = 0, time = observed	T1 = 1, time = observed
data used:	T0 and T1 only	
	treatment 2 (BIC)	
1	T2 = 0, time = 0	T2 = 1, time = 10
2	T2 = 0, time=avg. under T0	T2 = 1, time = 10
3	T2 = 0, time=observed	T2 = 1, time = observed
data used:	T0 and T2 only	

time = available time to eat, in minutes
observed = held at actual values for each observation

Results are given in Table 1.10. Focusing first on T1 and the sample as a whole, the table shows that T1 increases participation by approximately 16% for a student with previously zero time to eat (scenario one), and by 21% for a student with previously "typical" amount of available time (scenario 2).[9] The *pure quarantine effect* of T1 amounts to 5.5%, as indicated by scenario three results.

[9]The lower incremental effect for a previous non-participant under scenario one compared to a typical student under scenario two is due to the nonlinearity of the cumulative normal transformation of linear effects into probabilities. While both scenarios imply an equal linear increment along the support of the standard normal, the resulting changes in cumulative probability mass differ, since the first scenario implies a movement within the lower tail region of the normal, while the second scenario describes a movement

In summary, we conclude that the pure time to eat effect under T1 amounts to 10-15%, depending on how much time was already available under baseline, with the quarantine effect adding an additional 5-6% of increased probability of participation. Individual school effects are similar to full-sample effects for schools two and three. Marginal effects for school one appear to be higher, but do not emerge as significant, likely due to the considerably smaller sample size at that institution.

Shifting focus to T2, we can see from Table 1.10 that the BIC implementation increases participation probabilities by over 80% for a student with previously zero time to eat (scenario one). Moving a previously typical student with average time to eat from baseline to BIC (scenario two) brings an overall increase in participation of close to 60%. Keep in mind that this scenario de facto imposes a *time penalty* for the typical student, from a sample average of 18 minutes (see Table 1.5) down to 10 minutes, which explains the lower boost in participation under scenario two compared to scenario one. However, as indicated by scenario three results, the combined quarantine-plus-location-plus-price effect for T2 more than makes up for this time penalty, amounting to approximately 65% for the sample at large.

Overall, marginal effects under T2 mimic those for the sample at large for schools two and three, while exceeding the full-sample values by 10-20% for school one.

As discussed above, we can now derive a location-plus-price effect, or, alternatively put, an upper bound for the location effect, by comparing scenario three results across treatments. Focusing again on the full sample, this combined effect amounts to 65-5.5 = 59.5% for the typical student. Assuming that price effects are minimal, we therefore conclude that there still remains a very large location effect under BIC, after controlling for time-to-eat and quarantine effects.

Caloric intake equation

Results for the intake equation with fixed student effects are given in Table 1.11. We use robust standard errors that are clustered at the individual level. We first observe that the same student-specific variables that affect participation also play a significant role for intake, with effects going in the same direction. Specifically, having more time to eat and coming to school hungry increases consumption, while having had food or drinks before arrival has the opposite effect. For school one and the full sample, age is also a significant driver of intake, with a pronounced marginal effect per added year of over 360 kcals. In contrast, having walked or biked to school or BMI do not seem to play a role for caloric intake for our sample.

Since we explicitly control for time to eat, the binary indicator for T1 captures primarily the quarantine effect that comes with this treatment. This effect induces a significant, but moderate reduction in intake of 22 kcals, approximately 10% of the sample average under

within the central region, with much larger gains in probability mass.

Table 1.10: Marginal treatment effects for participation equation

	school 1		school 2		school 3		all	
T1, scenario 1	0.210		0.158	**	0.161	***	0.159	***
T1, scenario 2	0.257		0.181	**	0.251	***	0.212	***
T1, scenario 3	0.114		0.051		0.073	**	0.055	***
students	48		73		115		236	
observations	718		1034		1764		3516	
T2, scenario 1	0.828	***	0.729	***	0.798	***	0.806	***
T2, scenario 2	0.733	**	0.543	***	0.523	***	0.592	***
T2, scenario 3	0.729	**	0.589	***	0.620	***	0.651	***
students	93		149		199		441	
observations	1050		1465		2393		4908	

All values are averaged over observations
* (**) [***] significant at 10 (5) [1] percent

T0 for the typical student. This may capture a complacency effect, where marginally hungry students eat a “token breakfast” during quarantine because food is within convenient reach, to pass time, or to please teachers. Changes in caloric intake under T2 are two noisy to extract significant signals, as is evident from the second row of the table.

1.4 Conclusion

We conduct a unique field experiment using different implementation modes for school breakfast to study the separate effects of time, quarantine, and location on participation and nutritional intake for third to fifth graders at three schools in northern Nevada.

We find that even under a standard cafeteria setting and regular pricing participation rates can be increased substantially by allowing for more time to eat, in conjunction with an after-the-bell quarantine of all students in the cafeteria location. Realistically speaking, adding time after the bell may well be the only way breakfast participation can be increased for most schools, as it is unlikely that students with zero time to eat can be encouraged to arrive earlier, especially under given school bus schedules.

We believe that our C+10 findings are encouraging for schools that are unable to switch to universally free breakfast due to an insufficiently large share of fully subsidized students, as this mode can be implemented at very low costs, as long as the school cafeteria has sufficient capacity. Importantly, the C+10 mode also eliminates the share of students that have to go hungry all morning, i.e. that do not receive breakfast at home, and come to school too late

Table 1.11: Estimation results for school breakfast caloric intake

	school 1		school 2		school 3		all	
T1	-51.770		-6.545		6.503		-21.191	***
T2	-216.604		0.610		81.777		-201.397	
time to eat	1.824**	***	1.090	***	0.902	***	0.969	***
walk / bike	-1.581		4.586		3.150		3.649	
food before	-27.134	***	-19.006	***	-12.045	***	-17.002	***
hunger	-0.141		2.281	***	2.564	***	2.032	***
age	528.547	**	50.053		-184.901		361.094	**
BMI	9.979		0.039		10.593		0.984	
student FEs	yes		yes		yes		yes	
menu	yes		yes		yes		yes	
day of week	yes		yes		yes		yes	
students	90		147		193		430	
observations	982		1446		2214		4642	

FEs = fixed effects
* (**) [***] significant at 10 (5) [1] percent, based on robust standard errors clustered at the individual level

to participate in school breakfast. This is an important policy consideration.

If universally free breakfast is financially feasible, our findings make a compelling case for classroom implementation, with a total boost of 80% in participation probability for time-constrained students, and a still considerable increase of 60% for the typical student compared to standard provisions.

It would be interesting in future research to determine exactly what drives the sizable classroom-location effect we find for our sample. Even under our quarantined cafeteria setting participation rates are well below those observed for the classroom. Is it the calming influence of familiar surroundings, the absence of distractions that come with a larger group of students in the cafeteria, the low tolerance for disruptions in the classroom, or the encouraging presence of a familiar teacher? A better understanding of this “pure classroom effect” may help to design cafeteria implementation that adopts some of these factors, possibly allowing for further improvements in participation under standard location and pricing.

A second fruitful avenue of research would obviously be to tease out the pure price effect that comes with universally free implementation, controlling for location and - ideally - using individual-level data. This would provide further guidance as to the participation rates that could be achieved if universally free breakfast were to be served in the cafeteria.

Chapter 2

How Much Does Food Environment Matter: A Case Study of the Value of Food Environment in Dan River Region

2.1 Introduction

Food environment is defined by the Centers for Disease Control and Prevention (CDC) as the physical presence of food that affects individual food consumption, including the distribution of food stores, food service and any physical entity or system that affects individual access to food [18]. An area with less healthy food environment is often referred to as "food desert". In other words, the term "food desert", whose origination dates back to the early 1990s in Scotland [11], is used to describe an area with "limited access to affordable and nutritious food, particularly such an area composed of predominately lower-income neighborhoods and communities" [40]. In 2015, an estimated 19 million people in the U.S. lived in low-income and low access census tracts and live more than 1 mile (urban areas) or 10 miles (rural areas) from a supermarket [80]. In 2010, the Obama Administration announced the Healthy Food Financing Initiative (HFFI) to bring grocery stores and other healthy food retailers to food deserts [72]. The HFFI was motivated by the assumption that better food availability and accessibility will promote healthy food consumption.

However, the findings on the effects of food availability and accessibility are mixed. In the current literature, food environment is mainly measured as availability (e.g., the number or density of food outlets within a specified radius), accessibility (e.g., the distance to the nearest food outlet) and affordability (e.g., actual or perceived prices or price index of food) dimensions of food [21, 22, 28, 37, 77]. There is evidence to support that greater fast food availability and accessibility, is associated with more fast food meals consumed [37], less fruit and vegetable intake [64], and higher probability of being obese [36, 37]. At the same time, better access to fresh food or grocery stores was found to be positively related to fruit and vegetable intake, and negatively associated with obesity [21, 64]. These relationships are most evident for low-income minority adult residents [36, 37] and the majority of these studies are associations between existing food resources. A few quasi-experimental studies have explored the potential impact on diet and health outcomes when planting a new supermarket in deprived areas [30, 31, 35]. These studies found the addition of a supermarket did improve residents' perceived access to healthy food but the addition of a supermarket to the area did not lead to an increase in fruit and vegetable consumption for adults.

Another area of interest in the food environment literature includes the impact of food retailers such as fast food or supermarket for children and adolescents. For children in California public schools, the presence of a fast food restaurant within a small distance from school was found to be associated with increase in obesity rates [32]. However, there is no association found between fast food restaurant availability and adolescents' body weight outcomes based on national surveys [7, 76]. In addition, introducing a new supermarket in the community was not found to have a significant impact on fruit and vegetable intake for children [39].

The current HFFI policy does not consider the dimension of food affordability (i.e., food price) and acceptability (i.e., quality of food and the degree of fit into culture norm), two additional aspects of the food environment. A handful of studies have explored the impact of price on consumption of healthy foods such as fruit and vegetables and unhealthy foods (i.e., fast food). For example, Beydoun et al. [12] demonstrated that higher fast food price was associated with lower fast food consumption and higher intake of fruits and vegetables for children. Lower prices of fruits and vegetables is positively related to higher levels of fruits and vegetables consumption for US young adults [78]. At the same time, there is evidence showing that customers are more likely to purchase food of higher quality (heathier food, for example, vegetables and salads) in restaurants when nutrient information was displayed [29].

To improve the effectiveness of HFFI or HFFI-like policies, additional studies on food environment are needed. Taylor and Villas-Boas [88] estimated consumers' food outlet choice based on food outlet type and household attributes using a multinomial mixed logit model. By inferring the willingness-to-pay in distance traveled to shop at a particular type of food outlet, the authors found that households are willing to pay approximately \$15 per week for living one mile closer to superstores, supermarkets, and fast food restaurants. On the other hand, the households needed to be compensated with \$8-\$10 weekly for living closer to a grocery store and farmers market. However, they only considered the accessibility dimension and did not account for other important outlet features, such as, price and quality of food. In addition, willingness-to-pay in their study is a weekly measure. Our study aims to quantify the willingness-to-pay of residents through hedonic property methods based on the housing market, and provide a long-term measure of willingness-to-pay towards the three dimensionsavailability, accessibility and affordability of the food environment. The hedonic property method is a form of revealed preference valuation method based on housing prices. It treats a property as a composite good with associated housing and environmental attributes, and infers from housing prices the implicit price of, or marginal willingness-to-pay (MWTP) of homebuyer for its attributes. It is often applied to value non-market goods, including some environmental amenities and attributes, such as closeness to parks, air quality, and noise level. In this paper, we aim to quantify the preference of residents for food environmental attributes measured in multiple dimensions—availability, accessibility, acceptability (including the variety, quality and price of healthy food) of food using a comprehensive data sets for Dan River region near the border of Virginia and North Carolina.[1]

To our knowledge, this is the first study using housing market prices to assess and quantify consumers' preference towards food environment attributes. The potential omitted variable issue was controlled for by applying a spatial-lag model with a data-inferred spatial weight matrix. The weight matrix we choose is the one that provides the best fit in our data in terms of log-likelihood instead of an arbitrary weight matrix without justification as often used in the literature [6, 25, 61]. Furthermore, we examine the existence of willingness-to-pay heterogeneity between urban and rural areas, and among different types of food outlets, including healthy and less healthy food sources. Our study not only shows that a constant radius to define the relevant food environment for residents of any region is not appropriate for our application but also provides an empirical procedure to define the relevant food environment around a given residence's location.

Our findings suggest that, on average, the Dan River region residents do not prefer food environment improvement. To be specific, introducing a new fast casual restaurant or a new grocery store in a given neighborhood, or reducing the average traveling distance to grocery stores, or improving healthy food offerings in local food outlets, are found to be perceived as either null or negative. It provides one possible explanation to the ineffectiveness of food environment improving projects documented in the literature.

The remainder of this paper is organized as follows: the next section gives some background on the hedonic property method and its related econometric issues. This is followed by our econometric methods, and an empirical section that introduces the data, provides descriptive statistics, and presents estimation results and computed marginal benefits. The final section concludes.

[1] In this study, we define food acceptability as the variety, price, and quality of healthy food items supplied in food outlet. This is different from its traditional meaning in past literature, for example, see Cardello and Sawyer [17] and King et al. [62].

2.2 The Hedonic Property Method

Based on Rosen [81] hedonic hypothesis that goods are valued for their utility-bearing attributes or characteristics in a product differentiated market, the associated hedonic (implicit) prices for attributes can be estimated by regression analysis on the price function (product price regressed on characteristics). The consumer's marginal willingness-to-pay for a small change in quantity of a particular attribute can be inferred directly from an estimate of its implicit price, that is, its estimated coefficient, if we assume the market is in equilibrium [81]. The hedonic analysis has been mainly applied to value non-market goods, including school quality [14], hazardous waste sites [47, 53], open space [59], urban forest [92], mountain pine beetles [24], water quality [63], and air quality [8, 20, 60, 85]. As the fundamental need of humans, the environment of food plays a role in homebuyers' decision-making on home purchase if we assume that homebuyers have full information of food environment. Nevertheless, there is no explicit market for food environment as a whole. Therefore, the hedonic property method can help us to obtain the implicit prices for each dimension of the food environment.

However, several econometric issues common to the hedonic analysis need to be addressed: omitted variable problems and the price function's functional form choice. In this study, we mainly focus on the omitted variable problem, and choose a simple linear functional form for the price function to make estimation and interpretation tractable [6, 25, 60, 89]. To put the omitted variable problem into context: No matter how extensive the data set is, researchers may not observe all attributes of a home (either at the property or community level) that are valued by the homebuyers (therefore are relevant to the homebuyers' decision making). The omission of these unobserved yet relevant variables leads to biased estimates. This is especially true when the omitted variables are correlated with observed variables, for example, school quality is not observed in our study but it might be correlated with income level. The availability of panel data (i.e., repeated observation of the same property over several years) enables the property fixed effect model to address the issue if those omitted variables are time-invariant [96]. If they are time-variant the issue can be addressed with a quasi-experimental approach [53, 91] or a rational expectations approach [8].

