"Billions and Billions Served" Heterogeneous Effects of Food Source on Child Dietary Quality

Travis A. Smith*
Department of Applied Economics
University of Minnesota

May 29, 2013

Abstract

This paper estimates heterogeneous effects of food source (food away from home, at home and from school) on child dietary quality. Using a quantile estimator designed for panel data, two non-consecutive days of intake are used to identify the effect of food source across the unconditional distribution of dietary quality. Main results suggest that food away from home has a negative impact on dietary quality for all children except those falling in the very lowest portion of the unconditional distribution. As compared to home-prepared food, school food is found to increase dietary quality for children falling in the bottom quartile of the distribution. For children with a very high underlying proneness to consume a healthful diet, food from school has a negative effect. While food consumed under the National School Lunch and Breakfast Programs may not benefit every child (especially at the mean), it does improve the diets of many children whom otherwise would have poorer dietary quality. The implication is that U.S. schools are fertile grounds to improve nutrition skill formation, especially for the most disadvantaged.

Key Words: Unconditional quantiles, panel data, dietary quality, school food programs.

JEL Classification: C31, D39, I12, I18

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4–6, 2013.

*tasmith@umn.edu

"Billions and Billions Served"
Heterogeneous Effects of Food Source on Child Dietary Quality

1. Introduction

Early decisions in human capital accumulation and skill formation have direct consequences on the productivity of future investments (Cunha et al., 2006). Skills related to health capital, for example, quickly accumulate early on in life and have persistent impacts throughout adolescence and adulthood (McFadden, 2008). Therefore, it is of no surprise that the case for investing early in children, specifically the disadvantaged, is strong (Heckman and Masterov, 2007), and policymakers are particularly interested in programs that target such children. With nutrition in mind, two longstanding Federal programs have gained increasing attention in the United States: the School Breakfast Program (SBP) and the National School Lunch Program (NSLP).[1]

Offered in over 100,000 public and non-profit institutions, the SBP and NSLP serve millions of students every school day.[2] Together, these two Federally subsidized meal programs represent a substantial and repeated exposure to nutrition skill formation, which has strong implications for nutrition capital accumulation. For example, numerous experimental trials have demonstrated that infants and young children have the capability to learn and apply nutrition skills, but the ability to adopt new skills decreases as one matures into adulthood.[3] Outside of school and home, exposure to food-away-from-home (FAFH), such as fast-food and restaurant establishments, has become much more prominent in the daily diet of American children (Poti and Popkin, 2011). While the literature generally agrees that FAFH negatively impacts health, researchers are at odds

[1] Since its inception in 1946, the NSLP has served over 224 billion lunches in the U.S. (FNS-USDA, 2012a). Interestingly, it is estimated that McDonald's has sold over 247 billion hamburgers since its re-opening under its namesake in 1948.

[2] The NSLP is offered in 99% of all public schools and 94% of public and private schools combined (Ralston et al., 2008). Nearly 32 million lunches were served daily in 2011, with roughly two-thirds at a free or reduced-price (FNS-USDA, 2012a). The SBP served 12.1 million students in 2011, with 10.1 million receiving a free or reduced-priced breakfast (FNS-USDA, 2012b).

[3] Benton (2004) and Birch (1999) provide thorough reviews of such studies.

with respect to the impact of school food. The findings of this paper suggest that the conflicting results may be due to a focus on the average effect of food source, which considers both low and high dietary quality children to be homogeneous.

This study adds to the current literature by considering heterogeneous effects of food source across all levels of underlying dietary quality, rather than focusing on average diet quality. I focus on dietary quality because it correlates with body weight (Jennings et al., 2011) and academic achievement (Florence, Asbridge and Veugelers, 2008) in children, is a predictor for many chronic diseases in adulthood (Chiuve et al., 2013) and is at the forefront of Federal and State policies aimed at reforming nutritional standards in schools.

Three food sources are considered: food from home (FFH), from away from home (FAFH) and food from school (FFS). I define underlying dietary quality as a child's "proneness" to consume a healthful diet.[4] For example, a child that is prone to a very low quality diet, possibly due to parental or environmental factors, may exhibit large benefits from a school lunch and/or breakfast. On the other hand, children prone to high quality diets that consume a meal from school may experience a decrease in overall dietary quality. Therefore, examining the the average treatment effect (ATE) of participating in school food programs may mask important heterogeneous effects. This study expands on existing literature by estimating the quantile treatment effect (QTE) of food source on the *unconditional* distribution of child dietary quality.

Several studies have investigated the mean effects of food source on various aspects of child health. Many have found that FAFH increases calorie intake (Bowman et al., 2004; Powell and Nguyen, 2013) and reduces diet quality among children (Mancino et al., 2010). When examining the effect of participating in the SBP and the NSLP separately, several authors are in agreement that the former is beneficial (Bhattacharya, Currie and Haider, 2006; Millmet and Tchernis, 2012; Millimet, Tchernis and Husain, 2010; Schanzenbach, 2009) but the latter is not (Schanzenbach, 2009; Millimet et al., 2010). For example, Schanzenbach (2009) found that school lunches increased

[4]The term "proneness" was introduced by Doksum (1974). In the present study, proneness can be thought of as the (fixed) degree to which one consumes a healthful diet. Perhaps a more relatable example from the economics literature is ability, which is defined by how prone one is to more favorable labor market outcomes.

average daily calorie intake by about 40 calories and that children who consume school lunches have higher rates of obesity by 1 to 2%. Overall, while the consensus appears to be that child health is negatively impacted by FAFH, researchers are at odds when considering the average impact of a school breakfast or lunch. A potential driver behind this inconsistency is the fact that the ATE necessarily implies the relationship to be homogeneous for all children.

Several approaches to identifying the effect of food source on dietary outcomes have been used. A fixed effect, or first-differencing approach, is easily implemented by using two days of dietary intake typically found in U.S. nationally representative data sets (e.g., Mancino et al., 2010; Powell and Nguyen, 2013). When examining more long-term outcomes such as body weight, an individual fixed effect approach becomes more problematic (e.g., Schanzenbach, 2009). A second approach is to use instrumental variables (Hinrichs, 2010), which comes with limitations such as the exclusion restriction and access to a credible instrument.[5] Gundersen, Kreider and Pepper (2012) step back from identification and place bounds on the effect of participating in the free and reduced price lunch program. The bounding approach relies on the weaker *monotone instrumental variable* (MIV) assumption.[6] Bhattacharya, Currie and Haider (2006) used variation in the timing of the interview (i.e., if school is in session or not) coupled with SBP availability via difference-in-differences. Finally, Schanzenbach (2009) used regression discontinuity, which has assumptions similar in spirit to the MIV assumption. A drawback of regression discontinuity is that the effects are estimated only for those near the income eligibility cutoff, again, possibly masking any heterogeneous effects.

All of the aforementioned studies examined average effects. A major limitation of this approach is that it considers the effect of food source to be homogeneous for all children. This paper uses a quantile regression technique to determine the effect of food source on dietary quality across the entire distribution. Where this study differs methodologically from those previously published

[5] See Millmet and Tchernis, (2012) for an application to the SBP using alternative assumptions to estimate treatment effects without the exclusion restriction.

