The Economic Determinants and Cognitive Effects of Childhood Malnutrition in the United States

Susan L. Averett and David C. Stifel
Department of Economics and Business
Lafayette College
Easton, PA 18042

The U.S. is currently facing a two-pronged battle against child malnutrition -- the prevalence of overweight children has increased dramatically over the past two decades and the percentage of children who are underweight remains unacceptably high. Both forms of malnutrition create well-known public health problems. But, less is known about how childhood over- or underweight affect cognitive functioning, behavior or self-esteem. In this research, we use data from the NLSY to investigate the causes of childhood malnutrition using quantile regression methods. We then use these findings and instrumental variables methods to separately estimate the effects of child malnutrition on self-esteem, cognitive functioning and behavior problems. We use county and state level data on availability of fast food outlets and fast food prices, and school district level data on soda consumption and physical education requirements as instruments to identify the effect of malnutrition on these child outcomes.

1. The Economic Determinants and Cognitive Effects of Childhood Malnutrition in the United States

The United States is currently facing a two-pronged battle against child malnutrition. On the one hand, the prevalence of overweight children has increased dramatically over the past two decades (Hedley et al., 2004). On the other hand, the percentage of children who are underweight remains unacceptably high for such a wealthy country (Polhamus et al., 2003). Both forms of malnutrition create public health problems. For example, an overweight child is more likely to be obese as an adult and has a higher probability of suffering from Type 2-diabetes, high cholesterol, high blood pressure, some types of cancer, and heart disease than is a child who is not overweight (Schwimmer et al., 2003; Dietz, 1998). Furthermore, the Surgeon General has linked childhood overweight to social discrimination and depression (U.S. Office of the Surgeon General, 2001). At the other end of the weight distribution, children who do not get enough to eat are likely to suffer from stunted growth and hindered mental development. They are also more likely to experience emotional, academic and behavioral problems than are well-nourished children (Kleinman et al. 1998). Paradoxically, although one often associates underweight with poverty, many poor children today are overweight. This has led some researchers to describe the problem as one of "misnourishment," where instead of getting the necessary healthy food that their bodies need, children take in excessive amounts of inexpensive fats and calories (Bhattacharya and Currie, 2001). The strain that these consequences of child malnutrition will place on the health-care system, as well as concerns over the future economic effects of malnutrition-induced diminished productivity (Owens, 1989), has motivated researchers to try to understand the determinants and effects of child malnutrition.

As such, the objectives of our proposed research are twofold. First, we plan to study the determinants of child malnutrition using an econometric method that permits us to exploit the fundamental differences between undernutrition and overnutrition. Advances in the understanding of quantile regression methods and in computing power that make this technique more feasible, lead us to question the results of previous research that has focused primarily on either ordinary least squares (OLS) (Chou et al, 2002; Lakdawalla and Philipson, 2002; Anderson et al, 2003a; Cutler et al, 2003; Komlos and Baur, 2004) or probit methods (Ruhm, 2004; Anderson et al, 2003a; Cutler et al, 2003).[1] The former estimates the relationship between nutrition outcomes and explanatory covariates through the point of means. In other words, OLS models explain central tendencies or variations in nutrition outcomes for the average child. Since the average child in the United States is neither underweight nor overweight, this econometric method does not address the object of interest. Overweight children are at the upper tail of the weight distribution, while underweight children are at the lower tail. Neither is in the middle. Quantile regressions, which are described in more detail below, permit analysts to estimate the determinants of nutrition outcomes at any percentile of the weight distribution, including the upper and lower tails. While probit methods do estimate these relationships through the points of interest in the distribution, they throw out important (and available) distributional information on the dependent variable. We plan to improve upon previous research in this area by using quantile regression methods. In addition, since most of the recent economic research on the determinants of childhood malnutrition in the United States has focused on overnutrition (commonly referred to as obesity in adults and overweight when referring to children) at the expense of undernutrition[2], we will consider both over- and undernutrition in children in the United States. Like

[1] Abrevaya (2001) applied quantile regression methods to birth outcomes, but not to subsequent nutrition outcomes of children.

[2] Most of the research on determinants child undernutrition has concentrated on developing countries where its severity is admittedly much greater (see Strauss and Thomas, 1998 for a review of this literature). The studies that do use United States data either examined the determinants of low birthweight (Rous et al., 2004; Rosenzweig and Wolpin, 1991, Rosenzweig and Schultz, 1983) or

other recent research in this area, we will focus on disentangling causality from correlation. Furthermore, we will collect data on prices and availability of fast-food restaurants, proxies for exercise, and information on school contracts with soft drink manufacturers. This will complement an already rich dataset that we plan to use as our primary source of data in these estimations, the National Longitudinal Survey of Youth 1979 Cohort (NLSY79). These variables will allow us to identify the effects of malnutrition on various outcomes as discussed below.

Second, because there are concerns that undernutrition may seriously affect a child's cognitive and behavior functioning, and because overnutrition may lead to low self-esteem which may result in poor cognitive functioning, we plan to estimate the extent to which malnutrition (or perhaps more appropriately misnutrition) affects cognitive and behavioral functioning and self-esteem. We are interested in determining whether or not those children who are under or over nourished suffer cognitive or behavioral consequences. However, the potential for reverse causality is clear. Those children who have behavior problems or poor cognitive functioning may over or under eat to compensate for that making childhood malnutrition endogenous in these outcome models. Thus, we first estimate the models that predict whether or not a given child is malnourished and use these estimates as inputs into our second stage: whether or not malnutrition causes poor cognitive or behavioral functioning. In doing so, we remain cognizant of the fact that a simple binary indicator of malnutrition discards important information on the intensity of malnutrition. We therefore will also estimate how the severity of malnutrition affects child functioning including cognitive ability, behavior and self-esteem. This aspect of childhood over and underweight has been neglected but is particularly important to both parents and to policymakers. We describe how we plan to accomplish these goals after discussing trends in child malnutrition in the United States and the recent literature related to our study.