The omitted variable issue becomes particularly troublesome if only cross-sectional data are available. One common option is to include spatial fixed effects to control for unobserved neighborhood influences that vary across space. The spatial scale could be city [101] or county [59] depending on the study area, although these scales are considered to be too coarse by Abbott and Klaiber [1]. In the last decade, the increasing availability of geographic information systems (GIS) data and the breakthroughs in spatial econometrics enable researchers to capture the unobserved neighborhood effects by taking the spatial dependence structure of housing market into consideration beyond mere spatial fixed effects [4, 66]. Two spatial models are common in applications: the spatial error model and the spatial-lag model. The spatial error model assumes that the spatial correlation arises from the omitted variables that follow a spatial pattern [15, 63]. The spatial-lag model, on the other hand, captures the indirect effects of neighboring housing prices on the price of each property, in addition to the direct effects of property level and neighborhood level characteristics [6, 48, 60]. Alternatively, matching methods can be applied to cross-sectional data to form "treatment" and "control" groups that are matched by housing, neighborhood, and other temporal and spatial characteristics in order to recover the capitalization value of non-market goods [2].

Since our paper utilizes a cross sectional data set, we are employing these cross-sectional methods to address potential omitted variable issues. Furthermore, we examine the robustness of results across methods, and discuss the additional insights each reveals.

2.3 Methods

The spatial error model and the spatial-lag model are two basic ways to incorporate spatial dependence structure into a cross-sectional regression model. The spatial error model typically assumes that there are omitted variables with spatial patterns in the hedonic price function, which results in a spatially autocorrelated error term. A hedonic spatial error model in linear form is given as:

$$\begin{aligned} P &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3 + \epsilon, \\ \epsilon &= \lambda\mathbf{W}\epsilon + u, \end{aligned} \tag{2.1}$$

where P represents housing prices, $\mathbf{X}_1$ is a vector of housing structural characteristics, $\mathbf{X}_2$ is a vector of neighborhood characteristics, and $\mathbf{X}_3$ is a vector of environmental characteristics, which can include, for example, air quality, water quality, and noise level, this study focuses on food environment. The error term function captures the spatial pattern of omitted variables with an autoregressive coefficient, λ, a $n \times n$ spatial weight matrix, $\mathbf{W}$, and an independent and identically distributed (i.i.d.) error term u. The ordinary least squares (OLS) estimator remains unbiased if equation (2.1) is the true model, but is no longer efficient. Instead, efficiency can be gained by maximum likelihood or a generalized moments approach. The marginal implicit price from a spatial error hedonic model is the coefficient vector $\boldsymbol{\beta}_3$ as in the traditional linear hedonic model.

The spatial-lag model implicitly assumes that the spatially weighted average of housing prices in a neighborhood affects the price of each house, as in the process of the standard real-estate appraisal, in addition to the direct effects of housing and neighborhood characteristics. A hedonic spatial-lag model in linear form can be written as:

$$P = \rho\mathbf{W}P + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3 + \epsilon, \tag{2.2}$$

where ρ is a spatial autoregressive parameter, $\mathbf{W}$ is the spatial weight matrix, and ϵ is assumed to be a vector of i.i.d. errors. The OLS estimates are biased and inconsistent for the spatial-lag model due to the endogeneity introduced by the spatial lag $\mathbf{W}P$ term. We apply a spatial two-stage least squares estimator (S-2SLS) to obtain consistent estimates

[60]. The marginal implicit price of a spatial-lag hedonic model is $\boldsymbol{\beta}_k[I - \rho\mathbf{W}]^{-1}$, which can be further separated into a direct effect and an indirect effect, see LeSage and Pace [66] for details. Overall, in our application, the spatial-lag model outperforms the spatial error model in terms of model-fit (based on the log-likelihood), and spatial dependence tests.[2] Therefore, the optimal weight matrix decisions are made based on spatial-lag model only in our study.

To identify spatial models, the spatial weight matrix $\mathbf{W}$, which captures potential spatial interactions of properties, needs to be specified exogenously to the model by researchers [66]. We mainly focus on distance-based spatial weight matrices, rather than contiguity weight matrices that apply to areal data. Typical distance-based choices include binary distance weight matrix, inverse distance weight matrix, and inverse squared distance weight matrix [6, 15, 25, 60]. The properties within a specified critical distance d_w are considered as neighboring houses that can affect the price of the property in the center of the radius. The effect sizes as captured by the elements in the weight matrix can be set equal across all neighboring houses, or proportional to the inverse (squared) distance between properties, in the weight matrices mentioned above.[3] However, it is rare that a study offers theoretical or empirical guidance in the weight matrix specification. In this paper, we turn to our data in order to choose the optimal weight matrix that best capturing the spatial dependence structure of neighboring properties. We chose the optimal weight matrix that provides the maximum log-likelihood among those spatial-lag models using a binary distance weight matrix, an inverse distance weight matrix, and an inverse squared distance weight matrix in a range of critical distances from 0.1 to 10 miles with 0.1-mile increment.

One of the assumptions of the hedonic model is that the MWTP inferred is related to attribute changes that would not disrupt the market equilibrium. For an area with very few certain food sources, opening one more or closing one will have higher likelihood to result in a new market equilibrium as it might affect residents' shopping choices and travel cost, and then probably their housing choices. In our case, we are particularly concerned about grocery store since there are a limited number of grocery stores available in our study area. To assess the sensitivity of our findings to this market equilibrium assumption, we apply propensity score matching to examine whether there is a difference in housing prices for properties with grocery store access within the specified radius and properties without grocery store access within the radius. The radius is specified at the optimal cut-off distance of food environment measurements found for the alternative definition sample in the spatial models. The properties with grocery store access ("treatment" group) are matched to the properties without grocery store access ("control" group) on the housing structural characteristics, neighborhood characteristics, and food environment measurements for other types of food outlets. Only the nearest neighbor in the "control" group in terms of estimated

[2] This holds for most spatial weight matrices that we have tested. For few weight matrices, we may get better model fit with a spatial error model.

[3] By convention, the diagonal elements of weight matrix are set to be zero, and row elements are standardized such that they sum to one.

distance (i.e., propensity of having access to grocery store within the radius) is matched to each observation in the "treatment" group, with replacement. This is because we have more "treatment" observations than "control" observations. We force exact matching on the number of bedrooms and the number of bathrooms, and restrict the distance on other structural and neighborhood characteristics within 0.2 standard deviations for city sample and 0.25 standard deviations for county sample (caliper matching).

2.4 Empirical Application

2.4.1 Data

Our empirical study focuses on the Dan River region, which is located in south-central Virginia and north-central North Carolina, and consists of two cities (Danville and Martinsville) and three counties (Henry, Pittsylvania and Caswell). The Dan River region is one of the most health disparate regions in the Virginia. The County Health Rankings, based on work of the University of Wisconsin Population Health Institute [93], ranked the city of Danville, Henry county and Pittsylvania county as 132^{th}, 127^{th} and 72^{th} out of 134 counties in Virginia, respectively, and Caswell county as 72^{th} out of 100 counties in North Carolina in 2016. The rankings were based on health outcomes (length and quality of life) and health factors (health behaviors, clinical care, social and economic factors, and physical environment). The Dan River region is also an economic disparate area. According to the U.S. Census Bureau and Bureau of Labor Statistics, an estimated 19.6% of individuals in this area were below poverty line during the period 2010 to 2014 (the average in Virginia was 11.5%), and the unemployment rate was estimated at 9.6% in April of 2012 (the average in Virginia was 6%).

This paper utilizes unique first-hand surveillance food environment data of the Dan River region. These unique data were collected by research partners of Dan River Partnership for a Healthy Community (DRPHC) from 2011 to 2012.[4] The researchers enumerated and systematically audited all 483 food outlets of all types in the Dan River region based on the Nutrition Environment Measures Survey (NEMS) designed specifically for stores and restaurants [50, 82]. The food outlets in this region include 39 fast casual restaurants, 119 fast food restaurants, 135 sit-down restaurants, 38 grocery stores, 106 convenience stores, and 44 other-type (pharmacy or dollar) stores.[5] This rich data set provides the exact geographic location of each food outlet, and NEMS score calculated based on the variety, quality and price of healthy food provided in each food outlet (ranging from -12 to 39 with higher score

[4] The food environment auditing is part of a larger community-based participatory research initiative, which quantifies the built environment including all food and physical activity outlets, and collects health outcomes data of residents by random digit dialing. See Hill et al. [55], Chau et al. [19], and Hill et al. [56] for related studies.

[5] Two specialty restaurants were excluded from our study to keep the constructions of food environment measurements simple.

representing better healthy food offering).

Due to the non-market good nature of the food environment, we need property transaction data to infer and quantify resident's preference towards food environment attributes. We purchased the property transactions data from CoreLogic. The data set contains property characteristics information for those properties having transactions from 2010 to 2015 in the Dan River region, including sale price, exact address, lot size, living space in square feet, year built, number of bedrooms, and number of bathrooms. We only use arms-length transactions data for single-family residences.[6] Majority of properties in our data set were transacted after 2010. All food outlets were audited between 2011 and 2012, and all stores were still in existence two years later in another round of audit. We assume that this also holds true for restaurants. Therefore, we make the assumption that all food outlets exist during the property transaction period (2010-2015). The volume of property transactions in Martinsville is very low and only a small number of transactions are arms-length transactions on single-family residences, and thus these observations are not included in our study. The properties and food outlets are geocoded according to their addresses. Figure 2.1 shows how they are distributed across the Danville city and the surrounding three counties. The relevant neighborhood-level information, for example, population, race composition, average education and income levels, were obtained from the Census 2010 and the 2014 American Community Survey on the website of U.S. Census Bureau at the block group level.

2.4.2 Food Environment Measurements

Armed with the detailed geocoded food outlets and residence information, we are able to construct relevant food environment measurements for each property. We measure food environment in three dimensions—availability, accessibility, and acceptability. Specifically, we are interested in two sets of food environment measurements. We refer to the first set of food environment measurements as the "strict" definition of food environment. In the strict version of the definition, availability of food is captured by the number of food outlets within a specified cut-off driving distance d_s from each property. We use the driving distance (in 0.01 mile) to food outlets within the cut-off distance averaged over these food outlets to measure food accessibility.[7] For the acceptability dimension, we construct an inverse-distance

[6] The transactions happened between family members or companies with related shareholders are not considered as arms-length transactions. We restrict our sample to arms-length transactions to ensure that both parties in the deal are acting in their own self-interest and are not subject to any pressure or duress from the other party. At the same time, single-family residence is one of the most representative type of properties transacted in the housing market. It enables us to make inference on the preference of residents (living in a family) based on their location.

[7] We calculate average driving distance in the unit of 0.01 mile to better capture the effect of a small change in accessibility on housing price in hedonic analysis.

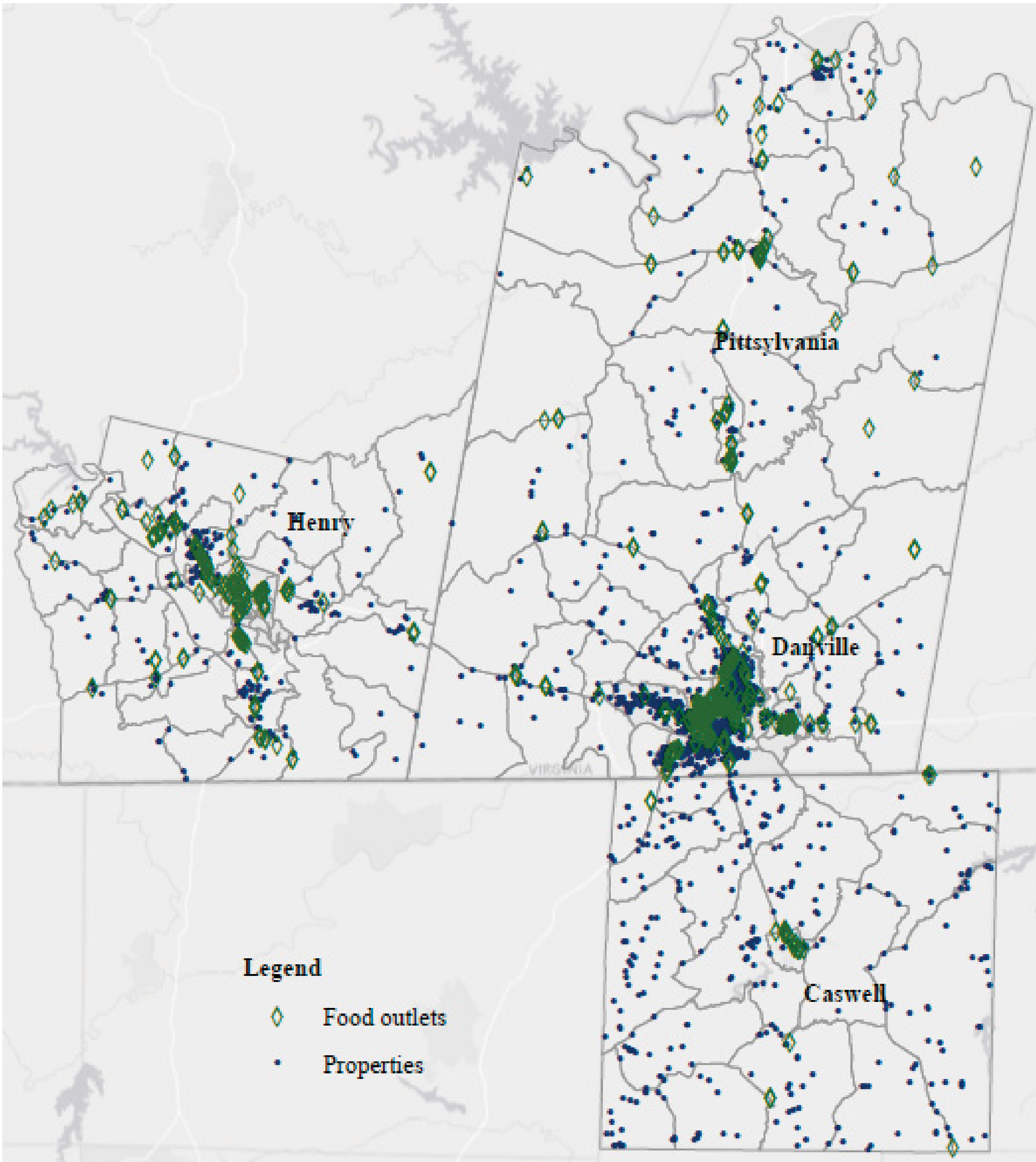

Figure 2.1: Food outlets and properties in the Dan River region

weighted score similar to that in Leggett and Bockstael [63]:

$$\text{weighted score} = \left(\frac{S_1}{D_1} + \frac{S_2}{D_2} + \cdots + \frac{S_m}{D_m}\right)/D,$$
$$D = \frac{1}{D_1} + \frac{1}{D_2} + \cdots + \frac{1}{D_m}, \tag{2.3}$$

where S_j is the audit score based on variety, quality and price of healthy food options in food outlet j, D_j is the driving distance from property to food outlet j, and m is total number of this type of food outlets within the cut-off driving distance d_s from the property. The inverse-distance weighted score is calculated for each property and for each type of food outlets. It quantifies the acceptability of nearby food outlets for a given property, in terms of selection, quality and price of healthy food options, and assumes that acceptability measure decays with distance. For example, assuming there are two grocery stores available for a given property within the cut-off driving distance, say, 5 miles. The driving distances to the two grocery stores are 1 mile and 3 miles, and the two grocery stores receive scores of 20 and 30, respectively. Then, we can calculate the weighted score of grocery store for this given property as $\left(\frac{20}{1} + \frac{30}{3}\right)/\left(\frac{1}{1} + \frac{1}{3}\right) = 22.5$. That is, the given property has acceptability of grocery store scored at 22.5, which is between 20 and 30 but closer to 20 since the grocery store with score 20 is closer to the property in distance and hence is considered as more relevant in the food environment measurements. The abovementioned food environment measurements are calculated for each type of food outlet—fast casual restaurant, fast food restaurant, sit-down restaurant, grocery store, convenience store, and other-type (pharmacy or dollar) store. Notice that it is not uncommon that a specific type of food outlet does not exist within the specified cut-off distance from a given property. In this case, we cannot obtain the average driving distance to this type of food outlets and the weighted score of this type of food outlets for the property. The observations with missing food environment measurements are not included in our strict definition sample.