[6] The MIV assumption relaxes the exclusion restriction (see, Manski and Pepper, 2000). Gundersen et al. (2012) assume participation in the free or reduced price programs is monotonically associated with income to overcome selection into free or reduced-price programs. They find receipt of free or reduced-price lunches improves child health outcomes.

is the identification procedure for examining distributional effects. Identification is facilitated by using within-person variation of dietary intake on two nonconsecutive days but maintains the nonseparable property of the disturbance term, also called the rank variable (Powell, 2012a). In other words, rank is determined by "total proneness," which is function of both the fixed effect and random error. This advancement allows the coefficient of interest to be interpreted as the effect on the unconditional distribution. Contrast this with location-shift quantile estimators that model the fixed effect as a separate additive term (Canay, 2011; Galvao Jr., 2011; Graham et al., 2009; Lamarche, 2010; Ponomareva, 2011). By separating the total disturbance, the coefficient of interest is now interpreted as the effect on the conditional distribution.

The paper proceeds as follows: The next section more formally defines dietary quality and introduces a widely used measurement, the Healthy Eating Index-2005, which forms the basis of the analysis. After a brief overview of the data, I use summary measures to motivate a more detailed analysis. I then discuss the identification and estimation strategy, followed by the main results. The final section discusses policy implications and conclusions.

2. Dietary Quality

The overall healthfulness of a child's diet can be distinguished by two factors: *energy balance* and *dietary quality*. Energy balance is relationship between calories consumed and calories expended, which results in body weight management (Hall et al., 2012). Dietary quality, on the other hand, represents the degree to which a child's diet is meeting a set of criteria, for example, eating the correct proportions of healthy foods while maintaining moderation in less-healthy foods. It is important to note that energy balance and dietary quality are interconnected: experimental studies have shown when children switch to higher-quality diets, as opposed to calorie-restriction diets, sustained weight control is observed (Epstein et al. 2001, 2008).

I quantify dietary quality using the Healthy Eating Index (HEI). The HEI was developed in 1995 to measure compliance to the U.S. Government's official recommendations for healthful eating, the

Dietary Guidelines for Americans (DGA). Every five years, based on an expert advisory panel, the DGA are revised by the U.S. Departments of Agriculture (USDA) and Health and Human Services (HHS). As such, the HEI has been updated several times to reflect the most current state of nutrition knowledge. This paper uses the HEI-2005 and will henceforth refer to the HEI-2005 as simply HEI.[7]

The HEI is the sum of 12 components based on the consumption of various foods or nutrients. Each component assigns a score ranging from 0 to 5 (total fruit, whole fruit, total vegetables, dark green/orange vegetables and legumes, total grains, whole grains), 0 to 10 (milk, meats and beans, oils, saturated fat, sodium) or 0 to 20 for the percentage of calories from solid fats, alcoholic beverages, and added sugars (SoFAAS) creating a maximum score of 100. Appendix table A.1 provides exact details of the scoring (see also, Guenther et al., 2008a).

The HEI has been widely used and evaluated as a valid measure of diet quality (Guenther et al., 2008b). In the medical literature, lower HEI-2005 scores are associated with higher risks of coronary heart disease, stroke and diabetes (Chuive et al., 2012), cardiovascular disease (Nicklas et al., 2012), breast cancer (Shahril, 2012), colorectal cancer (Reedy et al., 2008) and prostate cancer (Bosire et al., 2013). Economists have used the HEI as an indicator of well-being to analyze distributional trends (Beatty, Lin, and Smith, 2012) and to study the impacts of the Supplemental Nutritional Assistance Program (Gregory et al., 2013). It is important to reiterate that the HEI is a per-calorie measure of dietary quality and does not directly consider excessive calorie intake. Although at first glance this distinction may seem limiting, it is important and necessary to analyze the relative *quality* of foods consumed across various food sources.[8]

[7] Future work will use the HEI-2010. Note that the two have many similarities (see, National Cancer Institute, 2013).

[8] It is also worth noting that analyzing energy balance is problematic for several reasons: (a) quantifying energy expenditure is difficult (Crouter, Clowers, and Bassett, 2005); (b) calorie needs vary substantially for boys and girls of different ages, making distributional comparisons within the population difficult; (c) most importantly, calorie consumption is not monotonic; one would need to make an assumption about the asymmetric relationship between under- and over-calorie consumption.

3. Data

I use data from three waves of the National Health and Nutrition Examination Survey (NHANES) covering 2003-08. Each survey wave is an independently drawn sample, which is representative of the U.S. with the USDA overseeing the food intake component. The NHANES provides rich information on dietary intakes so that HEI scores can be calculated according to Guenther et al. (2008a) (see also, the Appendix table A.1). Each wave was conducted from November in the odd year to October in the even year. For the 2003-08 NHANES, respondents report 24-hour dietary intakes on two nonconsecutive days. Day-one intakes are administered in-person during the medical exam, and day-two intakes are conducted 3–10 days later in a follow-up telephone interview. All interviews are conducted by trained dietary specialists with the aid of three-dimensional measuring instruments.

A primary goal of this research is to understand how school food affects dietary quality. As such, I focus on school-aged children (4–19) that report attending kindergarten through high school during the school year and have complete dietary intakes on both days ($n = 7,009$). Thus, children that have dropped out of school or graduated are excluded. I also exclude those attending schools that do not offer a lunch ($n = 379$), as done elsewhere in the literature (Gleason and Suitor, 2003; Gunderson et al., 2012; Millmet et al., 2000; Schanzenbach, 2009). The final sample with complete information consists of $6,630$ children.

Broadly defined, food from home (FFH) are items bought at the grocery store, food from school (FFS) are meals received at school, and food away from home (FAFH) primarily consists of fast-food and full-service restaurant items. Also included in FAFH are items bought in vending machines, received as a gift, and street food. Thus, for example, a candy bar purchased from a vending machine at school is considered FAFH, not FFS. Appendix A.2 contains complete details for mapping the 25 original food source codes into one of the three categories.

4. Summary Measures

In specifying regression models, I will use home-prepared food (FFH) as the reference category. This will give FFH a control group interpretation, which is reasonable since children in the U.S. eat at home nearly every day (table 1). FFS and FAFH will be considered the policy variables of interest (i.e., treatments). Food served in schools and at away-from-home venues are sources of political debate and subject to policy interventions.[9] Variation in FFS and FAFH is considerable as evidenced by table 1. Roughly 40% of children ate at school on at least one day in 2003-08, and over three-quarters of children reported FAFH consumption on day-one, day-two or both.

Table 1: Variation in consumption by food source

	Percentage consuming on...		
Food Source	Neither day	Day 1 or 2	Both days
Home	0.19	2.20	97.61
School	58.72	29.41	11.88
Away	21.65	43.09	35.27

Source: Children aged 4-19 reporting two days of intake in the 2003-08 NHANES.

4.1. Mean Measures

Table 1 suggests using individual variation to estimate the impact food served in schools and away from home on dietary quality. By including individual fixed effects and assuming conditional exogeneity, unobservable characteristics associated with selection into the SBP and/or NSLP are no longer confounding. This suggests a general OLS specification such as

$$HEI_{it} = \alpha_i + \mathbf{d}'_{it}\beta + \varepsilon_{it} \tag{1}$$

[9] For example, the NSLP has undergone numerous regulatory changes with respect to minimal nutritional values (see Ralston et al., 2008). The most recent change to nutritional standards for the SBP and NSLP begin to take affect in the 2013 school year (FNS-USDA, 2012a and 2012b). In fast-food and restaurant establishments, nutritional facts are now mandatory. Moreover, some states, such as New York, are currently seeking to further regulate away-from-home venues.

where HEI_{it} is the natural log of HEI on day t for child i, α_i are fixed effects (e.g., individual and/or time), $\mathbf{d}_{it}$ represents the policy variables of interest and ε_{it} is a disturbance term. The underlying assumption is that changes in $\mathbf{d}_{it}$ are uncorrelated with changes in ε_{it}, so that β is consistently estimated. This assumption seems reasonable in the current context given that intake records are spread 3-10 days apart.