2. Child Malnutrition in the United States

2.1 *Measurement and Trends*

A standard for measuring child nutritional outcomes in developed countries is the Body Mass Index (BMI), which is defined as the ratio of weight in kilograms over height in meters squared.[3] A particular child's BMI can be compared to those on tables configured by the Centers for Disease Control (CDC), which established distributions for each sex by age because, for children, BMI levels in the reference population differ by age and gender. A child is typically considered overweight if his/her BMI for age is over the 95th percentile of the healthy reference population,[4] while he/she is considered at risk for overweight with a BMI for age above the 85th percentile. Children classified as underweight are those with BMI for age measures less than the fifth percentile of the reference population (CDC, 2000). In the population, prevalence rates for overweight, at risk of overweight, and underweight are calculated using these criteria.

---

evaluated assistance programs such as WIC (Rush et al, 1988; Currie, 2002; and Carlson and Senauer, 2003), Medicaid (Currie and Gruber, 1994; and Currie, 2002), food stamps (Davis, 1994; and Currie, 2002) and the government food programs (Battacharya et al., 2004; Currie et al., 2004).

[3] Other metrics include stature for age and weight for age for all children (see CDC, 2000; and Martorell and Habicht, 1986), and overall evaluation of child health by physicians for very young children (Wolfe and Sears, 1997). While stature for age is a commonly used measure of chronic malnutrition in developing countries, we do not use it in our analysis as we consider children up to the age of 15. Martorell and Habicht (1986) find that less than 10 percent of the worldwide variance in height can be ascribed to genetic or racial differences among children under the age of five. Genetic factors play a much larger role at older ages and as such, stature for age is not an appropriate measure in our analysis given our sample of children (described in more detail below).

[4] See CDC (2000) for a discussion of the reference population.

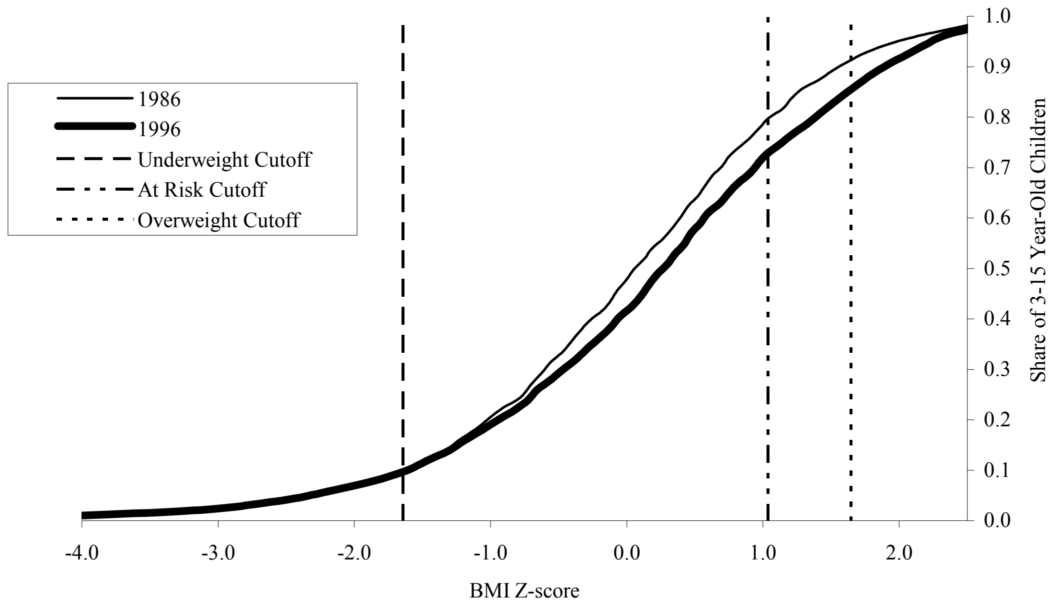
To compare BMI measurements across age and gender cohorts, normalized BMI z-scores (hereafter BMIZ) are calculated. The z-score for a child *i* is defined as follows:

$$BMIZ_i = \frac{BMI_i - BMI_{mean}}{\sigma_{BMI}}$$

where BMI_i is the child's BMI measurement, BMI_{mean} is the average BMI measurement for the healthy reference population of the same age and gender, and σ_{BMI} is the standard deviation of BMI measurements for the healthy reference population of the same age and gender. the BMIZ for the reference population has a standard normal distribution in the limit. Thus, there is a probability distribution on the expected value of a BMIZ for any given child – a standard normal distribution to be precise. This means that there is a five percent probability that a child from the reference population will have a BMIZ greater than 1.645. In other words, 1.645 is the BMIZ cutoff for the 95th percentile (overweight) in the distribution of BMI for age. Similarly, the cutoffs for the fifth (underweight) and the 85th (at risk of overweight) percentiles are -1.645 and 1.0365, respectively.

To illustrate, in Figure 1 we plot the 1986 and 1996 distributions of the normalized BMI for age measures for children in the NLSY79 dataset (described in more detail later in this proposal.)

Figure 1: Distribution BMI-for-Age
for Children of Age 3-15 in the United States



This figure illustrates that the prevalence of overweight children in this sample rose from 8.6 percent to 14.5 percent between 1986 and 1997 (as seen by the intersections of the distributions and the overweight cutoff). Further it shows that the prevalence of underweight children remained relatively constant at approximately 9.8 percent.[5] With regard to overweight children, the figure also

[5] These estimates are potentially biased because the age distribution in the 1996 sample is weighted toward older children relative to the 1986 sample. As noted later in Section 4, child's age has a negative effect on BMIZ at the 95th percentile. The implication of this is that the BMI z-scores of the

highlights a weakness of prevalence measures. Not only has the share of children who are overweight increased, but the degree to which these children are overweight has increased substantively (as seen by the 1996 distribution being considerably lower than the 1986 distribution in the region to the right of the overweight cutoff).