In the second set of food environment measurements, we, again, capture the availability of food by the number of food outlets within a specified cut-off driving distance from each property. The accessibility and acceptability measurements of food environment depend on whether the property has a particular type of food outlets available within the specified cut-off driving distance. If there are available food outlets of a particular type within the cut-off distance, the measurements follow the strict version definition, i.e., average driving distance and inverse-distance weighted score. If no food outlets of a particular type is available within the cut-off distance, we measure accessibility as the driving distance (in 0.01 mile) to the nearest food outlet of this type, and acceptability as the score of the nearest food outlet of this type. We refer to this set of food environment measurements as an "alternative" definition of food environment.[8] This alternative definition gives more flexibility to food environment measurements. It mimics the behavior of residents who live in a community without a certain type of food outlet in a specified radius. In this case, they have the option to go to the nearest food outlet of this type. In the following sections, we report estimation results and marginal benefits for strict and alternative set of food environment measurements separately.

For both strict and alternative sets of food environment measurements, we need to specify

[8]With the alternative definitions, food environment measurements are not missing for almost all observations except for some properties located in a county. For these properties, the distance between property and the nearest food outlet is greater than 10 miles, which is the upper bound of our distance measure.

the cut-off driving distance. We follow the similar optimal weight matrix selection approach stated above to choose the optimal cut-off distance of food environment measurements for given properties based on log-likelihoods. For the strict version of food environment measurements, the comparison samples we used to find the optimal cut-off distance are the properties with complete (non-missing) food environment attributes in the smallest cut-off distance that we consider, that is, 1 mile for properties in the city, and 3 miles for properties in the county. For the alternative definition, we follow the rule of USDA to have the comparison samples with food environment attributes calculated in 1 mile for city sample and 10 miles for county sample [80].[9]

2.4.3 Descriptive Statistics

In total, there are 2,162 properties with one transaction record, unique geocoded location, and complete property information.[10] Among them, 1,300 observations are located in the city of Danville, and 862 observations are distributed across Henry county, Pittsylvania county, and Caswell county. Due to heterogeneity in characteristics of housing, neighborhood, and food environment, the city and county samples are analyzed separately.

With the strict definition on food environment measurements, the final city sample and final county sample have 856 observations and 559 observations, respectively.[11][12] Table 2.1 presents the summary statistics for the final samples under the strict definition of food environment measurements. The average adjusted sale price (CPI-adjusted to 2010 price) for properties in the city final sample is $94,961, which is lower than the average adjusted sale price for properties in the county final sample, $108,743. Properties located in the county have, on average, larger lots, and are younger in age at time of sale than properties located in the city. Overall, the properties in the city and county have similar size of living space (approximately 1650 square feet), similar number of bedrooms (approximately 2.4 bedrooms),

[9] The optimal weight matrix should be specified for the comparison samples before comparing different cut-off driving distances for food environment measurements.

[10] For 255 properties, more than one sale records exist. We only kept the transaction that happened closer to the food environment auditing time period (2011-2012) since the food environment data is only cross-sectional.

[11] The comparison sample for properties in the city has 328 observations that have complete food environment measurements in 1-mile cut-off driving distance. The optimal weight matrix for this comparison sample is inverse squared distance weight matrix with critical distance 0.2-mile. For a range of cut-off distances from 1 mile to 10 miles, the comparison sample with food environment measurements in 1.5-mile cut-off distance has the largest log-likelihood. The observations with complete food environment measurements in 1.5 miles constitute the final city sample.

[12] The comparison sample for properties in the county has 156 observations that have complete food environment measurements in 3-mile cut-off driving distance. The optimal weight matrix for this comparison sample is binary weight matrix with critical distance 4.8 miles. For a range of cut-off distances from 3 miles to 10 miles, the comparison sample with food environment measurements in 8.6-mile cut-off distance has the largest log-likelihood. The observations with complete food environment measurements in 8.6 miles constitute the final county sample.

and similar number of bathrooms (approximately 1.8 bathrooms). The neighborhood attributes of properties in our final samples are not very different except for the fact that the city properties have a higher percentage of African American residents and more educated residents, on average. In terms of food environment measurements, the properties in the city and county are measured under different cut-off driving distances, that is, 1.5 miles and 8.6 miles, respectively, as we found in the optimization step. Although the measurements differ between properties in the city and county, we observe more fast food restaurants, more sit-down restaurants, and more convenience stores within the optimal cut-off driving distances for both city and county properties. The average driving distance to the six types of food outlets are very close. This hold for both city and county samples. Finally, fast casual restaurants and grocery stores have higher acceptability, in terms of variety, quality and price of healthy food options, out of the six types of food outlets for properties located in both city and county.

The alternative definitions of food environment measurements relax the requirement for the existence of at least one food outlet of each type within the specified radius.[13] We are able to retain the highest number of observations under the alternative definitions of food environment measurements for both city and county samples. The final city sample and final county sample has 1,300 observations and 626 observations, respectively.[14][15] Table 2.2 gives the summary statistics for the final samples under alternative definitions of food environment measurements. The properties in the city sample have an average adjusted sale price of $99,624, and the county counterparts have an average adjusted sale price of $106,979. The structural housing characteristics and neighborhood characteristics of city properties and county properties under alternative definitions of food environment measurements are not very different from that under strict definitions, in general. The number of food outlets of each type decreases under the alternative definitions since the cut-off driving distance is decreased from 1.5 miles to 1.1 miles for city sample, and from 8.6 miles to 4.3 miles for county sample. This also holds for the accessibility measure. However, we still observe higher acceptability measures for fast casual restaurants and grocery stores compared to other types of food outlets under the alternative definitions.

[13] The detailed percentages of observations that calculate food environment measurements based on the nearest food outlet of each type can be obtained from authors upon request.

[14] The comparison sample for properties in the city has 1,300 observations that have food environment measurements calculated under alternative definitions in 1-mile cut-off driving distance. The optimal weight matrix for this comparison sample is inverse distance weight matrix with critical distance 0.3-mile. For a range of cut-off distances from 1 mile to 10 miles, the comparison sample with food environment measurements in 1.1-mile cut-off distance has the largest log-likelihood. The observations with food environment measurements calculated in 1.1 miles constitute the final city sample.

[15] The comparison sample for properties in the county has 626 observations that have food environment measurements calculated under alternative definitions in 10-mile cut-off driving distance. The optimal weight matrix for this comparison sample is inverse squared distance weight matrix with critical distance 7.4-mile. For a range of cut-off distances from 1 mile to 10 miles, the comparison sample with food environment measurements in 4.3-mile cut-off distance has the largest log-likelihood. The observations with food environment measurements calculated in 4.3 miles constitute the final county sample.

Table 2.1: Summary statistics of strict definition sample

variable	city		county	
CPI-adjusted sale price in thousands	94.961	(66.182)	108.743	(66.041)
lot size in acres	0.287	(0.216)	1.730	(2.448)
living space in sqft	1699.939	(746.440)	1625.016	(633.384)
number of bedrooms	2.398	(1.218)	2.440	(1.258)
number of bathrooms	1.875	(0.886)	1.846	(0.763)
age of building at sale	52.786	(18.239)	38.242	(19.776)
total population in block	1324.988	(435.963)	1558.252	(532.341)
% African American in block	43.152	(20.701)	23.856	(15.232)
% some college or above education in block	55.740	(14.549)	45.941	(10.212)
yearly per capita income in thousands in block	22.414	(8.278)	20.792	(5.216)
median age in block	42.650	(7.220)	44.231	(3.358)
% households receiving SNAP benefits in block	22.371	(13.304)	17.360	(10.149)
number of fast casual restaurants	3.495	(1.551)	8.041	(6.300)
number of fast food restaurants	7.317	(3.954)	23.497	(15.457)
number of sit down restaurants	8.230	(5.712)	26.923	(20.435)
number of grocery stores	1.681	(0.687)	7.011	(4.372)
number of convenience stores	4.854	(2.314)	16.333	(10.735)
number of other-type stores	2.047	(1.222)	8.792	(5.957)
access to fast casual restaurants	101.950	(20.586)	542.133	(187.224)
access to fast food restaurants	103.597	(21.438)	546.409	(169.022)
access to sit down restaurants	99.454	(19.994)	553.147	(178.628)
access to grocery stores	93.187	(27.803)	540.540	(164.537)
access to convenience stores	94.247	(19.021)	545.103	(123.786)
access to other-type stores	92.285	(31.905)	531.535	(183.788)
acceptability of fast casual restaurants	12.016	(7.324)	13.213	(7.995)
acceptability of fast food restaurants	1.159	(2.652)	1.454	(1.502)
acceptability of sit down restaurants	2.623	(1.881)	2.266	(1.123)
acceptability of grocery stores	28.248	(6.178)	26.020	(5.284)
acceptability of convenience stores	6.439	(2.455)	5.940	(1.415)
acceptability of other type stores	6.904	(1.982)	7.906	(2.551)
number of observations	856		559	

Accessibility measures are in the unit of 0.01 mile. Standard deviation in parentheses.

2.4.4 Estimation Results

As mentioned above, we get a better model fit with a spatial-lag model compared to a spatial error model in terms of log-likelihood. Specification tests also indicate that the spatial-lag model better captures the local spatial dependence structure in our sample.[16] Therefore, we only report the estimation results for the spatial-lag model in the following sections. One

[16]The robust Lagrange multiplier (RLM)-error tests for error dependence in the possible presence of a missing lagged dependent variable, and the RLM-lag the other way round. For both city and county final samples under strict and alternative definitions of food environment, RLM-lag test favors the choice of spatial-lag model.

Table 2.2: Summary statistics of alternative definition sample

variable	city		county	
CPI-adjusted sale price in thousands	99.624	(69.530)	106.979	(64.724)
lot size in acres	0.391	(0.489)	1.843	(2.611)
living space in sqft	1728.817	(748.044)	1620.304	(631.588)
number of bedrooms	2.482	(1.172)	2.450	(1.244)
number of bathrooms	1.928	(0.903)	1.837	(0.752)
age of building at sale	49.454	(20.168)	37.885	(20.594)
total population in block	1331.315	(422.606)	1548.166	(539.530)
% African American in block	39.697	(22.449)	23.974	(14.986)
% some college or above education in block	54.588	(14.339)	45.487	(10.053)
yearly per capita income in thousands in block	23.755	(9.900)	20.791	(5.007)
median age in block	43.584	(7.157)	44.197	(3.244)
% households receiving SNAP benefits in block	22.497	(13.378)	17.243	(9.818)
number of fast casual restaurants	1.294	(1.338)	1.872	(2.493)
number of fast food restaurants	2.622	(2.880)	6.193	(8.158)
number of sit down restaurants	3.247	(3.315)	5.802	(8.381)
number of grocery stores	0.768	(0.702)	1.855	(1.922)
number of convenience stores	2.493	(1.742)	3.909	(3.933)
number of other-type stores	0.952	(1.102)	2.589	(3.408)
access to fast casual restaurants	124.183	(88.136)	428.352	(230.694)
access to fast food restaurants	120.189	(88.417)	369.723	(194.992)
access to sit down restaurants	88.719	(38.494)	379.153	(183.246)
access to grocery stores	126.900	(90.422)	406.641	(229.762)
access to convenience stores	82.428	(29.105)	318.126	(148.977)
access to other-type stores	131.246	(96.321)	418.458	(225.030)
acceptability of fast casual restaurants	10.961	(11.716)	14.057	(11.832)
acceptability of fast food restaurants	2.369	(6.273)	0.925	(4.119)
acceptability of sit down restaurants	2.356	(3.115)	2.018	(2.689)
acceptability of grocery stores	30.182	(7.369)	26.621	(9.388)
acceptability of convenience stores	6.271	(3.324)	6.359	(2.922)
acceptability of other type stores	7.853	(2.814)	7.574	(3.063)
number of observations	1300		626	

Accessibility measures are in the unit of 0.01 mile. Standard deviation in parentheses.

point worth noting is that the quantified preferences or willingness-to-pay of residents in our study are for the food environment and are not for their shopping or consumption choices. The preferences for a specific type of food outlet may include the preferences for healthy or less healthy food, traffic, noise, and other factors related to a food outlet.

Strict definition of food environment

Table 2.3 reports S-2SLS estimation results for the city and county samples under the strict definitions of food environment measurements.[17] For properties located in the city, the spatial autoregressive coefficient is found to be significantly positive, suggesting a strong spatial similarity in housing prices. We get expected signs for most structural housing characteristics except for the number of bedrooms. That is, we observe that the housing prices are significantly positively related to lot size, living space, and number of bathrooms, and negatively related to number of bedrooms and age of building at sale. The unexpected negative sign of number of bedrooms suggests that the residents in the city might prefer larger bedroom and/or living room and/or kitchen to more bedrooms for a given living space size. We also find that the properties located in block groups with fewer and older residents (in median age), higher percentage of African Americans, and higher percentage of at least college degree holders were sold at a higher price, on average, when controlling for their structural housing characteristics and food environment attributes.

For food environment characteristics, more other-type (pharmacy or dollar) stores around the area and being closer to sit-down restaurants and other-type stores are related to higher housing prices, and hence are preferred by the residents in the city. On the other hand, the number of fast casual restaurants and grocery stores in the area are negatively associated with housing price. The effect of the availability of grocery stores needs to be interpreted with caution, which will be discussed further in the marginal benefits section. Furthermore, the residents in the city attach much less value to homes that are, on average, closer to fast food restaurants (only at 10% significance level) and grocery stores (at 5% significance level). Recall that grocery store is the type of food outlet that has the best healthy food offerings, as shown on the acceptability measure in Table 2.1, out of all types of food outlets. The preferences of fewer and living further away from the relatively healthier food sources indicate that the residents in this city sample are not concerned about or even have an aversion to healthy food environment around their residences. This is also confirmed by the estimates of the acceptability measure on housing prices. Our results show that for given availability and proximity of food sources, the better acceptability of fast casual restaurants and sit-down restaurants, negatively affect housing price in the city sample. Overall, the residents in the city sample do not show a preference towards a healthy food environment, which is consistent with the focus group findings with local residents showing a desire for high fat and high energy-dense "comfort foods". Unpublished process data from focus groups with the DRPHC community coalition and residents find food preference themes and strong cultural norms around "comfort foods" and southern cooking that is characterized by energy dense and high fat. By comparison, there is very little role modeling or social/cultural norms around healthful foods.

[17] The optimal spatial weight matrix for the final city sample under the strict definitions of food environment measurements is inverse distance weight matrix with critical distance of 0.4 mile. The optimal spatial weight matrix for the final county sample is inverse squared distance weight matrix with critical distance of 0.5 mile.

For properties located in the county, the spatial autoregressive coefficient is significantly positive but with a much smaller magnitude than for the city sample. This indicates that the properties in the county sample are much less spatially dependent compared to the properties located in the city. All estimated coefficients of housing characteristics have the expected sign, including the number of bedrooms, while none of the effects of neighborhood characteristics are significant. In terms of food environment, only food availability attributes are significantly related to housing price. Similar to city residents, the residents in the county sample prefer fewer fast casual restaurants. However, county residents show opposite preference as compared to their city peers: they prefer more grocery stores and fewer other-type stores.