Some care must be taken in defining the policy variable. One approach is to simply include a dummy variable D_{it}^k that equals one if child i consumed any food from food source k on day t. There are several limitations to this approach. For example, a child consuming two meals away from home would be categorized in the same manner as a child consuming one FAFH meal. Moreover, the nutrient density (i.e., nutrient per calorie) varies widely depending on where the food was sourced (Lin and Guthrie, 2012) and should be considered when specifying the model.

I consider an alternative specification: the proportion of daily calorie intake from food source k, which I will refer to as d_{it}^k. Under this specification, I am capturing the extent to which a child is "exposed" to each food source. In other words, estimates using d_{it}^k have an elasticity interpretation: the effect of substituting some share of calories from one food source to another. In short, d_{it}^k will be my preferred definition but I report both for comparison.

Column (1) of table 2 reports results using equation (1) with individual fixed effects. Panels A and B report results using the two alternative policy variable definitions D_{it}^k and d_{it}^k, respectively. For example, under the dummy variable definition in panel A, children that consume FFS exhibit a 3.1% increase in dietary quality as compared to a child that does not eat at school. Using the alternative definition in panel B, the estimates can be used to calculate the elasticity. That is, for a 10% increase in the share of calories from FFS, a child sees a 0.64% increase in dietary quality.[10]

To put coefficient estimates from panel B in perspective, it is useful to note that on days when children eat a school meal, roughly one-third of their daily calorie intake comes from this food source. This implies an *average* increase in dietary quality of 2.1% if a child were to shift these calories from home to school. When children frequent food-away-from-home establishments, an average of

[10]For a given percentage increase p, the elasticity is calculated as $p(e^{\beta} - 1)$.

Table 2: Mean regressions with fixed effects

	(1)	(2)	(3)	(4)
Panel A				
D^{FFS}	0.031	0.030	0.008	0.007
	(0.007)	(0.007)	(0.008)	(0.008)
D^{FAFH}	−0.042	−0.033	−0.034	−0.028
	(0.006)	(0.006)	(0.006)	(0.006)
Panel B				
d^{FFS}	0.062	0.064	0.019	0.019
	(0.019)	(0.019)	(0.021)	(0.021)
d^{FAFH}	−0.098	−0.082	−0.080	−0.071
	(0.012)	(0.012)	(0.012)	(0.012)
Fixed Effects				
Individual	Yes	Yes	Yes	Yes
Sequential day	No	Yes	No	Yes
Day of the week	No	No	Yes	Yes
Observations	13,260	13,260	13,260	13,260
No. of children	6,630	6,630	6,630	6,630
R-squared	0.011	0.018	0.023	0.027

Notes: Dependent variable is the natural log of HEI. Standard errors are in parentheses and calculated via bootstrapping accounting for stratification and clustering. $D^k = 1$ if food source k was consumed on day t. d^k = percentage of calories consumed from food source k on day t.

about 40% of daily caloric intake is consumed there. Thus, using estimates from panel B, one can infer an average decrease in dietary quality by 3.15% when consuming FAFH. In comparing panel A and B, estimates using definition D^k_{it} tend to be slightly larger.

Column (2) includes a dummy for the sequential day of intake record. This variable may be necessary due to differences dietary recall on day-one (in-person interview) versus day-two (phone call). The estimated coefficient on FFS changes very little, and the effect of FAFH is slightly less.

Since NHANES surveys individuals on all days of the week, column (3) of table 2 includes an additional fixed effect for the day of the week. Perhaps unsurprisingly, estimates for FFS change dramatically and are no longer significant, most likely because children do not typically attend

schools on the weekend. Conversely, FAFH is more likely to be consumed on the weekend, although estimates remain relatively constant.

Finally, column (4) reports coefficient estimates with individual, sequential day, and day of the week fixed effects. This will be the preferred specification moving forward. In summary, there appears to a positive but insignificant average impact of school food on dietary quality and a robust, negative average impact on dietary quality with respect to FAFH.

4.2. Distributional Measures

To motivate a distributional analysis, I present some summary measures by selected quantiles. To this end, in table 3 I compare two-day average HEI scores for those that never select into FFS or FAFH (column 1 in table 1) and those that report consuming FFS or FAFH on at least one day of intake (columns 2 and 3 in table 1).[11]

Table 3: Differences in the distribution of two-day average HEI-2005 scores by food source

	Quantile					
Food Source	5	25	50	75	95	N
Food from school						
At least one day	36.13	44.27	51.27	58.30	67.72	3,298
Neither day	33.45	42.80	50.25	57.27	68.07	3,332
Difference (%)	7.43	3.31	1.99	1.78	-0.52	
Food away from home						
At least one day	33.94	42.76	50.02	56.82	66.34	5,163
Neither day	35.19	46.26	53.47	60.78	71.15	1,467
Difference (%)	-3.69	-8.20	-6.90	-6.97	-7.26	

Note: A Kolmogorov-Smirnov type test of stochastic dominance (Barrett and Donald, 2003) indicates first-order dominance of FFS (at least one day) over FFS (neither day) at a 5% significance level. FAFH (neither day) first-order dominates FAFH (at least one day) at a 1% significance level.

We can see that the effect of FFS drops precipitously across the distribution of HEI scores,

[11] Dividing the sample in this manner is similar to the definition of D_{it}^{k} used in panel A of table 2.

implying that children below the median are responsible for a large share of the mean effect. Interestingly, for children with the poorest dietary scores at the fifth percentile, FAFH has the smallest impact. This result is most likely due to the presumption that home-prepared food is more similar to FAFH for this group of children. Beyond the bottom quartile, the effect of FAFH is relatively constant and thus more closely reflects the mean regressions.

Of course, results from table 3 are confounded by both observable and unobservable individual characteristics. Moreover, given the results of table 2, we should also expect the day of the week to play an important role in identifying a causal interpretation. In the next section, I use an estimator developed by Powell (2012a) to estimate unconditional quantile effects controlling for individual fixed characteristics.

5. Unconditional Quantile Estimation under Individual Heterogeneity

In this paper an important departure from the previous literature is how I estimate and identify the impact of food source on dietary quality. Given the summary results of the previous section, it is likely that the impact of food source is heterogenous across the distribution of diet quality. Much of this heterogeneity is most likely due to both observable and unobservable individual characteristics.

Typically, panel data can be used to control for individual heterogeneity via fixed effects. With mean regression, an additive term α_i is included in the specification, and the estimated coefficient on the treatment (policy) vector $\mathbf{d}$ can be interpreted as the impact on the unconditional mean. With quantile estimation, an additive fixed effect alters the interpretation of the coefficient of interest. The intuition behind this result is rather straightforward: the τth quantile of $y_{it}|\mathbf{d}_{it}, \alpha_i$ is in most cases not equal to the τth quantile of $y_{it}|\mathbf{d}_{it}$. For example, a high-quantile child in the distribution of of $y_{it}|\mathbf{d}_{it}$ may become a low-quantile child after *conditioning* on fixed effects.

The estimator used in this study accounts for individual heterogeneity without specifying or even estimating an individual "fixed-effect" parameter. Rather, unobservable individual heterogeneity is incorporated into the model by using within-person variation for identification but maintains the

nonseparable property of the total disturbance $u_{it}^* = f(\alpha_i, u_{it})$. This allows the parameter vector $\beta(\tau)$ associated with $\mathbf{d}$ to be interpreted as the τth quantile treatment effect on the unconditional distribution of dietary quality. If we are interested in knowing how food source affects low dietary quality children separately from high dietary quality children, this is precisely what we want to estimate.