Discrete measures of over- or undernutrition such as prevalence rates are important for information-dissemination purposes as they are something that the general public can easily comprehend. However, focusing only on specific cutoffs such as being above the 95th percentile or below the fifth percentile can be misleading for two reasons. First, it puts undue emphasis on the admittedly arbitrary cutoff points. Marginal changes in the cutoff point can lead to categorical changes in the recorded health status of a child whose BMI for age measure is near the cutoff. Second, it ignores the distribution of BMI around the cutoff points. Thus, in our research, we borrow from the poverty literature by estimating not only prevalence rates, but also measures of the depth and severity of malnutrition (for overweight, see Jolliffe, 2004; and for underweight, see Sahn and Stifel, 2002).

The measures of the prevalence, depth and severity of malnutrition belong to a class of malnutrition measures that we refer to as M_α. These are defined as follows for underweight:

$$M_\alpha = \tfrac{1}{N}\sum_{i=1}^{N}(cut - BMIZ_i)^\alpha 1(BMIZ_i < cut),$$

and as

$$M_\alpha = \tfrac{1}{N}\sum_{i=1}^{N}(BMIZ_i - cut)^\alpha 1(BMIZ_i > cut),$$

for overweight, where *cut* is the under- or overweight threshold, and 1(.) is an indicator function that takes on a value of one when its argument is true, and zero otherwise. The parameter, α, can be interpreted as a malnutrition aversion parameter, similar to the poverty aversion parameter in the Foster-Greer-Thorbecke class of poverty measures (Foster, et al., 1984). When α is zero, M_0 is the prevalence of malnutrition. When α is one, M_1 is the average malnutrition gap, where a child's gap takes on a value of zero if he or she is not malnourished. We refer to this measure as the depth of malnutrition. M_2 can be interpreted as the severity of malnutrition as it is a weighted average of the malnutrition gaps where the weights are the gaps themselves. The prevalence of malnutrition (M_0) is related to the number of malnourished. The depth of malnutrition considers the distance that the malnourished children are from the threshold, but weights each child equally. The severity puts more weight on those who are furthest away from the threshold. As α approaches infinity, the social welfare function associated with the malnutrition measure is Rawlsian. In this extreme case, when comparing two distributions, the distribution with the most malnourished child is considered to have more malnutrition.

In Table 1, we present these types of malnutrition metrics applied to the NLSY79 data. Although underweight is typically thought of as a phenomenon only afflicting developing countries, it clearly occurs in the United States too. Indeed, we estimate that 9.8 percent of children between the ages of three and 15 in our sample had BMI levels that fell below the fifth percentile cutoff in 1986 and in 1996.[6] This is 4.8 percentage points greater than what would be expected in a healthy

---

children at the upper tail of the 1996 distribution are likely to be biased downward, as are the prevalence rates for 1996 compared to 1986. Conversely, we find that a child's age has a positive effect on the BMIZ at the lower tail of the distribution. This suggests that underweight measures are also likely to be biased downward for the 1996 distribution relative to the 1986 distribution. These potential biases reinforce our concerns about child malnutrition trends.

[6] This is consistent with Grigsby's (2003) estimate of an incidence rate less than 10 percent, though her estimate is a measure of protein-energy malnutrition (PEM), not underweight.

population. The prevalence, depth and severity of underweight children did not change substantially over the decade from 1986 to 1996. This can also be seen in the form of the stable lower tails of the BMI for age distributions that appear in Figure 1.

These estimates of undernutrition outcomes are paralleled in the literature on input measures such as "food insecurity" and hunger. For example, according to the United States Department of Agriculture (USDA, 2004), in 1999, 14 million children lived in "food insecure households," which means that their families lacked access to enough food to meet their basic steady state needs (Center on Hunger and Poverty, 1999). Another recent survey estimated that approximately 4 million American children experienced prolonged periods of food insufficiency and hunger each year. This is roughly 8 percent of all the children under the age of 12 living in the United States. The same study shows that an additional 10 million children are at risk for hunger (Kleinman et al. 1998). Finally, in a state by state analysis of food insecurity in the U.S., Nord et al. (1999) estimate that 9.7 percent of all households were food insecure during the years 1996-1998.

Table 1: Malnutrition Among American Children between Ages 3-15

Year	Prevalence	Depth	Severity
		Underweight	
1986	9.8%	0.09	0.16
1996	9.8%	0.09	0.14
		At risk of overweight	
1986	20.5%	0.14	0.16
1996	27.1%	0.21	0.24
		Overweight	
1986	8.6%	0.05	0.05
1996	14.5%	0.08	0.08

Source: Authors' calculations from NLSY79

Not surprisingly, food insecurity is most prevalent in poor families. The Center for Hunger and Poverty estimates that 35.4 percent of families below the poverty line are food insecure compared to only 10.2 percent of households nationwide. Paradoxically, however, children who live in poverty can also be overweight – perhaps because they lack access to healthy, nutritious low-fat foods (Center for Hunger and Poverty, 1999) – which adds to the confusion over the causes of under- and overnutrition.

Part of this paradox apparently stems from changes in food technologies and prices. As fast foods become more easily available and as the prices of high-calorie "junk" foods fall more quickly than the prices of fresh fruits and vegetables, the poor stretch their limited budgets by substituting out of the latter into the former (Currie et al., forthcoming; and Kennedy and Goldberg, 1995). Bhattacharya and Currie (2001) found that nearly 20 percent of their sample of food-insecure youths were overweight, with almost one-third consuming excess amounts of sweets. Even adolescents who are not "food insecure" are likely to be malnourished – a concept Bhattacharya and Currie (2001) refer to as "misnourishment." Indeed, the determinants of food insecurity and malnutrition outcomes (underweight and overweight) are quite different. It is because of this difference and the apparent poverty-obesity "paradox" that Currie et al. (forthcoming) conclude that, controlling for poverty, food insecurity is simply not a good predictor of poorer nutrition outcomes.