Alternative definition of food environment

Table 2.4 presents S-2SLS estimation results for city and county samples under the alternative definition of food environment measurements.[18] Notice that the main difference between the strict and alternative definitions of food environment lies with the accessibility and acceptability of food. The alternative definition of food environment takes the accessibility and acceptability measure of the nearest food outlet (beyond the specified radius but within 10 miles) into account when such type of food outlet is not available within the specified radius.[19]

With the alternative definition of food environment measurements, the residents in the city are found to prefer living in areas with more sit-down restaurants and other-type stores. Compared to the city sample under the strict definition, the residents in the alternative definition sample appear to be indifferent to the availability of fast casual restaurants and grocery stores. When considering the driving distance to the nearest food outlet when the outlet is not available within the specified radius, we found that the city residents dislike living too close to a convenience store, and at the same time, they do not have clear preferences for the accessibilities of other food outlets. Even with the more robust measure on acceptability of food, we still find that residents living in the city prefer sit-down restaurants with less healthy food offerings.

The residents living in the county are still found to enjoy being surrounded by more sit-down restaurants, but less fast casual and fast food restaurants. Under the strict definitions of food environment, they appeared to care less about the availability of any types of stores, including grocery stores. They, however, did prefer living closer to grocery stores when we take alternative definitions on food environment. Except for the negative effect of acceptability

[18] The optimal spatial weight matrix for the final city sample under the alternative definitions of food environment measurements is inverse distance weight matrix with critical distance of 0.3 mile. The optimal spatial weight matrix for the final county sample is inverse squared distance weight matrix with critical distance of 7.4 miles.

[19] The radiuses for city and county sample under the alternative definition are 1.1 miles and 4.3 miles, respectively.

Table 2.3: Estimation results for strict definition sample

variable	city		county	
ρ	0.702***	(0.093)	0.060*	(0.038)
lot size in acres	40.077**	(17.845)	2.073**	(1.200)
living space in sqft	0.021***	(0.004)	0.048***	(0.005)
number of bedrooms	-3.066*	(1.692)	3.941**	(1.413)
number of bathrooms	9.743***	(2.811)	11.563***	(3.400)
age of building at sale	-0.361**	(0.144)	-0.720***	(0.112)
total population in block	-0.024***	(0.007)	0.002	(0.004)
% African American in block	0.523**	(0.256)	0.012	(0.165)
% some college or above education in block	0.860**	(0.341)	0.247	(0.218)
yearly per capita income in thousands in block	-0.652	(0.498)	0.673	(0.577)
median age in block	1.049***	(0.389)	-0.655	(0.745)
% households receiving SNAP benefits in block	0.157	(0.275)	-0.306	(0.234)
number of fast casual restaurants	-4.375**	(2.186)	-4.375**	(1.684)
number of fast food restaurants	1.609	(1.240)	0.677	(0.828)
number of sit down restaurants	0.853	(0.575)	1.667***	(0.553)
number of grocery stores	-12.914***	(3.742)	4.878*	(2.453)
number of convenience stores	0.436	(1.525)	-1.705*	(0.937)
number of other-type stores	8.390**	(3.725)	-3.522**	(1.545)
access to fast casual restaurants	0.100	(0.120)	0.034	(0.035)
access to fast food restaurants	0.354*	(0.196)	-0.005	(0.038)
access to sit down restaurants	-0.308**	(0.145)	0.029	(0.037)
access to grocery stores	0.182**	(0.092)	-0.020	(0.035)
access to convenience stores	-0.063	(0.149)	-0.013	(0.037)
access to other-type stores	-0.198**	(0.086)	-0.030	(0.036)
acceptability of fast casual restaurants	-0.993*	(0.524)	-0.154	(0.321)
acceptability of fast food restaurants	-1.375	(1.702)	-0.950	(1.782)
acceptability of sit down restaurants	-4.703**	(1.870)	-0.559	(2.459)
acceptability of grocery stores	0.322	(0.433)	0.675	(0.742)
acceptability of convenience stores	-0.799	(1.034)	-1.730	(1.695)
acceptability of other type stores	1.329	(1.492)	1.250	(1.569)
observations	856		559	

Accessibility measures are in the unit of 0.01 mile. Robust standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

of convenience stores, better healthy food offerings in fast casual restaurants and other-type stores affect the housing price significantly positively. Consistent with the findings under strict definition, the county residents have weak preferences for a healthier food environment.

2.4.5 Marginal Benefits

For the spatial-lag model with simple linear functional form in our study, the marginal implicit price is the derivative of the hedonic price equation with respect to each explanatory

Table 2.4: Estimation results for alternative definition sample

variable	city		county	
rho	0.408***	(0.061)	0.211***	(0.073)
lot size in acres	1.433	(3.259)	1.930*	(1.035)
living space in sqft	0.027***	(0.004)	0.045***	(0.005)
number of bedrooms	-2.491*	(1.373)	3.859***	(1.398)
number of bathrooms	11.304***	(2.526)	13.792***	(3.029)
age of building at sale	-0.476***	(0.096)	-0.622***	(0.122)
total population in block	-0.020***	(0.006)	0.003	(0.004)
% African American in block	0.294**	(0.147)	-0.069	(0.147)
% some college or above education in block	0.440*	(0.251)	0.228	(0.203)
yearly per capita income in thousands in block	-0.038	(0.330)	0.489	(0.522)
median age in block	0.830**	(0.335)	-0.892	(0.688)
% households receiving SNAP benefits in block	0.178	(0.174)	-0.353	(0.229)
number of fast casual restaurants	-2.286	(1.939)	-4.422**	(1.880)
number of fast food restaurants	-0.905	(0.979)	-1.420*	(0.796)
number of sit down restaurants	1.433*	(0.760)	1.903***	(0.673)
number of grocery stores	-2.892	(2.770)	-0.067	(2.705)
number of convenience stores	1.441	(1.382)	0.671	(1.220)
number of other-type stores	4.130*	(2.327)	-0.118	(1.527)
access to fast casual restaurants	-0.022	(0.068)	-0.003	(0.015)
access to fast food restaurants	0.115	(0.074)	0.021	(0.016)
access to sit down restaurants	-0.029	(0.056)	-0.013	(0.016)
access to grocery stores	0.027	(0.056)	-0.031*	(0.018)
access to convenience stores	0.124**	(0.056)	-0.004	(0.019)
access to other-type stores	-0.060	(0.059)	0.005	(0.018)
acceptability of fast casual restaurants	0.258	(0.227)	0.401*	(0.206)
acceptability of fast food restaurants	0.248	(0.361)	0.680	(0.563)
acceptability of sit down restaurants	-1.880***	(0.712)	-1.231	(0.816)
acceptability of grocery stores	0.057	(0.236)	0.358	(0.269)
acceptability of convenience stores	-0.087	(0.496)	-1.703**	(0.744)
acceptability of other type stores	-0.099	(0.879)	1.417*	(0.827)
observations	1300		626	

Accessibility measures are in the unit of 0.01 mile. Robust standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

variable, that is, $\boldsymbol{\beta}_k[I - \rho\mathbf{W}]^{-1}$. It is composed of a direct effect on housing price due to the marginal change in the explanatory variable, and an indirect effect on housing price through the changes in neighboring housing prices, which again are due to the marginal change in their explanatory variables. The marginal implicit price can be interpreted as marginal willingness-to-pay if we assume the housing market is in equilibrium [81].

Strict definition of food environment

Table 2.5 gives direct effects, indirect effects and total effects of the marginal benefits derived from S-2SLS estimates of city and county samples under the strict definition of food environment. We mainly focus on discussing those statistically significant marginal benefits of food environment measurements in this section (only those statistically significant estimates are shown).[20] The marginal benefit, on average, of closing one fast casual restaurant within the radius (1.5 miles for city, and 8.6 miles for county) brings a direct impact of $4,738 for an average household in the city, and $4,381 in the county. Opening an additional other-type store within the radius brings direct benefits of $9,085 for an average household in the city. In contrast, closing an other-type store within the radius gives a direct benefit of $3,527 to an average household in the county. For properties located in the county, we also find that opening one more sit-down restaurant and closing one convenience store within the radius gives them direct benefit of $1,670 and $1,708, respectively. The marginal benefit of the availability of grocery stores should be interpreted carefully for the Dan River region as the total number of grocery stores in this region is limited. The correct inference on marginal willingness-to-pay is based on the assumption that the housing market is in equilibrium. The opening and closing of a grocery store in this region may affect its equilibrium status. Therefore, we use the current findings as a benchmark for the effects of grocery store availability, and discuss them further when examining alternative definitions of food environment measurements, and estimating the effects of grocery store availability with propensity score matching.

For proximity to food outlets, living 50 feet (approximately 0.01 mile) further away from fast food restaurants and grocery stores in the city brings direct marginal benefits of $383 and $198, respectively. Living 50 feet closer to sit-down restaurants and other-type stores, on the other hand, provides $334 and $215 direct marginal benefits to an average household in the city. Finally, lowering the acceptability measure of fast casual restaurants and sit-down restaurants by one point generates direct marginal benefits of $1,075 and $5,093 on average.

Alternative definition of food environment

Table 2.6 presents the marginal benefits for city and county samples under the alternative definition of food environment. As before, we only focus on these statistically significant marginal benefits of food environment measurements. The marginal benefit, on average, of opening one sit-down restaurant within the radius (1.1 miles for city, and 4.3 miles for county) brings a direct impact of $1,478 for an average household in the city, and $1,929 in the county. The results indicate that closing one fast casual restaurant and one fast food restaurant in the area provides direct marginal benefits of $4,483 and $1,439 for an average household in the county, respectively. The households in the city still obtain a value of

[20] The complete results of the marginal benefits under strict and alternative definitions of food environment measurements can be obtained from authors upon request.

Table 2.5: Impact measures of food environment measurements for strict definition sample

variable	city			county		
	direct	indirect	total	direct	indirect	total
number of fast casual restaurants	-4.738**	-9.960	-14.698	-4.381***	-0.191	-4.573***
number of sit down restaurants				1.670***	0.073	1.742***
number of grocery stores	-13.984***	-29.399	-43.383	4.886**	0.213	5.099**
number of convenience stores				-1.708*	-0.075	-1.782*
number of other-type stores	9.085**	19.100	28.185	-3.527**	-0.154	-3.681**
access to fast food restaurants	0.383*	0.805	1.188			
access to sit down restaurants	-0.334**	-0.702	-1.035			
access to grocery stores	0.198*	0.415	0.613			
access to other-type stores	-0.215**	-0.451	-0.666			
acceptability of fast casual restaurants	-1.075**	-2.261	-3.336			
acceptability of sit down restaurants	-5.093**	-10.706	-15.799			

Accessibility measures are in the unit of 0.01 mile. The p-values for the impact measures are based on 500 simulations. *** significant at 1%, ** significant at 5%, * significant at 10%.

$4,260 from an additional other-type store within the specified radius. However, they are found to be indifferent to the change of availability of grocery stores in the community under the alternative food environment definition. For proximity to food outlets, the households living in the county are found to not care about accessibilities to any type of food outlets. On the other hand, the households living in the city obtain a marginal direct benefit of $128 by living 50 feet closer to convenience stores, on average. The results on the acceptability measures are mixed. Decreasing acceptability measure by one point for sit-down restaurants in the city and convenience stores in the county, on average, add a direct benefit of $1,939 and $1,727 for an average household. In contrast, an average household in the county obtains direct benefits of $407 and $1,437 if the acceptability measures of fast casual restaurants and other-type stores increase by one point. At the same time, we do find some indirect effects of food environment measurements, caused by spatial dependence structure of properties, for households living in the city and county. For example, an average household living in the city can obtain an indirect marginal benefit of $1,226, on average, by having neighboring households living in areas with lower acceptability measures for sit-down restaurants by one point. Overall, we do not observe preference to healthier food environment for residents in this region based on their estimated marginal willingness-to-pay.

2.4.6 Capitalization Value of Grocery Store Access

We apply propensity score matching to investigate whether having a grocery store available within the specified radius (the optimal cut-off driving distance found for alternative definition samples, that is, 1.1 miles for city and 4.3 miles for county) makes a difference in housing price. This serves as a robustness check for our estimation results on the effect of

Table 2.6: Impact measures of food environment measurements for alternative definition sample

variable	city			county		
	direct	indirect	total	direct	indirect	total
number of fast casual restaurants				-4.483**	-1.120*	-5.602**
number of fast food restaurants				-1.439*	-0.359	-1.799*
number of sit down restaurants	1.478*	0.935*	2.413*	1.929***	0.482*	2.411***
number of other-type stores	4.260*	2.694	6.954*			
access to convenience stores	0.128**	0.081*	0.208**			
acceptability of fast casual restaurants				0.407*	0.102	0.509*
acceptability of sit down restaurants	-1.939***	-1.226**	-3.166***			
acceptability of convenience stores				-1.727**	-0.431	-2.158**
acceptability of other type stores				1.437*	0.359	1.796*

Accessibility measures are in the unit of 0.01 mile. The p-values for the impact measures are based on 500 simulations. *** significant at 1%, ** significant at 5%, * significant at 10%.

availability of grocery store. For properties located in the city under the alternative definition of food environment, 100 homes without grocery store available within the radius ("control" group) are matched to 456 homes with grocery store available within the radius ("treatment" group), with replacement. The t-test on the matched data shows that there is no significant difference between the housing prices of the two groups. For properties located in the county, 24 homes without grocery store available within the radius ("control" group) are matched to 129 homes with grocery store available within the radius ("treatment" group), with replacement. Similarly, we do not detect a significant difference for housing prices between the control and treatment groups in the county. The null capitalization value of one additional grocery store within the radius aligns with the findings from the spatial model estimation results under the alternative definition of food environment. Notice that the results of propensity score matching could vary with different specifications of cut-off driving distance, which is an arbitrary threshold we choose to divide the full sample into control and treatment groups. Even with this note in mind, the results of propensity score matching and spatial model under alternative definition suggest us to be cautious when interpreting the marginal effect of availability of grocery store. Overall, the residents in the Dan River region do not have a preference towards better availability of healthy food.

2.5 Conclusion

The topic of food environment has attracted substantial academic and policy attention in recent years. The implicit assumption that better food environment will result in healthier eating behaviors and better related health outcomes is behind all the policies and intervention programs that aim at improving food environment. It is vital to understand the current population's preference towards those food environment attributes in order to test, validate and improve those intervention programs. However, there is no existing market that can signal and quantify population preference towards food environment attributes such as availability, accessibility and acceptability. This paper takes a unique approach at addressing this need through creative usage of a first-hand food environment surveillance data set along with the associated local housing market transaction data. Through hedonic property models and spatial econometrics, we illustrate how this approach can quantify the food environment preference through a case study. Due to our data limitations, our results are not nationally generalizable but signal the existence of non-trivial local resistance towards healthy food environment. Another limitation is that, to infer the preferences of residents towards food environment from the housing market we have to make the assumption that homebuyers have full information of food environment.