5.1. Specification

Consider a cross-sectional quantile regression (QR) specification

$$y_i = \mathbf{d}_i' \beta(u_i^*), \quad u_i^* \sim U(0,1) \tag{2}$$

where for child i, y_i is dietary quality, $\mathbf{d}_i$ is a vector of the proportion of calorie intake from each food source and u_i^* is "total proneness" to consume a healthy diet.[12] Total proneness is a function of his or her unobservable proneness α_i (e.g., food preferences) and a disturbance term u_i (e.g., day-to-day randomness of food intake). In other words, u_i^* is a rank variable that incorporates heterogeneity into the model by allowing dietary quality to vary across children that have the same observed allocation of calories. It is necessary to assume the relationship between proneness and the outcome to be (weakly) monotonic. That is, children with a higher u_i^* are more prone to a healthier diet for a given allocation of calories across the three food sources.

Clearly, no cross-sectional distinction can be made between α_i and u_i but it is informative to see that the impact of covariates vary according to the nonseparable error term u_i^*. Moreover, an estimate of β using equation (2) assumes $\mathbf{d}_i$ is exogenous. If we believe individual-level characteristics influence a child's allocation of calories across food source, then $u_i^* | \mathbf{d}_i \nsim U(0,1)$.

To this end, it is useful to write down the Structural Quantile Function (SQF) introduced by

[12] One may also desire to include a set of controls $\mathbf{x}_i$, such as gender, age and race/ethnicity. The addition of "controls" alters the QTE interpretation of the estimates because some of u_i^* becomes observed through $\mathbf{x}_i$. Put differently, if $\mathbf{d}_i = (\mathbf{x}_i, \mathbf{z}_i)$ in equation (2) where $\mathbf{z}_i$ are the treatments of interest, $\mathbf{x}_i$ would also be interpreted as a treatment vector, a distinction that is not necessary in mean regressions. Specifically, estimates in this specification would provide the QTE on the distribution of $y_i | \mathbf{d}_i, \mathbf{x}_i$ rather than $y_i | \mathbf{d}_i$. See Powell (2013) for an in-depth discussion and estimation strategy of QTE in the presence of covariates.

Chernozhukov and Hansen (2005, 2008). The SQF of interest for identifying the unconditional QTE can be written as

$$S_y(\tau|\mathbf{d}) = \mathbf{d}'\beta(\tau), \quad \tau \in (0,1). \tag{3}$$

Equation (3) defines the τth quantile of the latent outcome $y_{\mathbf{d}} = \mathbf{d}'\beta(u^*)$ for a fixed allocation of calories and a randomly selected $u^* \sim U(0,1)$. This framework becomes important for describing the various "fixed-effect" quantile estimators and how they relate to the structural equation of interest.

Now consider a panel of students that report dietary intakes on multiple days. In this case, equation (2) can be rewritten as

$$y_{it} = \mathbf{d}_{it}'\beta(u_{it}^*), \quad u_{it}^* \sim U(0,1) \tag{4}$$

where $u_{it}^* = f(\alpha_i, u_{it})$. Again, α_i is the student's fixed level of proneness and u_{it} is an individual time-varying disturbance term. In this case, conditioning on individual fixed effects can overcome endogeneity concerns if u_{it} is uncorrelated with changes of $\mathbf{d}_{it}$. However, including an additive fixed effect in quantile regression, as done in mean regressions, alters the interpretation of the coefficients. For example, consider the two specifications

$$y_{it} = \alpha_i + \mathbf{d}_{it}'\beta(u_{it}) \quad \text{and} \quad y_{it} = \alpha_i(u_{it}) + \mathbf{d}_{it}'\beta(u_{it}) \tag{5}$$

of Koenker (2004) and Harding and Lamarche (2009), respectively. The underlying SQF for these two location-shift type specifications found in (5) take the form

$$S_{y_{it}}(\tau|\mathbf{d}_{it}, \alpha_i) = \alpha_i + \mathbf{d}_{it}'\beta(\tau), \quad \tau \in (0,1) \tag{6}$$

where τ now refers to the τth quantile of u_{it}, not u_{it}^*. In other words, the quantiles are now defined

relative to a the child's fixed level of diet quality. While correct in specification, it is not the ideal interpretation for our primary policy question: what is the effect of food source on low dietary quality children separate from high dietary quality children.

In the estimation strategy laid out below, the policy variables $\mathbf{d}$ are allowed unspecified correlation with individual fixed effects, $\alpha_i = h(\mathbf{d}_{i1}, \ldots, \mathbf{d}_{iT}, \varepsilon_i)$. This arbitrary correlation mirrors mean fixed effects, but α_i is not estimated. The estimator therefore assumes that the unconditional distribution of u^*_{it} is uniform but relaxes the conditional distribution assumption by allowing $u^*_{it}|\mathbf{d}_{it}, \alpha_i \nsim U(0,1)$. Specifically, I will estimate a specification that is related to the SQF taking the form

$$S_{y_{it}}(\tau|\mathbf{d}_{it}, \alpha_i) = \gamma_{ht}(\tau) + \mathbf{d}'_{it}\beta(\tau), \quad \tau \in (0,1) \tag{7}$$

where τ now refers to u^*_{it}, a child's total proneness to consume a healthy diet, which is precisely what we want to estimate. The parameter vector γ_{ht} plays the primary role for identification (sample moment 2 in the next section). The index h can refer to any set of exogenous characteristics that saturate the sample space over time t, or simply time itself. For example, it is likely that a "high-quality diet" on a weekday is much different than a "high-quality diet" on the weekend, if not for the simple fact that children do not attend school on the weekend. Therefore, I construct fixed effects based on the sequential survey day ($t = 1, 2$) and the day of the week in which the survey took place ($h = 1, \ldots, 7$).

6. Estimation

The SQF I will estimate is

$$S_{HEI_{it}} = \gamma_{ht}(\tau) + FFS_{it}\beta_1(\tau) + FAFH_{it}\beta_2(\tau), \quad \tau \in (0,1). \tag{8}$$

The underlying model corresponding to (8) is

$$HEI_{it} = \gamma_{ht}(u^*_{it}) + FFS_{it}\beta_1(u^*_{it}) + FAFH_{it}\beta_2(u^*_{it}) \tag{9}$$

where HEI_{it} refers to the natural log of HEI and γ_{ht} contains the 14 fixed effects as defined by the space $ht = \{h \times t\}$. Estimating equation (9) is not straightforward; the function is highly non-convex with many local optima, but it does have a well-pronounce global optimum. For brevity, I list the moment conditions here because they give intuition how estimation proceeds. See Powell (2012a) for full details of estimation.