As indicated in Table 1 and in Figure 1, the prevalence of underweight among children in the United States has remained stable in recent years. The same, however, cannot be said for the prevalence and degree of overweight children. Using the NLSY79 data, we find that the share of children who are overweight rose by nearly six percentage points. Further, the depth of overweight rose from an average of 0.05 standard deviations above the cutoff to 0.08 standard deviations. In other words, not only is there a larger share of children who are considered to be overweight, the degree to which they are heavier has grown substantially.

This rapid rise in overweight children has been particularly pronounced over the past 25 years. A Department of Health and Human Resources report (2002), estimates that for a similar age group (6 to 19), 15 percent (almost 9 million) were overweight in 1999-2000. This is triple the rate in 1980. Among a younger cohort of children between the ages of two and five, over 10 percent are overweight, representing a 7 percent increase from 1994 (Ogden et al., 2002).

2.2 *Why Has Childhood Overweight Been Increasing?*

A review of the recent literature on childhood overweight guides our empirical work on the determinants of childhood malnutrition. From a strictly accounting perspective, children will gain weight when the calories they take in are greater than the calories that they expend. This suggests that the percent of overweight children is rising either because more calories have been taken in and activity levels have remained the same or fallen, or because caloric intake has remained constant and activity levels have fallen off. Many researchers have used this framework to examine how caloric intake and activity levels have changed among adults.[7] Philipson and Posner (forthcoming) argue, for example, that technological changes allow us to expend far fewer calories than we take in, leading to a rise in obesity over time. Other researchers implicate anti-smoking campaigns, the falling prices of food, increasing prevalence of fast food restaurants, job strenuousness and unemployment as predictors of the upward trend in adult obesity (Rashad and Grossman, 2004; Lakdawalla and Philipson, 2002; Chou et al, 2002; Philipson, 2001; Ruhm, 2000)

Children, of course, do not have as much control over what they eat as adults do. Thus, many of the same factors that are implicated in rising adult obesity also likely contribute to the increase in childhood overweight. For example, television viewing, which is a sedentary activity, is often thought to be a major cause of childhood overweight as there is a strong positive correlation between a child's BMI and hours spent watching television (Lowry et al., 2002; Dennison et al., 2002; Gortmaker et al., 1996). Television also exposes children to a variety of advertisements for unhealthy food, which is also believed to be a factor explaining the rise in childhood overweight (Taras and Gage, 1995; Kraak and Pelletier, 1998). Finally, children may eat while watching television thus taking in more calories in a sedentary setting. Slyper (2004) notes, however, that evidence of declining physical activity leading to the rise in childhood obesity is scarce in large part because we lack good data on recent trends in the physical activity of children. He also notes that there is some evidence that if children are overweight, they may choose not to engage in physical exercise because it is uncomfortable for them. Indeed, the direction of causality here is not entirely clear – are kids becoming overweight because they are watching more television or are overweight children choosing to watch more television?

[7] There is a genetic component to overweight as well. But, since *changes* in childhood overweight occurred largely in recent decades, most researchers rule out genetics as a cause of the increase. Genetic change simply occurs too slowly over time to explain these changes in nutrition outcomes (Anderson, et al., 2003b).

Breastfeeding is thought to offer some protection against obesity as breastfeeding is more self-regulatory – the baby drinks until full rather than until the caregiver decides he is full. Formula, which is more energy dense, is also likely to lead to a greater caloric intake by infants. Several studies using U.S. data find a strong negative correlation between breastfeeding and BMI (see, for example, Gillman et al., 2001), and the American Academy of Pediatricians (2003) actively encourages women to breastfeed in part to protect their children against obesity. In contrast, a recent study using data from a British cohort concluded there is no evidence that breastfeeding influenced a child's BMI (Li et al., 2003). Changes in breastfeeding over time in the U.S. have not followed a clear pattern. The prevalence of breastfeeding increased between 1970 and the mid-1980s and then declined until 1990 when it then began slowly increasing (Ross Products, 2002). Thus, changes in breastfeeding habits are not likely an important explanation of the rise in the prevalence of overweight children.

Recent research has found evidence suggesting that the rise in maternal employment, which closely parallels the rise in childhood overweight, has contributed to the prevalence of overweight in children (Anderson et al., 2003a; and Ruhm, 2004). Researchers speculate that working mothers may rely on high-fat fast foods, childcare providers may not provide healthy snacks, and/or unsupervised children may indulge in unhealthy snacks or watch more television. Furthermore, for safety reasons, "latchkey" children may be confined to their houses while their parents work, thus limiting their access to physical exercise.

Fast food is almost always implicated in any discussion of childhood overweight. There is evidence that children eat far more fast food now than they did several decades ago. Children are more likely today to eat their meals away from home. In the late 1970s, children obtained about 2 percent of their total calories from fast food. In the late 1990s, this percent rose to 10 percent (Ebbling et al., 2002; Lin et al., 2001). It is well known that fast food is high in fat and that portion sizes tend to be large (Ebbling et al., 2002). Furthermore, the number of fast food outlets has increased dramatically over time and the prices of fast food have fallen. Chou et al. (2002) report that between 1972 and 1997 the number of fast food restaurants per capita doubled. Fast food is particularly appealing to children as the child's meal often comes with a toy or game, it tends to appeal to a child's palate, and, because it is fast, parents may choose it when dining out with children.

Consuming soda is also implicated in childhood overweight. Ludwig et al. (2001) found that BMI increases with each sugar-sweetened beverage a child drinks. Children who consume excessive amounts of "empty calories" (i.e., food with little nutritional value) such as sugar sweetened sodas tend to deprive their bodies of important nutrients such as calcium (Orlet et al., 2001). Anderson et al. (2003b) note that many schools have contracts with companies to supply soft-drinks and other "junk" foods to schools, and that the consumption of such foods in school may also lead to overweight. In response to such contracts, the American Academy of Pediatrics (2004) has called upon schools to restrict student access to soft-drinks. Data from the USDA (2000) shows that soda consumption increased since the late 1940s in the U.S. and that milk consumption decreased over the same time period which corresponds to the trends in childhood overweight. However, recent evidence indicates that children have not increased their caloric intake over time (with the exception of adolescent girls). Indeed, it is not necessarily the level of caloric intake that has increased. Rather, the type of calories ingested has changed with American children consuming more carbohydrates than before (Sylper, 2004).