In our study, the marginal implicit price for a small improvement in the food environment, for example, by introducing a new fast casual restaurant or a new grocery store in a given neighborhood, or reducing the average traveling distance to grocery stores, or improving healthy food offerings in local food outlets, is found to be either null or negative for residents in the city of Danville. In fact, the direct marginal willingness-to-pay for a small deterioration in healthy food offerings (one-point decrease in acceptability measure) of sit-down restaurants is as high as $5,093. Although the introduction of a new grocery store can bring a total marginal benefit of $5,099 to an average household in the surrounding three counties, the benefit gain is relatively small. Especially, it becomes null when we examine it under alternative definitions of food environment and by propensity score matching. The overall preference of residents in the Dan River region for an improved food environment is revealed as negative especially in the city sample. Our results provide one possible explanation to the ineffectiveness of opening a new supermarket on residents' fruit and vegetable intakes in other health and economic disparate regions found in the literature. Improving merely the physical food environment is not enough to change the individual eating behavior. The more fundamental and urgent change needs to happen with residents understanding and recognition of healthy eating. Meanwhile, providing more monetary and non-monetary incentives along with environment changes to offset those general aversions will be needed to achieve program goals.

Chapter 3

Humans vs. Machine: The Effects of Interpreter and Image Characteristics on the Accuracy of Cloud Interpretation for Satellite Images

3.1 Introduction

With the development of satellite technology, millions of Landsat images have been acquired. The images are a unique and valuable resource for research and applications in agriculture, cartography, forestry, regional planning, education and other fields. For example, Yuan et al. [99] utilize the multitemporal Landsat data in Twin Cities Metropolitan Area of Minnesota to map and monitor land cover change from 1986 to 2002 in this region; Liu et al. [67] reconstruct spatial and temporal patterns of cropland across China for the time period of 1990-2000 by using the relevant Landsat data. However, many of these Landsat images are covered by clouds and/or cloud shadows. This is especially true for multi-spectral image.[1] The brightening effect of the clouds and the darkening effect of cloud shadows could lead to misleading results in remote sensing analysis, including inaccurate atmospheric correction, biased estimation of Normalized Difference Vegetation Index (NDVI) values, mistakes in land cover classification, and false detection of land cover change [102]. There are also studies examining the actual clouds and atmospheric composition through satellite images (e.g., Arking and Childs [5], and Eck et al. [38]). Therefore, detecting clouds and cloud shadows is an important initial step before the Landsat images can be effectively utilized for further analysis.

[1] We can see through clouds and/or cloud shadows for image with large wavelength, for example, radar image.

However, accomplishing this for a large volume of satellite images is beyond the capability of typical remote sensing research institutions. Some automated algorithms have been developed for cloud, cloud shadow and snow masking purposes for Landsat images [41, 57, 102]. The algorithms are able to process images efficiently at a low cost, and at the same time, achieve reasonable interpretation accuracy. For instance, Zhu and Woodcock [102] report that their Function of mask (Fmask) algorithm, a commonly used automated algorithm in cloud and cloud shadow detection, reaches an average cloud interpretation accuracy as high as 96.4% for a globally distributed set of reference data, and is much less time consuming compared to manual screening of clouds and cloud shadows.

Nevertheless, the performances of Fmask for satellite images in some regions are unsatisfactory. For instance, Qiu et al. [79] find for a set of satellite images in South America, Europe, Asia and Australia, the overall accuracies of Fmask range between 5.51% and 51.55%. In addition, there are evidences showing that human interpretation is the most accurate method for image classification (e.g., Chuvieco [23]). Human interpreters recruited through Amazon.com's Mechanical Turk (AMT) platform with limited training were found to outperform Fmask for some satellite images by our earlier study [98].[2] Such practice of outsourcing work to the nonspecific, paid or unpaid, and generally large crowd, normally internet users, is often referred to as "crowdsourcing." It has been widely applied to many fields, including remote sensing, where geographic information is collected from volunteer crowds for land cover interpretation [46, 84], damage assessment [9, 49], and disaster management [52].

Building upon our earlier findings [98], this study investigates how the demographic characteristics of human interpreters on the AMT online platform and the scientific characteristics of satellite images relate to the accuracy of human cloud interpretation. In addition, we examine how human interpreters compare to CFmask (Fmask algorithm written in C) in terms of interpretation accuracy for twelve different satellite scenes. This allows us to determine whether human interpreters on the AMT platform are a reliable alternative to the automated algorithm CFmask for cloud interpretation work. Our results provide evidence that, overall, human interpreters on the AMT platform achieved higher accuracy compared to CFmask for an "average" image. Furthermore, we examine whether human interpreters and CFmask are more likely to miss an actual cloud and/or cloud shadow (false negative) or to incorrectly identify clouds (false positive) in a given cell. We find that human interpreters and CFmask are both more likely to make a false negative decision than a false positive decision without significant difference.

[2] For the rest of this paper, we will use "human interpreter" and "participant" interchangeably to refer to the individuals who participated in our online experiment.

3.2 Methods

The experiment in our study focuses on twelve satellite images in different regions with diverse backgrounds and cloud cover. We divided each image into eight-by-eight cells (64 cells for one image in total), and collected cloud interpretation result for each cell. Our analysis is separated into two parts: image level analysis, and cell level analysis. In the image level analysis, we mainly look at how the characteristics of human interpreters and image affect interpretation accuracy, and how human interpreters compare to CFmask. For the cell level analysis, we focus on whether human interpreters and CFmask are more likely to miss an actual cloud and/or cloud shadow (false negative) or to incorrectly identify clouds (false positive) in a given cell.

Specifically, in the image level analysis, we focus on image-level accuracy (ILA). Image-level accuracy is defined as the proportion of correctly identified cells out of 64 cells for a given image relative to the expert's interpretation result, that is, ILA = (number of correctly specified cells) / 64. It is a fractional dependent variable bounded between zero and one in our application. The most common choice is to model this as a log-odds ratio of a linear function [73]. However, as Papke and Wooldridge [73] pointed out, the log-odds type procedure would be problematic if the data takes on the extreme values zero or one with positive probability, as can be the case in our application. Therefore, we follow Papke and Wooldridge [73] and consider a fractional logit model:

$$E(Y_{ij}|\mathbf{X}_j, \mathbf{Z}_{ij}) = G(\beta_0 + \mathbf{X}'_j\boldsymbol{\beta}_X + \mathbf{Z}'_{ij}\boldsymbol{\beta}_Z), \tag{3.1}$$

where Y_{ij} is the image accuracy of human interpreter i for image j with $0 \leq Y_{ij} \leq 1$, $\mathbf{X}_j$ contains relevant image characteristics, and $\mathbf{Z}_{ij}$ includes demographic characteristics of human interpreter i and some human interpreter i-image j specific attributes.[3] In this study, we choose $G(\cdot)$ to be the logistic function that $G(z) \equiv \Lambda(z) \equiv exp(z)/[1 + exp(z)]$.

Alternatively, we include participant fixed effects for human interpreters in place of their observed characteristics to control for the effects of human interpreters on accuracy as a whole:

$$E(Y_{ij}|\mathbf{X}_j, \mu_i) = G(\beta_0 + \mathbf{X}'_j\boldsymbol{\beta}_X + \mu_i), \tag{3.2}$$

where μ_i represents the individual fixed effects for human interpreter i. This term captures the effects of both observed and unobserved characteristics of human interpreters. In a third approach, we replace the image attributes variables with image fixed effects to account for any potentially unobserved effects associated with an image:

$$E(Y_{ij}|\mathbf{Z}_{ij}, \xi_j) = G(\beta_0 + \mathbf{Z}'_{ij}\boldsymbol{\beta}_Z + \xi_j), \tag{3.3}$$

where ξ_j is the fixed effect for image j.

[3] In our application, the variables time spent on interpreting the image and number of previously completed images, are interpreter i-image j specific.

To compare the performance of human interpreters to that of CFmask on the cloud interpretation work in general, we treat CFmask as an additional "person" and combine the observations of human interpreters and CFmask. We then add a binary indicator term for CFmask:

$$E(Y_{ij}|\mathbf{X}_j, F_{ij}) = G(\beta_0 + \mathbf{X}_j'\boldsymbol{\beta}_X + \beta_F F_{ij}), \tag{3.4}$$

where F_{ij} equals one for the work finished by CFmask, and zero otherwise. The sign of the estimated coefficient β_F tells us about the relative performance of CFmask compared to human interpreters, with a negative sign indicating overall higher accuracy of human interpreters. The above mentioned fractional logit models can be consistently estimated by a quasi-maximum likelihood estimator (QMLE).[45]

For the cell level analysis, we investigate how the actual presence of clouds and/or cloud shadows, which are inferred from the expert's interpretation, affect the cell accuracy of human interpreters and CFmask when controlling for other image attributes. This allows us to know whether human interpreters and CFmask are more likely to miss an actual cloud and/or cloud shadow (false negative) or to incorrectly identify the cell as being impacted when in fact it is not (false positive). In the cell level analysis, we define cell accuracy as the probability of a cell being interpreted correctly by a participant based upon our remote sensing expert's interpretation result. A simple logistic regression is implemented for the full sample of human interpreters and CFmask:

$$Pr(Y_{ijk} = 1) = G(\beta_0 + \mathbf{X}_j'\boldsymbol{\beta}_X + \beta_F F_{ij} + \beta_C C_{jk} + \beta_{FC} F_{ij} \times C_{jk}), \tag{3.5}$$

where Y_{ijk} is a binary variable which takes values one or zero, capturing whether the cell is interpreted correctly or not by human interpreter i for image j cell k, and C_{jk} is a binary variable with one indicating the actual presence of cloud and/or cloud shadow for image j cell k, based on the expert's interpretation. The function $G(\cdot)$ is, again, chosen to be the logistic function. Similarly, we consider to include individual fixed effects for human interpreters and CFmask (treat CFmask as a "person") to control for any individual related effects:

$$Pr(Y_{ijk} = 1) = G(\beta_0 + \mathbf{X}_j'\boldsymbol{\beta}_X + \beta_C C_{jk} + \beta_{FC} F_{ij} \times C_{jk} + \mu_i), \tag{3.6}$$

where the individual fixed effects term μ_i absorbs the effect of the CFmask indicator (and thus the CFmask indicator F_{ij} is omitted from this regression). Whether human interpreters and CFmask are more likely to make a false negative decision or a false positive decision can be inferred directly from the sign of the estimated coefficient β_C in the above two regressions.

[4] We can estimate $E(Y_{ij}|\cdot)$ by assuming a particular conditional distribution and estimating the parameters by maximum likelihood. However, the estimates are not robust to distributional failure. See Papke and Wooldridge [73] for detailed discussion.

[5] We specify standard errors clustered at the individual level to allow for correlation of errors within the individual cluster for the above image level models except for the participant fixed effects model. No standard error is generated for average marginal effects when we specify clustered standard error for the participant fixed effects model. This is because of its highly singular variance matrix resulting from the inclusion of individual fixed effects. We instead use robust standard error for this model.

In addition, the coefficient β_{FC} of the interaction term between the CFmask indicator and the cloud indicator provides information on whether false negative decisions versus false positive decisions differs between CFmask and human interpreters, and if so, by how much. We estimate the logistic regressions at the cell level analysis with a maximum likelihood estimator (MLE).[6][7]

3.3 Data

The online experiment of cloud interpretation task was conducted on the AMT platform on January 22nd, 2016.[8] The participants were required to take a short online training on the cloud interpretation task, and pass a qualification test (with a score of 80% or higher) before they were authorized to start working on the actual task.[9] A remote sensing expert pre-selected twelve satellite images in different regions with diverse backgrounds and cloud cover. Table 3.1 lists the path and row of the satellite images utilized in this study, and describes approximate location and background for each image. Each image is divided into eight-by-eight cells, and participants needed to specify for each cell if it is impacted by cloud and/or cloud shadow. Each task contains one image, and there were twelve images available. We aimed to collect 100 completions for each type of task, with 1,200 completions over all types in total. After some data cleaning, 925 completions from 108 participants are included in our final image level analysis, and the corresponding $925 \times 64 = 59,200$ cell level results are utilized in our cell level analysis.[10] Qualified participants had the option to complete all twelve tasks of different types if they wished in this round, and earn $0.50 per task. The images with different types were randomly assigned to the participants as they accepted the task. The image count (number of images already completed in this round), and the time spent on image interpretation, in minutes, were recorded for each participant. After completing a given task, the participants were directed to take a short exit survey collecting their demographic information, including gender, age, education, student status, country of residence and whether they had any academic or professional background related to this

[6] No standard error is generated for average marginal effects when we specify robust or clustered standard error for the above cell level models. This is because of its highly singular variance matrix resulting from the inclusion of binary indicators and/or individual fixed effects. We instead use the default standard error for these models.

[7] The estimations in this study were implemented in Stata software package.

[8] This is a follow-up round for our previous rounds of cloud interpretation task on the AMT, for example, the one introduced in Yu et al. [98].

[9] The qualification does not expire. The participants have access to our current and all future tasks once they pass the qualification. This provides an additional incentive to participants to take the training.

[10] 275 completions had to be dropped out of the 1,200 completions for this analysis due to missing or incomplete participant demographic information. The usable completions are still evenly distributed over the twelve types of tasks.

task.[11][12]

In addition to the interpretation results and demographic information of participants collected in the experiment, we have interpretation results of CFmask and the remote sensing expert for the same twelve satellite images. We assume that the expert's interpretation results are the true results, and they are used to infer the image accuracy and the cell accuracy for human interpreters and CFmask. The image characteristics, including percentage of image covered by clouds, sun azimuth and sun elevation angles from acquisition location at the time of acquisition, the distance from earth to the sun at the time of acquisition (in thousandth of astronomical unit (AU)), number of ground control points (GCPs), root mean squared error (RMSE) of the geometric residuals (meter) measured on the ground control points used in geometric precision correction, and the maximum radiance detected for the aerosol band (watt/steradian/m^2), were obtained from the Landsat dataset for each image.[13][14]

Table 3.1: Satellite images in this study

image #	path/row	location	description
1	p007r005	Greenland	homogenous with little variation
2	p016r037	South Carolina, US	managed forest and urban
3	p045r030	Oregon, US	snow
4	p092r086	Australia	mountains and forests
5	p135r025	Mongolia/Russia	mountains and trees
6	p141r048	India	agriculture and forest, some mountain
7	p143r049	India	agriculture, less mountain
8	p196r044	Africa	Saharan desert
9	p196r052	Africa	sahel, variable vegetation and mositure
10	p196r056	Africa	forest with patches
11	p199r026	France	urban, agriculture and forest
12	p223r061	Brazil	forest

[11] Participants who had submitted work to us in previous rounds were flagged and exempted from taking the exit survey.

[12] The variables country of residence and related background, are not included in our analysis, due to insufficient variation. Specifically, only one out of 108 participants was from a country outside of the U.S., and only one participant had an academic or professional background related to this task.

[13] One thousandth of AU is approximately equal to 1.496×10^8 meters.

[14] We also collected data on the minimum radiance detected for the aerosol band, and the maximum and minimum radiance detected for the cirrus band, but these are not included in our analysis due to collinearity with the maximum radiance detected for the aerosol band.

3.4 Results

3.4.1 Descriptive Statistics

Table 3.2 presents summary statistics on the characteristics of the twelve images. The images have relatively low cloud coverage from 0.50% to 25.70% with a mean of 9.81%. The sun azimuth angle is 93.77 degrees to the east on average with large variation (SD 88.40 degrees) for the images. Meanwhile, the sun elevation angle is positive for all images with small variation (SD 9.85 degrees), indicating daytime scene. The distance from earth to the sun varies from the nearest 986.33 to the farthest 1016.31 thousandth of AU. In geometric precision correction process, there are 215 ground control points being used, on average, with the minimum 7 and the maximum 373. The RMSEs of the geometric residuals measured on the ground control points and the maximum radiance detected for the aerosol band have relatively small variations over the twelve images.