Referring to equation (9), let $\mathbf{d} \equiv (\gamma_1, \ldots, \gamma_T, \mathbf{x})$ where $\mathbf{x} = (FFS, FAFH)$ are the policy variables of interest. To simplify notation of the moment conditions, I will refer to γ_{ht} as simply γ_t but note that the fixed effects still refer to the tth day of intake on the hth day of the week. The sample moments are

$$g_i(b) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}\left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) - \sum_{s=1}^{T}\mathbf{1}(y_{is} \leq \mathbf{d}'_{is}b)\right] \tag{SM.1}$$

$$h_t(b) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) - \tau \quad \text{for all } t. \tag{SM.2}$$

The fixed effects force $h_t(b) = 0$ for all t, thus confining all "guesses" of b to the parameter set $\mathcal{B}$,

$$\mathcal{B} \equiv \left\{ b \;\middle|\; \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) = \tau \quad \text{for all } t \right\}. \tag{10}$$

By letting $\tilde{b}$ be the the coefficient vector on $\mathbf{x}_{it}$ we can write $\mathbf{d}'_{it}b = \gamma_t + \mathbf{x}'_{it}\tilde{b}$. Recalling that we have allowed arbitrary correlation between the fixed effects and the policy variables, we can define $\gamma_t(\tau, \tilde{b})$ as the τth quantile of the distribution $y_{it} - \mathbf{x}'_{it}\tilde{b}$ for each fixed-effect value t. Therefore,

$\hat{\gamma}_t(\tau, \tilde{b})$ solves

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(y_{it} - \mathbf{x}_{it}'\tilde{b} \leq \hat{\gamma}_t(\tau, \tilde{b})) = \tau \tag{11}$$

and it immediately follows that for any guess $\tilde{b}$, $\hat{\gamma}_t(\tau, \tilde{b})$ is known.

Estimation proceeds in a Generalized Method of Moments (GMM) framework:

$$\hat{\beta}(\tau) = \arg\min_{b \in \mathcal{B}} \left(\sum_{i=1}^{N} w_i g_i(b)\right)' W_n(b) \left(\sum_{i=1}^{N} w_i g_i(b)\right) \tag{12}$$

where w_i is the sample weight, which has been normalized to sum to 1. For surveys with stratification and clustering, the weighting matrix $W_n(b)$ is defined following Bhattacharya (2005, equation 6). With one or two treatment variables, grid searching is computationally achievable (Chernozhukov and Hansen, 2008) but can still be quite burdensome in practice. If bootstrapping is necessary for inference, the problem is exasperated.

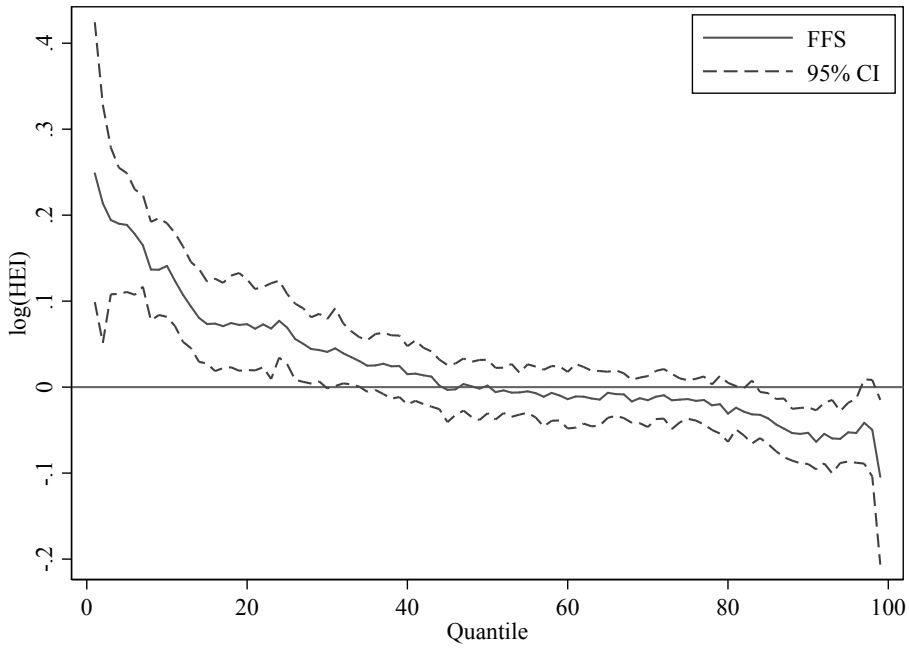
Yu and Moheed (2001) show that parameters in a quantile regressions can be estimated via Markov chain Monte Carlo (MCMC) algorithms. Chernozhukov and Hong (2003) generalized this technique into a GMM framework, which is suitable in this instance due to the complex survey design. Moreover, Chernozhukov and Hong (2003) show that inferences can be drawn from the posterior distribution. This key development dramatically reduces computation. To appreciate the degree to which MCMC speeds up computation, Chernozhukov and Hong (2003) note that computation is achieved at the parametric rate $1/\sqrt{B}$, where B is the number of draws. Grid-searching algorithms on the other hand have a nonparametric rate $(1/B)^{p/(d+2p)}$, where d is the parameter dimension and p is the smoothness order of the objective function.

In this paper, I use an adaptive MCMC algorithm. I follow the MCMC algorithm outlined in (Baker, 2013) by incorporating code provided by Powell (2012b). Appendix A.3 provides details of the algorithm. In short, estimators and 95-percent confidence intervals are taken from the mean and quantiles of the posterior distribution.

7. Results

Figures 1 and 2 plot coefficient estimates for food from school (FFS) and food away from home (FAFH), respectively, using equation (9). As a reminder, the policy variables are defined as the proportion of daily calorie intake from each food source. The results are interpreted as the marginal impact of reallocating calories from FFH to either FFS or FAFH. For example, the coefficient estimate for FFS at the fifth percentile of the HEI-2005 distribution is 0.189.[13] Thus, a 30-percent reallocation of calories from home to school, which is roughly the average proportion of caloric intake when school food is consumed, results in a 6.24% increase dietary quality. School meals are not of higher relative quality for all children; the coefficient estimate at the 90th percentile is -0.053, implying a 1.55% reduction in dietary quality when shifting 30% of calories from home to school.

Figure 1: Marginal Impact of Food from School (FFS) on the Unconditional Distribution of HEI-2005 Scores

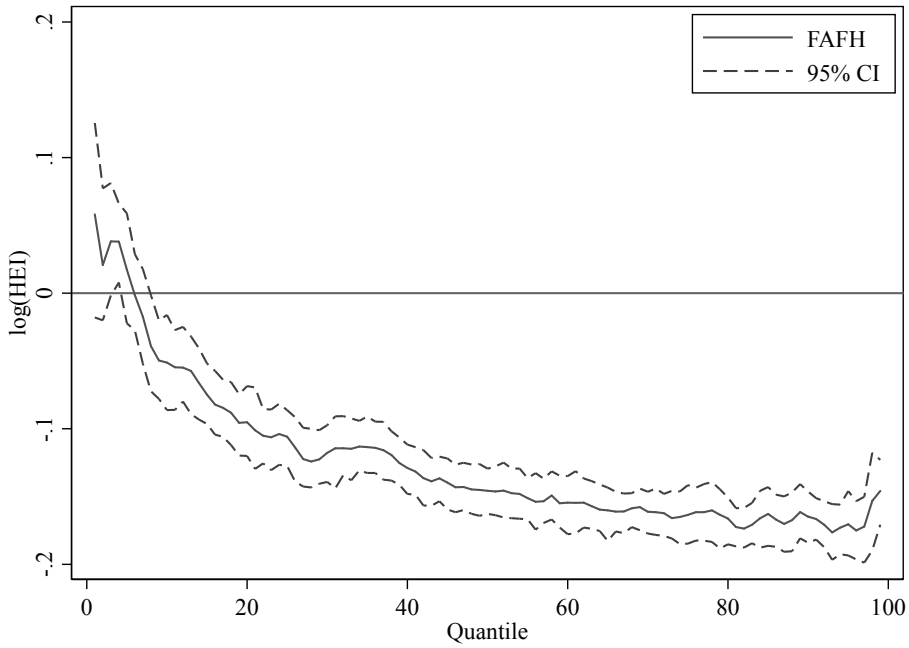


[13] In appendix table A.2, I report coefficient estimates and 95-percent confidence intervals for every fifth quantile. I also report results from a cross-sectional quantile regression as a basis of comparison (table A.3).