As is clear from the above studies, there are several possible explanations for the rapid increase in the prevalence and severity of overweight children. In order to sort these out, we will extend previous research in this area by, *inter alia*, including information on television watching, the availability and prices of fast food, soda consumption and physical activity.

3. Data

To investigate the correlates and possible causes of malnutrition in children, we plan to use data from the NLSY79 a panel study of approximately 12,000 individuals who were first interviewed in 1979 when they were between the ages of 14 and 22. The female respondents were re-interviewed annually from 1979-92 and bi-ennially since 1994. The data are a nationally representative sample of individuals born between 1957 and 1964 with an oversampling of the black, Hispanic, and low income white populations. Because of this, sampling weights are used when estimating summary statistics. The data include information about economic and demographic behavior and outcomes for the respondents and their families. In our preliminary analysis for this proposal, we used the sample of these data prepared by Anderson et al. (2003a). We are grateful to them for graciously providing us with their data. The NLSY79 is also used by Ruhm (2004) to examine the effects of maternal employment on a variety of outcomes including whether or not a child is overweight. We are currently creating our own sample for analysis.

Our initial analysis focuses on the original NLSY79 female respondents and includes data through the 1996 survey year. At this point, the mothers are between the ages of 31 and 41 and the children range in age from 3 to 15. It is worth noting that the sample of children in the NLSY79 are born disproportionately to younger mothers. This is potentially troubling because these women tend to have lower education and income levels. However, a great deal of information is available about the family circumstances of these children over time (e.g., prenatal care and birthweight, family income, household composition, family structure, and family background).

4. Estimating the Economic Determinants of Malnutrition with Quantile Regressions

As mentioned previously, the two most commonly used forms of estimation for models of the economic determinants of nutritional outcomes are OLS (Chou et al, 2002; Lakdawalla and Philipson, 2002; Anderson et al, 2003a; Cutler et al, 2003; Komlos and Baur, 2004) and probits (Anderson et al, 2003a; Cutler et al, 2003). Each method has its respective weakness.

A drawback to using OLS to estimate nutrition models is that the OLS parameter estimates reflect the effects of the explanatory variables at the point of the means of the independent variables and the dependent variable, the BMIZ. In a country with children suffering from being extremely overweight such that the mean BMIZ is 1.65 (the threshold below which 95 percent of the healthy reference population exists), OLS is an appropriate methodology to estimate the determinants of overweight. However, for the United States' data that we are using, the sample mean BMIZ is 0.113. If we are interested in explaining the socio-economic determinants of nutritional outcomes for children who are overweight, we should estimate regression lines/planes that go through, say, 1.65, rather than mean z-score. Quantile regression techniques permit us to do just that. Similarly, if we are interested in the determinants of undernutrition of children, we should estimate the regression through the 5th percentile of the healthy population (i.e., the z-score value of –1.65).

The benefit of using probit models to estimate the determinants of child overweight is that they permit the analyst to concentrate on the portion of the distribution in question. In this case, an indicator variable is defined to denote if a child is overweight (z-score above 1.65) or not overweight (z-score below 1.65). This dichotomous variable is then used as the dependent variable and the probit estimates the effect of explanatory variables on the probability of the child being overweight. The drawback to using probits to estimate such models is that important distributional information on the dependent variable is thrown out.

To understand quantile regressions[8] and how they differ from OLS and probits, it is instructive to start with the simplest case – median regressions. The median regression can be defined by minimizing the sum of the absolute value of errors, as opposed to the sum of the squared errors as in

[8] Much of the following discussion follows Deaton (1997) and Johnston and DiNardo (1997).

the OLS regression. As a means of illustrating what is known as the Least Absolute Deviations (LAD) estimator, let us define a model in which we postulate that a child's BMI z-score (y) is a function of observed individual, household and community characteristics (x),

$$y_i = x_i'\beta + \varepsilon_i . \tag{1}$$

The median regression coefficients can be obtained by solving the following,

$$\min_{\beta} \sum_{i=1}^{N} \left| y_i - x_i'\beta \right| = \min_{\beta} \sum_{i=1}^{N} (y_i - x_i'\beta)\,\mathrm{sgn}(y_i - x_i'\beta) , \tag{2}$$

where

$$\mathrm{sgn}(z) = \begin{cases} 1 & if \quad z > 0 \\ -1 & if \quad z \le 0 \end{cases} .$$

The first order conditions (or normal equations) necessary to choose the parameters that minimize (2), give some intuition of how this model works. These k equations for $j = 1,\ldots,k$, are

$$\sum_{i=1}^{N} x_{ij} \cdot \mathrm{sgn}(y_i - x_i'\beta) = 0 . \tag{3}$$

There are two items of note. First, if there is only a constant in x, then (3) defines β as the value of the dependent variable that has an equal number of points on either side of it. This is simply the median of y. Second, unlike the OLS normal equations, it is only the sign of the residuals, and not their magnitude that matters.

To generalize this to a regression of any quantile, let us first define the multiplier h_i

$$h_i = \begin{cases} q & if \quad y_i - x_i'\beta > 0 \\ 1-q & if \quad y_i - x_i'\beta \le 0 \end{cases} ,$$

where $0 < q < 1$ is the quantile of interest. Thus the quantile regression coefficients for quantile, q, can be found by solving

$$\min_{\beta} \sum_{i=1}^{N} h_i \left| y_i - x_i'\beta \right| = \min_{\beta} \left[q \sum_{y > x'\beta} (y_i - x_i'\beta) - (1-q) \sum_{y \le x'\beta} (y_i - x_i'\beta) \right] . \tag{4}$$

The first order conditions that correspond to (4) are now

$$\sum_{i=1}^{N} x_{ij} \left[q - 1(y_i - x_i'\beta \le 0) \right] = 0 , \tag{5}$$

where, once again, 1(.) is an indicator function that takes on a value of one when its argument is true and zero otherwise. This is equivalent to (3) when q is one half. Again, if x contains only a constant, then (5) defines β as the value of the dependent variable that has 100q percent of the sample above it, and 100(1-q) percent of the sample below it. This is simply the definition of the quantile, q.