Table 3.2: Summary statistics on image characteristics

	mean	(s.d.)	min	max
% cloud coverage	9.81	(8.45)	0.50	25.70
sun azimuth angle (degree)	93.77	(88.40)	-167.32	151.85
sun elevation angle (degree)	56.12	(9.85)	36.11	67.22
sun distance (thousandth of AU)	1004.76	(10.99)	986.33	1016.31
# GCPs in precision correction	215.00	(122.94)	7.00	373.00
RMSE (meter)	9.25	(3.40)	6.49	17.09
RMSE-Y (meter)	6.56	(2.63)	4.75	12.41
RMSE-X (meter)	6.50	(2.23)	4.36	11.75
max radiance for aerosol band	753.13	(16.61)	735.86	781.28

Summary statistics on the demographic information of participants are shown in Table 3.3. Close to two thirds of participants are male. Approximately half of participants are between 18 and 30 years old, and the other half are older than 30. Our participants are well-educated, with approximately 45% reporting to hold at least a college degree, and approximately one eighth indicating to be a current full-time or part-time student enrolled in a college, university, or vocational school. The cloud interpretation task was not new to approximately two fifths of participants, who had participated in our previous rounds of study.

It took approximately five minutes for a participant to complete one cloud interpretation task, on average, with slight fluctuations over the twelve different scenes, as shown in Table 3.4. The computed image accuracy, relative to the expert's interpretation results, is well above 80% on average for most images, except for image one, three, four and seven with accuracy ranging between 50% and 75%. The standard deviations of image accuracy across participants range from 8% to 32% for the twelve images.

Figures 3.1 to 3.3 show human interpretation and CFmask results relative to expert bench-

Table 3.3: Summary statistics on demographics

	percentage
male	65.74%
age 18-30 years old	49.07%
college or graduate degree	44.44%
current student	12.04%
previous participation	38.89%
number of participants	108

Table 3.4: Summary statistics on image completions

image #	time spent (min)		image accuracy	
	mean	(s.d.)	mean	(s.d.)
1	4.52	(7.12)	51.52%	(20.27%)
2	5.18	(7.80)	84.38%	(10.52%)
3	4.99	(7.57)	66.92%	(31.80%)
4	5.30	(7.69)	64.26%	(16.90%)
5	6.35	(9.32)	87.46%	(8.35%)
6	4.87	(6.80)	81.29%	(10.13%)
7	4.39	(7.02)	73.86%	(17.30%)
8	4.82	(6.97)	82.21%	(18.06%)
9	6.03	(9.38)	80.61%	(22.82%)
10	5.26	(7.59)	87.44%	(9.33%)
11	5.96	(8.31)	82.01%	(10.67%)
12	5.36	(7.64)	82.98%	(11.59%)

mark for all twelve images. For expert benchmark and CFmask results, a black shading suggests that the cell was interpreted as being impacted by cloud and/or cloud shadow. For human interpretation results, a darker shading indicates a higher percentage of human interpreters that identified the cell as being impacted. Naturally, a closer resemblance of the patterns for human interpretation and CFmask results with the expert benchmark indicates a higher degree of accuracy. Human interpreters performed especially well for images one, five, seven, and eight, compared to CFmask. CFmask had difficulty in differentiating impacted cells from cloud-free cells for these images. This is probably due to the fact that these images are either covered by very thin clouds or based on the background that is similar to cloud in the sense of pixel, such as desert. On the other hand, human interpreters can make some improvements for images three and nine that either has snow or sahel as a background. Overall, we do not observe obvious advantage of CFmask over human interpreters in accuracy for the twelve images. Whether human interpreters or CFmask, overall, achieved higher accuracy for the twelve images needs careful analysis, and is discussed in the following estimation results section.

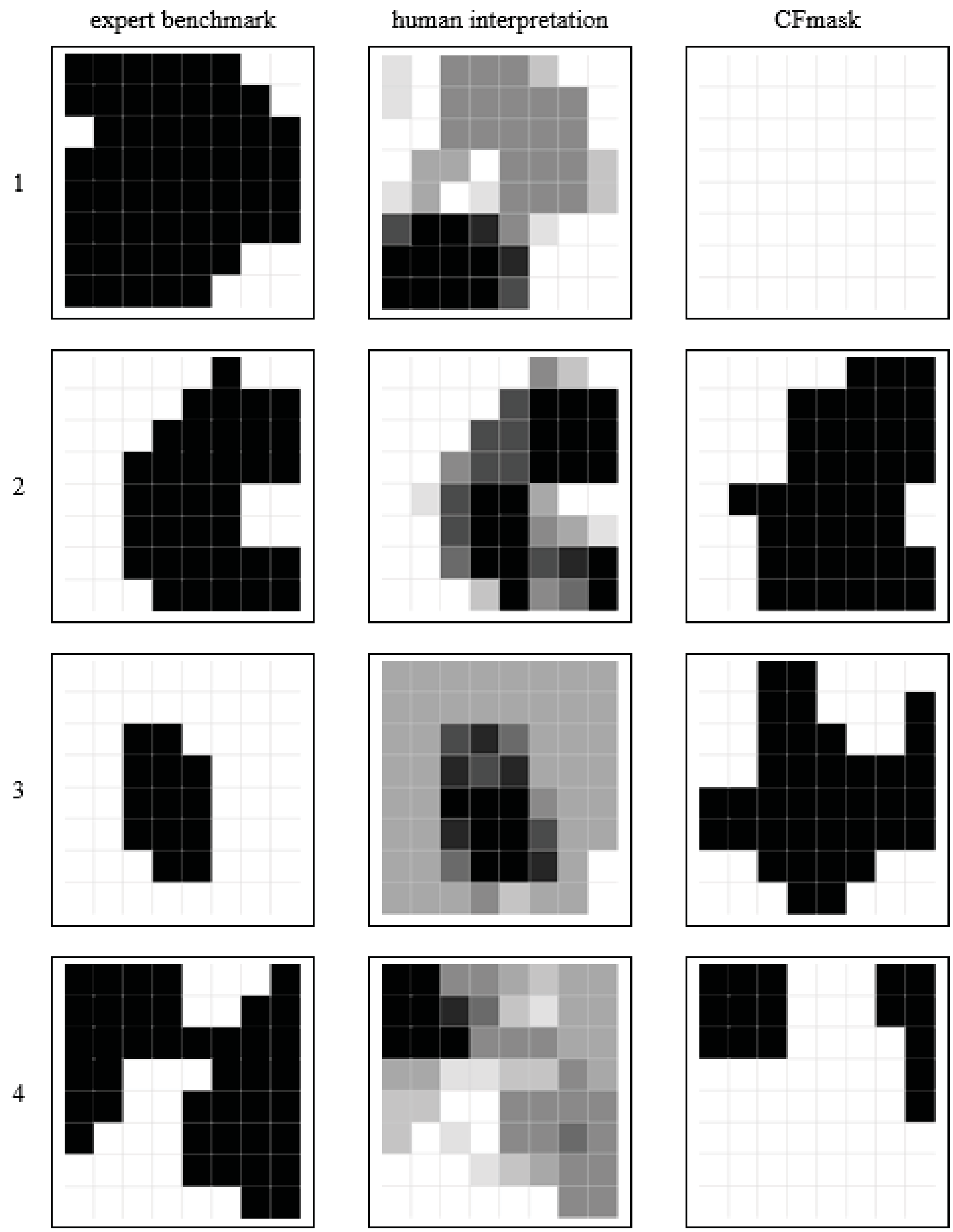

Figure 3.1: Visual comparison of human interpretation and CFmask results, image one to four

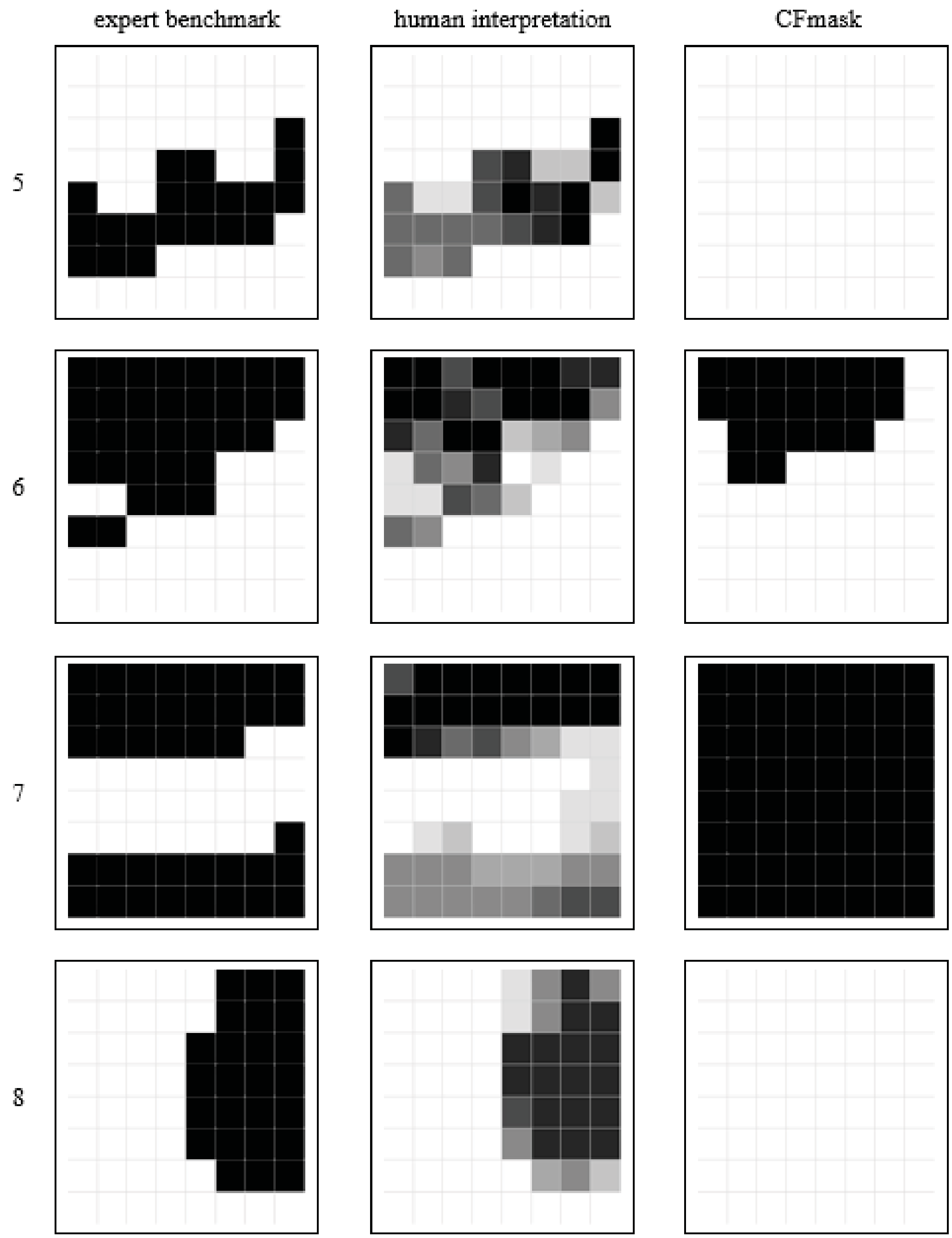

Figure 3.2: Visual comparison of human interpretation and CFmask results, image five to eight

3.4.2 Estimation Results

We first discuss estimation results for the image level analysis, followed by the cell level analysis.

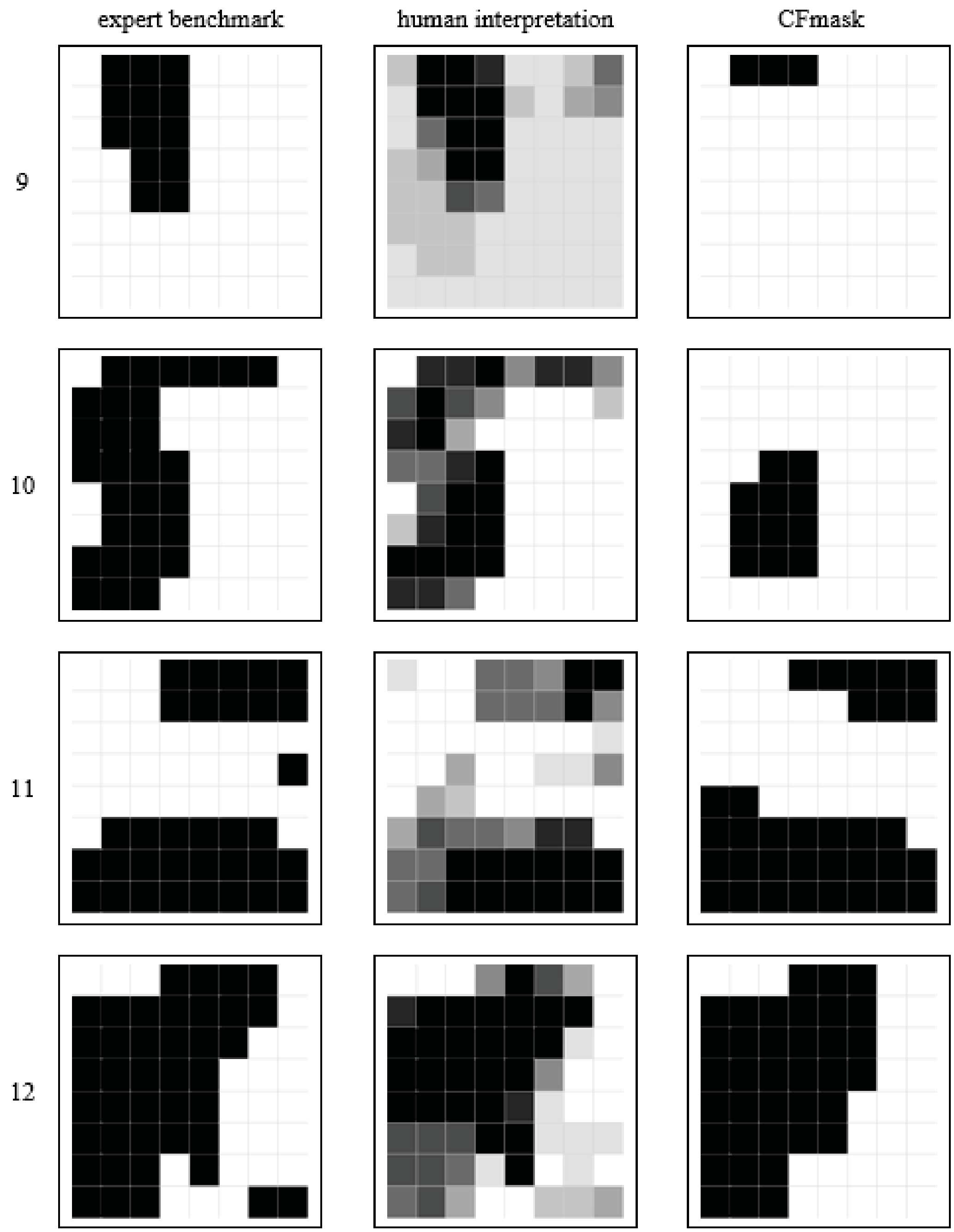

Figure 3.3: Visual comparison of human interpretation and CFmask results, image nine to twelve

Image level analysis

In the image level analysis, we examine how the characteristics of images and participants affect the cloud interpretation accuracy of human interpreters, and compare human interpreters to CFmask in terms of accuracy. We use the regression with the full set of image and participant characteristics, which is shown in equation (3.1), as the benchmark model to study the effects of image and participant. To control for the effects of unobserved characteristics of participant and image on accuracy, we estimate fractional logit models with fixed effects for participant and image, respectively, as shown in equations (3.2) and (3.3). We then compare the average marginal effects of image characteristics and participant attributes on accuracy from the benchmark model, the participant fixed effects model, and the image fixed effects model.