Although I do not estimate the long-run impacts on dietary quality, numerous experimental trials have shown that simple and repeated exposures to new and healthy foods have lasting impacts on dietary choices (Benton, 2004). For children falling in the lowest quartile of the HEI distribution, figure 1 implies a positive daily investment in nutrition skill formation, which could have long-run implications on nutrition capital accumulation.

Figure 2 clearly shows the negative effects of FAFH on dietary quality. One important finding from figure 2 is that home-prepared food is of no higher quality than FAFH for those falling in the bottom 10% of the HEI distribution. Coupled with the findings in figure 1, results suggest that a school meal is most likely the highest-quality meal these disadvantageed children receive.

Figure 2: Marginal Impact of Food Away from Home (FAFH) on the Unconditional Distribution of HEI-2005 Scores



8. Discussion and Conclusion

Food preferences reflect a complex cognitive structure rooted in early childhood experiences, exposures and environments. The formation of skills related to nutrition, such as ability to maintain energy balance and reach satisfactory levels of dietary quality, are learned and applied at early stages in life (Benton, 2004; Birch, 1999). The ability to adopt new skills, however, dissipates as one reaches adulthood (Morales et al., 2002). This insight into nutrition skill formation largely mirrors the expanding body of literature that has emerged in the past decade promoting skill formation and human capital development at early ages (Cunha et al., 2006).

The SBP and NSLP have undergone many reforms since the 1960's (Ralston et al., 2008). Originally aimed at alleviating hunger and malnutrition, these programs now strive to reach a balance between nutritional quality and caloric quantity. The most recent reform came from the Healthy, Hunger-Free Kids Act of 2010. Officially in effect for the 2012 school year, schools now have to meet new caloric and nutritional standards (USDA, 2012). Early evidence suggests higher standards for school meals improves child health outcomes (Taber et al., 2013). Moreover, localized experiments have shown that children are more likely to choose more nutritious meals after such a program is introduced and tend to make progressively healthier food choices the longer the program is in place (Grainger, Senauer and Runge, 2007).

Results of this study suggest there exists a large and meaningful impact of food served under National School Lunch and Breakfast Programs for children that exhibit low underlying dietary quality. These two Federal programs help children from disadvantaged environments to experience much needed dietary exposure and variety. The daily exposure to a higher quality meal potentially has a lasting and positive impact on nutrition capital accumulation. As policymakers and health advocates look to policy-amendable arenas to improve the American diet, this study suggests the NSLP and SBP are fertile grounds for intervention.

References

[1] Barrett, G.F. and S.G. Donald (2003). "Consistent Tests for Stochastic Dominance." *Econometrica*, 71(1): 71-104.

[2] Beatty, T.K.M., B-H Lin, and T.A. Smith, (2012). "Are Americans Eating Better? Distributional Changes in U.S. Dietary Quality 1989–2008." Selected paper for presentation at the American Agricultural Economics Association annual meeting, Seattle, WA.

[3] Benton, D., (2004). "Role of Parents in the Determination of the Food Preferences of Children and the Development of Obesity." *International Journal of Obesity* 28(): 858-869.

[4] Bhattacharya, D., (2005). "Asymptotic Inference from Multi-Stage Samples." *Journal of Econometrics* 126(1): 145-171

[5] Bhattacharya, J., J. Currie, S. Haider, (2006). "Breakfast of Champions? The School Breakfast Program and the Nutrition of Children and Families." *The Journal of Human Resources* 41(3): 445–466.

[6] Birch, L.L., (1999). "Development of Food Preferences." *Annual Review of Nutrition* 19: 41-62.

[7] Bosire, C., M.J. Stampfer, A.F. Subar, Y. Park, S.I. Kirkpatrick, S.E. Chiuve, A.R. Hollenbeck, J. Reedy. (2013) "Index-based Dietary Patterns and the Risk of Prostate Cancer in the NIH-AARP Diet and Health Study." *American Journal of Epidemiology* 177(6):504-513.

[8] Bowman, S.A., S. Gortmaker, C. Ebbeling, M. Pereira, D. Ludwig, (2004). "Effects of Fast-Food Consumption on Energy Intake and Diet Quality Among Children in a National Household Survey." *Pediatrics*, 113(1): 112–118.

[9] Canay, I.A., (2011). "A Simple Approach to Quantile Regression for Panel Data." *The Econometrics Journal* 14(3): 368–386.

[10] Center for Nutrition Policy and Promotion (CNPP), (2008a). "Diet Quality of Americans in 1994-96 and 2001-02 as Measured by the Health Eating Index-2005." U.S. Department of Agriculture. CNPP-37.

[11] Center for Nutrition Policy and Promotion (CNPP), (2008b). "Diet Quality of Low-Income and Higher Income Americans in 2003-04 as Measured by the Health Eating Index-2005." U.S. Department of Agriculture. CNPP-42.

[12] Chernozhukov, V. and H. Hong, (2003). "An MCMC Approach to Classical Estimation." *Journal of Econometrics* 115(2): 293–346.

[13] Chernozhukov, V. and C. Hansen, (2005). "An IV Model of Quantile Treatment Effects." *Econometrica* 73(1): 245-261.

[14] Chernozhukov, V. and C. Hansen, (2008). "Instrumental Variable Quantile Regression: A robust Inference Approach." *Journal of Econometrics* 142(1): 379-398.

[15] Chiuve, S.E., T.T. Fung, E.B. Rimm, F.B. Hu, M.L. McCullough, M. Wang, M.J. Stampfer, W.C. Willett. (2012). "Alternative Dietary Indices Both Strongly Predict Risk of Chronic Disease." *Journal of Nutrition* 142(6):1009-1018.

[16] Crouter, S.E., K.G. Clowers, and D.R. Bassett Jr., (2005) "A Novel Method for using Accelerometer Data to Predict Energy Expenditure." *Journal of Applied Physiology* 100(4): 1324-1331.

[17] Cunha, F., J.J. Heckman, L. Lochner, D.V. Masterov, (2006). "Interpreting the Evidence on Life Cycle Skill Formation." in *Handbook on the Economics of Education* vol. 1, eds. E.A. Hanushek and F. Welch.

[18] Doksum, K., (1974). "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case." *The Annals of Statistics* 2(2): 267–277.

[19] Epstein, L.H., C.C. Gordy, H.A. Raynor, M. Beddome, C.K. Kilanowski, R. Paluch, (2001). "Increasing Fruit and Vegetable Intake and Decreasing Fat and Sugar Intake in Families at Risk for Childhood Obesity." *Obesity Research*, 9(3): 171-178.

[20] Epstein, L.H., R.A. Paluch, M.D. Beecher, J.N. Roemmich, (2008). "Increasing Healthy Eating vs. Reducing High Energy-dense Foods To Treat Pediatric Obesity." *Obesity*, 61(2): 318-326

[21] Florence, M.D., M. Asbridge and P.J. Veugelers, (2008). "Diet Quality and Academic Performance." *Journal of School Health* 78(4): 209-215.

[22] Food Nutrition Service-U.S. Department of Agriculture (FNS-USDA). 2012a. "National School Lunch Program Fact Sheet." `http://www.fns.usda.gov/cnd/lunch/aboutlunch/nslpfactsheet.pdf`

[23] Food Nutrition Service-U.S. Department of Agriculture (FNS-USDA). 2012b. "The School Breakfast Program Fact Sheet." `http://www.fns.usda.gov/cnd/breakfast/AboutBFast/SBPFactSheet.pdf`

[24] Galvao Jr., A.F., (2011). "Quantile Regression for Dynamic Panel Data with Fixed Effects." *Journal of Econometrics* 164(1): 142–157.