The point estimates are made using linear programming, while the variance-covariance matrix of the parameter estimates is computed using bootstrapping methods with 500 replications (Efron, 1979). Deaton (1997) shows that in the presence of heteroskedasticity, the asymptotic formula for this matrix derived by Koenker and Basset (1982) performs very poorly, underestimating the standard errors. It is very rare to find regression functions estimated from household survey data that are homoskedastic.

Our preliminary estimates for four models appear in Table 2. Because we are using their data for the time being, the specification of these models is similar to those of Anderson et al. (2003a). The first set of columns represents estimates of the determinants of underweight. This is the quantile regression for the fifth percentile of the BMIZ distribution. This is followed by OLS estimates, which represent the standard approach of estimating the determinants of the average BMIZ. The two right-hand sets of columns are estimates for the quantile regressions through the 85th and the 95th percentiles, describing the determinants of weight for age for those at the thresholds of being at risk of overweight and overweight, respectively.

Table 2. Quantile Regressions of BMI Z-scores for Young Children in the United States

	5th Percentile		OLS		85th Percentile		95th Percentile	
Independent Variables	Coeff.	z-stat	Coeff.	t-stat	Coeff.	z-stat	Coeff.	z-stat
Avg hours per week working since birth - units of 10	0.060	3.09 **	0.059	6.65 **	0.028	2.75 **	0.017	1.37
Number of weeks worked since birth	0.007	0.62	-0.002	-0.43	0.007	1.03	0.003	0.39
African American dummy	0.041	0.64	0.045	1.25	0.072	1.82 +	0.116	3.20 **
Hispanic dummy	0.035	0.48	0.117	2.51 *	0.154	3.18 **	0.156	3.64 **
Mother's highest grade completed	0.009	1.37	0.006	2.01 *	-0.005	-1.30	-0.009	-2.04 *
Mother's AFQT Score	0.001	0.35	-0.001	-1.44	-0.002	-2.73 **	-0.002	-3.04 **
Child was first born	-0.032	-0.59	-0.022	-0.80	-0.015	-0.44	0.016	0.48
Number of children	-0.050	-1.99 *	-0.053	-4.29 **	-0.082	-5.26 **	-0.074	-6.00 **
Child was breastfed	-0.042	-0.87	-0.089	-3.65 **	-0.091	-2.89 **	-0.113	-3.50 **
Mother's BMI > 25 (overweight or obese)	0.266	5.45 **	0.266	9.85 **	0.275	7.81 **	0.264	7.91 **
Mother's BMI > 30 (obese)	0.214	2.90 **	0.193	5.45 **	0.330	7.29 **	0.288	6.90 **
Average family income since birth - units of $10,000	0.015	1.05	-0.001	-0.12	0.000	0.03	-0.005	-0.51
Percent of child's life mother was married	-0.044	-0.62	-0.106	-2.87 **	-0.096	-2.31 *	-0.074	-1.53
Mother reported child's height	0.253	2.51 *	0.410	9.49 **	0.579	8.64 **	0.550	9.69 **
Mother reported child's weight	-0.202	-2.29 *	0.099	2.69 **	0.079	1.42	0.073	1.66 +
Child's birthweight in pounds	0.174	9.41 **	0.099	11.10 **	0.067	5.39 **	0.038	3.07 **
Child's age in years	0.136	10.97 **	0.037	6.00 **	-0.012	-1.58	-0.047	-6.00 **
Child is female	0.025	0.58	0.015	0.69	0.006	0.24	-0.031	-0.98
Mother's age in years	0.021	1.73 +	0.025	4.38 **	0.020	2.58 **	0.017	2.01 *
1988 dummy	-0.008	-0.10	0.052	1.24	0.073	1.72 +	0.119	2.31 *
1990 dummy	-0.232	-2.71 **	-0.062	-1.38	0.102	2.08 *	0.190	3.26 **
1992 dummy	-0.013	-0.15	0.091	1.88 +	0.173	3.22 **	0.261	3.91 **
1994 dummy	-0.905	-8.11 **	-0.463	-8.54 **	-0.090	-1.25	0.060	0.71
1996 dummy	-0.513	-3.95 **	-0.311	-4.99 **	-0.024	-0.30	0.195	1.83 +
Constant	-4.525	-11.54 **	-1.711	-9.93 **	0.006	0.03	1.133	4.90 **
Number of observations	15,351		15,351		15,351		15,351	
Pseudo R^2	0.05		0.06		0.05		0.05	

Source: Authors' calculations from NLSY79

For each of the models, the R^2's are low at about 0.05. As noted by Chou et al. (2002), this is not surprising given the large genetic components that affect weight levels. Further, a number of potential explanatory variables are absent from these exploratory estimates (e.g., time spent watching television, and food prices). Nonetheless, several interesting results do emerge suggesting the need to further explore quantile regression analysis for nutrition models. We highlight several of them here.

The education level of a child's mother, for example, has differential effects depending on where the child is in the BMIZ distribution. Whereas, the OLS estimates suggest that the higher the level of mother's education, the heavier the child will be, the opposite is the case for those children who are just obese. Using standard OLS methods to understand overweight would have been misleading in this case. This would also be so for underweight, where we find no statistical evidence that the mothers' education levels have any independent effect on child nutrition.

The number of hours that the child's mother works contributes positively to the child's weights for underweight, average and at risk children, but does not have a significant effect for those children who are obese. So, while standard OLS models would predict that less time devoted to childcare because mothers are working would contribute to greater weight gain, we find that it has no statistical effect on the degree of overweight.