The average marginal effects of image and participant characteristics produced by the benchmark model are shown in Table 3.5. Table 3.6 shows the average marginal effects generated by the participant fixed effects model. As can be seen from these tables, average marginal effects of image characteristics flowing from the benchmark and the participant fixed effects models are similar, indicating that the observed participant characteristics are sufficient to control for any human effects. Specifically, the percentage of image covered by clouds affects image accuracy significantly, with one percentage increase in cloud coverage improving correctly interpreted cells by 0.31%, which is equivalent to approximately 0.20 cells. This may indicate that human interpreters were more likely to overlook or mis-interpret the image if clouds and/or cloud shadows are rare and/or thin and/or scattered. In addition, a marginal increase in sun azimuth angle, that is, one additional degree in angle to the east, has a similar positive effect of 0.15%. We do not detect significant effect of sun elevation angle. However, the combination of sun elevation and sun azimuth angles determines brightness of the image, which may affect human interpretation accuracy. Additionally, the distance from earth to the sun and the maximum radiance for the aerosol band are positively associated with image accuracy with effects of 19.61% and 12.70%, respectively. The surface around the clouds are brighter when earth is closer to the sun and/or the aerosol band is brighter. The results suggest that brighter clouds were easier to identify for human interpreters. In the mean time, the number of ground control points in precision correction has a negative effect of 0.03% on image accuracy. We found that the images with lower number of ground control points tend to have more homogenous background, like desert (image eight) or forest (image ten). Higher number of ground control points is associated with more complex background, for example, mountains and trees (image five). Clearly, it takes more effort for human interpreters to identify clouds and/or cloud shadows on an image with complex surface.

Marginal effects for the image fixed effects model are given in Table 3.7. As is evident from the table, the average marginal effects of participant characteristics estimated from the benchmark model and the image fixed effects model are also similar. This, in fact, suggests that controlling for the effects of unobserved image characteristics is unnecessary in our application. Specifically, a male participant was able to achieve 3.56% higher image accuracy than a female participant, which is equivalent to 2.28 more correctly interpreted cells on average. Being younger than 30 years, holding at least some college degree, and being a current part-time or full-time student had negative but insignificant effects on accuracy.

Table 3.5: Average marginal effects in image level benchmark model

	estimate	(s.e.)
% cloud coverage	0.0033**	(0.0015)
sun azimuth angle (degree)	0.0015**	(0.0006)
sun elevation angle (degree)	-0.0001	(0.0010)
sun distance (thousandth of AU)	0.1975*	(0.1014)
# GCPs in precision correction	-0.0003*	(0.0002)
RMSE (meter)	-0.9993	(0.6564)
RMSE-Y (meter)	0.6685	(0.4526)
RMSE-X (meter)	0.7639	(0.4850)
max radiance for aerosol band	0.1280*	(0.0662)
male	0.0355	(0.0244)
age 18-30 years old	-0.0266	(0.0215)
college or graduate degree	-0.0173	(0.0214)
current student	-0.0393	(0.0338)
previous participation	-0.0220	(0.0200)
time spent on the image (min)	0.0026*	(0.0015)
number of previously completed images	0.0052**	(0.0026)

Delta-method standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 3.6: Average marginal effects in image level participant fixed effects model

	estimate	(s.e.)
% cloud coverage	0.0031**	(0.0014)
sun azimuth angle (degree)	0.0015**	(0.0006)
sun elevation angle (degree)	0.0001	(0.0008)
sun distance (thousandth of AU)	0.1961**	(0.0997)
# GCPs in precision correction	-0.0003*	(0.0002)
RMSE (meter)	-0.9266	(0.6613)
RMSE-Y (meter)	0.6167	(0.4581)
RMSE-X (meter)	0.7131	(0.4865)
max radiance for aerosol band	0.1270*	(0.0651)

Delta-method standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

We also find that previous experience with our cloud interpretation task did not help for interpretation accuracy. In fact, the effect is negative, though it is not statistically significant different from no effect at a 10% significance level. This was probably due to the fact that our previous rounds were implemented more than six months before this round. It is possible that some returning participants who passed the qualification module in previous rounds (and thus did not have to take the qualification again in this round) started their interpretation work without reviewing the training materials carefully.[15] In this case, any benefit from prior

[15]Notice that the returning participants still had the option to reference back to the training module

experience with the cloud interpretation task may have likely faded through time, leading to an overall lack of recent training for this round. At the same time, one additional minute spent on interpreting the image increases image accuracy by 0.25%. Every additional image completed in this round boosts accuracy of the current image by 0.55%, indicating a positive learning effect.

Table 3.7: Average marginal effects in image level image fixed effects model

	estimate	(s.e.)
male	0.0356*	(0.0201)
age 18-30 years old	-0.0267	(0.0189)
college or graduate degree	-0.0169	(0.0203)
current student	-0.0384	(0.0300)
previous participation	-0.0210	(0.0181)
time spent on the image (min)	0.0025**	(0.0011)
number of previously completed images	0.0055***	(0.0016)

Delta-method standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

The results of the image level participant-CFmask comparison model, given in equation (3.4), are shown in Table 3.8. They indicate that human interpreters were, on average, 6.12% more accurate than CFmask in interpreting cloud and/or cloud shadow for the twelve images. That is, a typical human interpreter on the AMT platform can get approximately 3.92 more cells correct than CFmask for a given image, when the characteristics of the image are accounted for. This finding is consistent with what we observed in our earlier study [98], and provides further evidence on the potential effectiveness of human interpreters from the AMT platform for cloud interpretation tasks.

Cell level analysis

The results of the cell level benchmark model (given in equation (3.5)), are shown in Table 3.9. They, too, suggest that human interpreters performed better than CFmask in terms of cell accuracy. Overall, the probability of CFmask identifying a given cell correctly is lower than human interpreters by 5.08%. The results of cell level individual fixed effects model, displayed in Table 3.10, indicate that the presence of an actual cloud and/or cloud shadow in a cell decreases the probability of a participant or CFmask interpreting the cell correctly by approximately 22.45%. It suggests that both human interpreters and CFmask are more likely to miss an actual cloud and/or cloud shadow (false negative) than to incorrectly identify the cell as being impacted when in fact it is not (false positive). The insignificant coefficient for the interaction term between the cloud indicator and the CFmask indicator reveals that

during their interpretation work just like the new participants. Unfortunately, however, we do not have data on whether, and how many times the participants referenced back to the training materials.

Table 3.8: Average marginal effects in image level participant-CFmask comparison model

	estimate	(s.e.)
% cloud coverage	0.0034***	(0.0012)
sun azimuth angle (degree)	0.0015***	(0.0004)
sun elevation angle (degree)	-0.0002	(0.0010)
sun distance (thousandth of AU)	0.2022**	(0.0839)
# GCPs in precision correction	-0.0003***	(0.0001)
RMSE (meter)	-0.9559*	(0.5681)
RMSE-Y (meter)	0.6362	(0.3950)
RMSE-X (meter)	0.7329*	(0.4149)
max radiance for aerosol band	0.1310**	(0.0551)
Cfmask	-0.0612***	(0.0161)

Delta-method standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

there is no significant difference in false negative decisions between human interpreters and CFmask.

The average marginal effects of image characteristics, as shown in Table 3.9 and Table 3.10, are similar. Specifically, one percentage increase in cloud coverage for a given cell boosts the probability of getting the cell correctly by approximately 0.35%. Marginal increases in sun azimuth angle and sun elevation angle, have similar positive effects of 0.09% and 0.15%, respectively. Additionally, the distance from earth to the sun and the maximum radiance for the aerosol band are positively associated with cell accuracy with effects of approximately 30% and 20%, respectively. Meanwhile, the number of ground control points in precision correction has a negative effect of 0.07% on cell accuracy.

Table 3.9: Average marginal effects in cell level benchmark model

	estimate	(s.e.)
% cloud coverage	0.0036***	(0.0004)
sun azimuth angle (degree)	0.0009***	(0.0001)
sun elevation angle (degree)	0.0014***	(0.0003)
sun distance (thousandth of AU)	0.3173***	(0.0270)
# GCPs in precision correction	-0.0007***	(0.0000)
RMSE (meter)	-0.4549**	(0.2323)
RMSE-Y (meter)	0.2876*	(0.1647)
RMSE-X (meter)	0.3430**	(0.1656)
max radiance for aerosol band	0.2057***	(0.0179)
CFmask	-0.0508**	(0.0224)
actual cloud	-0.2262***	(0.0035)
actual cloud*CFmask	-0.0330	(0.0279)

Delta-method standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 3.10: Average marginal effects in cell level individual fixed effects model

	estimate	(s.e.)
% cloud coverage	0.0034***	(0.0004)
sun azimuth angle (degree)	0.0009***	(0.0001)
sun elevation angle (degree)	0.0016***	(0.0003)
sun distance (thousandth of AU)	0.3039***	(0.0266)
# GCPs in precision correction	-0.0007***	(0.0000)
RMSE (meter)	-0.4028*	(0.2273)
RMSE-Y (meter)	0.2517	(0.1611)
RMSE-X (meter)	0.3063*	(0.1621)
max radiance for aerosol band	0.1968***	(0.0176)
actual cloud	-0.2245***	(0.0034)
actual cloud*Cfmask	-0.0203	(0.0262)

Delta-method standard error in parentheses. *** significant at 1%, ** significant at 5%, * significant at 10%.

3.5 Conclusion

Crowdsourcing cloud interpretation work to human interpreters recruited on the AMT platform is one of the methods available to deal with the challenge of processing large amounts of satellite images that are potentially covered by cloud and/or cloud shadow. This study investigates how the characteristics of human interpreters and satellite scenes affect cloud interpretation accuracy, and how human interpreters compare to CFmask, a popular automated cloud-detection algorithm, in terms of accuracy and types of incorrect interpretation (false negatives versus false positives).

Specifically, we analyze several image level fractional logit models, some with participant fixed effects and image fixed effects, to study the effects of image characteristics and participant attributes. The results show that an image with higher cloud coverage, or larger sun azimuth angle, or farther distance to the sun received higher accuracy; a male participant achieved higher accuracy than a female participant on average; finally, the more time spent on interpreting the image, and the more images completed before the current image, was beneficial for improving accuracy, controlling for image and participant attributes.

The comparison between human interpreters and CFmask show that a typical human interpreter, on average, was able to achieve higher accuracy than CFmask when controlling for the effects of image attributes. In addition, the cell level analysis results suggest that human interpreters and CFmask are both more likely to miss an actual cloud and/or cloud shadow in a given cell (false negative) than to incorrectly identify a cell as being impacted while in fact it is not (false positive). The difference between human interpreters and CFmask on their false negative decisions is not statistically significant.

Bibliography

[1] Joshua K Abbott and H Allen Klaiber. An embarrassment of riches: Confronting omitted variable bias and multi-scale capitalization in hedonic price models. *Review of Economics and Statistics*, 93(4):1331–1342, 2011.

[2] Joshua K Abbott and H Allen Klaiber. The value of water as an urban club good: A matching approach to community-provided lakes. *Journal of Environmental Economics and Management*, 65(2):208–224, 2013.

[3] Sandra G Affenito, Douglas R Thompson, Bruce A Barton, Debra L Franko, Stephen R Daniels, Eva Obarzanek, George B Schreiber, and Ruth H Striegel-Moore. Breakfast consumption by African-American and white adolescent girls correlates positively with calcium and fiber intake and negatively with body mass index. *Journal of the American Dietetic Association*, 105(6):938–945, 2005.

[4] L Anselin. *Spatial Econometrics: Methods and Models*, volume 4. Springer Science & Business Media, 1988.

[5] Albert Arking and Jeffrey D Childs. Retrieval of cloud cover parameters from multispectral satellite images. *Journal of Climate and Applied Meteorology*, 24(4):322–333, 1985.

[6] Robert J Armstrong and Daniel A Rodriguez. An evaluation of the accessibility benefits of commuter rail in eastern massachusetts using spatial hedonic price functions. *Transportation*, 33(1):21–43, 2006.

[7] M Christopher Auld and Lisa M Powell. Economics of food energy density and adolescent body weight. *Economica*, 76(304):719–740, 2009.

[8] Patrick Bajari, Jane Cooley Fruehwirth, Kyoo Il Kim, and Christopher Timmins. A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution. *The American Economic Review*, 102(5):1898–1926, 2012.

[9] Luke Barrington, Shubharoop Ghosh, Marjorie Greene, Shay Har-Noy, Jay Berger, Stuart Gill, Albert Yu-Min Lin, and Charles Huyck. Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics*, 54(6), 2012.

[10] Bruce A Barton, Alison L Eldridge, Douglas Thompson, Sandra G Affenito, Ruth H Striegel-Moore, Debra L Franko, Ann M Albertson, and Susan J Crockett. The relationship of breakfast and cereal consumption to nutrient intake and body mass index: The national heart, lung, and blood institute growth and health study. *Journal of the American Dietetic Association*, 105(9):1383–1389, 2005.

[11] Julie Beaulac, Elizabeth Kristjansson, and Steven Cummins. Peer reviewed: A systematic review of food deserts, 1966-2007. *Preventing chronic disease*, 6(3), 2009.

[12] May A Beydoun, Lisa M Powell, Xiaoli Chen, and Youfa Wang. Food prices are associated with dietary quality, fast food consumption, and body mass index among us children and adolescents. *The Journal of nutrition*, 141(2):304–311, 2011.

[13] Jayanta Bhattacharya, Janet Currie, and Steven J Haider. Breakfast of champions? The school breakfast program and the nutrition of children and families. *Journal of Human Resources*, 41(3):445–466, 2006.

[14] Sandra E Black. Do better schools matter? parental valuation of elementary education. *The Quarterly Journal of Economics*, 114(2):577–599, 1999.

[15] Peter C Boxall, Wing H Chan, and Melville L McMillan. The impact of oil and natural gas facilities on rural residential property values: a spatial hedonic analysis. *Resource and energy economics*, 27(3):248–269, 2005.

[16] C.A. Cameron and P.K. Trivedi. *Microeconometrics: Methods and applications*. Cambridge University Press, 2005.

[17] Armand V Cardello and F Miles Sawyer. Effects of disconfirmed consumer expectations on food acceptability. *Journal of sensory studies*, 7(4):253–277, 1992.

[18] Centers for Disease Control and Prevention. General food environment resources, 2014. web site: http://www.cdc.gov/healthyplaces/healthtopics/healthyfood/general.htm, accessed April 16, 2016.

[19] Clarice N Chau, Jamie M Zoellner, and Jennie L Hill. Availability of healthy food: does block group race and income matter? *Journal of hunger & environmental nutrition*, 8 (1):22–38, 2013.

[20] Kenneth Y Chay and Michael Greenstone. Does air quality matter? evidence from the housing market. *Journal of political Economy*, 113(2):376–424, 2005.