[25] Grainger, C., B. Senauer and C.F. Runge (2007). "Nutritional Improvements and Student Food Choices in a School Lunch Program." *Journal of Consumer Affairs* 41(2): 265-284.

[26] Gregory, C., M. Ver Ploeg, M. Andrews, A. Coleman-Jensen, (2013). "Supplemental Nutrition Assistance Program (SNAP) Participation Leads to Modest Changes in Diet Quality" ERR-147, USDA, Economic Research Service.

[27] Gleason, P. and C. Suitor, (2003). "Eating at School: How the National School Lunch Program affects Children's Diets." *American Journal of Agricultural Economics* 85(4): 1047-1061.

[28] Guenther P.M., Reedy J., Krebs-Smith S.M. (2008a). "Development of the Healthy Eating Index-2005." *Journal of the American Dietetic Association* 108:1896-1901

[29] Guenther P.M., Reedy J., Krebs-Smith S.M., Reeve B.B. (2008b). "Evaluation of the Healthy Eating Index-2005." *Journal of the American Dietetic Association* 108:1854-1864.

[30] Hall, K.D., S.B. Heymsfield, J.W. Kemnitz, S. Klein, D.A. Schoeller, J.R. Speakman. (2012). "Energy Balance and its Components: Implications for Body Weight Regulation." *American Journal of Clinical Nutrition*, 95 (4), 989-994.

[31] Harding, M., and C. Lamarche, (2009). "A Quantile Regression Approach for Estimating Panel Data Models using Instrumental Variables." *Economic Letters* 104(3): 133–135.

[32] Heckman, J.J. and D.V. Masterov, (2007). "The Productivity Argument for Investing in Young Children." *Review of Agricultural Economics* 29(3): 446-493.

[33] Hinrichs, P., (2010). "The Effects of the National School Lunch Program on Education and Health." *Journal of Policy Analysis and Management* 29(3): 479-505.

[34] Hunter, M., (2013). "Adaptive Markov chain Monte Carlo Sampling and Estimation in Mata." *The Stata Journal* `http://arrow.hunter.cuny.edu/research/papers/HunterEconWP440.pdf`

[35] Koenker, R., (2004). "Quantile Regression for Longitudinal Data." *Journal of Multivariate Analysis* 91(1):24–89.

[36] Koenker, R. and G. Bassett, (1978). "Regression Quantiles." *Econometrica* 46(1):33–50.

[37] Jennings, A., A. Welch, E.M.F. van Sluijs, S.J. Griffin, and A.Cassidy, (2011) "Diet Quality is Independently Associated with Weight Status in Children Aged 9-10 Years." *Journal of Nutrition* 141(3): 453-459.

[38] Lin, B-H and J. Guthrie, (2012). "Nutritional Quality of Food Prepared at Home and Away From Home, 1977-2008." EIB-105, USDA-Economic Research Service.

[39] Mancino, L., J. Todd., J. Gutherie, B-H Lin., (2010). "How Food Away From Home Affects Children's Diet Quality." U.S. Department of Agriculture-Economic Research Service, ERR 104.

[40] Manski, C.F. and J.V. Pepper, (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68(4): 997-1010

[41] McFadden, D., 2008. "Human Capital Accumulation and Depreciation." *Review of Agricultural Economics* 30(3): 379-385.

[42] Millimet, D.L., R. Tchernis, M Husain, (2010). "School Nutrition Programs and the Incidence of Childhood Obesity." *The Journal of Human Resources* 45(3): 640–654.

[43] Morales M., D.K. Demory-Luce, T.A. Nicklas, T. Baranowski, (2002) "Consistency in food group consumption patterns from childhood to young adulthood: The Bogalusa Heart Study." Meeting of International Society of Behavioral Nutrition and Physical Activity, Seattle, WA.

[44] National Cancer Institute, (2013). "Comparing the HEI-2005 & HEI-2010" `http://riskfactor.cancer.gov/tools/hei/comparing.html`

[45] Nicklas, T.A., C.E. O'Neil, V.L. Fulgoni III. (2012) "Diet Quality Is Inversely Related to Cardiovascular Risk Factors in Adults." *Journal of Nutrition* 142(12):2112-2118.

[46] Ponomareva, M., (2010). "Quantile Regression for Panel Data Models with Fixed Effects and Small T: Identification and Estimation." Working paper, Northwestern Economics Department.

[47] Poti, J.M. and B.M. Popkin, (2011). "Trends in Energy Intake among US Children by Eating Location and Food Source, 1977-2006." *Journal of the American Dietetic Association* 111(8) 1156-1164.

[48] Powell, D., (2012a). "Unconditional Quantile regression for Panel Data with Exogenous or Endogenous Regressors." RAND Working Paper.

[49] Powell, D., (2012c). "Documentation for Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors code." RAND Working Paper.

[50] Powell, D., (2013). "A New Framework for Estimation of Quantile Treatment Effects: Non-separable Disturbance in the Presence of Covariates." RAND Working Paper.

[51] Powell, L.M., T.N. Nguyen, (2013). "Fast-Food and Full-Service Restaurant ConsumptionAmong Children and Adolescents." *Journal of the American Medical Association - Pediatrics* 167(1): 14-20.

[52] Ralston, K., C. Newman, A. Clauson, J. Guthrie, and J. Buzby, (2008). "The National School Lunch Program: Background, Trends, and Issues." ERR-61, USDA-Economic Research Service.

[53] Reedy J., P.N. Mitrou, S.M. Krebs-Smith, E. Wirfält, A. Flood, V. Kipnis, M. Leitzmann, T. Mouw, A. Hollenbeck, A. Schatzkin, A.F. Subar. (2008). "Index-based dietary patterns and risk of colorectal cancer: the NIH-AARP diet and health study." *American Journal of Epidemiology* 168(1):38-48.

[54] Taber, D.R., J.F. Chriqui, L. Powell, F.J. Chaloupka, (2013). "Association Between State Laws Governing School Meal Nutrition Content and Student Weight Status." *Pediatrics*: doi:10.1001/jamapediatrics.2013.399

[55] Schanzenbach, D.W., (2009). "Do School Lunches Contribute to Childhood Obesity?" *The Journal of Human Resources* 44(3): 684–709.

[56] Shahril, M.R., S. Sulaiman, S.H. Shaharudin, S.N Akmal. (2013). "Healthy Eating Index and Breast Cancer Risk among Malaysian Women." *European Journal of Cancer Prevention*, in print

[57] US Department of Agriculture (USDA), (2012). "Nutrition standards in the National School Lunch and School Breakfast Programs: Final Rule" http://www.gpo.gov/fdsys/pkg/FR-2012-01-26/pdf/2012-1010.pdf

A. Appendix

A.1. HEI-2005 Standards for Scoring

Table A.1: Healthy Eating Index-2005 standards for scoring.