Although we find that a child's birthweight is a good predictor of current weight, the effect is much stronger at the lower end of the distribution. In other words, we find evidence that the persistence of low birthweights leading to underweight as a child is much greater than the persistence of high birthweights leading to overweight as a child. Interestingly, Abrevaya (2001) finds that certain factors such as race, education and prenatal care have stronger effects at the lower end of the distribution of birthweights. Our preliminary analysis suggests that these effects persist in the form of low birthweights increasing the likelihood of child underweight. Of course, there are other factors that contribute to child underweight beyond low birthweights. Breastfeeding, however, does not appear to be one of them. In our sample, breastfeeding only has the expected negative impact on weight for average and overweight children. We find no effect of breastfeeding on the weights of underweight children. There is one important caveat here however. Breastfeeding is likely to be endogenous since it is correlated with the unobserved behavior of the mother. As such, these initial estimates need to be treated with care.

With everything else held constant, our estimates suggest a convergence toward levels of healthy nutrition as children get older. This is something that we would not have discerned using OLS methods. For underweight children, there is a strong and positive effect of the child's age on his/her weight for age.[9] Conversely, there is a negative effect of age on weight for those at the overweight end of the distribution. As mentioned in footnote 5, we would expect this effect to bias estimates of the malnutrition trends in our sample because the distribution of children gets progressively older. Indeed, *ceteris paribus*, the changing age distribution in the sample would lead to declining levels of underweight and overweight. As this is not what we see in the data, other factors must be leading to the persistence of underweight and to the increases in overweight that we observe.

These results, along with others such as African American and Hispanic children being more likely to be overweight but not underweight, give an indication that the point on the distribution of nutrition outcomes where the estimation takes place matters for understanding the determinants of malnutrition. This motivates us to further study the determinants of malnutrition empirically using quantile regression methods. We plan to extend this preliminary analysis in two ways.

---

[9] As a reminder, we note that these models explain deviations of BMI values from those of the healthy reference population. Thus, the impact of age estimated here is independent of the age effect that is present in the reference population.

First, the recent release of the 2002 NLSY79 data will permit us to expand our dataset so that we can document more recent trends in malnutrition in this sample and apply our methods to estimate their determinants. Second, we will concentrate on alternative specifications of the model starting with a clarification of the underlying theory. In particular, we will further pursue some of the explanations in the literature for rising overweight that we discussed previously – lack of exercise, increased women's labor force participation, greater access to cheaper fast foods, and increased consumption of sugar-sweetened beverages. This will include collecting external data that can be merged into the NLSY79. Such data that might help us address some of these causes include fast food prices from the American Chamber of Commerce Research Association (ACCRA, various years) to capture the effect of the price of fast foods, data from the Census of Retail Trade (Bureau of the Census, various years) on the number of fast-food restaurants per capita to capture the increased accessibility to fast-foods,[10] and state laws regarding physical education requirements as a proxy for exercise and calorie expenditure. The NLSY79 also has information on the average time that a child spends watching television. We plan to use these data as a proxy for sedentary lifestyle, but will remain cautious given their potential endogeneity.

5. Estimating the Effects of Malnutrition on Cognitive, Behavioral and Social Development

The second part of our proposed research involves estimating the effects of malnutrition on cognitive ability, behavior and self-esteem. One often-cited concern about undernutrition in children is that it may have negative consequences for cognitive development presumably because a lack of food deprives the brain of essential nutrients. Although this is generally an issue in the developing world, there is also a fairly sizeable literature on this topic in the medical field for the United States. Corman and Chaikind (1993), for example, find that low-birthweight children score lower on tests of academic performance. Alaimo et al. (2001) report that children aged 6 to 11 in food-insecure households scored lower on arithmetic tests, were more likely to have repeated a grade and to have seen a psychologist, and had difficulty getting along with other children. Winicki and Jemison (2003) also find that food insecurity negatively impacts the academic performance of kindergartners. Weinreib et al. (2002) report that severe child hunger is correlated with a greater incidence of behavior problems and a greater level of reported anxiety/depression. There is evidence that programs such as providing breakfast to school age children have been effective in mitigating these consequences (Murphy et al., 1998). This has become such a strongly held view that some schools purportedly manipulated the nutritional content of their lunches to improve their test scores (Figlio and Winicki, 2002).

At the other end of the weight distribution, there are concerns that overweight children may suffer from low self-esteem and that low self-esteem may lead to lower academic performance. For adults, it has been demonstrated that obese women have lower self-esteem than their non-obese counterparts (Averett and Korenman, 1999, 1996). This also appears to be the case for children (Eisenberg et al., 2003), with the effect increasing with age (Strauss, 2000). Furthermore, overweight children are more likely to be more socially isolated compared to adolescents who are not overweight (Strauss and Pollack, 2003). There is also evidence that overweight children have lower academic performance (Datar et al., 2004), and are more likely to have behavior problems (Datar and Sturm, 2004), to act as bullies, and to be bullied (Janssen et al., 2004). The social functioning of overweight children is likely to be reduced so much that Schwimmer et al. (2003) compare their qualities of life to those of children with cancer.

In our view, most of the previous research on children, weight and academic performance has not adequately addressed the issue of causality versus correlation. Interestingly, this is not true of the

[10] Chou et al., 2002 also use information on fast food availability and prices from these sources in their investigation of the causes of adult obesity.

research that examines the consequences of adult obesity where sophisticated econometric techniques have been used to determine causality (Averett and Korenman, 1996; Cawley, 2004). In an ideal world, an experiment could be run where some children are randomly "assigned" to be overweight, underweight, or well nourished. If the assignments were truly random and children who were either over or underweight performed lower on tests of cognitive ability, we would be confident that it was their nutritional status that caused the relatively poor performance. Of course, such an experiment is not feasible.[11] Indeed, it is quite plausible that being overweight in and of itself does not directly lead to lower academic achievement. Rather, depression, low-self esteem and other behavior issues that stem from being overweight could be what lead to lower academic achievement. It is also possible that the opposite is the case. In other words, because they may be depressed and/or isolated socially, overweight children may work harder in school and perform better. Depression itself may be a cause of obesity (Goodman and Whitaker, 2002). Finally, another line of reasoning is that parental behavior explains both malnutrition and cognitive functioning. Datar et al. (2004) found that overweight kindergartners were more likely to come from poor families in which the parents did not read to their children or encourage good academic performance. This makes it difficult to determine if being overweight is truly the cause of the poor academic performance, or if poor parenting or some other factor is the cause of both the overweight and the poor academic performance.