[21] Susan Chen, RJ Florax, Samantha D Snyder, et al. Obesity in urban food markets: Evidence from georeferenced micro data. *West Lafayette, IN: Purdue University*, 2009.

[22] Shin-Yi Chou, Michael Grossman, and Henry Saffer. An economic analysis of adult obesity: results from the behavioral risk factor surveillance system. *Journal of health economics*, 23(3):565–587, 2004.

[23] Emilio Chuvieco. *Fundamentals of satellite remote sensing*. CRC press, 2009.

[24] Jed Cohen, Christine E Blinn, Kevin J Boyle, Thomas P Holmes, and Klaus Moeltner. Hedonic valuation with translating amenities: mountain pine beetles and host trees in the colorado front range. *Environmental and resource economics*, 63(3):613–642, 2016.

[25] Jeffrey P Cohen and Cletus C Coughlin. Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science*, 48(5):859–878, 2008.

[26] Rachel Cooper, Madeleine Levin, and Jennifer Adach. School breakfast scorecard: School year 2007-2008. Technical report, Food Research and Action Center, 2009.

[27] Sean P Corcoran, Brian Elbel, and Amy Ellen Schwartz. The effect of breakfast in the classroom on obesity and academic performance: Evidence from New York City. In *Health & Healthcare in America: From Economics to Policy*. Ashecon, 2014.

[28] Chad Cotti and Nathan Tefft. Fast food prices, obesity, and the minimum wage. *Economics & Human Biology*, 11(2):134–147, 2013.

[29] David A Cranage, Martha T Conklin, and Carolyn U Lambert. Effect of nutrition information in perceptions of food quality, consumption behavior and purchase intentions. *Journal of Foodservice Business Research*, 7(1):43–61, 2005.

[30] Steven Cummins, Anne Findlay, Cassie Higgins, Mark Petticrew, Leigh Sparks, and Hilary Thomson. Reducing inequalities in health and diet: findings from a study on the impact of a food retail development. *Environment and Planning A*, 40(2):402–422, 2008.

[31] Steven Cummins, Ellen Flint, and Stephen A Matthews. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. *Health affairs*, 33(2):283–291, 2014.

[32] Janet Currie, Stefano DellaVigna, Enrico Moretti, and Vikram Pathania. The effect of fast food restaurants on obesity and weight gain. *American Economic Journal: Economic Policy*, 2(3):32–63, 2010.

[33] Barbara Devaney and Thomas Fraker. The dietary impacts of the school breakfast program. *American Journal of Agricultural Economics*, 71(4):932–948, 1989.

[34] J. Dickl. Breakfast in the classroom implementation: Dive into operations & financials, 2014. Presentation given at the national conference of the School Nutrition Association, Boston, MA, July 13-16, 2014.

[35] Tamara Dubowitz, Madhumita Ghosh-Dastidar, Deborah A Cohen, Robin Beckman, Elizabeth D Steiner, Gerald P Hunter, Karen R Flórez, Christina Huang, Christine A Vaughan, Jennifer C Sloan, et al. Diet and perceptions change with supermarket introduction in a food desert, but not because of supermarket use. *Health Affairs*, 34 (11):1858–1868, 2015.

[36] Richard A Dunn. The effect of fast-food availability on obesity: An analysis by gender, race, and residential location. *American Journal of Agricultural Economics*, 92(4): 1149–1164, 2010.

[37] Richard A Dunn, Joseph R Sharkey, and Scott Horel. The effect of fast-food availability on fast-food consumption and obesity among rural residents: an analysis by race/ethnicity. *Economics & Human Biology*, 10(1):1–13, 2012.

[38] TF Eck, PK Bhartia, PH Hwang, and LL Stowe. Reflectivity of earth's surface and clouds in ultraviolet from satellite observations. *Journal of Geophysical Research: Atmospheres*, 92(D4):4287–4296, 1987.

[39] Brian Elbel, Alyssa Moran, L Beth Dixon, Kamila Kiszko, Jonathan Cantor, Courtney Abrams, and Tod Mijanovich. Assessment of a government-subsidized supermarket in a high-need area on household food availability and childrens dietary intakes. *Public health nutrition*, 18(15):2881–2890, 2015.

[40] Farm Bill. Food, conservation, and energy act of 2008. 2008.

[41] Adrian Fisher. Cloud and cloud-shadow detection in spot5 hrg imagery with automated morphological feature extraction. *Remote Sensing*, 6(1):776–800, 2014.

[42] Rudolph Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3): 221, 1948.

[43] Food and Nutrition Service. *Provision 2 Guidance. National School Lunch and School Breakfast Programs*. USDA, 2002.

[44] Food Research and Action Center. Breakfast in the classroom fact sheet, 2016. web site: http://frac.org/federal-foodnutrition-programs/school-breakfast-program/breakfast-in-the-classroom/, accessed May. 16, 2016.

[45] David E Frisvold. Nutrition and cognitive achievement: An evaluation of the school breakfast program. *Journal of Public Economics*, 124:91–104, 2015.

[46] Steffen Fritz, Ian McCallum, Christian Schill, Christoph Perger, Roland Grillmayer, Frédéric Achard, Florian Kraxner, and Michael Obersteiner. Geo-wiki. org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, 1(3):345–354, 2009.

[47] Shanti Gamper-Rabindran and Christopher Timmins. Does cleanup of hazardous waste sites raise housing values? evidence of spatially localized benefits. *Journal of Environmental Economics and Management*, 65(3):345–360, 2013.

[48] Kishore Gawande and Hank Jenkins-Smith. Nuclear waste transport and residential property values: estimating the effects of perceived risks. *Journal of Environmental Economics and Management*, 42(2):207–233, 2001.

[49] Shubharoop Ghosh, Charles K Huyck, Marjorie Greene, Stuart P Gill, John Bevington, Walter Svekla, Reginald DesRoches, and Ronald T Eguchi. Crowdsourcing for rapid damage assessment: The global earth observation catastrophe assessment network (geo-can). *Earthquake Spectra*, 27(S1):S179–S198, 2011.

[50] Karen Glanz, James F Sallis, Brian E Saelens, and Lawrence D Frank. Nutrition environment measures survey in stores (nems-s): development and evaluation. *American journal of preventive medicine*, 32(4):282–289, 2007.

[51] Philip Gleason and Carol Suitor. Children's diets in the mid-1990s: Dietary intake and its relationship with school meal participation. Technical report, Mathematica Policy Research, 2001.

[52] Michael F Goodchild and J Alan Glennon. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3): 231–241, 2010.

[53] Michael Greenstone and Justin Gallagher. Does hazardous waste matter? evidence from the housing market and the superfund program. *The Quarterly Journal of Economics*, 123(3):951–1003, 2008.

[54] Jessie Hewins. School breakfast scorecard: 2014-2015 school year. Technical report, Food Research and Action Center, 2016.

[55] Jennie L Hill, Clarice Chau, Candice R Luebbering, Korine K Kolivras, and Jamie Zoellner. Does availability of physical activity and food outlets differ by race and income? findings from an enumeration study in a health disparate region. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1):105, 2012.

[56] Jennie L Hill, Wen You, and Jamie M Zoellner. Disparities in obesity among rural and urban residents in a health disparate region. *BMC public health*, 14(1):1051, 2014.

[57] M Joseph Hughes and Daniel J Hayes. Automated detection of cloud and cloud shadow in single-date landsat imagery using neural networks and spatial post-processing. *Remote Sensing*, 6(6):4907–4926, 2014.

[58] Scott A Imberman and Adriana D Kugler. The effect of providing breakfast in class on student performance. *Journal of Policy Analysis and Management*, 33(3):669–699, 2014.

[59] Elena G Irwin. The effects of open space on residential property values. *Land economics*, 78(4):465–480, 2002.

[60] Chong Won Kim, Tim T Phipps, and Luc Anselin. Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of environmental economics and management*, 45(1):24–39, 2003.

[61] Jungik Kim and Peter Goldsmith. A spatial hedonic approach to assess the impact of swine production on residential property values. *Environmental and Resource Economics*, 42(4):509–534, 2009.

[62] Silvia C King, Annette J Weber, Herbert L Meiselman, and Nan Lv. The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food quality and preference*, 15(7):645–653, 2004.

[63] Christopher G Leggett and Nancy E Bockstael. Evidence of the effects of water quality on residential land prices. *Journal of Environmental Economics and Management*, 39 (2):121–144, 2000.

[64] Tammy Leonard, Caitlin McKillop, Jo Ann Carson, and Kerem Shuval. Neighborhood effects on food consumption. *Journal of Behavioral and Experimental Economics*, 51: 99–113, 2014.

[65] Jacob Leos-Urbel, Amy Ellen Schwartz, Meryle Weinstein, and Sean Corcoran. Not just for poor kids: The impact of universal free school breakfast on meal participation and student outcomes. *Economics of Education Review*, 36:88–107, 2013.

[66] James LeSage and Robert Kelley Pace. *Introduction to Spatial Econometrics*. CRC Press, 2009.

[67] Jiyuan Liu, Mingliang Liu, Hanqin Tian, Dafang Zhuang, Zengxiang Zhang, Wen Zhang, Xianming Tang, and Xiangzheng Deng. Spatial and temporal patterns of china's cropland during 1990–2000: an analysis based on landsat tm data. *Remote sensing of Environment*, 98(4):442–456, 2005.

[68] R. McNulty. Breakfast after the bell: Lemon Grove school district case study. Technical report, Lemon Grove School District, 2016.

[69] A.F. Meyers, A.E. Sampson, M. Weitzman, B.L. Rogers, and H. Kayne. School breakfast program and school performance. *American Journal of Diseases of Children*, 143: 1234–1239, 2010.

[70] Yair Mundlak. On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society*, pages 69–85, 1978.

[71] Theo Nijman and Marno Verbeek. Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics*, 7(3):243–257, 1992.

[72] Office of Community Services. Healthy food financing initiative, undated. web site: http://www.acf.hhs.gov/programs/ocs/programs/community-economic-development/healthy-food-financing, accessed May 05, 2017.

[73] Leslie E Papke and Jeffrey M Wooldridge. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632, 1996.

[74] Partners for breakfast in the classroom. Where has breakfast in the classroom been implemented?, 2016. web site: http://breakfastintheclassroom.org/about-partners-for-breakfast-in-the-classroom/where-has-breakfast-in-the-classroom-been-implemented/, accessed May 16, 2016.

[75] Janet Poppendieck. *Free for all: Fixing school food in America*. Univ of California Press, 2009.

[76] Lisa M Powell. Fast food costs and adolescent body mass index: evidence from panel data. *Journal of health economics*, 28(5):963–970, 2009.

[77] Lisa M Powell and Yanjun Bao. Food prices, access to food outlets and child weight. *Economics & Human Biology*, 7(1):64–72, 2009.

[78] Lisa M Powell, Zhenxiang Zhao, and Youfa Wang. Food prices and fruit and vegetable consumption among young american adults. *Health & place*, 15(4):1064–1070, 2009.

[79] Shi Qiu, Binbin He, Zhe Zhu, Zhanmang Liao, and Xingwen Quan. Improving fmask cloud and cloud shadow detection in mountainous area for landsats 4–8 images. *Remote Sensing of Environment*, 199:107–119, 2017.

[80] Alana Rhone, Michele Ver Ploeg, Chris Dicken, Ryan Williams, and Vince Breneman. *Low-income and Low-supermarket-access Census Tracts, 2010-2015*. United States Department of Agriculture, Economic Research Service, 2017.

[81] Sherwin Rosen. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55, 1974.

[82] Brian E Saelens, Karen Glanz, James F Sallis, and Lawrence D Frank. Nutrition environment measures study in restaurants (nems-r): development and evaluation. *American journal of preventive medicine*, 32(4):273–281, 2007.

[83] Diane Whitmore Schanzenbach and Mary Zaki. Expanding the school breakfast program: Impacts on children's consumption, nutrition and health. Technical report, National Bureau of Economic Research, 2014.

[84] Linda See, Dmitry Schepaschenko, Myroslava Lesiv, Ian McCallum, Steffen Fritz, Alexis Comber, Christoph Perger, Christian Schill, Yuanyuan Zhao, Victor Maus, et al. Building a hybrid land cover map with crowdsourcing and geographically weighted regression. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:48–56, 2015.

[85] V Kerry Smith and Ju-Chin Huang. Can markets value air quality? a meta-analysis of hedonic property value models. *Journal of political economy*, 103(1):209–227, 1995.

[86] Karen E Spears and Heidi L Himler. Advantages and disadvantages of an automated weighed intake dietary assessment tool requiring minimal participant burden for use at congregate meal sites. *Journal of the Academy of Nutrition and Dietetics*, 113(1): 31–34, 2013.

[87] R. Taylor. Breakfast after the bell: Riverside school district case study. Technical report, Riverside School District, 2016.

[88] Rebecca Taylor and Sofia B Villas-Boas. Food store choices of poor households: a discrete choice analysis of the national household food acquisition and purchase survey (foodaps). *American Journal of Agricultural Economics*, 98(2):513–532, 2016.

[89] Marcel AJ Theebe. Planes, trains, and automobiles: the impact of traffic noise on house prices. *The Journal of Real Estate Finance and Economics*, 28(2-3):209–234, 2004.

[90] Maureen T Timlin, Mark A Pereira, Mary Story, and Dianne Neumark-Sztainer. Breakfast eating and weight change in a 5-year prospective analysis of adolescents: Project EAT (Eating Among Teens). *Pediatrics*, 121(3):638–645, 2008.

[91] David Tracey and Jonah E Rockoff. Estimates of the impact of crime risk on property values from megan's laws. *The American Economic Review*, 98(3):1103–1127, 2008.

[92] Liisa Tyrväinen and Antti Miettinen. Property prices and urban forest amenities. *Journal of environmental economics and management*, 39(2):205–223, 2000.

[93] University of Wisconsin Population Health Institute. County health rankings, 2016. web site: `http://http://www.countyhealthrankings.org/`, accessed May 06, 2016.

[94] U.S. Department of Agriculture, Agricultural Research Service. USDA National Nutrient Database for Standard Reference, Release 23. Nutrient Data Laboratory Home Page, `http://www.ars.usda.gov/ba/bhnrc/ndl`, 2010.

[95] U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans*. Washington, DC: U.S. Government Printing Office, 2010.

[96] Margaret Walls, Carolyn Kousky, and Ziyan Chu. Is what you see what you get? the value of natural landscape views. *Land Economics*, 91(1):1–19, 2015.

[97] Jeffrey M Wooldridge. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1):115–132, 1995.

[98] Ling Yu, Sheryl B Ball, Christine E Blinn, Klaus Moeltner, Seth Peery, Valerie A Thomas, and Randolph H Wynne. Cloud-sourcing: Using an online labor force to detect clouds and cloud shadows in landsat images. *Remote Sensing*, 7(3):2334–2351, 2015.

[99] Fei Yuan, Kali E Sawaya, Brian C Loeffelholz, and Marvin E Bauer. Land cover classification and change analysis of the twin cities (minnesota) metropolitan area by multitemporal landsat remote sensing. *Remote sensing of Environment*, 98(2):317–328, 2005.

[100] Jeffrey E Zabel. Estimating fixed and random effects models with selectivity. *Economics Letters*, 40(3):269–272, 1992.

[101] Jeffrey E Zabel and Katherine A Kiel. Estimating the demand for air quality in four us cities. *Land Economics*, pages 174–194, 2000.

[102] Zhe Zhu and Curtis E Woodcock. Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sensing of Environment*, 118:83–94, 2012.