	Score				
Component	0	5	8	10	20
Total fruit	0	$\longrightarrow$	$\geq$ 0.8 cup eq/1000 kcal		
Whole fruit	0	$\longrightarrow$	$\geq$ 0.4 cup eq/1000 kcal		
Total vegetables	0	$\longrightarrow$	$\geq$ 1.1 cup eq/1000 kcal		
Dark green/orange veg./legumes	0	$\longrightarrow$	$\geq$ 0.4 cup eq/1000 kcal		
Total grains	0	$\longrightarrow$	$\geq$ 3.0 cup eq/1000 kcal		
Whole grains	0	$\longrightarrow$	$\geq$ 1.5 cup eq/1000 kcal		
Milk	0	$\longrightarrow$		$\geq$ 1.3 cup eq/1000 kcal	
Meats and beans	0	$\longrightarrow$		$\geq$ 2.5 oz eq/1000 kcal	
Oils	0	$\longrightarrow$		$\geq$ 12 g/1000 kcal	
Saturated fat	$\geq$ 15	$\longrightarrow$	10 $\longrightarrow$	$\leq$ 7% of energy	
Sodium	$\geq$ 2.0	$\longrightarrow$	1.1$\longrightarrow$	$\leq$ 0.7 g/1000 kcal	
Calories from SoFAAS[a]	$\geq$50	$\longrightarrow$			$\leq$ 20% of energy

Source: Recreated from Guenther et al. (2008a).

[a]Solid Fat, Alcohol, and Added Sugar

A.2. Food Source Coding

In the descriptions that follow, bracketed numbers refer to the code found in the NHANES documentation. *Food at home (FAH)*: store [1], grown or caught by you or someone you know [19], and fish caught by you or someone you know [20]; *food from school (FFS)*: cafeteria at school [7]; *food away from home (FAFH)*: restaurant with waiter/waitress [2], restaurant fast food/pizza [3], bar/tavern/lounge [4], restaurant no additional information [5], cafeteria not at school [6], vending machine [14], common coffee pot or snack tray [15], from someone else/gift [16], mail order purchase [17], residential dining facility [18], sport, recreation, or entertainment facility [24], street vendor, vending truck [25], and fundraiser sales [26].

Contrary to other studies (e.g., Lin and Guthrie, 2013; Mancino et al., 2010), food from child care centers [8] is not included in FFS becuase this venue does not fall under the SBP or NSLP. "Other" food sources were also coded: Community food programs (family/adult day care center [9], soup kitchen/shelter/food pantry [10], Meals on Wheels [11], community food program - other [12], community program no additional information [13]), a catch-all other category (other, specify [91]) and unidentifiable responses (don't know [99]) made up a small proportion of total calorie intake. In preliminary analyses, I considered these items in a fourth "other" category and found them to be of relatively equal quality to FAFH. Therefore, these foods are considered to be FAFH

for this research. Please note however, that point estimates are robust to having a fourth category. Computationally, moving from 3 to 4 categories is not trivial, as the curse of dimensionality becomes a formidable problem for inference as discussed in the Estimation section.

A.3. Adaptive Markov chain Monte Carlo (MCMC) Algorithm

I heavily follow Hunter (2013, table 3). Specifically, I use a variant of the Metropolis-Hastings algorithm with vanishing adaptation. I use $t = 2500$ draws and discard (burn) the first 500. Starting values for the parameters are obtained from the standard quantile regression of Koenker and Bassett (1978), which allows for a smaller burn-in window. The initial variance matrix is taken to be the identity matrix. The proposal distribution is a multivariate Normal density with a targeted acceptance rate of 0.4. Adaptation is achieved through a damping parameter δ which controls how quickly the tuning mechanism $\rho_t = \frac{1}{(1+t)^\delta}$. Finally, the scaling parameter is $\lambda = \frac{2.83^2}{d}$ where d is the number of parameters to be estimated; $d = 2$ in this case.

A.4. Results for Selected Quantiles

Table A.2: Fixed Effects Quantile Regression Results

	FFS		FAFH	
Quantile	Estimate	95% CI	Estimate	95% CI
5	0.189	(0.111, 0.249)	0.017	(-0.022, 0.059)
10	0.141	(0.082, 0.191)	-0.051	(-0.086, -0.016)
15	0.073	(0.027, 0.123)	-0.075	(-0.097, -0.052)
20	0.073	(0.020, 0.125)	-0.095	(-0.120, -0.069)
25	0.069	(0.027, 0.108)	-0.106	(-0.128, -0.086)
30	0.041	(-0.002, 0.079)	-0.118	(-0.139, -0.097)
35	0.025	(-0.005, 0.055)	-0.114	(-0.133, -0.091)
40	0.015	(-0.020, 0.048)	-0.129	(-0.148, -0.112)
45	-0.003	(-0.041, 0.026)	-0.140	(-0.160, -0.122)
50	0.002	(-0.030, 0.032)	-0.146	(-0.163, -0.129)
55	-0.005	(-0.030, 0.027)	-0.151	(-0.167, -0.136)
60	-0.014	(-0.048, 0.018)	-0.154	(-0.178, -0.135)
65	-0.007	(-0.036, 0.018)	-0.160	(-0.183, -0.143)
70	-0.015	(-0.046, 0.013)	-0.161	(-0.177, -0.146)
75	-0.014	(-0.037, 0.008)	-0.164	(-0.185, -0.142)
80	-0.031	(-0.064, 0.005)	-0.166	(-0.185, -0.151)
85	-0.036	(-0.066, -0.007)	-0.163	(-0.186, -0.143)
90	-0.053	(-0.089, -0.024)	-0.165	(-0.184, -0.146)
95	-0.053	(-0.086, -0.018)	-0.170	(-0.193, -0.146)

Note: Dependent variable is log(HEI). Estimates from equation (9).

As a basis of comparison, I also report estimates from a cross-sectional quantile regression with fixed-effects for sequential day (α_d) and day of the week (α_w) in table A.3. Note that these estimates do not control for the individual fixed effect (α_i) and consider the sequential day and day of the week to be treatments (see Powell, 2013). In general, estimates tend to be larger under this specification than the fixed-effects quantile estimator of Powell (2012), although the two are not directly comparable.

$$HEI_i = \alpha_d(u_i^*) + \alpha_w(u_i^*) + FFS_i\beta_1(u_i^*) + FAFH_i\beta_2(u_i^*) \tag{13}$$

Table A.3: Cross-sectional Quantile Regression Results

	FFS		FAFH	
Quantile	Estimate	95% CI	Estimate	95% CI
5	0.227	(0.151, 0.304)	-0.064	(-0.119, -0.009)
10	0.150	(0.038, 0.262)	-0.098	(-0.161, -0.036)
15	0.119	(0.018, 0.221)	-0.115	(-0.168, -0.062)
20	0.082	(0.003, 0.161)	-0.150	(-0.195, -0.105)
25	0.046	(-0.025, 0.117)	-0.166	(-0.207, -0.125)
30	0.028	(-0.041, 0.096)	-0.186	(-0.227, -0.145)
35	0.013	(-0.053, 0.078)	-0.180	(-0.219, -0.141)
40	0.013	(-0.038, 0.064)	-0.178	(-0.209, -0.146)
45	-0.002	(-0.066, 0.062)	-0.181	(-0.222, -0.141)
50	-0.006	(-0.051, 0.040)	-0.192	(-0.221, -0.163)
55	-0.006	(-0.048, 0.036)	-0.204	(-0.231, -0.177)
60	-0.019	(-0.071, 0.034)	-0.201	(-0.235, -0.166)
65	-0.019	(-0.078, 0.039)	-0.198	(-0.237, -0.159)
70	-0.024	(-0.074, 0.026)	-0.196	(-0.230, -0.162)
75	-0.040	(-0.092, 0.012)	-0.198	(-0.234, -0.162)
80	-0.042	(-0.092, 0.008)	-0.206	(-0.241, -0.171)
85	-0.052	(-0.103, -0.001)	-0.207	(-0.240, -0.173)
90	-0.061	(-0.098, -0.024)	-0.188	(-0.214, -0.162)
95	-0.063	(-0.112, -0.013)	-0.191	(-0.227, -0.154)

Note: Dependent variable is log(HEI).