In our research we propose to study whether children who are malnourished have lower self-esteem, perform poorer on tests of academic ability, or are more likely to have behavioral problems than well-nourished children. In this analysis, we emphasize the use of econometric techniques which will help us determine if malnutrition has causal effects on the outcomes we describe below. The NLSY79 data are ideally suited for such an investigation because various developmental measures are available for the children.[12] These child assessments have been administered bi-ennally since 1986 (Baker and Mott, 1989; Chase-Lansdale et al., 1991). We propose to use two measures of cognitive development: (1) the Peabody Picture Vocabulary Test (PPVT) (Dunn and Dunn, 1981), administered to children ages 3 and older, and (2) an average of the math and reading recognition scores from the Peabody Individual Achievement Tests (PIAT) (Dunn and Barkwardt, Jr., 1970), administered to children ages 5 and over. The PPVT provides an estimate of the child's receptive vocabulary and verbal ability. The PIAT is among the most widely used brief assessments of academic achievement. The PIAT Mathematics assessment begins with early skills (recognizing numerals) and progresses to measuring more advanced concepts. The Reading Recognition assessment measures word recognition and pronunciation ability, which are considered essential components of reading achievement. For both the PPVT and PIAT, we use the standardized score which has a mean of 100 and a standard deviation of 15. Reliability of the PPVT is quite high; Dunn and Dunn (1981) report a median split-half reliability of 0.80. Test-retest reliability for the Reading Recognition test is 0.89 for children from kindergarten through twelfth grade (Baker et al., 1993). The one month test-retest reliability for the PIAT mathematics assessment is 0.74, with lower levels of reliability for children in the lower grades (Dunn and Markwardt, Jr., 1970).

In addition to these two high quality and often used cognitive measures, we will also use the Behavior Problems Index (BPI), a measure that was administered to children age 4 and older and derived from the Child Behavior Checklist and other child behavior scales (Achenback and Edelbrock, 1981; Peterson and Zill, 1986). For this measure, parents report the frequency with which a child

---

[11] Interestingly, some experimental studies similar in design to this have been carried out and have found that students who fasted before school scored lower on tests of cognitive ability (Pollitt et al. 1998).

[12] These data have been used extensively to examine the effect of maternal employment on cognitive ability (Ruhm, 2004) and the effect of paternal child care on children's cognitive ability (Averett et al., forthcoming).

exhibited 28 specific problems. Responses to individual items ("often true," "sometimes true," and "not true") were summed to produce an index score for each child. Using the Spearman-Brown formula to estimate reliability, an r of 0.92 was obtained for the 28 item BPI (Baker et al., 1993).

Finally, to assess self-esteem, we will use the Self-Perception Profile for Children. This profile can be used to construct a score of general self-esteem in addition to scores in perceived physical appearance, scholastic competence, behavioral conduct, social acceptance, and athletic competence. The scale is self-administered and Harter (1995) reports an internal reliability of 0.8.

In this part of the analysis, the child's BMIZ now becomes the explanatory variable of interest in models estimating the determinants of self-esteem, achievement and behavior problem scores. As we noted in the first part of this proposal, one of our main concerns is that the determinants of BMIZ may be very different depending upon where in the distribution a child falls. The same is likely true here. For example, a severely overweight child is potentially more likely to suffer from low-self esteem than one who is only a little bit above the recommended weight for height. Thus, a simple OLS estimate of the relationship between BMIZ and our outcomes is going to give us information about how BMIZ affects these outcomes on average. Furthermore, using BMIZ as a continuous variable confines the effect of changes in BMIZ on our outcomes to be linear—i.e., each unit increase in the BMIZ changes the outcomes the same amount. However, it is unlikely that this relationship is linear. We propose to use the malnutrition gap (described earlier) as our independent variable that measures BMIZ. This gap will be zero for those who are in the recommended range. This would capture the effect of the degree to which children are malnourished, not just if they are malnourished. Furthermore, this would allow for differences between a 10 percent change in BMIZ at the mean and a 10 percent change in BMIZ for a child who is already overweight (or underweight).

As discussed in the introduction, BMI is potentially endogenous. We propose several empirical methods for dealing with this endogeneity. First, we will use a wide array of control variables (including parental education, race, ethnicity, child's birthweight, mother's cognitive ability, number of siblings, whether or not the parent works).

Another approach we will use is to instrument the child's BMI. However, a suitable instrument must be highly correlated with the child's BMI but uncorrelated with the outcome of interest. An obvious predictor for a child's BMI is his/her mother's BMI since it is well-documented that overweight children are likely to have obese parents. However, this is not likely to be a good instrument because choices that mothers make regarding food and exercise are also likely passed down to the child and affect the outcomes. We plan to make use of the data on fast food outlets and prices, and physical fitness requirements as instruments to help us identify variation in malnutrition that is not correlated with cognitive functioning or behavioral functioning. These factors should be determinants of a child's nutritional status but be unrelated to an individual child's cognitive ability, behavior, or self-esteem.

Our research on the effects of BMI on cognitive ability, behavior, and self-esteem improves on past research in several ways. First, many previous researchers did not use nationally representative samples, thus limiting the degree to which their results can be generalized. Second, previous researchers often focused on either underweight or overweight rather than both extremes of the distribution. Third, we plan to examine the effects of BMI at various points in the distribution rather than as just an indicator of over or underweight by using the malnutrition gap as our measure of malnutrition. Finally, we will examine to what extent, if any, BMI *causes* low cognitive ability, behavior problems and low self-esteem.